



AKDENİZ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ



Abdullah AKGÜN

VERİ MADENCİLİĞİ SINIFLANDIRMA YÖNTEMLERİNİN ETKİNLİĞİNİN
KARŞILAŞTIRILMALI OLARAK İNCELENMESİ: SEYAHAT ACENTASI ÖRNEĞİ

Turizm İşletmeciliği ve Otelcilik Ana Bilim Dalı
Doktora Tezi

Antalya, 2019



AKDENİZ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ



Abdullah AKGÜN

VERİ MADENCİLİĞİ SINIFLANDIRMA YÖNTEMLERİNİN ETKİNLİĞİNİN
KARŞILAŞTIRILMALI OLARAK İNCELENMESİ: SEYAHAT ACENTASI ÖRNEĞİ

Danışman

Prof. Dr. Beykan ÇİZEL

Turizm İşletmeciliği ve Otelcilik Ana Bilim Dalı

Doktora Tezi

Antalya, 2019

T.C.
Akdeniz Üniversitesi
Sosyal Bilimler Enstitüsü Müdürlüğüne,

Abdullah AKGÜN'in bu çalışması, jürimiz tarafından Turizm İşletmeciliği ve Otelcilik Ana Bilim Dalı Doktora Programı tezi olarak kabul edilmiştir.

Başkan	: Prof. Dr. Melih GÜNAY	(İmza)
Üye (Danışmanı)	: Prof. Dr. Beykan ÇİZEL	(İmza)
Üye	: Prof. Dr. Cem Oktay GÜZELLER	(İmza)
Üye	: Doç Dr. Murat ÇUHADAR	(İmza)
Üye	: Dr. Öğr. Üyesi Ceylan ALKAN	(İmza)

Tez Başlığı:

VERİ MADENCİLİĞİ SINIFLANDIRMA YÖNTEMLERİNİN ETKİNLİĞİNİN
KARŞILAŞTIRILMALI OLARAK İNCELENMESİ: SEYAHAT ACENTASI ÖRNEĞİ

Onay: Yukarıdaki imzaların, adı geçen öğretim üyelerine ait olduğunu onaylarım.

Tez Savunma Tarihi : 14 / 06 / 2019

Mezuniyet Tarihi : 25/07/2019

(İmza)

Prof. Dr. İhsan BULUT

Müdür

AKADEMİK BEYAN

Doktora Tezi olarak sunduđum “Veri Madenciliđi Sınıflandırma Yöntemlerinin Etkinliđinin Karşılaştırılmalı Olarak İncelenmesi: Seyahat Acentası Örneđi ” adlı bu çalışmanın, akademik kural ve etik deđerlere uygun bir biçimde tarafımda yazıldıđını, yararlandıđım bütün eserlerin kaynakçada gösterildiđini ve çalışma içerisinde bu eserlere atıf yapıldıđını belirtir; bunu şerefimle dođrularım.

09 / 07 / 2019

İmza

Abdullah AKGÜN





T.C.
AKDENİZ ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
TEZ ÇALIŞMASI ORJİNALLİK RAPORU
BEYAN BELGESİ



SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜ'NE

ÖĞRENCİ BİLGİLERİ	
Adı-SOYADI	Abdullah AKGÜN
Öğrenci Numarası	20135204019
Enstitü Ana Bilim Dalı	Turizm İşletmeciliği ve Otelcilik ABD
Programı	Turizm İşletmeciliği ve Otelcilik
Programın Türü	() Tezli Yüksek Lisans (X) Doktora
Danışmanın Unvanı, Adı-SOYADI	Prof. Dr. Beykan ÇİZEL
Tez Başlığı	Veri Madenciliği Sınıflandırma Yöntemlerinin Etkinliğinin Karşılaştırılmalı Olarak İncelenmesi: Seyahat Acentası Örneği
Turnitin Ödev Numarası	1150135963

Yukarıda başlığı belirtilen tez çalışmasının a) Kapak sayfası, b) Giriş, c) Ana Bölümler ve d) Sonuç kısımlarından oluşan toplam 237 sayfalık kısmına ilişkin olarak, 09 /07/2019 tarihinde tarafımdan Turnitin adlı intihal tespit programından Sosyal Bilimler Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nda belirlenen filtrelemeler uygulanarak alınmış olan ve ekte sunulan rapora göre, tezin/dönem projesinin benzerlik oranı;

alıntılar hariç % 7.

alıntılar dahil % 14 'tür.

Danışman tarafından uygun olan seçenek işaretlenmelidir:

(X) Benzerlik oranları belirlenen limitleri aşmıyor ise;

Yukarıda yer alan beyanın ve ekte sunulan Tez Çalışması Orijinallik Raporu'nun doğruluğunu onaylarım.

() Benzerlik oranları belirlenen limitleri aşıyor, ancak tez/dönem projesi danışmanı intihal yapılmadığı kanısında ise;

Yukarıda yer alan beyanın ve ekte sunulan Tez Çalışması Orijinallik Raporu'nun doğruluğunu onaylar ve Uygulama Esasları'nda öngörülen yüzdelik sınırlarının aşılmasına karşın, aşağıda belirtilen gerekçe ile intihal yapılmadığı kanısında olduğumu beyan ederim.

Gerekçe:

Benzerlik taraması yukarıda verilen ölçütlerin ışığı altında tarafımda yapılmıştır. İlgili tezin orijinallik raporunun uygun olduğunu beyan ederim. 09 / 07 / 2019

Prof. Dr. Beykan ÇİZEL

İÇİNDEKİLER

İÇİNDEKİLER	v
ŞEKİLLER LİSTESİ	viii
TABLolar LİSTESİ	ix
KISALTMALAR	xi
ÖZET	xii
SUMMARY	xiii
GİRİŞ	1

BİRİNCİ BÖLÜM VERİ MADENCİLİĞİ

1.1. Veri İle İlgili Kavramlar	8
1.1.1. Veri, Enformasyon ve Bilgi	8
1.1.2. Enformasyon Sistemleri	9
1.1.3. Veri Tabanları	11
1.1.4. Veri Ambarları	12
1.2. Veri Madenciliği.....	14
1.3. Veri Madenciliği Süreci	18
1.3.1. Problemin Tanımlanması	20
1.3.2. Verinin Anlaşılması	21
1.3.3. Verilerin Hazırlanması	22
1.3.3. Modelin Kurulması	24
1.3.4. Modelin Değerlendirilmesi	24
1.4.2. Birliktelik Kuralları	28
1.3.5. Modelin Kullanılması ve İzlenmesi	32
1.4. Veri Madenciliği Yöntemleri.....	32
1.4.1. Kümeleme	33
1.4.3. Sınıflandırma.....	34

İKİNCİ BÖLÜM SINIFLANDIRMA YÖNTEMLERİ

2.1. Karar Ağaçları	36
2.1.1. Turizmde Karar Ağaçları	39
2.1.2. Karar Ağacı Algoritmaları	41
2.1.2.1. ID3 (iteratif dikotomizer) Algoritması	42
2.1.2.2. C4.5 ve C 5 Algoritması.....	43
2.1.2.3. CART veya C&RT (Sınıflandırma ve Regresyon Ağaçları) Algoritması.....	43
2.1.2.4. CHAID (Otomatik Ki-kare Etkileşim) Algoritması	44
2.1.2.5. Diğer Karar Ağacı Algoritmaları.....	44

2.1.3. Örnek Uygulama	45
2.2. Bayes Ağları	49
2.2.1. Turizmde Bayes Ağları	50
2.2.2. Örnek Uygulama	53
2.3. Yapay Sinir Ağları	54
2.4. En Yakın Komşu	57
2.5. Destek Vektör Makineleri	58
2.6. Kaba Kümeler	59
2.7. Lojistik Regresyon	62
2.8. Genetik Algoritmalar	62

ÜÇÜNCÜ BÖLÜM **TURİZMDE VERİ VE VERİ KAYNAKLARI**

3.1. Anketler, Yüz Yüze Görüşmeler	69
3.2. Seyahat Günlükleri	71
3.3. Operasyonel Sistemler	72
3.4. MİY Sistemleri	75
3.5. Coğrafi Bilgi Sistemleri	76
3.6. Küresel Konumlandırma Sistemi	77
3.7. Web Siteleri	79
3.8. Sosyal Medya	85
3.9. Cep Telefonları	88
3.10. Kullanıcı Tarafından Oluşturulan İçerik	89
3.10.1. Kullanıcı Yorumları	91
3.10.2. Coğrafi Etiketli Öğeler	94
3.10.3. Kullanıcı Değerlendirmeleri	95
3.10.4. Kullanıcı Fotoğrafları	96
3.10.5. Kullanıcı Videoları	97
3.11. Veri Hazırlama	97

DÖRDÜNCÜ BÖLÜM **TURİZMDE VERİ MADENCİLİĞİ**

4.1. Müşteri İlişkileri Yönetimi	108
4.2. Ürün Yönetimi	114
4.3. Turist Memnuniyeti	118
4.4. Turist Tercihleri	120
4.5. Turist Davranışı	124
4.6. Öneri Sistemleri	129

4.7. Turist Tahmini	136
---------------------------	-----

BEŞİNCİ BÖLÜM ARAŞTIRMA

5.1. Araştırmanın Amacı	141
5.2. Kapsam ve Yöntem	142
5.3. Araştırma Süreci	143
5.4. Araştırmanın Problemleri	143
5.5. Verinin Toplanma Süreci	144
5.5.1. Veri Tabanının Oluşturulması	145
5.5.2. Verinin Birleştirilmesi ve Dönüştürülmesi	150
5.5.3. Değişken Seçimi.....	156
5.7. Verilerin Analiz Süreci	161
5.7.1. Modellerin Performans Analizleri.....	162
5.7.1.1. En Yakın Komşu	163
5.7.1.2. Karar Ağaçları	164
5.7.1.3. Bayes Sınıflandırma	168
5.7.1.4. Yapay Sinir Ağları.....	169
5.7.1.5. Regresyon	170
5.7.1.6. Lojistik Regresyon.....	170
5.7.1.7. RapidMiner Otomatik Model	171
5.7.2. Tur Bazlı Analizler	172
5.8. Araştırmanın Bulguları	174
5.4.1. Algoritmaların Performansları	175
5.4.2. Müşteri Profilleri	177
5.4.3. Bölgelere Göre Müşteri Profilleri	180
5.4.4. Turlara Göre Müşteri Profilleri	181
5.9. Tartışma ve Sonuç	184
KAYNAKÇA.....	192
Ek-1. Karar Ağacı Kuralları	211
Ek-2. Budanmış Karar Ağacı Kuralları.....	221
ÖZGEÇMİŞ	223

ŞEKİLLER LİSTESİ

Şekil 1.1 Veri Madenciliği Süreci	19
Şekil 1.2 Basit bir çakışım matrisi.....	26
Şekil 2.1 Transfer Satın Alma Konseptinde Bir Karar Ağacı	37
Şekil 2.2 Örnek Uygulama İçin Karar Ağacı	49
Şekil 2.3 Çekirdek İşlevi İle Verinin Dönüştürülmesi	59
Şekil 2.4 Kaba Küme Analizinin Süreç Haritası ve Ana Süreçleri	60
Şekil 4.1 Üç Ana Nicel Tahmin Metodu Kategorileri.....	137
Şekil 5.1 Araştırma Süreci.....	143
Şekil 5.2 Verinin Microsoft SQL Server'dan Microsoft Access Veritabanına Aktarılması ..	145
Şekil 5.3 Veritabanından Analiz İçin Seçilen Tablolar	146
Şekil 5.4 Müşteriler Tablosundaki Ham Veriler	153
Şekil 5.5 Analiz İçin Hazırlanan Müşteri Verisi	154
Şekil 5.6: Adım Adım Geriye Doğru Eleme Yöntemiyle Seçilen Değişkenler	158
Şekil 5.7 Sınıflandırma Modelleri Süreci	162
Şekil 5.8 K- En Yakın Komşu Algoritmasının Performansının Değerlendirilmesi	164
Şekil 5.9 Karar Ağaçları Modeli.....	164
Şekil 5.10 ID3 Algoritmasının Eğitim ve Test Süreci.....	165
Şekil 5.11 CHAID Algoritmasının Eğitim ve Test Süreci	166
Şekil 5.12 C4.5 Algoritmasının Eğitim ve Test Süreci	166
Şekil 5.13 Gradyan Arttırılmış Ağaçlar Algoritmasının Eğitim ve Test Süreci.....	167
Şekil 5.14 Rastgele Orman Algoritmasının Eğitim ve Test Süreci	167
Şekil 5.15 Rastgele Ağaç Algoritmasının Eğitim ve Test Süreci.....	168
Şekil 5.16 Naive Bayes Algoritması İçin Eğitim ve Test Verilerinin Düzenlenmesi	169
Şekil 5.17 Deep Learning Algoritması İçin Eğitim ve Test Verilerinin Düzenlenmesi.....	169
Şekil 5.18 Genelleştirilmiş Doğrusal Model İçin Eğitim ve Test Verilerinin Düzenlenmesi	170
Şekil 5.19 Lojistik Regresyon İçin Eğitim ve Test Verilerinin Düzenlenmesi	170
Şekil 5.20 Rapidminer Programının Auto Model Sonucu.....	171
Şekil 5.21 Günlük Turların Sınıf Değişkeni Olarak Belirlenmesi.....	173
Şekil 5.22 Günlük Turlara Göre Profil Belirleme Modelinin Süreci	173
Şekil 5.23 Günlük Turlara Göre Profil Belirleme Modelinin Eğitim ve Test Bölümü	174
Şekil 5.24 Karar Ağacı	178
Şekil 5.25 Bölgelere Göre Veri Filtreleme	180

TABLOLAR LİSTESİ

Tablo 1.1 Turizm Çalışmalarında Kullanılan Veri Madenciliği Tanımları.....	16
Tablo 1.2 Turistlerin Aldığı Günlük Turlar.....	30
Tablo 2.1 Karar Ağacı Algoritmalarının Kullanıldığı Çalışmalar.....	40
Tablo 2.2 Turistlerin Tur Satın Alma Durumu.....	45
Tablo 2.3 Gelir Düzeyinin "Yüksek" Olduğu Durumlar.....	47
Tablo 2.4 Gelir Düzeyi "Orta" için oluşturulan tablo.....	48
Tablo 2.5 Bayes Sınıflandırıcının Kullanıldığı Çalışmalar.....	52
Tablo 2.6 Bilgi Verilen Turistlerin Listesi.....	53
Tablo 2.7 Tur alma ile ilgili Tüm olasılıklar.....	53
Tablo 3.1 Anketler aracılığıyla Veri Toplayan Araştırmalar.....	70
Tablo 3.2 Operasyonel Sistemlerden Veri Toplayan Araştırmalar.....	74
Tablo 3.3 Web Sitelerinden Veri Toplayan Araştırmalar.....	84
Tablo 3.4 Sosyal Medyadan Veri Toplayan Bazı Araştırmalar.....	88
Tablo 3.5 Coğrafi Etiketli Öğeler Üzerine Çalışan Araştırmalar.....	95
Tablo 3.6 Fotoğraflar Üzerine Çalışan Araştırmalar.....	97
Tablo 4.1 MİY Konuları İçin Veri Madenciliği Tekniklerinin Kullanıldığı Araştırmalar.....	113
Tablo 4.2 Ürün Yönetimi İçin Veri Madenciliği Tekniklerinin Kullanıldığı Araştırmalar.....	117
Tablo 4.3 Turist Memnuniyeti İçin Veri Madenciliği Tekniklerinin Kullanıldığı Araştırmalar.....	120
Tablo 4.4 Turist Tercihleri İçin Veri Madenciliği Tekniklerinin Kullanıldığı Araştırmalar..	124
Tablo 4.5 Turist Davranışı için Veri Madenciliği Tekniklerinin Kullanıldığı Araştırmalar ..	128
Tablo 4.6 Ürün Önerileri İçin Veri Madenciliği Tekniklerinin Kullanıldığı Araştırmalar	134
Tablo 4.7 Tahmin İçin Veri Madenciliği Tekniklerinin Kullanıldığı Araştırmalar	138
Tablo 5.1 Seyahat Acentası Veri Tabanından Alınan Tablolar.....	146
Tablo 5.2 Rezervasyon Tablosundaki Alanlar ve İçerdiği Veri.....	147
Tablo 5.3 Müşteriler Tablosundaki Alanlar ve İçerdiği Veri.....	148
Tablo 5.4 Ürünler Tablosundaki Alanlar ve İçerdiği Veri.....	148
Tablo 5.5 Aksiyonlar Tablosundaki Alanlar ve İçerdiği Veri.....	149
Tablo 5.6 Biletler Tablosundaki Alanlar ve İçerdiği Veri.....	149
Tablo 5.7 Müşteri Tipi Değişkeni ve Düzeyleri.....	153
Tablo 5.8 Çocuk Bilgisi Değişkeninin İçeriğinin Düzenlemesi Şablonu.....	154
Tablo 5.9 Veri Setindeki Değişkenler.....	157
Tablo 5.10 Değişkenlerin Teker Teker Analize Alınması Durumundaki Model Performansı	159

Tablo 5.10 Değişkenler İkişer - İkişer Analize Alındığında Model Performansı	160
Tablo 5.12 Sınıflandırma Algoritmaları Karşılaştırması	176
Tablo 5.13 Tur Satın Alan Müşteri Grupları	179
Tablo 5.14 Tur Satın Almayan Müşteri Grupları	179
Tablo 5.16 Turlara Göre Müşteri Profilleri	182
Tablo 5.16 Tur Satışlarını Arttırmak İçin Yapılabilecekler	184



KISALTMALAR

AHP	: Analitik Hiyerarşi Süreci
BSM	: Temel Yapısal Yöntem
CART	: Sınıflandırma ve Regresyon Ağaçları
CBS	: Coğrafi Bilgi Sistemleri
CRISP-DM	: Veri Madenciliği için Sektörler Arası Standart Süreç
DMO	: Destinasyon Yönetim Organizasyonları
DT	: Karar Ağaçları
DVM	: Destek Vektör Makinesi
DW	: Veri Ambarı,
ET	: Enformasyon Teknolojileri
e WOM	: Elektronik Ağızdan Ağıza İletişim
FFBN	: İleri Besleme Geri Yayılım Ağları
GPS	: Küresel Konumlandırma Sistemi
GA	: Genetik Algoritma
KDS	: Karar Destek Sistemi
KOİ	: Kullanıcının Oluşturduğu İçerik
LDA	: Lineer Ayırt Edici Analiz
LR	: Lojistik Regresyonlar
KTSA	: Konum Tabanlı Sosyal Ağlar
MİY	: Müşteri İlişkileri Yönetimi
ÇHİ	: Çevrimiçi Hareket İşleme
ÇAİ	: Çevrimiçi Analitik İşleme
ÖS	: Öneri Sistemleri
PMS	: Tesis (Otel) Yönetim Sistemleri
POS	: Satış Noktası Sistemleri
PC	: Kişisel Bilgisayar
RBF	: Radyal Temelli Fonksiyon
RBFN	: Radyal Temelli Fonksiyon ağları
RO	: Rastgele Orman
SA	: Sinir Ağları
SOM	: Kendi Kendini Düzenleyen Haritalar
DVM	: Destek Vektör Makinesi
VYS	: Veri Tabanı Yönetim Sistemi
VM:	: Veri Madenciliği
VTBK	: Veri Tabanlarında Bilgi Keşfi
WOM	: Ağızdan Ağıza İletişim
URL	: Standart Kaynak Bulucu
YSA	: Yapay Sinir Ağları
YES	: Yönetim Enformasyon Sistemi
YÜG	: Yeni Ürün Geliştirme

ÖZET

Turizm sektöründe bir seyahatin karar aşamasından bitimine kadar, hem turistler açısından hem de işletmeler açısından veriye bağımlı bir süreç yaşanmaktadır. İşletmeler tüm işlemlerini sürecin en başından en sonuna kadar enformasyon teknolojilerini kullanarak takip etmektedirler. Turistlerde aynı şekilde, tatil yeri ile ilgili yapılan aramalardan tatil sonrası deneyim paylaşımına kadar sürecin tamamında enformasyon teknolojilerini kullanmaktadırlar. Bunun sonucunda işletmelere ait veri tabanlarında ve turizm ile ilgili web sitelerinde ve sosyal medyada bol miktarda veri birikmektedir.

Güçlü veri analizi araçlarına duyulan ihtiyaçla birleştirilen veri bolluğu, aranan bilgiye ulaşma konusunda bazı sıkıntılara neden olmaktadır. Bir turist için aradığı şeye ulaşamama, bir işletme için o kadar veri içinde bilgiden bağımsız kararlar verme durumuna yol açmaktadır. Rekabet üstünlüğü elde edebilmek için önemli olan veriye sahip olmak değil, o verilerden işletme için gerekli olan yararlı bilginin çıkartılmasıdır. Bu konuda veri madenciliği teknikleri işlemelerin imdadına yetişmektedir. Yapılan çalışmalar veri madenciliği tekniklerinin devasa veri yığınları arasından yararlı bilgiyi çıkarma konusunda güvenilir olduklarını ispatlamıştır.

Bu doktora tezinin amacı veri madenciliği yöntemlerinden sınıflandırma teknikleri kullanılarak günlük tur satın alan ve almayan müşterilerin profilinin çıkarılmasıdır. Bu amaç için Antalya’da faaliyet gösteren A sınıfı bir seyahat acentasının rezervasyon ve günlük tur ile ilgili işlemsel verilerinin kaydedildiği bir veri tabanındaki veriler kullanılmıştır. Rezervasyonu yapılan bir turistin günlük tur satın alma durumu “SATIN ALAN” veya “ALMAYAN”dır, bu da bir sınıflandırma problemidir. Dolayısıyla çalışmada veri madenciliğinin sınıflandırma fonksiyonuna ilişkin yöntemler kullanılmıştır. Araştırma sürecinde seyahat acentası veri tabanındaki farklı tablolarda yer alan rezervasyon, günlük tur bileti, ürünler ve müşteriler hakkındaki veriler bir tabloda birleştirilmiştir. Elde edilen veriler üzerinde 11 farklı sınıflandırma algoritması çalıştırılmıştır. Bu algoritmalarından en iyi performansı veren karar ağacı sınıflandırma tekniklerinden C4.5 algoritması veri içindeki müşteri profillerinin keşfi için kullanılmıştır.

Seyahat acentası veri tabanında müşteri ile ilgili çok fazla özellik olmamasına rağmen, yapılan analizde günlük tur satın alan müşterilere ait bilgilere ulaşılmıştır. Günlük turun satıldığı bölgeye göre tur satın alan grupların farklılaştığı ve en fazla yaş grubunun ve turist tipinin (aile, grup, tek vb.) tur satın alma ile ilgisinin olduğu tespit edilmiştir. Araştırma süreci ve bulgularının ilgili alanda çalışan araştırmacı ve uygulamacılara ışık tutacağı öngörülmektedir.

Anahtar kelimeler: Veri madenciliği, Sınıflandırma algoritmaları, Seyahat acentası veri tabanı, Müşteri profili keşfi

SUMMARY

**COMPARATIVE ANALYSIS OF THE EFFECTIVENESS OF DATA MINING
CLASSIFICATION METHODS: A CASE OF TRAVEL AGENCY**

In tourism sector, for both tourists and businesses, a data-dependent process is experienced from the decision-making stage to the end of a travel. Businesses do all their transactions by using information technologies from the beginning to the end. In the same way, tourists use information technologies throughout the process, from searching for holiday to sharing of experiences after holidays. As a result, large amounts of data are accumulated on the databases of the businesses, and websites and social media related to tourism.

The abundance of data, coupled with the need for powerful data analysis tools, causes some problems in reaching the searched information. It leads to inability to reach what a tourist is looking for and the situation of making knowledge-independent decisions in the abundance of data for an enterprise. In order to obtain competitive advantage, it is not important to have the data, but to discover useful information necessary for the business from the data. In this regard, data mining techniques deal with this problem. Studies have proven that data mining techniques are reliable in discovering useful information from huge amount of data.

The aim of this PhD dissertation is to profile customers who buy and don't buy daily tours using classification techniques from data mining methods. For this purpose, the data -in a database where the transactional data regarding the booking and daily tour are recorded- of A class travel agency operating in Antalya, were used. A booked tourist's daily tour purchase status is "BUYER" or "NON-BUYER", which is a classification problem. Therefore, methods related to the classification function of data mining were used in the study. During the research process, the data in different tables in the travel agency database about reservation, daily tour ticket, products and customers were aggregated in a table. 11 different classification algorithms were run on the preprocessed data. C4.5 algorithm from decision tree classification techniques, which gave the best performance and correct estimation rate among these algorithms, was used for the discovery of customer profiles.

Although there are not enough features related to the customer in the travel agency database, in the analysis, some profiles related to the customers who bought the daily tour were reached. It was determined that groups buying the tours according to the region where the daily tour was sold differed, and age group and tourist type (family, group, etc.) were the most related feature for buying daily tours. The research process and findings are expected to show the way to researchers and practitioners working in the relevant field.

Key words: Data mining, Classification Algorithms, Travel agency database, Customer profile discovery

GİRİŞ

Turizm, ülke ve bölgelerin ekonomisine yaptığı katkı ile dünya ekonomisinin en hızlı büyüyen sektörlerinden biri haline gelmiştir. Sektörde destinasyon ve işletmeler pazar paylarını arttırabilmek için birbirleri ile kıyasıya bir mücadelenin içine girmiş durumdadır. Günümüzde bu mücadelede yeni iletişim ve enformasyon teknolojisi cihazları, bulut bilişim, yapay zekâ ve otomasyonlar işletmelere rekabet etme gücü vermektedir. Bu teknolojilerin başarısı için de işletmenin alanı ile ilgili veriye sahip olması gerekmektedir.

Verinin en önemli güç olduğu bu rekabet ortamında, herhangi bir konuda tüketicinin temel istek veya duygularını yansıtan verilere sahip olan ve bunun üzerine akıllı uygulamalar geliştiren şirketler başarılı olabilecektir. Turizmden aldığı payı arttırmak isteyen destinasyon, otel ve seyahat acentaları, daha fazla turiste ulaşabilmek, ulaştığı turistleri memnun edebilmek için elindeki teknolojiyi etkin ve verimli kullanmak zorundadır. Turistlere dikkatlerini çekebilecek ürünler sunabilmek için turistlerin istek ve ihtiyaçlarını göz önünde bulundurarak ve bu ihtiyaçları karşılayabilmeli gerekirse kişiselleştirilmiş yeni ürünler geliştirerek pazardaki yerlerini korumalıdır. Tüm bunları yapabilmek için de turist hakkında detaylı veriye sahip olmak gerekmektedir.

Veri bolluğu açısından turizm sektörü en zengin sektörlerden birisidir. Geleneksel veri toplama yöntemi olan anketlere ek olarak sektörde işletmelerde yaygın olarak kullanılan enformasyon sistemleri ile birlikte sosyal medyada devasa veri yığınları oluşmaktadır. Seyahat kararı veren bir turist ilk yapacağı işlemlerden birisi internetten gezilecek yerler hakkında bilgi toplamaktır (Puri vd., 2017). Turistlerin gezdiği sayfa ve baktığı resim ilgili sistemlerin veri tabanlarında kayıt altına alınmaktadır ve bu kayıtlar turizm yöneticilerine turistlerin ilgilendiği ürün ve hizmetler hakkında fikirler vermektedir. Turistler bir bölge ve otelde karar verdiği zaman ister klasik seyahat acentası isterse çevrimiçi seyahat acentası ile rezervasyon yaptırın, milliyet, adres, belki tatil için ayırdığı bütçe, ödeme şekli vb. bilgiler ilgili sistemlerde kayıt altına alınmaktadır. Üstelik bu veriler tur operatöründen, uçuş firması, karşılayıcı seyahat acentası, konaklayacağı otel ve turlar için hizmet alınan işletmelere kadar dağıtılmakta ve her işletme aynı turist için daha detaylı bilgileri sistemlerine dâhil etmektedir.

İşletmelerin bünyesinde toplanan verilerden çok daha fazlası sosyal medya platformlarında toplanmaktadır. Turizm, teknolojik ilerlemeleri benimsemiş ve turistlerin çok farklı deneyimler arasından daha iyi bir seçim yapabilmeleri için sosyal medyada güçlü bir konuma sahip olan ilk sektörlerden birisidir (Nave vd., 2018: 694). Turistler genellikle satın aldıkları ürün ve hizmetler hakkında diğer (potansiyel) turistlerle etkileşime girmektedir. Turistler ve

potansiyel turistler arasındaki etkileşim müşteri satın alma ve geri alım kararlarını, hatta turizm işletmelerinin performansını da etkilemektedir. Bir işletme müşterilerin davranışlarını ve bu tür davranışlar için olası tetikleyicileri anlayabilmesi durumunda müşterileri elde tutabilmektedir. Tüketici davranışını anlama ve modelleme ile ilgili sosyal bilim ve pazarlamada çeşitli teoriler vardır. Tüketici davranışının bir ürün veya hizmet üzerindeki tercihini değiştirmesi sonucunda müşteri davranışı değişebilir. Tüketicilerin tercihlerini değiştirmesinin arkasındaki konsepti anlamak için bir takım yaklaşımlar uygulanmıştır. Bu yaklaşımlardan birisi de veri madenciliği tekniklerinin kullanılmasıdır.

Turizm alanında başlangıçta segmentasyon, pazarlama ve tahmin konularında (Bach vd., 2013) veri madenciliği teknikleri kullanılmıştır. Bu dönem turizm araştırmacı ve yöneticilerinin veri madenciliği tekniklerini tanımaları (örneğin; Law, 1998; Lau vd., 2001; Olmeda ve Sheldon, 2001) konusundaki çabalara sahne olmuştur. Pazarlama konusunda ağırlık olarak web sitesi pazarlama (Lau vd., 2001; Olmeda ve Sheldon, 2001) ve veri tabanı pazarlama (Cho ve Leung, 2002; Ha ve Park, 1998) üzerinde durulmuştur. Bir bölgeye olan seyahat talebinin tahmin edilebilmesi için (Cho, 2003; Goh ve Law, 2003) yapay sinir ağları ve kaba kümeler gibi veri madenciliği teknikleri kullanılmıştır. İlk dönemler araştırmacıların yüz yüze görüşme, anket ve veri tabanlarından elde ettikleri verilere, özellikle Web 2.0'nin tanıtılmasından sonra web siteleri ve sosyal medyada yayınlanan kullanıcı yorumları ve fotoğrafları, GPS, cep telefonu, vb. teknolojilerden elde edilen veriler eklenmiştir. Bu da araştırmacıların daha fazla konuda (öneri sistemleri gibi) veri madenciliği tekniklerini kullanmalarına olanak sağlamıştır.

2014 ve sonrası yapılan araştırmalarda yoğun bir şekilde kullanıcıların oluşturduğu içerikler (KOİ) üzerinde çalışmalar yapılmaktadır. Bunun en büyük nedenlerinden birisi işletmeler tarafından saklanan verilere ulaşım sorunu KOİ için ortadan kalkmıştır. Son dönemlerde turistler seyahat öncesi bilgi araştırmak, ürün ve hizmetleri satın almak ve seyahat sonrası deneyimlerini paylaşmak için web sitelerini ve sosyal medyayı daha yoğun olarak kullanmaya başlamışlardır. Turist davranışı, tercihi, memnuniyeti gibi alanlarda özellikle sosyal medya üzerinde yoğun bir veri yığılması olmuştur. Bu da özellikle 2018'den sonraki yapılan araştırmalarda çok yoğun bir artışa neden olmuştur. KOİ'nin yaygınlaşması ve kullanıcılar tarafından yazılan yorumların turist tercihleri ve memnuniyetlerinin en büyük göstergeleri haline gelmesi metin madenciliğinin önünü açmıştır. Son dönem çalışmalarının çoğunda metin madenciliği en çok tercih edilen yöntemlerden birisi olmuştur.

Turizmde veri madenciliği yöntemlerinin kullanıldığı alanlardan birisi de öneri sistemleridir. Figueredo vd. (2018) turizm alanında önerilerde bulunabilmek için kullanılan bir teknik olan işbirlikçi filtreleme yaklaşımını geliştirmişlerdir. Zhu vd. (2017) veri madenciliği

tekniklerini kullanarak yeni bir seyahat ürünü öneri motoru geliştirmişlerdir. Leal, vd. (2017) makine öğrenimi ve veri madenciliği yöntemlerinin turist kitlesi kaynaklı verilere dayanarak geçici seyahat aşamaları önereceğini savunmaktadırlar. Alt yapısında veri madenciliği teknikleri kullanılan öneri sistemleri bir destinasyonun (Dietz vd., 2018; Leal dv., 2017; Zheng vd., 2018), otelin (Nilsashi vd., 2015; Zisos vd., 2018;), turistik ürünlerin (Zhang ve PIRAMUTHU, 2018; Zhu vd., 2017), seyahat güzergâhının (Hui vd., 2017; Guo vd., 2017; Wan vd., 2017) potansiyel turistlere ya da kişiye özel (Sun vd., 2018; Liu vd., 2011; Kesorn vd., 2018) önerilmesinde kullanılmaktadır.

Turizm yöneticilerinin, destinasyon müşterilerini daha uygun ürünler tasarlamak ve pazarlama stratejilerine odaklanabilmeleri için tanımlamaları çok önemlidir. Bu görev, farklı gruplar için derlenecek tam tercih profillerini gerektirmektedir. Ancak, az sayıda çalışma, müşterileri veya müşterilerin profillerini etkileyen faktörleri değerlendirmiştir ve hiçbir çalışma henüz tam bir tercih profili sağlayamamıştır (Li vd., 2013: 322). Bu, turist ürünlerinin ve tercih eden turistlerin kendine özgün özelliklerinin olmasından kaynaklanabilir. Müşterileri tanıma konusunda işletmelerin operasyonel sistemlerinde (Talón-Ballestero vd., 2018-) ve çevrimiçi site ve sosyal medyada yeterince veri bulunmaktadır. Hatta, bazı araştırmacılar (Valls vd., 2018; Chiang vd., 2018) geleneksel anket yöntemiyle ilgili verileri toplamışlardır.

Günümüz turizm işletmeleri tercih edilen olmak için birbirleri ile rekabet halindedir. Turistler bir ürünü tercih ederken genel olarak o ürünü daha önce deneyimleyenlerin görüş ve önerilerine büyük önem vermektedirler. Sosyal ağ platformlarının internetteki kullanımının artmasıyla, turist piyasasının eğilimini izlemek amacıyla sosyal medya verisi kullanmak mümkündür. Sosyal medya veri analizi, e-turizm çağında değerlendirme sonuçlarının doğruluğunu ve kesinliğini büyük ölçüde artırmıştır (Sun vd., 2017). Sosyal medya üzerinde paylaşılan yorumlar ve coğrafi etiketli içerikler (Hausmann vd., 2018; Shi vd., 2017; Zhang vd., 2018) turist tercihlerinin ve turist memnuniyetinin belirlenmesinde araştırmacılar tarafından güvenilir olarak kullanılmaktadır. Dijital platformlarda yayınlanan olumlu değerlendirmeler müşteri memnuniyetini gösterirken negatif değerlendirmeler müşteri memnuniyetsizliğini göstermektedir (Xu, 2017: 1663). Bazı araştırmacılar (örneğin: Zhao vd., 2019; Xu, 2017) genel müşteri memnuniyetini ölçmek için çevrimiçi müşteri yorumlarını kullanırken, bazıları da (örneğin: Nuankaew vd., 2017; Shapoval vd., 2018) müşteri davranışlarını anlayabilmek için anketler aracılığıyla veri toplamışlardır.

Turizm ve konaklama endüstrisinde, talebi etkileyen değişkenleri ve nihayetinde endüstrinin (belirli bir otel veya restorani) performansını anlamak çok önemlidir. Geçmişte, yöneticiler ve karar vericiler bu değişkenlerin etkisini araştırmak için basit veri analizi

formlarını (basit doğrusal regresyon ya da çoklu regresyon) kullanmaktaydı (Al Shehhi ve Karathanasopoulos, 2018: 1624). Son zamanlarda, veri madenciliği, değişkenler veya nedenler arasındaki korelasyonun altını çizme ve nihayetinde performansın tahmin edilmesi gibi karmaşık iş durumlarını inceleme potansiyelini ortaya çıkarmıştır. Talebi tahmin etmek için Yapay sinir ağları (YSA) en çok kullanılan makine öğrenme tekniğidir. Philips vd. (2015), Pattie ve Snyder (1996), Govers vd. (2007) ve Cho (2003) turistlerin geçmiş varış verilerini veya oda doluluk verilerini kullanarak turizm veya otel talebini tahmin etmek için YSA'yı kullanmıştır. Bulanık sistemler (Chao ve Chao 2017; Chiang, 2017; Logesh vd., 2018; Majewska ve Truskolaski, 2018), Kaba kümeler (Goh ve Law, 2003; Goh vd., 2018;) ve destek vektörleri (Claveria vd., 2018; Cankurt ve Subaşı, 2016) turizm talep tahminini yapmak için kullanılan diğer veri madenciliği teknikleridir. Birçok çalışma bu tekniklerin birkaçını aynı anda kullanarak (Al Shehhi ve Karathanasopoulos, 2018; Claveria vd., 2018; Jun vd., 2018) daha doğru tahmin elde etmeye çalışmış ya da yöntemlerin performanslarını karşılaştırmışlardır. Özetle veri madenciliği turizm sektöründe, müşteri ve pazar bölümlenme, müşteri memnuniyetinin ya da memnuniyetsizliğinin tespiti, seyahat davranışı modelleme, turist tahmini, turizm talebi tahmini, müşteri tercihleri, yeni ürün oluşturma ve ürün değerlendirme gibi turizm alanında yapılabilen tüm araştırma konularında kullanılabilecek tekniklere sahiptir.

Turizm alanında yapılan veri madenciliği çalışmaları incelendiğinde veri madenciliğinin ilk çıkış mantığına ters bir durum ortaya çıkmaktadır. Veri madenciliği büyük veri tabanlarından yararlı bilgiyi keşfetmek ya da bu keşif sürecindeki bir adımı ifade etmek için kullanılmış bir terimdir. Turizm alanındaki çalışmalarda çoğunlukla birkaç yüz, bazen daha az anket gözleminden oluşan verilerin veri madenciliği teknikleri ile analiz edildiği gözlemlenmektedir. Son dönemlerde yaygın olarak kullanılan metin madenciliği için aynı durum söz konusu değildir, çünkü sosyal medya bu konuda yeterli veri desteğini sağlamaktadır. Turizm işletmelerinde kullanılan enformasyon sistemlerinde biriken veriler ise turizm çalışmalarında yeterince yer alamamaktadır.

Turizm işletmelerinde ön büro sistemleri (acenta ve oteller için), POS sistemleri, MİY sistemleri, arka büro (muhasabe ve insan kaynakları) sistemleri, yiyecek ve içecek sistemleri, çağrı sistemleri, enerji sistemleri gibi enformasyon sistemleri turist ile ilgili işlemlerin yürütülmesinde yaygın olarak kullanılmaktadır. Bu sistemlere ait veri tabanlarında turist seçtiği ve seçme ihtimali olan ürünler hakkında, turist tercihleri hakkında, ulaşım ve transferi ile ilgili, konaklayacağı otel ve konaklama esnasında yaptığı her türlü işlem ile ilgili yığınla veri depo edilmektedir. Bu veriler günlük işlem boyunca değerlendirilmekte, müşterinin dönüşünden sonra bazen geçmişte gelip gelmediğine bakma dışında kullanılmamaktadır. Aslında bu

verilerde işletmenin müşteri ile ilgili yararlanabileceği yararlı bilgiler gizlidir. Geçmişteki satın alma kararları, daha etkili pazarlama iletişimleri için farklı turist gruplarını segmentlere ayırmak için yararlı girdiler olmaktadır (Cho ve Leung, 2002: 126). Bu bilgilerin işletme karar vericilerinin kullanımına sunulması işletme için hayati öneme sahip olabilmektedir.

Veri madenciliği çalışmalarında veri toplamak için genellikle anket yöntemi son zamanlarda da açık veri konumundaki çevrimiçi kullanıcı içerikleri değerlendirilmektedir. İşletme bünyelerinde ve işletmeler arası gerçekleşen müşteri ile ilgili işlemlerden elde edilen veriler göz ardı edilmiştir. Bu tez çalışması literatürdeki bu açığı kapatmak için bir seyahat acentasının operasyonel kayıtlarının tutulduğu bir veri setini kullanmaktadır. Daha önceki çalışmalar seyahat acentası veri tabanlarında yeterli veri özelliği olmadığı için (Liao vd., 2010), çalışmalarında veriyi anket ile toplamıştır. Bu çalışmada ise seyahat acentası veri tabanını tercih edilmiştir.

Bu doktora tezinin amacı veri madenciliği yöntemlerinden sınıflandırma teknikleri kullanılarak günlük tur satın alan ve almayan müşterilerin profilinin çıkarılmasıdır. Bu amaç için Antalya’da faaliyet gösteren A sınıfı bir seyahat acentasının rezervasyon ve günlük tur ile ilgili işlemsel verilerinin kaydedildiği bir veri tabanında tutulan veriler kullanılmıştır. Rezervasyonu yapılan bir turistin günlük tur satın alma durumu “SATIN ALAN” veya “ALMAYAN”dır, bu da bir sınıflandırma problemidir. Dolayısıyla çalışmada veri madenciliğinin sınıflandırma fonksiyonuna ilişkin yöntemler kullanılmıştır. Araştırma sürecinde seyahat acentası veri tabanındaki farklı tablolarda yer alan rezervasyon, günlük tur bileti, ürünler ve müşteriler hakkındaki veriler bir tabloda birleştirilmiştir. Elde edilen veriler üzerinde 11 farklı sınıflandırma algoritması çalıştırılmıştır. Bu algoritmalarından en iyi performansı ve doğru tahmin oranını veren karar ağacı sınıflandırma tekniklerinden C4.5 algoritması veri içindeki müşteri profillerinin keşfi için kullanılmıştır.

Çalışmada ilk önce veri madenciliği konusunda bilgiler verilmiş ve ardından veri madenciliğinin sınıflandırma algoritmaları ile ilgili detaylı açıklamalar yazılmıştır. Turizm sektöründe veri kaynakları ve veri türleri üçüncü bölümde tanıtılmıştır. Dördüncü bölümde turizm alanında veri madenciliği ile ilgili yapılan araştırmalar yer almaktadır. Son bölümde seyahat acentası veri tabanında yapılan veri madenciliği uygulaması, tartışma ve sonuç yer almaktadır.

BİRİNCİ BÖLÜM VERİ MADENCİLİĞİ

Teknolojik gelişmeler bireylerin, işletmelerin ve toplumların birbirleri ile ilişkilerinin bir bölümünü iletişim ve enformasyon teknolojileri üzerinden yürütebilmelerine olanak sağlamıştır. Bu da bireylerin iletişimlerini sürdürebilmeleri ve işletmelerin operasyonel işlerinin yürütülebilmesi için büyük miktarda veri ve enformasyona ihtiyaç duymalarına neden olmuştur. Birey ve işletmeler enformasyon sistemlerinden ihtiyaçları olan enformasyonu elde ederlerken, aynı zamanda birer veri kaynağı işlevi de görmekte, büyük bir veri havuzu olan internete ya da bireysel veya kurumsal enformasyon sistemlerine veri katkısında bulunmaktadır. Enformasyon sistemlerinin kullanımına olan bu büyük katılım, şirketlerin müşterileri hakkında işletmelerin faturalama gibi günlük faaliyetlerini yaparken yüzlerce gigabayt veya terabayt veri toplamalarına yardımcı olmaktadır.

Günümüzün rekabet ortamında işletmelerin başarılı olabilmeleri için mevcut müşterilerini elde tutmaları, bunun içinde onlara özel ürünler ve fırsatlar sunmaları gerekmektedir. Bunun da ilk adımı müşteriye ve ne yaptığını tanımaktır. Büyük ticari işletmelerde, müşterinin ne yaptığını fark etmenin ilk adımı büyük ölçüde otomatikleştirilmiştir (Berry ve Linoff, 2004:3). Buna rağmen çoğu işletme müşterileri hakkında hiçbir şey öğrenmeden yığınla veriyi sadece operasyonel işlemlerde kullanıp, veri tabanlarında çürümeye (Han ve Kamber, 2006:4) terk etmektedir. Müşteri odaklı bir işletme ise müşteri ya da müşteri adayıyla yapılan her türlü işlemi (müşteri hizmetlerine yapılan her çağrı, her satış noktası işlemi, her katalog siparişi, her web sitesi ziyaretini) müşterileri hakkında bir şeyler öğrenebilme fırsatı olarak görmektedir. Fakat öğrenme sadece veri toplamaktan daha fazlasını gerektirmektedir (Berry ve Linoff, 2004:4).

Sistemlere kaydedilenlerden bir şeyler öğrenebilmek için, birçok kaynaktan gelen veriler - fatura kayıtları, tarayıcı verileri, kayıt formları, başvurular, çağrı kayıtları, kupon ödemeleri, anketler - önce bir araya getirilmeli ve tutarlı ve yararlı bir şekilde organize edilmelidir (Berry ve Linoff, 2004:5). Muhtemelen harici verilerle desteklenen, işletmenin birçok operasyon destek sisteminden alınmış verileri içeren sistemler veri ambarı (Danubianu vd., 2009: 293) olarak bilinmektedir. Veri ambarı, işletmelerin müşterileri hakkında ne fark ettiğini hatırlamasını sağlar. Veri ambarlarının doğru tarihsel verileri desteklemesi gerekir, böylece veri madenciliği bu kritik eğilimleri yakalayabilir. Veri ambarı kuruluşu bir bellek sağlamaktadır. Ancak, zekâsı olmadan bellek pek kullanışlı değildir. Zekâ, geçmiş verileri taramayı, kalıpları fark etmeyi, kuralları tasarlamayı, yeni fikirlerle gelmeyi, doğru soruları bulmayı ve geleceğe

dair öngörülerde bulunmayı sağlamaktadır. Veri ambarına zekâ katan araçlar ve teknikler, onları daha iyi tanımak için müşteriler ve potansiyel müşterilerle olan etkileşimler sonucu ortaya çıkan geniş veri yığınlarından yararlanmayı mümkün kılmaktadır (Berry ve Linoff, 2004:6).

Veri ambarının en önemli yönlerinden biri, müşteri davranışlarını zaman içinde izleme kabiliyetidir. Müşteri ilişkileri yönetimi için ilgi kalıplarının çoğu sadece zamanla belirginleşmektedir. Müşteri ile ilişkileri sadece yapılan işlemler bazında değil toplam hayat döngüsü bazında yönetmeyi sağlayacak veri yönetimi yapısının ve buna ilişkin entegrasyon gereksinimlerinin tasarlanması ve müşteri ilişkileri yönetimi sisteminin kurulması memnun ve sadık müşteri grubunun oluşturulmasında çok katkı sağlayacaktır. İşletme içindeki analitik yeteneklerini artırmak veya bunların en olası kombinasyonlarına imkân tanımak üzere veri ambarları kullanılabilir. Turizm sektöründeki yöneticilere mevcut veriler hakkında bilgi sağlamak ve vermek veri ambarı sistemlerinin temel işlevidir (Mirela vd., 2009: 922). Sadık bir müşteri olarak kalması muhtemel olan ve tekrar gelme ihtimali olan kimdir? Hangi ürünler hangi pazarlarda pazarlanmalıdır? Bir turistin belirli bir promosyona cevap verip vermeyeceğini belirleyen nedir? Bir müşterinin isteyeceği bir sonraki ürün veya hizmet nedir? Bu gibi sorulara verilen cevaplar, kurumsal verilerin içine gömülmüş durumdadır. Onlara ulaşmak için güçlü veri madenciliği araçları kullanılmalıdır.

Veri madenciliği, verilerdeki geçerli, yeni, potansiyel olarak yararlı ve nihayetinde anlaşılabilir örüntüleri belirleyen bir süreç (Ha ve Park, 1998: 4) ve anlamlı kalıpları ve kuralları keşfetmek için büyük miktarlardaki verilerin araştırılması ve analizidir. Kurumsal verilerde yakalanan müşteri davranışları rastgele değil, müşterilerin farklı ihtiyaçlarını, tercihlerini ve eğilimlerini yansıttığı için çalışmaktadır. Veri madenciliğinin amacı, bu ihtiyaçlara, tercihlere ve eğilimlere ışık tutan tarihsel verilerdeki kalıpları bulmaktır. Veri madenciliği de bu bilgilere, kalıplara vb. ulaşabilmek için belirli bir süreç içerisinde çeşitli yöntemler kullanmaktadır. Bu yöntemler, genel olarak tanımlayıcı (Descriptive) ve tahmin edici (Predictive) olmak üzere iki ana başlık altında incelenmektedir (Rud, 2001: 4). Belirli bir durumda uygulanacak belirli bir yöntem kombinasyonunun seçimi, veri madenciliği görevinin doğasına, mevcut verilerin yapısına ve araştırmacının becerilerine ve tercihlerine bağlıdır (Berry ve Linoff, 2004:7).

Bir veri madenciliği sisteminden istenebilecek her soru için, uygulanabilecek birçok araç vardır. Bazı durumlarda, tek bir aracın uygulanmasıyla sonuca ulaşılabilirken bazen, bir çözüme ulaşmak için birden fazla araç kullanılmak durumunda kalınmaktadır (MacLennan vd., 2009: 6). Sınıflandırma, kümeleme, birliktelik, regresyon, tahmin, dizi analizi ve sapma analizi genel veri madenciliği görevlerindedir.

Veri madenciliğinin birçok ticari faaliyetin geliştirilmesinde son derece etkili olduğu kanıtlanmıştır. Veri madenciliği süreci, veri depolama desteği (veri ambarları, veri martları ve / veya çevrimiçi analitik işleme araçları) ve ayrıca veri analiz yazılımı (veri madenciliği yazılımı) formundaki bilgi teknolojilerine dayanmaktadır. Ancak, veri madenciliği süreci, bu veri madenciliği yazılım araçlarını bir firmanın verilerine uygulamaktan çok daha fazlasıdır. Model tiplerinin seçiminde, belirli bir problemle ilgili verilerin seçiminde ve dönüştürülmesinde ve sonuçların yorumlanmasında analistin zekâsına ihtiyaç vardır (Olson ve Delen, 2008:8).

1.1. Veri İle İlgili Kavramlar

Veri madenciliğinin temelini veri oluşturmaktadır. Veri madenciliği, büyük veri havuzlarında yararlı bilgileri otomatik olarak keşfetme sürecidir. Verinin bilgiye dönüşme sürecinde verinin elde edilmesinden analize hazır hale getirilmesine kadar bazı işlemlerden geçmesi gerekmektedir. Bu süreçte sık karşılaşılan terimler bu konuda açıklanmıştır.

1.1.1. Veri, Enformasyon ve Bilgi

Dünyadaki varlıklar sözlü veya yazılı ifadeler ile ya da şekiller, resimler, rakamlar, semboller ile tanınmakta ve tanımlanmaktadır. Yaşı, boyu, göz rengini, kiloyu, işi bazen duyguları, algıları ölçmek veya anlatabilmek için sayısal ya da sayısal olmayan değerler kullanılır. İşletmelerde rezervasyon bilgileri, ön büro işlemleri, kasa işlemleri, operasyon işlemleri ve faturalama işlemleri esnasında bu işlemleri tanımlayan karakterlerden ve rakamlardan bazen görüntü ve videolardan oluşan sembol ve şekiller ile bilgisayar sistemlerine kaydedilir. Yukarıdaki örneklerde olduğu gibi kayıt altına alınmış her türlü olay, durum ve fikir veri olarak adlandırılır. Veri (Data) Latince vermek anlamında kullanılan dare fiilinin geçmiş zaman çekimi olan datum kelimesinin çoğuludur (Akpınar, 2014: 15) Türk Dil Kurumu veriyi “Olgu, kavram veya komutların, iletişim, yorum ve işlem için elverişli biçimli gösterimi” şeklinde tanımlamıştır.

Veri, temel olarak varlığı bilinen, ilişkilendirilmemiş, düzenlenmemiş, özümlememiş, yorumlanmamış gözlemler ve işlenmeye hazır ama işlenmemiş ham haldeki sayı, rakam veya numara gibi figürlerdir. Veri, objeleri, olayları veya onların özelliklerini temsil eden sembollerini içerir. Bilgisayar ortamında ise bilgisayarın alabildiği, işleyebildiği, sonuç üretebildiği ve saklayabildiği sayısal, alfa sayısal veya mantıksal biçimlerde ifade edilebilen her türlü değer tanımla ifade edilebilir. Müşteriye ait voucher, rezervasyon, tur ve transfer bilgileri gibi genellikle fiziksel bir olay veya işletme faaliyetlerini ilgilendiren, kendi başlarına çok az anlam taşıyan harf, rakam, özel karakterler gibi ham gerçekler veya gözlemler veriye örnek olarak verilebilir.

Verinin belirli bir sorun çerçevesinde, ilişkilendirilmesi, anlamlandırılması, işlenmesi, düzenlenmesi ile enformasyon oluşur. Enformasyon, verinin işlenmiş, içeriği olan ve insanlara anlam taşıyan halidir şeklinde de tanımlanmıştır. İşletmeler için pazar ve ekonomi ölçümleri /göstergeler gözlem yolu ile elde edildiğinden veri sayılırken, bunların analizleri ve pazarın ekonomik davranış modellerinin anlaşılır bir yapıya dönüştürülmesi enformasyon olarak ifade edilir (Çizel, 2005). Veriden daha çok zengin bir içeriğe sahip olan enformasyon, alan kişinin algılamasının değişmesi ve yargısı üzerinde etki yapmayı hedefleyen yazılı, sözlü veya görsel bir mesajdır.

Enformasyon otomatik olarak bilgiye dönüşmez, bunun için enformasyon ayrıca bilişsel bir süreçten geçirilerek öznel yargıya dönüşmelidir. Enformasyonun bilgiye dönüşebilmesi için, onu bireyin algılaması, özümsemesi ve sonuç çıkarması ve mevcut olan belirsizliğin ortadan kalkmış olması gerekmektedir. Bilgi (Knowledge), enformasyonun belirli bir amaç için ilişkili olarak kullanılmasıyla ortaya çıkmaktadır. Bu durumda bilgi yorumlanmış, analiz ve bağlam ile zenginleştirilmiş, kişisel anlamda düzenlenmiş, özümsemiş enformasyondur. Bilgi insanın beyni tarafından oluşturulan bir anlam (Çizel, 2005: 44) olduğundan enformasyonun bilgiye dönüşmesi insanın kişisel niteliklerinden (algılama yeteneği, yaratıcılık, deneyim, kabiliyet, kültür, karakter, kişilik, duygu, sezgi, algı, güdü, eğitim, ortam) doğrudan etkilenmektedir.

1.1.2. Enformasyon Sistemleri

1980'lerde güçlü kişisel bilgisayarların (PC) geliştirilmesi, bireylerin ana bilgisayarlara veya mini bilgisayarlara yatırım yapmak zorunda kalmadan güvenilir bir bilgi işlem yapmasını sağlamıştır. Kişisel bilgisayarlarla birlikte bilgisayarlı sistemler ve uygulamaları hızla geliştirilmiş, daha geniş işlev ve etkinliklerde kullanılmaları sağlanmıştır (Praničević, 2006: 925). Bu gelişmeler sonunda birey ve işletmelerin günlük yaşamı tamamen değişmiş, özellikle işletmelerin iş yöntemleri yeni teknolojilere uygun hale gelmiştir. Zamanla bu teknolojiler insanlar ve işletmeler için vazgeçilmez bir yaşam aracı haline dönüşmüştür.

Günümüz teknolojisi bir işletmenin tüm faaliyet alanlarında kullanılabilecek kapasitede ve ortaya çıkan verileri depo edebilecek yazılım ve donanıma sahip durumdadır. Enformasyon teknolojileri, bilgisayar hızını sürekli artırarak, donanım boyutunu küçülterek; donanım ve yazılım maliyetlerini azaltarak ve sayısız terminal ve uygulamanın güvenilirliğini, uyumluluğunu ve birbirine bağlanabilme özelliğini geliştirerek yeteneklerinde köklü değişiklikler yapmaktadır (Buhalis, 1998: 1). Bu değişiklikler enformasyon sistemlerini işletmeler için vazgeçilmez bir rekabet aracı haline getirmiştir. İşletmeler başarılı bir şekilde hem müşteri ve hem de firma değeri arttırmak için bu teknolojileri kullanmak durumundadır. Enformasyon

teknolojilerinin veri işleme ve verinin alış verişi konusundaki üstünlüklerinden, artan veri hacmi ve işlemlerin karmaşıklaşmasının yanında taleplere anında cevap verme ve değişen şartlara uyum sağlamanın gerekliliğinden dolayı kullanımları zorunlu hale gelmiştir.

Enformasyon teknolojileri, enformasyonun toplanması, işlenmesi, saklanması, erişilmesi, iletilmesi ve uygulaması (kullanılması) için kullanılan yöntemlerdeki (bilgisayarlar ve iletişim teknolojileri) en son gelişmelere verilen ortak terimdir (Buhalis, 1998: 1). Enformasyon sistemleri ise örgütün koordinasyon, kontrol, analiz, karar verme gibi fonksiyonlarını desteklemek için enformasyonun toplanması, işlenmesi, depolanması ve dağıtılması için çalışan, birbirleri ile ilişkili unsurlar olarak tanımlanır (Çizel, 2005: 46). Bir enformasyon sistemi, toplum ve örgüt için hayati olan enformasyonu kullanmak isteyen yönetici, personel, müşteri, herkes için erişilebilir ve servis edilebilir bir şekilde toplayan, depolayan, koruyan, işleyen ve dağıtan bir sistemdir. Bir diğer deyişle enformasyon sistemi, enformasyon teknolojilerini kullanan aktif toplumsal bir sistemdir (Pilepić ve Šimunić, 2009: 413).

Örgütün günlük rutin faaliyetleri ile ilgili verilerin işlenmesi ve bir veri tabanına kaydedilmesi için Operasyon Destek Sistemi yazılımları kullanılır. Yönetim Enformasyon sistemi (YES) karar verme sürecini kolaylaştırmak için gerekli, zamanlı ve doğru bilgiyi yönetime sağlayan, örgütün planlama kontrol ve operasyonel fonksiyonlarının etkin bir şekilde yürütülmesine imkân veren enformasyon sistemidir. Yöneticilere yarı yapısal ve yapısal olmayan kararları almalarında destek sağlamak amacıyla, karar modellerine ve verilere kolay erişimi sağlayan etkileşimli ve esnek enformasyon sistemleri de Karar Destek Sistemleri olarak adlandırılır (Çizel, 2005). Bu sistemlere ait yazılımlar yazılım şirketleri tarafından bütünleşmiş bir sistem halinde yazılarak işletmelerin kullanımına sunulmaktadır. Günümüz işletmelerinde enformasyon teknolojisine sahip olmak stratejik bir kaynağa sahip olmaktan ziyade hayatta kalmaya yarayan bir olgu olarak algılanmaktadır.

Turizm işletmelerinde operasyon destek sistemleri olarak ön büro (Danubianu vd., 2009) ve arka büro programları yaygın olarak kullanılmaktadır Seyahat acentalarında rezervasyon, operasyon günlük tur ve muhasebe işlemlerinin takibi için operasyon destek sistemleri kullanılmaktadır. Otellerde tesis yönetim sistemleri otelin rezervasyon, ön büro, kat hizmetleri, müşteri hesapları, POS işlemleri, yiyecek ve içecek kayıtlarını (Kasavana, 2011: 6) tutmaktadır. Kullanılan programların hepsi gerekli olan verinin depolanması için veri tabanı yönetim sistemlerinden yardım almaktadırlar. Günümüzde hiçbir yazılım verileri kendine ait bir veri tabanında tutmamaktadır.

1.1.3. Veri Tabanları

Dijitalleşmenin tüm hızı ile sürdüğü günümüzde işletmelerin rekabet ortamında ayakta kalabilmeleri için (Magnini vd., 2003: 95) veriye ulaşım ve kullanım kilit rol oynamaktadır. Verilerin bir işletme tarafından elde edilmesi, korunması, gerektiğinde paylaşılması ve işlenmesi hayati öneme sahiptir. Bir seyahat acentasında müşteri bilgileri, rezervasyon bilgileri, tur ve transfer bilgileri, ödeme bilgileri, bir otel için personel ve müşteri bilgileri, oda ve konaklayan bilgileri, stok bilgileri, satılan ürün bilgileri, doluluk bilgisi, oda durumu bilgileri, ödeme bilgileri, folyo ve fatura bilgileri vb. çok çeşitli veriler sistemlerde tutulmaktadır. Bir otelde ise tesis yönetim sistemleri ve POS sistemleri sayesinde bir müşterinin aldığı tüm ürünler, bir müşteri ya da ürünün otele girişinden çıkışına ya da tüketilişine kadar tüm hareketlerinin bütün detayları saklanabilmekte ve hareketleri izlenebilmektedir. Turizm sektöründe dijitalleşme ve iş modellerinin değişmesi ile birlikte son dönemde gündemde olan konulardan biri de önceki başlıklar altında bazı örnekleri verilen verilerin yönetimidir.

Sistemlerde depolanan verinin artması ve veri tabanı yönetim sistemlerinin daha güvenli olmasından dolayı, bir zamanlar kendi bünyelerindeki kütüklerde verileri depo eden otel ve acenta yazılımları (örneğin Fidelio), günümüzde veri depolama işlemini veri tabanı yönetim sistemlerine devretmiş durumdadır. İşletmelerin kullandığı operasyon destek sistemleri işletmeler tarafından girilen ürün ve müşteri verilerini kendi bünyelerinde tutmak yerine asıl görevi veriyi saklamak ve yönetmek olan başka bir yazılım tarafından oluşturulmuş veri tabanını kullanmaktadır. Bir bilgisayarda sistematik şekilde saklanan, programlarca kullanılabilen, birbirleri arasında tanımlı ilişkiler olan, yönetilebilir, güncellenebilir, taşınabilir, düzenli veri yığınlarına veri tabanı denilmektedir.

Veri tabanı sistemleri belirli bir kuruluş tarafından bir ya da birkaç uygulamada kullanılmak üzere, gereksiz yinelemelerden arındırılmış olarak, veri kümelerinin düzenli bir şekilde tutulduğu ve bu verinin çeşitli yazılımlar aracılığıyla yönetildiği bir ortamdır (Özkan, 2008: 14). Veri tabanı; seyahat acentası, tur operatörü, otel, restoran, okul, üniversite, banka, üretim şirketi, hastane, kamu kuruluşu gibi bir kuruluşun uygulama programlarının kullandığı operasyonel verilerin yönetim alanıdır. Bir veri tabanı, bir yazılım paketi tarafından bir bilgisayarda düzenlenmiş ve kullanıcının bu tür girdileri dâhili olarak mantıksal kriterlere göre ilişkilendirmesine, harmanlamasına, özetlemesine ve çoğaltmasına izin vermektedir (Law, 1998: 59). İstenildiğinde verilerin tümü veya istenilen özelliklere uyanları görüntülenebilmekte, yazdırılabilmekte hatta verilerden yeni bilgiler üretilerek çeşitli amaçlar için kullanılabilir.

Bir veri tabanından beklenen özellikler; verileri koruması, onlara erişilmesini sağlaması ve başka verilerle ilişkilendirilmesi, sorgulama ve raporlama gibi işlemlerin yapılmasına izin

vermesidir. Bir veri tabanı sayesinde veriler bir merkezde toplanmakta ve kontrol altında tutulmakta, herkesin bu verilere yetkileri ölçüsünde erişmesi, düzeltilmesi, silmesi veya görübilmesi sağlanmaktadır. Verilerin bir merkezde toplanması sonucunda verilerin tutarlı olması ve erişimlerdeki tutarsızlıkların ortaya çıkması önlenmektedir. Bu aynı zaman da bazı verilerin tekrar tekrar girilmesini önlemektedir. Böylece veri girişinde ve veriye erişimde etkinlik ve güvenilirlik sağlanmaktadır.

1.1.4. Veri Ambarları

Günümüz yöneticileri çok değişken olan iş dünyasında satışlardan rakiplerine, müşterilerden tedarikçilere kadar her türlü bilgiye her zamankinden daha çok ihtiyaç duymaktadırlar. İhtiyaç duyulan bilgiye ulaşabilmek için müşteriler başta olmak üzere işletme ile ilgili taraflar ve süreçler hakkında veri toplamak gerekmektedir. Başarılı bir turizm işletmesi olabilmek için müşteri, ürün ve hizmet, tedarikçi, satış ve doluluk ile ilgili tüm verilere doğru zamanda ve doğru bir şekilde ulaşabilmek oldukça önemlidir. İşletmelere veri toplama konusunda enformasyon teknolojileri yardımcı olmaktadır. Bu sistemler her gün işletmeler için milyarlarca byte müşterileri, iş hareketleri, ürünleri, çalışanları vs. hakkında veri üretmekte ve depolamaktadırlar.

Veri işletmelerin önemli bir varlığı olarak kabul edilmekte ve müşteriye dönük sunulacak birçok hizmette de yapılacak analizin ana girdisini oluşturmaktadır. Yalnız işletmelerde veriler bazen bir departmanın bazen de tüm işletmenin tüm iş süreçlerinin başından sonuna kadar yürütülmesini desteklemek için tasarlanmış veri tabanlarında (ÇHİ: Çevrimiçi Hareket İşleme) tutulmaktadır. Veri tabanlarındaki verilerden karar vericilerin yararlanabilecekleri bilginin çıkartılmasında iki büyük sorun oluşmaktadır. Veri tabanlarında veriler farklı tablolarda hatta bazen farklı veri tabanlarında tutulmaktadır. İkinci sorun veri yığınının büyüklüğünün artmasıyla birlikte, işlem yapmak zorlaşmakta ve özellikle yeni veri girişlerinin yapıldığı şu anda işlem gören ve geçmişe dönük verilerin aynı ortamda bulunduğu durumlarda ciddi performans problemleri yaşanabilmektedir. Bundan dolayı karar mekanizmasını oluşturan yöneticiler ve danışmanlar, günlük operasyonlara ait verilerin kaydedildiği veri tabanlarını kullanarak işletmede zaten mevcut olan kritik bilgilere ya da enformasyonlara ulaşamamaktadırlar. Bunun sonucu olarak uzmanlar, mevcut verilerden elde edilmiş, işlenmiş ve depolanmış sadece küçük bir parça veriye ulaşabilmekte ve o kadar yoğun veri içinde bilgisiz kalmaktadırlar (Ha ve Park, 1998). ÇHİ uygulamaları girdi/çıkıtlı yoğunluklu işlemlere odaklandığından karar destek sistemlerinin gereksinim duyduğu karmaşık sorgulama ve çözümleme işlemlerinde yetersiz kalmaktadır.

Yönetim kararları verme sürecini kolaylaştırmak amacıyla işletmenin günlük rutin işlemlerinin kaydının yapıldığı veri tabanlarından çok sayıda heterojen veri kaynağının bir yerde toplanarak verinin üzerinde daha etkili ve daha kolay sorguların yapılması sağlanmalıdır. Bunun için veri yönetimi ve bağlantılı konuların sadece bilişim teknolojileri yetkisinde bırakılmaması, iş birimi bazında bir aksiyon olarak düşünülüp hareket edilmesi gerekmektedir. İşletme içindeki analitik yeteneklerini artırmak veya bunların en olası kombinasyonlarına imkân tanımak üzere veri ambarı oluşturulmalıdır. Bir veri ambarı, müşteriler, satışlar ve kârlar gibi kurumsal çaptaki konulara odaklanarak geleneksel enformasyon görüşlerinin ötesine geçmektedir. Bu konular hem organizasyonel hem de süreç sınırlarını kapsar ve tam bir görüntü sağlamak için birden fazla kaynaktan bilgi gerektirmektedir (Danubianu vd., 2009: 293).

Veri ambarı kararların doğrulanması için gerekli bilgileri elde etmek amacıyla yöneticilerin emrinde tutarlı ve kalıcı bir veri kaynağı oluşturmak ihtiyacından ortaya çıkmıştır (Danubianu vd., 2009: 293). Veri ambarı, karar verme sürecinde yöneticilere destek vermek amacıyla konuya yönelik, bütünlük, zaman boyutu olan ve sadece okunabilen veri topluluğudur (Özkan, 2008: 22). Çevrimiçi analitik işlemeyi (HAI) desteklemek için tasarlanmıştır (Danubianu vd., 2009: 290). Analitik modeller tipik olarak veri ambarlarında depolanan büyük veri kümelerini kullanmaktadır (Bell ve Mgbemena, 2018: 196). Inmon, bir veri ambarını “karar vermede gerekli olan bilgiyi sağlamak için tasarlanmış bütünlük, konuya odaklı, veri tabanlarının bir koleksiyonu” olarak tanımlamaktadır (akt: Han ve Kamber, 2006: 106). Bir veri ambarı, yönetimin karar verme sürecini desteklemek için konuya yönelik, bütünlük, zaman aralıklı, kalıcı bir veri topluluğudur (Ha ve Park, 1998:1). Ayrıca, bir veri ambarı, çeşitli kaynaklardan elde edilen ve nihai kullanıcılara anlayabilecekleri ve kullanabilecekleri şekilde sunulan tek, eksiksiz ve tutarlı bir veri deposudur (Pyo vd., 2002: 397).

Bir veri ambarı, muhtemelen harici verilerle desteklenen, işletmenin birçok operasyon destek sisteminden alınmış verileri içermektedir. Bu operasyonel sistemlerin her biri, farklı türde ticari işlemleri kaydeder ve bu işlemlere ilişkin işletmenin politikalarını uygulamaktadır (Danubianu vd., 2009: 293). İdeal olarak, depodaki veriler birçok kaynaktan toplanmış, temizlenmiş, birleştirilmiş ve çeşitli yararlı yollarla özetlenmiştir. İyi bir veri ambarı, verilerin oluşturulduğu operasyon destek sistemlerindeki saklanma biçiminden çok daha dostça bir biçimde bilgiye erişilmesini sağlamaktadır (Berry ve Linoff, 2004). Ayrıca veri ambarı geçerliliği daha uzun sürecek bilgiye de ulaşımı sağlamaktadır, çünkü veri tabanı içerisindeki veriler genelde anlıktır, yani o an için güncelliğini koruyan ancak belirli bir süre sonunda güncelliğini kaybeden verilerdir. Ancak veri ambarı içerisindeki veriler genelde yığılarak

birikmekte ve geçerlilikleri çok daha uzun sürmektedir. Veri ambarı içerisinde ne kadar çok kayıt olursa yapılan incelemelerin sonucu da o kadar doğru olacaktır.

1.2. Veri Madenciliği

Günümüzde enformasyon teknolojileri alanındaki gelişmeler, baş döndürücü bir hızla ilerlemekte ve her gün bir başka yenilikle insanların hizmetine sunulmaktadır. Turizm ve seyahat endüstrisi de bu yeniliklere yabancı kalmamış ve özellikle seyahat ve turizm faaliyetlerinin doğası gereği en büyük enformasyon teknolojileri (Buhalis, 1998; Pyo vd., 2002) kullanıcılarından birisi olmuştur. Kullanılan yüksek performanslı enformasyon teknolojileri işletmelerdeki iş akışını değiştirmiştir. Artık işletmelerin her türlü günlük rutin işlemleri enformasyon sistemleri ile yapılmakta ve günlük tüm veriler sayısal ortamda saklanmaktadır. Turistlere ait kişisel bilgiler (Shapoval vd., 2018: 311), konaklama, transfer, tur ve aldıkları tüm hizmet ve ekstralara ait veriler ilgili işletmelerin kullandığı programlar aracılığıyla veri tabanlarında depolanmaktadır.

Günümüz işletmeleri yaptıkları iş ile ilgili, müşterileri ile ilgili ve faaliyette buldukları pazar ile ilgili yığınla veriyi kendi bünyelerinde depolamaktadırlar, ayrıca istedikleri takdirde dış ortamlardan elde etme şansına sahiptirler. Veri artık işletmelerin önemli bir varlığı olarak kabul edilmekte ve müşteriye dönük sunulacak birçok hizmette de yapılacak analizin ana girdisini oluşturmaktadır. Bunun için veri yönetimi ve bağlantılı konuların sadece bilişim teknolojileri yetkisinde bırakılmaması, iş birimi bazında bir aksiyon olarak düşünülüp hareket edilmesi gerekmektedir. Veri yığınlarının içerisinde işletmenin karar vericilerinin yararlanabileceği anlamlı bilgilerin çıkartılması gerekmektedir.

İşlenmeden veri tabanlarında yığılan veriler, günlük rutin faaliyetlerin raporlanmasından başka ne işe yaracaklardır? Veri tabanlarında yığılan veriler, eğer karar destek sistemleri aracılığıyla yöneticilerin karar almasını sağlayacak şekilde yönetilmezlerse bilgisayarlarda yer almaktan başka hiçbir işe yaramayacaklardır. Bu da günümüzde özellikle turizmde kişisel ürün oluşturma ve pazarlama çabası içinde olan işletmelerin aslında ellerindeki bir hazineden mahrum kalmalarına neden olacaktır. Günümüzde bazı firmalar çok miktarda veriye sahip olmalarına karşın, bilgi bakımından oldukça fakir (Han ve Kamber, 2006: 4) durumundadır. Buna karşın, gelişmiş yazılım ve iletişim teknolojisi, hem potansiyel hem de gerçek turistlerin beklentileri, kişilikleri ve sürekli değişen ihtiyaçları ile tutarlı benzersiz maceralar bulmalarına yardımcı olmaktadır (Pyo vd., 2002: 396). Bu gelişmeler gelecekte daha agresif bir destinasyon yönetim ve planlama ve pazarlama ihtiyacına da yol açacaktır. Turist özellikleri, imajları, tutumları ve tercih edilen destinasyon nitelikleri dâhil olmak üzere hâlihazırda entegre olan

bilgi, destinasyonları etkili bir şekilde pazarlamak için kullanılmalıdır (Buhalis, 1998). Bu tür bilgiler pazarlamacılara da sunulmalıdır, böylece pazarlamacılar bu değişen çevreye göre yanıt verebilirler (Pyo vd., 2002: 396).

Bir işletmenin günümüz rekabet ortamında eski müşterilerini koruması ve yeni pazarlar bulması için, kurumsal veri yönetimi konusunda yatırımlar yapması ve işletme içindeki analitik yeteneklerini artırması gerekmektedir. Müşteri ile ilişkileri sadece yapılan işlemler bazında değil toplam hayat döngüsü bazında yönetmeyi sağlayacak veri yönetimi yapısının ve buna ilişkin entegrasyon gereksinimlerinin tasarlanması ve müşteri ilişkileri yönetimi sisteminin kurulması memnun ve sadık müşteri grubunun oluşturulmasında çok katkı sağlayacaktır. Bunun için artan miktardaki veriden yararlı bilginin çıkarılmasına yardımcı olacak ve onu bilgiye dönüştürmeye yardımcı olacak araçlara olan ihtiyaç giderek artmaktadır (Shapoval vd., 2018: 311). Bilgi dediğimiz bu gizli enformasyon, veri tabanlarında gizlidir ve genellikle veri öğeleri arasındaki ilişkiler biçimindedir. Bu ilişkiler fonksiyonel veya kısmi işlevsel bağımlılıklar şeklinde olabilmektedir. Bilginin keşfedilmesi ve tanımlanması çeşitli tekniklerin kullanımını içerebilir (Cho ve Leung, 2002: 117). Veriler üzerinde çözümlemeler yapmak amacıyla çeşitli istatistiksel ve matematiksel yöntemler kullanılabilir.

İşletmeler için hayati öneme sahip verilerden bilgi elde etme işlemi için geleneksel sorgu veya raporlama araçlarının ve istatistik yöntemlerinin veri yığınları karşısında yetersiz kalması yeni teknoloji arayışlarına neden olmuştur. Bu arayışlardan birisi de büyük ölçekli veri tabanlarından veri içinde gizli, önceden bilinmeyen, anlamlı ve yararlı örüntülerin otomatik biçimde elde edilmesini sağlayan Veri Tabanlarında Bilgi Keşfi (VTBK)'dir. VTBK gürültülü ve tutarsız verilerin temizlenmesi, varsa birden fazla veri kaynağının birleştirilmesi, ilgili verilerin veri tabanından alınması, özet veya toplama işlemleri yaparak verilerin madenciliğe uygun formlara dönüştürülmesi, veri madenciliği, örüntü değerlendirmesi ve bilgi sunumu adımlarının yinelemeli bir dizisinden oluşmaktadır (Han ve Kamber, 2006: 7). Araştırmacılar tipik olarak veri madenciliğini de kapsayan tüm süreçleri veri tabanları veya VTBK bilgi keşfi olarak ifade ederlerken (Örneğin: Pyo vd., 2002), veri piyasasında veri madenciliğinin kullanılması, genellikle bilgi keşfinin tüm sürecini ifade etmektedir (Ha ve Park, 1998: 4). Junping vd. (2008), veri madenciliğini büyük veri tabanları veya veri setlerinden dikkate değer bilgi çıkarma işlemi olarak tanımlamakta ve bilgi çıkarımı, bilgi keşfi, akıllı veri analizi, bilgi edinme ve veri tabanlarında bilgi keşfi gibi isimlerle de adlandırılabilceğini belirtmektedirler. Bu çalışmada da veri madenciliği bilgi keşfinin tüm sürecini ifade etmektedir.

Yazında “Veri Madenciliği” teriminin farklı tanımlarının yapıldığı makale ve çalışmalar mevcuttur. Veri madenciliğinin en eski tanımlarından biri “örtük, önceden bilinmeyen ve

potansiyel olarak yararlı bilginin veriden ayrılması”dır (Shapoval vd., 2018: 311). Veri madenciliği geliştikçe, aşağıdaki gibi yeni bir tanım önerilmiştir: “Veri tabanlarında bilgi keşfi, verilerdeki geçerli, yeni, potansiyel, kullanışlı ve nihai olarak anlaşılabilir kalıpları tanımlamada ihmal edilemeyen bir süreçtir” (Fayyad vd., 1996: 30). Veri madenciliği, anlamlı model ve kuralları keşfetmek için büyük miktarlardaki verilerin keşifsel analizidir (J. Wong vd., 2006: 359). Maimon ve Rokach (2010) göre veri madenciliği desenleri, verideki ilişkileri ve diğer bulguları (beklenmeyen şaşırtıcı ve yararlı) keşfetmek için büyük bir mevcut veri tabanlarının analizini içermektedir. Tablo 1.1’de turizm alanında veri madenciliği yöntemlerinin kullanıldığı bazı çalışmaların yer verdiği veri madenciliği tanımları listelenmiştir.

Tablo 1.1 Turizm Çalışmalarında Kullanılan Veri Madenciliği Tanımları

Çalışma	Kullanılan Veri Madenciliği Tanımı
Magnini vd., 2003	Veri madenciliği, büyük veri setlerinden yararlı, belirgin olmayan ve önceden bilinmeyen desenleri veya veri eğilimlerini tespit etmek için istatistiksel analizler kullanan, daha önce keşfedilmemiş ilişkilerin bilgisayar tabanlı keşfini (genelde yalnızca sınırlı insan müdahalesi gerektiren "makine öğrenme" yöntemlerini kullanarak) vurgulayan büyük ölçüde otomatikleştirilmiş bir işlemdir.
Pyo vd., 2002	Veri tabanlarından bilgi keşfi, verileri karar desteği için yararlı hale getiren kalıpları keşfetme sürecidir.
Karathiya vd., 2012	Veri madenciliği, daha iyi karar verme için bilgi üretmek amacıyla daha önce bilinmeyen desenleri, eğilimleri ve ilişkileri araştırmak ve bulmak için genellikle büyük veri kümelerini analiz etme sürecidir.
Ha ve Park, 1998	Veri tabanlarında bilgi keşfi olarak da adlandırılan veri madenciliği, modern istatistik ve yapay zekâ teknolojilerini büyük veri tabanlarından geçerli, önceden bilinmeyen, anlaşılır ve uygulanabilir bilgiler çıkarma sorununa uygulayan ve önemli iş kararları vermek için kullanan bir bilimdir.
Bose, 2009	Tan, Steinbach ve Kumar’a (2006) göre “Veri madenciliği, büyük veri havuzlarında yararlı bilgileri otomatik olarak keşfetme sürecidir”.
Danubianu ve Hapenciuc, 2008	Veri madenciliği, anlamlı modelleri ve kuralları keşfetmek için büyük miktarlardaki verilerin otomatik ya da yarı otomatik yöntemlerle keşif ve analizidir.
Cho ve Leung, 2002	1990'larda başlayan veri madenciliği genellikle Cho (1998), Smyth vd. (1996) ve Wüthrich (1995), “örtük, önceden bilinmeyen ve potansiyel olarak yararlı bilginin veriden ayrıştırılması” olarak tanımlanmıştır.
Pyo vd., 2002	VM, trendleri, kalıpları ve korelasyonları ayıklamak için özel bir algoritma uygulayan KDD sürecindeki bir faaliyettir ve aynı zamanda bir veri ambarından örtük bilgiyi keşfetme sürecidir (Chou ve Chou 1999).
Čech ve Bureš, 2009	Veri madenciliği, bilinmeyen ve insan kapasitesinin ötesindeki potansiyel olarak yararlı desenleri, bağlantıları, ilişkileri ve kuralları keşfetme sürecini ifade etmektedir.
Koyun ve Yangeç, 2018	Veri madenciliği, veri yığınlarının belli bir amaç doğrultusunda işlenerek anlamlı bilginin elde edilmesidir.
Öztürk ve Tanrısevdi, 2017	Veri madenciliği, önceden bilinmeyen, veri içinde gizli, anlamlı ve yararlı örüntülerin büyük ölçekli veri tabanlarından elde edilmesini sağlayan veri tabanlarında bilgi keşfi süreci için bir adımdır

Veri madenciliği veri ambarındaki gizli önemli bilgileri otomatik olarak algılayan, keşif odaklı bir veri analizi teknolojisidir (Pyo vd., 2002: 397). Sadece sınırlı insan katılımı gerektiren, makine öğrenme yöntemlerini (yapay sinir ağları, ilişki kuralları, karar ağaçları ve yapay zekâya dayanan genetik algoritmalar gibi), veriden örüntü veya bilgi elde etmek için kullanılmaktadır. Veri madenciliği, gerekli miktarda veri depolamak için gelişmiş teknolojileri, veri içindeki spesifik özelliklerin nasıl tanımlandığı ve açıklandığına dair yaklaşımlar sunan istatistikleri ve verilerdeki potansiyel olarak yararlı desenleri fark etmek için akıllı yaklaşımlar uygulayan yapay zekâ ve makine öğrenme teknikleri kullanan disiplinler arası bir alandır.

Verilerden çıkarımlar yapması bakımından benzer (Shapoval vd., 2018: 311) olmalarına rağmen veri madenciliği, geleneksel istatistiksel modellemeden çeşitli şekillerde farklılık göstermektedir (Magnini vd., 2003: 95). Veri madenciliği, genellikle büyük ölçekli veri kümelerine uygulanan istatistiksel ve / veya yapay zekâ analizini içerir. Geleneksel istatistiksel analiz, belirli bir beklenen sonuç kümesinin mevcut olduğu, genellikle yönlendirilen bir yaklaşımı içermektedir. Klasik istatistikler kestirimci modeller olarak bilinen, temel hedefleri parametreleri tahmin etmek ve / veya hipotezleri doğrulamak veya reddetmek olan büyük öznel bileşenlere sahiptirler (Shapoval vd., 2018: 312). İstatistiksel teknikler, herhangi bir analiz gerçekleştirilemeden önce verilerin dağılımı ile ilgili çeşitli varsayımların yapılmasını gerektirmektedir. Bu varsayımlar ihlal edilirse, sonuçların geçerli olacağına dair bir garanti yoktur (Bose, 2009: 937). Bundan dolayı klasik istatistiksel araçlar, neyin inceleneceğine dair önceden tasarlanmış bir fikre sahip olan eğitimli bir araştırmacı tarafından denetlenmektedir (Magnini vd., 2003: 95). Bu yaklaşım denetimli olarak adlandırılmaktadır (hipotez geliştirme ve test) (Olson ve Delen, 2008: 4).

Öte yandan veri madenciliği, yeni ve faydalı şeyler öğrenmektedir (Olson ve Delen, 2008:4) ve veri madenciliği açısından doğru model bilinmemektedir. Aslında, analizin amacı doğru modeli bulmaktır. Klasik istatistiklerde modeller belirlenirken, veri madenciliğinde ise bir dizi rakip model belirlenerek ve veri incelemesine göre seçim yapılmaktadır (Shapoval vd., 2018: 312). Karar ağacı analizinde olduğu gibi, bilgi keşfinin çoğu otomatik yollarla gerçekleştirilebilir. Bu yaklaşımda bilgi keşfi denetimsiz olarak adlandırılır (Olson ve Delen, 2008:4). Ayrıca, istatistiksel analizler genellikle bağımsız ve bağımlı değişkenler arasındaki ilişkilerin doğrusal olduğunu varsayar ki bu çoğu zaman geçerli değildir. Bu nedenle, bu varsayımların ihlali, yanlış ve yanıltıcı istatistiksel sonuçlara yol açabilir. Veri madenciliği teknikleri (ör., Sinir ağları) bu sınırlamaların üstesinden gelir ve bu varsayımların geçerli olmadığı durumlarda geleneksel istatistiksel analizlerden daha iyi performans gösterir (Magnini vd., 2003:95). İstatistiksel bir ön analiz ile konu ile ilgili ilişkiler gözden kaçırılabilir. Bununla birlikte, sadece

doğrulamak yerine bağımlılık hipotezleri kurarak, veri madenciliği teknikleri önemli bağlantıları ortaya çıkarmaktadır (Magnini vd., 2003: 95). İstatistiksel yöntemlerin bu sınırlaması, araştırmacıların, turizm veri analizi için makine-öğrenme tabanlı veri madenciliğini kullanmalarını sağlamıştır (Bose, 2009: 937).

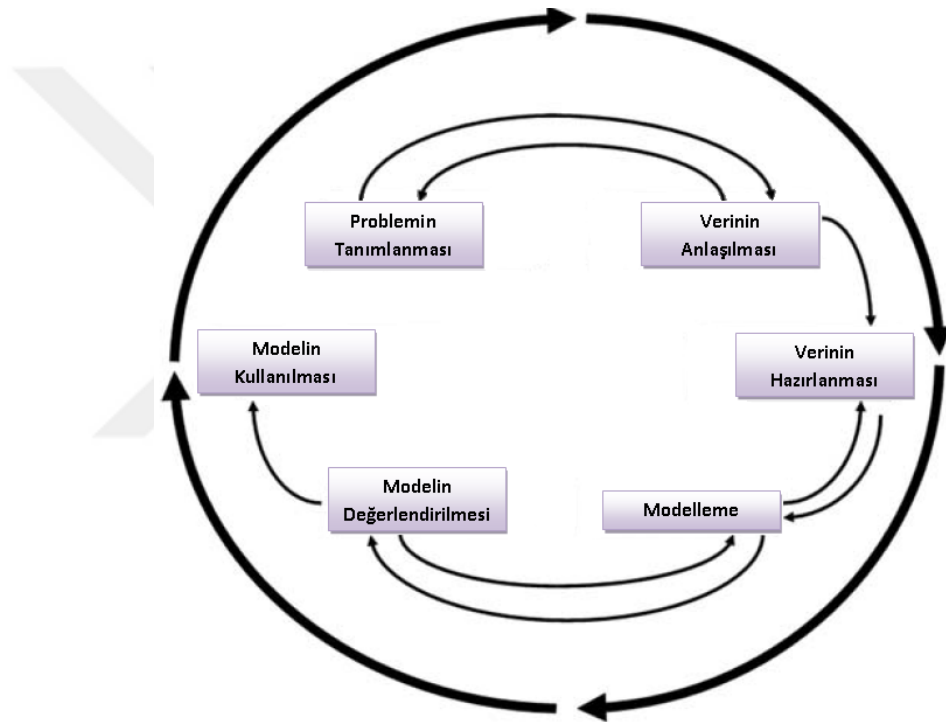
Geleneksel istatistiksel modellere göre bir başka önemli avantaj, veri madenciliğinin büyük ve karmaşık veri kümelerini kolayca işlemesidir (Magnini vd., 2003: 95). Çok büyük veri kümelerini yönetmek, klasik istatistiksel analizde kullanılanlardan farklı beceriler gerektirir. Veri madenciliği, bu tür problemleri büyük miktarlardaki verilerin verimli özetleriyle, önceki verilerin kalıplarını ve ilişkilerini tanımlayarak ve geleceğe yönelik öngörücüler oluşturarak yönetir (Shapoval vd., 2018: 311).

Veri madenciliği, performans, kullanım hızı ve kullanıcı dostu olma açısından da büyük kazanımlar sağlamaktadır (Magnini vd., 2003). Başarılı veri madenciliği için istatistiksel kavramların temel olarak anlaşılması zorunludur (Olson ve Delen, 2008). Veri madencileri istatistiksel ilkeleri anlamalıdır, ancak karar verme süreçlerini incelemek, anlamak ve iyileştirmek için son derece uzmanlaşmış bir istatistik bilgisi gerekli değildir (Magnini vd., 2003). Klasik istatistiksel yaklaşımlar veri madenciliği için esastır. Bununla birlikte, klasik istatistiksel yöntemlerle sistematik keşif hala veri madenciliğinin temelidir. İstatistiksel analiz alanı tarafından geliştirilen araçlardan bazıları, verilerle uğraşırken otomatik kontrol (bazı önemli insan rehberliği ile) yoluyla sağlanmaktadır (Olson ve Delen, 2008).

1.3. Veri Madenciliği Süreci

Veri madenciliği, bir problemin anlaşılmasından ve tanımlanmasından başlayarak, sonuçların analizi ve sonuçların strateji avantajı ile rekabet avantajı elde etmek için kullanılmasıyla sonuçlanan, kullanıcı tarafından yapılan birçok kararla ilgili çok sayıda adımı içeren etkileşimli ve yinelemeli bir süreçtir (Ha ve Park, 1998: 4). Veri madenciliği analizini sistematik bir şekilde yapmak için genellikle CRISP (Sektörler Arası Standart Süreç) ve SEMMA (Örnekleme, Keşfetme, Değiştirme, Modelleme ve Değerlendirme) gibi genel bir süreç takip edilir (Olson ve Delen, 2008: 9). Bu yaklaşımların her bir adımına her analizde ihtiyaç duyulmasa da, bu süreç, veri araştırması, veri toplama, veri işleme, analiz, değerlendirme ve uygulamadan başlayarak gereken adımların iyi bir şekilde kapsanmasını sağlar. Veri madenciliği sürecinin bir aşamasının sonucu bir sonraki aşamanın girdisi olduğu için her bir aşamasının dikkatle izlenmesi gerekmektedir. Bu çalışmada genellikle bir veri madenciliği çalışmasına katılan bir dizi adımdan oluşan bir endüstri standardı süreci olan Sektörler Arası Standart Süreç, CRISP tercih edilmiştir.

Veri Madenciliği için Sektörler Arası Standart Süreç (CRISP-DM), Şekil 1.1'de görüldüğü gibi altı fazdan oluşan bir yaşam döngüsüne sahiptir. Bir sonraki aşama genellikle önceki aşama ile bağlantılı sonuçlara bağlıdır. Aşamalar arasındaki en önemli bağımlılık oklarla gösterilmiştir. Örneğin, modelleme aşamasında, modelin davranışına ve özelliklerine bağlı olarak, model değerlendirme aşamasına geçmeden önce daha fazla ayrıntılandırma (rafine) için veri hazırlama aşamasına geri dönülmesi gerekebilir (Larose, 2006:265). CRISP – DM, veri madenciliğinin, probleminin tanımlanmasından, veri toplama ve yönetmeye, veri ön işleme, model oluşturma, model değerlendirme ve son olarak model dağıtımına kadar bütün bir süreç olarak görülmesini talep etmektedir. Altı aşama aşağıdaki gibidir:



Şekil 1.1 Veri Madenciliği Süreci

Kaynak: (Larose, 2006: 266)

Veri madenciliği uygulamalarını kullanmaya başlarken atılacak ilk adım problemi tanımlamaktır. Bu adım iş hedeflerini belirlemeyi, mevcut durumu değerlendirmeyi, veri madenciliği hedeflerini oluşturmayı ve bir proje planı geliştirmeyi içerir. İş hedefleri ve proje planı oluşturulduktan sonraki adım, verinin anlaşılması veri gereksinimlerini dikkate alır. Bu adım, ilk veri toplama, veri açıklaması, veri keşfi ve veri kalitesinin doğrulamasını içerebilir (Olson ve Delen, 2008: 9). Mevcut veri kaynakları tanımlandıktan sonra seçilmeleri, temizlenmeleri, istenen formata yerleştirilmeleri ve biçimlendirilmeleri gerekmektedir. Veri temizliği ve veri dönüşümü verinin hazırlanması aşamasında gerçekleştirilir.

Veriler anlaşılıp analiz için hazır hale getirildiğinde, işin amacına ve veri tipine uygun modeller uygulanır. Ardından model sonuçları, birinci aşamada belirlenen iş hedefleri bağlamında değerlendirilmelidir. İş anlayışı kazanmak, veri madenciliğinde, çeşitli görselleştirme, istatistiksel ve yapay zekâ araçlarının sonuçlarının, kullanıcılara örgütsel işlemlerde daha derin bir anlayış sağlayan yeni ilişkiler gösterdiği, yinelemeli bir prosedürdür. Bu, sıklıkla CRISP-DM'nin önceki aşamalarına geri dönerek diğer ihtiyaçların (genellikle örüntü tanıma yoluyla) tanımlanmasına yol açacaktır. CRISP-DM sürecinin daha önceki aşamalarında keşfedilen bilgiler sayesinde, daha sonra kilit durumların öngörülmesi veya tanımlanması da dâhil olmak üzere pek çok amaç için ticari faaliyetlere uygulanabilecek güvenilir modelleri elde edilebilir. Bu modellerin çalışma koşullarındaki değişimler için izlenmesi gerekir, çünkü bugün doğru olanlar, sonra doğru olmayabilir. Önemli değişiklikler olursa, model yeniden yapılmalıdır (Olson ve Delen, 2008: 10).

Bu altı aşamalı işlemde çok fazla geri izleme olur. Ek olarak, deneyimli analistlerin her çalışma için her aşamayı uygulamasına gerek olmayabilir. Ancak CRISP-DM, veri madenciliği için yararlı bir çerçeve sunmaktadır (Olson ve Delen, 2008: 11). Veri madenciliğinin birçok ticari faaliyetin geliştirilmesinde son derece etkili olduğu kanıtlanmıştır. Veri madenciliği süreci, veri depolama desteği (veri ambarları, veri martları ve / veya çevrimiçi analitik işleme araçları) ve ayrıca veri analiz yazılımı (veri madenciliği yazılımı) formundaki enformasyon teknolojilerine dayanmaktadır. Ancak, veri madenciliği süreci, bu veri madenciliği yazılım araçlarını bir firmanın verilerine uygulamaktan çok daha fazlasıdır. Model tiplerinin seçiminde, belirli bir problemle ilgili verilerin seçiminde ve dönüştürülmesinde ve sonuçların yorumlanmasında analistin zekâsına (aklına) ihtiyaç vardır (Olson ve Delen, 2008: 8).

1.3.1. Problemin Tanımlanması

Veri madenciliği çalışmalarında başarılı olmanın ilk şartı, uygulamanın işletmenin hangi amacı için yapılacağına açık bir şekilde tanımlanmasıdır. Bu, yeni bilgiye yönetsel ihtiyaç duyulması ve yapılacak çalışma ile ilgili iş hedefinin bir ifadesiyle başlar. İş hedefi, problem üzerine odaklanmalı ve elde edilecek sonuçların nasıl değerlendirileceği tanımlanmalıdır. “Ürünlerimizin her biriyle ne tür müşteriler ilgileniyor?” veya “Müşterilerimizin tipik profilleri nelerdir ve her birinin bize sağladığı değer nedir?” gibi hedefler açıkça belirtilmelidir (Olson ve Delen, 2008: 11). Müşteri değerini ve bağlılığını arttırmak isteyen bir turizmcinin tanıtımlara cevap verme olasılığı yüksek olan müşterileri belirlemek, yeni turistik ürün stratejileri geliştirmek veya sadık müşteriler oluşturarak pazar payını arttırmak gibi hedefleri bulunabilir.

Bu aşamada, veri madenciliği uygulamasının sonuçlarını etkileyebilecek önemli faktörler tespit edilmelidir. İlk önce işletmede üretilen sayısal verilerin boyutlarını, proje için yeterlilik düzeyini ve iş süreçlerini iyi analiz etmesi gerekmektedir. İşletme veri tabanları ya da yararlanılabilecek diğer veri kaynakları yapılacak çalışmanın veri kaynağı olacak ve buralarda saklanan gizli bilgilerin keşfi yapılacaktır. Onun için bu aşamada elde edilecek bilginin işletme için ne ifade ettiğinin doğru analiz edilmesi gerekmektedir. Yapılacak uygulama belirlenen hedeflere ulaşmaya çalışacağından problemin doğru bir şekilde tanımlanması önemlidir. Problemin yanlış belirlenmesi veri madenciliği çalışmasının yararlı ve faydalı sonuçlara ulaşmasını engelleyecek, hatta yanlış sonuçlara ulaşılmasına neden olacaktır. Uygulama ile doğru tahminlerde kazanılacak faydalara ve yanlış tahminlerde katlanılacak olan maliyetlere ilişkin tahminlere de bu aşamada yer verilmelidir.

1.3.2. Verinin Anlaşılması

Veri madenciliği görev odaklı olduğundan, farklı iş görevleri için farklı veri kümeleri gerekir. Veri madenciliği sürecinin ilk aşaması, belirli bir işletme görevini doğru şekilde tanımlamak için mevcut veri tabanlarından ilgili verileri seçmektir. Veri seçiminde dikkate alınması gereken en az üç konu vardır (Olson ve Delen, 2008: 11). Öncelikle problem kısa ve net bir şekilde tanımlanmalıdır. Örneğin bir destinasyon yönetimi veri madenciliği projesi, destinasyon içinde müşterilerin hareket ve harcama davranışlarını belirlemeye çalışabilir. İkinci konu, probleminin açıklaması için ilgili verileri tanımlamak olacaktır. Demografik, zamansal ve mekânsal veriler ve harcama verileri (turistin güzergâhı, oralarda yaptığı alışverişler, çektiği fotoğraflar, yaptığı paylaşımlar) analizler için seçilebilir. Üçüncü konu, ilgili veriler için seçilen değişkenlerin birbirinden bağımsız olması gerektiğidir. Değişkenlerin bağımsızlığı, değişkenlerin örtüşen bilgiler içermediği anlamına gelir. Bağımsız değişkenlerin dikkatli bir şekilde seçilmesi, veri madenciliği algoritmalarının hızlı bir şekilde faydalı bilgi kalıplarını keşfetmesini kolaylaştırabilir.

Veri seçimi için veri kaynakları değişebilir. Normalde, iş uygulamaları için veri kaynağı türleri demografik veriler (gelir, eğitim, hane sayısı ve yaş gibi), sosyo grafik veriler (hobi, kulüp üyeliği ve eğlence gibi) ve işlem verileri (satış kayıtları, kredi kart harcamaları, çekler) gibi verilerden oluşmaktadır (Olson ve Delen, 2008: 12). Veri türü nicel ve nitel olabilir. Sayısal değerler kullanılarak kantitatif veriler ölçülebilir. Ayrık (tamsayılar gibi) veya sürekli (gerçek sayılar gibi) olabilir. Kategorik veriler olarak da bilinen nitel veriler hem nominal hem de ordinal verileri içermektedir. Nominal veriler, iki değeri olan cinsiyet verileri (erkek ve kadın) gibi sıralı olmayan son değerlere sahiptir. Nicel veriler bir çeşit olasılık dağılımı ile kolayca

temsil edilebilir. Bir olasılık dağılımı, verilerin nasıl dağıldığını ve şekillendiğini açıklar. Örneğin, normal olarak dağıtılmış veriler simetriktir ve genel olarak çan şeklinde ifade edilir. Niteliksel veriler önce sayılara kodlanabilir ve daha sonra frekans dağılımları ile tanımlanabilir. Veri madenciliği işletme hedefine göre ilgili veriler seçildikten sonra, veri ön işleme (verinin hazırlanması) tabii tutulmalıdır.

1.3.3. Verilerin Hazırlanması

Veri madenciliği veri yığınları üzerinde yapılmaktadır. İşlem sonucunda elde edilecek bilgi, verinin ne derece problemin gereklerine uygun olarak hazırlandığı ile doğru orantılıdır. Dolayısıyla veri madenciliği süreçleri içinde en fazla zaman ayrılması ve en dikkatli olunması gereken süreçtir. Elde edilen verinin tamamının ön işlemeden geçirilmesi, amaca uygun olanların yeni bir veri tabanı ya da veri ambarında analize hazır hale getirilmesi gerekmektedir. Burada hazırlanan verinin kalitesi direkt sonuçları etkilemektedir. Bazen modelin kurulması esnasında yaşanan sorunlar, bu aşamaya sık sık geri dönülmesini ve verilerin yeniden düzenlenmesini zorunlu kılmaktadır.

Verilerin hazırlanması aşamasında yapılan ilk işlem verinin toplanmasıdır. Tanımlanan problem için gerekli olduğu düşünülen veriler ve bu verilerin toplanacağı veri kaynakları belirlenir. Verilerin toplanmasında işletmenin kendi veri kaynaklarından ve başka kuruluşların kullanıma açık veri tabanlarından faydalanılabilir. İlk aşamada belirlenen probleme uygun olarak, ilgili veri tabanından ve veri ambarından değişkenler seçilir. Burada verinin anlaşılması önem kazanmaktadır. Gereğinden az veri kaynağı veri madenciliği çalışmasını eksik bırakmakta ve istenen sonuçlara ulaşılmasını engellemekte, gereksiz veri kaynağı da sürecin uzamasına neden olabilmekte ya da veri kirliliğine yol açabilmektedir.

Modeller ancak iyi verilerin üzerine kurulduğunda iyi sonuç verecekleri için, toplanan verilerin ne ölçüde uyumlu oldukları veriyi toplamanın ardından incelenerek değerlendirilir. İşletmelerin günlük operasyonel işlemlerinde kullanılan büyük çaplı veri yığınları ilgili oldukları konuya özel değerler ve analiz esnasında kullanımı mümkün olmayan ham veriler içermektedir. Ham veri, içerisindeki tutarsızlıklar ve problemlerin giderilerek veri madenciliği uygulamalarına uygun hale gelmeleri amacıyla bir takım işlemlere tabii tutulmaktadır.

Farklı kaynaklardan toplanan verilerin uyumlu ve aynı biçimde olması beklenemez. Verilerin farklı zamanlara ait olmaları, kodlama farklılıkları ve farklı ölçü birimlerinin kullanılması uyumsuzluklardan bazılarıdır. Söz konusu uyumsuzluklar veri madenciliği uygulaması için bir veri ambarı altyapısı hazırlanırken giderilmiş olmaktadır. Veri tabanı veya veri ambar-

larında özet veya birbirleriyle bağlantılı bulunan veriler, daha anlamlı bir yapıya dönüştürülürler. Böylece düzensiz veri yığınları gerekli düzenlemeler yapılarak veri madenciliği uygulaması için daha işlevsel hale getirilmektedir. Veri ambarı kullanmadan da veri madenciliği yapılabilir. Fakat çok büyük miktarda veri içeren operasyonel veri tabanları üzerinde veri madenciliği uygulamalarını çalıştırmak performans problemlerine yol açabilmektedir.

Farklı kaynaklardan hatta bazen aynı veri tabanına ait farklı tablolardan toplanan verilerde uyumsuzlukların olması çok normaldir. Uyumsuzlukların giderilerek verilerin birleştirilmesi ve temizlenmesi gerekmektedir (Gürsoy, 2009: 33). Veri ön işlemenin amacı, seçilen verileri daha iyi kalite için temizlemektir. Seçilen bazı veriler, farklı veri kaynaklarından seçildiğinden farklı biçimlerde olabilir. Seçilen veriler düz dosyalardan, sesli mesajdan ve web metninden geliyorsa, tutarlı bir elektronik formata dönüştürülmeleri gerekir. Genel olarak, veri temizleme, eksik değerleri filtrelemek, birleştirmek ve doldurmak anlamına gelmektedir. Verileri filtreleyerek, seçilen veriler aykırı değerler ve fazlalıklar için incelenir. Aykırı değerler, verilerin çoğunluğundan veya açıkça seçilen veri gruplarının kapsama alanı dışındaki verilerden büyük ölçüde farklılık gösterir (Olson ve Delen, 2008: 12). Aykırı değerler, insan hataları veya teknik hatalar gibi birçok nedenden kaynaklanabilir veya doğal olarak aşırı olaylar nedeniyle bir veri kümesinde oluşabilmektedir.

Yedekli veriler, birkaç farklı şekilde kaydedilen verilerdir. Belirli bir ürünün günlük satışları mevcut ise, aynı ürünün mevsimsel satış verileri gereksizdir, çünkü satışlar günlük verilerden veya mevsimsel verilerden alınabilir. Toplanan verilerden bilgileri elde etmek için veri boyutları küçültülür (Olson ve Delen, 2008: 13). Rezervasyon kaydının girildiği bir tabloda hem doğum tarihi hem de yaş alanı, hem sezon bilgisi hem de tarih bilgisi bulunmaktadır. Analizin durumuna göre bu alanlardan birisi kullanılmayabilir. Bazen de farklı değişkenlerin birleştirilerek tek bir değişken olarak işleme alınmaları daha uygun olabilir.

Modelde kullanılan algoritma, verilerin gösteriminde önemli rol oynamaktadır. Örneğin bir uygulamada bir yapay sinir ağı algoritmasının kullanılması durumunda kategorik değişken değerlerinin evet/hayır olması modelin etkinliğini artıracaktır. Deneyimler, tüm giriş ve çıkış değerleri 0 ile 1 arasında olduğunda, sinir ağlarının en iyi sonucu verdiğini göstermiştir. Bu, en yaygın aktarım işlevinden dolayı 0 ile 1 arasında yeni değerler almak için, hem sürekli hem de kategorik olan tüm değerlerin normalleştirilmesini gerektirmektedir (Ha ve Park, 1998: 10). Aynı şekilde bir karar ağacı algoritmasının kullanılması durumunda ise örneğin, gelir değişken değerlerinin yüksek / orta / düşük olarak gruplanmış olması; kredi riskinin tahmini için geliştirilen bir modelde, borç/gelir gibi önceden hesaplanmış bir oran yerine, ayrı ayrı borç ve

gelir verilerinin kullanılması modelin etkinliğini artıracaktır. Çünkü karar ağacı normalleştirilmemiş değerler ile doğrudan çalışabilmektedir. Bazı karar ağaçları, kodlanmış kategorik özellikleri açık bir şekilde birden çok doğru veya yanlış özellik olarak hazırlamadan işleme yeteneklerine sahiptir (Ha ve Park, 1998: 10).

1.3.3. Modelin Kurulması

Veri modelleme, veri madenciliği yazılımının çeşitli durumlar için sonuçlar üretmek için kullanıldığı yerdir (Olson ve Delen, 2008: 15). Veri madenciliğinde veri kaynaklarından en fazla verimin elde edilebilmesi için modelin kurulması aşaması, analiz sonucunda elde edilecek sonuçların kalitesini etkileyeceğinden çok önemlidir. Eğer model doğru kurulmazsa veri seti içerisindeki önemli örüntüler ve yararlı bilgi tespit edilemez ve kritik ilişkiler doğru bir şekilde sunulamaz. Bu da modelden başarılı sonuç elde etme olasılığını azaltmaktadır (Gürsoy, 2009: 35).

Model kurma aşamasında ilk önce küme analizi ve verilerin görsel olarak keşfi uygulanır. Veri türüne bağlı olarak, daha sonra çeşitli modeller uygulanabilir. Görev verileri gruplandırmaksa ve gruplara verilenler varsa, ayırıcı analiz uygun olabilir. Eğer amaç tahmin ise, veri sürekli ise regresyon uygundur (ve değilse lojistik regresyon). Her iki görev için sinir ağları uygulanabilir. Karar ağaçları verileri sınıflandırmak için başka bir araçtır (Olson ve Delen, 2008: 15). Diğer modelleme araçları da mevcuttur ve veri tipine uygun daha detaylı modeller uygulanabilir. Sonraki bölümde bu farklı modeller daha ayrıntılı olarak anlatılmıştır.

1.3.4. Modelin Değerlendirilmesi

Tüm veri analizleri belirli bir koşul veya özniteliğin mevcut olduğunu yanlış bir şekilde gösterme potansiyeline sahiptir. Bir model yanlış sonuçlar vermese bile, modelde işlenen verilerde ilginç herhangi bir model bulunmayabilir. Bunun nedeni, modelin kalıpları bulmak için iyi ayarlanmamış olması, yanlış tekniğin kullanılmış olması ya da modelin bulması için verilerde ilginç bir durumun olmaması olabilir. CRISP-DM'nin Değerlendirme aşaması, modelin ne kadar değerli olduğunu ve onunla ne yapılacağını belirlemeye yardımcı olmaktadır (North, 2012: 10). Model sonuçları, problemin anlaşılması aşamasında belirlenen iş hedefleri bağlamında değerlendirilir. Bu, sıklıkla CRISP-DM'nin önceki aşamalarına geri dönerek diğer ihtiyaçların (genellikle örüntü tanıma yoluyla) tanımlanmasına yol açmaktadır. Probleminin çözümüne ulaşabilmek veri madenciliğinde, çeşitli görselleştirme, istatistiksel ve yapay zekâ araçlarının sonuçlarının, kullanıcılara örgütsel işlemlerde daha derin bir anlayış sağlayan yeni ilişkiler gösterdiği, tekrarlanan işlemleri içermektedir (Olson ve Delen, 2008: 10).

Tanımlanan problemin çözümüne ulaştıracak en uygun modeli bulabilmek için, olabildiğince çok sayıda modelin kurularak denenmesi gerekebilmektedir. En iyi çözüme ulaşabilmek için veri hazırlama ve model kurma aşamaları tekrarlanmaktadır. Kurulan her modelin amaca ne kadar hizmet ettiğinin değerlendirilmesi için bazı teknikler vardır. Bu teknikler her modelde kullanılarak ve sonuçlar değerlendirilerek en uygun olan modele ulaşmak için sayısız deneme yapılmasında yarar bulunmaktadır. Değerlendirme için kullanılacak tekniklerin ve diğer önlemlerin ötesinde, model değerlendirmesi aynı zamanda bir insani yönü içermelidir (North, 2012: 10). Bir modelin kullanılabilirliğini belirlemek için hem veri güdümlü hem de içgüdüsel değerlendirme teknikleri kullanılabilir.

Tarihsel veriler kullanılarak tahmin edici bir model geliştirildiğinde, modelin gelecek için (model oluşturma sürecinde görmemiş olduğu verilerde) nasıl bir performans göstereceği tanımlanan problemin çözümüne ne kadar katkı sağlayacağı açısından önemlidir. Bazen aynı problemin çözümü için birden fazla tahmin modeli kurulabilmektedir. Bu modeller arasında gerçek tahmin karar durumu için hangi modelin kullanılacağına, modellerin tahmin performansları (örneğin, doğruluk) karşılaştırılarak karar verilebilmektedir. Bir tahminin performansını ölçmek için kullanılan performans ölçütleri ve bu ölçütleri doğru değerlendirmenin bazı metodolojileri bulunmaktadır.

A) Performans Ölçütleri

Sınıflandırma problemlerinde, birincil performans ölçütleri kaynağı bir çakışım matrisidir (sınıflandırma matrisi veya olasılık tablosu). Şekil 1.2, iki sınıflı bir sınıflandırma problemi için çakışım matrisini göstermektedir. Köşegen sol üsten sağ alt köşeye doğru olan sayılar alınan doğru kararları temsil etmektedir ve bu köşegen dışındaki sayılar hataları gösterir. Bir sınıflandırıcının doğru pozitif oranı (aynı zamanda isabet oranı veya hassasiyet olarak da bilinir), doğru sınıflandırılmış pozitifleri (doğru pozitif sayısı) toplam pozitif sayıya bölerek tahmin edilmektedir. Sınıflandırıcının yanlış pozitif oranı (aynı zamanda yanlış alarm oranı olarak da bilinir) yanlış sınıflandırılmış negatifleri (yanlış negatif sayısı) toplam negatiflere bölerek tahmin edilmektedir. Bir sınıflandırıcının genel doğruluğu, doğru şekilde sınıflandırılmış toplam pozitifleri ve negatifleri toplam örnek sayısına bölmek suretiyle tahmin edilmektedir. Hassasiyet, kesinlik ve F-ölçümü gibi diğer performans ölçütleri, diğer toplu performans ölçütlerini hesaplamak için kullanılmaktadır (Olson ve Delen, 2008: 137).

		Doğru Sınıf	
		Pozitif	Negatif
Tahmin Edilen Sınıf	Pozitif	Doğru Pozitif Sayısı (TP)	Yanlış Pozitif Sayısı (FP)
	Negatif	Yanlış Negatif Sayısı (FN)	Doğru Negatif Sayısı (TN)

Şekil 1.2 Basit bir çakışım matrisi

Kaynak: Olson ve Delen, 2008: 137.

Çakışım matrisinden hesaplanabilen en yaygın kullanılan metriklerin denklemleri de aşağıda verilmiştir.

$$\text{Doğru Pozitif Oranı} = \frac{TP}{TP+FN}$$

$$\text{Doğru Negatif Oranı} = \frac{TN}{TN+FP}$$

$$\text{Doğruluk} = \frac{TP+TN}{TP+FP+TN+FN}$$

$$\text{Kesinlik} = \frac{TP}{TP+FP}$$

$$\text{Hassasiyet} = \frac{TP}{TP+FN}$$

$$F \text{ ölçütü} = 2 \cdot \frac{\text{Kesinlik} \cdot \text{Hassasiyet}}{\text{Kesinlik} + \text{Hassasiyet}}$$

Sınıflandırma problemi ikili olmadığında, çakışma matrisi biraz daha karmaşıklaşmaktadır. Bu durumda performans ölçütlerinin terminolojisi “sınıf başına doğruluk oranları” ve “genel sınıflandırıcı doğruluğu” ile sınırlandırılmıştır (Olson ve Delen, 2008: 139). Bu formüller aşağıda verilmiştir. Formüllerde i sınıf numarası, n ise toplam sınıf sayısıdır.

$$\text{Doğru Sınıflandırma Oranı} = \frac{(\text{Doğru Sınıflandırma})}{\sum_{i=1}^n (\text{Yanlış Sınıflandırma})}$$

$$\text{Genel Sınıflandırıcı Doğruluğu} = \frac{\sum_{i=1}^n (\text{Doğru Sınıflandırma})}{(\text{Toplam Gözlem Sayısı})}$$

B) Değerlendirme Metodolojisi

Bazı denetimli öğrenme algoritmaları tarafından indüklenen bir sınıflandırıcının doğruluğunu tahmin etmek üç nedenden dolayı önemlidir. İlk olarak, sınıflandırma sistemindeki sınıflandırıcının çıktısının güven seviyesini belirleyen gelecekteki tahmin doğruluğunu değerlendirmek için kullanılabilir. İkinci olarak, verilen bir setten bir sınıflandırıcı seçmek için kullanılabilir (iki veya daha fazla yeterlilik modelinden “en iyi” modelini seçmek). Son olarak, bir birleştirme sınıflandırıcısının sonucunun optimize edilebilmesi için birden fazla sınıflandırıcıya güven seviyeleri atamak için kullanılabilir. Kombine sınıflandırıcılar, bireysel tahminlere göre daha sağlam ve daha doğru tahminler üretmelerini gösteren deneysel sonuçlar nedeniyle gittikçe daha popüler hale gelmektedir. Bir sınıflandırıcının nihai doğruluğunu değerlendirmek için, düşük yanlılık ve düşük varyansa sahip bir tahmin yöntemi gerekmektedir. Bazı uygulama alanlarında, bir sınıflandırıcı seçmek veya sınıflandırıcıları birleştirmek için mutlak doğruluklar daha az önemli olabilmektedir. Sınıflandırma tipi veri madenciliği modelleri için kullanılan en popüler değerlendirme metodolojileri basit bölme, N katlı çapraz geçerlilik testi, Ön yükleme ve katlama (Bootstrapping and Jackknifing) yöntemleridir (Olson ve Delen, 2008: 140).

Basit Bölme

Basit bölme (holdout veya test örneği tahmini) verileri eğitim seti ve test seti (veya bekleme seti) olarak adlandırılan birbirini dışlayan iki alt gruba ayrılmaktadır. Verilerin $2/3$ 'ünü eğitim seti, geri kalan $1/3$ 'ünü test seti olarak belirlemek yaygındır. Eğitim seti, indükleyici (model üreticisi) tarafından kullanılır ve yerleşik sınıflandırıcı daha sonra test setinde test edilir. Sınıflandırıcının yapay bir sinir ağı olduğu bu kuralın bir istisnası vardır. Orada veriler birbirini dışlayan üç alt gruba ayrılmıştır; Eğitim, doğrulama, test etme. Doğrulama seti, model oluşturma sırasında aşırı uyumu önlemek için kullanılmaktadır.

Bu yöntemin ana eleştirisi, iki alt kümedeki verilerin aynı türden (aynı özelliklere sahip olduğu) olduğu varsayımına sahip olduğu gerçeğidir. Bu basit bir rastgele bölümlenme olduğu için, sınıflandırma değişkeninde verilerin çarpık olduğu çoğu gerçekçi veri setinde, böyle bir varsayım doğru olmayabilir. Bu durumu iyileştirmek için tabakanın çıktı değişkeni olduğu tabakalı örnekleme önerilmektedir. Bu basit bölünme üzerinde bir iyileşme olmasına rağmen, tek rastgele bölümlenmeyle ilişkili bir önyargıya sahiptir.

“K” Katlı çapraz geçerlilik testi

İki veya daha fazla yöntemin öngörme doğruluğunu karşılaştırmak için eğitim ve test verisi (holdout data) örneklerinin rastgele örnekleme ile ilişkili önyargıyı en aza indirmek için, k-katlı çapraz geçerlilik adı verilen bir metodoloji kullanılabilir. Döndürme değerlendirme

olarak da adlandırılan k-katlı çapraz geçerlilikte, tüm veri seti rastgele olarak yaklaşık eşit büyüklükteki birbirini dışlayan alt kümelere bölünmektedir. Sınıflandırma modeli k kez eğitilip, test edilmektedir. Her seferinde bir kat dışında hepsi eğitilir ve kalan tek kat üzerinde test edilir. Bir modelin genel doğruluğunun çapraz geçerlilik değerlendirmesi, k bireysel doğruluk ölçümlerinin ortalaması alınarak hesaplanmaktadır.

$$CVA = \frac{1}{k} \sum_{i=1}^k A_i$$

Formülde CVA, çapraz geçerlilik doğruluğunu ifade etmektedir, k, kullanılan kat sayısıdır ve A, her katın doğruluk ölçüsüdür (örneğin, isabet oranı, kesinlik, özgüllük vb.).

Çapraz geçerlilik doğruluğu, bireysel vakaların k farklı katlara rastgele atanmasına bağlı olacağından, ortak bir uygulama katların kendilerini katmanlaştırmaktır. Tabakalı k-kat çapraz geçerlilikte katlar, orijinal veri kümesi ile yaklaşık olarak aynı oranda tahmin edici etiketler (yani sınıflar) içerecek şekilde oluşturulur. Ampirik çalışmalar, tabakalı çapraz geçerliliğin, normal çapraz geçerliliğe oranla düşük sapma ve düşük varyans ile karşılaştırma sonuçları üretme eğiliminde olduğunu göstermiştir.

K-kat çapraz geçerlilik ayrıca 10 katlı çapraz geçerlilik olarak da bilinmektedir, çünkü 10 değerini alan k en yaygın kullanım şeklidir. Aslında, deneysel çalışmalar onun (10) en uygun kat sayısı olduğunu göstermiştir (bu, testi tamamlamak için gereken süreyi ve geçerlilik süreci ile ilgili sapma ve varyansı optimize etmektedir). 10 katlı çapraz geçerliliğin geleneksel tekli bölme (2/3 eğitim, 1/3 test) deneyine kıyasla daha fazla veri gerektirmemektedir. Aslında, veri madenciliği çalışmalarında, nispeten daha küçük veri kümelerine sahip yöntem karşılaştırma çalışmaları için, k-katlı çapraz geçerlilik yöntemleri önerilmektedir (Olson ve Delen, 2008: 143). Temel olarak, 10 kat (veya herhangi bir sayıda kat) çapraz geçerliliğin temel avantajı, her seferinde ayrı bir bölümünü kullanarak deneyi her seferinde 10 kez tekrarlayarak, eğitim ve test veri örneklerinin rastgele örneklenmesiyle ilişkili önyargıyı azaltmaktır. Bu metodolojinin olumsuz tarafı, sadece bir kez yerine, k kez eğitim ve test yapılması gerektiğidir.

1.4.2. Birliktelik Kuralları

Birliktelik Analizi veri tabanında sıklıkla birlikte veya aynı süre içerisinde gerçekleşen durumları bulmak üzere kullanılan metottur. Amaç, ortaklaşa giden değişkenleri belirlemektir. Bir öge kümesi arasında gruplandırma veya kalıp arayan bir araçtır (Karathiya vd., 2012: 115). Bununla birlikte, perakende endüstrileri için, birliktelik kuralı perakendecilerin müşteri alışverişini davranış kuralları oluşturmasına yardımcı olabilecek (Chiang, 2018: 667) ve veri

kümelerinden sık sık desenler, korelasyonlar, birliktelikler veya nedensel yapılar bulmak için yararlı bir veri madenciliği teknolojisidir. Aynı zamanda birliktelik kuralı, başkalarının oluşumuna bağlı olarak bir öğenin oluşumunu öngörecektir kuralları bulan bir dizi işlem vermektedir (Nuankaew vd., 2017: 94). Herhangi bir ürün alındığında bu ürünün yanında başka ürünün de satın alınması bir birliktelik kuralı vermektedir.

Birliktelik analizinden elde edilen yararlı kurallar, pazarlama stratejisini etkin bir şekilde uygulamak için pazarlama veya MİY sistemlerine uygulanabilir. Müşteri satın alma geçmişlerinden elde edilen bilgiler, sonraki satın alımlar için olasılık kurallarını formüle etmek için kullanılır (Magnini vd., 2003: 96). Örneğin, eğer bir turist Nepall'i ziyaret ederse, Hindistan'ı da ziyaret etmesi ihtimalinin 0.66 olması gibi muhtemel ifadeler üretmektedir (Karathiya vd., 2012: 115). Konaklama hizmeti satın bir müşterinin aynı alışverişte konaklamamanın yanında transfer alma olasılığı verir. Bu özelliğinden yararlanarak bir seyahat acentası tekne turu ile birlikte dalış satın alınma oranı yüksekse, acenta yöneticileri tekne turu ile dalış turunu aynı paket içine koyarak dalış turu satışlarını arttırabilirler (Akgün ve Çizel, 2017). Bu teknik lokanta operasyonlarında yiyecek ve içecek ürünlerinin çapraz satışını yapmak için fast-food zincirlerinde yaygın olarak kullanılmaktadır (Ting vd., 2010: 492). Bilinen diğer bir uygulaması, müşterilerin belirli öğelere göz atarken “ilişkili ürünler” ile sunulacağı çevrimiçi kitapçılardadır. Seçilen bir kitabın hemen ardından ilgili diğer kitaplar listelenmektedir.

Birliktelik analizi bir veri setindeki birden fazla özellik arasında gizlenmiş ilginç ilişkileri keşfetmek için bir veri madenciliği tekniğidir (Liu vd., 2013: 360). Olumlu ve olumsuz birliktelik olmak üzere iki birliktelik türü keşfedilebilmektedir. Olumlu birliktelik, bir kural biçiminde kesin olarak tanımlanır, örneğin $A \Rightarrow B$, şöyle yorumlanır: “Eğer A ön koşulu bir veri kaydında gerçekleşirse, o zaman büyük bir olasılıkla B koşulu aynı kayıta gerçekleşir”. A ve B, $A \cup B$ bir öğe seti oluşturduğu ve $A \cap B = \emptyset$ olduğu veri setinin iki öğesidir. Olumsuz birliktelik, “A bir veri kaydında bulunursa, B aynı kayıta OLMAYACAK” şeklinde ifade eden $A \Rightarrow B$ şeklinde sunulmaktadır. Soldaki A niteliği kuralın öncüsüdür ve sağdaki B niteliği bunun sonucudur yani, olay A (önceden belirlenmiş olay olarak adlandırılır) meydana gelirse, olay B (sonuç olay olarak adlandırılır) belirli bir olasılıkla gerçekleşir. Sembolik olarak, $A \Rightarrow B$ şeklinde yazılır ve A B yi gerektirir (ima eder, kasteder) şeklinde okunur.

Birliktelik kurallarının ana hatlarını belirleyen iki ana kavram vardır: doğruluk ve kapsam. Doğruluk, öncül doğruysa ardılın gerçekleşeceği ihtimalini ifade eder. Yüksek doğruluk, bunun son derece güvenilir bir kural olduğu anlamına gelir. Kapsam, kuralın uygulandığı veri tabanındaki kayıt sayısını ifade eder. Yüksek kapsam, kuralın çok sık kullanılabileceği anlamına gelir. Bazı literatürde kapsam için destek ve doğruluk için de güven

denir (Al-Salim, 2007:616). Destek, bir veri tabanından rastgele seçilen bir işlem grubunun A ve B maddelerini içermeye olasılığını temsil eden bir yüzde olarak ifade edilir (Ting vd., 2010: 493). Matematiksel olarak ifade edildiğinde formül $P(A \cap B)$ 'dir. Rastgele seçilen bir işlem kümesinden A'yı içerenlerin B'yi içermeye olasılığına güven denir ve $P(B | A)$, $(P(A \cap B) / P(A))$ formülüyle ifade edilmektedir.

Tablo 1.2 Turistlerin Aldığı Günlük Turlar

Turist	Aldığı Turlar
1	Pamukkale, Damlataş, Alanya Kalesi
2	Damlataş, Alanya Kalesi
3	Alış - veriş, Tekne Turu, Hamam
4	Alanya Kalesi, Hamam
5	Hamam, Alış veriş, Pamukkale
6	Tekne Turu, Hamam, Alanya Kalesi
7	Pamukkale, Alış veriş
8	Alış veriş, Hamam, Damlataş
9	Alanya Kalesi, Damlataş, Hamam
10	Alış - veriş, Tekne Turu, Hamam

Tablo 1.2'de turistlerin aldığı günlük tur kayıtları listelenmektedir. Bu kayıtlar kullanılarak yapılan birliktelik analizi uygulaması aşağıda verilmiştir.

Apriori algoritması kullanılarak birlikte satın alınan turların tespiti yapılacaktır. Bunun için;

Destek = % 40

Güven = % 50;

Destek %40 ve tüm müşteri sayısı 10 olduğuna göre eşik destek sayısı $0.40 \times 10 = 4$ dür.

Ürün kümesi={Hamam, Alış veriş, Pamukkale, Tekne Turu, Alanya Kalesi, Damlataş }

Destek sayıları

Hamam : 7

Alış veriş : 5

Pamukkale : 3

Tekne Turu : 3

Alanya Kalesi : 5

Damlataş : 4

Bu durumda göre eşik destek sayısı 4 ve üzeri olan ürünler çözümlenmeye dâhil edilecektir.

Cözümlemeye Katılacak Ürünler

Hamam : 7

Alış veriş : 5

Alanya Kalesi : 5

Damlataş : 4

Yukarıda ürünlerin tek tek alınması durumundaki destek değerlerine bakılmıştır. Aşağıda ürünlere ikili ve üçlü gruplar halinde bakıldığı durumda oluşacak destek değerleri verilmiştir.

İkili ürün gruplarının destek değerleri

Hamam, Alış veriş : 4

Hamam, Alanya Kalesi : 3

Hamam, Damlataş : 2

Alış veriş, Alanya Kalesi : 0

Alış veriş, Damlataş : 1

Alanya Kalesi, Damlataş : 3

Üçlü ürün gruplarının destek değerleri

Hamam, Alış veriş, Pamukkale : 1

Hamam, Alış veriş, Tekne Turu : 2

Hamam, Alış veriş, Alanya Kalesi : 0

Hamam, Alış veriş, Damlataş : 1

Destek eşik sayısı 4 olduğundan üçlü gruplardan hiçbir satırın bu koşula uymadığını görülmektedir.

Hamam, Alış veriş : 4

$$\text{Destek} = \frac{\text{sayı (Hamam, Alış veriş)}}{N} = \frac{4}{10} = 0,4$$

Eşik değerden küçük olmadığına göre bu kural kullanılabilir.

$$\text{Güven(Hamam} \rightarrow \text{Alış veriş)} = \frac{\text{Hamam, Alış veriş}}{\text{hamam}} = \frac{4}{7} = 0,57$$

Birliktelik Kuralı

Hamam \rightarrow Alış veriş, Hamam'ın bulunduğu ürün kümesinde Alış-Verişin bulunma olasılığı %57

1.3.5. Modelin Kullanılması ve İzlenmesi

Başlangıçta belirlenen sorular başarıyla tanımlanmış, bu soruları cevaplayabilecek veriler hazırlanmış ve ilginç ve yararlı olma testini geçen bir model oluşturulmuşsa, sonuçlar gerçekten kullanılma noktasına gelmiştir (North, 2012: 10). Veri madenciliği çalışması, orijinal veri madenciliği proje hedeflerine bağlanması gereken yeni bilgileri ortaya çıkarmıştır (Olson ve Delen, 2008: 18). Bu aşamadaki etkinlikler arasında modeli otomatik hale getirme, modelin çıktılarının tüketicileriyle buluşturma, mevcut yönetim veya operasyonel bilgi sistemleriyle bütünleşme, modelden yeni öğrenmeyi besleme, doğruluğunu ve performansını artırmak için tekrar modele geri dönme ve model kullanımının sonuçlarının izlenmesi ve ölçülmesi bulunmaktadır (North, 2012:10). Tahmin edilen ve gözlenen değişkenler arasındaki farklılığı gösteren grafikler model sonuçlarının izlenmesinde kullanılan yararlı bir yöntemdir.

Belirli bir veri madenciliği çalışmasından elde edilen bilgilerin değişime karşı izlenmesi önemlidir. Müşteri davranışı zaman içinde değişmektedir ve verilerin toplandığı dönem boyunca doğru olan şey çoktan değişmiş olabilir. Temel değişiklikler olursa, açığa çıkan bilgi artık doğru değildir. Bu nedenle, ilgilenilen alanın yayılma süresi boyunca izlenmesi kritik öneme sahiptir (Olson ve Delen, 2008: 18). Sistem özelliklerinde ve dolayısıyla ürettikleri verilerde zamanla ortaya çıkan değişiklikler, kurulan modellerin sürekli olarak izlenmesini ve gerekiyorsa yeniden düzenlenmesini gerektirebilmektedir.

Kurulan ve geçerliliği kabul edilen model doğrudan bir uygulama olabileceği gibi, bir başka uygulamanın alt parçası olarak kullanılabilir. Veri madenciliği modellerinin etkin bir şekilde konuşlandırılması için dengenin sağlanması gerekir. Bir modelin işlevini ve faydasını paydaşlara açıkça ileterek, modeli iyice test edip kanıtlayarak, daha sonra uygulanmasını planlayıp izleyerek veri madenciliği modelleri etkin bir şekilde kurumsal akışa dâhil edilebilir. Ancak konuşlandırmanın dikkatli ve etkili bir şekilde yönetilmemesi, en iyi ve en etkili modelleri bile batırabilir (North, 2012: 11).

1.4. Veri Madenciliği Yöntemleri

Temel olarak, birçok işletme için veri madenciliğinin amacı geçmiş deneyimlere ve şimdiki eğilimlere dayanarak kurum kapasitesini geliştirmek, anormal desenleri bulmak ve geleceği öngörebilmektir. Veriler arasındaki ilişkileri ortaya koymak ve gerektiğinde ileriye yönelik kestirimler de bulunmak için veri madenciliği belirli yöntemler kullanmaktadır. Birçok analitik yöntem, veri madenciliği araçları olarak sınıflandırılabilirken, veri madenciliği araç setinin bir parçası olarak hangi tekniklerin tam olarak ele alınması gerektiği konusunda fikir birliği sağlanamamıştır (Magnini vd., 2003: 95). Bu çalışmada en yaygın olarak kullanılan

yöntemlerin tanıtımı yapılmıştır. Genel olarak bu yöntemler tanımlayıcı ve tahmin edici olmak üzere iki ana başlık altında incelenmektedir (Rud, 2001, s. 4).

Tanımlayıcı modeller açıklayıcı kategoriler halinde nesnelere gruplamak için kullanılan kurallar oluşturmaktadır (Rud, 2001, s. 5). Veri setindeki bağıntı ve profilleri belirleyen ve yorumu kullanıcıya bırakan tanımlayıcı modeller, karar vermeye rehberlik etmede kullanılacak mevcut verilerdeki bağıntıların tanımlanmasını sağlamaktadır. Veri tabanı içindeki verilerden bilgi çıkarmak için kullanılırlar. Kümeleme (Clustering) ve Birliktelik Kuralları (Association Rules) modelleri tanımlayıcı modellerdir.

Sınıflandırma ve tahmin, önemli veri sınıflarını tanımlayan modelleri ayıklamak veya gelecekteki veri trendlerini tahmin etmek için kullanılacak iki tür veri analizidir (Han ve Kamber, 2006: 285). Tahmin edici modellerde sonuçlar bilinen verilerden hareket edilerek bir model geliştirilmesi ve kurulan bu modelden (Berry ve Linoff, 2004: 9) yararlanılarak, sonuçları bilinmeyen veri kümeleri için sonuç değerlerinin tahmin edilmesi amaçlanmaktadır (Han ve Kamber, 2006: 285). Modelde bağımsız değişkenler bağımlı değişkenin tahmin edilmesi için kullanılır.

1.4.1. Kümeleme

Kümeleme, heterojen bir popülasyonu birkaç homojen alt gruba veya kümeye ayırma görevidir (Berry ve Linoff, 2004: 11). Kümeleme, verilerdeki doğal gruplandırmayı keşfetmeye çalışan keşifsel bir tekniktir (Karathiya vd., 2012:115). Kümelemeyi sınıflandırmadan ayıran şey, kümelemenin önceden tanımlanmış sınıflara dayanmamasıdır. Sınıflama modelinde, verilerin sınıfları bilinmekte ve yeni bir veri geldiğinde bu verinin hangi sınıftan olabileceği tahmin edilmektedir. Kümelemede önceden tanımlanmış sınıflar ve örnekler yoktur. Kümeleme analizinde, temel olarak birbirine benzer özellik taşıyan gözlemlerin kümelenebilmesi ve benzer özellik göstermeyen öğelerin ayrıştırılması hedeflenmektedir (Çelik vd., 2017: 150). Kayıtlar kendi kendine benzerlik temelinde gruplandırılmıştır. Aynı kümedeki elemanlar birbirleriyle benzerlik gösterirken, başka kümelerin elemanlarından farklıdırlar. Belirli bir gözlemin belirli bir kümeye ait olması durumunda, özel bir kümeleme algoritmasında, başka bir kümeye dâhil edilemez. Uzaklığa dayalı kümelenede, iki veya daha fazla nesne belirli bir mesafeye göre “yakın” ise, bunlar aynı kümeye ayrılır (Bose, 2009: 936). Elde edilen kümelere ne gibi bir anlam çıkarılacağına belirlenmesi kullanıcıya bağlıdır (Berry ve Linoff, 2004: 11).

Kümeleme tekniklerinin temel amacı anlamlı veri yapılarını belirlemek olduğundan, metin madenciliği, istatistiksel öğrenme ve örüntü tanıma gibi birçok veri madenciliği uygulamasında merkezi rol oynamaktadır (Jun vd., 2018: 138). Kümeleme genellikle başka bir veri madenciliği veya modelleme biçiminin bir önışlemi gibi görev alabilmektedir. Örneğin,

kümeleme, pazar bölümlenme çalışmasının ilk adımı olabilmektedir (Berry ve Linoff, 2004: 11). Kümeleme genellikle pazar bölümlendirmesini yapmak ve farklı bölümlerin küme profilini tanımlamak için kullanılmaktadır. Pazarın nasıl bölündüğünü ve farklı bölümlerin özelliklerinin bilinmesi, kuruluşlar ve hedeflenen bölümlere ulaşmak için kullanılabilir yollar ve iletişim gibi kararlara yardımcı olmaktadır (Karathiya vd., 2012: 115). “Müşterilerin en iyi şekilde ne tür bir promosyona yanıt vereceği” sorusuna cevap ararken tek bir boyutta herkese uyan bir kural belirlemeye çalışmak yerine, müşteriler benzer satın alma alışkanlıklarına göre gruplandırılır ve ardından her grup için en iyi ne tür bir tanıtım yapılabileceğine bakılabilir.

Kümeleme veri madenciliği, istatistik, biyoloji ve makine öğrenimi gibi pek çok alanda kullanılır. Yapay öğrenme, yapay sinir ağları ve istatistikler dâhil olmak üzere çeşitli alanlarda kümeleme algoritmaları yaygın olarak çalışılmıştır. Literatürde pek çok kümeleme algoritması bulunmaktadır. Son birkaç on yılda, k-ortalamlar kümeleme ve karışım modelleri gibi geniş bir kümeleme algoritmaları ailesi geniş çapta incelenmiştir; Veri kümesi yapılarını ortaya çıkarmak ve görselleştirmek için kendi kendini organize eden harita yöntemleri ve çekirdek tabanlı kümeleme, vektör kümelemesini destekleme, spektral kümeleme ve grafik tabanlı kümeleme gibi kümeleme yöntemleri de doğrusal olmayan küme yapılarını yakalamak için geliştirilmiştir (Jun vd., 2018: 138). Kullanılacak olan kümeleme algoritmasının seçimi, veri tipine ve amaca bağlıdır.

1.4.3. Sınıflandırma

Sınıflandırma, veri öğelerini önceden belirlenmiş bir sınıf etiketine göre sınıflandıran kontrollü öğrenme tekniğidir (Septiadi vd., 2016: 211). Sınıflandırma, yeni bir nesnenin niteliklerini inceleme ve bu nesneyi önceden tanımlanmış bir sınıfa atamaktır. Sınıflandırma görevi, sınıfların iyi tanımlanmış bir tanımı ve sınıflandırılmış örneklerden oluşan bir eğitim seti ile karakterize edilir. Görev, sınıflandırılması için sınıflandırılmamış verilere uygulanabilecek bir tür model oluşturmaktır (Berry ve Linoff, 2004: 9). Sınıflandırma ve tahmin, önemli veri sınıflarını tanımlayan modelleri ayıklamak veya gelecekteki veri trendlerini tahmin etmek için kullanılabilir iki tür veri analizidir.

Sınıflandırma işlemi için istatistik ve makine öğrenimi temelli çeşitli yöntemler geliştirilmiştir (Akpınar, 2014: 69). Lineer regresyon analizi, lojistik regresyon analizi, diskriminant analizi ve bayes sınıflandırma istatistik temelli sınıflandırma yöntemleridir. Karar ağaçları, en yakın komşu yöntemi, yapay sinir ağları, destek vektör makineleri başlıca makine öğrenimi temelli sınıflandırma yöntemleridir.

Sınıflandırma ve tahmin yöntemleri 2. Bölümde (Sınıflandırma Algoritmaları) daha detaylı olarak anlatılacaktır.

İKİNCİ BÖLÜM

SINIFLANDIRMA YÖNTEMLERİ

Sınıflandırma günlük hayatın vazgeçilmezlerindedir, çünkü dünyayı anlayabilmek ve iletişim kurabilmek için sürekli olarak sınıflandırma yapılmaktadır (Berry ve Linoff, 2004: 8). Canlılar (hayvan, bitki vb.), hayvanlar (memeli, sürüngen vb.), insanlar (ırk, renk, din vb.) gruplandırılmaktadır. Sınıflandırma, yeni sunulan bir nesnenin özelliklerini incelemek ve onu önceden tanımlanmış bir sınıf kümesine atamaktan ibarettir. Bir hayvanın, insanın özelliklerine bakılarak, daha önce bilinen bir sınıfa atanmasıdır. Havanın bir sonraki gün nasıl olacağı ya da bir kutuda ne kadar kırmızı top olduğunun tahmin edilmesi aslında bir sınıflandırma işlemidir.

Veri sınıflandırması, benzer şekilde, yeni sunulan bir nesnenin bir veri tabanındaki özelliklerini tespit eden ve onu bir sınıflandırma modeline göre bir dizi önceden tanımlanmış sınıftan birine atayan işlemdir. Sınıflandırma görevi, sınıfların tanımı ve sınıflandırılmış örneklerden oluşan bir eğitim seti ve sınıflandırılması için sınıflandırılmamış verilere uygulanabilecek bir model oluşturma ile karakterize edilmektedir (Ha ve Park, 1998: 8). Sınıflandırılacak nesnelere genellikle bir veri tabanı tablosunda veya bir dosyada kayıtlarla temsil edilir ve sınıflandırma işlemi bir tür sınıf koduna sahip yeni bir sütun eklemekten oluşmaktadır (Berry ve Linoff, 2004: 9).

Sınıflandırma ve tahmin, önemli veri sınıflarını tanımlayan modelleri ayıklamak veya gelecekteki veri trendlerini tahmin etmek için kullanılacak iki tür veri analizidir. Bu tür bir analiz, verileri daha iyi anlamamıza yardımcı olabilir. Sınıflandırma kategorik (ayrık, sırasız) etiketleri öngörürken, tahmin modelleri sürekli değerli fonksiyonlardır (Han ve Kamber, 2006: 285). Belirli bir profile sahip bir müşterinin günlük tur satın alıp alamayacağını tahmin etmeye yardımcı olmak için ya da her hangi bir günlük tur için o gün tura çıkılıp çıkılmayacağını ya da müşterinin harcama gurubunu tahmin etmek için analizlere ihtiyaç vardır. Bu örneklerin her birinde, veri analizi görevi günlük tur satın alıp alamayacağı için “EVET” veya “HAYIR”, tur için “ÇIKILIR” veya “ÇIKILMAZ”; veya müşteri harcamaları için “Az”, “Orta” veya “Çok” gibi kategorik etiketleri öngörmek için bir model veya sınıflandırıcının yapıldığı sınıflandırma görevidir. Pazarlama yöneticisinin, belirli bir müşterinin konaklama süresince ne kadar harcama yapacağını ya da bir seyahat acentasından sezon boyunca ne kadar müşteri geleceğini tahmin etmek isteyebilir. Bu veri analizi görevi, sayısal modelin bir örneğidir, burada model, kategorik bir etiketin aksine sürekli değerli bir işlevi veya sıralı değeri tahmin eder. Bu model bir öngörücüdür. Regresyon analizi, sayısal tahminde çoğunlukla kullanılan istatistiksel bir metodolojidir, dolayısıyla iki terim sıklıkla eşanlamlı olarak kullanılır (Han ve Kamber, 2006:

286). Aslında sayısal tahmin için başka birçok yöntem kullanılabilir. Sınıflandırma ve sayısal tahmin, iki temel tahmin problemidir.

Sınıflandırma tekniği, veri öğelerini önceden belirlenmiş bir sınıf etiketine göre sınıflandıran kontrollü öğrenme tekniğidir (Septiadi vd., 2016: 211). Sınıflandırma, yeni bir nesnenin niteliklerini inceleme ve bu nesneyi önceden tanımlanmış bir sınıfa atamaktır. Burada önemli olan, her bir sınıfın özelliklerinin önceden net bir şekilde belirlenmiş olmasıdır. Sınıflandırma sonrasında her kategoride yer alan kayıtların, alanların, kişilerin, nesnelerin, kurumların özelliklerini ortaya çıkarmaktır. Bu sınıflar çeşitli müşteri grupları için oluşturulmuş ise her bir grubun özelliğini ortaya çıkarmak, profilini belirlemek en önemli amaçtır. Sınıflandırma bir ürünün özellikleri ile müşteri özelliklerinin eşleşmesi için kullanılarak bir müşteri için ideal ürün veya ürüne göre ideal müşteri profili çıkarılabilir. Müşteriler bu şekilde gruplandıktan sonra amaç, her bir grubun özelliklerini analiz etmek, profilini ortaya çıkarmak ve bu grupların özelliklerini, tutum ve davranışlarını içeren bir davranış geliştirebilmektir.

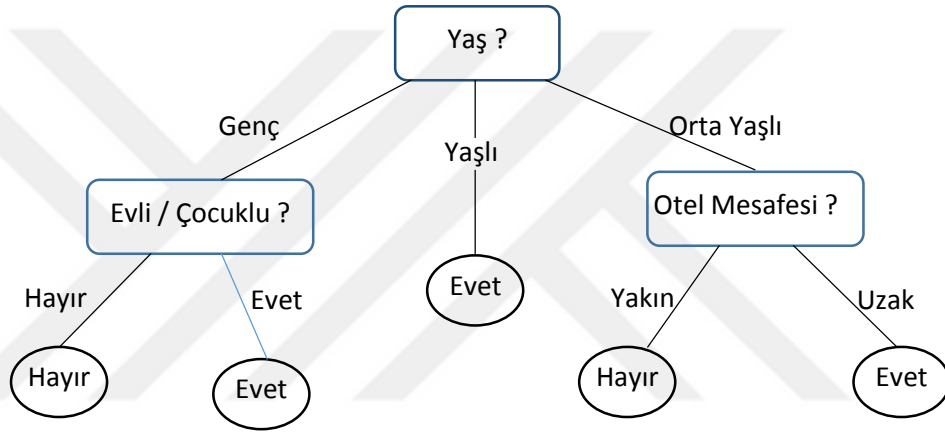
Bir müşteri kümelenme araştırması gerçekleştirildiğinde, müşteriler topluca benzer özellikler veya davranışlara sahip oldukları için belirli bir kümeye ait olurlar. Tahminli modelleme, kimin sorularını cevaplayan modelleri tanımlamaktadır. Örneğin, bir ürünü veya hizmeti kim satın alacak? Müşteri kaybı bağlamında, tahmin modelleri, kimlerin çalkalamasının muhtemel olduğu konusunda fikir verebilmektedir. Tahmin modelleri genellikle geçmişteki davranışlarına dayanarak gelecekteki müşteri davranışlarını tahmin etmektedir. Son olarak, tanımlayıcı ve tahmine dayalı modelleme genellikle veri madenciliği yaklaşımları kullanılarak gerçekleştirilmektedir (Bell ve Mgbemena, 2018: 197).

2.1. Karar Ağaçları

Seçenek sayısının fazla olduğu ve seçeneklerin gerçekleşmesinin belirli ya da belirsiz olduğu bir problem ile ilgili en iyi karara ulaşmak amacıyla bir takım yöntemlere veya araçlara gereksinim duyulmaktadır. Karar ağacı, ardışık aşamalarda karar almanın söz konusu olduğu problemlerin analizinde ve çözümlenmesinde karar vericilerin kullanabileceği popüler bir veri madenciliği tekniğidir. Gürsoy (2009), karar ağacını “olası tüm eylemlerin yönlerini, eylemlerin yönlerine etkisi olabilecek tüm olası faktörleri ve tüm bu faktörlere dayanan her bir olası sonucu, verilere bağlı olarak değerlendiren, çizgi, kare, daire gibi geometrik semboller kullanımı yoluyla karar vericiye sorunu anlamada kolaylık sağlayan düzenleme” biçiminde tanımlamıştır.

Karar ağaçları sınıflandırma ve tahmin problemlerinde sezgisel açıklanabilirlik özellikleri nedeniyle yaygın olarak kullanılan (Vajirakachorn ve Chongwatpol, 2017: 82), güçlü ve

popüler (Berry ve Linoff, 2004: 165) yöntemlerden birisidir. Karar ağacı tabanlı sınıflandırma, bir dizi örnekte karar ağaçlarını oluşturan denetimli bir öğrenme yöntemidir. Bir karar ağacı çeşitli olası soru dizilerini (testler), cevapları ve sınıflamaları gösteren yönlendirilmiş bir grafikdir (Ha ve Park, 1998: 8). Bir karar ağacının analiz sonuçları genellikle okuyucunun anlaşılmasını geliştirmek için kolayca anlaşılabilir ağaç dallanma grafikleriyle temsil edilmektedir (J. Wong vd., 2006: 362). Tipik bir karar ağacı Şekil 2.1'de gösterilmiştir. Konsept transfer satın almayı temsil etmektedir, yani, seyahat acentası müşterilerinin konaklama ile birlikte transfer satın alıp almayacağını tahmin etmektedir. İç düğümler dikdörtgenler ile gösterilir ve yaprak düğümleri ovaler ile gösterilmektedir (Han ve Kamber, 2006: 292).



Şekil 2.1 Transfer Satın Alma Konseptinde Bir Karar Ağacı

Bir karar ağacı, bir dizi basit karar kuralını uygulayarak, geniş bir kayıt koleksiyonunu birbirini izleyen küçük kayıt kümelerine bölmek için kullanılabilir bir yapıdır. Her ardışık bölünme ile sonuç kümelerinin üyeleri birbirine daha fazla benzer hale gelmektedir. Bir karar ağacı modeli, büyük heterojen bir popülasyonu belirli bir hedef değişkene göre daha küçük, daha homojen gruplara ayırmak için bir dizi kuraldan oluşur (Berry ve Linoff, 2004:166). Bir karar ağacının amacı, giriş değişkenlerine dayanarak bir hedef değişkeni tahmin eden bir model oluşturmaktır. Bir karar ağacındaki her yaprak, hedef değişkene bir bağlantı göstermektedir (Bell ve Mgbemena, 2018: 200). Hedef değişken genellikle kategoriktir ve karar ağacı modeli, verilen bir kaydın her bir kategoriye ait olma olasılığını hesaplamak veya kaydı en olası sınıfa atayarak sınıflandırmak için kullanılmaktadır (Berry ve Linoff, 2004:166).

Karar ağaçları, farklı ölçüm seviyelerinde çok sayıda bağımsız değişkeni ele alacak şekilde tasarlanmıştır. Karar ağaçları, bağımlı değişkeni bağımsız değişkenlere bölerek grupları tanımlamaktadır. Karar ağacı analizinin sonuçları, bağımlı değişkenle en çok hangi bağımsız değişkenlerin ilişkisi olduğunu göstermektedir (Byrd ve Gustke, 2007: 179). Karar ağaçları

ayrıca sürekli bir değişkenin değerini tahmin etmek için kullanılabilir (Berry ve Linoff, 2004:166). Karar ağaçları ayrıca bir hedef değişkene sahip bir dizi potansiyel girdi değişkeni arasında gizli ilişkiler bulabilmektedir. Buna ek olarak, bir karar ağacı veri araştırmalarını ve modellemeyi birleştirebilir, bu nedenle modelleme sürecinde ilk adım olarak (Septiadi vd., 2016:211) ve diğer veri madenciliği araçlarından daha az veri hazırlığı gerektirirler (Shapoval vd., 2018: 316).

Karar ağaçları, ikili sınıflandırma bağlamında pek çok avantaja sahiptir ve bunlar, yorumlanması ve anlaşılması nispeten kolay olduğu için popülerdir (Bell ve Mgbemena, 2018: 202). Karar ağacı sınıflandırıcılarının oluşturulması herhangi bir alan bilgisi veya parametre ayarı gerektirmemektedir ve bu nedenle de bilgi keşfi için uygundur. Karar ağaçları yüksek boyutlu verileri işleyebilir. Ağaç biçimindeki edinilmiş bilginin temsili sezgiseldir ve genellikle insanlar tarafından asimile edilmesi kolaydır. Karar ağacı induksiyonunun öğrenme ve sınıflandırma adımları basit ve hızlıdır (Han ve Kamber, 2006: 292). Farklı ölçüm seviyelerinde ölçülmüş olan ortak değişkenleri idare etme kabiliyetine sahiptir ve hem sayısal hem de kategorik verileri işleyebilirler. Ancak, karar ağaçları bazı sınırlamalara sahiptir; Yüksek derecede kararsızlığa sahiptirler. Verilerdeki küçük bir değişiklik, eğitilmiş modeli onaylarken genellikle tahmin kalitesini etkileyen farklı bölünme serileriyle sonuçlanabilir (Bell ve Mgbemena, 2018: 202). Genel olarak, karar ağacı sınıflandırıcıları iyi bir doğruluğa sahiptir. Bununla birlikte, başarılı kullanım eldeki verilere bağlı olabilir. Karar ağacı induksiyon algoritmaları tıp, üretim ve üretim, finansal analiz, astronomi ve moleküler biyoloji gibi birçok uygulama alanında sınıflandırma için kullanılmıştır. Karar ağaçları, çeşitli ticari kural induksiyon sistemlerinin temelidir (Han ve Kamber, 2006: 292).

Bir karar ağacı, her iç düğümün (yapraksız düğüm) bir öznelik üzerinde bir testi ifade ettiği, her dalın, testin dışında bir durumu temsil ettiği ve her yaprak düğümünün (veya terminal düğümü) bir sınıf etiketi içerdiği bir akış şeması benzeri ağaç yapısıdır. Bir ağaçtaki en üstteki düğüm kök düğümüdür (Han ve Kamber, 2006: 291). Kök düğümünü seçtikten sonra, bir sonraki adım kök düğümünden çıkan dallara bakmaktır ve hangi çocuk düğümü ile karşılaşacağını belirlemek için bir test uygulanır. Bu testler için farklı algoritmalar vardır, ancak amaç her zaman aynıdır. Bu işlem, veriler artık bölünemez hale gelinceye kadar tekrarlanır, yani ağaç son yaprak düğümüne doğru büyümüştür (J. Wong vd., 2006: 362). Ağaç budama işlemi, hata eğilimli dalların ortadan kaldırılmasını içerir. Budanmış bir ağaç, bir sınıflandırıcının performansını artırabilir ve modelin bilgi edinimi amacıyla daha fazla analizini kolaylaştırabilir. Budama işlemi, sınıflandırıcının öngörücü parçalarını asla çıkarmamalıdır (Bell ve Mgbemena, 2018: 201).

2.1.1. Turizmde Karar Ağaçları

Karar ağaçları en açık yorumlanabilirliğe sahip sınıflandırma yöntemi olduğundan turizmde hem araştırmacılar hem de diğer turizm profesyonelleri tarafından anlaşılması daha kolaydır. Bu özelliğinden dolayı turizmde en yaygın kullanılan yöntemlerden birisidir. Turizm sektöründe yerel halkın sınıflandırılması, müşteri segmentasyonu ve müşteriye uygun ürünlerin tespit edilmesi ve tahmin işlemlerinde bazen tek genelde başka bir veri madenciliği yöntemiyle birlikte kullanılmıştır. Diğer sınıflandırma yöntemleri ile test edildiği çalışmalarda bazen en iyi sonucu vermese de (Delen ve Sirakaya, 2006), birçok çalışmada en yüksek tahmin verimliliğini ve en açık yorumlanabilirliği (Xie vd., 2003) göstermektedir.

Turizm planlayıcılarının turizm planlama sürecine paydaşları dâhil etmesi için bu paydaşların ve çıkarlarının belirlenmesi gerekmektedir. Byrd ve Gustke, (2007) paydaşları tanımlamak için karar ağacı (CHAID) kullanmıştır. Karar ağacının sonuçlarından, bölgelerindeki sürdürülebilir turizm gelişimi için verdikleri destekler açısından paydaş grupları belirlenmiştir. Alexander vd. (2018) doğal kaynaklar açısından zengin olan bölgelerde doğa temelli turizm faaliyetlerinin hızla artmasının o bölgede yaşayan yerel halkın geçim kaynaklarına etkisini araştırdıkları çalışmalarında hane halkının sınıflandırılması için kümeleme ile birlikte CART algoritmasını kullanmışlardır. Geleneksel liderlik yapılarıyla geçim kaynakları ilişkisinin belirlendiği çalışmada arazi düzenlenmesindeki planlamalarının turizm girişimlerinden önce yapılması, kapsayıcı toplum konsensüsünün geliştirilmesi ve arazi tahsislerinin bir topluluğun daha savunmasız ve sesini duyuramayan kısımlarının ihtiyaçlarına göre yapılması önerilmektedir.

Karar ağaçlarının turizmde kullanım alanlarından birisi de turistlerin sınıflandırılmasıdır. Ha ve Park, (1998) müşterileri “sadık” veya “kaçabilecekler” olarak sınıflandırmak için karar ağacından (C4.5) yararlanmıştır. Sadık bir ziyaretçi ilişkisi kurmak ve sürdürmek için ziyaretçilerin tercihlerini ve ziyaretçilerle etkileşimlerini anlamak için Karathiya vd. (2012) de karar ağaçlarını kullanmışlardır. Elde ettikleri sonuçlar Gujarat Devleti Turizm Bakanlığı tarafından seçilmiş ziyaretçilere uygun tasarlanmış tanıtım materyalleri tahsis etmek için kullanılacaktır. Nuankaew vd. (2017) turizm cazibe merkezlerinin turistlere tanıtılması ve teşvik edilmesi için gerekli önemli faktörleri keşfetmek için veri madenciliği yöntemi tercih etmişlerdir. Faktörlerin keşfi karar ağaçları ve birliktelik kurallarının kullanımı ile yapılmış ve turistik yerler ile ilgili 7, devlet politikaları ile ilgili 4 ve bireysel davranış ile ilgili 3 olmak üzere 14 önemli faktör belirlenmiştir. Yotsawat ve Srivihok, (2014) turist segmentlerini oluşturmak için karar ağacı ile iki seviyeli kümeleme uygulamıştır. Deneysel sonuçlar, karar ağacı uygulamasının kümelenmelerin kalitesini artırabileceğini göstermiştir.

Turizm sektöründe iş zekâsı ve iş analitiğinin etkilerini araştıran Vajirakachorn ve Chongwatpol (2017), Tayland'daki bir yerel yemek festivalinde bir vaka çalışması yapmıştır. Yapılan çalışma yalnızca “ziyaretçilerin festival sitesini tekrar ziyaret etme niyetini etkileyen en önemli faktörler nelerdir” araştırma sorusuna cevap vermekle kalmayıp, aynı zamanda ziyaretçilerin yerel festival etkinliğinde yaşadıkları hizmetler ile ilgili memnuniyetlerini etkileyen kilit özellikleri anlamada da yardımcı olmaktadır. Bevanda vd. (2008)’nin ID3 algoritması kullanarak kullanıcıların web sitesi tasarımı kalite faktörleri algısını keşfettikleri çalışmalarında elde ettikleri sonuçlar da yönetim ve web tasarımcılarına iş performanslarını iyileştirmek ve geliştirmek için yararlı bilgiler sağlamaktadır.

Pyo vd. (2002), veri madenciliği analizine ihtiyaç duyulduğunu ve piyasa bilgilerinin yanı sıra turistler ve destinasyon hakkındaki bilgilerin dağıtımına uygulanması gerektiğini vurgulamıştır. Shapoval vd. (2018)’de bir destinasyon olarak Japonya’ya gelen turistlerin gelecekteki destinasyon pazarlaması amacıyla davranışlarını analiz etmek için karar ağaçlarını kullanmışlardır. J. Wong vd. (2006) Kuzey Tayvan turistlerinin seyahat modellerini analiz etmek için üç veri madenciliği tekniğinin kullanımını önerdiler. Araştırmacılar, değerli yolcuları tanımlamak için RFM (Yenilik, Sıklık ve Parasal Değer) uygulamış, daha sonra seyahat modelini ve kuralını keşfetmek için C4.5 karar ağacı ve birliktelik kuralı uygulamıştır.

Karar ağaçları turizm sektöründe tahmin amaçlı da kullanılmıştır. Septiadi vd. (2016) kullanarak uçuş biletleri ve otel odaları satış verilerinden tahmin modelleri oluşturmuştur. Delen ve Sirakaya (2006), üç popüler veri madenciliği yönteminin (yapay sinir ağları, karar ağaçları ve kaba kümeler) etkinliklerini, turizm politikası amacıyla toplanan kumar oyu verilerini kullanarak karşılaştırmıştır. Bu karşılaştırmalı çalışma için altmış benzersiz tahmin modeli oluşturulmuş ve modeller arasından, kaba kümeler, yapay sinir ağlarının ardından karar ağaçları (% 76.7) en iyi tahmin aracı olduğunu göstermektedir. Xie vd. (2003)’nin çalışmasında da karar ağaçları en yüksek tahmin verimliliğini ve en açık yorumlanabilirliği göstermektedir.

Tablo 2.1’de turizm alanında yapılan çalışmalarda karar ağacı yöntemini kullanan çalışmalardan bazıları ve karar ağacı yöntemini kullanma amaçları verilmiştir.

Tablo 2.1 Karar Ağacı Algoritmalarının Kullanıldığı Çalışmalar

Araştırma	Kapsam	Algoritma Kullanım Amacı
Alexander vd., 2018	Kuzey Botswana - Chobe Bölgesinde hane düzeyinde, anket formu kullanılarak toplanan veri	Kümeleme ile gruplandırılan geçim (bağımlılığı) ile reislik arasındaki ilişkiyi değerlendirmek
Bevanda vd., 2008	Pula'daki Juraj Dobrila Üniversitesinden 30 öğrencinin üç çevrimiçi seyahat acentesinin web sitesini karşılaştırmaları	Kullanıcıların web sitesi tasarımı kalite faktörleri algısını keşfetmek
Nuankaew vd., 2017	Wat Nong Waeng ve Wat Phrathat Nadun ziyaretinde bulunan turistler	Turizm cazibe merkezlerinin turistlere tanıtılması ve teşvik edilmesi için gerekli önemli faktörleri keşfetmek

Vajirakachorn ve Chongwatpol, 2017	Tayland'ın Pattaya kent merkezindeki bir festivale katılanlara ait demografik veriler ile itme ve çekme faktörleri, algılanan değerler, algılanan kalite ve memnuniyet ile ilgili toplam 30 özellik	Festival alanlarını tekrar ziyaret etme niyetinin tahmin edilmesi
Shapoval vd., 2018	Japonya Turizm Ajansı için üstlenilen turist harcamaları anket serisinin bir parçası olarak, Japonya'daki uluslararası havalimanlarında ve limanlarda toplanan veri	Japonya'da gelen turistlerin davranışsal ve satın alma modellerini özetlemek ve yorumlamak için karar ağaçları (iki bağımlı değişken ile) kullanılmıştır.
Byrd ve Gustke, 2007	Kuzey Carolina'nın doğusunda iki kırsal bölgedeki turizm paydaşları (İlçe sakinleri, işletme sahipleri, yerel kamu görevlileri ve turistler)	Turizm paydaşlarının turizm gelişimine ilişkin tutumları ve algılarının, ayrıca yerel politik ve turizm faaliyetlerine katılımları ve turizmin çevresel etkilerine bakışlarının tespit edilmesi
Karathiya vd., 2012	Gujarat'ı en az beş günlük tur paketleri ile ziyaret eden 1000 ziyaretçinin görüşü	Sadık bir ziyaretçi ilişkisi kurmak ve sürdürmek için ziyaretçilerin tercihlerini ve ziyaretçilerle etkileşimlerini anlamak
J. Wong vd., 2006	Kuzey Tayvan'dan yurt dışına çıkan turistlerin demografik özellikleri; karar verme bilgileri; ve ziyaret süresi dahil olmak üzere ziyaret ettikleri yerler hakkında toplanan veri	Demografik bilgiler, satın alma ve karar verme davranış modelleri ve ziyaret edilen yerler gibi değerli turistlerin turist özelliklerini keşfetmek
Septiadi vd., 2016	8853 adet bilet satış ve 4912 adet otel satış verisi	Uçuş biletleri ve otel odaları satış verilerinden tahmin modelleri oluşturmak ve en ucuz uçuşları veya otel odalarını aramak için harcanan zamanı azaltmak
Ha ve Park, 1998	Gümrüksüz satış mağazasına üye olan müşteriler	Mağazanın müşterilerini sadık veya kaçabilir olarak sınıflandırmak
Min vd., 2002	Güney Kore, Seul'deki 11 lüks otelin müşterilerinin demografik verileri ve otel servislerinden memnuniyet seviyeleri	Mevcut müşteri veri tabanını belirli segmentasyon türlerine göre sınıflandırmak ve seçilen belirli bir otel, oda ve ödeme seçenekleri vb. de müşteri davranışlarını tahmin etmek
Xie vd., 2003	San Francisco Bay Area Seyahat Araştırması (BATS) 2000 içerisinde 34.680 katılımcı içeren toplam 15.064 konut (dokuz ilden) rastgele örnekledi.	Veri madenciliği yöntemlerinin çalışma seyahat modu seçim modellemesi üzerindeki yeteneklerini ve performansını incelemek
Üngüren ve Doğan, 2010	Alanya bölgesinde bulunan 23 adet beş yıldızlı konaklama işletmelerindeki çalışanların iş tatmin düzeyini etkileyen faktörler	Konaklama işletmelerinde çalışanların iş tatmin düzeylerine ait demografik değişkenlerin iş tatmini üzerindeki etkilerini bir bütün olarak sınıflandırmak
Boz vd., 2018	Beş farklı otele ait toplam 38.826 kayıttan oluşan otel rezervasyon verisi	Otellerin gelecekteki rezervasyonlarının iptal durumlarını tahmin etmek

2.1.2. Karar Ağacı Algoritmaları

Karar ağaçları oluşturmak için bir dizi algoritma kullanılmaktadır. Bu algoritmalarından bazıları CART (sınıflandırma ve regresyon ağaçları), Otomatik ki-kare etkileşim dedektörü (CHAID), iteratif dikotomizer (ID3) ve ID3'ün ardılı olan C4.5 algoritmalarıdır (Bell ve Mgbemena, 2018: 202). 1970'lerin sonlarında ve 1980'lerin başında, makine öğreniminde araştırmacı olan J. Ross Quinlan, ID3 (Iterative Dichotomiser) olarak bilinen bir karar ağacı algoritması geliştirmiştir. Bu çalışma, E. B. Hunt, J. Marin ve P. T. Stone tarafından tanımlanan

konsept öğrenme sistemleri üzerinde daha önceki çalışmalarda genişlemiştir. Quinlan, daha sonra yeni denetlenen öğrenme algoritmalarının sıklıkla karşılaştırıldığı bir kriter haline gelen C4.5'i (ID3'ün devamı) sunmuştur (Han ve Kamber, 2006: 292). 1984'te bir grup istatistikçi (L. Breiman, J. Friedman, R. Olshen ve C. Stone), ikili karar ağaçlarının oluşumunu tanımlayan Sınıflandırma ve Regresyon Ağaçları (CART) adlı kitabı yayınladılar. ID3 ve CART aynı zamanda birbirlerinden bağımsız olarak icat edilmiştir, ancak eğitim ağaçlarından gelen karar ağaçlarını öğrenmek için benzer bir yaklaşım izlemiştir.

ID3, C4.5 ve CART, karar ağaçlarının yukarıdan aşağıya tekrarlanan bir bölme-ve-fethetme tarzında oluşturulduğu geri izlemesiz bir yaklaşımı benimsemektedir. Karar ağacı indüksiyonu için çoğu algoritma, bir dizi eğitim seti ve bunlarla ilişkili sınıf etiketleri ile başlayan yukarıdan aşağıya doğru bir yaklaşımı izlemektedir. Eğitim seti, ağaç inşa edildiğinden, daha küçük alt kümelere tekrar tekrar bölümlendirilmektedir. Bunu yaparken, giriş gözlemini iki veya daha fazla alt gruba ayıran değişken için bir değişken ve karşılık gelen eşik tanımlamak için matematiksel algoritmalar (ör., Bilgi kazancı, Gini İndeksi ve ki-kare testi) kullanılmaktadır. Bu adım, tüm ağaç düğümünde tam ağaç oluşturuluncaya kadar tekrarlanmaktadır. Ayırıştırma algoritmasının amacı, ortaya çıkan iki veya daha fazla alt grubun homojenliğini (sırasını) maksimize eden bir değişken eşik çifti bulmaktır. Bölme için en yaygın kullanılan matematiksel algoritma entropi tabanlı bilgi kazanımını (ID3, C4.5, C5'de kullanılır), Gini İndeksini (CART'da kullanılır) ve ki-kare testini (CHAID'de kullanılır) içermektedir (Delen ve Sirakaya, 2006: 317).

2.1.2.1. ID3 (iteratif dikotomizer) Algoritması

ID3 algoritması, karar ağaçları ile sınıflandırma işlemlerini yerine getirmek üzere Quinlan tarafından geliştirilen algoritmalarından birisidir. ID3 algoritması entropiye dayalı bölümlenmeyi kullanmaktadır (Özkan, 2008: 54). Algoritma sadece daha önce bölen olarak kullanılmamış öznitelikleri seçerek ağacın yapraklara doğru homojenleşmesine çalışmaktadır. Algoritmada yeni dallara ayrılma, seçilecek tüm özniteliklerin tüketilmesi durumunda ve seçilen özneliğin kategorisinde tanımlanan kritere uygun gözlemlerin tüketilmesi durumunda sona ermektedir (Akpınar, 2014:221).

Bir sistemdeki belirsizliğin ölçüsüne entropi denir (Özkan, 2008: 55). Bir S kaynağında $\{m_1, m_2, \dots, m_n\}$ olmak üzere n adet mesaj üretilebilmesi durumunda m_i mesajlarının üretilme olasılıkları P_i 'dir. $P = \{P_1, P_2, \dots, P_n\}$ olasılık dağılımına sahip mesajları üreten S kaynağının entropisi $H(S)$ aşağıdaki gibidir.

$$H(S) = - \sum_{i=1}^n P_i \log_2(P_i)$$

T veri tabanını X testine göre bölmekle elde edilen bilgileri ölçmek için “kazanç ölçütü” adı verilen bir ifadeye başvurulmaktadır (Özkan, 2008: 58). Bu ölçüt aşağıdaki formül ile ifade edilmektedir.

$$\text{Kazanç}(X, T) = H(T) - H(X, T)$$

2.1.2.2. C4.5 ve C 5 Algoritması

C4.5, bilgi teorisinden türetilen ID3 algoritmasından geliştirilmiştir. C4.5, şube sınırlaması olmaksızın sürekli bağımlı değişkenler ile ilgilenme avantajına sahiptir (J. Wong vd., 2006: 362). CLS ailesinin özelliklerini bünyesinde barındıran C4.5, en iyi bölen öznelik seçiminde entropi ve enformasyon kazancı kriterlerini kullanmaktadır.

C4.5 algoritmasında tüm örnekler aynı sınıfa ait olduğunda, bir yaprak düğüm oluşturularak, bu düğüm bu sınıfın adı ile isimlendirilmektedir. Özneliklerin hiç birisinin enformasyon kazancına sahip olmaması ve daha önce görülmemiş bir sınıfın ortaya çıkması durumunda, C4.5 sınıfın beklenen değerini kullanarak bir karar düğümü oluşturmaktadır (Akpınar, 2014: 221).

C4.5 ve C5.0 algoritmaları ID3 algoritmasının bazı eksik yönlerini gidermek için geliştirilmiştir. C4.5 algoritmasından 1997 yılında karar ağaçlarının ölçeklenebilirliğini büyük ölçüde geliştiren C5.0 geliştirilmiştir (Septiadi vd., 2016: 211). C5.0 algoritmasının C4.5'e göre tahmin doğruluğu daha da artırılmış ve özellikle büyük sayıda verilerden oluşan veri setlerinde kullanılmaktadır. C5.0, bellek ve hesaplama süresi açısından C4.5'ten daha etkilidir. Bazı durumlarda C5.0 tahmin performansındaki artışa ek olarak, C4.5 için gerekli olan yarım saat, bir saatlik işlem hızını sadece 3.5 saniyede gösterebilmektedir.

2.1.2.3. CART veya C&RT (Sınıflandırma ve Regresyon Ağaçları) Algoritması

CART - Sınıflandırma ve Regresyon Ağaçları - en çok kabul gören ve kullanılan karar ağacı yöntemlerinden birisidir. CART algoritması 1984 yılında Breiman vd. (aktaran: J. Wong vd., 2006: 362) geliştirilmiştir.

CART'ın avantajları diğerlerinden farklı olarak parametrik olmayan doğasını, aykırı değerlere sahip olma yeteneğini, ilgilenilen değişkenleri seçme yeteneğini ve daha önemlisi değişkenleri tekrar kullanma yeteneğini içermektedir. Bu avantajlar, etmen ve kuralların ortaya çıkarılması için etkili bir yaklaşım sunmaktadır (Bell ve Mgbemena, 2018: 202). CART hem ayrık hem de sürekli bağımlı değişkenlerle ilgilenebilir, ancak sadece bir ikili ağaç

oluşturmaktadır (J. Wong vd., 2006: 362). Algoritma veriyi ilk düğümde ve her iç düğümde bir önceki düğümde daha homojen (çıktı değişkenlerinin benzer değerler almadığı) düğümler elde edecek şekilde ikiye böler. Homojenlik en küçük kareler sapması ölçütü ile değerlendirilir. Ayırma işlemi homojenlik kriteri sağlanıncaya veya başka bir durdurma kuralına (zaman, düğümdeki minimum kayıt sayısı vb.) ulaşıncaya kadar devam etmektedir. Büyük verilerle çalışıldığında sonuç bulması uzun sürmektedir (Gürsoy, 2009: 75).

2.1.2.4. CHAID (Otomatik Ki-kare Etkileşim) Algoritması

Kass (1980) tarafından geliştirilen (akt. Byrd ve Gustke, 2007: 179) CHAID - Chi-squared Automatic Interaction Detection algoritması, veriyi bağımlı değişkendeki varyasyon grupları içinde minimum, gruplar arasında ise maksimum olacak şekilde farklı alt gruplara ayırmayı hedeflemektedir. CHAID, AID (otomatik etkileşim tespiti)'nin (Sonquist ve diğerleri, 1971) bir uzantısıdır ve sınırsız dalları ve ayrık ve sürekli bir tahmin edici türü olması avantajlarıdır. CHAID'in sınırlaması, yalnızca ayrık bağımlı değişkenlerle ilgilenmesidir (J. Wong vd., 2006: 362).

Nominal ve ordinal değişkenler için ki-kare analizleri kullanılır ve aralık ve oran değişkenleri için bir varyans analizi kullanılır (Byrd ve Gustke, 2007: 179). Bölme, hangi değişkenin en düşük p değerine sahip olduğuna dayanır. İki veya daha fazla değişken arasında bir bağ oluşursa, en yüksek F değerine sahip değişken, tahmin edici olarak seçilmektedir. CART algoritmasının aksine CHAID düğümleri ikiden fazla alt gruba bölebilmektedir. CHAID yönteminin gücü, farklı ölçüm düzeylerinden oluşan büyük veri kümelerindeki kalıpları tespit etmeye çalışırken bulunur. CHAID, her ölçüm seviyesinde bağımsız ve bağımlı değişkenleri ele alabilir. Ayrıca, tüm bağımsız değişkenlerin aynı ölçüm seviyesinde olması gerekmez. Diğer bir deyişle, bir veri seti, nominal, kategorik, aralık ve oran olan bağımsız değişkenlerden oluşuyorsa, CHAID karar ağacı analizi, analizdeki tüm değişkenleri içerebilmektedir.

2.1.2.5. Diğer Karar Ağacı Algoritmaları

Yukarıdaki bölümlerde detaylı olarak anlatılan karar ağacı algoritmalarından başka literatürde kullanılan başka algoritmalar da mevcuttur. Burada bazıları özetlenmiştir.

Bu algoritmalarından birisi QUEST algoritmasıdır. 1997 yılında Loh ve Shih tarafından geliştirilen Quest (Quick, Unbiased, Efficient, Statical Tree) algoritması ikili büyüyen bir algoritmadır (Gürsoy, 2009: 76). Quest algoritmasının CART algoritmasına benzer avantajları vardır, ancak ikili olduğu için ağaç çok geniş olabilir ve yavaş büyüyebilir.

Bir başka karar ağacı algoritması 1995 yılında Tin Kam Ho tarafından önerilen Rastgele Orman (Random Forest - RF) algoritmasıdır (Boz vd., 2018). Hem sınıflandırma hem de regresyon problemlerinde kullanılabilen bu algorithmada bir karar ağacı yerine birden fazla karar ağacı türetilerek daha güçlü sonuçlar elde edilebilmektedir. Rastgele Orman (Random Forest) karar ağacı, veri setinde en iyi niteliklerden seçilen düğümleri dallara ayırmak yerine, her bir düğümden rastgele alınan niteliklerin en iyisini seçerek tüm düğümleri dallara ayırmaktadır. Her veri kümesi asıl veri setinden yer değiştirmeli olarak üretilmektedir. Rastgele özellik seçimi kullanılarak ağaçlar geliştirilir ve budama işlemi yoktur (Akçetin ve Çelik, 2014).

Hem sınıflandırma hem de regresyon problemlerinde kullanılabilen algoritmalarından birisi de Gradyan Arttırılmış Karar ağaçları (GradientBoostedTrees) algoritmasıdır. Topluluk yöntemlerinden bir tanesi olan Gradyan Arttırılmış Karar ağaçları algoritması, türevlenebilir kayıp fonksiyonlarının artırma genelleştirilmesi için kullanılan sınıflandırma performansı yüksek ve etkili bir algoritmadır. Bu algoritma kullanılarak oluşturulan modeller, web arama sıralaması ve ekoloji dahil olmak üzere çeşitli alanlarda kullanılmaktadır.

Rastgele Ağaç (Random Tree) her düğümden belli bir sayıda rastgele seçilmiş özellikleri alarak ağaç oluşturan bir sınıflama algoritmasıdır. Rastgele Orman algoritmasında olduğu gibi budama işlemi yapılmamaktadır. Ayrıca tutulan veri setine dayalı sınıf olasılıklarının tahminine izin veren bir opsiyonu bulunmaktadır (Akçetin ve Çelik, 2014).

2.1.3. Örnek Uygulama

Tablo 2.2’de seyahat acentasının rezervasyon kaydını yaptığı turistlerin listesi bulunmaktadır. Turistlerden bazıları günlük tur satın almışlar (EVET), bazıları da satın almamışlardır (HAYIR). Bu örnek verileri kullanarak ID3 algoritması yardımıyla oluşturulan karar ağacı örneği aşağıdaki gibidir.

Tablo 2.2 Turistlerin Tur Satın Alma Durumu

Turist	Yaş Durumu	Gelir düzeyi	Evli -Çocuklu	Tur Satın Alma
1	Genç	Orta	Hayır	Evet
2	Genç	Yüksek	Hayır	Evet
3	Orta	Orta	Evet	Hayır
4	Orta	Yüksek	Evet	Evet
5	Genç	Yüksek	Hayır	Hayır
6	Genç	Orta	Hayır	Hayır
7	Orta	Yüksek	Evet	Evet
8	Orta	Yüksek	Evet	Hayır
9	Genç	Orta	Evet	Hayır

Verilen tabloda Tur Satın Alma özelliği hedef sınıf değerlerini içermektedir. Tur Satın Alma özellik değerlerinden oluşan küme aşağıdaki gibidir.

Tur Satın Alma= {Evet, Evet, Hayır, Evet, Hayır, Hayır, Evet, Hayır, Hayır}

Yukarıdaki kümede |Tur Satın Alma|= 9'dur. Beş adet "Hayır" için ve dört adet "Evet" için olasılık değerleri aşağıda verilmiştir.

$$P_1 = \frac{5}{9} \text{ Hayır olasılığı}$$

$$P_2 = \frac{4}{9} \text{ Evet olasılığı}$$

Olasılık Dağılımı $P_{\text{Tur Satın Alma}} = \left(\frac{5}{9}, \frac{4}{9} \right)$ şeklindedir.

$$H(\text{Tur Satın Aldı}) = - \sum_{i=1}^n P_i \log_2(P_i)$$

eşitliğini kullanarak Tur Satın Alma kümesi için entropi aşağıdaki gibi hesaplanır.

$$H(\text{Tur Satın Alma}) = - \left(\frac{5}{9} \log_2 \frac{5}{9} + \frac{4}{9} \log_2 \frac{4}{9} \right) = 0,99$$

1. Dallanma

Karar ağacının oluşturulmasında seçilecek özelliğin belirlenebilmesi için her bir özellik için kazanç ölçütleri belirlenir.

A) Gelir düzeyi için kazanç ölçütü

Gelir (yüksek): 5

Gelir (Orta): 4

Gelir (yüksek) Entropisi: 0,97

Gelir (Orta) Entropisi: 0,81

Gelir Entropisi = 0,90 > $H(\text{gelir, tur satın alma}) = \frac{5}{9} \times 0,97 + \frac{4}{9} \times 0,81 = 0,54 + 0,36 = 0,90$

Gain=Kazanç= 0,99-0,90=0,09

B) Yaş düzeyi için kazanç ölçütü

Yaş (Orta): 4

Yaş(Genç):5

Yaş (Genç) Entropisi: 0,97

Yaş (Orta) Entropisi: 1

Yaş Entropisi = 0,98 > $H(\text{Yaş, tur satın alma}) = \frac{5}{9} \times 0,97 + \frac{4}{9} \times 1 = 0,54 + 0,36 = 0,98$

Gain=Kazanç= 0,99-0,98=0,01

C) Evli –Çocuklu için kazanç ölçütü

Çocuklu (Evet): 5

Çocuklu(Hayır):4

Çocuklu (Evet) Entropisi: 0,97

Çocuklu (Hayır) Entropisi: 1

Çocuklu Entropisi =0,98 > H(Yaş, tur satın alma): $\frac{5}{9} \times 0,97 + \frac{4}{9} \times 1 = 0,54 + 0,36 = 0,98$
Gain=Kazanç= 0,99-0,98=0,01

Gelir, Yaş, Evli-Çocuklu özelliklerinin kazançları karşılaştırıldığında en yüksek kazancın 0,09 ile Gelir özelliğinde olduğu görülmektedir. Bu durumda ilk dallanma için Gelir özelliği kullanılacaktır.

2. Dallanma

İkinci dallanma Gelir özelliğinin durumları için yapılacaktır. Gelir düzeyinde “Yüksek” ve “Orta” olmak üzere iki farklı durum söz konusudur.

Adım 1: Gelir düzeyi “Yüksek” için dallanma

Bu aşamada Gelir Düzeyi özelliğinin “Yüksek” değeri için alt karar ağacı düzenlenecektir. Bu değer için düzenlenen tablo aşağıda görülmektedir. Tablo hazırlandıktan sonra bu değer için Tur Satın Alma entropisini hesaplamak gerekmektedir.

Tablo 2.3 Gelir Düzeyinin "Yüksek" Olduğu Durumlar

Turist	Yaş Durumu	Gelir düzeyi	Evli -Çocuklu	Tur Satın Alma
2	Genç	Yüksek	Hayır	Evet
4	Orta	Yüksek	Evet	Evet
5	Genç	Yüksek	Hayır	Hayır
7	Orta	Yüksek	Evet	Evet
8	Orta	Yüksek	Evet	Hayır

$$H(\text{Tur Satın Alma}) = - \left(\frac{3}{5} \log_2 \frac{3}{5} + \frac{2}{5} \log_2 \frac{2}{5} \right) = 0,97$$

a) Yaş için kazanç ölçütü

Yaş (Orta): 3

Yaş(Genç):2

Yaş (Genç) Entropisi: 1

Yaş (Orta) Entropisi: 0,70

Yaş Entropisi =0,966 > H(Yaş, tur satın alma): $\frac{3}{5} \times 1 + \frac{2}{5} \times 0,70 = 0,50 + 0,46 = 0,88$

Gain=Kazanç= 0,97-0,88=0,09

a) Evli -Çocuklu için kazanç ölçütü

Çocuklu (Evet): 3

Çocuklu (Hayır):2

Çocuklu (Evet) Entropisi: 0,918

Çocuklu (Hayır) Entropisi: 1

Çocuklu Entropisi =0,966 > H (Çocuklu, tur satın alma): $\frac{3}{5} \times 0,918 + \frac{2}{5} \times 1 = 0,50+0,46=0,95$

Gain=Kazanç= 0,97-0,95=0,02

Değerlere bakılarak en yüksek kazancın Evli Çocuklu özelliğini seçilerek elde edilebileceği söylenebilir.

Adım 2: Gelir düzeyi “Orta” için dallanma

Gelir Düzeyi özelliğinin “Orta” değeri için alt karar ağacı düzenlenecektir. Bu değer için düzenlenen tablo aşağıda görülmektedir. Tablo hazırlandıktan sonra bu değer için Tur Satın Alma entropisini hesaplamak gerekmektedir.

Tablo 2.4 Gelir Düzeyi "Orta" için oluşturulan tablo

Turist	Yaş Durumu	Gelir düzeyi	Evli -Çocuklu	Tur Satın Alma
1	Genç	Orta	Hayır	Evet
3	Orta	Orta	Evet	Hayır
6	Genç	Orta	Hayır	Hayır
9	Genç	Orta	Evet	Hayır

$$H (\text{Tur Satın Alma}) = - \left(\frac{1}{4} \log_2 \frac{1}{4} + \frac{3}{4} \log_2 \frac{3}{4} \right) = 0,81$$

a) Yaş için kazanç ölçütü

Yaş (Orta): 1

Yaş(Genç):3

Yaş (Genç) Entropisi: 0,918

Yaş (Orta) Entropisi: 0

Yaş Entropisi =0,966 > H (Yaş, tur satın alma): $\frac{3}{4} \times 0,918 + \frac{1}{4} \times 0 = 0,50+0,46=0,688$

Gain=Kazanç= 0,81-0,68=0,12

a) Evli -Çocuklu için kazanç ölçütü

Çocuklu (Evet): 2

Çocuklu (Hayır):2

Çocuklu (Evet) Entropisi: 0

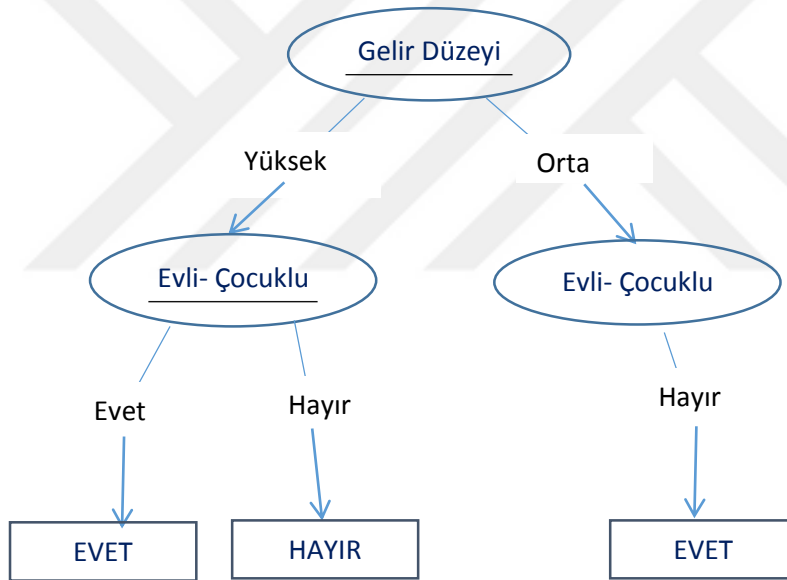
Çocuklu (Hayır) Entropisi:0

Çocuklu Entropisi =0,966 > $H(\text{Çocuklu, tur satın alma}): \frac{2}{4} \times 0 + \frac{2}{4} \times 0 = 0+0=0$

Gain=Kazanç= 0,97-0=0,97

Değerlere bakarak en yüksek kazancın Evli Çocuklu özelliğini seçerek elde edilebileceğini söyleyebiliriz. Bu durumda karar ağacı

Şekil 2.2'deki gibi olmaktadır.



Şekil 2.2 Örnek Uygulama İçin Karar Ağacı

2.2. Bayes Ağları

Bayes ağları, anahtar değişkenler arasındaki neden-sonuç ilişkilerini haritalayarak belirli bir alandaki bilgi gövdesini düzenleyen diyagramlardır (Olmeda ve Sheldon, 2001: 6). Başka bir deyişle, bir bayes ağı, ilgilenilen değişkenler arasındaki olasılıksal ilişkileri kodlayan grafiksel bir modeldir (Heckerman, 1997: 79). Bayes ağları, makine öğrenmesi için teşhis gibi performans görevlerinde faydalı olduklarından dolayı umut verici bir sunum sağlar: açıkça herhangi bir induksiyon görevinde merkezi problem olan belirsizlik ve gürültü konularıyla ilgilenmektedirler (John ve Langley, 1995: 338).

Bayes ağı, istatistiksel tekniklerle birlikte kullanıldığında, grafik modelin veri modelleme için birkaç avantajı vardır (Heckerman, 1997: 79). Birincisi, model tüm değişkenler arasındaki bağımlılıkları kodladığından, Bayes ağları eksik veri setlerini kolayca yönetebilir. İkincisi, bir Bayes ağı nedensel ilişkileri öğrenmek için kullanılabilir ve bu nedenle bir problem alanı hakkında bilgi edinmek ve müdahalenin sonuçlarını tahmin etmek için kullanılabilir. Üçüncüsü, modelin hem nedensel hem de olasılıksal bir semantiği olduğu için, önceki bilgileri (genellikle nedensel biçimde gelir) ve verileri birleştirmek için ideal bir sunum şeklidir (gösterim). Dördüncüsü, Bayes ağları ile bağlantılı olarak Bayes istatistik yöntemleri, verilerin aşırı yüklenmesinden kaçınmak için etkili ve ilkeli bir yaklaşım sunmaktadır.

Bayes ağlarını ve ilgili veri madenciliği tekniklerini anlamak için, Bayes olasılık ve istatistik yaklaşımını anlamak önemlidir (Heckerman, 1997: 80). Bayes teoremi olasılıklar üzerinde önemli bir yere sahiptir. Bayes teoremine dayanarak sınıflandırma işlemi yapmak mümkündür. B_1 ve B_2 ortak noktaları olmayan iki farklı olay ve A olayını bu iki olay cinsinden ifade etmek için koşullu olasılık bağıntısından yararlanılmaktadır (Özkan, 2008: 168). Temel küme B_1, B_2, \dots, B_n olaylarında oluşuyorsa ve olasılıkları sıfırdan farklı ise bir A olayının B_j olayı içinde gerçekleşme olasılığı aşağıdaki formül ile hesaplanmaktadır.

$$P(B_i / A) = \frac{P(A \cap B_i)}{P(A)} = \frac{P(A/B_i) P(B_i)}{\sum_{i=1}^k P(A/B_i) P(B_i)}$$

2.2.1. Turizmde Bayes Ağları

Tekne turu ya da yat turu düzenlemeyi planlayan bir turizmci tura çıkmadan önce hava durumu, fırtına gibi turun sağlıklı olarak yürütülmesine engel olacak olaylar hakkında bir araştırma yapmakta ve tecrübesine de dayanarak gerekirse turu ertelemektedir. Aynı şekilde bir golf turu için hava durumu, sıcaklık, nem ve rüzgâr durumu verileri elde edilerek her bir kıstasın golf oyununun oynanıp oynanması üzerindeki etkilerinin olasılık olarak hesaplanması yapılmaktadır. Bir turizmci ya da oyuncu kendi beyninde analizi yaparak oyunun oynanmayacağına ya da tura çıkılamayacağına karar vermektedir.

Yukarıdaki örneklerde de görüldüğü gibi aslında bayes yaklaşımı turizmde genel olarak karar verme aşamasında kullanılmaktadır. Daha önce öğrenilen durumlara göre yeni karşılaşılan durum için karar verme işlemi bayes sınıflandırıcının yaptığı işlem ile aynıdır. Naïve Bayes sınıflandırmasında sisteme belirli bir oranda sınıfı/kategorisi belli olan veri sunulmaktadır. Öğretilmiş veriler üzerinde yapılan olasılık işlemleri ile sisteme sunulan yeni test verileri, daha önce elde edilmiş olasılık değerlerine göre işletilmekte ve verilen test verisinin hangi kategoride olduğu tespit edilmeye çalışılmaktadır. Bu özelliği bu algoritmanın turizm

araştırmalarında yaygın olarak kullanılmasının önünü açabilecektir. Bayes ağları turizm araştırmalarında sınıflandırma, filtreleme ve tahmin için kullanılmaktadır.

Tahmin çalışmaları turizm alanında en yaygın olarak yapılan çalışmalardan birisini oluşturmaktadır. Bayes ağları da tahmin için tercih edilen yöntemlerden birisi olmuştur. Knoblock, (2004), tarihsel hava durumu bilgileriyle tarihsel uçuş gecikme verilerini ilişkilendiren bir tahmin edici geliştirmiştir. Tahmin edici tarihsel veri kaynaklarından öğrenen bir Naive-Bayes sınıflandırıcıdır. Tahmin, haftanın günü, günün saati, havayolu, kalkış ve iniş havaalanları ve hava durumu tahminlerine dayanmaktadır. Sistem, belirli bir uçuş ve tarih girişi olarak almakta ve daha sonra kalkış ve iniş havaalanlarında bu tarih için hava tahminlerini kontrol etmektedir. Günlük hava tahminini öğrendikleri ile birleştiren sistem, “Uçuşunuzun, 15 dakikadan fazla bir gecikme olması ve % 75'lik bir güven ile olması beklenmektedir” gibi ifadeler sunabilmektedir. Benzer bir tahmin çalışmasını Arabi vd. (2018) gelecek yıllarda ulaştırma durumunu (talep kapasitesinin oranı) tahmin edebilmek için yapmıştır. Çalışmalarının sonuçlarına göre, farklı bölgelerin ve bölümlerin trafik tasarımı için kullanılacak bir arz ve talep modeli önermiştir. K. Wong vd. (2006) sınırsız VAR (various vector autoregressive) modeline farklı kısıtlamalar getirerek üç Bayes VAR modeli geliştirmiş ve tahmin doğruluğunda önemli gelişmeler sağlamıştır. Ancak, ECM (error correction model) ve TVP (time varying parameter) modelleri gibi diğer modern ekonometrik modellere göre tahmin performansları bu çalışmada değerlendirilmemiştir ve bu gelecekteki araştırmacılar için bir çalışma alanı olabilir.

Turizm sektöründe son zamanlarda kullanıcı tarafından oluşturulan içerikler üzerinde yapılan araştırmalara ilgi giderek artmaktadır. Kullanıcı içeriklerinden birisi de kullanıcı fotoğraflarıdır. Fotoğraf turist deneyimi ile ilgili birçok özelliği bünyesinde barındırmaktadır. Fotoğrafların incelenmesi ile turist davranışı, tercihleri, memnuniyet gibi birçok konuda bilgi elde etmek mümkündür. Fotoğraflar ile ilgili diğer bir çalışmada Deng ve Li (2018), destinasyon yönetim işletmelerinin kullanıcılar tarafından çekilen fotoğrafları destinasyon tanıtımında kullanabilmek amacıyla seçebilmesi için Naive Bayes sınıflandırıcıyı kullanmışlardır. Wan vd. (2018) ise kullanıcıların fotoğraflarından turistlerin ilgi noktalarını ve tercihlerini keşfetmişlerdir. Bu keşfi daha sonra bayes sınıflandırıcı yardımıyla potansiyel turistlere uygun kişiselleştirilmiş doğal noktaları tahmin etmek için kullanmışlardır. Sosyal medyadan elde ettikleri büyük miktardaki veriden kültür turizmine odaklanan Nguyen vd. (2017), yaptıkları kümeleme analizinden sonra aslında kültürel miras ile ilgisi olmayan birçok fotoğrafın bu küme içerisinde yer aldığını fark etmişler ve farklı kategoriye ait olan fotoğrafların çıkartılması için algoritma önermişlerdir. Önerilen algoritma, coğrafi etiketli her

bir fotoğrafın kültürel miras kaynağına ait olup olmadığının olasılığını hesaplamak için Naive Bayes yöntemine dayanılarak oluşturulmuştur.

Naive bayes algoritmasının turizm sektöründe en yaygın kullanıldığı alanlardan birisi de öneri sistemlerinde müşteriye en uygun ürünün önerilmesidir. Liu vd. (2011) kişiselleştirilmiş seyahat paketi önerisi için çevrimiçi seyahat verilerinden yararlanmışlar ve seyahat paketlerinin benzersiz özelliklerini analiz ettikten sonra seyahat paketi ve turiste önerilmesi için Bayes ağı kullanan bir model geliştirmişlerdir. Siering vd. (2018) çevrimiçi kullanıcı yorumlarından ve havayolu şirketi öneren kullanıcıların göstergesinden denetimli öğrenmeyi gerçekleştiren farklı makine öğrenme tekniklerinin öneri kararını öngörmedeki performanslarını değerlendirmişlerdir. Çalışmalarında Yapay Sinir Ağı (YSA) ve Destek Vektör Makinesi (DVM) gibi karmaşık öğrenme algoritmaları ile birlikte Naive Bayes algoritmasını da kullanmışlardır. Çalışmada Naive Bayes sınıflandırıcılarının oldukça basit olmasına ve potansiyel olarak gerçekçi olmayan varsayımlara dayanmasına rağmen, yine de genel olarak iyi performans gösterdikleri kanıtlanmıştır. Üstelik bu algoritmanın düşük hesaplama çabası gerektirme ve dolayısıyla daha fazla zaman kazanma avantajı bulunmaktadır (Siering vd., 2018: 57).

Bjørkelund vd. (2012) duygu analizi skorlarını 5 kategoride toplamış ve kategorileri sınıflandırmak için kullandığı makine öğrenme tekniklerinden birisi bayes ağları olmuştur. Benzer şekilde Xiang vd. (2018)'nin da metin sınıflandırmak için kullandığı algoritmalarından birisi de Multinomial Naive Bayes algoritmasıdır. Çalışmada performansı en iyi olan algoritma Multinomial Naive Bayes olmuştur. Puri vd.'nin (2017) yaptığı karşılaştırmada Gauss Naive Bayes algoritması çok iyi performans göstermiştir.

Tablo 2.5 Bayes Sınıflandırıcının Kullanıldığı Çalışmalar

Çalışma	Kapsam	Amaç
Arabi vd., 2018	Tahranda belirlenen 20 bölgede ulaşım durumu	Gelecek yıllarda ulaşım talep kapasitesinin oranını tahmin etmek
Liu vd., 2011	Veri seti 10 ülke, 139 şehir ve 1063 manzara içermektedir.	Farklı turistik paketler ile turistlerin ortak noktalarını (benzerlikleri) tespit etmek
Bjørkelund vd., 2012	TripAdvisor'dan elde edilen seyahat yorumları	Duygu analizi skorlarından oluşturulan kategorileri sınıflandırmak
Deng ve Li, 2018	2004 -2014 arası NewYork'ta çekilen Flickr'da paylaşılan 99,2 milyon fotoğraf ve 0,8 milyon videoya ait meta veri	Fotoğrafları uygunluk düzeylerine göre sıralamak.
Knoblock, 2004	Eski tarihli hava durumu ve uçuş verileri	Belirlenen bir uçuşta ertelenmenin olup olmayacağını tahmini
Nguyen vd., 2017	Sosyal medyadan elde edilen coğrafi etiketli sosyal medya verisi	Coğrafi etiketli her bir fotoğrafın kültürel miras kaynağına ait olup olmadığının olasılığını hesaplamak
Siering vd.,2018	Havayolu değerlendirme platformu airlinequality.com'da yayınlanan havayolu yorumları	Kullanıcıların havayolu önerileri ve çevrimiçi yorumlardan öğrenerek havayolu firması önerisinde bulunmak

Wan vd., 2018	2015 -2016 arası coğrafi etiketli fotoğraflar ve Pekin Hava durumu verileri	Bir turist için kişiselleştirilmiş doğal noktaları tahmin etmek
Xiang vd., 2018	TripAdvisor'dan elde edilen 1.94 milyon otel yorumu	Metin sınıflandırması
Puri vd., 2017	TripAdvisor'dan elde edilen 4000 otel yorumu	Metin sınıflandırması

2.2.2. Örnek Uygulama

Seyahat Acentasının otel temsilcisinin bilgilendirme yaptığı turist listesi Tablo 2.6'de verilmiştir. Tabloda müşterinin gelir ve yaş durumları vardır. Bunların dışında tur satın almaya etki eden özellikler veri tabanından alınabilir. Turistlerden bazıları bilgilendirme sonrası günlük tur satın almışlar bazıları da satın almamışlardır. Seyahat acentasının Tablo 2.6'de verilen veri seti eğitim seti olarak kullanılacaktır. Bu veri setine göre Orta yaşlı, gelir düzeyi yüksek olan yeni bir turistin tur satın alma ihtimali hesaplanacaktır.

Tablo 2.6 Bilgi Verilen Turistlerin Listesi

Turist	Yaş Durumu	Gelir düzeyi	Tur Satın Aldı
1	Genç	Orta	Evet
2	Genç	Yüksek	Evet
3	Orta	Orta	Hayır
4	Orta	Yüksek	Evet
5	Genç	Yüksek	Hayır
6	Genç	Orta	Hayır
7	Orta	Yüksek	Evet
8	Orta	Yüksek	Hayır
9	Genç	Orta	Hayır

Yukarıdaki eğitim kümesini ele alarak, Bayes sınıflandırıcılarını kullanmak suretiyle X_1 : YAŞ = ORTA X_2 : GELİR DÜZEYİ: YÜKSEK bir turistin tur alma durumunun hangi sınıfa ait olduğunu belirleyelim.

TUR ALMA DURUMU: ? (Evet ya da Hayır)

Tablo 2.7 Tur alma ile ilgili Tüm olasılıklar

		Evet		Hayır	
Yaş	Genç	2	2/4	3	3/5
	Orta	2	2/4	2	2/5
Gelir	Orta	1	1/4	3	3/5
	Yüksek	3	3/4	2	2/5

Bayes sınıflandırma yapabilmek için her bir hipotez için bayes olasılıkları tek tek hesaplanır. Bunun için,

C_1 : SATIN ALMA = EVET

C_2 : SATIN ALMA = EVET

olmak üzere $P(X|C_1) P(C_1)$ VE $P(X|C_2) P(C_2)$ ifadeleri hesaplanır. İfadelerden büyük olanı sınıfı verir.

A) $P(X|C_1) P(C_1)$ OLASILIĞI

$P(X_1 | C_1)$ = YAŞ =ORTA | TUR ALMA DURUMU: EVET = $\frac{2}{4}$

$P(X_2 | C_1)$ = GELİR DÜZEYİ: YÜKSEK| TUR ALMA DURUMU: EVET = $\frac{3}{4}$

$P(X | C_1)$ = TUR ALMA DURUMU: EVET = $\frac{3}{4} \times \frac{2}{4} = \frac{6}{16}$

$P(C_1)$ = TUR ALMA DURUMU: EVET = $\frac{4}{9}$

$$\frac{6}{16} \times \frac{4}{9} = 0,375 \times 0,444 = 0,166666$$

B) $P(X|C_2) P(C_2)$ OLASILIĞI

$P(X_1 | C_2)$ = YAŞ =ORTA | TUR ALMA DURUMU: HAYIR = $\frac{2}{5}$

$P(X_2 | C_2)$ = GELİR DÜZEYİ: YÜKSEK| TUR ALMA DURUMU: HAYIR = $\frac{2}{5}$

$P(X | C_2)$ = TUR ALMA DURUMU: HAYIR = $\frac{2}{5} \times \frac{2}{5} = \frac{4}{25}$

$P(C_2)$ = TUR ALMA DURUMU: HAYIR = $\frac{5}{9}$

$$\frac{4}{25} \times \frac{5}{9} = 0,16 \times 0,5555 = 0,088$$

Bu durumda orta yaşlı gelir düzeyi yüksek bir turistin golf turu satın alması beklenmektedir. Sınıfı EVET'tir.

2.3. Yapay Sinir Ağları

Bir sinir ağı, insan beyninininkine benzer, karmaşık, akıllı hesaplamalar yapabilen (Olmeda ve Sheldon, 2001: 6) ve ağırlıklandırma ile birbirlerine bağlanmış birçok işlem elemanından (nöron) oluşan matematiksel sistemlerdir. Biyolojik nöronlar devresi olarak da adlandırılan yapay sinir ağı (YSA), bir öğrenme süreci boyunca verilerdeki modeli tanımak veya sınıflandırmak için kullanılan matematiksel ve hesaplamalı bir modeli ifade etmektedir (Vajirachorn ve Chongwatpol, 2017: 81). YSA'lar, son derece karmaşık lineer olmayan fonksiyonları modelleme yeteneğine sahip, oldukça karmaşık analitik teknikler olarak bilinmektedir. Biçimsel olarak tanımlanmış olan YSA'lar, bilişsel sistemde ve beyin nörolojik fonksiyonlarında öğrenme süreçlerinden sonra modellenen ve bir süreci yürüttükten sonra diğer

gözlemlerden (aynı veya diğer değişkenler üzerinde) yeni gözlemleri (özel değişkenler üzerinde) tahmin edebilen analitik tekniklerdir (Delen ve Sirakaya, 2006: 317).

Sinir ağları, beyindeki sinir hücrelerinin işlevinden ilham alan bilgisayar cihazlarıdır. Birçok paralel, birbirine bağlı bilgi işlem biriminden oluşmaktadır (Cho ve Leung, 2002: 123). Yapay sinir ağları, başlangıçta, birbirine bağlı bir sistemde, basit hesaplama elemanlarının (nöronlar) birleşimi yoluyla insan beyninin nörofizyolojisini taklit etmeyi denemek için makine öğrenimi alanında geliştirilmiştir. Sonradan önemli bir veri madenciliği yöntemi haline gelmiştir. Sinir ağlarının, gerçek “zekâ” konusundaki tartışmalara rağmen, artık faydalı istatistiksel modeller olduklarından şüphe yoktur. Özellikle, gözlemlenen verilere, özellikle eksik bilgi, hata veya yanlışlıklar ile karakterize edilen veri setleri ve yüksek boyutlu veri tabanlarına uyum konusunda kayda değer bir yetenek göstermektedir (Giudici ve Figini, 2009: 76). YSA'lar, karmaşık veya kesin olmayan verilerden anlamlı örüntü tanıma veya veri sınıflandırması türetme yeteneklerinden dolayı birçok alanda başarıyla kullanılmıştır (Vajirakachorn ve Chongwatpol, 2017: 81). Örneklerden ve deneyimlerden öğrenmekte ve eğitim için kullanılan verilerin ötesinde genelleme yapabilmektedir. Etkin bir tahmin metodudur ve tahmin problemlerinde kullanılabilir. Bir sinir ağı, veri kümelerindeki doğrusal olmayan ilişkilere yaklaşabilir ve bu nedenle de örüntü tanıma veya verilerin sınıflandırılmasında yardımcı olmaktadır (Olmeda ve Sheldon, 2001:6).

Bir sinir ağı, ağırlıklı bağlantılar yoluyla birbirine bağlanmış nöron adı verilen bir dizi temel hesaplama biriminden oluşmaktadır. Bu birimler, tabakalar halinde düzenlenir, böylece bir tabakadaki her nöron, yalnızca önceki katmanın nöronlarına ve sonraki katmana bağlanmaktadır. Düğüm olarak da adlandırılan her nöron, özerk bir hesaplama birimini temsil eder ve girişleri, aktivasyonunu belirleyen bir dizi sinyal olarak almaktadır. Aktivasyonu takiben, her nöron bir çıkış sinyali üretmektedir. Tüm giriş sinyalleri aynı anda nörona ulaşır, böylece nöron birden fazla giriş sinyali alır, ancak yalnızca bir çıkış sinyali üretir. Her giriş sinyali bir bağlantı ağırlığı ile ilişkilidir. Ağırlık, giriş sinyalinin nöron tarafından iletilen son itkiyi üretmedeki önemini belirlemektedir. Bağlantılar, karşılık gelen ağırlıkların sırasıyla pozitif, negatif veya boş olup olmadığına göre uyarıcı, engelleyici veya boş olabilir. Ağırlıklar, biyolojik modellerle benzer şekilde, ağda dolaşan çeşitli sinyallere cevap olarak uygun bir öğrenme algoritmasına göre değiştirilen uyarlanabilir katsayılardır. Yanlılık denilen bir eşik değeri genellikle tanıtılır. Yanlılık, bir regresyon modelindeki bir kesişmeye benzemektedir (Giudici ve Figini, 2009: 76).

Bir sinir ağının nöronları katmanlar halinde düzenlenmiştir. Ağın ilk bölümünde girdi katmanı ve bu katmanda her biri farklı bağımsız değişkenleri temsil eden girdi düğümleri

mevcuttur. İkinci bölümde gizli katman(lar) ve gizli düğümler, son bölümde ise çıktı katmanı ve çıktı düğümleri bulunmaktadır. Girdi katmanı yalnızca dış ortamdan bilgi alır; içerisindeki her bir nöron genellikle açıklayıcı bir değişkene karşılık gelmektedir. Girdi katmanı herhangi bir hesaplama yapmaz; bilgileri bir sonraki seviyeye iletir. Çıktı katmanı, ağ tarafından sistemin dışına gönderilen son sonuçları üretir. Nöronlarının her biri bir cevap değişkenine karşılık gelmektedir. Bir sinir ağında genellikle iki veya daha fazla yanıt değişkeni vardır. Girdi ve çıktı katmanları arasında, dış ortamla doğrudan temas halinde olmadığı için gizli katman adı verilen bir veya daha fazla orta seviye olabilmektedir. Bu katmanlar sadece analiz amaçlıdır; onların işlevi, giriş değişkenleri ile çıkış değişkenleri arasındaki ilişkiyi almak ve verileri daha yakından uyarlamaktır (Giudici ve Figini, 2009: 79). Girdi düğümleri ağırlıklı bağlantılarla gizli düğümlere, gizli düğümler ise yine ağırlıklı bağlantılarla çıktı düğümlerine bağlanırlar. Çıktı katmanı bir veya daha fazla çıktı değişkenini temsil edebilmektedir. Literatürde bir sinir ağındaki katman sayısını hesaplamak için standart bir kural yoktur. Bazı yazarlar, nöronların tüm katmanlarını sayarken, bazıları da ağırlıklı nöronların katman sayısını saymaktadır.

İleri besleme geri yayılım ağları (feed-forward back propagation networks – FFBN), radyal temelli fonksiyon ağları (radial based function networks - RBFN) ve Kohonen Kendini düzenleyen haritalar (self-organizing maps) literatürde var olan bazı sinir ağları algoritmalarıdır. Geri yayılım, ağın gizli katmanlarına bağlı olan ağırlıklarını değiştirmek için kullanılan denetimli bir öğrenmedir. Bir gözlem için ileri besleme aşaması tamamlandığında, tahmin değeri ile gerçek değer arasındaki fark kullanılarak bir hata değeri hesaplanır. Daha sonra geri besleme ile bu hata değeri ağ yapısına gönderilir ve benzer bir örnek geldiğinde daha az hata oluşturacak şekilde ağırlıklar değiştirilir.

Quick, Dynamic, Prune ve RBFN ileri besleme geri yayılım algoritmasına dayanan tekniklerden bazılarıdır. Quick tekniği, ağ topolojisi için genel kurallardan ve verinin özelliklerinden faydalanır. Bu yöntemde ağ parametreleri için ilgili yazında kabul görmüş genel parametre değerleri kullanılır ve veri üzerinde bir defa öğrenme gerçekleştirilerek sinir ağı modeline ulaşılır. Dynamic yöntemi, istenilen doğruluğa erişinceye kadar, performansı iyileştirmek için network topolojisini değiştirerek öğrenme gerçekleştirir. Öncelikle bir başlangıç topolojisi oluşturur, öğrenme süresince gizli düğümler ekleyip çıkararak son modelini oluşturur. Prune yöntemi, kavramsal olarak, dinamik yönteminin tam tersi işlev görür. Küçük bir ağ ile başlayıp onu geliştirmek yerine, büyük bir ağ ile başlar ve gereksiz sinir hücrelerini girdi ve gizli katmanlarından atarak aşamalı olarak budar. RBFN yöntemi ise çıktı değerlerine göre veriyi bölmek için k-ortalama yöntemine benzer bir yöntem kullanan özel bir yapay sinir ağıdır.

Yapay Sinir Ağlarından insan beyninin çalışma mekanizmasını taklit etmeye çalışan bir yöntem de derin öğrenmedir (Deep Learning). Derin öğrenme yöntemleri oldukça etkili bir yöntem olarak ortaya çıkmaktadır çünkü ön işlem yapmadan verilerden özniteliklerin çıkarılmasını sağlayan bir yapıya sahiptirler. Verilerden klasik yöntemlerle özniteliklerin alınması oldukça yorucu bir süreçtir. Bununla birlikte, bunu otomatik olarak yapan derin öğrenme yöntemleri daha etkili sonuçlar üretebilmektedir. Derin öğrenme görüntü analizi, robotik, gen analizleri, ses analizi, otonom araçlar, sanal gerçeklik ve kanser teşhisleri vb. birçok alanda kullanılmaktadır. Bu derece çok yaygın bir alanda kullanılması en büyük sebebi problemlerin çözümünde elde ettiği yüksek doğruluktur. Hatta ses tanımlama, görüntü tanımlama gibi bazı problemlerde insan performansının üzerine çıkmıştır.

2.4. En Yakın Komşu

Bir vakayı sınıflandırmanın bir yolu, sınıfı bilinen bir vakayı hatırlamak ve yeni vakanın aynı sınıfa sahip olacağını tahmin etmektir. Bu, benzer ve hatırlanan vakalara atıfta bulunarak, görünmeyen vakaları sınıflandıran en yakın komşu sistemlerinin altında yatan felsefedir (Cho ve Leung, 2002: 117).

Her ne kadar en yakın komşu öğrenme algoritmasının birçok çeşidi mevcut olsa da, bu algoritmalar, tüm (veya seçilen) eğitim örneklerini saklar ve test örneğinin sınıfını tahmin ederken saklanan eğitim örnekleriyle bir test örneğinin benzerliğini (veya mesafesini) ölçmek için benzerlik (veya mesafe) işlevini kullanmaktadır. Yani, herhangi bir eğitim süreci olmadığı için, eğitim süreci boyunca bir kavram açıklaması oluşturmazlar. Bunun yerine, test örneklerini eğitim örneklerine dayanarak tahmin etmektedirler. En yakın komşu sınıflandırıcıya en popüler örnek, k-en yakın komşu (k-NN) algoritmasıdır.

En yakın komşu modelleri, aynı zamanda, bilinmeyen örneklem için bir gerçek değerinin tahmininde de kullanılabilir. Tahmini sınıflandırma için en yakın komşu yöntemleri de kullanılabilir. Bir gözlemi sınıflandırmak için, komşuluğu eskisi gibi belirlenir ve her bir kategorideki uygun olasılıklar komşulukta nispi frekanslar olarak hesaplanır. Olasılığı en uygun olan sınıf gözlemin sınıfı olarak seçilir. Ağaç modelleri gibi, en yakın komşu modelleri olasılık dağılımı gerektirmez. Ancak, sınıflandırma ağaçları, ağaç yolları açısından açık tahmin kuralları sağlayarak, verileri özel sınıflara ayırırken, en yakın komşu modellerinde elde edilen değerler, açık kurallara göre değil, örtüşen gözlem gruplarına dayanmaktadır (Giudici ve Figini, 2009: 89).

Bu yöntemler ayrıca, hazırlanması gereken bir model veya tahmin edilmesi gereken bir fonksiyon gerektirmediklerinden bellek tabanlı modeller olarak da bilinmektedir. Bunun yerine,

tüm gözlemler bellekte tutulmakta ve bir tahmin gerektiğinde, bellekten maddeleri geri çağırmakta ve gerekeni hesaplamaktadırlar.

K-en yakın komşu (k-NN) algoritması, eğitim veri seti içindeki en benzer veri girişlerine dayanarak yeni bir veri girişinin hedef niteliğini tahmin etmektedir. Bu yöntem, sınıfları belli olan bir örnek kümesindeki gözlem değerlerinden yararlanarak, örneğe katılacak yeni bir gözlemin hangi sınıfa ait olduğunu belirlemek amacıyla kullanılmaktadır (Özkan, 2008:117). Sınıflandırma yapılırken veri setindeki her bir kaydın diğer kayıtlara olan uzaklığı hesaplanmaktadır (Silahtaroglu, 2008: 65). Sınıflandırma örnek kümedeki gözlemlerin her birinin sonradan belirlenen bir gözlem değerine olan uzaklıklarının hesaplanması ve en küçük uzaklığa sahip k sayıda gözlemin seçilmesi esasına dayanmaktadır (Özkan, 2008: 117). Sınıflandırmada aktif rol alan komşuların sayısı k ile gösterildiği için, algoritma k-en yakın komşu algoritması olarak isimlendirilmektedir.

2.5. Destek Vektör Makineleri

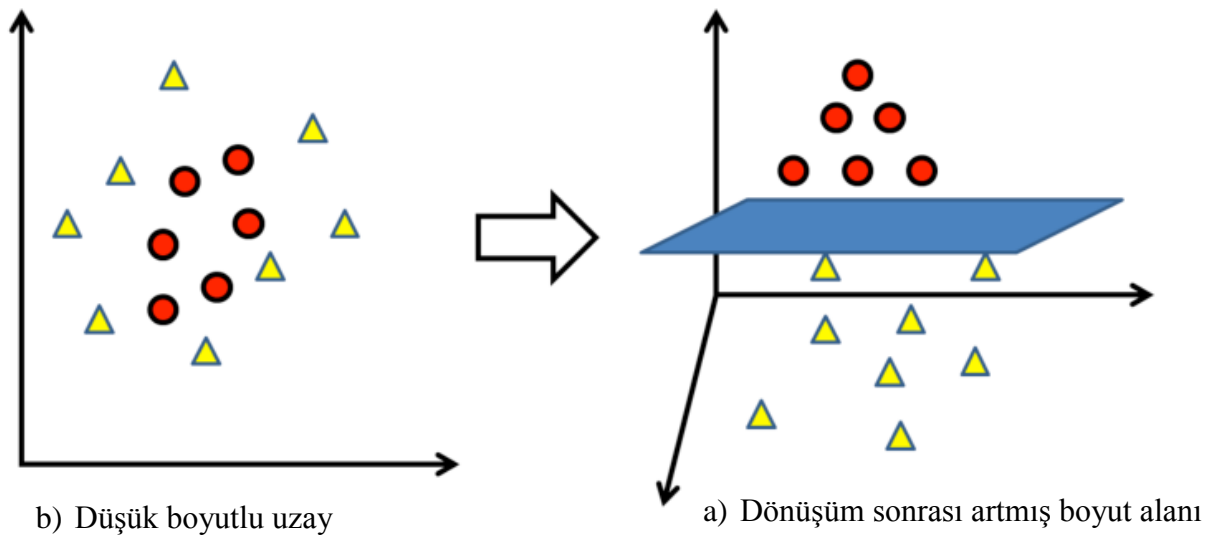
Destek vektör makineleri (DVM), karar sınırlarını belirleyen karar düzlemleri kavramına dayanan en iyi denetimli öğrenme algoritmalarının birisidir. Bir DVM, sınıflandırma görevlerinin verilerini kategorilere ayırmak veya regresyon görevlerinde istenen çıktının sayısal değerini tahmin etmek için çok boyutlu bir alanda aşırı düzlemler oluşturmak için haritama işlevini kullanan bir sınıflandırma yöntemidir (Vajirakachorn ve Chongwatpol, 2017: 81). Bir haritada iki bölgeyi birbirinden doğru şekilde ayırabilmek için aralarındaki sınır çizgisinin en doğru yerde çizilmelidir. DVM bu sınırın nasıl çizileceğini belirlemektedir.

DVM hiçbir parametre almayan bir sınıflandırıcıdır. Doğrusal ve doğrusal olmayan verilerde sınıflandırma yapabilmektedir, ancak verileri genellikle doğrusal olarak sınıflandırmaya çalışmaktadır. Sınıflandırma işlemi bir düzlemde bulunan iki grup arasına bir sınır çizilerek yapılmaktadır. Araya çizilecek doğru iki sınıfın elemanlarına en uzak yerden geçecek şekilde ayarlanmaktadır.

DVM verileri doğrusal olarak sınıflandırmaya çalışmaktadır, ama bu her zaman mümkün olmamaktadır. Gözlemlerin doğrusal bir şekilde ayrılması her zaman mümkün olmadığından, dönüşümler, gözlemlerin atanan sınıflarına göre ayrılmasını sağlayan çekirdek işlevleriyle gerçekleştirilmektedir (Siering vd., 2018: 57). Girdi verilerini dönüştürmek için doğrusal, polinomsal, radyal tabanlı fonksiyon (RBF) ve sigmoid gibi çeşitli çekirdek işlevleri kullanılmaktadır. Böylece girdi verileri daha fazla ayrılabilir hale gelmektedir (Vajirakachorn ve Chongwatpol, 2017: 81). Şekil 2.3 (a)'daki grafikte kırmızı noktaları doğrusal bir çizgi ile

ayırmak mümkün değildir. Çekirdek işlevi ile kırmızı noktalar biraz yukarı kaldırılıp (z eksenini) 3. bir boyut oluşturularak DVM ile doğrusal bir çizgi oluşturmak mümkün olmaktadır.

Destek Vektör Makinesi, farklı sınıf örnekleri arasındaki mesafeleri maksimize eden maksimum aşırı düzlem sınırını bulma prensibine dayanmaktadır (Siering vd., 2018: 57). Aşırı düzlemde sınır çizgileri arasındaki daha büyük bir mesafe, giriş verilerinin sınıflandırılmasında daha az hata olduğunu göstermektedir. Sınıflandırma görevleri için, örneğin, yeni gözlemler aynı alanda haritalandırılmakta ve ait oldukları bölge tarafında tahmin edilmektedir. (Vajirakachorn ve Chongwatpol, 2017: 81).



Şekil 2.3 Çekirdek İşlevi İle Verinin Dönüştürülmesi

Turizmde özellikle havaalanı servisleri için çevrimiçi yorumlarda verilen önerileri analiz eden Siering vd. (2018) YSA ve DVM'yi Naive bayes algoritmasına göre daha başarılı bulmuştur. Al Shehhi ve Karathanasopoulos, (2018) tahmin için hem RBF hem de DVM kullanmışlar ve bunun performansta önemli gelişmelere neden olduğunu gözlemlemişlerdir. Ayrıca farklı modellerin ve sekiz farklı kentin tahmin doğruluğunu karşılaştırırken, DVM ve RBF'nin ARIMA veya basit hareketli ortalamadan daha iyi performans gösterdiği bulmuştur. Turizm sektöründe iş zekası (BI) kullanımına değinen Vajirakachorn ve Chongwatpol (2017)'in kullandığı algoritmalarından birisi de DVM'dir.

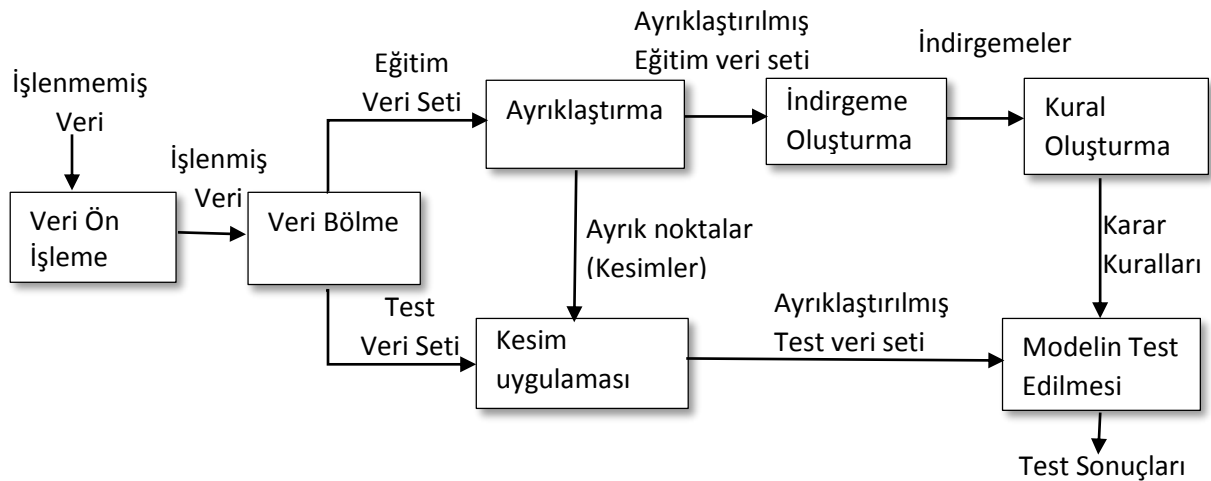
Bir DVM, birkaç bin eğitim örneğiyle başa çıkabilmektedir.

2.6. Kaba Kümeler

Kaba küme analizi, Pawlak'ın (1982) ilk olarak, kesin olmayan ve belirsiz veri kaynaklarından gelen iş kuralları biçiminde bilgiyi keşfetmek için ortaya attığı kaba küme

teorisine dayanan matematiksel bir yaklaşımdır (Delen ve Sirakaya, 2006: 318). Özellik seçimi, özellik çıkarma, özellik azaltma ve özellikle belirsiz ve kesin olmama durumunda veriden karar kurallarının çıkarılması gibi çeşitli amaçlar için kullanılan bir veri madenciliği yaklaşımıdır (Chen ve Tsai, 2016: 198). Gerçek hayatta, kümelerle uğraşırken, sınırlı çözünürlük (tam ve ayrıntılı bilgi eksikliği) ve ölçülebilir özellikleriyle ilişkili belirsizlik nedeniyle, özgün küme öğelerini birbirinden tam olarak ayırt etmek için hiçbir araca sahip olamama durumu bu yaklaşımın temelinde yatmaktadır (Delen ve Sirakaya, 2006: 318). Kusurlu verilerin ele alınmasına yönelik bir yaklaşım olarak, kaba küme analizi olasılık teorisi, kanıt teorisi ve bulanık set kuramı gibi diğer geleneksel teorileri tamamlamaktadır.

Kaba küme felsefesi, bazı enformasyonun (veri, bilgi) araştırma evreninin her nesnesiyle ilişkilendirilme varsayımı üzerine kurulmuştur. Aynı bilgilerle karakterize edilen nesnelere, haklarındaki mevcut bilgiler ışığında ayırt edilemez. Bu şekilde üretilen vazgeçilmezlik ilişkisi, kaba küme teorisinin matematiksel temelidir. Tüm ayrılmaz nesnelere kümesine temel küme denir ve evren hakkında temel bir bilgi granülü (zerre- çekirdek mi) oluşturur. Bazı temel kümelerin birleşmesiyle oluşan herhangi bir nesne kümesi net veya kesin bir küme olarak adlandırılır, aksi takdirde küme kaba (kesin veya belirsiz) olmaktadır. Sonuç olarak, her kaba küme, küme üyeleri olarak tamamen kesin bir şekilde sınıflandırılmayan sınır çizgisi durumlarına (yani nesnelere) sahiptir. Bu nispeten yeni tekniği daha iyi anlamak için, kaba küme analizi yürütme ana adımları Şekil 2.4'de verilmiştir.



Şekil 2.4 Kaba Küme Analizinin Süreç Haritası ve Ana Süreçleri

Kaynak: Delen ve Sirakaya, 2006: 319

İlk adım, eksik ve boş değerlerin ele alınmasını ve veri setini bir karar tablosuna dönüştürmeyi içeren verileri önceden işlemektir. Daha sonra, ikinci adımda, veriler eğitim ve testlere ayrılır. 10 kat çapraz doğrulama yöntemine dayanarak, veri kümesi 10 ayrı altkümeyle

bölünür ve bu 10 alt kümeden biri, 10 deneyin her biri için test kısmı olarak kullanılır. Veriler eğitim ve test alt kümelerine ayrıldıktan sonra, eğitim alt kümesi, sürekli sayısal olarak değer verilen değişkenleri anlamlı aralıklı etiketlere ayırmak için kullanılır. Kaba kümeler algoritması sadece nominal değerli değişkenlerle çalışabileceğinden, doğru model oluşturma için ayrıklaştırma süreci çok önemlidir. Değişkenler ayrıldıktan sonra indirgemeler üretilir. Bu indirgemeler kullanılarak, kurallar oluşturulur. Bu noktaya kadar test veri setinin, ayrıklaştırma ve indirgeme ve kural oluşturma süreçleri için hiçbir şekilde veya biçimde kullanılmamıştır.

Kaba kümeler teorisinin önemi, koşul ve karar özellikleri arasındaki bağımlılıkları tanımlayan karar kurallarının oluşturulmasıdır. Karar kuralları mantıksal ifadelerle ifade edilmektedir (Goh vd., 2008: 328):

EĞER [Koşul (lar)] SONRA [Karar (lar)].

Kaba kümeler analizinde karar kurallarının bu modüler yapısı, uygulayıcıların ve araştırmacıların gerektiğinde, genel sistemi etkilemeden mevcut karar kurallarını benimsemelerini ve değiştirmelerini sağlamaktadır.

Kaba set modelinin tahmin doğruluğunu elde etmek için, test örneği, model oluşturma aşamasında üretilen ayrı noktalar (yani kesimler) kullanılarak ayrıştırılır ve sınıflandırma kurallarına uygulanır. Mevcut katın sonuçları, bir sınıflandırma matrisine (aynı zamanda gerçek pozitiflerin, gerçek negatiflerin, yanlış pozitiflerin ve yanlış negatiflerin dört matris hücresinde listendiği 2 x 2 matris olarak gösterilen karışıklık matrisi olarak da bilinir) kaydedilir ve deney 10 kez tekrarlanır (10 katın her biri için).

Kaba kümelere yöneltilen başlıca eleştiri, tüm değişkenlerin nominal formatta olmasını gerektirmesidir. Kaba kümeler, sayısal sürekli değerli değişkenlerle çalışamaz. Bu değişkenlerin kullanılabilmesi için ayrıklaştırma yapılmalıdır. Sürekli sayısal değişkenleri aralık etiketlerine dönüştürme işlemi olan ayrıklaştırma, eşit genişlik, eşit kova / derinlik gibi basit sıralama algoritmalarından genetik algoritmalar ve tavlama benzetimi gibi çok karmaşık arama algoritmalarına kadar değişen farklı yöntemler kullanılarak yapılabilir. Kullanılan gelişmişlik seviyesine bakılmaksızın, herhangi bir veri seti için ayrıklaştırmanın tek bir en iyi yöntemi yoktur, bu da onu deneye dayalı bir uzun öğrenme süreci haline getirmektedir. Bazıları ayrıca, sürekli değerli sayısal değerlerin birkaç aralıklı etiketlere dönüştürülmesinin, değişkenin bilgi içeriğini azaltacağını iddia edebilir. Bazıları, sürekli değerli sayısal değerlerin birkaç aralık etiketine dönüştürülmesinin değişkenin bilgi içeriğini azaltacağını da iddia edebilir. YSA'lar ve karar ağaçları, kaba kümelerle karşılaştırılırsa, hem sürekli hem de nominal değerli değişkenleri işleyebilmektedir (Delen ve Sirakaya, 2006: 329).

Kaba küme teorisi, kalite mühendisliği, insan kaynakları yönetimi, sağlık bakımı, özellik seçimi (çeşitli) gibi alanlarda uygulanmıştır (Chen ve Tsai, 2016: 198). Turizmde Chen ve Tsai (2016) yer seçimi kararlarını desteklemek için kaba küme teorisine (RST) dayanan bir veri madenciliği çerçevesi geliştirmiştir. Law vd. (2006), veri madenciliğinde iş seyahatindeki kişilerin temel bilgilerini yakalayabilecek (iş seyahatinde olanları modellemek için) kaba kümeler temelli bir model geliştirmeye çalışmaktadır. Delen ve Sirakaya (2006), yapay sinir ağları, karar ağaçları ve kaba kümelerin üç popüler veri madenciliği yöntemini test etmişler ve kaba kümeler algoritmasının üçü arasında en iyi tahmin aracı olduğunu bulmuşlardır.

Ayrıca, kaba kümeler yaklaşımı, bir veri tabanındaki kalıpları karma verilerle (hem nicel hem de nitel değişkenlerden oluşan bir veri kümesi) ortaya çıkarmak için özellikle uygun bir araç sunmaktadır. Örneğin, turistlerin bir destinasyona seyahat etme kararı sadece fiyat ve gelir gibi nicel ekonomik faktörlerden değil, aynı zamanda bir varış yerinin algılanması, kültürel geçmişi ve iklim durumu ve seyahat süresi kısıtlamaları gibi niteliksel ekonomik olmayan faktörlerden de etkilenmektedir. Bu nedenle, kaba kümeler tekniği, turizm talep analizi için kullanışlı ve benzersiz bir araçtır (Goh vd., 2008: 328).

2.7. Lojistik Regresyon

Lojistik regresyon analizi diskriminant analizi ile birlikte sınıflandırma amacıyla kullanılan istatistik yöntemlerden biridir. Ancak diskriminant analizi tüm öngörü değişkenlerinin normal dağıldığı varsayımı, nominal öngörü değişkenlerinin kullanılamaması ve sınıfları ayıran sınırların doğrusal olması gibi sebeplerle veri madenciliğinde pek yaygın olarak kullanılmamaktadır.

Bir kategorik bağımlı değişkenin öngörülmesinde, bütün bağımsız değişkenler sürekli ve normal dağılıma sahip ise diskriminant analizi, bütün bağımsız değişkenler kategorik ise lojit analizi, bir takım değişkenler sürekli diğerleri kategorik ise ve sürekli değişkenlerin dağılımı hakkında varsayımda bulunulmuyorsa lojistik regresyon analizi kullanılmaktadır.

2.8. Genetik Algoritmalar

Veri madenciliği, kalıpları veya kilit ilişkileri saptamak için insan anlayışını verilerin otomatik analiziyle birleştirebilmektedir. Bir dizi değişken üzerinde tanımlanmış büyük bir veri tabanı göz önüne alındığında, amaç veri tabanındaki en ilginç kalıpları verimli bir şekilde bulmaktır. Bazı uygulamalarda ilginç kalıpları tanımlamak için genetik algoritmalar uygulanmıştır. Genetik algoritmalar, genetik kalıtım sürecini kullanan matematiksel prosedürlerdir (Olson ve Delen, 2008: 125). Genetik algoritmalar, problemlerin çözümü için doğada gözlemlenen evrimsel süreci bilgisayar ortamında taklit etmektedir. En iyi olanın yaşamına devam

edebilmesi ilkesine göre bütünsel en iyi çözümü aramaktadır. Genetik algoritmalar, temelde doğal evrim sürecini örnek alarak geliştirilen, ilginç ve başarılı sonuçlar üretebilen, algoritmanın amacını çok iyi anlatan tanımıyla da bir arama ve optimizasyon yöntemidir.

Veri madenciliğine genetik algoritmalar iki şekilde uygulanmıştır (Olson ve Delen, 2008: 131). Genellikle veri madenciliğinde karar ağaçları ve birliktelik kuralları gibi diğer algoritmaların performansını iyileştirmek için kullanılmıştır. Dış destek, başka bir öğrenme sistemi için bir parametrenin değerlendirilmesi veya optimizasyonu, genellikle kümeleme veya karar ağaçları gibi diğer veri madenciliği araçlarını kullanan hibrit sistemlerdir. Bu anlamda, genetik algoritmalar diğer veri madenciliği araçlarının daha verimli çalışmasına yardımcı olmaktadır. Genetik algoritmalar, genetik algoritmanın genellikle karar kuralları veya karar ağaçları olarak tanımlamalar ürettiği analize doğrudan da uygulanabilmektedir.

Genetik algoritmalar belirli veri yapıları gerektirmektedir. Kategorik olarak ifade edilen özelliklere sahip bir popülasyon üzerinde çalışırlar. Genetik ile analogi, popülasyonun (genlerin) karakteristiklerden oluşmasıdır. Genetik algoritmaları uygulamanın bir yolu, potansiyel olarak daha iyi kombinasyonların üretilmesini arttırmak için mutasyon özellikli operatörler (üreme, çaprazlama, seçim) uygulamaktır. Genetik algoritma süreci aşağıdaki gibidir (Olson ve Delen, 2008: 125):

1. Rastgele bireyler (ebeveynler) seçilir.
2. Çapraz geçişle çoğaltılır. Üreme, hangi bireysel varlıkların hayatta kalacağını seçen operatördür. Başka bir deyişle, hayatta kalmayı belirlemek için bazı objektif işlevler veya seçim özellikleri gerekir. Genetik değişim gelecekteki varlıkların nesillerindeki değişikliklerle ilgilidir.
3. Bir spor işleviyle gelecek nesil için hayatta kalanlar seçilir.
4. Mutasyon, müteakip operasyonlarda rastgele seçilen varlıkların rastgele seçilen niteliklerinin değiştirildiği işlemdir.
5. Belirli bir zindelik seviyesine ulaşıncaya kadar ya da önceden ayarlanan yineleme sayısına ulaşılan kadar tekrarlanır.

Genetik algoritma parametreleri arasında popülasyon büyüklüğü, geçit hızı (bireylerin geçme olasılığı) ve mutasyon oranı (belirli bir varlığın mutasyona uğrama olasılığı) bulunmaktadır. Evrimin temel hipotezlerinden biri, bu algoritmanın da temel dayanağını oluşturmaktadır; her zaman sadece "en iyi" olan hayatta kalmakta ve çoğalmaktadır. Başarısız bireyler çoğalamaz ve nesilleri devam edemez. Algoritmanın uyarlanabilir bu yüzeysel parametreleri (evrimin ilkeleri) dışında tüm problemlerde kullanılacak belirgin standartları bulunmamaktadır. Algoritma çözülecek probleme göre şekillenmektedir.

Genetik algoritmalar sistematik olarak deęişken deęerlerin daha iyi (ya da muhtemelen en uygun) kombinasyonlarını arayarak ayırık veri üzerinde alıřmaktadır. zellikle doęrusal olmayan fonksiyonlar söz konusu olduęunda karmařık etkileřimlerdeki problemlerde ok etkili olabilmektedir. Genetik algoritmalar, veri madencilięinde kesin olmayan ve tutarsız bilgilerle bařa ıkma yeteneklerinden dolayı deęerlidir.

Genetik algoritmaların geliřtirilmesi ve geerlilięi ok kolaydır, bu avantajları onları ekici kılmaktadır. Ayrıca algoritma paraleldir, yani byk poplasyonlara verimli bir řekilde uygulanabilir. Algoritma, zayıf bir orijinal czmlle bařlarsa hızlı bir řekilde iyi czmlere ilerleyebilmesi iin de etkilidir. Mutasyon kullanımı, yntemi ok doęrusal olmayan problem alanlarında bile kresel optimumu belirleme yeteneęine sahiptir. Yntem, verilerin daęıtımı hakkında bilgi gerektirmez.

Genetik algoritmalar, veri setlerinin zelliklerin genetik algoritmanın alıřması iin ayırık deęerlere sahip olduęu bir forma eřlenmesi gerektirir. Bu genellikle mmkndr, ancak srekli deęişkenlerle uęrařırken ok fazla ayrıntılı bilgiyi kaybedebilir. Verileri kategorik forma kodlamak istemeden veride yanlılıęa neden olabilir. Ayrıca genetik algoritmalarla analiz edilebilecek veri setinin boyutu da sınırlandırılmıřtır. ok byk veri kmeleri iin, rnekleme gerekli olacak ve bu, aynı veri seti zerinde farklı alıřmalarda farklı sonulara yol aacaktır.

Veri madencilięi ierisinde birok genetik algoritma uygulaması iřletmecilik dıřında uygulanmıřtır. zel rnekler arasında tıbbi veri madencilięi ve bilgisayar aęı izinsiz giriř tespiti bulunmaktadır. İřletmelerde genetik algoritmalar mřteri blmlendirmesine, kredi puanlamasına ve finansal gvenlik seimine uygulanmıřtır.

ÜÇÜNCÜ BÖLÜM

TURİZMDE VERİ VE VERİ KAYNAKLARI

Sanayi devrimi, tam olarak “toplumu ve ekonomiyi temelden değiştiren gelişme” olarak anlaşılmakta ve rekabet avantajı yaratmak için standart iş modellerini köklü bir şekilde yeni alanlara dönüştürmektedir. Dördüncü sanayi devriminin ana araçları (veya sonuçları) yeni iletişim ve enformasyon teknolojisi cihazları, bulut bilişim, yapay zekâ ve otomasyondur. Ana yakıt (veya yeni sanayi devriminin ana nedeni) veridir (Petković vd., 2018: 151). Veri işletmelerin önemli bir varlığı olarak kabul edilmekte ve müşteriye sunulacak birçok hizmette de yapılacak analizin ana girdisini oluşturmaktadır.

Turizm sektöründe dijitalleşme ve iş modellerinin değişmesi ile birlikte son dönemde gündemde olan konulardan biri de veri yönetimidir. Dijitalleşmenin tüm hızı ile sürdüğü günümüzde işletmelerin rekabet ortamında ayakta kalabilmeleri için (Magnini vd., 2003: 95) veriye ulaşım ve kullanım kilit rol oynamaktadır. Sapphire Ventures Başkan Yardımcısı Max Wessel’e göre, dördüncü sanayi devriminin en büyük faydalanıcıları “herhangi bir konuda tüketicilerin temel istek veya duygularına erişim sağlayan verilere sahip ve bunun üzerine akıllı uygulamalar geliştiren” şirketler olacaktır (akt. Petković vd., 2018: 151). Verilerin bir işletme tarafından elde edilmesi, korunması, gerektiğinde paylaşılması ve işlenmesi hayati öneme sahiptir. Verinin hem içeriden hem dışarıdan düzgün bir şekilde temini, korunması ve yönetimi kurumun müşterileri için değer üretmesini sağlayacak analitik çalışmaların başlangıç noktasını oluşturmaktadır. Verilerin işletme içinde müşterek biçimde anlaşılması, departmanlar arası analiz ve kararlara imkân tanıyan ortak bir iş yapısı olması ve hangi verinin nerede ve nasıl kullanıldığına dair güçlü yönetimin olması gerekmektedir.

Müşteri ile ilişkileri sadece yapılan işlemler bazında değil toplam hayat döngüsü bazında yönetmeyi sağlayacak veri yönetimi yapısının ve buna ilişkin entegrasyon gereksinimlerinin tasarlanması ve müşteri ilişkileri yönetimi sisteminin kurulması memnun ve sadık müşteri grubunun oluşturulmasında katkı sağlamaktadır. Dolayısıyla, veri ve bağlantılı konuların sadece bilişim teknolojileri yetkisinde bırakılmaması, iş birimi bazında bir aksiyon olarak düşünülüp hareket edilmesi gerekmektedir. Buna ek olarak, gerçek zamanlıya yakın veriyi daha derin analitikleri desteklemek için büyük bir veri ortamı inşa edilebilir, bunların en olası kombinasyonlarına imkân tanımak üzere yoksa bir veri ambarı kurulabilir, varsa mevcut veri ambarı modernize edilebilir.

Bilimsel araştırmaların temel dayanağını veri teşkil etmektedir. Veriler işlenerek anlam çıkarmada, problemin çözümünü sağlayacak sonuçlara varmada kullanılmaktadır. Bir araştırma

sağlıklı sonuçlara ulaşılabilmesi için analiz edilen verilerin sağlıklı (güvenli) olması gerekmektedir. Analiz tekniği ne kadar güçlü olursa olsun, veriler sağlıklı değilse sağlıklı sonuçlara ulaşmak mümkün değildir. Sağlıklı veri elde etmek için de veri toplama yöntemi ve veri kaynağının sağlıklı olması gerekmektedir.

Aslında turizmde veri kaynağı olarak bizzat turistin kendisi ve turistin kendi yaptığı ve hakkında yapılan işlemlerin tutulduğu veri tabanları Chakrabarti vd. (2009: 37) olmak üzere iki birincil veri kaynağı bulunmaktadır. Chakrabarti vd. (2009) veri madenciliği açısından veri kaynağı olarak veri tabanlarını görmekte ve veritabanlarının operasyonel sistemler, veri ambarları, çevrimiçi analitik işleme uygulamaları, anketler ve hanehalkı ve demografik veri tabanları için veri havuzu olarak hizmet verdiklerini belirtmektedir. Veri tabanları günümüzde tüm enformasyon sistemlerinin verilerini depo etmek için kullanılmaktadır. Operasyon destek sistemlerinin, Müşteri İlişkileri Yönetimi sistemlerinin kendi bünyelerinde verileri depo etmek için kullanılan veri tabanları vardır. Web siteleri ve sosyal medya aslında kendi bünyelerinde veri barındırmazlar. Bu sistemler verileri veri tabanlarında depolamakta ve kullanıcıların veriyi sanki site içerisindeymiş gibi kullanmasını sağlamaktadır.

Enformasyon teknolojilerindeki gelişim insan yaşamıyla birlikte turist yaşamını değiştirmiş ve turizm ve turist ile ilgili verilerin depo edildiği ortamların gelişmesine neden olmuştur. Enformasyon sistemleri her seyahatin başlangıcından bitimine kadar her alanda hem işletmeler hem de turistler tarafından kullanılmaktadır. Bu da sektörde müşterilere ve ürün ve hizmetlere ve bunlar için yapılan operasyonel işlemlere dair devasa bir veri akışına ve birikimine neden olmaktadır. Sektörde kullanılan enformasyon teknolojileri bünyelerindeki veri tabanı ve veri ambarlarına turistik ürün, turist ve bunlarla ilgili işlemlere ait verileri depo etmektedirler. Ayrıca, içte oluşan veriden belki de çok daha fazlası dış kaynaklardan elde edilebilmektedir. Web siteleri, sosyal medya gibi platformlarda turizm pazarlaması, ürün yönetimi ve operasyonel aktiviteler için değerli bilgiler elde edebilecekleri veriler yönetici ve araştırmacıların hizmetine sunulmuş durumdadır. Dolayısıyla enformasyon sistemleri turistler hakkında veri toplamak için güvenilir bir veri kaynağı haline gelmiştir.

Turizm yapısı gereği, satış aşamasından satış sonrası işlemlere kadar veriye bağımlı olarak çalışmaktadır. Turizmde veri akışı üç yönlü olmaktadır. Veri akışının ilki, bilet, otel odaları, eğlence ve benzeri gibi turistlerin tükettiği ürünler hakkında tedarikçilerden müşterilere veya turistlere doğrudur. Diğeri de turistler hakkında toplanmış bilgilerden oluşan ters yönde servis sağlayıcılara doğru olan bilgi akışıdır (Bose, 2009: 936). Günümüzde ise bu iki yönlü bilgi akışından daha yoğun ve hızlı bir akış müşteriden müşteriye. Bu akış kontrolsüz, yapılandırılmamış bir veri yığına şeklindedir. Üç yönlü veri akışının olduğu turizmde genel

anlamda veri, ilk olarak bizzat turistlerin kendilerinden işletme ve arařtırmacılar tarafından anket ve görüřme yapılarak toplanır. İkinci olarak işletmelerin veri tabanlarında turistten işletmeye ve işletmeden turiste yönelik iki farklı veri türü bulunmaktadır. Üçüncü olarak turistlerin deneyimlerini paylaşmaları sonucu ortaya çıkan kullanıcı tarafından oluşturulan içeriklerdir (KOİ). Burada veri setlerinden bahsedilmiştir. Bu veri setlerinin içinde turistin algısı, memnuniyeti, davranışları vb. ile ilgili çok detaylı veriler olabilmektedir.

Turizm sektöründe geleneksel olarak arařtırmalar ve yöneticiler, turistleri tanımak ve onların davranış ve tercihlerini öğrenerek, ürün geliřtirmek için gerekli veriyi anket ve görüřme yöntemleri ile yine turistlerden toplamaktadırlar. Bu yöntemde belirlenen bir ölçek ile yapılandırılmış veriler toplanmakta ve bir yazılım aracılığıyla analiz edilmektedir. Teknolojik gelişmelerin ardından turizm sektörünü ilgilendiren devasa veri yığınları web sitelerinde, sosyal medyada ve ilgili diđer platformlarda birikmeye başlamıştır. Bu kadar çok veri, bununla birlikte veri işleme konusunda bazı zorlukları beraberinde getirmiştir. Daha önceleri yapılandırılmış bir anketle veri toplayan arařtırmacılar, birbirinden bağımsız ve hiç bir kontrolün olmadığı farklı platformlardaki verileri analiz etmek için bir sürü zorluklar ile karşılaşmaya başlamışlardır. Bu zorluklardan en belirginini veri yığınlarının yapılandırılmamış olmasıdır. Yapılandırılmamış bir veriden bilgi çıkartmak gerçekten çok zor bir uğraştır.

Anket ve görüřme yapılarak elde edilen veriler, işletmelerin müşteri ile ilgili işlemleri kayıt ederek elde ettiđi veriler ve nispeten enformasyon teknolojilerinin kullandığı veri tabanlarındaki veriler yapılandırılmıştır. Bu veri türlerinde toplanılması düşünölen veri ve özellikleri yöntemin tasarım aşamasında yapılandırılmaktadır. Bu veriler üzerinde belli kurallar çerçevesinde analizler yürütmek daha kolaydır. Burada da özellikle veri tabanlarındaki verilere ulaşım sorunu vardır. Veri tabanlarındaki veriler bir işletme için hayati öneme sahip olduğundan yöneticiler bu verileri kimse ile paylaşmak istememektedirler. Bu tür veriler genelde sahipli oldukları için başkaları tarafından ulařılmaz ya da kullanılması istenmesi durumunda veri sahibinin izninin alınması gerekmektedir.

İşletmelerde turist ile ilgili yapılan işlemlerin kayıtlarının tutulduğu operasyon destek sistemlerinin veri tabanları, bir işletme için müşteri verisinin depolandığı paha biçilemez bir kaynaktır. Otel işletmelerinde tesis yönetimi ve MİY sistemlerinde (Talón-Ballesteros vd., 2018 : 187), acenta ve tur operatörlerinin veri tabanlarında dâhili ve yapılandırılmış veya harici ve yapılandırılmamış (fikir platformları, sosyal ağlar veya cođrafi konumlandırma gibi pek çok bilgi) çok sayıda veri mevcuttur. Teknolojik gelişmeler ile birlikte turizmde kullanılmaya başlayan cođrafi bilgi sistemleri, özellikle turist davranışı konusunda arařtırmacılara yeterli veri sağlayabilmektedir. İnternet kullanımının yaygınlaşması ile birlikte web siteleri günlük hayatın

tam da içinde yer almaktadır. Cep telefonlarının akıllı hale gelmesi ile daha da çok yaygınlaşan sosyal medya kullanımı turizm ile ilgili, kullanıcı yorumları, değerlendirmeleri, coğrafi etiketli öğeleri ile araştırmacılara bir veri deryası sunmaktadır. Web, Sosyal Medya ve cep telefonlarının yanında, bluetooth, GPS gibi cihazlar turizm alanında veri kaynakları olarak yaygın olarak kullanılmaktadır.

Ancak turist ve turizm ile ilgili verinin çoğu yapısal olmayan formatta saklanmaktadır. İşletme içindeki yapılandırılmış verilerde dahi farklı firmalardan temin edilen yazılımlarda (örneğin, önbüro yazılımı ile muhasebe yazılımı farklı firmalardan) farklı şekilde veri yapısı olabilmektedir. Aynı işletmeye ait sistemler arasında bile veri uyumsuzlukları varken, dış kaynaklardan özellikle sosyal medyadan elde edilen veriler aynı kaynak içinde bile yapılandırılmamış olarak gelmektedir. Kullanıcılar tarafından oluşturulan ve paylaşılan veriler (KOİ) ise anonim ve yapılandırılmamış yorumlar, ürün ve hizmet değerlendirmeleri, konum bilgisi, fotoğraflar ve videolardan oluşmaktadır. Farklı platformlarda tutulan bu veriler genelde hiç bir denetim olmadan kullanıcıların yayınladığı ya da paylaştığı halleriyle tutulmaktadır. Bu verilerin genel özelliği, genelde sosyal ağ içerisinde paylaşıldıkları için ulaşılmaları kolaydır, halka açık veriler olduklarından izin alma ve veri sahibi ile irtibata geçme zorunluluğu yoktur. Ancak, web sitesi, sosyal medya gibi platformlardaki ulaşılmaları kolay olan, yetkilendirilmemiş verilerin yapılandırılmamış oldukları için analizleri zordur.

Veri kaynaklarındaki bu sıkıntı enformasyon teknolojilerini faaliyetlerin neredeyse tamamında kullanılmakta olan turizm sektöründe maalesef yönetici ve araştırmaların değerli bilgiye ulaşmasını zorlaştırmaktadır. Bazı işletmeler yoğun veriye sahipken müşterileri ve bazı işlemleri hakkında pek bilgi sahibi olamamaktadırlar (Ha ve Park, 1998: 1). Veri tabanlarında ve dış kaynaklarda kendileri ve müşterileri ile ilgili veri yığınları arasından pazarlamada, ürün yönetiminde kullanabilecekleri faydalı bilgileri elde edememektedirler. Verinin bilgiye, bilginin karara dönüşmesi için veri bütünlüğünü sağlayan, veri tutarlılığını garanti altına almada yardımcı olan teknolojileri doğru bir şekilde kullanmak önem taşımaktadır. Veri yığınlarından anlamlı yararlı bilgiler elde edebilmenin en bilinen yolu veri madenciliği yöntemlerinin kullanılmasıdır.

Veri madenciliği - anlamlı modeller ve kurallar keşfetmek için büyük miktarlardaki verilerin araştırılması ve analiz edilmesi için teknikler - işletmelerin görünüşte alakasız verilerden anlamlı ilişkiler çıkarmasına yardımcı olmaktadır (Mirela vd., 2009: 922). Veri madenciliği teknolojileri, tüketicilerin bilgisini keşfetmede kullanılmaktadır. Uzun süredir çalışan bir işletmenin müşteri alışveriş kayıtlarını içeren geniş ölçekli bir veritabanı olabilmektedir. Bu

nedenle, müşteri alışveriş niyetlerini anlamak için, pek çok uluslararası işletme, veri madenciliği teknolojileri ile büyük verileri (müşteri alışveriş kayıtları) ile POS (satış noktası) sistemlerinden müşteri bilgisini keşfetmektedir (Chiang, 2018: 665). Yalnız işletme bünyesinde kullanılan operasyon destek sistemlerinin ve müşteri ilişkileri yönetimi sistemlerinin veri yapısı ve depolanan veriler şirketler için stratejik bir varlıktır ve gizlidir. Bu nedenle, araştırma amaçlarına yönelik veri erişimi, veri istismarını önlemek için ortak ilgi alanları olan şirketler ve üniversiteler arasındaki işbirliği anlaşmalarıyla mümkün olabilmektedir (Talón-Ballesteros vd., 2018).

Bir konu üzerinde araştırma yapacak bir araştırmacının öncelikle o konu ile ilgili verilere ulaşması gerekmektedir. Bunun içinde o alandaki veri kaynağını ve veri özelliklerini bilmek zorundadır. Bu bağlamdan yola çıkılarak bu bölümde turizm ile ilgisi olan verilerin kaynakları, türleri, özellikleri ve bu verileri elde etmenin yolları üzerinde durulmuştur.

3.1. Anketler, Yüz Yüze Görüşmeler

Anket, araştırmacılar tarafından oluşturulan, bilgi alınacak kişilere doğrudan doğruya okuyup cevaplandırılacakları soruların hazırlanması ile yapılan bilgi edinme, bir çeşit yazılı iletişim tekniği ile uygulanan veri toplama yöntemidir. Soru ve cevapların yapılandırılmış ve yazılı olmasının bazı avantajları vardır. Verilerde standartlaşma olduğu için çeşitli karşılaştırmalara imkân vermektedir. Ayrıca kasıt olmadan verilen eksik ve yanlış bilgiler katılımcı tarafından düzeltilebilmektedir. Anketler ile toplanan verilerin ilgili programlar yardımıyla kısa sürede değerlendirilebilmesi anket yönteminin yaygınlık kazanmasında etkili olmuştur.

Görüşme (Mülakat - Interview) ise anketlerde olduğu gibi belirlenen sorulara cevap olarak veri toplama yöntemidir. Her ikisi de önceden belirlenmiş insanlara bir takım sorular sorularak uygulanmaktadır. Ankette katılımcı ya da anketi yapan kişi soruları okumaktadır. Görüşme biraz daha karşılıklı konuşma şeklinde olmaktadır. Bu da bazen olayların ya da durumların etkisi altında kalınarak görüşmecinin ya da görüşülen kişinin tutum ve tavırlarını etkileyebilmektedir. Görüşme yoluyla veri toplayabilmek için konuya yönelik yeteri kadar araştırma yapılmalıdır. Aksi takdirde elde edilen veriler doğruluktan uzaklaşabilmektedir.

Anket yöntemi ile insan davranışları, tercihleri, inançları, tutumları, duyguları, iş performansları, bilgi düzeyleri, v.b. gibi çok farklı türde veri toplamak mümkündür. Turizm sektöründe araştırmalar ve yöneticiler uzun yıllar boyunca, turistleri tanımak ve onların davranış ve tercihlerini öğrenerek ürün geliştirmek için gerekli veriyi anket ve görüşme yöntemleri ile toplamaktadırlar. Anket, turizm sektöründe hem yöneticiler hem de araştırmacılar tarafından en fazla kullanılan veri toplama yöntemlerinden birisidir. İşletmeler misafirlerine

ürün ve hizmetleri hakkında sorular sorarak müşteri memnuniyetini arttırmayı amaçlamaktadır. Araştırmacılar da araştırma konularına dair verileri elde edebilmek amacıyla anket düzenlemektedir. Anketlerle elde edilen veri anket hazırlığı aşamasında yapılandırıldığı için üzerinde analiz yapmaya ya da değerlendirilmeye hazır durumdadır.

Veri madenciliği büyük veri yığınlarından bilgi ve desenleri çıkartmak için kullanılmaktadır. Bununla birlikte turizm araştırmalarında bazı araştırmacılar anketler ile topladıkları veriler üzerinde veri madenciliği tekniklerini kullanarak analiz yapmışlardır. Bunun nedenlerinden birisi bazı ülkelerde bulunan Kişisel Verileri Koruma Yasası'nın kısıtlamasından dolayı araştırmacılar işletmelerin müşteri veri tabanına ulaşamamaktadır. Dolayısıyla, anket ile yapılan görüşme, veri toplamının tek yolu olmaktadır (Liao vd., 2010: 4214). Bu çalışmalardan bazıları Tablo 3.1'de verilmiştir.

Tablo 3.1 Anketler aracılığıyla Veri Toplayan Araştırmalar

Araştırma	Veri Seti	Toplanan Veri
Min vd., 2002	Anket	Güney Kore, Seul'deki 11 lüks otelin 281 müşterisine demografik veriler ve otel servislerinden memnuniyet seviyeleri
Bloom, 2005	Anket	Güney Afrika'da Cape Town'u ziyaret eden uluslararası turistlerin turizm gezisi, demografik, sosyoekonomik ve coğrafi özelliklerinin geniş bir karışımı
Bevanda vd., 2008	Anket	Öğrenciden yaygın olarak kullanılan Hırvat çevrimiçi seyahat acentesini temsil eden üç web sitesini karşılaştırmaları
Byrd ve Gustke, 2007	56 sorulu anket	İlçe sakinleri, işletme sahipleri, yerel yönetim görevlileri ve turistlere, turizm geliştirme ve yerel toplumla ilgili çevre algısı, yerel toplumdaki turizm etkisi algıları, toplum katılımı türleri ve miktarı, bireylerin rekreasyon faaliyetlerine katılımı konusundaki görüşleri ve temel demografik bilgileri
Delen ve Sirakaya, 2006	E-posta anketi	Oyun oynamanın yasallaştırılması için oylama verileri
Emel vd., 2007	Anket	Bursa'da seyahat acentası müşterilerine 32 sorudan (motivasyon, sosyo-demografik ve konaklama özellikleri) oluşan bir anket uygulanmış ve 224 anket elde edilmiştir.
Karathiya vd., 2012	Anket	Gujarat'ı en az beş günlük tur paketleri ile ziyaret eden 1000 ziyaretçinin tekrar ziyaret etme ve başkalarına tavsiye etme üzerine görüşleri
Liao vd., 2010	Beş bölümlü anket	Müşteri profili ile ürün segmentasyonu arasındaki ilişkileri tanımlamak için müşterilerin temel verileri, tercih verileri ve satın alma verileri
Wu vd., 2013	Telefonla Görüşme	Ankete katılanların demografik verileri ve otel web sitesi kullanım tecrübeleri, özellikle de bir otel web sitesinden arama veya satın alma durumları
Karl vd., 2015	Kişisel Görüşme	Veriler kişisel görüşmelerde standart bir anket (3 bölüm 22 soru) kullanılarak 2013 yılında Münih / Almanya'da toplanmıştır.
Li ve Kwortnik 2016	Çevrimiçi Anket	Katılımcıların 7 seyir hattında yolculuk yapıp yapmadıkları ve hatlar ile ilgili 5 alanda 25 soru soruldu
Nuankaew vd., 2017	Anket	Wat Nong Waeng ve Wat Phrathat Nadun ziyaretinde bulunan turistlerden (genel bilgiler, memnuniyet düzeyleri ve geri bildirimler) elde edilen veriler

Shapoval vd., 2018	Anket	Japonya Turizm Ajansı için yürütülen turist harcamaları anket serisinin bir parçası olarak Japonya'daki uluslararası havaalanları ve limanlarında toplanan (turist memnuniyetinin özellikleri, genel memnuniyet, geri dönüş niyeti ve demografik veriler) veri
Vajirakachorn ve Chongwatpol, 2017	Kayıt formları ve Anketler	Festival alanlarını tekrar ziyaret etme niyetini açıklamak ve tahmin etmek için demografik bilgiler, itme faktörleri, çekme faktörleri, algılanan değerler, algılanan kalite ve memnuniyet ile ilgili toplam 30 özellik hakkında veri
J. Wong vd., 2006	Anket	Kuzey Tayvan'dan yurt dışına çıkan turistlerin demografik özellikleri; karar verme bilgileri; ve ziyaret süresi dahil olmak üzere ziyaret ettikleri yerler hakkında veri
Alexander vd., 2018	Anket	Doğal kaynak kullanım alanlarına ilişkin mekânsal veriler elde edilmiştir ve bu veriler ArcGIS'te arazi alanı kullanım taslak haritasını hazırlamak için kullanılmıştır
Chiang, 2018	Anket	Tayvan'ın en büyük üç şehrindeki (Taipei, Taichung ve Kaohsiung) kahve tüketicileri hakkında veri
Hausmann vd., 2018	Anket	Kruger National Park içinde çekilip Instagram and Flickr'da paylaşılan fotoğraflara ek olarak parka gelen 563 turist ile anket yapılmıştır
Law vd., 2011	Anket	Hong Kong giden turizmi yıllık yurt içi anketlerinden (seyahat etkinlikleri, motivasyon ve demografik veriler) toplanmıştır.
Valls vd., 2018	Anket	Turistlerin karar vermesinde itme-çekme faktörlerinin rolünü belirlemek için, İngiliz turistlerin tatildeki motivasyonları ve anlamlarına dair deneysel araştırmalar, Güney Katalonya'da bir İspanya kıyı turizm bölgesi olan Tarragona'da gerçekleştirildi. Anket ile, kişisel bilgiler, motivasyonlar, turistik etkinlikler, destinasyon özellikleri ve memnuniyetleri konusunda veri toplanmıştır.

3.2. Seyahat Günlükleri

Bir seyahat günlüğü, turistik faaliyetler ile ilgili ayrıntılı bilgiler sağlayabilmektedir. Günlükler, turistlerin günlük hareketlerini, ziyaret edilen yerleri, etkinlikleri, günün saatini, her bir yerde harcanan süreyi ve varış noktasında gezinmek için ulaşım türünü kaydetmek için kullanılmaktadır. Seyahat günlüğü, seyahat aktiviteleri için umut verici bir veri kaynağıdır çünkü geleneksel anketlerle karşılaştırıldığında turist deneyimleri hakkında daha detaylı bilgi elde edilebilir (Vu vd., 2017:). Bir seyahat günlüğü turistin seyahat ettiği her destinasyon ile ilgili detay bilgiyi (yazı, fotoğraf), kendi ve çevresindeki turistlerin duygu ve düşüncelerini, destinasyon içindeki tüm aktiviteler hakkında bilgileri içermektedir.

Seyahat günlüğü yapımında önemli bir problem, veri toplama sürecine katılımlarını sağlamak için turistlerle doğrudan iletişime geçilme zorunluluğudur. Turistlerin bu kadar kapsamlı bir günlük yazmaya ikna edilmesi ya da kendileri için yazdıklarının ve topladıklarının araştırmalar için alınıp kullanılması zor olmaktadır. Bunun için geleneksel seyahat günlükleri turizm araştırmalarında veri kaynağı olarak kullanılmamıştır. Kâğıt tabanlı seyahat günlüklerine ek olarak, seyahat verileri küresel konumlandırma sistemi (GPS) teknolojisine sahip mobil cihazlar kullanılarak da toplanabilmektedir. Fakat bunun için de turistlerin seyahatlerini

manuel olarak belgelemek için zaman harcaması gerekir ve mekânsal - zamansal seyahat verilerini kaydetmek için mobil cihazların satın alınmasında ekstra maliyetler oluşmaktadır. Böylece, toplanan veriler sınırlı bir süre için küçük ölçekli turizm yerlerinde minimum cevaplarla sınırlı kalmaktadır (Vu vd., 2017).

Seyahat günlüklerini yayınlayan web siteleri üzerinde ise turistlerin yazıları, fotoğrafları, videoları bulunmaktadır. Bu yazarlar kendi istekleri ile ya da belli bir ücret karşılığı seyahat anılarını okuyucuları ile paylaşmaktadırlar. Bu veriler bir seyahat ile ilgili tüm detayları içermekle birlikte bir bölgeyi ziyaret edenlere oranla o bölgede günlük tutanların oranı genelleme yapılamayacak kadar az olmaktadır.

3.3. Operasyonel Sistemler

Bilgi, turizm sektörü için hayati olduğundan (Mirala vd., 2009), ilgili teknolojilerin etkin kullanımını gerektirmektedir. Ancak turizmde ilgili teknolojilerin kullanımı bilgisayar ağlarının ortaya çıkışına kadar, muhasebe kayıtlarını tutma, fatura yazdırma, vergi beyannameleri bastırma, personel maaşlarını hesaplama gibi işletme içi güncel uygulamalar ile sınırlı kalmıştır (Mirela vd., 2009: 923). Bilgisayar ağlarının ortaya çıkmasıyla birlikte turizm alanında enformasyon teknolojileri daha yoğun olarak kullanılmaya başlanmıştır. Son yıllarda enformasyon teknolojilerinde yaşanan gelişmeler sonucunda bilgisayara dayalı otomasyon sistemleri turizm sektörünün en önemli bir altyapısı haline gelmiştir

Seyahat kararı veren bir müşterinin satın alacağı ürünü seçerken, ürün hakkında bilgileri dağıtmak için, müşterinin satın aldığı ürüne yolculuğu esnasındaki bilgilerin kayıt edilmesi için, konaklayacağı otel ve konaklama esnasında yaptığı her türlü işlemin kaydı ve bu kayıtların raporlanması, faturalanması ve muhasebeye aktarılması için enformasyon teknolojilerine ihtiyaç vardır. Bazı işletmelerin yüz binlerce yolcunun kaydını tuttuğu düşünülürse, klasik yöntemlerde bu kadar yolcu bilgisinin saklanması, tasnif edilmesi ve özellikle ihtiyaç duyulduğu anda hızlı bir şekilde ulaşılması neredeyse mümkün değildir. Geleneksel yöntemlerle dağınık bir şekilde olan ve çeşitli aşamalardan sonra bulunabilen veri, enformasyon ve bilgiler, enformasyon teknolojileri sayesinde daha kolay, hızlı ve güvenli biçimde bulunabilmektedir.

Turizm sektöründeki hemen her işletme günlük faaliyetlerinin kayıt altına alınması için enformasyon sistemleri kullanmaktadır. Oteller, acentalar, lokantalar, alışveriş merkezleri, araç firmalarına veri tabanlarında, firmalar arası işlemlerin ve turistlere yönelik yapılan işlemlerin kayıtlarından oluşan verileri depo etmektedirler. Bu veriler işlem anında kaydedilmekte, sonrasında gerekli liste ve raporların alınmasından sonra veri tabanında yaşlanmaya terk

edilmektedir. Sektörde operasyon destek sistemlerinde tutulan verilere yönelik yeterince çalışma bulunmamaktadır.

Sektörde farklı amaçlara yönelik iş yapan farklı işletmelerin bünyelerinde farklı operasyon destek sistemleri çalışmaktadır. Otellerde tesis yönetim sistemi ve müşteri ilişkileri yönetim sistemleri, seyahat acentaların incoming ve outgoing acenta yazılımları yaygın olarak kullanılmaktadır. Lokantalarda menü, reçete ve stok takibi ve günlük servis takibi yapan yazılımlar ağırlıklı olarak kullanılmaktadır. Fatura ve muhasebe işlemlerin takibi tüm işletmelerde operasyon destek sistemleri aracılığıyla yapılmaktadır.

İşletmelerin operasyon destek sistemlerine olan ihtiyacı karşılamak için birçok yazılım firması yazılım sistemleri geliştirmiştir. Bazı işletmelerde yazılım firmalarına ait otomasyon sistemlerini satın almakta ya da kiralamakta, bu da önemli bir gider kalemi olarak karşılımlarına çıkmaktadır. Bazı işletmeler de yazılım şirketlerinin piyasaya sürdüğü yazılımları satın almak yerine kendi bünyelerinde yazılım ekibi kurarak kendi otomasyonunu kendileri kurup, geliştirmektedir.

Seyahat işletmelerinde enformasyon sistemleri rezervasyon, operasyon, günlük tur, fatura ve muhasebe gibi günlük operasyonel işlemlerin kayıtlarının tutulması, saklanması, gerektiğinde girilen bu kayıtlardan raporlar oluşturulması için yaygın olarak kullanılmaktadırlar. Girilen rezervasyonlardan otel giriş listeleri, operasyon listeleri gibi karşılayıcı acentanın en çok yaptığı işlemlerden olan transfer işlemleri için rehber, transferman, şoför ve müşteri bilgilerinin baz alındığı raporlar çıkartılır. Bu raporlarla alanda çalışan karşılama (arrival), uğurlama (departure) ve tur operasyon işlerini takip eden acenta personelinin işlerini kolaylaştırmaktadır. Son nesil acenta yazılımlarında çıktı yerine alandan da rahatça acenta veri tabanına ulaşan cihazların kullanımına geçilmiştir. Bu sistemler hem kâğıt için harcanan paradan tasarruf yaptırmakta hem de son dakika eklenen müşterilerin havaalanında unutulmalarını engellemektedir. Acentanın günlük tur ile ilgili bilet satışı, turların düzenlenmesi, toplama ve tur araçlarının belirlenmesi, tur esnasında yapılan masrafların kaydı, özellikle alışveriş turlarında oluşan komisyonların takibi, tur rehberlerine ve turdan komisyon alan diğer kişi ve kurumlara ait komisyonların takip edilmesi ilgili yazılımlar aracılığıyla yapılmaktadır.

Enformasyon sistemleri otel işletmelerinde rezervasyon, önbüro, kasa, kat hizmetleri, seyahat acentalarında rezervasyon, operasyon, günlük tur için günlük operasyonel işlemlerin kayıtlarının tutulması, saklanması, gerektiğinde girilen bu kayıtlardan raporlar oluşturulması için yaygın olarak kullanılmaktadırlar. Tesis yönetim sistemlerinde rezervasyon modülünde müşteriye ait kişisel bilgiler, tercih ettiği odaya ait bilgiler, faturalama modülünde ise, harcama ve ödeme bilgileri, otel içerisinde tercih ettiği ürünler ve bu ürünlere ait işlemler tutulmaktadır.

Fatura ve muhasebe kayıtlarının tutulması için tüm işletmelerde muhasebe programları yasa gereği zorunlu olarak kullanılmaktadır. Yukarıda bahsedilen günlük operasyonel ve finansal işlemlerin son durumları ya da istenen bir andaki durumları için kurum yöneticilerinin bilgilendirilmesi için gerekli liste ve raporlarda ilgili sistemlerin için de genelde uyumlu olarak kullanıcılara sunulmaktadır.

Bazı programlar kurum içindeki haberleşme ve iş takibinin sistem üzerinden yapılmasına olanak sağlamaktadırlar. Kurumda yapılacak tüm işler sisteme girilerek, departman yöneticileri aracılığıyla ilgili elemanlara dağıtılır. İşin başlama ve bitiş saatleri, işin yapılış aşaması ve bu esnadaki tüm safhalar sistem üzerinden takip edilebilmektedir. Bu sistem sayesinde işlerin ortalama ne kadar sürede bitirilebildiği, personelin günlük, haftalık, aylık ve yıllık bazda ne kadar iş yaptıklarının takibi, müşterilerin şikâyetleri, en fazla şikâyet ettikleri konu, şikâyetlerinin ortalama ne kadar sürede çözüldüğü gibi müşteri memnuniyetini etkileyecek kalite takipleri yapılabilmektedir. Bu tür sistemlerde kurum içinde kimse sözlü iş yapmamakta ve herhangi bir müdür ya da şef bir elemana yapması için bir iş vermemektedir. İşler sadece sisteme girilerek yapacak kişinin amirinin ataması ile yapılmaktadır. Bu da kurum içinde görev dağılımı anlamında bir eşitlik sağlayarak, personelin herkes tarafından ilgili ilgisiz işlerde görevlendirilmesini engellemektedir.

Tablo 3.2 Operasyonel Sistemlerden Veri Toplayan Araştırmalar

Araştırma	Veri Tabanı	Toplanan Veri
Septiadi vd., 2016	Otel ve bilet bilgileri	Bilet satış verileri: işlem tarihi, havayolu, rota, fiyat, ödeme yöntemleri ve müşterinin cinsiyeti. Otel satış verileri: misafir cinsiyet, check-in tarihi, check-in günü, otel isimleri, otel-yıldız, oda ve gün miktarı, oda fiyatları ve ödeme şekli. Bilet satış 8853 ve otel satış 4912 veri içermektedir.
Brunk vd., 2018	Müşterilerin asansör kullanımı	Avrupa'daki en büyük kayak merkezlerinden ikisinden iki ayrı veri seti alındı. Veri setleri 2015 / 2016 kış sezonunda her müşterinin tüm asansör faaliyetlerini içermektedir. Toplamda, iki veri kümesi birlikte 30 milyondan fazla asansör yolculuğu ve dolayısıyla olay günlüğü girişini içermektedir.
Grinberger ve Shoval, 2019	Turistin Konumu	Veri tabanı, Kudüs Ulaştırma Master-Plan Ekibi (JTMT) tarafından yapılan bir turist seyahat anketi kapsamında toplanan tüm ziyaretleri boyunca İsrail'de turistlerin hareketini gösteren GPS izlerini içermektedir.
Scherrer vd., 2018	Kullanıcı Hareketleri kaydı	1 Ocak - 31 Ocak 2016 tarihleri arasında Avustralya'da 71.207 benzersiz kullanıcının hareketlerinin kaydedildiği veri seti. Veriler pozisyonlu (enlem / boylam) ve zaman damgalı GPS kayıtlarından oluşmaktadır.
Ha ve Park, 1998	Müşteri Bilgileri	Gümrüksüz satış mağazasına üye olanlara ait müşteri verilerinden RFM (yenilik, sıklık, parasal) değerleri, eğitim ve test verileri olarak alınmıştır.

Kurum içindeki insan kaynaklarına ait evrakların, primlerin, beyannamelerin takibi için de enformasyon teknolojileri kullanılmaktadır. Fatura, irsaliye, sipariş gibi resmi ve resmi olmayan evrakların takibi, bunların muhasebeleştirilmesi ve vergi, bütçe, gibi işletmenin finansal bilgilerinin takibi de yeni ilgili yazılımlar aracılığıyla yapılmaktadır.

3.4. MİY Sistemleri

Yeni pazarlama felsefesi; müşteri temeline dayanan ve bilgi teknolojisi ile desteklenen, organizasyonun içinde ve dışında birlik yaratan grupların oluşturulmasıyla yönetilmektedir. Bu bağlamda müşterilere kaliteli hizmet vermek amacıyla, öncelikle müşteri hizmetleri anlayışı yeniden yapılandırılmalı ve müşterilerin dahi kullanabileceği müşteri ilişkileri yönetimi sistemi kurulmalıdır. Bir MİY sistemi “tüketicilerle karlı ilişkiler geliştirmek, sürdürmek ve güçlendirmek için tüketici bilgisini geliştirmek ve güçlendirmek için teknolojiye dayalı sağlam bir araçtır” (Talón-Ballesteros vd., 2018: 188). İşletmelerde MİY sistemleri genelde operasyon destek sistemleri bünyesinde ya da operasyon destek sistemleri tarafından beslenen farklı bir sistem olarak hizmet vermektedirler. Bir MİY sistemi, çoklu kanal müşteri hizmetleri, anlık veri analizi, kullanılan sistemlerde entegrasyon, yönetmeliklere uyum ve sertifikasyon, raporlamalar, deneyim ve bilgi bankası hizmetleri verebilmektedir. MİY sistemi, tüm müşteri talep ve önerilerini tek kanaldan, en etkin şekilde, bilişim programları üzerinden yönetmeye imkân verebilmektedir.

MİY yazılım ve donanım uygulamalarının iş operasyonlarının, iş performans süreçlerinin ve iletişim otomasyonu ve süreçlerin koordinasyonu olmak üzere üç anahtar süreci vardır (Güleş, 2004: 237). MİY sistemleri ile iş performans süreçlerinin otomasyonu çerçevesinde veri depolama, verinin analizlere uygun hale getirilmesi ve veri madenciliği işlemlerinin yapılabilmesi için hazırlanması işlemleri yapılır. MİY uygulamalarında kullanılan veri madenciliği, müşterilere ilişkin elde edilen bilgilerin kullanımını kapsamaktadır.

MİY'in turizm sektöründe uygulanması, henüz gelişim aşamasındadır. Şirketlerin MİY'in turizmde uygulamasının gücünü ve önemini fark edebilmeleri belli bir süre gerektirmektedir. Sınırlayıcı faktörlerden biri, bu teknolojilerin kullanılması ve kullanılması için gereken finansal kaynaklardır. MİY sistemlerinin turizm ve konaklama işletmelerinde gelişmesinin diğer bir sınırlayıcı faktörü de, misafir ihtiyaç ve davranışlarının izlenmesi için temel teşkil eden sağlık ve spa tesisleri, spor tesisleri ve diğer hizmetler ile ilgili eksiklikler ve kötü bilgisayar ekipmanlarıdır (Ivanovic vd., 2011: 62). Yönetimi desteklemek için enformasyon sistemleri nadiren kullanılmaktadır ve MİY sistemlerinin turizm sektöründe kullanımı asgari

düzeydedir ve çoğunlukla lüks otellerde nadiren uygulanır. Diğer küçük oteller ve alt kategorilerdeki oteller için MİY, çoğunlukla eski konukların verilerini incelemeye odaklanmaktadır.

Günümüzün rekabet ortamında işletmelerin başarılı olabilmeleri için mevcut müşterilerini elde tutmaları, bunun için de onlara özel ürünler ve fırsatlar sunmaları gerekmektedir. Bunun da ilk adımı müşteriye ve ne yaptığını tanımaktır. Müşteri odaklı bir işletme müşteri hizmetlerine yapılan her çağrıyı, her satış noktası işlemi, her katalog siparişini, her web sitesi ziyaretini müşterileri hakkında bir şeyler öğrenebilme fırsatı olarak görmektedir (Berry ve Linoff, 2004: 4). Bu verilerin toplanması ve ardından analiz edilebilmesi için iyi tasarlanmış bir MİY sistemine ihtiyaç bulunmaktadır.

Talón-Ballesteros vd. (2018)'nin çalışmasında kullandığı veritabanı uluslararası çapta geniş çapta bilinen bir uluslararası otel şirketi tarafından rutin olarak kullanılan bir MİY sisteminden derlenmiştir. MİY sisteminde 2013 ve 2014 yıllarında 4.935.806 farklı müşterinin (zincirdeki otellerde bir gece kalanlar) bilgileri kaydedilmiştir. Demografik değişkenler, davranış değişkenleri ve işlem verileri de dâhil olmak üzere toplam 18 değişken ve bunların ilgili kategorileri analiz edilmiştir. Elde edilen sonuçlar, MİY'in veri yapısını ve olası değişikliklerini yeniden gözden geçirmek amacıyla otel zincirinin yöneticilerine yararlı bilgiler sağlamıştır.

3.5. Coğrafi Bilgi Sistemleri

Coğrafi bilgi sistemleri (CBS), dünyanın yüzeyi ve yakını hakkında bilgi oluşturmak, elde etmek, birleştirmek, dönüştürmek, görselleştirmek, analiz etmek, modellemek ve arşivlemek için çeşitli işlevler sağlayan çok büyük yazılım paketleridir (Goodchild, 2006: 251). Uzayda ve çoğunlukla üç boyutlu uzayda yerleri sıcaklık, nüfus yoğunluğu, arazi kullanımı veya yükseklik gibi özelliklerle ilişkilendirmektedirler. Günümüzde coğrafyada ve dünya yüzeyiyle ilgilenen diğer disiplinlerde araştırmayı desteklemek için yaygın olarak kullanılmaktadır. CBS uygulamaları çeşitli kullanım alanlarına ve tematik konulara yönelik olarak geliştirilmektedir. Kent, Orman, Karayolları, Arazi, Tapu ve Kadastro, Lojistik, İç Güvenlik, Araç İzleme, Trafik, Kampus, Deprem ve Harita Bilgi Sistemi CBS uygulamalarına örnek olarak gösterilebilir.

CBS ile ilgili çalışmalar 1970'li yıllarda başlamış olsa da o dönemde bilgisayar maliyetlerinin yüksek olması kullanımlarını sınırlamıştır. 1990'ların yarısından sonra, internetin de popülerleşmesinden sonra CBS'ler kişisel bir analitik motor olmaktan ziyade bilgi iletişimi için bir araç olarak görülmeye başlanmıştır. Özellikle CBS tasarımcılarına ölçüm ve analiz üzerinde yoğunlaşmalarını sağlayan veri yönetiminin ayrıntılarını devralan ilişkisel veritabanı yönetim

sistemlerinin geliştirilmesi ve ardından kişisel bilgisayarların piyasaya sürülmesi ile fiyatların düşmesi CBS'lerin ticari olarak kullanılmasının önünü açmıştır.

Bugün, tipik bir CBS, başlıca coğrafi bilgi türlerinin tümünün ele alınması ve görselleştirilmesi ve dönüşümünden detaylı analiz ve modellenmeye kadar çok çeşitli işlevleri gerçekleştirebilmektedir. Bir CBS'nin gücü (1) karmaşık veri setlerine erişmek ve almak, (2) karşılaştırma ve analizlerine imkân vererek farklı iki seti birbirine eklemek, (3) verilere sayısız istatistiksel ve diğer prosedürler uygulamak (4) güçlü bir görselleştirme yöntemleri dizisi listeleme yeteneğinde yatmaktadır. CBS, bu gücü ve elde edilmesinin ucuz olması nedeniyle, eğitimden, bir bilimsel araştırmanın yönetilmesinden politika oluşturmaya kadar çeşitli amaçlar için kullanımı yaygın olarak benimsenmiştir. Genel olarak, dünyanın yüzeyi ve yüzeyiyle ilgilenen herhangi bir alan artık CBS'i başarısı için temel kabul etmektedir. CBS teknolojisi bilimsel araştırmalar, alt yapılar (elektrik, doğalgaz, telefon, su), askeri uygulamalar, kaynak yönetimi, maden haritalama, varlık yönetimi, arkeoloji, çevresel etki değerlendirmesi, peyzaj mimarlığı, kentsel planlama, arama kurtarma, lojistik, kartografya, kriminoloji, coğrafi tarih, pazarlama, haritacılık, tarım alanlarının tespiti, araç takip sistemleri, meteoroloji ve diğer amaçlar için kullanılabilir.

CBS veritabanları, haritaların üretilmesinden ürün ve mağaza konumunun piyasa alanlarının analizine kadar çeşitli amaçlar için kullanılabilir. Geleceğe bakıldığında, konum araştırmalarını desteklemede CBS'in gittikçe artan bir rolü olabilecektir. Örneğin, CBS doğal afetlerde acil müdahaleleri kolaylaştırmak için hesaplamalar yapmaya olanak tanımaktadır. CBS yeni sulak alanların bulunmasında ve bu alanların korunmasında kullanılabilir. Turizmde ise yıllardır turist hareketliliğini, dolayısıyla, turistlerin tercih ettiği bölgeleri tespit etmek için kullanılmaktadır. Seyahat davranışını anlamak, piyasayı tahmin etmeyi veya uygun seyahat önerileri sunmayı amaçlayan turizm uygulayıcıları için önemlidir. GPS, Cep telefonu ve coğrafi etiketli öğeler kullanılarak turistlerin hareket verileri elde edilebilmektedir. Araştırmacılar, bu verilerin mekânsal özelliklerini analiz etmek, ziyaretçilerin mekânsal davranışlarının farklı yerler ve peyzaj konfigürasyonları ile nasıl bağlantılı olduğunu anlamak ve rekreasyon alanlarının farklı gruplar tarafından nasıl kullanıldığını incelemek için coğrafi bilgi sistemlerini kullanmıştır (Orellana vd., 2012: 672).

3.6. Küresel Konumlandırma Sistemi

Küresel konumlandırma sistemi (GPS), dünyadaki enlem ve boylamlardan oluşan mevcut pozisyonu tespit eden bir sistemdir (Nagao vd., 2004). Sistemi kullanan tipik uygulamalardan birisi de araç navigasyon sistemidir. Son zamanlarda PDA (Kişisel Dijital Yardımcı)

ve cep telefonlarında kullanılabilir hale geldi, böylece GPS kullanan birçok mobil uygulama geliştirilmeye başlanmıştır.

GPS'te, mevcut konum, jeosenkron yörüngede yirmi dört uydu yeryüzüne gönderilen radyo sinyalleri kullanılarak gerçek zamanlı olarak hesaplanmaktadır. Bu uydular, ABD'deki Savunma Bakanlığı tarafından yönetilmekte ve işletilmektedir. Radyo sinyalleri, GPS uydularının sahip olduğu yörünge ve zaman bilgilerini içermektedir. GPS alıcısı, radyo sinyallerini alır ve içindeki bilgiyi kullanarak mevcut konumu algılar. GPS alıcısı, üç GPS uydusundan radyo sinyallerini alabildiğinde mevcut bir pozisyonu tespit edebilmektedir.

Başlangıçta tamamen askeri amaçlar için kurulan sistem, 1980'lerde sivil kullanıma da açılmıştır. Sivil GPS uygulamasının radyo sinyali, SA adlı bir hatayı içeriyordu çünkü GPS, askeri uygulamalar için tasarlanmıştır. SA, GPS uydularının gönderdiği zaman ve yörünge bilgisine çarpıtma verilerek üretilmektedir. SA 2 Mayıs 2000'de Japonya'da iptal edilmiştir. Dolayısıyla günümüzde GPS alıcısı doğru konum algılamasını uygulayabilmektedir. Bu da her bir turist için turist varış yerinin konumunu, varış ve kalkış saatini ve güzergâhını içeren kişisel turizm faaliyeti bilgisinin GPS kullanılarak elde edilmesini sağlamıştır ve turist hareketliliğini, dolayısıyla, turistlerin tercih ettiği bölgeleri tespit etmek için kullanılmaktadır. İnsanların gerçek zamanlı hareketlerini kaydedebilen bir teknik olarak, turist yönetimi için çok yeni bir potansiyel getirmektedir (Li vd., 2016: 31). Yalnız kayıt cihazlarının dağıtımı, takip edilen bireyin doğrudan işbirliğini gerektirmektedir (Versichele vd., 2014: 68). Ayrıca bu cihazların temini anket gibi diğer veri toplama yöntemlerine göre daha pahalı olmaktadır.

GPS alıcısındaki konum tespiti, GPS uyduları arasındaki mesafeler kullanılarak yapılmaktadır. Mesafe, radyo sinyalinin GPS alıcısına ulaşana kadar geçen süre esas alınarak hesaplanmaktadır. Radyo sinyalinin GPS alıcısına ulaştığı anda iyonosfer ve troposferin etkileri sonucunda bozulma ortaya çıkmaktadır. Ayrıca, GPS alıcısının yakınındaki nesnelere tarafından yansıtılan radyo sinyalini alma ihtimali de mevcuttur. Mesafe hesaplamasının doğruluğu, GPS alıcısının durumuna da bağlıdır. Bu nedenlerden dolayı GPS alıcısı tarafından nihayet tespit edilen pozisyon, SA iptal edilmesine rağmen hata içermektedir. Standart GPS'de, konum tespitinde hata önlenemez. Konum tespitindeki hatayı azaltmak için DGPS (Diferansiyel GPS) ve RTK-GPS (Gerçek Zamanlı Kinematik GPS) geliştirilmiştir. DGPS ve RTK-GPS kullanılarak son derece hassas konum tespiti gerçekleştirilebilmektedir. Ancak, kişisel turizm faaliyeti bilgileri araştırması için bunları kullanmak zordur, çünkü cihazın boyutu ve başlama zamanı bakımından sorunları vardır (Nagao vd., 2004: 3). Bu nedenle, GPS tespit verilerinden doğru kişisel turizm etkinliği bilgisini elde etmek için konum saptamadaki hataya karşı gerekli önlemler alınmalıdır.

Mobil internet teknolojisinin ve coğrafi bilgi teknolojisinin son zamanlardaki gelişimi ile bireysel turistlerin mekân-zaman yolunu doğru bir şekilde izlemek ve kaydetmek mümkündür. Son zamanlarda, seyahat spatial-zamansal yörünge verileri elde etmek için GPS izleme teknolojisi, turist davranışının çalışmasında uygulanmıştır, ancak kullanımı kümelenme ve model analizi düzeyinde açıklamayla sınırlı kalmıştır. Kişinin tercihlerine ve geçmiş turistlerin deneyimlerine dayanarak yeni gelenlerin potansiyel lokasyonlarını tahmin etmek ve önermek, hem turistler hem de turizm ürün tedarikçileri için bir anlam ifade etse de, son zamanlardaki az sayıda çalışma turist hareketinin trendlerini turistik cazibe merkezi seviyesinde tahmin etmeye çalışmıştır (Zheng vd., 2017: 268).

Zheng vd. (2017) Pekin, Çin Yaz Sarayı'nda GPS kullanan 117 turistin hareket bilgilerini toplamış ve 111 geçerli yörünge elde etmişlerdir. Bir odak turistin, o anki yörüngesine ve önceki turistlerin deneyimlerine dayanarak potansiyel konumunu tahmin etmek için veri madenciliğine dayalı bir sezgisel algoritma tasarlamışlardır. Birenboim vd. (2013), İspanya'nın Katalonya kentindeki PortAventura tema parkındaki ziyaretçilerin zaman-mekân yörüngelerini takip etmek ve kaydetmek için GPS (Global Positioning System) teknolojisini kullanmıştır. Nagao vd. (2004) turizm amaçlı kiralanan araçların GPS loglarını kullanarak kişisel turizm aktivite bilgilerinin çıkartılmasını gerçekleştirmiştir.

3.7. Web Siteleri

Bir bilgisayar ağı, iki veya daha fazla bilgisayarın kablolu ya da kablosuz iletişim araçları üzerinden yazılım ve donanım bileşenleri ile birlikte bağlanması ile meydana getirilen sistem bütünüdür. Burada asıl amaç, kaynakların ve bilginin (veri, ses, görüntü veya video) paylaşılması ve kişiler arasında iletişimin sağlanmasıdır. İletişimin sağlanabilmesi için üç elemana ihtiyaç vardır. Ağ servisleri, aktarım ortamı ve protokoller. Ağ servisleri iletişim hâlindeki bilgisayarların paylaştıkları olanaklardır. Ağda bazı bilgisayarlar servis sağlayan (Server: Sunucu), bazı bilgisayarlar da servis isteyen (Client: İstemci) konumdadırlar.

Bilgisayar ağının dünya çağında olması, dünyadaki bilgisayarların birbirine bağlanabilmesi durumunda adı Internet olmaktadır. İnternet, birbiriyle tüm dünya üzerinde yayılmış bilgisayar ağlarının birleşiminden oluşan devasa bir bilgisayar ağı olarak tanımlanabilir. Diğer ağlarda olduğu gibi internette de servis sunan, servis alan makineler mevcuttur ve bunların arasında iletişimin sorunsuz olmasını sağlayan protokoller vardır. Protokoller, bilgisayarlar arası iletişimde kullanılan ağ dilleridir (kurallarıdır). Ağ protokolleri verinin cihazlar arasında nasıl taşınacağını ve veri ile hangi bilgilerin gönderileceğini belirler. En sık kullanılan ve bilinen protokol TCP/IP protokol grubudur.

Ağdan yararlanabilmek için ağ üzerinde hizmet veren servislerden birisini kullanmak gerekmektedir. Bu servislerden en yaygın olarak kullanılanı da web (www)'tir. Web, internet üzerinde yazı, ses, grafik, resim ve hareketli görüntülerden oluşan belgeleri ağ üzerindeki diğer bilgisayarların erişimine açmaktadır. World Wide Web veya www, internette yayınlanmış olan birbirleriyle ilişkili hiper-metin dokümanlarından meydana gelen bir enformasyon sistemidir. World Wide Web, belli bir hiper-metin dili kullanan ve internet tarayıcısı (web tarayıcısı) adı verilen, kullanıcı dostu bir grafiksel arayüz yardımıyla kullanılır.

World Wide Web üzerinde bulunan her bir doküman veya sayfaya kısaca "Web Sayfası" denir. Web sayfaları içerisinde metin, görsel ve video klipleri içerebildikleri gibi diğer web sayfalarına gitmeyi kolaylaştıran bağlantılar da içerebilir. Web sayfalarının toplamını bir araya getiren oluşuma da web sitesi adı verilmektedir. Her web sitesinin internet üzerindeki yerini belirleyen bir adresi vardır. Buna URL (Standart kaynak bulucu: Uniform Resource Locator) denilmektedir. Web sitelerine IP adresi yerine site ismi yazılarak ulaşılmaktadır. Bir web sitesine ait IP adresine ulaşmak ve sunucuya erişmek için IP adresinin yerine ismini yazarak ulaşılmasını sağlayan sistem alan (Domain) adıdır. Her ülkede bulunan domain server aracılığıyla IP adreslerine bir isim atanır. Domain server isim ile yapılan aramayı web sitesi dosyalarının barındırıldığı sunucuya (Hosting) yönlendirir. Kullanıcı sunucuda paylaşım sunulan bilgileri web tarayıcısı aracılığıyla kullanır.

Web sitelerinin en önemli yanı bilgi vermektir. İşletmelerin tüm faaliyetlerini, müşterilerine sunduğu hizmetleri veya ürünleri ve tüm iş faaliyetlerini müşterilere sunmaya yardımcı olmaktadır. Geniş bir çevreye bilgi paylaşımı yapmayı sağlar. İşletme ile ilgili olaylar, ilanlar, duyurular ve haberler web üzerinde paylaşılabilir ve çalışan ya da diğer kesimlere ulaştırılabilir. E-Ticaret ile internet üzerinden alışveriş yapmayı, ürünleri satmayı sağlar. Web siteleri çağımızın silahıdır. İnternet kullanımının yaygınlaşması bu gücü daha da arttırmaktadır.

Tüm ürün ve hizmetlerde olduğu gibi web'de belirli süreçlerde kendini yenileyerek günümüze gelmiştir. Şu anda, web evrimini Web 1.0'dan Web 5.0'a kadar gerçekleştirmiştir. Web 1.0, tüketiciler ve web siteleri arasında minimum etkileşim sağlayan statik ve salt bir web biçimidir. Bu biçim, kullanıcıların web sayfası içinde bilgi yayınlamasına veya paylaşmasına izin vermemektedir (Tavakoli ve Mura, 2018). Bir sonraki nesil olan Web2 ise "Sosyal Web" olarak bilinmektedir ve kullanıcıların bilgi yayınlamasına ve paylaşmasına izin vermektedir. İkinci nesil internet (Web 2.0) olarak da adlandırılan bu teknoloji, genellikle insanların çevrimiçi işbirliği yapmasını ve bilgi paylaşmasını sağlayan araçları içermektedir. Bunların

örnekleri arasında, bunlarla sınırlı olmamak üzere, sosyal ağ, anlık mesajlaşma, sosyal etiketleme, mash-up'lar, bloglar, sanal dünyalar, podcast'ler, web videoları ve wiki'ler bulunmaktadır (Berezina vd., 2016: 5).

2006 yılında John Markoff tarafından başlatılan web'in üçüncü kuşağı, "Semantik Web" olarak da bilinmektedir. Web 3.0 fikri, veri madenciliği sürecinin iyileştirilmesi, önerilerle birlikte akıllı arama, geliştirilmiş yazılım araçları ve yenilikçi kişiselleştirilmiş teknikler gibi mevcut hizmetlerin işlevlerinin iyileştirilmesine dayanmaktadır. Dahası, çok kullanıcı sanal dünyalar, bu web sürümünün altında sınıflandırılmaktadır. Bu tür ortamlarda kullanıcılar kendilerini avatar olarak temsil edebilir ve diğer kullanıcılarla etkileşime girebilirler (Patel, 2013). Örneğin, Starwood Hotels, sanal turistlerin sanal oteli ziyaret edebilecekleri ve sanal ortamda gerçek mekânın tasarımını birlikte yaratabilecekleri Second Life'da "sanal Aloft"u başlatmıştır (Tavakoli ve Mura, 2018).

'Yaygın web', 'simbiyotik web' ve 'Ultra Akıllı Elektronik Ajan' Web 4.0'ı tanımlamak için kullanılan terimlerdir (Patel, 2013: 416). Bu seviyede teknoloji, makinelerin ve insanların simbiyotik olarak etkileşime girmesine izin verir. Kısacası, makineler ağın içeriğini okumak ve analiz etmek için çok ilerlemiştir ve sonuç olarak ne yürüteceğine karar verir. Örneğin, "seyahatin dijital arkadaşı" çok yakında akıllı cihazlara entegre edilecek ve turistlerin seyahatle ilgili tüm konuları ele almalarına yardımcı olacaktır (Soava, 2015, s. 112). Buna KLM "Mavi Bot" - kendi kendine öğrenen bir sistem - ek bir örnek olarak verilebilir. Kullanıcıların Messenger veya Google Asistan'da bir bilet rezervasyonu yapmasına yardımcı olabilmektedir (Tavakoli ve Mura, 2018).

Web'in Web 5.0 olarak bilinen en son sürümü, makineler ve insanlar arasında karmaşık ve etkileyici etkileşimler sağlamaktadır. Web 5.0 bilgisayarların insanlarla etkileşime girmesini sağlayan duyuşal ve duygusal bir web olarak ifade edilmektedir. Örneğin, yüz ifadeleri özel kulaklıklar kullanılarak gerçek zamanlı olarak kişiselleştirilebilir ve avatlara eklenebilir. Ayrıca, insanlarla sanal insanlar arasındaki gerçek zamanlı etkileşimler, yapay otomatik insan benzeri davranıştan dolayı mümkün olmaktadır (Tavakoli ve Mura, 2018). Boston'daki Bilim Müzesi'ndeki sanal insan müzesi rehberi, ziyaretçilerin sanal makinelerle nasıl etkileşimde bulunabileceğinin somut bir örneğidir.

Özellikle Web 2.0'ın ortaya çıkmasından sonra web sitelerini ziyaret etme ve kullanma amacı "salt okunur"dan "okuma-yazma" olarak değişmiştir. Bu evrim, diğerleriyle etkileşime giren ve sosyal ağlar, çevrimiçi topluluklar, bloglar, wiki'ler ve diğer işbirlikçi medya aracılığıyla bilgi paylaşımı yapan hevesli kullanıcıları yaratmıştır. Gerçekten de, web büyük bir

iletifim kanalı haline gelmiŒtir (Kim vd., 2017: 362). Bu deęiŒimden en ok etkilenen sekt6rlerden birisi de turizm olmuŒtur. Web, turistlerin davranif kalıplarında b6y6k bir deęiŒiklięi teŒvik etmektedir. Turistler sadece otel ve uak bileti rezervasyonu yapmakla kalmayıp, aynı zamanda evrimii inceleme siteleri ve kiŒisel seyahat blogları aracılıęıyla (Kim vd., 2017: 362) seyahat deneyimlerini, Œikayet ve memnuniyetlerini dięer insanlarla paylaŒmaya baŒlamıŒlar (Hu ve Chen, 2016: 929).

G6n6m6zde birok potansiyel turist seyahate ıkmadan 6nce, gideceęi yerle ilgili bilgi almanın yanında aynı destinasyona ya da iŒletmeye daha 6nce gidenlerin deneyimlerini de takip etmektedir (Nave vd., 2018: 695). Seyahat anında ise buldukları ortamlara ve deneyimlere ait bilgi ve fotoęrafları paylaŒmaktadırlar. 6zellikle mobil cihazların yaygınlaŒması ile birlikte turizme y6nelik birok kullanıcılara ait yorum, deęerlendirme, fotoęraf ve videoların yayınlandığı web sitesi ortaya ıkmıŒtır (Phillips vd., 2015: 131). Booking.com, Trivago, TripAdvisor ve Hotels.com gibi birok seyahat platformu web sitesi B2C (İŒletmeden T6keticiciye) pazarlama ve sosyal aę fonksiyonları saęlamaktadır. Restoranlar, seyahat acenteleri ve dięer ilgili iŒletmelerin m6Œterileri, paylaŒım bilgileri, rezervasyon odaları, seyahat rehberleri, fiyatların karŒılaŒtırılması gibi hizmetler hakkında yorum yapabilmektedirler. Bu web siteleri, t6keticideneyimlerinin ardından bir geri bildirim mekanizması saęlar ve kullanıcıların otel rezervasyonunu etkilemek iin yorumları y6nlendirmez.

Otellerin derecelendirilmesine adanmıŒ olan evrimii yorum web siteleri, k6resel GSYİH'nın % 9,4'6ne katkıda bulunan turizmin artan etkisinden dolayı b6y6k 6l6de pop6lerlik kazanmıŒtır (Geetha vd., 2017: 44). Bu sitelerden birisi olan TripAdvisor, seyahat alanında yaklaŒık 7,7 milyon iŒletmeyi kapsayan, 661 milyonu aŒkın yorum ve g6r6Œ6 turistlere sunmaktadır. TripAdvisor, turistlerin, tercihlerine en uygun otelde, en uygun fiyatı bulabilmeleri iin ayrıca 200'6 aŒkın otel rezervasyon sitesindeki fiyatları da karŒılaŒtırmaktadır. TripAdvisor markalı siteler, 49 pazarda sunulmakta olup aylık ortalama 456 milyon farklı ziyaretiden oluŒan, d6nyanın en b6y6k seyahat topluluęunu barındırmaktadır (TripAdvisor, 2017). Bu ve benzeri siteleri takip eden yolcular yapılan deęerlendirmeleri g6zden geirmekte ve ihtiya duydukları bilgileri aramaktadır (Puri vd., 2017: 237).

TripAdvisor gibi 6neri siteleri, bunu kullanan turistlerin% 60'ının yeni bir 6r6n veya hizmet satın almadan 6nce evrimii olarak incelemesi ve bu t6keticilerin% 80'inin bu t6rden fikirlerden etkilenmesi aısından ok 6nemli hale gelmiŒtir (O'Connor, 2010; Nave vd., 2018: 695). Seyahat ve konaklama hizmetlerinde t6keticidenoluŒturulan yorumların 6zellikle turistler iin kritik bir bilgi kaynaęı olduęu bulunmuŒtur (Duan vd., 2016: 284). Park vd. (2007), t6keticilerin daha d6r6st bilgi saęlayıcıları oldukları d6Œ6n6ld6ę6, evrimii t6keticiden

yorumlarının genellikle ürün ve hizmet sağlayıcıları tarafından sağlanan bilgilerden daha inandırıcı ve güvenilir olarak kabul edildiğini ileri sürmektedir. Seyahat yorumlarına bakan kullanıcılarının % 84'ü çevrimiçi yorumların satın alma kararlarını önemli ölçüde etkilediğini bildirmişlerdir (Nave vd., 2018: 695). Çevrim içi yorumlar, diğer turistlerin karar verme süreçlerini ve gelecekteki seyahat motivasyonlarını ve gelecekteki davranışlarını doğrudan etkileyebilir. Son zamanlarda yapılan araştırmalar, çevrimiçi kullanıcılar tarafından üretilen geri bildirimlerin genel olarak satışlarda (Nave vd., 2018: 695) ve özellikle turizmde önemli bir etkiye sahip olduğunu göstermiştir.

Turistler, seyahat dönüşü ya da deneyimlerinin ardından yaşadıkları olumsuz ve olumlu olayları ilgili platformlarda paylaşmakta ve bu paylaşımlar potansiyel ziyaretçilerin en değerli başvuru kaynağı olmaktadır. Bu platformlarda eski turistlerin tercihleri, memnuniyetleri, memnuniyetsizlikleri, ürün ve hizmet değerlendirmeleri, deneyimleri yer almaktadır. Bu siteler, tüketicilere değerli bilgiler sağlamak için ana kanallardır. Potansiyel turistler de bu yorum gözden geçirmekte ve ihtiyaç duydukları bilgileri aramaktadır (Puri vd., 2017: 237). Bakılan yorum ve değerlendirmeler, potansiyel turistlerin karar verme süreçlerini ve gelecekteki seyahat motivasyonlarını etkileyebilmektedir. Bunun sonucu olarak, turizm endüstrisinde, otel işletmeleri en çok sosyal ağlar, çevrimiçi seyahat toplulukları ve yorum siteleri üzerinde paylaşılan KOİ'den etkilenmektedir (Phillips vd., 2015), çünkü bu yorumlar rezervasyon olasılıklarını gerçek zamanlı olarak doğrudan etkilemektedir (Nave vd., 2018: 693). Çevrimiçi değerlendirmelerdeki artış, e-dağıtım yoluyla otel odası rezervasyonlarının büyümesine benzer bir eğilim ortaya koymaktadır (Phillips vd., 2015:131). Torres vd.'nin (2015) yaptığı bir çalışmada, otelin derecelendirmesinin ve yorum sayısının çevrimiçi işlemlerle yaratılan değer üzerindeki olumlu etkisi araştırılmaktadır. Araştırma, çevrimiçi yorum web sitesi puanlarının ve yorum sayısının, her çevrimiçi rezervasyon işleminin ortalama boyutuyla pozitif bir ilişki olduğunu göstermiştir (akt. Geetha vd., 2017: 45) .

Web 2.0 siteleri ve sosyal medya platformları, son on yılda tüketicilerin şirketlerle ve birbirleriyle çevrimiçi olarak kurdukları ilişki nedeniyle muazzam miktarda değerli bilgi üretmiştir. Çoğu bilgi, tüketicilerin çevrimiçi incelemeler yoluyla başkalarıyla paylaşmaya istekli olduğu deneyimlerin bireysel yorumlarını (Nave vd., 2018: 693), ürün ve hizmetler için yaptıkları değerlendirmeleri ve verdikleri puanları, şikayetlerini, memnuniyetlerini, fotoğraf ve videolarını yansıtmaktadır. Bazı web siteleri, derecelendirme sistemini veya diğer verileri yıldız / sayı veya metin olarak derecelendirirken, bazı web siteleri hem metin hem de derecelendirme sistemine hitap eden derecelendirme sistemiyle donatılmıştır. Sadece sayısal derecelendirme

sistemi kullanmak, değerlendirmeyi zorlaştıran birçok başka inceleme problemi olduğundan yeterli veri sağlamamaktadır (Puri vd., 2017: 237).

Seyahat ile ilgili çok fazla web sitesi olmasına ve bu sitelerde devasa verilerin birikmesine rağmen yöneticilerin ve araştırmacıların değerli bilgilere ulaşması kolay olmamaktadır, çünkü çok sayıda yorumu inceleyip, değerlendirmek gerçekten göz korkutucudur (Hu ve Chen, 2016: 929). Kullanıcılara ait içeriklerin hacmi elle ve kapsamlı bir şekilde analiz edilemeyecek kadar büyüktür ve yapılandırılmamış oldukları için istatistik tekniklerinin kullanılması da zordur. Tüketici tarafından üretilen çevrimiçi veriler büyük veri ve veri madenciliği teknikleri ile analiz edilebilir.

Başlangıçta web sitelerinin yapısı ve içeriğinin analizini içeren çalışmalar yapılmış ve önerilmiş ise de (Olmeda ve Sheldon, 2001), web 2.0'nin yaygınlaşmasından sonra web sitelerinde yayımlanan kullanıcı içeriklerinin analizine yönelilmiştir. Çeşitli çalışmalar gelişmiş metin madenciliği tekniğini kullanmış ve çevrimiçi yorum veri kümelerinden bilgi edinmek için büyük veri setini analiz etmiştir (Kim vd., 2017: 362). Verileri web sitelerinden elde eden ve veri madenciliği ve büyük veri teknikleri kullanarak analiz yapan araştırmaların bir kısmı Tablo 3.3'de listelenmiştir.

Tablo 3.3 Web Sitelerinden Veri Toplayan Araştırmalar

Araştırma	Web Sitesi	Toplanan Veri
Kim vd., 2017	virtualtourist	14 kategoride müşteri yorumları
Bogicevic vd., 2017	AirlineQuality	2010-2013 yılları arasında yayımlanan 901 turist yorumu
Xie vd., 2017	TripAdvisor	Teksas eyaletinin 427 şehrindeki 2.652 otel için yapılan çevrimiçi yorumlara yönetim yanıtı bilgisi
Guo vd., 2017	TripAdvisor	16 ülkede bulunan 25.670 otel için 266,544 çevrimiçi yorum
Berezina vd., 2016	TripAdvisor	ABD'de Florida, Sarasota'daki oteller hakkında kullanıcıların yaptığı 2,510 yorum
Liu vd., 2013	TripAdvisor	4 yıldız ve üstü puan alan 93 Melbourne otelinin değerlendirmelerini içeren (otel faktörleri hakkında yorumlar ve puanlar), toplam 6196 kayıttan oluşan bir veri kümesi
Hargreaves, 2015	TripAdvisor	Otel değerlendirmelerini içeren (kullanıcının otel deneyimi yorumları ve konum, uyku kalitesi, odalar, hizmet, değer, temizlik gibi nitelikler için genel derecelendirmeleri (1 - 5) ve değerlendiricinin temel demografik bilgileri) toplam 14 175 çevrimiçi kayıt
Shi vd., 2017	Baidu ya da TripAdvisor	Şangay turizm yıllığı (2015) içinde listelenen 675 turizm destinasyonu ve diğer eğlence alanları (örneğin, tiyatrolar, ticari caddeler ve meydanlar) hakkında 446.273 coğrafi etiketli veri.
Liu vd., 2017	TripAdvisor	Çin'deki popüler seyahat destinasyonlarındaki 10.149 otel için kullanıcılar tarafından yapılan 412.784 yorum
Zhao vd., 2019	TripAdvisor	San Francisco'daki 217 otelden 155'i için bireysel düzeyde 127.629 yorum
Xu, 2018	Booking.com	6 farklı yıldız grubu için toplam 4.800 yorum
Gal-Tzur vd., 2018	TripAdvisor	Dört şehir (Tel Aviv, Hayfa, Stuttgart ve Atina) forumlarında yayımlanan sorular

3.8. Sosyal Medya

Son yıllarda, world wide web'in modern toplumun hemen hemen tüm alanları üzerindeki etkisi muazzam bir şekilde artmıştır (Neidhardt vd.,2017: 102). Web 2.0 teknolojisinin ortaya çıkışı, kişisel bilgi, fikir ve deneyimin çevrimiçi olmasının yeni bir yolunu açmıştır (Hu ve Chen, 2016: 929). Web 2.0 teknolojileri, Kullanıcı Tarafından Oluşturulan İçerik (KOİ) sitelerini (ücretli içerikler yerine genel kamu tarafından yaratılan veya üretilen medya içeriği) geliştirmiş ve insanların arama yapma, bulma, okuma, toplama, geliştirme ve bilgi kullanma şeklini değiştirmiştir (Nave vd., 2018: 695). Daha önce web üzerinde sadece web yöneticilerinin oluşturduğu içerik ve onlara ait veriler mevcut iken Web 2.0'dan sonra kullanıcılara ait içerikler yayınlanmaya ve paylaşılmaya başlamıştır. Akıllı telefonların da yaygın bir şekilde kullanılmasıyla birlikte bir dizi çevrimiçi topluluk ve önemi artmıştır. Bugün, çevrimiçi topluluklar, insanların hem özel hayatlarında hem de iş ortamlarında iletişim kurmaları ve etkileşimde bulunmaları için hizmet etmektedirler (Neidhardt vd., 2017:102).

Sosyologlar ve piyasa araştırmacıları, bir bireyin hayatının, bireyin sosyal bağlantılar ağı ile nasıl etkileşime girdiğine bağlı olduğuna inanmaktadır. “Sosyal ağ” terimi, çevrimiçi bir topluluğu temsil eden Facebook, LinkedIn ve Twitter gibi sosyal ve profesyonel ağ sitelerini ifade etmek için kullanılmaktadır. Çevrimiçi bir topluluk, bir sosyal yapının örneğidir. Bu çevrimiçi topluluklara üye olan kişiler (aynı zamanda düğümler olarak da bilinir), aralarındaki ilişkiyle birlikte bir sosyal ağı temsil etmektedirler (Bell ve Mgbemena, 2018:197) Sosyal ağ üzerinde kullanıcılar deneyimlerini yayınlamakta ve paylaşmaktadırlar. Kullanıcılar tarafından oluşturulan bilginin basit, anlık ve çift taraflı olarak paylaşılmasını ve ulaşılmasını sağlayan ortamlara da “Sosyal medya” denilmektedir.

Sosyal medya, metinler, fotoğraflar ve video mesajları gibi kullanıcı tarafından oluşturulan içerik içeren Web 2.0 uygulamalarıdır (Vu vd., 2017: 4). Sosyal medya, Twitter (www.twitter.com), Instagram (www.instagram.com) ve Facebook (www.facebook.com) gibi internet platformlarında kullanıcıların oluşturduğu içeriklerin iletişimini, oluşturulmasını ve paylaşılmasını sağlayan etkileşimli platformlardan oluşmaktadır (Thomaz vd., 2017: 787). Sosyal medya kullanıcılara arkadaşlarının, ailelerinin ve takipçilerinin iletişim kurup ulaşabilecekleri platformlar sağlamakta (Brandt vd., 2017: 704) ve etkileşim, sesini duyma becerisini sağlayarak tüketicinin güçlendirilmesine yol açmaktadır (Duan vd., 2016: 284).

Sosyal medya teknolojileri, bloglar, forumlar, fotoğraf paylaşımı, ürün / hizmet yorumları, sosyal oyun, video paylaşımı ve sanal dünyalar gibi farklı biçimlere sahiptir (Vu vd., 2017: 4). Sosyal medya üzerinde günde milyonlarca metinsel yorum, fotoğraf ve video kullanıcılar tarafından paylaşılmaktadır. Yorum, fotoğraf ve videolar içerinde kullanıcıların duygularını,

tercihlerini, tavsiyelerini, konumlarını da içermektedir. Kullanıcılar arasındaki bu iletişim ve etkileşim internette kamuoyuna duyurulduğunda, bu günün satın alma kararlarının önemli bir parçası (Hu ve Chen, 2016: 931) olan eWOM etkisi yapmaktadır. eWOM yani "electronic WOM" insanların tüm çevrimiçi mecralarda (sosyal ağlar, bloglar, forumlar vd.) bir ürün ya da hizmet hakkındaki iletişim ve etkileşimlerine verilen genel isimdir. WOM, mal ve hizmetlerin değerlendirilmesinde özel taraflar arasında gayri resmi bir iletişimdir. Kusurlu ürünlere sahip müşterilerin yüzde yetmiş beşi bir ya da daha fazla müşteriye olumsuz bir WOM yaymaktadır (Bell ve Mgbemena, 2018:198). Günümüzde hizmet sağlayıcılar ve tüketiciler, tüketiciler ve tüketiciler arasında en etkili iletişim kanalı olarak tanınmıştır (Hu ve Chen, 2016: 929). Hatta sadece müşterinin arkadaşları üzerinde değil, aynı zamanda arkadaşlarının arkadaşları üzerinde de bir etkiye sahiptir (Bell ve Mgbemena, 2018: 198).

Son birkaç yıldır, mobil cihazların yaygın kullanımı, sosyal web hizmetlerinin büyümesini büyük ölçüde artırmıştır. Sosyal medya, mobil teknolojilerin kullanımı veya web tabanlı cihazlar aracılığıyla müşterilerle güçlü ilişkiler kurmak için en güçlü iletişim araçlarından biri olarak kabul edilmektedir. Chatterjee ve Wang (2012) turistlerin% 47'sinin zaten internette seyahat hedefleri ve otelleri aradığını ve seçtiğini, % 40'ının tatil hedeflerini önceden keşfetmek ve öğrenmek için interneti kullandığını, % 33'ünün seyahat edeceği havayolunu seçemeye karar verdiğini, %32'sinin hedef ülkede kültür, olay ve miras hakkında bilgi edinmek için kullandığını göstermiştir (akt: Nave vd., 2018: 695).

Turizm, teknolojik ilerlemeleri benimsemiş ve turistlerin çok farklı deneyimler arasından daha iyi bir seçim yapabilmeleri için sosyal medyada güçlü bir konuma sahip olan ilk sektörlerden birisidir (Nave vd., 2018: 694). Turizm, bilgisayar rezervasyon sistemlerinin gelişimi, müşterilerle iletişim, etkileşim, araştırma araçları, toplu veri depolama ve stratejik karar alma desteği ile başlayarak, otuz yılı aşkın süredir teknolojiyi benimsemiştir. İnternet aracılığıyla ve daha spesifik olarak sosyal medya aracılığıyla, bireyler bugün diğer turistlerin fikirlerini ve görüşlerini kolayca okuyabilir ve yorumlayabilir (Nave vd., 2018: 695). Dolayısıyla son yıllarda bilgi aramak ve çevrimiçi rezervasyon yapmak için interneti kullanan turist sayısı önemli ölçüde artmıştır.

Sosyal medya, turizm deneyimleri ve hayallerindeki anlamın üretimi ve dolaşımı için yeni kanallar sağlamaktadır (Shi vd., 2017). Restoranlar, seyahat acenteleri ve diğer ilgili işletmelerin müşterileri, paylaşım bilgileri, rezervasyon odaları, seyahat rehberleri, fiyatların karşılaştırılması gibi hizmetler hakkında yorum yapabilirler ve deneyimlerini paylaşabilirler. Seyahat özellikleri (ör. Fiyatlar, hava koşulları ve diğer ilgi çekici yerler) hakkındaki görüşlerin yanı sıra, deneyimlerin paylaşılması ayrıca, çevrimiçi iletişimdeki fotoğraflar, ifadeler ve dilsel

belirteçler aracılığıyla duyguları, hayal gücü ve fantezileri iletmeyi de içermektedir (Shi vd., 2017).

Sosyal aktivite verilerinin insanların etkinlik çalışması konusundaki temsili olduğuna dair bazı kaygılar olsa da ya da kullanıcılar kendileri hakkında tweet attığı zamanlarda ilgili faaliyetlerde bulunmamış olma ihtimalleri olsa da (Brandt vd., 2017: 704), önceki araştırmalar (Tilly vd., 2015; Bendler vd., 2014) sosyal medya verilerinin turizm araştırmaları için güvenilir bir veri kaynağı olduğunu göstermiştir (Shao vd., 2017: 77). Tilly vd. (2015), çevrimiçi seyahat incelemelerinden elde edilen, resmi kaynaklar tarafından sağlanan bilgilerle daha önceki mevcudiyeti ve yüksek korelasyonu vurgulayan turizme ilişkin makro-seviye bilgileri analiz etmiştir. Bendler vd. (2014) sosyal medya etkinliği (Twitter mesajları) ve çevrelerindeki çevresel özellikler (barlar veya müzeler gibi) arasındaki güçlü mekânsal ve zamansal ilişkileri göstermiştir.

Turizm endüstrisinin temel endişelerinden biri, bir turist için unutulmaz bir deneyim yaratan ve bu beklentiler sırasında ilk beklentilerin gerçekleşip gerçekleşmediğinin anlaşılmasıdır. (Nave vd., 2018: 695) Bunu anlamanın yollarından birisi de sosyal medyayı takip etmektir. Mevcut büyük hacimli verilerle, sosyal medya, turistik faaliyetler, algılar ve kaygılar hakkında değerli ve önceden edinilmesi zor olan bilgiler sağlayabilir (Vu vd., 2017: 4). Sosyal medyadaki bu yüksek miktarda veri araştırma için değerli bir kaynaktır, çünkü insanların davranışlarını ve etkileşimlerini incelemeye olanak sağlar. Dahası, büyük miktarda veri turizm şirketlerinin önemli bir varlığı haline gelmiştir. Verilerin doğru bir şekilde ele alınmasının müşteri ilişkileri yönetiminin iyileştirilmesinden (hem yeni yolcuları cezbetmek hem de mevcut olanları korumak açısından), işletmedeki mevcut sorunların tespit edilmesine ve iyileştirilecek noktaların tespitine kadar avantajları vardır (Neidhardt vd., 2017: 102)

Sosyal medya verilerinin avantajlarının yanında verilerin nasıl yönetileceği ve kalitesinin nasıl sağlanacağı, müşterilerin mahremiyetinin nasıl muhafaza edileceği ve bundan nasıl değerli bilgiler alınacağı gibi yeni zorluklar ortaya çıkmaktadır (Neidhardt vd., 2017: 102). Sosyal medyada paylaşılan bilgiler kamuya açıktır ve bu nedenle farklı ülkelerdeki gizlilik mevzuatından bağımsız olarak erişilebilmekte ve analizi yapılabilmektedir (Thomaz vd., 2017: 787). Burada yapılandırılmamış ve büyük miktarda veri kümeleri ile karşılaşmaktadır ve bu verinin klasik istatistiksel yöntemler ile analiz edilmesi mümkün olmamaktadır. Bununla birlikte, büyük miktarda veriyi işleyebilen ve analiz edebilen yeni hesaplamalı, matematiksel ve istatistiksel yöntemlerin geliştirilmesiyle, artık metinsel ve ilişkisel verileri analiz etmek için çok sayıda teknik bulunmaktadır (Neidhardt vd., 2017: 102)..

Bu tekniklerden veri madenciliği, büyük miktarlardaki verinin içinden geleceğin doğru tahmin edilmesinde yardımcı olacak anlamlı ve yararlı bağlantı ve kuralların bilgisayar programlarının aracılığıyla aranması ve analizidir. Ayrıca, çok büyük miktardaki verilerin içindeki ilişkileri inceleyerek aralarındaki bağlantıyı bulmaya yardımcı olan veri analizi tekniğidir. Verileri sosyal medyadan elde eden ve veri madenciliği ve büyük veri teknikleri kullanarak analiz yapan araştırmaların bir kısmı Tablo 3.4’de listelenmiştir.

Tablo 3.4 Sosyal Medyadan Veri Toplayan Bazı Araştırmalar

Araştırma	Sosyal Medya	Toplanan Veri
Deng ve Li, 2018	Flickr	Kullanıcılar tarafından New York City’de çekilip paylaşılan 20,974 fotoğraf
Nguyen vd., 2017	Örneğin Flickr ve Instagram	Sosyal Medyadan coğrafi etiketli fotoğraflar (enlem ve boylam) ve etiketleri
Hausmann vd., 2018	Instagram ve Flickr	Kruger National Park içinde 563 turist ile yapılan anketle elde verilere ek olarak Instagram (9,059 adet) ve Flickr (4,598)’da paylaşılan fotoğraflar
Zhang vd., 2018	Flickr	Çin'deki beş ünlü şehire (Örneğin, Pekin veri kümesi 81.867 fotoğraf) ait coğrafi etiketli fotoğraflar
Vu vd., 2015	Flickr	Hong Kong'a gelen batılı ve asyalı yolcuların çektiği 29,443 adet fotoğraf
Dietz vd., 2018	Foursquare	77 ülkedeki 415 şehirden 3.680.126 mekânda 266.909 kullanıcının 18 aylık check-in verisi
Hu vd., 2018	Twitter	New York City büyükşehir bölgesi için 5.019.637 adet coğrafi etiketli tweet (tweetler kullanıcı kimliğini, enlem, boylam, tarih, metin ve diğer bilgileri içermektedir)
Mor ve Dalyot, 2018	Flickr	Manhattan, New York’da yaklaşık 210 kilometrekarelik bir alanda 22.665 kullanıcı tarafından çekilen toplam 358.691 adet coğrafi etiketli fotoğraf
Vu vd., 2018	Flickr	Avustralya giden turizm ile ilgili 3.623 yolcu tarafından çekilen 890.000'den fazla fotoğraftan oluşan bir veri seti

3.9. Cep Telefonları

Başlangıçta sadece haberleşme amaçlı kullanılan cep telefonları, akıllı telefonların yaygınlaşmaya başlaması ile birlikte bilgisayarlar gibi kullanılmaya başlanmıştır. Hatta taşıma kolaylığı, şarjlarının daha uzun süre dayanması gibi avantajları ile birlikte bilgisayarlardan daha fazla kullanılır duruma gelmişlerdir.

GPS uygulaması ve cep telefonları üzerinde çalışan çeşitli uygulamalar ile cep telefonları turizm amaçlı kullanılan cihazlardan birisi haline gelmiştir. GPS ile cep telefonun yerinin tespit edilmesi turist hareketliliğinin takip edilmesi için büyük fırsatlar vermektedir. Ayrıca konum belirlemeye yarayan bazı uygulamalar turistlerin hangi mekânlarda bulunduğunu, ne kadar vakit geçirdiğini takip etmede kullanılabilirlerdir.

Bu özelliklerinin dışında cep telefonları, fotoğraf ve video çekme özellikleri turist deneyiminin belgelenmesi açısından bir ıgır açmış ve çekilen bir fotoğraf ya da videonun paylaşılmasının önünü açmıştır. Bu paylaşımlar deneyimlerin, müşteri davranışının, müşteri tercihlerinin (Li vd., 2016: 31) vb. takip edilmesi konularında arařtırmacılara veriler sunmaktadır.

Cep telefonu lokasyonlarının büyük veri madenciliđi ile kaydedilen gerçek zamanlı turist hareket davranışı, morfolojik özelliklerin ve turist tercihlerinin niceliksel olarak anlaşılmasının doğrulanmasında başka bir bakış açısı sağlamak için kullanılmaktadır. Turist tercihlerini cep telefonu veya GPS teknolojisiyle izlemek son yıllarda ortaya çıkan yeni bir tekniktir. İnsanların gerçek zamanlı hareketlerini kaydedebilen bir teknik olarak, turist yönetimi için çok yeni bir potansiyel getirmektedir (Li vd., 2016: 31). Bu kaynaklardan veri elde edebilmek için cep telefonu sahibi ile irtibat halinde olmak gerekmektedir, bu da turistlerin kolay kolay yaklaşmıyacağı bir durum olabilmektedir. Alternatif bir yaklaşım, cep telefonlarının telefon sahibinin doğrudan katılımı olmadan bir hücre kulesi ađı üzerinden hareketini izlemektir. Özellikle Estonya'da, bu yöntem zaten turistlerin bölgesel hareket modellerini incelemek için yaygın olarak kullanılmaktadır (Versichele vd., 2014: 68).

3.10. Kullanıcı Tarafından Oluřturulan İçerik

Web teknolojilerinin ilk sürümü Web 1.0 kullanıcıların web sayfası içindeki bilgileri yayınlamasına veya paylaşmasına izin vermemiřti. Daha sonra sosyal web olarak ortaya çıkan Web2.0 ise toplulukları ortak ilgi alanlarına bağlayabilmekte ve iki yönlü etkileşimi kolaylařtırmaktadır (Tavakoli ve Mura, 2018). Web 2.0 ve daha sonraki sürümleri kullanıcıların web sitelerinde yorum yapmalarına, yorumlarını paylaşmalarına, kısacası ortamın bir medya olarak kullanılmasına izin vermektedirler. Web 2.0 teknolojisinin ortaya çıkışı, mobil cihazların yaygın kullanımı, sosyal web hizmetlerinin büyümesini büyük ölçüde artırmıştır (Brandt vd., 2017:703). Kullanıcılarının çevrimiçi etkinlikleri, herkesin rahatça ulaşabileceđi kullanıcı tarafından oluřturulan içeriklerden (KOİ) oluřan muazzam bir veri kaynađını oluřturmuřtur.

KOİ, tüketiciler tarafından sađlanan, her zaman ve her yerde kolayca ulaşılabilen, ücretsiz veya düşük maliyetli, spontane, anlayışlı ve tutkulu bir geri bildirim olarak kabul edilebilir (Guo vd., 2017: 468). Bu çevrimiçi geri bildirimler, hem yorumlar hem de derecelendirmeler biçiminde, gelecekteki müşteri talebini ve otellerin finansal performansını etkileyen ve dolayısıyla önemli bir iş deđerine sahip olan bir elektronik-word-of-mouth (eWOM) etkisi yaratmaktadır (Zhao vd., 2019: 111). WOM, tüketicilerin tanıdıkları arasındaki memnuniyet veya memnuniyetsizlik düzeyini bildiren sözlü bir bildiri olarak açıklanabilir, hizmet veya ürün tüketiminden sonra gerçekleşen mantıksal bir satın alma davranışı olarak tanınmaktadır (akt:

Bogicevic vd., 2017: 514). İnternet medyasının ve iletişimin çağdaş dünyasında, WOM, elektronik WOM (eWOM) olarak bilinen bir çevrimiçi öneri biçimi olmuştur (Bogicevic vd., 2017: 514).

eWOM, tüketiciler gibi ticari olmayan bir iletişimcinin, ürün ve hizmet özelliklerine veya kullanımına ilişkin yorumlarını içeren, genellikle çevrimiçi eleştiri, çevrimiçi öneriler ve çevrimiçi görüşler biçiminde olan, internet tabanlı informal iletişim olarak tanımlanabilir (Litvin vd., 2008; Xu, 2017:1666). Litvin vd.'e (2008: 461) göre, elektronik ağızdan ağıza iletişim (eWOM), “belirli mal ve hizmetlerin kullanımı ya da özelliklerine veya onların satıcılarına bağlı olarak internet tabanlı teknoloji ile tüketicilere yönelik tüm informal iletişim” olarak tanımlanmaktadır (Hu ve Chen, 2016: 929). İnternet, Sözlü WOM'un aksine eWOM, sosyal yakınlığın ve coğrafi yakınlığın sınırlarını aşarak mesajların sadece arkadaşlara ve ailelere değil, ilgilenen herhangi bir tüketiciye ulaştırılabildiği sanal bir ortam sağlamaktadır. Geleneksel ağızdan ağıza iletişim ile karşılaştırıldığında, eWOM daha hızlı bir yayılma hızına ve daha geniş bir yayılma aralığına sahiptir ve müşterilerin gerçek zamanlı olarak otel algısını göstermektedir (Xu, 2017: 1666). İnternet pazarlamasının gücü ile birlikte, eWOM, tüketicilerin karar alma sürecini önemli ölçüde etkileyen, kullanıcı tarafından oluşturulan bir içerik kaynağı olarak algılanabilir (Bogicevic vd., 2017: 514). Günümüzde, farklı elektronik iletişim türleri, hizmet içi iletişim, görüş ve öneriler gibi, hizmet sağlayıcılar ve tüketiciler arasında tüketiciler arasında en etkili iletişim kanalı olarak tanınmıştır (Cantalops ve Salvi, 2014; Hu ve Chen, 2016: 929).

Son birkaç yıldır, mobil cihazların yaygın kullanımı, sosyal web hizmetlerinin büyümesine ve kullanıcılarının çevrimiçi etkinliklerinin muazzam veri kaynağı olmasına neden olmuştur. Mart 2013'te, Flickr kullanıcıları günde 3,5 milyondan fazla yeni görüntü yüklemişlerdir; Aralık 2014'te Facebook'un günlük ortalama 890 milyon aktif kullanıcısı vardı; bunların 745 milyonu mobil cihazlardan hizmete erişiyordu (Brandt vd., 2017: 703). Ortalama bir günde, 500 milyondan fazla Twitter mesajı gönderilmekte ve 80 milyon fotoğraf, Instagram'da paylaşılmakta ve 3,5 milyar beğeni toplamaktadır. Bu sitelerden birisi olan TripAdvisor, seyahat alanında yaklaşık 7,7 milyon işletmeyi kapsayan, 456 milyon farklı ziyaretçiye ait, 661 milyonu aşkın yorum ve görüşü turistlere sunmaktadır (TripAdvisor, 2017). Sonuç olarak, sosyal medya analitiği araştırmacılar ve şirket yöneticileri için giderek daha ilgili bir konu haline gelmiştir (Brandt vd., 2017:704).

Web sitesi ve sosyal medyada kullanıcı tarafından oluşturulan içeriklerin en yaygın olarak kullanılanı metin temelli çevrimiçi yorumlardır. Turistler deneyimlerini, memnuniyet ve memnuniyetsizliklerini, görüşlerini, bazen önerilerini web sitelerinde ve sosyal medyada

paylaşmaktadır. Özellikle akıllı cep telefonlarının kullanımının yaygınlaşması ile birlikte fotoğraf paylaşımı yoğunlaşmış ve sadece fotoğraf paylaşımı yapılabilen platformlar ortaya çıkmıştır. Buna konum bazlı sosyal medya platformlarının da etlenmesiyle coğrafi etiketli ögeler turistlerin mekânsal- zamansal hareketlerinin takibi için kaynak sağlamaya başlamıştır. Aşağıda kullanıcı tarafından oluşturulan içerikler hakkında detaylı bilgiler verilmiştir.

3.10.1. Kullanıcı Yorumları

Turizm sektöründe özellikle kullanıcıların turistik ürünler hakkında yorum yapabileceği, ürünü ve hizmeti değerlendirebileceği siteler mevcuttur. Bu siteler ve sosyal medya üzerinde turistler artık deneyimlerini paylaşabilir ve oteller, restoranlar ve turistik yerler için (örneğin, müşteri hizmetleri, otopark ve temizlik hakkında yorumlar yoluyla) başkalarına özel önerilerde bulunabilmektedirler (Guo vd., 2017: 468).

Müşteriler çevrimiçi yorumları, otellerle olan memnuniyetini ve memnuniyetsizliğini yansıtmak (Berezina vd., 2016) için yazmakta ve paylaşmaktadırlar. Bununla birlikte, müşterilerin psikolojik ihtiyaçları (örneğin, sosyal kimliğin kazanılması), bir topluluğa ait olma hissi ve şirketlere ve diğer müşterilere yardımcı olmak gibi diğer motivasyon kaynakları çevrimiçi yorum yazma davranışlarını teşvik etmektedir (Cantalops ve Salvi, 2014; Xu, 2017: 1668). Müşteriler genellikle çevrimiçi yorum yazmak için dört farklı güdü ile motive olurlar (Zhao vd., 2019:112). Birincisi fedakârlık ve karşılıklıdır. Bu güdü ile çevrimiçi yorumları yayımlayan müşteriler, gelecekteki otel ziyaretçilerinin otel hakkında daha iyi kararlar vermelerine ve otellerin hizmet işlemlerini iyileştirmelerine yardımcı olmayı hedeflemektedir. İkincisi, müşterilerin psikososyal ihtiyaçlarını karşılamaktadır. Bu nedenle çevrimiçi yorum gönderen müşteriler memnuniyetlerini ve hayranlıklarını ya da bir otele karşı memnuniyetsizliklerini ve şikâyetlerini gösterebilmektedirler (Cantalops ve Salvi, 2014). Üçüncüsü müşterinin sosyal ihtiyaçlarıdır. Çevrimiçi toplumda “yardımcı” olarak seçilmek, seyahat topluluğunda sosyal kimlik kazanmak gibi olumlu bir üne sahip olmak isterler ya da otel yöneticilerinin çevrimiçi yanıtlarını beklerler. Dördüncü teşvik ekonomiktir, yorumları yayınladıklarında çevrimiçi yorum platformundan ödüller kazanmaktadırlar. Müşterilerin dil tarzı, çevrimiçi yorumu yazdıran güdülerden etkilenmektedir (Zhao vd., 2019: 112). Bunlara ek olarak, çevrimiçi metinsel yorumların yapısı açıktır ve içeriği, yorumcuların eğitim geçmişleri, duyguları, tedarikçi-müşteri ilişkileri, hoşgörü ve keyif boyutu, yanıt beklentisi, zaman maliyetlerini gözden geçirme ve internet alışkanlığı gibi birçok faktöre bağlıdır (Schuckert vd., 2015; Xu, 2017: 1668).

Çevrimiçi yorumlar, müşteri memnuniyeti ve memnuniyetsizlik kaynaklarını incelemek için fırsatlar sunmaktadır. Çevrimiçi seyahat yorumları, tüketicilerin gerçek duygularını ve turistlerin ürün veya hizmetlerden memnuniyetini ya da memnuniyetsizliğini yansıtmaktadır (Berezina vd., 2016). Olumlu yorum yapan müşteriler sunulan ürün ve hizmetleri övmek, gelecekteki müşterilere tavsiyelerde bulunmak ve duygusal açıklamanın olumlu etkilerini almak istemektedir. Bu nedenle, müşteriler tarafından bahsedilen olumlu içerik, müşterilerin ürün ve hizmet sağlayıcılarla ilgili genel memnuniyetini tam olarak göstermeyebilir (Xu, 2017: 1668). Aynı şekilde olumsuz yorumlar müşterilerin tamamen memnuniyetsizliklerini göstermez. Müşterilerin olumsuz yorumları genellikle ürünlerin ve hizmetlerin başarısızlığının belirli bir yönünün bir sonucudur (Xu ve Li, 2016).

Jiang vd. (2010) tüm olumsuz deneyimlerin memnuniyetsizliğe yol açmadığını bulmuştur. Müşterilerin bahsettiği deneyimlerden memnuniyetin sadece küçük bir bölümünü bulmuşlardır. Olumsuz yorumları, müşterilerin utançlarını ve hayal kırıklıklarını hafifletmek için olumsuz duygular açıkladığını göstermiştir. Birçok müşteri, gelecekteki müşterileri uyarmak için ürün ve hizmetlerin olumsuz tarafını tanımlamış veya kaliteyi iyileştirmek için ürün ve servis sağlayıcısına başvurmuştur (Cantalops ve Salvi, 2014). Müşterilerin olumsuz yorumları, bu nedenle, beklenti ve algılanan kalitenin karşılaştırılmasıyla belirlenen müşteri genel memnuniyetini tam olarak yansıtmamaktadır (Xu, 2017: 1668).

Tüketicilerin neredeyse % 79'u, kullanıcı tarafından oluşturulan içeriklere (ör. Bloglar, forumlar, çevrimiçi yorumlar) ve satın alma kararlarını kesinleştirirken kişisel önerilere eşit güven duymaktadır (Bogicevic vd., 2017: 514). Bu güven, gelecekteki müşterilerin satın alma niyetlerini, müşteri güvenini, müşteri talebini ve otellerin finansal performansını etkilemektedir (Zhao vd., 2019; Xie vd., 2014; Xu, 2017). Xie vd., (2014) tarafından yürütülen araştırmada tüketici yorumlarının çeşitliliği ve hacmi ile yönetim yanıtlarının sayısının otel performansı ile önemli ölçüde ilişkili olduğu bulunmuştur. Ek olarak, tüketici değerlendirmelerinin çeşitliliği ve hacmi genel değerlendirme ve otel performansı arasındaki ilişkiyi yönlendirmektedir. Ayrıca, bir otelin finansal performansı müşteri çevrimiçi derecelendirmelerinden etkilenebilmektedir (Geetha vd., 2017: 44).

Dolayısıyla turizm işletmeleri kendi bünyelerinin dışında gerçekleşen ama kendilerini ilgilendiren bu olgudan uzak durmamalıdır. İşletmelerde verilen hizmet ile ilgili olarak müşterilerden anket düzenleyerek ve sosyal medya üzerinden geri bildirim alınabilmektedir. Anketlerde bazı müşterilerin gerçek bilgi vermediği, bazılarının da bilgi vermek istemediği daha önceki araştırmalarla tespit edilmiştir (Berezina vd., 2016). Oteller, müşterilerin beklentilerini ve ihtiyaçlarını anlamak ve ilgili ürünleri ve hizmetleri geliştirmek için çevrimiçi müşteri

yorumlarını kullanmaktadır (Zhao vd., 2019: 112). Olumsuz müşteri yorumları ve bunlara otellerin cevapları konusunda potansiyel müşterilerin algıları ve değerlendirmeleri üzerine yapılan çalışma (Sparks, So ve Bradley (2015)) çevrimiçi bir cevabın verilmesinin potansiyel tüketicilerin, işletmenin güvenilirliği ve müşterileri ne derece umursadığı konusundaki algılarını arttırdığını bulmuştur (akt: Geetha vd., 2017: 44). Dolayısıyla, KOİ, özellikle konaklama endüstrisinde (ör. Oteller ve restoranlar) tüketicilerin taleplerini önemseyen işletmeler için yararlı bir bilgi kaynağı olarak hizmet edebilir (Guo vd., 2017: 468).

KOİ işletmeler üzerinde olduğu kadar potansiyel turistler üzerinde de ekilidir. İnternet teknolojisinin yaygın kullanımı ile elektronik ağızdan ağza [eWOM] iletişiminin ürün ve hizmetlerin çevrimiçi yorumlar yoluyla tüketicinin davranışları ve tercihleri üzerinde güçlü bir etkisi vardır (Rong vd., 2012: 731). Küresel turistlerin yaklaşık % 89'u ve küresel otelcilerin % 64'ü, otel rezervasyonlarının çevrimiçi otel yorumlarından etkilendiğine inanmaktadır. Ady, TrustYou ve Quadri-Felitti (2015) tarafından yürütülen çalışma, turistlerin yaklaşık % 95'inin rezervasyon kararlarını vermeden önce çevrimiçi otel yorumlarını okuduğunu ve yorumların turistlerin üçte birinden fazlasının otel seçiminde kararlarını etkileyen en kritik etkenlerinden biri olduğunu bulmuştur (akt: Hu ve Chen, 2016: 929). Turistik üründe yanlış seçim yapılması durumunda maddi ve psikolojik hasar büyük olduğundan, turist kullanmadan önce yeterli bilgiyi toplayarak en uygun kararı vermeye çalışmalıdır. Bu nedenle, turistler sadece servis sağlayıcısından alınan bilgileri değil, aktif kullanıcılardan gelen yorumları da dikkate almak durumundadır (Kim ve Yu, 2017: 4442).

Web üzerinde kontrolsüz bir şekilde yayılan kullanıcı yorumlarının güvenilirliğinin sorgulanmasına rağmen, turist ürünlerine daha yüksek düzeyde güven oluşturmaya katkıda bulunmuştur (Tavakoli ve Mura, 2018). Tilly vd. (2015) ve Bendler vd. (2014), sosyal medya platformlarında görüntülenen kullanıcı tarafından oluşturulan içeriğin güvenilirliğini destekleyen kanıt sağlamıştır (Brandt vd., 2017: 704). Schuckert vd. "Bu bilgi kaynakları [çevrimiçi otel değerlendirmeleri] daha objektif, muazzam ve örnek önyargısız olarak değerlendirilmektedir, çünkü yorumlar geleneksel anketlerden farklı olarak laboratuvar etkilerine sahip olmaksızın kendiliğinden yayınlanmaktadır (Liu vd., 2017: 556).

Son birkaç yıldır, mobil cihazların yaygın kullanımı, sosyal web hizmetlerinin büyümesini artırarak kullanıcılarının çevrimiçi etkinliklerini (Brandt vd., 2017: 703) muazzam bir hazineye dönüştürmüştür. Araştırmacılarda bu hazineye kayıtsız kalmamışlar ve turizm araştırmaları, kullanıcı tarafından oluşturulan incelemelerin, araştırma yapmak için bir veri kaynağı olarak kullanılmasının artmasına tanık olmuştur (Liu vd., 2017: 556). Özellikle müşterilerin otel değerlendirmeleri, müşterilerin memnuniyetini incelemek için mükemmel bir enstrüman

gibi görünmektedir ve kullanıcı tarafından oluşturulan yorumlar, otel memnuniyetinin sürücülerini anlamak için (Guo vd., 2017: 468) kullanılabilir ve zengin bir veri kaynağı sağlamaktadır (Liu vd., 2017: 556). Sonuç olarak, geçtiğimiz on yılda, çevrimiçi tüketici değerlendirmeleri olgusunu inceleyen çalışmalar giderek artmaktadır (Guo vd., 2017: 468) ama işletme yöneticilerinin eWOM etkilerini kendi iş planlama ve karar verme süreçlerine etkin bir şekilde dâhil etmeleri bir sorun olmaya devam etmektedir (Rong vd., 2012: 731).

3.10.2. Coğrafi Etiketli Öğeler

Son zamanlarda, çoklu ortam ve mobil teknolojilerdeki ilerlemeler, seyahat fotoğrafları gibi büyük miktarlarda kullanıcı tarafından oluşturulan verilerin oluşturulmasını ve paylaşılmasını sağlamıştır. Bu teknolojiler, GPS teknolojisi ve mobil fotoğraf yakalama cihazları, artık coğrafi bilgilerin bir fotoğraf etiketinde bulunmasına ve depolanmasına izin vermektedir. Bu nedenle, bilgisayar bilimi araştırmacıları, bu coğrafi etiketli fotoğrafları kullanmak için yeni teknikler geliştirmeye başlamışlardır (Vu vd., 2015: 224). Fotoğraflar, tweetler ve konum tabanlı sosyal ağlarda (KTSA) yayınlanan coğrafi bilgiler, turist hareket ve davranışı konusunda araştırmacılara çok değerli bir veri yığını sunmaktadır. Veri yığınlarındaki coğrafi etiketli öğeler, turist davranış verilerinin verimli ve zamanında yakalanması için yeni araştırma alanı sunmaktadır. Küresel ölçekte milyonlarca coğrafi etiketli ve zaman damgalı fotoğrafla birlikte, fotoğraf paylaşım siteleri, veri elde etmek ve seyahat davranışlarını araştırmak isteyen araştırmacılar için potansiyel altın madenleridir.

Akıllı telefonlar ve tabletler gibi birçok fotoğraf çekme cihazı, coğrafi bilgilerin (enlem ve boylam koordinatları) kullanıcının aldığı her fotoğrafla meta veri olarak depolanmasını sağlayan yerleşik küresel konumlandırma sistemlerine (GPS) sahiptir. Bu gelişmeyle birlikte sosyal medya, konum kavramını da içerisine dâhil etmiştir. Sosyal medya ve Konum Tabanlı Hizmetlerin kesişimi, kullanıcının konum bilgilerini veya coğrafi konum içeriklerini coğrafi veriler olarak yayınlamasına izin veren Konum Tabanlı Sosyal Ağlar veya Konum Paylaşımı Hizmetlerinin geliştirilmesine yol açmıştır (Shi vd., 2017: 1189). Sosyal medyaya ve web tabanlı etkinliklere katılan kullanıcıların oluşturduğu yetkilendirilmemiş kaynaklardan elde edilen coğrafi etiketli veriler, temel olarak coğrafi mekânsal olabilmekte ya da coğrafi olmayan bilgilere mekânsal özelliklerin eklenmesi şeklinde olabilmektedir.

Özellikle, bir destinasyon içindeki ziyaretçi davranışı ve seyahat akışı modelleri coğrafi konumla yakından ilişkilidir (Shi vd., 2017: 1189). Turistlerin coğrafi konumlarını bilmek, bir destinasyon içinde turist grupları için tercih edilen yerlerin, buraları ne zaman ziyaret ettiklerinin ve farklı yerleri ziyaret ederken tercihe edilen güzergahların tespit edilebilmesi açısından

önemlidir. Coğrafi etiketli öğeler, zaman ve coğrafi bilgilerin gömülü olmasıyla, kullanıcının mekânsal-anlık hareket yörüngesinin çıkarılmasına olanak tanımaktadır (Vu vd., 2015: 223).

Tablo 3.5 Coğrafi Etiketli Öğeler Üzerine Çalışan Araştırmalar

Araştırma	Veri
Deng ve Li, 2018	Kullanıcılar tarafından New York City’de çekilip paylaşılan 20,974 fotoğraf
Nguyen vd., 2017	Sosyal Medyadan coğrafi etiketli fotoğraflar (enlem ve boylam) ve etiketleri
Hausmann vd., 2018	Kruger National Park içinde 563 turist ile yapılan anketle elde verilere ek olarak Instagram (9,059 adet) ve Flickr (4,598)’da paylaşılan fotoğraflar
Shi vd., 2017	Şangay turizm yılı (2015) içinde listelenen 675 turizm destinasyonu ve diğer eğlence alanları (örneğin, tiyatrolar, ticari caddeler ve meydanlar) hakkında 446.273 coğrafi etiketli veri.
Zhang vd., 2018	Çin’deki beş ünlü şehire (Örneğin, Pekin veri kümesi 81.867 fotoğraf) ait coğrafi etiketli fotoğraflar
Cai vd., 2018	Avustralya’nın Queensland bölgesinde, 2014’te çekilmiş 63,559 adet coğrafi etiketli fotoğraf.
Hu vd., 2018	New York City büyükşehir bölgesi için 5.019.637 adet coğrafi etiketli tweet (tweetler kullanıcı kimliğini, enlem, boylam, tarih, metin ve diğer bilgileri içermektedir). Ayrıca, veri seti TripAdvisor’ın NYC’deki ilk 50 gözde mekan listesi ile desteklenmiştir.
Mor ve Dalyot, 2018	Manhattan, New York’da yaklaşık 210 kilometrekarelik bir alanda 22.665 kullanıcı tarafından çekilen toplam 358.691 adet coğrafi etiketli fotoğraf

3.10.3. Kullanıcı Değerlendirmeleri

Özellikle turizm işletmelerinin değerlendirildiği sitelerde müşterilerinin yorumları ile birlikte o işletmeye verilen puanlar da bulunmaktadır. Turistler konakladıkları işletmeleri sitenin belirlediği skalada puanlamaktadırlar. Bu puanların ortalamaları site üzerinde sürekli olarak yayınlanmakta ve potansiyel turistler için işletme hakkında bir ön fikir vermektedirler. Ayrıca müşterilerin verdiği otel puanları müşteri memnuniyetini incelemek için mükemmel bir araç (Hargreaves, 2015; Liu vd., 2017: 556) durumundadır.

Araştırmacılar turistlerin verdiği puanların gerçekten onların duyguları ile örtüşüp örtüşmediği konusunu araştırmışlardır. Otel hakkındaki yorumların teknik tarafı ile çevrimiçi müşteri derecelendirmeleri arasındaki ilişki konusunda Geetha ve ark. (2017), çevrimiçi müşteri incelemelerinin duygu polaritesine odaklanmış ve müşteri puanlarını etkilediğini tespit etmiştir. He vd. (2017) çevrimiçi otel metin incelemelerini analiz etmek için doğal dil ön işleme, metin madenciliği ve duyarlılık analiz tekniklerini kullanmışlar ve çevrimiçi müşteri değerlendirmelerinin başlığının ve içeriğinin duygu puanlarının otellerin genel müşteri puanlarıyla yüksek korelasyon gösterdiklerini bulmuştur. Elde ettikleri sonuçlar, metin incelemelerinden elde edilen özellik görüşlerinin çoğunun, müşterinin genel değerlendirmesiyle önemli ölçüde ilişkili olduğunu gösteren Qu vd. (2008) çalışmasını onaylamaktadır. Bu aynı zamanda, Kim vd. (2015) çalışmasından elde ettikleri bulguları da destekleyerek, genel puanların otel performansının en kritik yordayıcısı olduğunu göstermektedir (Zhao vd., 2019: 113).

3.10.4. Kullanıcı Fotoğrafları

Fotoğraflar insanın hayatındaki önemli anları ölümsüzleştirmektedir. Ortaya çıktığı günden bu yana teknolojik gelişmelerle paralel olarak fotoğraf da değişmiş, kullanım alanları çoğalmış, çoğaltma teknikleri farklılaşmıştır. Günümüzde ise cep telefonları kadar yaygınlaşmıştır.

Turizm sektörü de ortaya çıkmasıyla birlikte fotoğrafla tanışmış ve turistik ürünlerin tanıtımı ve seyahat anılarının ölümsüzleştirilmesi için kullanılmaya başlanmıştır. Dolayısıyla fotoğraflar, turizm sektöründe tanıtım ve seyahat fotoğrafları olmak üzere iki farklı şekilde karşımıza çıkmaktadır. Geçmişte, destinasyon yönetim organizasyonları destinasyon imajının tasarım ve sunumunda ön sıralarda yer almışlardı. Genellikle profesyonellerce veya profesyonel kalitede alınmış çok sayıda fotoğraf toplanmakta ve bunları reklamlarda ve tanıtım materyallerinde kullanılmaktaydı. Bu resimli görüntüler kitle ya da profesyonel medya aracılığıyla hedef pazara özel olarak gösterilmekteydi.

Web 2.0 ile birlikte birçok konuda olduğu gibi fotoğraf basımı ve paylaşımı da büyük değişime uğramıştır. Turizm sektöründe de dikkatler yavaş yavaş yöneticilerin reklam amaçlı çektiği kaliteli fotoğraflardan sosyal ağlarda paylaşılan seyahat fotoğraflarına kaymıştır (Deng ve Li, 2018:267). Pan vd. (2014) turistlerin çektiği fotoğrafların destinasyonun tanıtımında kullanılabileceği sonucuna varmışlardır. Akrabaların veya arkadaşların paylaştığı içeriğin, pazarlamacıların yayınladığı bilgilere kıyasla genellikle görülme ve güvenilir olma olasılığı daha yüksektir. Bu nedenle, KOİ (örneğin, genel halk tarafından alınan ve paylaşılan fotoğraflar), potansiyel destinasyon ziyaretçiler için TDI (turizm destinasyon imajı) oluşturmanın giderek daha etkili bir aracı haline gelmiştir (Deng ve Li, 2018:268).

“Bir resim bin kelimeye bedeldir”. Kelimelerle anlatılamayan özellikle duygular bir fotoğrafla hatta hiçbir kelime kullanmadan bile anlatılabilmektedir. Seyahat fotoğrafları, fotoğrafçıların içsel duygularını yansıtan semboller olabilmektedir. Ayrıca fotoğrafçıların seyahat deneyimlerini saklayan kayıtlar olarak da hizmet etmektedirler (Pan vd., 2014: 59). Bu kayıtlar turistlerin tercihini belirlerken, potansiyel turistlerin de tercihleri üzerinde etkili olmaktadır. Ayrıca dijital cihazlarla çekilen fotoğraflar genellikle tanımlayıcı bilgiler, yani meta veriler taşırlar. Bu bilgiler genellikle JPG fotoğraflarında gömülü değiştirilebilir görüntü dosyası olarak görünür. Fotoğraflar bir web sitesine yüklendiğinde meta veriler ayrıştırılabilir ve kaydedilebilir (Deng ve Li, 2018: 271). Bu özellik fotoğrafçının duygusal olarak incelenmesinin yanında mekânsal olarak da incelenmesini sağlamaktadır.

Tablo 3.6 Fotoğraflar Üzerine Çalışan Araştırmalar

Araştırma	Veri	Yapılan İşlem
Nguyen vd., 2017	Sosyal Medyadan coğrafi etiketli fotoğraflar (enlem ve boylam) ve etiketleri	Kültürel miras kaynaklarının yerlerinin belirlenmesi
Deng ve Li, 2018	Kullanıcılar tarafından New York City’de çekilip Flickr’da paylaşılan 20,974 fotoğraf	Fotoğraflardaki meta veriler üzerinde metin madenciliği ve yapılan yorumlar üzerinde duygu analizi, tanıtımda kullanılacak fotoğrafları seçmek için bayes sınıflandırıcı kullanılmıştır.

3.10.5. Kullanıcı Videoları

Turistlerin ziyarette buldukları destinasyon ve konakladıkları işletmelerde çektikleri videolar, onların duygularını, davranışlarını anlamada çok önemli bir veri kaynağı olarak sosyal medyada yer almaktadır. Ancak videolar üzerinde kapsamlı bir şekilde analiz işlemleri henüz başlangıç aşamasındadır. İlerleyen süreçte videolar turist davranışı ve algısı konusunda vaz geçilmez veri kaynaklarından birisi olmaya adaydır.

3.11. Veri Hazırlama

Bunca veriye rağmen turizm sektöründe veri madenciliğinin yeteri kadar yaygınlaşmamasının temelinde araştırmacıların verilere ulaşmasındaki zorluklar yatmaktadır. Çünkü veri olmadan veri madenciliği olamamaktadır. Veri toplanabildiği sürece, veri madenciliği yararlı olacaktır (Law, 1998: 66). Aslında merkezi rezervasyon ve tesis yönetim sistemlerinin çoğalmasının tetiklediği, turizm şirketlerinde büyük miktarda tüketici verisi birikmektedir. Sektörde işletme bünyesinde biriken devasa verilerden yararlanma konusunda araştırmacıların karşılaştığı iki büyük problem bulunmaktadır. Bunlardan birisi işletmelerin bünyesindeki sistemlerde biriken verilere ulaşım zorluğudur. İşletme izin vermediği sürece araştırmacının oradaki verilere ulaşması mümkün değildir. Diğer bir problemde verilerin birbirinden ayrı çalışan tablo veya veritabanlarında tutuluyor olmasıdır. Verilerin birleştirilmesi, dönüştürülmesi, hazırlanması bilgisayar ve yazılımları konusunda yeterli bilgiye sahip olmayı gerektirmektedir. Turizm alanında çalışan araştırmacıların bu alanda yeterli olmadıkları görülmektedir. Veri tabanlarından veri alabilmek için SQL, Oracle gibi veri yönetim sistemlerini en azından istenen veriyi alabilecek kadar bilmek gereklidir.

İşletme bünyesindeki veri tabanları dışında web siteleri ve sosyal medya platformlarında bulunan kullanıcı tarafından oluşturulan içeriklerden veri elde edilebilir. Bu verilere ulaşmak araştırmacılar açısından veri tabanlarındaki verilere ulaşmaktan daha kolaydır. Veri tabanı enformasyon sisteminin çalışma sistemine göre tablo ve alanlar ile yapılandırılmıştır. Kullanıcı bazlı içerikler ise yapılandırılmamış verilerdir. Kullanıcı bazlı içeriklerden oluşan veri setlerini

analize hazırlamak için bir hayli emek harcamak gerekmektedir. Bu konuda da araştırmacıların imdadına metin madenciliği teknikleri yetişmektedir.

Veri madenciliği, büyük veri havuzlarında yararlı bilgileri otomatik olarak keşfetme sürecidir ve bilgiyi, insanlar için kolayca anlaşılabilir bir biçimde keşfetmek ve sunmak için makine öğrenimi ve istatistiksel ve görselleştirme tekniklerini kullanmaktadır (Bose, 2009: 936). Veri madenciliğinin dört temel adımından ilki “veri toplama”dır. Veri toplama sırasında en uygun verilerin en uygun kaynaklardan toplanması gerekir (Bose, 2009: 936). Veriler bazen birden fazla kaynaktan toplanarak birleştirilebilir. Verilerin niteliği (zenginlik, doğruluk, eksiksizlik ve temsili de dâhil olmak üzere) veri madenciliği modellerinin uygulanabilirliğini sınırlandıran konuların başında gelmektedir (Delen ve Sirakaya, 2006:328). İyi bir sonuç elde edebilmek için verilerin kaliteli olması gerekmektedir.

Sürdürülebilir bir turizm endüstrisi oluşturmak için stratejik planlama ve karar verme ile uğraşan yöneticiler için turist seyahat davranışlarına bakış açısı çok önemlidir. Ancak, uluslararası turistlerin davranışlarını tam olarak yakalama ve anlamada önemli zorluklarla karşılaşmaya devam ediyorlar. Zorluklar temel olarak kullanımda olan verimsiz veri toplama yaklaşımlarından kaynaklanmaktadır (Vu vd., 2015: 222). Anketler ve görüşmeler turistlerden seyahat verilerini toplamak için popüler yöntemlerdir. Bu yöntemler zaman alıcı ve sınırlıdır. Ayrıca toplanan veri sadece belli bir grup için geçerli olmaktadır, farklı kültürlerden gelen turistlerin tercihleri de farklı olabilmektedir.

Bu bölümde veri madenciliği için kullanılacak verilerin öncelikli olarak hangi kaynaklardan elde edilebileceği üzerinde durulmuştur. Problemin çözümü için gerekli verinin kaynaklardan elde edilmesi ve ardından analize hazırlanması için veri kaynağı veri türüne göre takip edilebilecek adımları vardır. Başarılı bir uygulama yapabilmek için verilerin en iyi şekilde hazırlanmış olması gerekmektedir.

DÖRDÜNCÜ BÖLÜM TURİZMDE VERİ MADENCİLİĞİ

Son yıllarda turizm, yaygın olarak bölgesel ve ulusal ekonomik kalkınmaya katkısıyla dünya ekonomisinin en hızlı büyüyen sektörlerinden biri haline gelmiştir. Turizm ürün tasarımı ve geliştirme, iç ve dış kazancın büyüyen bir kaynağı olarak, ilgili ülkelerde yarattığı ekonomik sonuçlar açısından değerlendirildiğinde, bütün ülkelerin/bölgelerin bu sonuçlardan yararlanmak için büyük çaba gösterdiği önemli faaliyetler haline gelmiştir (Liao vd., 2010: 4212). Özellikle ödemeler dengesine olan katkısı ve doğurduğu ekonomik canlılık, pek çok ülke için turizmi cazip bir hale getirmiş, kalkınma planlarında geniş bir yer verilmesini sağlamıştır (Çuhadar vd., 2009: 100).

Turizmden aldığı payı arttırmak isteyen turizm tedarikçileri, daha fazla turiste ulaşabilmek, ulaştığı turistlere ait işlemlerin takibini yapabilmek için son teknolojiyi kullanmaktadır. Günümüzde turizm sektörü, enformasyon sistemlerini en yaygın olarak kullanan sektörlerin başında gelmektedir. Bu sistemler başlangıçta finans ve muhasebe alanlarını destekleyen bir arka ofis işlevi olarak kullanılırken (Mirela vd., 2009: 923), son yıllarda sektörün her alanında kullanılmaktadır. Özellikle internet teknolojisinin en baskın iletişim kanalı haline gelmesiyle hem talep hem de arz tarafında değişiklikler yaşanmıştır. Sektör yöneticileri, rekabet gücünü koruyabilmek için değil, sadece hayatta kalmak için değişen iş ortamına ayak uydurmanın gerekliliğinin farkına varmışlardır (Magnini vd., 2003:95). Turistler, turizm şirketlerinin değer zincirinin bir parçası olarak modern teknolojileri aktif olarak kullanmasını beklemektedirler. Zamanla turistlerin beklentileri sadece turizm hizmetlerinin en iyi şekilde verilmesiyle yetinmeyip, işletmelerin son teknolojiyi kullanıp kullanmamalarını da kapsamaya başlamıştır (Bach vd., 2013: 26).

Turizm ürünlerinin özelliğinden dolayı, müşteriler ürünü satın almadan önce web üzerinden satın alabilecekleri ürünlerin araştırmasını yaparlar. Bunu bilen seyahat sektörü yöneticileri web üzerinden ürün ve hizmetleri ile ilgili bilgileri sürekli yayınlarlar (Lim vd., 2010: 141). Bu durumda seyahatten önce turistlerin ürün seçiminde kullandığı B2C sistemler ya da B2B sistemler üzerinde, turistik ürünler ile ilgili veriler tutulmaktadır. Bu verilerden turistlerin ya da işletmelerin satın aldığı ürünlerin kaydı ilgili firmaların operasyon destek sistemlerine kaydedilmektedir. Operasyon destek sistemlerinde de ürünlerin kayıtlarının yanında bu ürünlerin servisleri ile ilgili çok daha detaylı bilgiler daha sonradan ilgili firmalar tarafından düzenlenmektedir. Seyahat esnasında uçak, araç, rehber, vb ile ilgili veriler operasyon elemanı tarafından sistemlere eklenmektedir. Konaklama operasyon sistemlerinde

müşterilere ait kişisel ve aldıkları ürünlere ait bilgiler kaydedilmektedir. Bir turistik gezi ile ilgili olarak seyahat ve konaklama sektörlerinde kayıtlar açılmakta ve bu kayıtlar seyahatin ve konaklamanın her anında güncellenmektedir.

Web siteleri, sosyal medya ve bazı uygulamalarda oluşturulan ve veri ambarlarında saklanan büyük miktarlardaki bu veriler ilginç gizli bilgileri bulmak için bir potansiyeli temsil etmektedir. Olmeda ve Sheldon (2001)'a göre turizm sektörü, müşteri hizmetlerini geliştirmenin yeni yöntemlerini sürekli araştırmaya ihtiyacı olan sektörlerden birisidir ve daha da kişiselleştirilmiş deneyimleri müşterine sağlamaktadır. Veri madenciliği teknikleri, bu kişiselleştirmeye ulaşmak için gerekli araçları sağlayabilmektedir. Turizm işletmelerinin pazarlama uygulamaları için potansiyel olarak yararlı olan tüketici davranışı eğilimlerini tahmin etmeye yarayan veri madenciliği teknikleri, planlama ve yüksek kalite pazarlama araştırması yapabilmek için uygulanabilmektedir. (Mirela vd., 2009).

Özetle veri madenciliği turizm sektöründe, müşteri ve pazar bölümlenme, müşteri memnuniyetinin ya da memnuniyetsizliğinin tespiti, seyahat davranışı modelleme, turist tahmini, turizm talebi tahmini, müşteri tercihleri, yeni ürün oluşturma ve ürün değerlendirme gibi turizm alanında yapılabilen tüm araştırma konularında kullanılabilir tekniklere sahiptir. Veri madenciliği teknik ve yöntemlerini kullanarak yöneticilerin arzu ettikleri satış miktarına ve karlılık oranlarına ulaşması mümkündür. Ne var ki bu kadar faydalı olduğu birçok araştırmacı tarafından teyit edilen bu yöntemler muhasebe, tıp, hukuk ve benzeri birçok alanda (Yotwasat ve Srivihok, 2014) başarılı bir şekilde uygulanıyor olmasına rağmen turizm alanında hak ettiği kullanım oranını yakalayamamıştır (Delen ve Sirakaya, 2006).

Bach vd. (2013) turizm alanında araştırmalar üzerinden yaptıkları literatür taramasında çalışmaların iki ayrı zaman diliminde farklılık gösterdiğini bulmuşlardır. 1995-2005 yılları arasında yapılan çalışmalar uyguladıkları yöntemler açısından 2006-2013 yılları arasında yapılan çalışmalardan ayrılmış ve makalelerde geçen anahtar kelimeler farklılık göstermiştir. 2005 yılına kadar turizm çalışmalarında veri madenciliği konusunda segmentasyon, tahmin ve pazarlama gibi anahtar kelimeler bulunmuştur (Bach vd., 2013: 28). Karşılaşma dönemi de denilebilecek bu dönemde turizm alanında çalışan araştırmacılar veri madenciliği ile karşılaşmışlar, geleneksel istatistiksel yöntemlerinde kullanılan analizlerin veri madenciliği teknikleri ile yapılabileceği konusu ile ilgilenmişlerdir. Bu dönemde yapılan çalışmaların bazılarında veri madenciliği uygulaması yapılmamış ve nasıl yapılabileceği konusunda bilgiler (örneğin, Olmeda ve Sheldon, 2001; Cho ve Leung, 2002) veya hangi alanlarda veri madenciliği yapılabileceği konusunda öneriler (Magnini vd., 2003) verilmiştir.

Bu dönem (1995-2005 arası) veri madenciliği konusunda turizm yöneticilerini ve araştırmacıları veri madenciliği tekniklerini işletmelerinde ve çalışmalarında kullanmaları konusunda bilgilendirme ve ikna sürecidir. Law (1998) veri madenciliği yöntemlerinin geleneksel yöntemlerden çok daha farklı olmadığını, aksine, veri madenciliği yöntemleri onlarca yıldır var olan analitik tekniklerin doğal uzantıları ve genellemeleri olduğunu iddia etmektedir. Yazara göre (s:62) iş dünyasındaki aşırı ham verilerin anlaşılma problemi göz önünde bulundurulduğunda, bilgisayar bilimi alanındaki araştırmacılar veri tabanlarından pratik, önceden bilinmeyen ve dolaylı olarak yararlı bilgileri almak için veri madenciliği adı verilen bir yaklaşım geliştirdiler. Bilgisayar biliminden, özellikle yapay zekâdan ödünç alınan teknik yaklaşımlar, veri madenciliği sürecine dâhil edilmiştir. Veri madenciliğinde kullanılan yaygın yapay zekâ metodolojileri, büyük paralel dizi işlemcileri, bulanık mantığı ve sinir ağını içermektedir.

O dönemde Uysal (2004) araştırmacıların moda olanı değil, hedeflerine en uygun olanı seçmeleri gerektiğini belirterek veri madenciliği ve yapay sinir ağları üzerinde durmaktadır. Olmeda ve Sheldon, (2001) veri madenciliği tekniklerinin pazarlamada ve elektronik müşteri ilişkileri yönetiminde potansiyel kullanımlarını incelemektedir. Çeşitli veri madenciliği teknolojileri tanımlanmış ve seyahat endüstrisindeki potansiyel kullanımlarını açıklamıştır. Magnini vd., (2003) veri madenciliğini tanıttıktan sonra yöneticilerin otellerini en üst seviyelere taşımak için pazarlama stratejisi geliştirme potansiyeline sahip olduğunu vurgular. Bu tezini pekiştirmek için Harrah'ın otel ve casinolarının 1997 yılında, müşterilerin satın alma etkinliklerini izleyen ve Harrah'ın tesislerinde harcamaları teşvik eden çalışmalarının başarısından bahsetmektedir. Yazarlar ayıca otel endüstrisi için veri madenciliği uygulamaları ve etkili bir veri madenciliği için kılavuzlar sunmakta ve karşılaşılabilecek olası sınırlılıklar ve zorluklar konusunda da uyarmaktadır. Otelcilikte veri madenciliği zorunluluktur diyen Law (1998), veri madenciliği hakkında doğru olmayan bazı söylem ve uygulamaların doğrusunu, olması gerekeni bildirmektedir.

Magnini vd., (2003) veri madenciliği tarafından gerçekleştirilen görevleri sınıflandırma, kümeleme, sapma tespiti, ilişkilendirme ve tahmin olarak beş kategoride gruplandırmışlardır. İlk dönemde bu görevler ağırlıklı olarak müşteri ilişkileri yönetimi (pazarlama, bölümlenme) ve ulaşım (taşımacılık) konularında kullanılmışlardır.

Pazarlama konusunda ağırlık olarak web sitesi pazarlama (Lau vd., 2001; Olmeda ve Sheldon, 2001) ve veritabanı pazarlama (Cho ve Leung, 2002; Ha ve Park, 1998) üzerinde durulmaktadır. Lau vd. (2001) turizm büroları ve potansiyel yolcular arasında doğrudan pazarlama kanalları kurabilmek için web sitelerini önermekte ve metin madenciliği teknikleri

ile potansiyel bir yolcunun kişisel web sitesinden ilgi alanlarının ve ihtiyaçlarının belirlenebileceğini savunmaktadır. Yazara göre potansiyel bir müşteriye, elektronik postalar gönderilebilir, müşteri için tasarlanmış özel hizmet paketleri sunulabilir. Aynı şekilde Olmeda ve Sheldon, (2001: 7) da seyahatle ilgili çok sayıda web sitesi veri madenciliği tekniklerin çok yararlı olabileceğini ve zengin bir bilgi birikimi sağlayabileceğini ifade etmektedir. Yazarlara göre alınan bilgiler, web sayfasının müşteri bilgileri, kullanım bilgileri veya içerik bilgileri olabilir. Yalnız web kullanımı madenciliği, web yapısı madenciliği ve web içeriği madenciliği arasında ayırım yapmak önemlidir.

Yaygın olarak kullanılmasına rağmen, veri madenciliği yalnızca yapılandırılmış, sayısal veritabanlarına uygulanabilir. Bununla birlikte, işletme bilgilerinin çoğunluğu, otelin dâhili veritabanlarında veya web tabanlı veri kaynaklarında saklananlar gibi yapılandırılmamış veya yarı yapılandırılmış metin belgeleri biçimindedir. Metin bilgilerinin işlenmesinin geleneksel yolu, bilgi toplama, analiz etme ve yayma konusunda insani eylemleri içermekte ve para, zaman ve insan kaynaklarına ciddi yatırım gerektirmektedir. Ayrıca, nitel metin verilerini nicel sayısal verilerle iş analizlerinde birleştirmek zordur (Lau vd., 2005: 344). Bu nedenle, müşteri ilişkileri ve pazarlama konusunda çalışan araştırmacılar, iş zekâsını geniş metin koleksiyonlarından doğru bir şekilde çıkarabilen ve parçalanmış bilgileri iş zekâsı veri tabanlarına entegre edebilen bir yöntem olarak web madenciliği (Olmeda ve Sheldon, 2001;) ya da metin madenciliğini (Lau vd., 2005) önermektedirler.

Cho ve Leung, (2002) şirketlerin veri tabanlarında bulunan verilerden elde edilen bilgileri pazarlama stratejileri planlamada, pazar bölümlenmesi yapmada ve yeni ürün/hizmeti piyasaya sürmede kullanarak turizmde niş pazarları desteklemeyi önermektedir. Yazarlar diğerleri gibi veri madenciliği teknikleri ve turizmde veri tabanı pazarlamanın muhtemel uygulamalarından bahsetmektedirler. Ha ve Park, (1998) da veri madenciliği araçları ve yeni pazarlama stratejilerini tanıtmakta ve tamamen veri madenciliğinden elde edilen bilgileri kullanan yeni bir pazarlama stratejisi önermektedir. Law ve Au, (2000) turizm alışverişi için özel bir pazar segmentine sahip olan karar vericiler, daha doğru planlama faaliyetlerini gerçekleştirebilir diyerek, bir dizi karışık sayısal ve sayısal olmayan turizm alışverişi verileri arasında var olan ilişkileri modellemek için kaba kümeler teorisini içeren bir yaklaşım sunmaktadır. Ayrıca Kaba küme teorisinin turizm alışverişi ilişkileri modelleme ve tahminine dâhil edilmesi gerektiğini savunmaktadır.

Dolničar, (2004) turizmin segmentasyon çalışmalarını gözden geçirmekte, bölümlendirme yaklaşımlarının sistematikliğini önermektedir ve bu sistematikten ortaya çıkan yeni

yaklaşımların yönetsel yararlılığını göstermektedir. Ana amaç, akademisyen ve uygulayıcılara pazar anlayışını arttırmak için kullanılabilir bir keşif tekniği menüsü sunmaktır. Min vd. (2002) hedef müşterilerin ve promosyon verilecek müşteri profillerinin tespiti için karar ağaçlarını (if – then) kullanmışlardır.

Veritabanlarından bilgi keşfi olarak bakılan veri madenciliği için veri ve veritabanı otel tesis yönetim sistemleri ve müşteri ilişkileri yönetim sistemlerinde biriken verilerden yararlanılmıştır. Pyo vd., (2002) veri tabanındaki bilgi keşfinin (VTBK) çeşitli yönlerini tartışmak ve uygulamadaki yerini turistik yerlere aktarmayı amaçlamışlar ve özellikle VTBK'nin operasyonel konuları, VTBK uygulamaları, veri ambarı ve destinasyon yönetimi, bilgi dağıtımı ve depolama konularını ele alınmışlardır.

Turizmin bel kemiğini oluşturan taşımacılık, ulaşım ve seyahat davranışı konuları bu dönemde veri madenciliği uygulamalarının en yoğun yapıldığı alanlardan birisi durumundadır. Hensher ve Ton, (2000), yapay sinir ağlarının (YSA) birçok trafik ve ulaşım sorununun incelenebileceği bir çerçeve olarak kullanılmasını araştırmaktadır. Nijkamp vd. (1996) İtalya'daki yüksek hızlı trenin etkisini karayolu ve demiryolu taşımacılığı modları arasındaki rekabete özel bir odaklanma ile değerlendirmek için hem mantık modellerini hem de sinir ağını kullandılar. Özellikle, sinir ağ modelinin, uygulamasında kritik problemlerin olmasına rağmen, daha iyi bir performansa sahip olduğunu gördüler. Aynı şekilde Cantarella ve Luca, (2003)'ün elde ettiği sonuçlar da Yapay Sinir Ağları'nın (YSA) seyahat talebi analizi için etkili bir araç olduğunu göstermektedir. İsrail'deki kadın ve erkeklerin seyahat talebi modellerinin karşılaştırılması sinir ağı metodolojisi ile yapılmıştır (Shmueli vd., 1996: 152).

Knoblock, (2004) hafta boyunca bir havayolu uçuşunun erteleneceğini tahmin eden bir veri madenciliği sistemi geliştirmişlerdir. Geliştirilen sistem, hava koşullarından kaynaklanan uçuş gecikmeleri hakkında tarihsel bilgiler, gecikmelerin meydana geldiği havaalanlarındaki gerçek hava koşulları hakkında tarihsel bilgiler ve mevcut hava durumu tahminlerinden elde edilen bilgileri bir araya getirerek muhtemel bir uçuş için gecikme tahminleri yaptı. Nagao vd., (2004) doğru kişisel turizm etkinliği bilgilerini elde etmek için bir GPS log madenciliği yöntemini önerdiler ve araç kiralayan turistlerden elde edilen GPS log verisinden kişisel turizm aktivite bilgilerini çıkarttılar. Xie vd. (2003) veri madenciliği yöntemlerinin (karar ağaçları (DT) ve sinir ağları (SA)) çalışma seyahat modu seçim modellemesi üzerindeki yeteneklerini ve performansını incelemektedir.

2000 yılından sonra özellikle enformasyon teknolojilerinde ve web teknolojilerinde inanılmaz gelişmeler yaşanmıştır. Çevirmeli modemlerin yerini alan ADSL modemler internetin kapasitesini ve hızını eskiye oranla birkaç kat arttırmıştır. Bu dönemde firmalar çok sayıda

farklı işlemlerde çok fazla veriyi yakalayan özellikle müşteri ilişkileri yönetimi (MİY) sistemleri ve enformasyon teknolojilerine büyük yatırımlar yaptılar. Bunlar arasında havayolu veya otel check-in kayıtları, rezervasyon şekilleri, çeşitli tesislerin kullanımı, kredi kartı kullanım şekilleri, sık kullanıcı veya sadakat kartı kayıtları ve kablosuz ses ve veri kayıtları bulunmaktadır (Verma, 2010: 473). Kullanıcılar ise özellikle Web 2.0 ile birlikte önceleri sadece bilgi edinmek için girdikleri internette bilgi veren konumuna gelmişlerdir. Tüm gelişmelerden en fazla etkilenen sektör yine turizm olmuştur.

Turizm araştırmacı ve yöneticileri için tanışma dönemi diyebileceğimiz 2006 ile 2013 yılları arasında önceki dönemdeki araştırma yöntemleri (segmentasyon, tahmin ve pazarlama) kullanılmaya devam ederken daha gelişmiş sistemlere ve teknolojilere (örneğin, teklif veren ve bulanık) yönelik anahtar kelimeler (Bach vd., 2013:28) araştırmalara yansımıştır. Daha önce araştırmacıların elinde yüz yüze görüşmelerden, anketlerden ve veri tabanlarından elde ettikleri veriler varken, bu dönemde bu verilerin yanında web siteleri ve sosyal medyada yayınlanan kullanıcı yorumları, CBS vb teknolojilerden elde edilen veriler eklenmiştir. Araştırmacıların araştırmalarına yeni yöntemler eklenmiş olmasına rağmen bilişim teknolojilerinin gelişimi yanında çok geride kalmıştır. Devasa veri yığınlarının biriktiği turizm sektöründe veri madenciliği uygulama ve araştırmaları aynı yoğunlukta olmamıştır.

Bunca veriye rağmen turizm sektöründe veri madenciliğinin yeteri kadar yaygınlaşmamasının temelinde araştırmacıların verilere ulaşmasındaki zorluklar yatmaktadır. Çünkü veri olmadan veri madenciliği olamamaktadır. Aslında merkezi rezervasyon ve tesis yönetim sistemlerinin çoğalmasının tetiklediği, turizm şirketlerinde büyük miktarda tüketici verisi birikmektedir. Sektörde biriken devasa verilerden yararlanma konusunda araştırmacıların karşılaştığı iki büyük problem bulunmaktadır. Bunlardan birisi işletmelerin bünyesindeki sistemlerde biriken verilere ulaşım zorluğudur. İşletme izin vermediği sürece araştırmacının oradaki verilere ulaşması mümkün değildir. Bazı ülkelerde bulunan Kişisel Verileri Koruma Yasası'nın kısıtlamasından dolayı araştırmacılar işletmelerin müşteri veri tabanına ulaşamamaktadır (Liao vd., 2010: 4214). Diğer bir problemde verilerin birbirinden ayrı çalışan tablo veya veri tabanlarında ya da yapılandırılmamış halde farklı platformlarda tutuluyor olmasıdır. Verilerin toplanması, birleştirilmesi, dönüştürülmesi, hazırlanması bilgisayar ve yazılımları konusunda yeterli bilgiye sahip olmayı gerektirmektedir. Turizm alanında çalışan araştırmacıların veri tabanları üzerinde araştırma yapabilmeleri için veri tabanı yönetim sistemlerini iyi şekilde kullanabiliyor olmaları gerekmektedir.

Bu ve benzeri sorunlardan dolayı bu dönemde ilk dönemde olduğu gibi bazı araştırmacılar tarafından bilgilendirme ve ikna süreci dönemin ortalarına kadar devam etmiştir.

Yönetici ve araştırmacıların karşılaştıkları zorluklar ya da işlerini daha iyi yapabilmeleri için gerekli ihtiyaçları anlatıldıktan sonra veri madenciliğinin bu sorunlara çözüm üreteceğinden bahsedilmektedir. Bu dönemde genelde başka alanlarda yapılan araştırmalardan esinlenerek turizmde de benzerlerinin yapılabileceğine dair çok öneri çalışması yayınlanmıştır.

Danubianu ve Hapenciuc, (2008), veri madenciliği ve müşteri ilişkileri yönetimi konularında detaylı bilgiler vermişler ve veri madenciliğinin aşırı veri yüklemesini azaltarak karar vermeyi kolaylaştırdığını ve yüksek kar potansiyeline sahip müşterileri içeren pazar segmentlerini tanımlamasına yardımcı olabileceğini belirtmişlerdir. Ayrıca veri madenciliğinin müşteri ilişkileri yönetimi kapsamında hangi amaçlar için kullanılabilceğini açıklamışlardır. Čech & Bureš, (2009), e-turizmde kullanılacak ileri teknolojilerin eleştirel bir incelemesini sunmakta ve özellikle veri madenciliği, semantik web ve çok ajanlı teknolojiler üzerine odaklanmaktadır.

Bose (2009) istatistiksel teknikler ile herhangi bir analiz gerçekleştirilemeden önce verilerin dağılımı ile ilgili çeşitli varsayımların yapılması gerektiğini, bunların ihlal edilmesi durumunda, sonuçların geçersiz olabileceğini, bundan dolayı, araştırmacıların, turizm veri analizi için makine-öğrenme tabanlı veri madenciliğini kullanmalarını sağladığını belirtmiştir. Bose (2009)'a göre turizm endüstrisinde veri madenciliği teknikleri; turistlerin harcamalarını tahmin etmek, turist profillerini analiz etmek ve turist sayısı tahmini için kullanılabilir. Aynı çalışmada, veri madenciliği tekniklerinin bu etkinlikleri desteklemek için nasıl kullanıldığını göstermek için örnekler de sunulmuştur.

“Turizmde veri madenciliği teknikleri nasıl kullanılabilir?” diye soran Mirela vd. (2009), Magnini vd. (2003) gibi veri madenciliği tekniklerinin turizmde kullanım alanlarını tanıtmaktadırlar. Yazarlara göre; pazar gruplarının büyüklüğünü ve yapısını izlemek, tahmin edici modellerle faaliyetleri sınıflandırmak için sınıflandırma, demografik özellikler, yaşam biçimleri ve gerçek ürün alımları gibi bilgilerle müşterilerin kim olduklarını anlama konusunda kümeleme, bir müşterinin satın almasındaki normal dışı davranışlarını tespit etmek için sapma tespiti, müşteri satın alma geçmişleri ile ilgili özel bilgileri sonraki alışverişe ilişkin olasılıklı kuralları formüle etmek için birliktelik kuralları, veri içindeki desenlere ve eğilimlere dayalı sürekli değişkenlerin gelecekteki değerini öngörmek için tahmin teknikleri turizm verilerinin analizi için kullanılabilir.

2005-2007 yılları arasında konaklama ve turizmde enformasyon teknolojileri uygulamaları konusundaki yayınlarda veri madenciliği tekniklerinin ana kullanım alanları tahmin ve pazarlamadır (Law vd., 2009: 608). Delen ve Sirakaya (2006), yapay sinir ağları, karar ağaçları

ve kaba kümelerin üç popüler veri madenciliği yöntemini test etmişler ve kaba kümeler algoritmasının üçü arasında en iyi tahmin aracı olduğunu bulmuşlardır. Benzer şekilde, Kon ve Turner (2005), sinir ağlarının ve temel yapısal yöntemin tahmin doğruluğunu karşılaştırmış ve BSM'nin turizm talebini tahmin etmede daha yüksek bir doğruluk sağladığını teyit etmiştir. Petropoulos vd. (2006), turizm talebini tahmin etmek için bir teknik analiz sistemi geliştirmişlerdir. Bloom (2005), sinir ağları uygulamalarının, pazar segmentlerindeki ve içindeki turistlerin değişen davranışlarını izlemek için kullanılmasını önermiştir.

Seyahat ve turizm endüstrisinde bu dönemin tahmin modelleme uygulamaları, müşteri ilişkileri yönetiminden oluşmaktadır. Müşteri ilişkileri yönetimi (MİY) sistemleri ve enformasyon teknolojileri ile kayıt altına alınan havayolu veya otel check-in kayıtları, rezervasyon şekilleri, çeşitli tesislerin kullanımı, kredi kartı kullanım şekilleri, sık kullanıcı veya sadakat kartı kayıtları ve kablosuz ses ve veri kayıtları müşteri tercihlerini uzun bir süre boyunca izlemek için kullanılabilir (Verma, 2010). Müşteri ilişkileri yönetimi kapsamında veri madenciliği doğrudan postalama hareketine cevap vermesi muhtemel olan ziyaretçileri sınıflandırmak için kullanılabilir. Ayrıca, veri madenciliği, her ziyaretçinin muhtemel değerini tahmin etmek, ziyaretçileri elde tutmak için pazar görüşlerini planlamak ve ziyaretçinin bağlılığı hakkında bilgi üretmek için kullanılabilir (Karathiya vd., 2012: 116). Bu tür işlem verilerinin etkin kullanımı, kuruluşların ürün-hizmet sunumlarını çeşitli müşterilerin kullanım alışkanlıklarına göre özelleştirebilmelerine ve böylece memnuniyet, elde tutma ve sadakatin artmasına izin verebilmektedir (Loveman, 2003; Verma, 2010). Kümeleme, karar ağaçları vb. gibi veri madenciliği teknikleri, çeşitli sektörlerdeki müşterileri başarılı bir şekilde bölümlere ayırıp hedeflemek için yaygın olarak kullanılmaktadır. Veri madenciliği, müşteri davranışlarındaki kalıpları keşfetmek ve anlamak için etkili bir yaklaşım sağlar ve böylece karar vericinin müşterileri daha iyi gruplandırmasına yardımcı olmaktadır (Hoontrakul ve Sahadev, 2008: 60).

Hızlı büyüyen turizm endüstrisinin önemli ekonomik katkıları, ziyaretçilerin davranışsal ve demografik modellerini anlama konusunda dünya çapında dikkat çekmiştir. Law vd., (2009: 608)'in tahmin ile birlikte tespit ettiği bir diğer çalışma konusu pazarlamadır. Bir DMO'nun yöneticisi, turist profillerini ve seyahat modellerini ne kadar iyi anlarsa, hedeflerini daha iyi pazarlayabilirler. J. Wong vd. (2006), Kuzey Tayvan yolcularının seyahat modellerini analiz etmek için veri madenciliği tekniklerini kullanmışlar ve Asya ülkelerindeki DMO'ların Tayvan'daki varış yerlerini tanıtmaları gerektiğini önermişlerdir. Law vd., (2006: 135), veri madenciliğinde akademik araştırmacılar tarafından daha önce gözden kaçırılan bir pazar segmenti olan iş seyahatindeki kişilerin temel bilgilerini yakalayabilecek kaba kümeler temelli bir model geliştirmeye çalışmaktadır. Hoontrakul ve Sahadev, (2008: 75) çevrimiçi sorgulama

modellerini anlamak için bir kümeleme tekniğinin kullanıldığı bir örneğe bakmaktadır. Bu bağlamda uygulanan kümeleme süreci iki ana amaca sahiptir: müşteri profilini anlamak ve dönüşüm oranını artırmak için eyleme geçirilebilir stratejiler üretmek. Byrd ve Gustke, (2007) diğer çalışmalar gibi müşteri segmentasyonu yerine, paydaş segmentasyonu üzerinde durmuş ve turizm planlama, geliştirme ve yönetimde paydaşların katılımını araştırmaktadır. Karar ağaçları ile topluluklarında sürdürülebilir turizm gelişimi için destekleriyle ilgili paydaş grupları belirlemişlerdir.

Ayrıca, çeşitli araştırmacılar, turizm sitelerinin pazarlama etkinliğini artırmak için modeller önermişlerdir. Örneğin, Law (2005), pazarlama etkinliğini artırmak için Asya'nın otel endüstrisinde uygulama için Nesne Tabanlı bir Veritabanı Pazarlaması modeli başlatmış ve Mills vd. (2007), bir web sitesinin reklam etkinliğini artırabilecek bir Modifiye İnteraktif Turizm Reklamcılığı Modelini tanıtmıştır. Al-Salim, (2007) bir yandan hizmet sağlayıcının işleyişini ve işlem maliyetlerini en aza indirgeyen seyahat paketlerini tasarlamak, diğer yandan da müşteri memnuniyetini en üst düzeye çıkarmak için paketlerin bileşenlerini hizalamak üzere bir kitlesel özelleştirme stratejisi kullanmaktadır. Metodolojisi, ilgili seyahat paketi bileşenlerinin müşterilerin daha önceki satış kayıtlarına dayanarak bir araya gelme olasılıklarının yüksek olduğu yumuşak bir gruptandırmadan oluşmaktadır. Birliktelik kuralları tekniği bir veri madenciliği tekniği olarak bu yumuşak gruptandırmayı gerçekleştirmek için bir araç olarak uygulanmaktadır. Hizmet sektöründeki kitlesel özelleştirmenin, günümüzde rekabetçi pazarda rekabet etmek için gerekli olan pazarlama ve işletme kararlarını iyileştirmek için kullanılacak güçlü bir taktik olduğunu savunmaktadır.

E-dağıtım kanalları aracılığıyla otel odası rezervasyonlarının aşırı büyümesi, doğal olarak çevrimiçi yorumların eşi benzeri görülmemiş bir oranda artışına neden olmuştur (Duan vd., 2016: 284). O'Connor (2010), otellerin yorumları takip etmek ve çevrimiçi itibarı yönetmek için sosyal ağlar ve kullanıcı tarafından oluşturulan içerik kavramını aktif olarak benimsemesini önermektedir. İnternet teknolojisinin yaygın kullanımı ile elektronik ağızdan ağıza [eWOM] iletişiminin ürün ve hizmetlerin çevrimiçi yorumlar yoluyla tüketicinin davranışları ve tercihleri üzerinde güçlü bir etkisi bulunmaktadır (Rong vd., 2012: 731). İşletmelerin ve yöneticilerin internette yayılan yorumlara kayıtsız kalması düşünülemez. Daha çok metin içerikli bu yorumları analiz etmek için sayısal verilerin analiz edilmesine alternatif olarak, metin madenciliği başka iyi bir seçimdir. Lau vd.(2005), metin madenciliğinin çevrimiçi metin analizi için nasıl bir araç olarak kullanılabilceğine dair üç örnek göstermiştir. Metin madenciliği tekniklerinden yararlanarak, Singh vd. (2007), 1994 ve 2003 yılları arasında Uluslararası Ağrlama Yönetimi Dergisi'nde yayınlanan insan kaynakları yönetimi üzerine araştırmaları

incelemişler ve ortaya çıkan dokusal içerikleri dokuz araştırma temasına ayırmışlardır. Birlikte-lik kuralı tekniğini uygulayan Rong vd. (2012: 731), Hong Kong'da toplanan bir yurt içi turizm veri setini kullanarak turizm sektörü bağlamında eWOM'u araştırmışlardır. Çalışmada profiller ve çevrimiçi deneyim paylaşımcıları ile seyahat web sitesi tarayıcılarının ilişkileri araştırılmaktadır. Ampirik sonuçlar, turizm yöneticilerinin yeni potansiyel müşterileri tanımlamalarına ve daha etkili pazarlama stratejileri planlamalarına yardımcı olmada faydalıdır.

Son dönemlerde turistler seyahat öncesi bilgi araştırmak için, ürün ve hizmetleri satın almak için ve seyahat sonrası deneyimlerini paylaşmak için web sitelerini ve sosyal medyayı daha yoğun olarak kullanmaya başlamışlardır. Turist ve turist davranışı, tercihi, memnuniyeti gibi alanlarda özellikle sosyal medya üzerinde yoğun bir veri yığılması olmuştur. Bu da turizm alanında 2016 yılından sonra web ve sosyal medyada kullanıcı tarafından oluşturulan içerikleri üzerinde yapılan araştırmalarda inanılmaz bir artışa neden olmuştur. Metin madenciliği en fazla kullanılan tekniklerden birisine haline gelmiştir.

Turizm araştırmaları ile aynı anda başlamış olan öneri sistemleri son dönemlerde alt yapılarında veri madenciliği yöntemlerini kullanmaya başlamışlardır. Figueredo vd. (2018) turizm alanında önerilerde bulunabilmek için kullanılan bir teknik olan işbirlikçi filtreleme yaklaşımını geliştirmişlerdir. Zhu vd. (2017) veri madenciliği tekniklerini kullanarak yeni bir seyahat ürünü öneri motoru geliştirmişlerdir. Leal, vd. (2017) makine öğrenimi ve veri madenciliği metodolojilerinin turist kitlesi kaynaklı verilere dayanarak geçici seyahat aşamaları önereceğini savunmaktadırlar.

Son dönemde daha önceki dönemlerde yaygın olarak kullanılan konular üzerindeki araştırmalara devam edilmiştir. Ancak bu dönemde yapılan çalışma adetlerinde artış söz konusudur. Aşağıdaki bölümlerde turizm konularında yapılan veri madenciliği çalışmaları hakkında detaylı bilgiler verilmiştir. Burada müşteri ilişkileri yönetimi, ürün yönetimi ve müşteri memnuniyeti genel anlamda birbirleri ile ilişkilidir ve birçok çalışmada bu üç konu ya da ikisi birlikte kullanılmıştır. Burada çalışmada ağırlık verilen konulara göre çalışmalar sınıflandırılmıştır.

4.1. Müşteri İlişkileri Yönetimi

Müşteri İlişkileri Yönetimi (MİY), tüketici davranışını anlamak ve etkilemek için farklı teknikler kullanan kurumsal müşteri odaklı bir yaklaşımdır. Tüketici ilişkilerini tüm yönleriyle etkilemek (müşteri memnuniyetini artırmak, müşteri sadakatini artırmak veya karlılığı arttırmak) ve bir kurumdaki çalışanların MİY araçları kullanmasını sağlamak gibi iki amacı olan bir süreçtir (Danubianu ve Hapenciuc, 2008: 3261). Müşteri ilişkileri yönetimi, uzun süreli ve

başarılı ilişkiyi başlatmaya, kurmaya, sürdürmeye ve geliştirmeye yönelik tüm iş faaliyetlerini ifade etmektedir ve organize bir şekilde kurumsal müşteri ilişkilerini yönetmeye yardım eden yöntem ve araçlar kümesidir (Liao vd., 2010: 4212). Özellikle otellerde müşterileri hakkında çok miktarda değerli bilgi ürettikleri için müşteri memnuniyeti ve birikimini iyileştirmek için kilit bir strateji haline gelmiştir.

MİY özellikle konaklama sektörü için bir işletmenin kazançlarını arttırmak ve uzun süreli hayatta kalmalarını sağlamak için en iyi stratejilerden birisidir (Talón-Ballestero vd., 2018:187). İşletmeler de aşırı rekabetçi ve küresel bir pazarda ayakta kalmanın, müşterilerle daha yakın ilişkiler gerektirdiğini fark etmişlerdir. Müşterileri edinmek, ilişki kurmak ve elde tutmak en öncelikli konular haline gelmiştir. Artan pazar rekabetçiliği ve yeni müşteri alımına kıyasla eski müşteriyi tutmanın daha az maliyetli olması bunun başlıca nedenlerindedir (Talón-Ballestero vd., 2018: 188). Mevcut bir müşteriyi elde tutma maliyeti, yeni bir müşteri geliştirme maliyetinin yaklaşık 1 / 5'i kadardır (Chiang, 2018: 673). Bu nedenle, müşteri deneyimini kişiselleştirmek ve memnuniyetlerini arttırmak için MİY önemli bir strateji haline gelmiştir (Talón-Ballestero vd., 2018: 188). Birçok şirket için müşteri ilişkilerinin kalitesi, diğer işletmelere karşı rekabet avantajı sağlamaktadır. Ayrıca, müşterinin tanımı, doğrudan tüketicileri, ortakları ve satıcıları kapsayacak şekilde (başka bir deyişle, katılan, bilgi veren veya şirketten hizmet isteyen herkes) genişletilmiştir (Danubianu ve Hapenciuc, 2008:3261). Ayrıca işletmeler çağrı merkezleri ve yardım masaları gibi hizmetlerle müşterilerini korumaktadırlar (Danubianu ve Hapenciuc, 2008: 3262). Müşterileri anlamak ve onlarla olan ilişkileri ve ilişkileri yönetmek, müşteri ilişkileri yönetiminin (MİY) hayati bir parçasıdır. İyi MİY'e sahip bir işletme, müşteri davranışlarındaki olası değişiklikleri tahmin edebilmelidir. Müşteri davranışını tahmin etme, müşteri davranış modelleri hakkında daha iyi bir fikir edinmek ve gelecekteki davranışları tahmin etmek için müşteri davranış modellemesi, araç ve teknikler uygulayarak elde edilebilir (Bell ve Mgbemena, 2018: 196).

Genel olarak MİY, dört unsurdan oluşan bir çerçeve ile tanımlanabilir: Bilmek, Hedeflemek, Satmak ve Hizmet Vermek. Bu bağlamda MİY öncelikli olarak işletmelerin kendi pazarlarını ve müşterilerini tanımasını ve anlamasını gerektirir (Danubianu ve Hapenciuc, 2008: 3262). Müşteri bilgisi, konaklama sektörü için hayati önem taşımaktadır ve bu, daha kaliteli hizmetlerle (yani, daha fazla uyarlanmış ve özelleştirilmiş) müşteriyle ilişkilerde ve pazarlama stratejilerinin yaklaşımıyla teklifin iyileştirilmesinde önemli bir rol oynamaktadır. Bu işlemler de sadakati arttıran ve müşterileri tekrarlayan, daha yüksek kârlılık sağlayan daha iyi müşteri memnuniyeti ile sonuçlanır (Talón-Ballestero vd., 2018: 187). Bunun için, işletmelerin müşteri bilgilerine sahip olması gerekmektedir. Bu bilgiler iç kaynaklardan (müşteriler,

müşteri anketleri veya işlem sistemlerinde yer alan davranışsal veriler) özet olarak alınabilir veya veriler dış kaynaklardan elde edinilebilir (Danubianu ve Hapenciuc, 2008: 3262).

Müşterilere ait iç kaynaklardan bilgi edinmenin yolu işletme bünyesinde bir MİY sistemi kurmaktır ve kurumsal bir veri ambarına sahip olmaktır. Bir MİY sistemi “tüketicilerle karlı ilişkiler geliştirmek, sürdürmek ve güçlendirmek için tüketici bilgisini geliştirmek ve güçlendirmek için teknolojiye dayalı sağlam bir araçtır”. Bir MİY sistemi küresel bir MİY stratejisinin önemli bir parçasıdır ve müşteriler ile uzun vadeli ilişkilerin kurulmasını kolaylaştıran altyapı sağlar (Talón-Ballesteros vd.,2018: 188). MİY sistemlerinin uygulanması, misafir davranış kalıplarını tanımlamaya ve bunları uzun vadede muhafaza etmeye olanak tanımaktadır (Talón-Ballesteros vd., 2018: 187).

İşletmelerin müşteri verilerine kendi bünyelerindeki sistemlerin dışında da ulaşması mümkündür. İnternet'in artan yaygın kullanımı, hem tüketiciler hem de işletmeler için temel bilgi kaynağı haline gelen çeşitli sosyal medya sitelerinde kullanıcı tarafından oluşturulan içeriğin hızlı gelişmesini sağlamıştır. Özellikle, internette yayınlanan ürün ve hizmetlerin kullanıcı tarafından oluşturulan yorumları, dijital müşteri-müşteri etkileşimleri için en güçlü bilgi paylaşım aracı haline gelmiştir (Duan vd., 2016: 282). Özellikle otellerin derecelendirilmesine adanmış olan web siteleri, küresel GSYİH'nın% 9,4'üne katkıda bulunan turizmin artan etkisinden dolayı büyük ölçüde popülerlik kazanmıştır (Geetha vd., 2017: 44). Sosyal Medya ve web sitelerinde (Yorum siteleri: TripAdvisor ve Yelp, sosyal ağlar ve mikroblog siteleri: Facebook ve Twitter, sosyal paylaşım: YouTube ve satın alma / yorum siteleri: Amazon ve Travelocity) kullanıcılar tarafından oluşturulmuş ve bir işletmenin müşterisini tanımasını sağlayacak hazine değerinde veriler mevcuttur.

Çevrimiçi kullanıcılar, ürün ve hizmetleri değerlendirmek için bu tür ağızdan ağıza iletişim (WOM) bilgilerine giderek daha fazla güvenmektedir. Yakın zamanda yayınlanan bir raporda, on tüketiciden yaklaşık dokuzu, yerel bir işletmenin kalitesini belirlemek için çevrimiçi yorumları okuduğunu ve tüketicilerin yüzde 88' inin, kişisel önerilerin yanı sıra çevrimiçi yorumlara güvendiğini göstermektedir. İşletmeler, tüketici davranışlarını ve beklentilerini daha iyi anlamak ve pazar fırsatlarını tanımlamak için bilgileri, kendi ve rakiplerinin ürün ve hizmetlerinin geri bildirimlerini ve değerlendirmelerini toplamak için kullanmanın önemini giderek daha fazla fark etmektedir (Duan vd., 2016: 282).

Müşteri bilgileri müşteri içinde gizli olduğu için (Liao vd., 2010: 4213), potansiyel değeri için toplanması gereken verinin tamamını keşfetmek gerçekten kolay değildir. Bu nedenle, müşteri verilerinin ne kadar etkili işlendiği ve kullanıldığı giderek daha da önemli hala

gelmektedir. İç kaynaklardan elde edilen verilerin yapısı işletme amaçlarına uygun olarak tasarlanmış olabilir ama dış kaynaklardan gelen veriler için bu neredeyse mümkün değildir. Veriler alt yapısı farklı olan birden fazla kaynaktan gelecektir ve bu veriler üzerinde analiz yapmak gerçekten zordur. Üstelik bu kaynaklardan gelen veri, anketler ile toplanan veriler gibi birkaç yüz görüşmeden / gözlemden değil, belki de milyonlarca kullanıcı yorumundan oluşacaktır.

Bu da iş ve bilimsel uygulamalarla elde edilen büyük miktardaki depolanan verilerin analizi, anlaşılması hatta görselleştirilmesi için yardımcı olacak yeni teknikleri gündeme getirmiştir. Geliştirilen yeni teknikler arasındaki veri madenciliği, desenler, ilişkiler, değişiklikler ve veritabanlarında depolanan büyük miktarda verilerde önemli yapılar gibi önemli müşteri bilgilerinin keşif sürecidir (Liao vd., 2010). Pek çok uluslararası işletme müşteri alışverişi niyetlerini anlamak için, veri madenciliği teknolojileri ile büyük verileri (müşteri alışveriş kayıtları) ile POS (satış noktası) sistemlerinden (Chiang, 2018: 665) ve sosyal medya üzerinden müşteri bilgisini keşfetmektedir.

Bununla birlikte, MİY sistemlerinde hâlihazırda mevcut olan büyük miktardaki bilginin etkin müşteri profili oluşturmada oynayabileceği rol, henüz basit ve iyi bilinen istatistiksel açıklamalar için bile yeterince çalışılmamıştır. Buna ek olarak, otel MİY sistemlerinden toplanan kapsamlı verilerden müşterileri etkili bir şekilde takip etmek için gelişmiş analiz tekniklerinin henüz otel endüstrisinde uygun bir şekilde kullanılmadığına dair kanıtlar bulunmaktadır (Dursun ve Caber, 2016). Oteller, MİY sistemlerinin potansiyelini tam olarak kullanmamaktadır, devam eden çalışmalar ve güçlü bir ilgi vardır (Talón-Ballesteros vd., 2018: 188). Başarılı bir çevrimiçi MİY müşterinin niyetlerini ve ihtiyaçlarını anlamak için müşterilerin işlem verilerini analiz etmelidir (Chiang, 2018: 668). Önemli olan müşterilere ait verilerin bir veri ambarında depolanması değil, bu verilerden işletme için hayati öneme sahip bilgi ve müşteri davranışını belirleyen kalıpların ortaya çıkartılmasıdır.

Müşteri ilişkileri ve turizm bağlamında aşağıdaki araştırma konularını incelenebilir. Tam olarak turizm pazarı için müşteri profilleri nelerdir? Müşterilerin turizm bilgisi ve ürünün kendisi pazarın ihtiyaç ve isteklerine yansıtılıyor mu? Turizm ürün tasarımı ve ürün karması için planlama müşterilerinin bilgisine göre geliştirilebilir mi? Müşteri bilgisi, yeni ürün geliştirme ve müşteri ilişkileri yönetimi için firmanın bilgi varlıklarına dönüştürülebilir mi? Hangi ürünlerin hangi müşterilere ve hangi kanal üzerinden satılacağı, en karlı müşteriler ve artık hedeflemeye değer olmayanların tespiti konusunda pazarlama departmanına destek verebilir. Turizm sektöründe, misafirlerin nereli olduklarını, harcamayı ne kadar, nerede ve ne olarak yaptıklarını bilmek, pazarlama stratejileri hazırlama ve kârı arttırmada yardımcı olabilir (Magnini vd., 2003). Bir bölgeye turistlerin niçin geldiğini tespit ederek reklam kampanyası

düzenlemek ve bir sonraki sezonda turist sayısını arttırmak için yapılması gerekenler saptanabilir.

İşletmelerin başarılı bir MİY uygulaması için gerekli bilgilere ulaşması konusunda veri madenciliği teknikleri kullanılabilir. Veri madenciliği, aşırı bilgi yüklemesini azaltmaya ve karar vermeyi iyileştirmeye yardımcı olabilir. Veri madenciliği uygulamaları, şirketlerin satın alma davranışlarının etkili yordayıcıları olarak hizmet eden farklı değişkenler arasında kalıpları arayarak yüksek kar potansiyeline sahip müşterileri içeren pazar segmentlerini tanımlamasına yardımcı olabilir. Pazarlamacılar daha sonra, hedeflenen bir segmentin satın alma kararlarını artıracak kampanyaları tasarlayabilir ve uygulayabilir. Bu aktiviteyi kolaylaştırmak için pazarlamacılar, veri üreten çıktıları tanımlanmış pazar segmentlerine odaklanan kampanya yönetim yazılımlarında kullanılmaktadırlar (Danubianu ve Hapenciuc, 2008). Veri madenciliği yoluyla çıkarılan müşteri bilgisi, araştırmalardan çıkan ürün ve pazar bilgisi ile entegre edilebilir. Böylece ürün geliştirme, ürün tanıtım ve müşteri ilişkileri yönetimi için bir referans olarak hizmet verebilir (Liao vd., 2010). Çıkarılan bilgiler, verileri tahmin etmek, sınıflandırmak, modellemek ve özetlemek için kullanılabilir.

Başarılı bir MİY; satış, pazarlama ve müşteri hizmetleri gibi müşteri odaklı süreçleri geliştirerek, kârlı müşterilerle uzun vadeli ilişkiler kurarak ve tatmin ederek, şirketin iş başarısını arttırarak (Talón-Ballesterro vd., 2018: 188) işletmeler için yatırımlarda daha yüksek getiri vaat etmektedir (Danubianu ve Hapenciuc, 2008: 3262). Gelişmiş müşteri ilişkileri karlılığı daha uygun müşterileri çekerek, çapraz satış ve müşteriyi daha fazla ürün alması konusundaki ikna faaliyetleri ile kar elde ederek ve müşteriyi elde tutma yoluyla arttırılmaktadır (Danubianu ve Hapenciuc, 2008: 3261). Müşteri ilişkileri yönetimi ile kârlılığı artırmanın üç yolu ile ilgili olarak, veri madenciliği teknikleri aşağıdaki gibi kullanılabilir (Danubianu ve Hapenciuc, 2008):

- Daha uygun müşterileri çekmek için: Veri madenciliği, firmaların hangi müşterilerin belirli ürünleri ve hizmetleri satın alma olasılığının yüksek olduğunu anlamalarına yardımcı olabilir, böylece işletmelerin daha yüksek yanıt oranları ve daha iyi yatırım getirileri için hedeflenmiş pazarlama programları geliştirmelerini sağlar.
- Daha iyi çapraz satış ve müşteriyi teşvik etmek için: İşletmeler, müşteri tarafından arzu edilen ek ürün ve hizmetler sunarak, değer seviyelerini arttırabilir ve satın alma alışkanlıklarını güçlendirir.
- Daha iyi müşteri korumak için: Veri madenciliği teknikleri, hangi müşterilerin kusurlu olduğunu ve neden daha olası olduğunu belirleyebilir. Bir şirket bu bilgileri, bu müşterileri sürdürmelerine izin veren fikirler üretmek için kullanabilir.

Veri madenciliği, ayrıca, ek bazı MİY girişimlerini de destekleyebilmektedir (Danubianu ve Hapenciuc, 2008: 3263). Veri madenciliği, veri tabanı pazarlamacılarının, müşterilerinin hedeflenen ihtiyaçlarına, isteklerine ve tutumlarına daha yakın kampanyalar geliştirmesine yardımcı olmaktadır. Veri tabanındaki daha önceki veriler analiz edilerek geniş bir müşteri aktivitesi modellenenmektedir. Ayrıca veri madenciliği yeni müşterileri keşfedebilmek ve kampanyalarda geri dönüş oranını arttırabilecek çoklu segmentasyon çözümleri sunmaktadır. Veri madenciliği, müşterilerin pazarlama tekliflerine kanala özgü yanıtlarını modelleyerek, kampanya optimizasyon süreçlerinin etkinliğini artırabilmektedir. Bir pazarlama kampanyasının optimizasyonu süreci, bir kuruluşun özellikleri ve kısıtlamalarını karşılayan, kullanılan pazarlama kanallarını tanımlayan ve ilgili zaman parametrelerini belirleyen, bir dizi teklif seti ile bir dizi müşteri arasında bir haritalama oluşturmaktadır.

Tablo 4.1’de MİY konularının araştırılması için veri madenciliği tekniklerini kullanan bazı çalışmalara yer verilmiştir. Aslında sonraki konularda işlenen ürün yönetimi, turist memnuniyeti gibi konular MİY konularıdır, ancak bu bölümde sadece MİY ve MİY sistemleri konusunda araştırma yapanlar listeye alınmıştır.

Tablo 4.1 MİY Konuları İçin Veri Madenciliği Tekniklerinin Kullanıldığı Araştırmalar

Araştırma	MİY Konusu	Kullanılan Teknik	Yapılan İşlem
Ha ve Park, 1998	Segmentasyon	Kümeleme (Self-Organizing Map)	Müşterileri satın alma alışkanlıklarına göre segmentlere ayırmak
Cho ve Leung, 2002	Segmentasyon	Tekniklerin Tanıtımı Uygulama Yok	Veri tabanı pazarlaması
Bloom, 2005	Pazar Segmentasyonu	Yapay Sinir Ağları	Cape Town uluslararası turizm pazarının bölümlendirilmesi
J. Wong vd., 2006	Müşteri profili belirleme	Karar Ağaçları C4.5 Algoritması	Kuzey Tayvan'dan “değerli yolcuları” tanımlamak
Danubianu ve Hapenciuc, 2008	MİY Sistemleri	Tekniklerin Tanıtımı Uygulama Yok	Veri Madenciliği tekniklerinin MİY uygulamaları için kullanımının açıklanması
Liao vd., 2010	Müşteri profili belirleme	Kümeleme Birliktelik Kuralları	Müşteri profili ile ürün segmentasyonu arasındaki ilişkileri tanımlamak
Chiang, 2018	Müşteri Edinimi ve Korunması	FCM algoritma Pazar Sepet Analizi	Kazançlı pazarların ve pazarlama kurallarının keşfi
Talón-Ballesteros vd., 2018	Müşteri profili belirleme	Büyük Veri Bootstrap yeniden örnekleme teknikleri	CRM sistemindeki genel bilgileri kullanarak uluslararası bir otel zincirindeki müşteri profilini belirlemek
Valls vd., 2018	Motivasyon itme-çekme faktörleri	Ontolojik temelli bir kümeleme yöntemi	Motivasyonlar ve anlamlar arasındaki ilişkiyi anlamak

4.2. Ürün Yönetimi

Turizm ile ilgili çok hızla gelişen teknoloji ve değişen müşteri istekleri, işletmeleri ağır rekabet koşulları içinde bırakmaktadır. İşletmeler ayakta kalabilmek için son teknolojiyi kullanmak zorunda ve mümkün olduğu kadar kişiselleştirilmiş ürün ve hizmetlerle piyasaya girmek durumundadırlar. Üstelik bu rekabet ortamında işletmeler, şikâyetini bir tıkla gelecekteki potansiyel müşterilerine duyurabilen ve onların satın alma davranışlarını etkileyen çevrimiçi müşterilerle karşı karşıyadır. Dolayısıyla işletmeler, yeni ürün gelişimlerini, tüm hizmet ve servis olanaklarını tümüyle müşterilerinin beklentileri doğrultusunda oluşturmaktadır. Yeni bir ürünün oluşturulması ve piyasaya sunulması için beklentileri ve memnuniyetleri işletmeler için en önemli başarı kriterleri içinde yer almaktadır. Teknoloji ve etkilediği müşteri isteklerindeki gelişmeler hayli karmaşık olan Yeni Ürün Yönetimi kavramını da beraberinde getirmiş bulunmaktadır (Sağır, 2010). Ürün yönetimi başlangıçta “yeni ürün geliştirme” ile neredeyse aynı kullanılırken, günümüzde geleneksel ürün yönetimi yaklaşımı neredeyse tümüyle müşteri odaklı ve ürünün tüm hayat döngüsünü kapsayan bir yaklaşımına dönüşmüştür. Feşel (2013)’e göre ürün yönetimi bir firmanın örgütsel hayat döngüsü fonksiyonu olup, müşterileri memnun eden en kârlı ürün karmasını oluşturmak için yeni ürün, teknoloji, pazar geliştirmek ve/veya mevcut ürünleri, teknolojileri iyileştirmeye yönelik yönetimsel ve organizasyonel faaliyetleri kapsamaktadır. Burada amaç, firmanın karını en üst düzeye çıkartabilecek ürün karmasının planlaması ve seçilmesini sağlamaktır.

Fikirden üretim aşamasına geçene dek işletmeler ürün yönetimi, pazarlama, satış ve Ar-Ge bölümlerinin tüketici araştırmaları sonuçlarını benimseyerek üretim safhası aşamalarında nihai üretimi hedeflemektedir (Balyemez vd., 2005). Tüketiciler turizm sektöründe diğer sektörlere göre biraz daha belirgin şekilde kendini göstermekte ve işletmelerin finansal performansını etkilemektedirler (Phillips vd., 2015: 131). Turizm sektöründe, misafirlerin nereli olduklarını, harcamayı ne kadar, nerede ve ne alarak yaptıklarını bilmek, pazarlama stratejileri hazırlama ve karı arttırmada yardımcı olabilmektedir (Magnini vd., 2003). Bu bağlamda, pazarlama kararları ve turizm yeni ürün geliştirme stratejik planlaması, destinasyon seçimini, müşteri tercihini / gücünü, ürün özelliklerini, gezilerin türünü ve kısa ve uzun vadede turizm akımlarının tahminini etkileyen müşteri talep ve piyasa arz desenleri, özellikleri, faktörleri hakkında bilgiyi gerektirmektedir.

Turistlerin satın alma davranışı ile ilgili verilere ulaşma konusunda sosyal medya güvenilir bir veri kaynağı durumundadır. Burada yapılan yorum ve değerlendirmeler, tüm sektörün etkin yönetimine ve turizm işletmelerinin rekabet avantajının sağlanmasına katkıda bulunabilmektedir. Kullanıcı tarafından oluşturulan içerikler, şirketler için önemli bir bilgi

kaynağıdır ve bilgilerin analizi, ürünlerin / hizmetlerin kalitesinin iyileştirilmesini, müşteri ihtiyaçlarının tanımlanmasını ve yeni pazarlama stratejilerinin uygulanmasını kolaylaştırabilmektedir (Phillips vd., 2015: 131). Otel servis sağlayıcısı, müşteri hizmetleri ile ilgili genel müşteri eğilimlerini ve sorunlarını tanımlamak için çevrimiçi yorumları periyodik olarak analiz edebilir (Kim ve Yu, 2017: 4442). Mevcut ve potansiyel müşteriler hakkında değerli bilgiler sağlayabilen büyük veriler arasında, KOİ özellikle turizm ve turizm pazarlaması konusunda yüksek bir değere sahiptir, çünkü kullanıcı tarafından oluşturulan veriler, turizm faaliyetlerinin temsilcisi olduğu için, mal / hizmet deneyimini yaşayan müşterinin vermiş olduğu verilere göre, potansiyel müşterinin karar vermesi için daha güvenilirdir (Kim vd., 2017: 363).

Böylece, turizm yeni ürün geliştirme (YÜG) çalışmalarının amacının, seyahat davranışını, turist memnuniyetini ve turizm yönetimini tahmin ve /veya tasarlama ve anlama yeteneğini geliştirmek olduğu ileri sürülebilir. Buna ek olarak, turizm ürünü talep ve geliştirme analizinin önemli bir hedefi özellikle müşteri alım profilleri ve desenlerine karşı kamunun anlayışını düzeltmektir. Bu nedenle, müşterilerin turizm ürünlerini nasıl seçtiklerini bilmek ve seçimlerini belirleyen faktör ve nitelikleri ortaya çıkarmak, sadece turizm talebini anlama da değil, aynı zamanda turistik ürün geliştirmede de yararlıdır.

Turizm politika yapımcıları, perakende iş yöneticileri, mesire yeri yönetim şirketlerinin yöneticileri ve kamu kuruluşları, konaklama alanları ve ulaşım gibi gerekli turizm altyapıları için planlama yapabilmek için turizm aktiviteleri ve turist tercihleri arasındaki ilişkiyi bilmek isterler. Ayrıca, programlama ve kadro oluşturma, tur broşürleri hazırlama ve yatırımlar gibi operasyonel, taktik ve stratejik kararlar almalarına yardımcı detaylı analizlere gereksinim duyarlar (Bose, 2009: 937). Bu analizlerin yapılması için kullanılacak istatistiksel yöntemlerin yetersiz kalması araştırmacıların turizm veri analizi için makine-öğrenme tabanlı (machine-learning-based) veri madenciliği tekniklerine yönelmesine neden olmuştur. Veri madenciliği tekniklerinin kullanılması, şirketlerin imajlarını ve müşterilerin çevrimiçi medyadaki tepkilerini incelemek için etkili bir yöntem olarak piyasa analizi için değerli bir araç olacaktır. Üstelik veri madenciliği teknikleri, genellikle “çevrimiçi sitelerin”, yani “nefret siteleri” olarak adlandırılan çevrimiçi yayınların en yeni konularının içeriğini keşfetmek için yararlı bir yöntem olabilir (Bogicevic vd., 2017: 523).

Veri madenciliği teknikleri turizmin her alanında olduğu gibi yeni ürün geliştirme, planlama, olan ürünlerde iyileştirme ve pazarlama stratejilerinin oluşturulması konusunda vazgeçilmezdirler. Özellikle turistler tarafından oluşturulan içeriklerin analiz edilmesi ve işlet için hayati öneme sahip bilgi ve ilişkilerin tespit edilmesinde rakipsizdirler. Birliktelik kuralları için bir yöntem olarak Apriori algoritmasını ve Tayvan’daki Phoenix Tours International

firmasından müşteri bilgisi çıkarımı için uygulanan veri madenciliği için kümeleme analizini kullanan Liao vd.(2010); veri madenciliği sonuçlarından çıkan bilgiler ışığında firmaya yeni ürün geliştirme ve müşteri ilişkileri yönetimi için öneriler ve çözümler sunmaktadır. Araştırma sonuçları, turistler tarafından dikkate alınan ana faktörlerin güvenlik, seyahat acentesi itibarı, seyahat tarzı ve kabul edilebilir bir fiyat olduğunu göstermiştir. Sonuçlar firmaya, itibarını nasıl koruyacağını bilmesi, daha güvenli seyahat destinasyonları ve bölgeleri seçmesi, canlı (örneğin, uzak ada turu, doğa turu, kaplıca ve SPA turu) yolculuk tarzı tasarlaması ve kabul edilebilir bir yolculuk ücreti olan paketleri yeni ürün geliştirmede dikkate alması gerektiğini önermektedir.

Veri madenciliği teknikleri kullanımı ile belirli hedef müşterilere özel özelleştirilmiş servis teklifleri, promosyonlar ve pazarlama avantajlarından yararlanabilir (Kim ve Yu, 2017: 4442). Al-Salim (2007), hem hizmet sağlayıcının işleyişini ve işlem maliyetlerini en aza indirgeyen seyahat paketlerini tasarlamak, hem de müşteri memnuniyetini en üst düzeye çıkarmak için paketlerin bileşenlerini hazırlayan bir kitlesel özelleştirme stratejisini kullanmaktadır. Seyahat acentelerinin büyük veri tabanlarındaki gizli ilişkiler, seyahat paketlerini müşterilerin gereksinimlerine göre özelleştirmek için birliktelik kuralları tekniği kullanılarak ortaya çıkarılmıştır. Ayrıca, bu modelin sonuçları seyahat acenteleri tarafından sunulan hizmetler için fiyatlandırma mekanizması olarak kullanılabilir. Sonuç olarak, bu araştırma, hizmet sektöründeki kitlesel özelleştirmenin, günümüzde rekabetçi pazarda rekabet etmek için gerekli olan pazarlama ve işletme kararlarını iyileştirmek için kullanılacak güçlü bir taktik olduğunu göstermiştir.

Veri madenciliği tekniklerin yeni ürün oluşturulma esnasında yararlanılabileceği gibi piyasada var olan ürünlerin eksiklerinin giderilmesi veya daha da geliştirilmeleri için kullanılabilir. Memnuniyet araştırmalarından elde edilen veriler bir ürünün müşteriler tarafından olumsuz görülen yerlerinin düzeltilmesinde kullanılabilir (Bogicevic vd., 2017; Vajirachorn ve Chongwatpol, 2017). Hatta bir ürünün deneyimleyenler gözüyle yeniden kategorileştirilmesi ya da fiyatlarının belirlenmesi yapılabilir (Li ve Kwornik, 2016). Bu durumda ürün ve hizmet için daha sadık müşteri grubunun oluşturulması mümkün olabilir. Coğrafi etiketli sosyal medyadan yararlı kültürel miras kaynaklarını tanımlamak (Nguyen vd., 2017) ve kullanıcı bağlamına (ör., yer) göre sıralama yapılabilir. Misafirlerin davranışları izlenerek ilgi çekici yerler ve bölgeler tespit edilerek, daha iyi planlama ve tasarım yapılabilir. Veri madenciliği yoluyla elde edilen müşteri bilgileri, araştırmadan elde edilen ürün ve pazarlama bilgileriyle bütünleştirilebilir ve alt sektördeki perakendecilerin yanı sıra turizm akışı tedarikçilerine de sağlanabilir. Böylece ürün geliştirme, ürün tanıtım ve müşteri ilişkileri yönetimi için referans olarak kullanılabilir (Liao vd., 2010: 4213).

Tablo 4.2’de yeni ürün geliştirme, ürün planlama, ürün değerlendirme ve iyileştirme ve ürün pazarlaması konularının araştırılması için veri madenciliği tekniklerini kullanan bazı çalışmalara yer verilmiştir. Çalışmaların çoğunda ürün oluşturma ve geliştirme önerileri, müşteri davranış ve tercihlerine bakılarak yapılmıştır. Dolayısıyla bazı çalışmalar her iki kategoride değerlendirilebilmektedir.

Tablo 4.2 Ürün Yönetimi İçin Veri Madenciliği Tekniklerinin Kullanıldığı Araştırmalar

Araştırma	Kullanılan Teknik	Yapılan İşlem
Liao vd., 2010	Kümeleme ve Birliktelik Analizi	Yeni ürün geliştirmede dikkate alınması gereken konular hakkında seyahat acentasına öneriler sunulmaktadır.
Versichele vd., 2014	Birliktelik Kuralları	Bu şekilde, bu çalışma, küçük ölçekli hareketlerin katılımcı olmayan ölçümlerinde ad-hoc algılama ağlarının potansiyelini göstererek, mekânsal turizm davranış araştırması alanına bir metodolojik katkı sağlamayı amaçlamaktadır.
Li ve Kwortnik 2016	Kümeleme	Yolcunun algılanan deneyim ile cruise hatları kategorize edildi.
Kim vd., 2017	Duygu Analizi	14 yolcu deneyimi kategorisindeki hizmet performansı ölçülerek kategoriler arasındaki hizmet seviyelerinin algılanması karşılaştırıldı.
Xie vd., 2017	Big data analytics	Otel performansına çevrim içi yorumlara verilen yönetsel cevapların etkisi üzerindeki otel sınıfının düzenleyici etkisi değerlendirildi.
Bogicevic vd., 2017	İçerik Analizi	Temel havayolu kalitesi özelliklerinin belirlenmesi
Guo vd., 2017	LDA Latent Dirichlet Analizi	Otel ziyaretçileri tarafından dile getirilen müşteri hizmetinin anahtar boyutlarının belirlenmesi
Nguyen vd., 2017	Kümeleme ve Naive Bayes Sınıflandırıcı	Kümeleme ile coğrafi etiketli öğeler gruplandırıldı ve Naive bayes ile bunların arasından kültürel mirasa ait olmayan fotoğraflar elendi
Vajirakachorn ve Chongwatpol, 2017	Kümeleme ve 4 Farklı Sınıflandırıcı	Festival alanlarını tekrar ziyaret etme niyetinin tahmin edilmesi
Deng ve Li, 2018	Metin Madenciliği Duygu Analizi Naive Bayes	Kullanıcılar tarafından çekilen fotoğrafların destinasyon tanıtımında kullanılmak üzere seçimi
Alexander vd., 2018	Kümeleme ve Regresyon Ağacı (CART)	Kümeleme ile hanehalkları alt gruplara ayrıldı ve CART ile gruplandırılan geçim türü ile reislik arasındaki ilişki değerlendirildi
Keçeci vd., 2018	Rastgele Orman	Otel için hedef kitle belirlenmiş ve yapılan kampanyaya önceki yıllara göre geri dönüş oranında yüzde 94’lük bir artış elde edilmiştir.
Özdemir ve Orçanlı, 2012	İki Aşamalı Kümeleme Algoritması	Türkiye’de genel bir pazar bölümlenmesi yapılmış, elde edilen bölümlenmelerdeki etkili tüketici profilleri belirlenmiş ve ayrıca mevcut bilgiler çerçevesinde olabilecek niş pazar bölümleri tespit edilmeye çalışılmıştır.

4.3. Turist Memnuniyeti

Turizmin tanıtımı için önemli bir faktör olan turist memnuniyetini, pazarlama ve tüketici davranışı ile ilgili daha önce yapılan çalışmalar, müşterinin beklenti ve gerçek performansa dayalı olarak sunulan bir hizmet veya ürünün öznel değerlendirmesi olarak tanımlamıştır (Guo vd., 2017: 467). Rekabetçi ve dinamik bir ortamda faaliyet gösteren oteller için turist memnuniyetini ölçmek hayatidir. Turist memnuniyeti, turistlerin bir destinasyona seyahat etmelerine ya da geri dönmelerine yönelik seçimlerini etkilediğinden bir destinasyonun başarısı için çok önemlidir (Nuankaew vd., 2017: 90). Çünkü memnun müşteri memnuniyetini yakın tanıdıklarına anlatmakta, internette paylaşarak bunu konuyla ilgilenen herkesin duymasına neden olmaktadır. Memnuniyetsizlik kronik bir hastalık gibidir. Memnuniyetsiz müşteri ömür boyu en kötü reklamdır. Her fırsatta kendi başına gelen olayı ballandıra ballandıra anlatarak diğer insanları uyarmayı bir toplumsal sorumluluk olarak görmektedir.

Memnuniyet boyutları ile ilgili düşük performans, müşteri memnuniyetsizliği yaratabilmekte, ancak yüksek performansları müşteri memnuniyeti yaratmamaktadır (Xu, 2017: 1668). Müşteri memnuniyetinin yüksek düzeyde olmasının, müşteriler arasında otel imajının olumlu algılarını ortaya çıkarmak için değil, aynı zamanda tavsiyelerde bulunma, daha fazla ödeme yapma ve tekrar otelde kalma isteğini de teşvik ettiği bulunmuştur (Liu vd., 2017: 554). Müşteri memnuniyeti müşteri sadakatini etkilemektedir. Müşteri sadakati, otellerin başarısı için önemli bir faktördür. Müşteri sadakati, müşteri memnuniyeti ve kat hizmetleri, resepsiyon, yiyecek-icecek ve fiyat gibi otel hizmetleri ile pozitif yönde ilişkilidir. Bu nedenle, müşteri sadakat düzeyini doğru anlamak, otellerin müşteri sadakatinden yararlanmalarına yardımcı olabilmektedir (Geetha vd., 2017: 44).

Otel müşteri memnuniyetinde anahtar faktörlerin incelenmesi, turizm çalışmalarında uzun süreli bir araştırma konusudur (Liu vd., 2017: 555). Oteller, müşteri memnuniyetini ve davranışsal niyetlerini değerlendirmek ve ele almak için farklı araçlar kullanmaktadırlar. Bu araçlar konuk odalarına yorum kartları yerleştirmeyi, hizmet içi arızaları gidermek için hizmet kurtarma tekniklerinin kullanılmasını, hizmet sonrası misafir memnuniyet anketlerinin dağıtılmasını ve kurum içinde çözülemeyen sorunlara yönelik takip tedbirlerinin alınmasını içerebilmektedir (Berezina vd., 2016: 1). Otel yönetimi ve araştırmacılar tarafından yapılan anket ve görüşmelerde müşteriler, memnuniyet boyutlarını temsil eden en önemli gördükleri ürün özelliklerine ilişkin algılarına dayanarak memnuniyetlerini belirtirler (Xu, 2017: 1668). Yalnız önceki çalışmalar müşterilerin çoğunluğunun aldıkları tatmin edici olmayan hizmet konusunda hareket etmediklerini ve hizmet sağlayıcıya şikâyetle bulunmadıklarını göstermektedir.

Müşteriler deneyimlerini paylaşma ve otellere geri bildirim sağlama konusunda isteksizdirler (Berezina vd., 2016: 1).

Otel içerisinde yapılan anketlere yaşadığı her deneyimi yansıtmayan müşteriler internet ortamında deneyimlerini, otellerle olan memnuniyetini ve memnuniyetsizliğini yansıtmak (Berezina vd., 2016: 2) için yazmakta ve paylaşmaktadırlar. Bunu da, müşterilerin psikolojik ihtiyaçları (örneğin, sosyal kimliğin kazanılması), bir topluluğa ait olma hissi ve şirketlere ve diğer müşterilere yardımcı olmak gibi diğer motivasyon kaynakları çevrimiçi yorum yazma davranışlarını teşvik etmektedir (Cantalops ve Salvi, 2014; Xu, 2017: 1668). İnternet web siteleri, sosyal medya, bloglar ve diğer çevrimiçi platformlar aracılığıyla müşterilerin deneyimlerini paylaşmasını kolaylaştırmaktadır (Berezina vd., 2016: 1). İnternette yayınlanan müşteri yorumlarının bolluğu, müşteri memnuniyeti ve memnuniyetsizlik kaynaklarını incelemek için fırsatlar sunmaktadır. Otel yöneticilerinin, müşterilerini daha iyi anlayabilmek ve otel performansını iyileştirmek için çevrimiçi olarak sunulan müşteri yorumlarını ve değerlendirmelerini kullanmaları önemlidir (Berezina vd., 2016: 3). Otel endüstrisini geliştirmenin bir yolu, derecelendirmeler ve yorumlar yoluyla müşterilerin daha iyi anlaşılmasıdır. Çevrimiçi müşteri puanları, konaklama sektöründe önemli bir rol oynamaktadır. Çevrimiçi otel değerlendirmeleri de müşteri memnuniyeti ile ilgili karşılaştırmalı ve kıyaslamalı bilgiler sunmaktadır (Geetha vd., 2017: 44).

Kullanıcı yorumlarının açık yapısı müşterilerin tüketim deneyimlerinin ve algılarının daha kapsamlı ve doğru bir şekilde yansıtılmasını sağlamaktadır; bu nedenle, akademisyenler kısa süre önce metin incelemelerine daha fazla dikkat çekmişlerdir (Berezina vd., 2016; Xu, 2017: 1663). Son yıllarda, turizm araştırmaları, kullanıcı tarafından oluşturulan yorumların, araştırma yapmak için bir veri kaynağı olarak kullanılmasının artmasına tanık olmuştur (Liu vd., 2017:556). Schuckert ve ark., (2015:143) çevrimiçi otel değerlendirmeleri için “Bu bilgi kaynakları daha objektif, muazzam ve önyargısızdır, çünkü yorumlar geleneksel anketlerden farklı olarak laboratuvar etkilerine sahip olmaksızın kendiliğinden yayınlanmaktadır” şeklinde değerlendirmektedirler. Müşterilerin otel değerlendirmeleri, otel memnuniyetinin sürücülerini anlamak için kullanılabilir bir malzeme kaynağı sağlamakta ve müşterilerin memnuniyetini incelemek için mükemmel bir enstrüman gibi görünmektedir (Hargreaves, 2015; Liu vd., 2017: 556). Ancak, uzun metrajlı, uzun süreli, önemli sayıda ve çevrimiçi metinsel incelemelerin açık yapısı nedeniyle, metin incelemelerinden anahtar noktaların çıkarılması karmaşık ve zorlayıcı olabilmektedir (Xu, 2017: 1663). Çevrimiçi ortam, yöneticilerin tüm bunları gözden geçirip değerlendirmesinin zor olabileceği kadar büyük miktarda veri üretmektedir (Berezina vd., 2016: 3). Büyük hacimli metinsel bilgilerden anlamlı kalıpların

çıkarılmasına olanak tanıyan metin madenciliği yaklaşımı birçok araştırmada kullanılmıştır (Berezina vd., 2016: 3).

Tablo 4.3’de turist memnuniyeti çalışmalarında veri madenciliği tekniklerinin kullanıldığı çalışmaların bir kısmının listesi verilmiştir. Sosyal medya üzerindeki turist yorumlarının artışı ve bu alandaki veri yığını alan çalışmada yeterli veri kaynağını sağlamaktadır. Dolayısıyla bu alandaki çalışmaların artarak devam edeceği öngörülebilir.

Tablo 4.3 Turist Memnuniyeti İçin Veri Madenciliği Tekniklerinin Kullanıldığı Araştırmalar

Araştırma	Kullanılan Teknik	Yapılan İşlem
Liu vd., 2013	Duygu Analizi Birliktelik Kuralları	Çevrimiçi yorumlar olumlu / olumsuz olarak sınıflandırıldı. Seyahat modunda müşteri beklentilerindeki değişiklikleri tespit etmek için bir CTR algoritması kullanıldı.
Hargreaves, 2015	Metin Madenciliği Duygu Analizi	Altı otel özelliği ile ilgili derecelendirmelerle müşteri memnuniyeti ölçülmüştür.
Berezina vd., 2016	Metin Madenciliği	Metin madenciliği tekniği, müşterileri hangi faktörlerin tatmin edebileceğini veya memnuniyetsiz bırakabileceğini ve müşterilerin oteli başkalarına tavsiye etmek veya önermemek gibi niyetlerini nelerin şekillendirdiğini anlamak için kullanılmıştır.
Liu vd., 2017	Regresyon	Müşterilerin dil gruplarına göre ayrımını yapılarak otel müşteri memnuniyetinin belirleyicileri hakkında yeni bilgiler sunma amacıyla kullanıcı tarafından oluşturulan içeriklerin avantajlarından yararlanılmıştır.
Guo vd., 2017	LDA Latent dirichlet allocation	LDA ile otellerin ziyaretçilerle etkileşimlerini yönetmek için kullanabilecekleri 19 kontrol edilebilir boyut ortaya çıkarılmıştır
Nuankaew vd., 2017	Karar Ağacı Birliktelik Kuralları	Turistleri seyahat hedeflerini seçmeye teşvik eden özel ve en önemli faktörlerin belirlenmesi
Shapoval vd., 2018	Karar ağaçları	Özellikle, ankette kullanılan ikili ve sıralı ölçekler nedeniyle, Japonya’da gelen turistlerin davranışsal ve satın alma modellerini özetlemek ve yorumlamak için iki adımlı modelleme ile karar ağaçları (iki bağımlı değişken ile) kullanılmıştır.
Xu, 2017	LSA Latent Semantic Analysis	Çevrimiçi müşteri değerlendirmesi davranışını ve çeşitli seyahat grubu bileşimlerindeki turistlerin otelleri ile genel memnuniyetin belirleyicilerinin araştırılması
Zhao vd., 2019	Büyük Veri	Genel müşteri memnuniyetini tahmin etmek. Çevrimiçi metin incelemeleriyle müşterinin inceleme topluluğuna katılımının teknik özellikleri kullanılarak genel müşteri memnuniyeti öngörülmektedir.
Dirsehan, 2016	Metin Madenciliği	Yabancı turistlerin yorumlarından faydalanarak konakladıkları işletmelere yönelik mikro ve makro düzeyde beğenilerini ve problemlerini metin madenciliği yaklaşımıyla ortaya çıkarmaktır.

4.4. Turist Tercihleri

Turizm endüstrisi birçok ülkede hızla büyümekte ve hükümetler tarafından desteklenmektedir. Yabancı ülkelerden gelen turistleri ve turizm yatırımcılarını çekmek için destinasyonlar ve etkinlikler teşvik edilmektedir. Dolayısıyla turistler, en ilginç yerleri seçerken

önlerinde çok fazla seçenek bulunmaktadır. Bundan dolayı, seyahat acenteleri veya turizm organizasyonları iyi bir planlamaya sahip olmalı ve turistlerin ihtiyaçları için kampanyalar sağlamalıdır. Bunun için de turistlerin davranış kalıplarını ve tercihlerini bilmek zorundadırlar (Yotsawat ve Srivihok, 2014: 1197). Seyahat davranışının doğru ve derinlemesine bir şekilde anlaşılması, turizm endüstrisinin büyük ekonomik faydalarından yararlanmak için hayati önem taşımaktadır. Seyahat davranışını daha iyi anlayarak, turizm uygulayıcıları seyahat edenlerin ihtiyaçlarına cevap vermek için daha uygun iş stratejileri ve seyahat hizmeti / ürünleri formüle edebilir ve bu da iş yatırımında dikkate değer bir getiri sağlayabilir (Vu vd., 2018: 399).

Turist tercihi, genellikle ziyaretçi sayısı, memnuniyet derecesi ve ziyaretçilerin bir destinasyona doğru olan farkındalığı ile değerlendirilir (Sun vd., 2017: 3). Turist tercihi, ziyaretçilerin, gerçek bir ziyaretten sonra hedeflerle ilgili algılarını ve yorumlarını ifade eder, bu nedenle olumlu, olumsuz ve tarafsız olabilir. Turist tercihi, farklı destinasyonların turistleri cezbetme ve turizm kaynaklarını etkin bir şekilde kullanarak turizm pazar paylarını kazanma kabiliyetini yansıtmaktadır (Sun vd., 2017: 2). Turist tercihinin ölçülmesinin nedeni, ziyaretçi memnuniyetinin üretilmesinde çok önemli olduğu için turizm ürünlerinin turist tercihlerine uygunluğunun tespit edilmesidir (Sun vd., 2017: 3).

Çoğu kişi, bir gezi planlarken, doğru seçeneği bulabilmek için planlama zamanının çoğunu harcayarak konaklamaya öncelik vermektedir. Yolcuların hedeflerine, amaçlarına ve seyahat tarzlarına ve önceki konaklama deneyimlerine bağlı olarak farklı beklenti ve / veya tercihleri bulunmaktadır (Liu vd., 2013). Müşteri gereksinimlerinin kapsamlı bir şekilde anlaşılması, otel yöneticilerinin stratejik planlama, pazarlama ve ürün geliştirme açısından pazarda liderlik etmelerine yardımcı olabilir (Li vd., 2017: 311). Bununla birlikte, karmaşık karar verme süreci ve çok çeşitli seçim kriterleri nedeniyle bu gibi önemli bilgileri tespit etmek zordur (Li vd., 2013). Bir turist oda kalitesi ve servis kriterleri için düşük puan verdiği bir odayı, temiz olduğu için seçebilmektedir. Bazı turistler de, sadece hem oda kalitesini hem de hizmet kriterlerini karşılar bir oteli seçebilmektedirler (Li vd., 2013: 322). Müşteri tercihlerini tahmin etmede temel sorun, satın alma kararlarının, marka, kalite, performans, fiyat, özellikler ve kanal dâhil olmak üzere pek çok farklı kriteri eş zamanlı olarak dikkate alması temelinde gerçekleşmesidir. Örneğin, müşteriler, maliyet, hizmet kalitesi, gıda kalitesi, gıda çeşitliliği, mutfak ve ambiyans gibi bir veya daha fazla faktöre dayalı bir restoran seçebilirler. Benzer şekilde, müşteriler, bulunduğu yere (havaalanına, turistik yere veya şehir merkezine yakın), marka ismine, çeşitli tesislere (yüzme havuzu, golf sahası ve spa ve fitness merkezi gibi), hizmet kalitesine, fiyatına, sadakat programına ve eski misafirlerin kalite derecelendirmelerine göre bir otel seçebilir (Verma, 2010: 470). Bu ölçütlerin, müşterinin karar verme niyetlerini

yönlendirmede nasıl etkileşimde bulunduğunun anlaşılması, yöneticilere müşteri tercihleri hakkında bilgi verebilir (Li vd., 2013: 322).

Turizmin gelişmesi, yerel ekonominin refahını arttırmak ve sosyal refahı iyileştirmek için en fazla değeri sağlayabileceğinden, hem kasıtlı hem de kasıtsız olarak turistler arasındaki popülerliğini arttırmayı amaçlayan, aynı bölgede bulunan farklı turizm destinasyonları arasındaki rekabet artmaktadır. Bu nedenle, turist tercihinin kazanmak rekabetin nihai amacıdır. Bununla birlikte, turizm bölgeleri arasındaki kısır rekabet olgusu turistlerin gerçek taleplerinin göz ardı edilmesine neden olmaktadır. Ne yazık ki, rakiplerini rekabet avantajından kurtarmak için her zaman bir başka destinasyonun etkili yeniliklerini taklit edip kopyalamaktadırlar. Bu da ziyaretçi sayısında ciddi bir düşüşe neden olmaktadır. Mümkün olduğunca çok sayıda turisti barındırabilmek için, Çin'in Dapeng yarımadasındaki bazı doğal noktalar, turistlerin gerçek talebini göz önünde bulundurmadan, sadece ekolojik dengesizliğe ve kaynak atıklarına yol açmakla kalmayıp, aynı zamanda turistik mekanların genel manzarasını da baltalayarak, yüksek yoğunluklu otel ve restoran inşaatı yapmaktadırlar (Sun vd., 2017: 1). Sonuç olarak, yapılanlar turizm piyasası talebine uymamakta ve birçok doğal nokta iflas etmektedir. Bu nedenle, turizmin gelişmesini turizmin tercihinin göre bölgesel bağlamda teşvik etmek bir zorunluluktur (Sun vd., 2017: 2).

Günümüzde, turist tercihi ile ilgili mevcut araştırmaların büyük çoğunluğu, analitik hiyerarşi süreci (AHP), önem performans analizi (IPA), uzman puanlama ve anket anketleri gibi küçük grup fikrine dayalı yöntemleri benimseyerek turist tercihinin değerlendirilmiştir. Her bir model niteliğinin puanlarını aramak için kullanılan temel veriler, genellikle dosyalanmış anketler ve kısmi turistler ve ilgili paydaşlarla yapılan görüşmelerle toplanmaktadır. Bu yöntemlerin turist tercihinin tanımlamak ve anlamlı araştırma sonuçları elde etmek için bir çözüm sağladığı inkâr edilemez. Bununla birlikte, bu yöntemler tüm etkili faktörleri kapsayacak kadar objektif değildir, çünkü veri tabanının ölçeği sınırlıdır (Sun vd., 2017: 3). Ayrıca müşteri tercihlerini anlamak için, şirketler genellikle müşterilerinin bireysel özelliklerine odaklanmaktadır. Bununla birlikte, tipik olarak kullanıcılar izole aktörler değildir, ancak birbirleriyle ilişkilidir (Neidhardt vd., 2017: 117). Sosyologlar ve piyasa araştırmacıları, bir bireyin hayatının, bireyin sosyal bağlantılar ağı ile nasıl etkileşime girdiğine bağlı olduğuna inanmaktadır (Bell ve Mgbemena, 2018: 197). Dolayısıyla, müşteri tercihlerini anlamak için, bireysel seviyenin ötesine geçmek ve ağ seviyesini göz önünde bulundurmak mantıklıdır. Bu, Neidhardt vd., (2017) sonuçları tarafından da desteklenmektedir. Bireyi ve ağ seviyesini birleştirmek, daha karmaşık bir kullanıcı modelinin oluşturulmasına yardımcı olur; bu da bir işletmenin hitap etmek istediği toplulukları daha iyi anlamak için kullanılabilir. Dahası, kazanılan içgörüler daha

iyi ürünler tasarlamaya veya belirli ürünleri belirli kullanıcı gruplarına önermeye yardımcı olabilir (Neidhardt vd., 2017: 117).

Otel yöneticileri, stratejik planlama, pazarlama ve ürün geliştirmelerini iyileştirmek amacıyla turist tercihlerini anlama yollarını bulmaya devam etmektedirler. Sorun şu ki mevcut çalışmaların çoğu gelişmekte olanlara değil mevcut talep özelliklerine odaklanmaktadır. Böylece, otel yöneticileri, yolcuların endişelerindeki değişikliklere karşılık olarak uygun kararlar vermede zorlanmaktadır (Liv d., 2017: 311). Yeni ürün ve hizmetlerin hızla çoğalması, işletmelerin müşterilerinin neyi arzuladığını ve neyi ödemeye istekli olduğunu anlamasını zorlaştırmaktadır. Konaklama sektöründeki potansiyel piyasa teklifleri, müşterilerin farklı alternatiflerin göreceli maliyetlerini ve faydalarını karşılaştırmasını ve stratejik olarak değerlendirmesini sağlayan enformasyon teknolojisindeki ilerlemelere büyük ölçüde bağlı olarak giderek karmaşıklaşmıştır. Örneğin, otel müşterileri rekabetçi teklifleri Expedia, Orbitz ve Travelocity gibi çevrimiçi rezervasyon kanallarında kolayca karşılaştırabilirler. Ayrıca geçmiş müşterilerin yorumlarını ve önerilerini Tripadvisor.com gibi sosyal medya sitelerinde okuyabilmektedirler (Verma, 2010: 470). Bunlar da müşterilerin tercihlerini etkilemektedir. Tercihlerdeki değişiklikler otel işletmelerinin performansını etkileyerek, misafirlerinin taleplerini tespit etme ve bunlara cevap verme ihtiyacını doğurmaktadır (Liv d., 2017: 311).

E-turizm, farklı yerlere yönelik turist tercihini ölçmek için sosyal medyada büyük miktarda veriyi kullanmayı mümkün kılmaktadır. E-turizmi, turistlerin web sitelerinde seyahat planlarını düzenlemek, turizm destinasyonlarını seçmek, otel ve araç rezervasyonu yapmak ve seyahat rotalarını düzenlemek gibi etkili bir kanal sunmaktadır. E-turizmin gelişiyile, dünya sosyal medya verilerinin patlamasını yaşamıştır. Günümüzde, sosyal medya verileri sadece turistlere talep doğrultusunda çeşitli seyahat bilgileri sağlamakla kalmamakta, aynı zamanda turist destinasyonlarına göre karar vermek için turizm destinasyonlarına da ilham vermektedir (Sun vd., 2017: 2). Bu artan pazar şeffaflığı, bu iş ortamında çalışacak hizmet şirketleri için hem fırsatlar hem de riskler yaratmaktadır. Bu nedenle, hizmet sağlayıcılar bilgili ve sofistike tüketicilerle yüz yüze geldikçe, tüketicinin seçimleriyle ilişkili satışların anlaşılmasını sağlamak için uygun yönetim kararlarının alınabilmesi aciliyet arz etmektedir (Verma, 2010: 470).

Turizm politika yapıcıları, perakende iş yöneticileri, mesire yeri yönetim şirketlerinin yöneticileri ve kamu kuruluşları, konaklama alanları ve ulaşım gibi gerekli turizm altyapıları için planlama yapabilmek için turizm aktiviteleri ve turist tercihleri arasındaki ilişkiyi bilmek istemektedirler. Ayrıca, programlama ve kadro oluşturma, tur broşürleri hazırlama ve yatırımlar gibi operasyonel, taktik ve stratejik kararlar almalarına yardımcı detaylı analizlere gereksinim duymaktadır (Bose, 2009:937). Bu analizlerin yapılması için kullanılacak istatistiksel

yöntemlerin yetersiz kalması araştırmacıların turizm veri analizi için makine-öğrenme tabanlı veri madenciliği tekniklerine yönelmesine neden olmuştur.

Sosyal medya verileri turistlerin tercihlerini yansıtırma konusunda, veri madenciliği de bu verilerden anlamlı ilişkiler çıkartılması konusunda araştırmacılara ve işletmelere fırsatlar sunmaktadır. Son zamanlarda sosyal medya verilerinden turist tercihi konusundaki çalışmalarda gözle görülür bir artış kaydedilmiştir. Bu çalışmalardan bir kısmı Tablo 4.4’de verilmiştir.

Tablo 4.4 Turist Tercihleri İçin Veri Madenciliği Tekniklerinin Kullanıldığı Araştırmalar

Araştırma	Kullanılan Teknik	Yapılan İşlem
J. Wong vd., 2006	Karar Ağaçları	Demografik bilgiler, satın alma ve karar verme davranış kalıpları ve ziyaret edilen destinasyonlar gibi değerli turistlerin turistik özelliklerini keşfetmek için karar ağacı uygulandı.
Law vd., 2011	Birliktelik Kuralları	Hong Kong'dan giden yolcuların davranış kalıplarını daha iyi anlama
Karl vd., 2015	Kümeleme ve Sınıflandırma	Turistlerin destinasyon seçim süreçleri deneysel olarak araştırılmıştır
Li vd., 2016	Mekan Dizimi Analizi	Sokak ağı entegrasyonu ile kentsel yapı arasındaki ilişki ve veri madenciliğinden elde edilen turist tercihleri arasındaki ilişkiler araştırılmış ve doğrulanmıştır.
Li vd., 2013	Choquet Integral	Choquet Integral kullanarak Hong Kong gelen yolcuların otel seçimi tercihlerini keşfetmek
Shi vd., 2017	Duygu Analizi	Sina Weibo coğrafi etiketli veriler kullanılarak şehirdeki turistik mekanların popülerliği incelendi.
Sun vd., 2017	Duygu Analizi	Dapeng Yarımadası'nda yakın bir yerde bulunan 34 doğal noktaya yönelik nispi turist tercihi hesaplanmıştır.
Valls vd., 2018	Kümeleme	İngiliz turistlerin Katalan bölgesinde belirli bir destinasyona gitmeye karar verme nedenleri incelenmektedir
Zhang vd., 2018	Mekansal Kümeleme	Coğrafi etiketli fotoğraflar kullanılarak popüler doğal noktaların tespit edilmesi

Turist tercihleri konusunda yapılan çalışmalar genelde turist memnuniyeti ya da turist davranışı ile birlikte yapılmaktadır. Özellikle sosyal medya üzerindeki yorumlar ve coğrafi etiketli veriler araştırmacıların aynı turist memnuniyetine, tercihinine ve davranışı konularında araştırma yapmalarına izin vermektedir.

4.5. Turist Davranışı

Bir turizm bölgesi, “çok çeşitli turist ihtiyaçlarını ve beklentilerini karşılayabilecek büyüklükteki” doğal ortamların veya kültürel etkinliklerin sergilendiği birçok turistik mekândan oluşmaktadır. Bu mekânlar, bazı bölgelerde kümelenmiş bir şekilde bazı bölgelerde dağınık bir şekilde (Shao vd., 2017: 66) karşımıza çıkabilmektedirler. Turizm yöneticileri de bu bölgelerde cazibe merkezlerini ve çevrelerini geliştirerek küresel pazarda turist çekmek için birbirleri ile rekabet etmektedir (Shi vd., 2017: 1188). Bu bölgelere gelen turistler, gezi, anıtlar,

rekreasyon, müzeler veya kültürel etkinliklerden alışveriş, yemek ve diğer insanlarla etkileşime kadar bir dizi deneyimden yararlanmaktadır.

Turistlere deneyimler sağlayan bir turizm bölgesi turizm uygulamaları ve ilgili altyapılardan oluşmaktadır. Altyapılar ulaşım, konaklama, yemek, eğlence ve benzeri dâhil olmak üzere turistlere çeşitli hizmetler sunmaktadır. Bir bölgenin fiziksel (örneğin, iklim, kültür, bitki örtüsü veya manzara) veya kültürel (tiyatro gösterisi, müzeler veya festivaller gibi) özellikleri turistlerin başlıca ziyaret nedenidir (Shao vd., 2017: 66) Özellikle çevrimiçi olarak bir turizm bölgesini araştıran potansiyel turistler genellikle konaklayacakları oteller, ziyaret edilecek yerler ve otelleri ve turistik yerleri birbirine bağlayan yollar hakkında araştırma yapmaktadırlar (Zhou vd., 2017: 1). Ek olarak potansiyel bir turistin, seyahatle ilgili zaman ve maliyeti azaltmak için tek bir seyahatte birden fazla destinasyona gitme olasılığı yüksektir (Zheng vd., 2017: 268). Bu da bir bölgeye gelen bir turistin aynı bölge içinde ya da yakın bölgeler arasında seyahatler yapacağı anlamına gelmektedir.

Bölge içinde ya da bölgeler arasında yapılan seyahatler turist hareketliliğine neden olmaktadır. Bildiğimiz gibi, turistlerin tüm seyahat süreci genellikle birkaç gün sürmektedir. Çoğu durumda, her günün güzergahı, ara düğümleri genellikle temel ve eğlence ihtiyaçlarını karşıladıkları restoranlar, turistik yerler, benzin istasyonları, dükkanlar, vb. yerler olmakla birlikte ikamet ettikleri yerden başlamakta ve yine aynı yerde bitmektedir (Shao vd., 2017: 66). Turistik mekanların mekânsal dağılımı (örneğin kümelenmiş, dağılmış), turistlerin destinasyon içinde geniş çaplı veya dar hareket edip etmediklerini etkilemektedir (Hu vd., 2018: 6).

Haldrup'a göre (2004: 434) “turist hareketliliği çoğu zaman, araştırılması ve açıklanması gereken bir fenomen değil, belirli turizm ve turizm davranış biçimlerinin karakterini açıklayan bir kara kutuya dönüşmüştür”. Hareket, hem bireyin yaşadığı ortamından belirli bir destinasyona ulaşmak hem de o turist destinasyonunda dolaşmak için turizmin kilit bir yönünü temsil etmektedir (Versichele vd., 2014: 67). Turizm tedarikçilerinin çok istedikleri bölge içindeki hareketlilik iyi analiz edilmediği ve alt yapıların iyi planlanmadığı durumlarda, özellikle ziyaretçi sayılarının artması ile birlikte, trafik, gürültü ve belli bölgelerde kalabalıklaşmaya neden olmaktadır. Bir turizm destinasyonunda aşırı kalabalıkların yol açtığı tıkanıklık olumsuz bir faktör olarak kabul edilmektedir (Shi vd., 2017: 1187). Turistler kalabalık ortamlarda ya da ortamın kalabalık olma ihtimalinde, en çok tercih ettikleri yerlerde ya da zamanlarda olsa bile yer değiştirmektedirler (Shi vd., 2017: 1188). Dahası, zaman ve ulaşım gibi kısıtlamalar ve destinasyona ait bilgi eksikliği turistlerin günlük tur programlarını da etkileyebilmektedir (Shao vd., 2017: 66).

Turistlerin rahatsız olmasına neden olan taşıma kapasitesi sorununu çözebilmek için o bölgedeki ziyaretçilerin mekânsal ve zamansal dağılımını bilmek gereklidir (Orellana vd., 2012: 672). Turizm organizasyonları arasındaki akışlardaki darboğazları ve gereksiz engelleri tespit etmek turizm yöneticileri için değerlidir (Prideaux 2000; Vu vd., 2018). Seyahat rotalarının ayrıntılı kayıtlarına erişim, ulaşım altyapısında aşırı yüklenmeyi engelleyen politikaların tasarlanmasını sağlamakta ve turist destinasyonları arasındaki seyahat engellerini gidermektedir (Prideaux, 2000; Chua vd., 2016).

Turistlerin bir bölgedeki akışını izlemek ve analiz etmek turist davranışını anlamak için de gereklidir. Turistlerin mekânsal davranışlarının en önemli yönlerinden biri, bölgedeki hareketleridir. Bir destinasyon içindeki hareketlerin, destinasyon yönetim aktiviteleri paketine doğrudan uygulanabilecek şekilde anlaşılması önemlidir (McKercher ve Lau, 2008). Turist hareketliliğini anlamak, destinasyonların idaresi ve tasarımı, yerinde hareketin planlanması ve turistik mekanların pazarlanmasında önemli bir rol oynamaktadır (Zheng vd., 2017: 267). Bir bölgeye yaptıkları ziyaretler sırasında insanların hareketlerinin izlenmesi, hangi yerleri en çok ya da en az ziyaret ettiklerini, her bir yerde ne kadar zaman geçirdiklerini ve farklı hedef grupların tercih ettiği yerleri belirlemede yardımcı olabilmektedir. Turistlerin hareketlerini izleme ve analiz etme ve alan kullanımı, farklı gruplar arasındaki potansiyel kalabalıklaşma ve çatışmalar hakkında bilgi sağlayabilir (Orellana vd., 2012: 672). Benzer şekilde, farklı turistik demografik değişkenlerin tercihi göre konumlar iyileştirilebilir veya geliştirilebilir (Lew ve McKercher, 2006; Chua vd., 2016). Turizm sektörünün planlama taleplerini karşılamak için turist akışını izlemek ve analiz etmek gittikçe önem kazanmaktadır.

Seyahat davranışını anlamak, piyasayı tahmin etmeyi veya uygun seyahat önerileri sunmayı amaçlayan turizm uygulayıcıları için önemlidir. Seyahat davranışını daha iyi anlayarak, turizm uygulayıcıları seyahat edenlerin ihtiyaçlarına cevap vermek için daha uygun iş stratejileri ve seyahat hizmeti / ürünleri formüle edebilir ve bu da iş yatırımında dikkate değer bir getiri sağlamaktadır (Vu vd., 2018: 399). Buna rağmen turist hareketliliğine yönelik ampirik çalışmalar doğrudan gözlem ya da kişisel görüşmeler gibi geleneksel yöntemlerin emek yoğun ve çoğu zaman pahalı olmasından dolayı oldukça azdır (Versichele vd., 2014:67). Bu konuyla ilgili çalışmalar Campbell (1967) ile başlamasına (akt: Zheng vd., 2017: 268) rağmen hala emekleme aşamasındadır (Shi vd., 2017: 1188). İlerlemenin yetersiz olmasının bir nedeni, veri edinimi ile ilgili zorluklardan (Shao vd., 2017:66) kaynaklanmaktadır çünkü turizm, genellikle kentsel ve kırsal alanlar arasındaki seyahatleri içermektedir. Operasyonel bir perspektiften bakıldığında, mekânsal - zamansal hassasiyet ve maliyetle ilgili özel zorluklar ortaya çıkarılmaktadır (Chua vd., 2016: 295).

Davranış modeli keşif süreci, geleneksel olarak anketler, anketler ve kamuoyu anketleri (surveys, questionnaires and opinion polls) kullanılarak toplanan farklı veri türleri tarafından desteklenmekteydi (Vu vd.,2015: 224). Ancak geleneksel veri toplama yöntemleri genellikle turistlerle doğrudan iletişime ihtiyaç duyduğundan, toplanan verilerin cevap sayısı az olmakta ve kapsam dâhilindeki coğrafi alan ile sınırlı kalmaktadır (Vu vd., 2018: 400). Ayrıca telefon anketleri veya turistik mekânsal-zamansal davranışlar üzerinde yapılan saha incelemeleri gibi geleneksel veri toplama yöntemlerinin de zaman ve enerji tüketimi gibi kısıtlamalar (Shao vd.,2017: 67) ve maliyet, ölçeklenebilirlik, örnekleme gibi bu yöntemlerin etkinliğini azaltan engeller (Shi vd., 2017: 1188) vardır. Oysa bu alanda yapılacak çalışmalar turistlerin mekansal etkinliklerinin hem ayrıntılı hem de bol miktarda olmasını gerektirmektedir (Shao vd., 2017:67).

Son dönem çalışmaları geleneksel yöntemlerin maliyet (Shi vd., 2017: 186) ve verilerin güvenilirlik (Versichele vd., 2014: 67) sorunlarını aşmak için yeni kaynaklar önerilmektedir. Dijital çağda konum duyarlı cihazların popülerliği ve sosyal ağ hizmetinin geliştirilmesi, kullanıcı merkezli bireysel mekânsal davranışlara ve bunların içerik bilgilerine (kronolojik, hassas ve muazzam) erişmeyi mümkün kılmıştır (Shao vd., 2017: 66). Geleneksel olarak veri toplama yöntemlerine ek olarak günümüzde web, sosyal medya, GPS, bluetooth, cep telefonu ve uygulamaları yeni veri kaynakları olarak karşımıza çıkmaktadır. Buralardan tüketim kayıtları, coğrafi etiketli veriler hacimli ama ucuz ve kolay elde edilebilmektedir (Shao vd., 2017:67). Özellikle sosyal medya, seyahatle ilgili bilgi almak için gittikçe artan sanal topluluklara, kişisel bloglara ve Flickr, Twitter ve Instagram gibi ağlara gittikçe artan bir şekilde, turizm endüstrisi üzerinde güçlü bir etki yaratmıştır (Xiang ve Gretzel, 2010). Kullanıcılar tarafından oluşturulan bu veriler, turistlerin hareketlerini incelemek ve seyahat tercihlerini anlamak için değerli bir bilgi kaynağı olarak hizmet etmektedir (Chua vd., 2016: 296).

Kullanıcıların ilgi alanlarını ve etkinlikleri paylaşımlarına ve dolayısıyla yeni ya da daha yakın sosyal ilişkiler kurmasına ilham veren çevrimiçi hizmetler, araştırmacıların çok miktarda veriden oluşan sosyal medya kullanıcılarının daha doğru bir resmini çekmelerini sağladığı için tutulmaktadır. Toplumsal olarak üretilen metinsel ve mekânsal-zamansal meta verilerle birlikte, bu zenginleştirilmiş multimedya veri noktaları, daha fazla bilgi birikimine ve toplumsal kalıpları anlamağa yol açtıkları için pek çok araştırmayı teşvik etmişlerdir(Shi vd., 2017:-1189). Günümüzde bir çok araştırmacı turist altyapısının optimizasyonu, ürünlerin pazarlanması ve turizm hareketliliğinin çevre üzerindeki etkilerinin daha iyi yönetilmesi için turistlerin mekânsal-zamansal davranışına odaklanmıştır (Versichele vd., 2014: 67). Son birkaç yılda, akademisyenler ve uygulamacılar, turistlerin hareket modellerini ve bu modeller

üzerindeki uygulamalara nasıl rehberlik edeceklerini, örneğin, pazar bölümlenmesini (Xia vd., 2010) ve davranış analizini kullanarak giderek daha fazla incelemektedirler (Zheng vd., 2017:267).

Fuchs vd. (2014), destinasyon yönetiminin yanı sıra turizm sektörünün iş zekasına da yardımcı olan gerçek zamanlı turist davranışı hakkında daha fazla bilgi edinmek için büyük veri madenciliğini kullanarak bir destinasyona yardımcı olan bir model sunmaktadır. Bu alanda yapılan çalışmaların bir kısmı da Tablo 4.5’de verilmiştir.

Tablo 4.5 Turist Davranışı için Veri Madenciliği Tekniklerinin Kullanıldığı Araştırmalar

Araştırma	Kullanılan Teknik	Yapılan İşlem
Shmueli vd., 1996	Yapay Sinir Ağları	İsrail'deki kadın ve erkeklerin seyahat talebi modellerinin karşılaştırılması
Xia vd., 2010	EM algoritması Beklenti - Maksimizasyon	EM algoritması Phillip'e giden turistlerin pazar bölümlendirmesini (sosyo-demografik ve seyahat davranış değişkenleri açısından) sağlamak için kullanıldı
Orellana vd., 2012	Sıralı Örüntü Madenciliği	Yerlerin ziyaret edildiği bağıl zamansal sırayı gösteren GSP'(Genelleştirilmiş Sıralı Desenler)leri çıkarmak için bir veri madenciliği algoritması kullanıldı
Vu vd., 2015	Yoğunluk kümeleme Markov Zinciri	Coğrafi etiketli fotoğraflara dayalı olarak turistlerin ilgi çekici yerleri, seyahat düzenleri ve günlük etkinlikleri keşfedildi
Belcastro vd., 2016	Yoğunluk kümeleme	Sosyal medyadan toplanan coğrafi etiketli verilerden turistler tarafından ziyaret edilen ilginç yerlerin keşfi
Zheng vd., 2017	HPA Heuristic Prediction Algorithm	Odaklanmış bir turist yörüngesini ve geçmişteki ziyaretçilerin hareketlerini dikkate alan sezgisel bir yöntem ile Pekin, Çin Yaz Sarayı'nda bir vaka çalışması yapıldı.
Shao vd., 2017	Yoğunluk kümeleme	Sosyal medya verilerine dayalı olarak kent turizm bölgelerinin çıkarılması ve analizi
Shi vd., 2017	Duygu Analizi	Şanghai'da şehir turizmi yoğunlaşmasını, coğrafi kaynaklardan elde edilen kitle kaynaklarını kullanarak keşfetmek
Brandt vd., 2017	LDA	Sosyal medya analitiğinin, şehir içinde kullanıcıların varlığına ve çevresel ve topikal katılımına ilişkin mekansal kalıpları yakaladığını gösteriyoruz.
Brunk vd., 2018	Süreç madenciliği algoritması	Kayak merkezindeki kayakçıların hangi asansörleri kullanacaklarının tahmini yapıldı.
Cai vd., 2018	OPTICS algorithm	Anlamsal açıdan zenginleştirilmiş yörüngeleri ele almak ve anlamsal yörüngeler için yeni bir benzerlik ölçütü sunmak için genişletilmiş bir OPTICS kümeleme algoritması kullanılmıştır.
Dietz vd., 2018	Sezgisel Algoritmalar	Öneri sisteminin arka planında kullanılmak üzere check-in verilerinden seyahat düzenleri öğrenilmiştir.
Gal-Tzur vd., 2018	Sınıflandırma Algoritmaları	Soru-Cevap forumlarında yayınlanan ulaşım ile ilgili soruları otomatik olarak kategorize etmek ve seyahat ile ilgili bilgilendirme isteyen soruları çıkarmak. Ulaşım ile ilgili sosyal medya içeriğini analiz etmek için metin madenciliği teknikleri uygulanmıştır.
Gao vd., 2018	Kümeleme	Turistik toplanma alanlarının tespit edilebilmesi için bulanık seviye ağırlık endeksini MST modeli ile birleştiren bir model (F-AMST) geliştirmiştir.
Grinberger ve Shoval, 2019	Kümeleme	Turistlerin İsrail'in Kudüs ve Tel Aviv kentlerinde gözlemlenen ziyaret kalıplarının arkasındaki faktörlerin analizi yapılmıştır.

Hu vd., 2018	Yoğunluk kümeleme	Turist cazibe merkezlerinden DBSCAN algoritması ile turist grafiği oluşturulmuş ve cazibe merkezleri ve en popüler tur güzegahları dahil olmak üzere turist hareketi modellerini tespit etmek için ağ analizi algoritmaları (Markov kümeleme algoritmaları) uygulanmıştır
Majewska ve Truskolaski, 2018	Bulanık küme sınıflandırma	Turizmde, mekansal yığılma bölgesel birimlerin sınırlarının ötesine yayılmaktadır. Bölgeyi daha kesin olarak ölçmek için turizm işletmelerinin küresel konumlandırma sistemi (GPS) koordinatları ve coğrafi ağırlıklı bulanık kümeleme yaklaşımı (FGWC) kullanılmıştır.
Mor ve Dalyot, 2018	Sınıflandırma	Popüler bölgeleri ve ilgi çekici yerleri bulmak için fotoğraflarda hücre tabanlı kümeleme, fotoğrafçılar arasında en popüler turistik yerleri ziyaret eden rotaları hesaplamak için en Yakın Komşu rota hesaplama algoritması kullanılmıştır.
Scherrer vd., 2018	Kümeleme	Mobil navigasyon uygulamasından toplanan, insan hareketleri sonucunda oluşan veriye dayanarak farklı hareketli popülasyon alt grupları tanımlandı.
Vu vd., 2018	Sıralı Kural Madenciliği	Turistlerin davranışları coğrafi etiketli fotoğraflar ve sıralı kural madenciliği yoluyla analiz edildi.
Z. Wang vd., 2018	Kümeleme Kural tabanlı algoritmalar	Cep telefonu verilerini kullanan mevcut seyahat davranış çalışmalarını gözden geçirmekte ve bugüne kadar kaydedilen ilerlemeyi sunmakta, daha sonra seyahat davranışı araştırmasını ilerletmek için cep telefonu verilerinin potansiyelini tartışmakta ve bazı zorlukları gündeme getirmektedir.
Zee vd., 2018	Kümeleme	KOI'nin mekansal analizinin turist davranışları hakkında bilgi edinmenin ve destinasyon yönetimi ile ilgili DMO'lar arası tartışmalar için bir girdi görevi görmesinin bir aracı olarak uygulanabilirliği test edildi.
Nagao vd., 2004	Log Madenciliği	Turizmde toplanan araçlardan elde edilen GPS log verisinden kişisel turizm aktivite bilgileri çıkartılmıştır.

4.6. Öneri Sistemleri

Turistler, seyahatleri planlarken ya da konaklama gibi turizm ürününden memnuniyetlerini ya da şikâyetlerini bildirirken ya da deneyimlerini paylaşırken, destinasyonlar hakkında sorular sormak ve yapılan yorumları görmek için genellikle çevrimiçi hizmetleri ve sosyal medyayı kullanmaktadırlar. Web'deki verilerin katlanarak artması nedeniyle oluşan “enformasyon patlaması” ya da “aşırı enformasyon yüklemesi”, değerli bilgilerin bulunmasının çok zor olduğu pek çok işe yaramaz veriyle karşılaşmaya neden olmaktadır. Bu da e-ticaret sitelerinde satışa hazır birçok ürün olmasından kaynaklanmaktadır. Bunun sonucunda, katalogun tamamında istenen ürünleri aramak kullanıcılar için sıkıcı bir iş haline gelmekte ve muhtemelen ilgilendikleri ürünleri bulamamaktadırlar. Seçici ve kişiselleştirilmiş ürünlerin tanıtımını yapan mekanizmalara sahip e-ticaret sistemlerine sahip olma ihtiyacı, tüketicilerin ürün bulma ve satın

almalarına yardımcı olan “öneri sistemleri”ni ortaya çıkarmıştır (Lucas vd., 2013: 1). Bu sistemler çoğunlukla e-ticaret ortamında kullanılmasına rağmen, sanal kütüphaneler, haber siteleri, bilimsel portallar, e-öğrenme sistemleri vb. gibi diğer alanlarda da kullanılmaktadırlar.

Seyahat edenlerin istedikleri destinasyonlarda hangi turizm merkezlerini ziyaret etmeleri gerektiği konusunda uzmanlara, yerlilere veya arkadaşlara danışmaları genellikle önemlidir (örneğin, nereye gidecekleri, nerede kalacakları ve oraya nasıl gidileceği, gümrük ve göçmenlik kuralları, uyarılar ve diğerleri). İnsanlar çevrimiçi sosyal ağlarda muazzam miktarda seyahat bilgisine erişebilmektedir, ancak en alakalı bilgiyi bulmak ve / veya tanımlamak zordur. Seçim sürecinde ilerlemenin olmaması, kullanıcılar için yaygın bir problemdir, çünkü büyük miktarda bilgi ile karşılaşacaklardır, bu yüzden en sevdikleri öğeyi bulmaktan sıkıntı çekeceklerdir. Bu nedenle bir öneri sistemi (ÖS) bu problem için kilit bir teknolojidir. Bir ÖS, kullanıcı tercihlerini analiz eden ve işlevlerini bireysel kullanıcılara uyarlayan bir enformasyon filtreleme sistemidir ve kullanıcı çıkarlarını veya tercihlerini bulma bu nedenle bir ÖS'nin kilit bir işlemidir (Kesorn vd., 2018: 26703). Öneri sistemleri, akıllı özelliklerinden dolayı müşteri davranışlarını analiz ederek kişisel tavsiye oluştururlar. Öneri işlemlerinin yapılabilmesi için bazı yöntemler geliştirilmiştir. İçerik tabanlı, işbirlikçi filtreleme ve hibrid çözüm gibi yöntemler, genel hizmet ve ürün önerileri için yaygın olarak kullanılmaktadır (Zhu vd., 2017: 17596).

Öneri sistemlerinde kullanılan en popüler teknikler, içerik tabanlı filtreleme ve işbirlikçi filtreleme'dir. Bu tür sistemlerde her kullanıcı, tipik olarak kullanıcının derecelendirdiği (veya satın aldığı) öğeleri içeren bir kullanıcı profili ile temsil edilmektedir. İçerik tabanlı filtrelemede, kullanıcının profilinde zaten bulunan öğelere benzerliklerine göre yeni öğeler önerilir. Bunu başarmak için her bir öğeye ilişkin daha fazla ayrıntıya ihtiyaç vardır. Öte yandan, işbirlikçi filtreleme yaklaşımları öğelere “agnostik” dir. Bunun yerine, kullanıcı tarafından atanan dereceleri, derecelendirme modellerini temel alan diğer benzer kullanıcıları (veya öğeleri) bulmak için kullanırlar. İşbirlikçi filtreleme sistemlerinin genel yapısı geniş başarılarının sebebidir (Eirinaki vd., 2018: 413).

Öneri sistemleri, e-ticaret, sosyal ağlar, e-turizm ve internet reklamcılığı, vb. gibi çok sayıda alanda hizmet kalitesini arttırmak için başarıyla uygulanmaktadır. Seyahat ürünleri için öneri sistemlerinin geliştirilmesi doğal bir yoldur (Zhu vd., 2017: 17596). Ancak, çok sayıda araştırma (Liu vd.,2011; Tan vd., 2014) seyahat ürünleri önerileri ile geleneksel ürün önerileri arasında önemli farklılıklar olduğunu göstermektedir. Ayrıca, seyahat ürünleri için bir öneri sisteminin tasarlanması ve uygulanmasında birçok zorluklar bulunmaktadır. İlk olarak, seyahat verileri, filmler, kitaplar ve müzikler gibi genel malların verisinden çok daha azdır (Liu vd., 2014). Seyahat masrafları çok daha pahalı olduğu için, turistler genellikle seyahat ürünlerini

çok sık satın almazlar. İkincisi, seyahat ürünleri arasındaki iç ilişkiler çok daha karmaşık ve yakındır. Her seyahat ürünü birçok karmaşık ve çeşitli bilgilerden oluştuğu için tur güzergâhı, program, otel ve araç gibi kendine özgü karmaşık ilişkileri vardır. Ayrıca, bu bilgilerdeki küçük değişiklikler farklı seyahat ürünlerine yol açsa da, İçsel önemi olan bu farklı seyahat ürünleri, turistlerin ortak tercihlerini yansıtabilir. Üçüncü olarak, geleneksel öneri sistemleri genellikle kullanıcı derecelendirmelerine dayanmaktadır. Bununla birlikte, turistler tarafından seyahat ürünlerine yönelik puanlar genellikle kolayca elde edilemez (Liu vd., 2011). Son olarak, seyahatlerde soğuk başlangıçlı kullanıcıların oranı, genel ürün verilerinden daha büyüktür. Turistler genellikle seyahat ürünlerine uzunca bir süre fazla ilgi göstermezler, bu nedenle web erişim günlükleri turizm ürünlerinin satışının yapıldığı web sitelerinde diğer sektörlerde olduğundan daha azdır. Örneğin, turistler genellikle seyahat ettikleri yerler ve seyahat programları planlandıktan sonra turizm web sitesindeki seyahat ürünlerinin sayfalarını ziyaret etmeye başlarlar, bu nedenle çevrimiçi seyahat verilerinde soğuk başlangıçlı kullanıcıların varlığı daha fazla olmaktadır.

Büyük ve çok sayıda veri havuzunda toplanan ve saklanan E-Turizm'deki hızlı büyüyen, muazzam miktarda veri, kullanıcıların kendi başlarına faydalı bilgileri yakalama yeteneğini çok aşmıştır. Kullanıcılar, özellikle de E-turizm web sitelerine göz atarken, nerenin popüler, önce nereye ve ardından nereye gidileceğini bilmek istemektedirler (Hui vd., 2017: 1121). Bununla birlikte, geleneksel E-turizm siteleri yalnızca gün sayısı veya destinasyona göre sıcak rota listesi için sorgu sunmaktadır. Kullanıcılar bir rota verisi açısından zengin ancak bilgi açısından fakir durumda kalabilmektedirler. Bir turist'in kaynakları araştırabileceği, yorumlayabileceği, paylaşabileceği ve değerlendirebileceği çok sayıda iyi bilinen Web tabanlı turizm portalı (örn. TripAdvisor, Expedia, airbnb, Wikivoyage, vb.) bulunmaktadır. Bu platformlar gezi planlaması kapsamında çevrimiçi turizm alanı için hayati bir rolü temsil etmektedir (Xiang ve Gretzel 2010). Bu işbirlikçi platformlar, kullanıcılar önerileri değerlendirerek ve yaparak saygınlıklarını arttırabildiklerinden, itibar tabanlı kitle kaynaklı platformlar olarak düşünülebilirler. Genel olarak, turistlerin kendi dijital ayak izlerini oluşturmalarına ve profil oluşturma ve öneri mekanizmalarını uygulamalarına olanak tanımaktadırlar, yani turistler kendilerine akıllı “kitle kaynaklı” öneri sistemleri oluşturmak için dijital ayak izlerine katkıda bulunurlar. (Leal, vd., 2017).

Son yıllarda, turist öneri sistemi, kullanıcıların turistik yerlerdeki tercihlerini desteklemek için yaygın olarak kullanılmaktadır (Fang vd., 2017: 347). Turizm uygulamaları için tasarlanmış öneri sistemleri aynı zamanda Seyahat Öneri Sistemi veya Destinasyon Öneri Sistemi olarak da adlandırılmaktadır. Bu sistemler müşteri ve seyahat acentası arasında bir aracı

olarak sınıflandırılabilir (Lucas vd., 2013: 3). Dahası, bu sistemler sayesinde müşteriler insan faktörü tarafından kısıtlanmama avantajına sahiptirler, tatil yerlerini kendi görüş ve beğenileri ile belirleyebilmektedirler. Turizm bağlamı özellikle ilginçtir çünkü tavsiyeler, önerilen seyahatin anlamlı bir resmini sağlamak için lokasyonlar, turistik yerler, konaklama yerleri ve uçuşlar gibi çeşitli ürünlere atıfta bulunabilmektedir (Werthner ve Ricci, 2004). Turizm için mevcut öneri sistemleri genellikle kullanıcının seyahatini planlamasına yardımcı olmak (yani bir veya daha fazla varış yeri seçmek) ve kullanıcının bir bölgede (genellikle bir şehir) ne yapması gerektiğini planlamasına yardımcı olmak için tasarlanmıştır.

Mevcut öneri yaklaşımlarının çoğu, temel olarak anahtar kelime arama ve semantik tabanlı aramaya odaklanan içerik eşleştirme yöntemleridir. Ancak, anahtar kelime aramada verimlilik yoktur ve semantik tabanlı yöntemin uygulanması pahalıdır (Gong vd., 2018: 93). İçerik tabanlı turizm öneri sistemleri, kullanıcılara geçmişte sevdiklerine benzer noktaları tavsiye etmektedirler. Geçmişten, kullanıcının profili tercihini temsil edecek şekilde oluşturulmuştur. Öte yandan, tavsiyeye karar vermek için kullanıcının profili ile eşleşmesi için noktaların özellikleri karakterize edilmektedir (Fang vd., 2017: 348). Kullanıcıya zaman ve maliyet gibi kısıtlamalarını karşılamak için uygun bir tur planı sunulabilir. Bazen tur planının önerisi, seyahat yolunun veya hareket esnasında harcanan zamanın en aza indirildiği bir noktalar kombinasyonuna yaklaşmak için bir tam bir programlama problemine veya seyahat satıcısı problemine dönüşebilmektedir.

Öneri yöntemlerinden işbirlikçi filtreleme (Zheng vd., 2010; Nilashi vd., 2015) de turizm önerisi alanlarında yaygın olarak kullanılmaktadır. İşbirlikçi filtreleme sistemleri, benzer tercihleri olan kullanıcı gruplarına göre önerilerde bulunur. Kullanıcılar arasındaki benzerlik normalde bazı öğelere verdikleri derecelendirmeler karşılaştırılarak hesaplanır. Sistem, geçerli kullanıcıyla benzer ilgi alanlarını kimlerin paylaştığını belirlediğinde, bu kişilerin beğendiği öğeler bu kullanıcıya önerilmektedir. Bu yaklaşımda, kullanıcının hangi öğeleri sevdiğini veya beğenmediğini (örneğin, ziyaret etmekten hoşlandığı yerleri) bilmek için verilen öneriler hakkında bazı geri bildirimler gerekmektedir (Patil ve Kölhe, 2014: 1557). Yaygın olarak kullanılan işbirlikçi filtreleme algoritmalarının, birçok nedenden ötürü seyahat ürünlerini tavsiye etmek için kullanılması genellikle zordur. (1) Seyahat ürünlerinin içeriği çok karmaşıktır, (2) kullanıcı-madde matrisi son derece seyrek ve (3) soğuk başlangıçlı kullanıcılar yaygın olarak bulunmaktadır (Zhu vd., 2017: 17595).

Turizm ve seyahat sektörü, farklı kaynaklardan toplanan çok miktarda verinin kullanımı yoluyla hizmetlerini geliştirmektedir. Farklı turistlerin değerlendirme, derecelendirme ve deneyimlerine erişim kolaylığı, turistik planlamayı zengin ve karmaşık hale getirmiştir. Bu

nedenle turizm sektörünün karşı karşıya kaldığı büyük bir sorun, turistlerin tercihlerini tespit etmek ve her bir turist için kişiye özel güzergah sağlamak amacıyla toplanan verileri kullanmaktır (Figueredo vd., 2018: 85). Turistlerin tercihlerini yansıtabilecek olan seyahat verilerinin giderek artması, satışların artmasını sağlamak için seyahat şirketleri farklı turistlerin tercihlerini anlamak ve daha çekici seyahat ürünleri sunmak zorundadır. Bu nedenle, seyahat ürünleri için akıllı bir kişiselleştirilmiş öneri sistemi tasarlamak acil bir taleptir (Zhu vd., 2017: 17596). Bu nedenle, farklı kullanıcıların ihtiyaçlarının özelliklerini tanımlayarak, bilgi önerisi bu sorunu çok iyi çözecektir. Öneri sistemlerinin (ÖS) ilk olarak kullanıcıların beğenilerine göre tavsiye etmeleri önerilmektedir (Hui vd., 2017: 1121).

Turistler, turistlerin geri bildirimlerini ve memnuniyetlerini paylaşmak için olumlu ve kritik olan sosyal katılım gücüne sahiptir. Ziyaretçiler, yalnızca hedeflere olumlu yönde odaklanan reklamlardan çok, yolcuları turizm ürünlerini yaratmanın bir parçası haline getiren iki yönlü iletişime ve akıllı etkileşime güvenmektedir. Kitle kaynaklı bilgiler, karar vermeyi etkilemesine rağmen, tipik olarak, çeşitli platformların ve kaynakların karmaşıklığı nedeniyle, bir turist kendi seçeneklerinin geliştirilmesi için kendi kitle kaynaklı ayak izini izleyemez veya kontrol edemez. Leal, vd. (2017) makine öğrenimi ve veri madenciliği metodolojileri, bilgi keşfi için özel algoritmalar sağladıkça, turizmde kitle kaynaklı bilgi ve makine öğrenimi metodolojilerinin kombinasyonunun, turist seyahat döngüsünün kişiselleştirilmesini daha da kolaylaştırması gerektiğini ve sonuçta turist kitlesi kaynaklı verilere dayanarak geçici seyahat aşamaları önereceğini savunmaktadırlar. Patil ve Kölhe (2014) turizm tavsiye sistemlerinde sıklıkla kullanılan veri madenciliği tekniklerini araştırmışlardır. Literatürde belgelenen turizm tavsiyesi sistemlerinden bazıları, turist tercihlerini, sık davranışları, yeni eğilimleri veya bağlamları saptamak için makine öğrenimi algoritmalarını kullanmasına rağmen, çoğu sistem kitle kaynaklı bilgileri kapsamlı olarak kullanmamaktadır.

Mevcut öneri sistemlerinin hız, heterojenlik veya eksik veri yönü gibi verileri işlemek için açık bir şekilde yeterli olmaması nedeniyle, veri madenciliğinin desteğinde öneri verilebileceği konusunda kapsamlı bir anlayış geliştirilmiştir. Daha hızlı öneriyi sağlamak önemlidir. Dahası, veri ile bilgi arasındaki uçurum, veri mezarlarını bilginin “altın külçelerine” dönüştürebilen veri madenciliği araçlarını gerektirmektedir. Veri madenciliği, büyük veri tabanından geçerli ve belki de bilinmeyen bilgilerin çıkarılması ve daha sonra kritik iş ve stratejik kararlar vermek için bilgi kullanılması sürecidir (Hui vd., 2017: 1121). Veri madenciliğinin örnekleme teknikleri ve boyut azaltma teknikleri, işleme öncesi adımda uygulanabilir; sınıflandırma yöntemi, model bazlı bir RS veya içerik tabanlı RS üretmek için kullanılabilir; RS'in performansını iyileştirmek için kümeleme algoritmaları kullanılır; İlişkilendirme kuralları, bir

işlem olduğunda öğeleri önermek için sezgisel bir çerçeve sunmaktadır (Fu vd., 2000). Kullanıcının gelecekteki seçimi, geçmiş tecrübesi ile tahmin edilmektedir. E-turizm alanında, büyük miktarlardaki işlem verileri, veri madenciliği yoluyla yararlı bilgiler elde etmek için kolay bir fırsat sunmaktadır (Liao vd., 2010; Lucas vd., 2013; Li vd., 2015).

Tablo 4.6 Ürün Önerileri İçin Veri Madenciliği Tekniklerinin Kullanıldığı Araştırmalar

Araştırma	Kullanılan Teknik	Öneri Konusu
Siering vd., 2018	Duygu analizi	Turizmde özellikle havaalanının servisleri için çevrimiçi yorumlarda verilen önerileri analiz etmişler ortaya çıkarmışlardır.
Dietz vd., 2018	Kümeleme	Destinasyon Öneri Sistemi - Önerilen seyahat güzergahlarının kişiselleştirilmesini iyileştirmek için karakteristik turist türlerinin derinlemesine bir analizi ve kümelenmesi gerçekleştirilmiştir.
Zhu vd., 2017	LDA (Latent Dirichlet Allocation) Kümeleme – K-means	Seyahat ürünleri için bir öneri motoru geliştirilmesi - Kişiselleştirilmiş seyahat ürünlerini önermek için hem kullanıcı tıklaması hem de kullanıcı satın alma davranışından yararlanılmıştır.
Gong vd., 2018	Big Data Birliktelik Kuralları	Cazibe merkezi öneri yöntemi - Cazibe merkezi tanımlamalarından ve turist paketlerinden elde edilen bilgi miktarlarının avantajlarını kullanılmıştır.
Figueredo vd., 2018	Deep Learning and Fuzzy Logic	Turizm danışmanı sistemi - Fotoğraflardan gizli kişisel tercihleri tespit edilmiş, turistler sınıflandırılmış ve kentin ilgi çekici yerleri turist profili ile eşleştirilmiştir.
Guo vd., 2018	Algoritma kodları verilmiş	Seyahat güzergâhı önerisi - Çok kaynaklı sosyal medya verilerinden yararlanılarak seyahat için çok yönlü bir bilgi algı ve öneri sistemi
Hui vd., 2017	Birliktelik Kuralları	Turistlere tur güzergâhı önerisi - Her güzergâh, birlikte bulunma yüzdesi yüksek olan birkaç doğal noktadan oluşmaktadır.
Leal, vd., 2017	Latent Dirichlet Allocation	Destinasyon Önerisi - Kullanıcılar ve lokasyonların profilleri çıkartılarak anlamsal konu benzerliğine dayalı gelecekteki kullanıcı destinasyonlarına ilişkin kişiselleştirilmiş öneriler üretir.
Liu vd., 2011	Bayes Ağları	Kişiselleştirilmiş Seyahat Paketi Önerisi - Turist kategorisine ve paket için planlanan seyahat mevsimlerine göre seyahat paketi seçilmektedir.
Logesh vd., 2018	Fuzzy c-means	Akıllı Seyahat Öneri Sistemi - İlgi çeken noktaların listesini sağlayarak seyahat eden kullanıcıyı desteklemek için tasarlanmıştır.
Lucas vd., 2013	İlişkisel sınıflandırma ve bulanık mantık	Turizm sistemi için karma öneri yaklaşımı - Sistem, kullanıcı özelliklerine göre turistik noktalar hakkında bilgi sağlar. Bu özellikler demografik bilgiler, seyahat geçmişi ve belirli turistik noktalara yönelik kullanıcı etkileşimleri ile karşılaştırılmaktadır. Bu noktalar sistem tarafından anlamsal gruplar olarak ele alınmakta ve aktif kullanıcı, profilindeki verilere göre bu gruplardan birinde sınıflandırılmaktadır.
Fu vd., 2000	Birliktelik Kuralları	Web Sayfası Önerisi - Sistem kullanıcının navigasyonunu aktif olarak izler ve gözlemler. Kullanıcının navigasyon geçmişi yakalandıktan sonra, geçmişe ait gizli bilgiler keşfedilir. Bilgi daha sonra potansiyel olarak ilginç web sayfalarını kullanıcılara önermek için kullanılır.

Kesorn vd., 2018	Rocchio algoritması	Kişiselleştirilmiş Cazibe Öneri Sistemi - Öneri Facebook check-in verilerinden elde edilen bilgilere dayanmaktadır. Ayrıca, "soğuk başlangıç" adı verilen sorun, kullanıcıların arkadaşlarının giriş verileri kullanılarak çözülmüştür.
Miah vd., 2018	Çekirdek Algoritma - (Depth First Search) algoritması	Akıllı Tur Planlama Öneri Sistemi - Verilen bütçeye ve zamana göre kullanıcının destinasyonunu baz alarak tüm alt yerlerin en yakın olduğu bir yeri ziyaret etmek için maksimum öneri sağlayacak bir özelleştirme algoritması sunulmuştur.
Nilashi vd., 2015	EM, ANFIS, PCA	Otel Önerisi - Kullanılan algoritmalar kullanıcı tercihlerini çeşitli açılardan değerlendirerek turistler için uygun öneriler sunmaktadır. Boyut azaltma ve tahmin tekniklerinin kullanıldığı otel önerileri için yeni bir karma yöntem sunulmuştur
Sutjiadi vd., 2018	Fuzzy C-Means Algorithm	Surabaya Turizm destinasyon önerisi - Surabaya şehri ziyaret etmek isteyen turistlerin kriterlerine ve bütçesine uygun en yakın otel ve restoranı bulmaları için yardımcı olmaktadır.
Veloso vd., 2018	Büyük Veri	Ölçeklendirilebilir Öneri Motoru - Restoran verilerindeki derecelendirmeler, yorumlar ve konum kullanılarak iki farklı profilleme yaklaşımı incelenmiştir. Nihai önerileri doğru bir şekilde hesaplamak için, mekan uygunluğuna, paranın değerine ve duyarlılığına dayalı olarak öneri filtreleri uygulanmıştır.
Wan vd., 2018	K-En Yakın Komşu --- Bayes	Turist güzergâhi önerisi - Kullanıcıların lokasyona dayalı sosyal ağlardan çıkardıkları coğrafi tercihleri, doğal noktaların mekansal dağılımını ve hava koşulları gibi diğer seyahat faktörlerini göz önünde bulundurarak akıllı turist güzergah önerileri yapmak için karma bir yöntem geliştirilmiştir.
Zhang ve Piramuthu, 2018	Latent Dirichlet Allocation	Ürün Önerisi - Geçmiş müşterilerinin oteller hakkında yazılan yorumlardan elde edilen bilgiler ile ilgilenilen konular bakımından benzer düzeyde tüketici yararı sunan ürünler kümelenmiştir. Potansiyel herhangi bir müşteri için, ilgi konusu veya ilgi alanı özellikleri göz önüne alındığında daha iyi bir değere sahip ürün / hizmet önerileri oluşturulmaktadır.
Zheng vd., 2018	Tekil değer ayrışımı (TDA)	Turizm destinasyon öneri sistemi - Kullanıcı duygularını rafine etmek için fikir madenciliği teknolojisini kullanan, zamanla kaybolan destinasyon popülaritesini ve kullanıcı tercihlerini temsil etmek için zamansal dinamikleri kullanan bir turizm destinasyon öneri sistemi sunulmuştur.
Zisos vd., 2018	Duygu analizi	Otel Öneri Sistemi - Kullanıcılar ve ürünler için dinamik profiller tanımlanıyor ve bunlar en etkili şekilde eşleştiriliyor. Bu sayede hedeflenen müşteri segmentlerine daha iyi ve kişiselleştirilmiş pazarlama yapılabilir. Uyumlu profiller daha sonra işlenerek ve kullanıcılara ürün / alternatif önerileri sağlamak için kullanılıyor.
Sun vd., 2018	Destek Vektör Makineleri	Kişiselleştirilmiş Cazibe Öneri Sistemi Coğrafi etiketli fotoğraflar kullanılarak, bir kullanıcının tercihleriyle eşleşen cazibe merkezi önerileri sağlamak için kişiselleştirilmiş bir öneri sistemi oluşturulmuştur

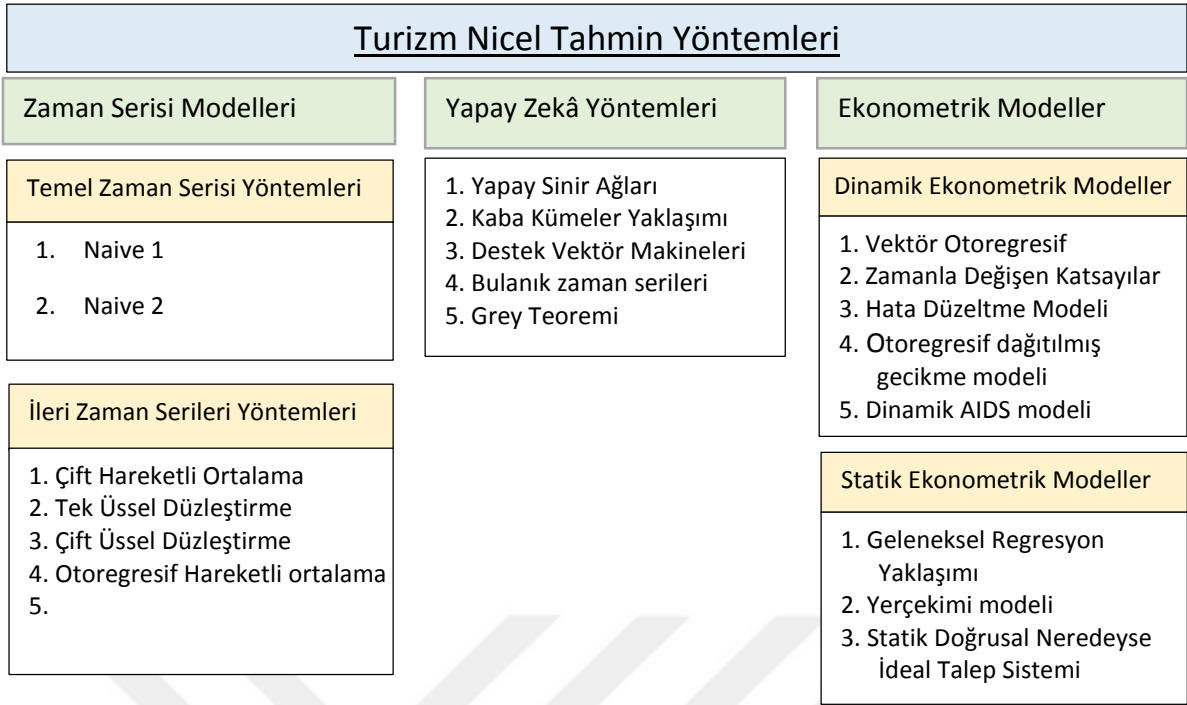
4.7. Turist Tahmini

Turizm talebi tahmini, ileri turizm planlaması için esastır (Jun vd., 2018: 136). Turistlerin hızlı akışına ayak uydurabilmek için bir destinasyonun kaynaklarının kullanımının planlaması gerekmektedir. Planlama sürecindeki kilit ve ön unsurlardan biri, turist gelişlerine olan talebi hem hacim hem de belirleyiciler açısından incelemektir (Goh ve Law, 2003: 511). Uygun kamu politikaları geliştirmek ve sağlam iş yatırım kararları almak için, hem devlet idareleri hem de özel sektör işletmeleri gelecekteki operasyonları planlamak ve tesisler ve altyapı yatırımlarına olan ihtiyacı değerlendirmek için temel turist talebi tahminini kullanmaktadır. Bu nedenle, tahminler turizm yönetimi için vazgeçilmez hale gelmiştir (Jun vd., 2018: 136).

Tahmin, turizm endüstrisi için hayati öneme sahiptir. Bununla birlikte, turizm pazarının karmaşık, evrimsel doğası nedeniyle, turist talebi serisi doğası gereği gürültülü, koşullu olarak durağan olmayan ve bazı durumlarda kaotik olma eğilimindedir. Bu dinamik, durağan olmayan talep serilerinin modellenmesi zordur, çünkü daha az gürültü ve karmaşıklıkla daha doğru tahminler yapılmasını sağlayan bir sisteme ihtiyaç vardır (Jun vd., 2018: 136).

Turizm araştırmalarında önemli alanlardan biri olan turizm talebi modellemesi ve tahmini, hem akademisyenlerin hem de pratisyenlerin dikkatini çekmiştir (Höpken vd., 2017: 189). Li vd. (2005) tarafından yapılan kapsamlı bir incelemeye göre, 1960–2002 döneminde bu konuyla ilgili 420 çalışma yayınlanmıştır. Bu çalışmaların çoğu, çeşitli yerlerde turizm talebini modellemek ve tahmin etmek için hem niteliksel hem de niceliksel olan farklı tekniklerin uygulanmasına odaklanmaktadır (akt: Song ve Li, 2008: 203). Song ve Li (2008), sosyal bilimler alıntı endeksi (SSCI), Google Akademik gibi çeşitli veri tabanlarında ve yayınlanan makalelerden yapılan alıntılarda bir araştırma yaptılar. Bu çalışmada, 2000-2007 yılları arasında yayınlanan turizm talep modellemesi ve tahmini hakkında yüz yirmi bir çalışma tespit edilmiştir. Bu 121 çalışmada bazı genel gözlemlere ek olarak, en son metodolojik gelişmeler, tahmin rekabeti, kombinasyon ve entegrasyon, turizm döngüleri, dönüş noktaları, yönlü değişiklikler ve mevsimsellik analizi, olayların etki analizi ve risk tahminleri konularına vurgu yapılmıştır.

Bu alanda yapılan çalışmalarda, temel turizm dalgalanma mekanizmasını araştırmayı ve turist talebi veya varışları için uygun öneri geliştirmeyi amaçlayan birçok turizm tahmin modeli önerilmiştir. Bu çalışmaların çoğu, Şekil 4.1’de de açıklandığı gibi (1) nedensel ekonometrik modellere, (2) zaman serisi modellerine ve (3) yapay zeka (AI) modellerine ayrılabilen nicel yaklaşımları benimsemiştir (Jun vd., 2018: 136).



Şekil 4.1 Üç Ana Nicel Tahmin Metodu Kategorileri

Zaman serisi yöntemleri, birden fazla olmayan geçmiş veri dizisinden tahmin ederek gelecekteki eğilimleri önceden belirlemektedir. Farklı turist varış tahmini modellerinin doğruluğunu test ederken, zaman serisi modellerinin düşük maliyetle ve makul faydalarla kabul edilebilir tahminler üretebileceği bulunmuştur. Genel olarak, zaman serisi modellerinden ARIMA en yaygın olarak kullanılmaktadır. Zaman serisi yaklaşımlarının turizm talebi tahmininde faydalı olduğu kanıtlanmış olmasına rağmen, temeli turist karar verme süreçlerini içeren herhangi bir ekonomik teoriye dayanmaması gibi bazı önemli sınırlamaları vardır. Bu nedenle turist davranışlarını analiz edememektedirler ve belirli turizm odaklı strateji ve politikaların etkinliğini değerlendirmekte yetersiz kalmaktadırlar.

Nedensel ekonometrik modeller, bir varsayımsal açıklayıcı faktörler seti kullanarak, turizm talebi (bir varış yerindeki turist varış yerleri tarafından ölçülen) ve turist harcamaları gibi değişkenler arasındaki ilişkileri belirlemeye çalışmıştır (Jun vd., 2018: 136). Ekonometrik yaklaşımların zaman serisi modellere göre en büyük avantajlarından biri, turizm talebi (bağımlı) değişkeni ile etkileyen faktörler (açıklayıcı değişkenler) arasındaki nedensel ilişkileri analiz edebilmelerinde yatmaktadır (Höpken vd., 2017: 189). Turizm talebinde yapılan son ekonometrik çalışmalar, tüketici fiyat endeksinin (yani, menşe ülkesindeki ülkelere göre bir varış noktasında turizm fiyatları, rekabet eden destinasyonlarda turizm fiyatları [ikame fiyatlar]), gayri safi yurtiçi hasılanın (turistlerin gelirini temsilen), döviz kurlarının, aynı zamanda faiz ve işsizlik oranının, aynı zamanda faiz ve işsizlik oranı, para arzının ve ihracat / ithalat oranlarının turizm talebinin en önemli (ekonomik) belirleyicileri arasında olduğunu göstermiştir (Cho

2001; Song ve Li 2008: 211). Ayrıca, insan yapımı olaylar (özellikle mega olaylar), reklam yatırımları, ayrıca krizler (örneğin finansal krizler, terör saldırıları) ve doğal afetler (SARS, ayak ve ağız hastalığı vb.) turizm talebini (yani dış şokları) önemli ölçüde belirlemektedir (Höpken vd., 2017: 189).

Oto regresyon dağıtılmış gecikmeli modeller (ADLM), hata düzeltme modeli (EDM), vektör otoregresyon (VAR) modeli ve Zamana göre değişen parametre (TVP) modeli temel ekonometrik modeller olarak ortaya çıkmıştır (Peng vd., 2014). Ayrıca, yapısal eşitlik modellemesi turizm talebi modellemesi için kullanılmıştır (Höpken vd., 2017: 189).

Turizm ve konaklama endüstrisinde, talebi ve nihayetinde endüstrinin (veya belirli bir otel veya restoranı) performans faktörlerini anlamak çok önemlidir. Geçmişte, yöneticiler ve karar vericiler bu değişkenlerin etkisini araştırmak için basit veri analizi formlarına (basit doğrusal regresyon ya da çoklu regresyon) güveniyorlardı. Daha yakın zamanlarda, büyük veri analizi, değişkenler veya nedenler arasındaki ilişkiyi anlamak ve sonunda performansı tahmin etmek için karmaşık iş durumlarını inceleme potansiyelini ortaya çıkarmıştır. Makine öğrenme teknikleri, zamana bağlı bir karar verme sürecini şekillendirmeye yardımcı olmak için depolanmış veya canlı akışa tabi tutulmuş on veya yüzlerce (hatta binlerce) veri değişkenini analiz etmeyi mümkün kılmaktadır (Al Shehhi ve Karathanasopoulos, 2018: 1624). Daha doğru tahminlere duyulan ihtiyaç, destinasyonlara turist taleplerinin daha ayrıntılı tahminlerini elde etmek için makine öğrenme tekniklerinin kullanımının artmasına neden olmuştur (Claveria vd., 2016: 111).

Tablo 4.7’de turizm talep tahmini için veri madenciliği tekniklerinin kullanıldığı çalışmaların bir kısmının listesi verilmiştir.

Tablo 4.7 Tahmin İçin Veri Madenciliği Tekniklerinin Kullanıldığı Araştırmalar

Araştırma	Kullanılan Teknik	Yapılan İşlem
Cho, 2003	Yapay Sinir Ağları	Seyahat Talep Tahmini Farklı ülkelerden Hong Kong'a seyahat talebini (gelenlerin sayısını) tahmin etmek için üç zaman serisi tahmin tekniğinin (üstel düzleştirme, tek değişkenli ARIMA ve Elman'ın Yapay Sinir Ağları Modeli) uygulamasını incelemektedir.
Wu vd., 2010	Bağımsız Bileşen Analizi	Otel Doluluk Oranı tahmini Otel doluluk oranlarını etkileyen baskın faktörlerin belirlenmesi için bağımsız bileşen analizi (ICA) tercih edilmiştir.
Goh ve Law, 2003	Kaba Kümeler	Seyahat Talep Tahmini Ayrılmazlık ilişkileri (indiscernibility relations), küme yaklaşımı ve nitelik azaltma kavramlarına dayanarak, 131 eğitim verisi analiz edilmiş ve karar kuralları oluşturulmuştur. Bu kurallar, farklı menşe ülkelerinden / bölgelerinden Hong Kong'a turistlerin geldiğini tahmin etmek için kullanılmıştır
Wong vd., 2006	Bayes Modeli	Turizm Talep Tahmini Çeşitli vektör oto regresyon (VAR) modellerinin tahmin performansını inceleyerek turizm literatüründeki mevcut tahmin doğruluğu tartışmalarını genişletmektedir. Özellikle, bu çalışma,

		Bayesian kısıtlamalarının (önceki) sınırsız VAR sürecine dahil edilmesinin, daha yüksek bir doğruluk derecesi elde etmek için performans tahmininde bir iyileşmeye yol açıp açmayacağını araştırılmasını amaçlamaktadır
Goh vd., 2008	Kaba Kümeler	Turizm Talep Tahmini Turizm talep analizi ve tahmini için nicel ekonomik ve nitel ekonomik olmayan faktörler içeren bir dizi hibrid veriden karar kurallarının otomatik olarak keşfedilmesi için kaba kümeler yaklaşımı uygulanmıştır.
Wu vd., 2013	Lojistik regresyon	Otel İnternet Sitelerinin Tarayıcılarını ve Satın Alanlarını Tahmin Etme Otel web sitelerinin tarayıcılarını ve alıcılarını tahmin etmek için değişkenlerin optimal alt kümesini araştırmak için lojistik regresyon modeli uygulanmıştır.
Hassani vd., 2015	İleri Beslemeli Sinir Ağı	ABD Turist gelişlerinin tahmini ABD'ye gelen turistlerin zaman serisi, toplam ve menşei ülkeye göre analizi yapılmış ve Vector SSA modelinin tahmin doğruluğu, yaygın olarak kullanılan çeşitli tahmin modellerinin (ARIMA, ETS ve ileri beslemeli Sinir Ağı) tahmin doğruluğu ile karşılaştırılmıştır.
Claveria vd., 2016	Destek Vektör Regresyon Sinir Ağı	İspanya'ya Turizm Talebi Tahmini Makine öğrenme tekniklerinin tahmin performansı karşılaştırılmıştır. İspanya'nın on yedi bölgesine uluslararası turizm talebi tahmini yapılmıştır.
Pan ve Yang, 2017	Büyük Veri EM algoritması	Haftalık Otel Kapasitesinin Tahmini Bir varış yeri için haftalık otel doluluk oranını tahmin etmek için birçok büyük veri kaynağı (ilgili arama motoru sorguları, yerel turizm bürosu web sitesi trafiği ve ayrıntılı hava durumu bilgileri) ve iki farklı yöntem (ARMAX ve MSDF) kullanılmıştır. Çalışmada bir varış yeri için haftalık otel doluluk oranını tahmin etmenin en iyi yolu araştırılmıştır.
Cankurt ve Subaşı, 2016	MLR ANN Destek vektör Regresyon	Turizmde talep modellemesi ve tahmini Çoklu doğrusal regresyon, çok katmanlı algılayıcı (MLP) regresyonu ve destek vektör regresyonu (SVR) Türkiye için çok değişkenli turizm tahmini yapmak için kullanılmıştır. Turizm talebini tahmin etmek için aylık veri puanlarıyla çok değişkenli regresyon modeline dayanan veri madenciliği teknikleri karşılaştırılmıştır.
Çuhadar vd., 2009	Yapay Sinir Ağları	Turizm talep tahmini Antalya iline yönelik aylık dış turizm talebi tahminleri yapılmıştır
Hassani vd., 2017	ARIMA, ETS, NN, TBATS ve ARFIMA	Turizm talebini tahmin etmek Seçilen Avrupa ülkelerinde turizm talebini tahmin etmek için çeşitli parametrik ve parametrik olmayan tahmin tekniklerinin kullanımını değerlendirmiştir.
Höpken vd., 2017	K-en yakın komşu	Turist gelişimini tahmin etmek Geçmiş gelişleri girdi olarak kullanarak turizm talebini tahmin etmektedir ve istatistiksel yöntemlerle veri madenciliği yöntemleri karşılaştırılmıştır.
Li vd., 2017	Büyük Veri	Turizm talebini tahmin etmek Arama verileriyle Pekin turist hacimlerinin tahmininde uygulanan yeni bir tahmin çerçevesi öneriliyor. İlk olarak, daha kapsamlı bir endeks oluşturmak için arama eğilimi verilerinin ortak bileşenlerini kullanan bir GDFM tanıtılıyor sonra bu yeni endeksi geleneksel zaman serisi modeli ve geçmişte yapılan çalışmalarda yaygın olarak kullanılan PCA tabanlı endeks modeli ile karşılaştırılıyor. Ekonometrik modellerin performanslarını statik ve dinamik testlerle farklı endekslerle değerlendiriliyor

Pantano vd., 2017	Rastgele Orman	Turistlerin bir turistik yere tepkisini tahmin etmek Bu araştırma, turistlerin belirli bir varış noktasına yönelik puanlamalarını tahmin etmek için açık verilerin kullanımı araştırılmıştır.
Li vd., 2018	Geri Yayılım Sinir Ağları	Etkili turist hacmi tahmini Baidu endeksine dayalı turist hacmini tahmin etmek için PCA-ADE-BPNN adlı etkili bir model önerilmiştir. r
Al Shehhi ve Karathanasopoulos, 2018	DVM/SVR, RBF	Oda Fiyatı Tahmini En iyi tahmin modelini seçmek için dinamik günlük ortalama otel odası fiyatlarının tahmini yapılmıştır. Doğrusal (basit hareketli ortalama ve ARIMA) ve doğrusal olmayan (RBF ve DVM) modellerden, en iyi modeli veya bunların bir kombinasyonunu belirlemek için model performans ölçüleri karşılaştırılmıştır.
Arabi vd., 2018	Bayes Modeli	Ulaşım Arz Ve Talep Tahmini Gelecekteki ulaşım durumunu tahmin etmek için Bayesian modeli kullanılmıştır.
Jun vd., 2018	Yapay sinir ağı K-Ortalama Kümeleme	Turizm talep tahmini Yapay bir sinir ağı (YSA) ve verilen veri serisinin iki yönünü dikkate alan bir kümeleme algoritması kullanan bir kombine turizm tahmin modeli önerilmiştir. Önerilen modelin üstünlüğünü ve doğruluğunu doğrulamak için, diğer üç YSA tabanlı model ve en popüler ARIMA modeliyle üç doğrusal olmayan, durağan olmayan turist varış veri serileri kullanılarak karşılaştırıldı.
Boz vd., 2018	Rastgele Orman Destek Vektör Mak. En Yakın Komşu C4.5	Rezervasyon iptali tahmini Otel rezervasyon verisi kullanılarak otellerin gelecekteki rezervasyonlarının iptal durumları tahmin edilmeye çalışılmıştır. Algoritmaların performans karşılaştırılması yapılmıştır
Küçük Yılmaz vd., 2017	Karar Ağaçları	Yolcu talep tahmini Eskişehir'de havayolu yolcu taşımacılığına olan potansiyel yolcu talebi ölçülmüştür.

BEŞİNCİ BÖLÜM ARAŞTIRMA

5.1. Araştırmanın Amacı

Bu doktora tezinin amacı veri madenciliği yöntemlerinden sınıflandırma teknikleri kullanılarak günlük tur satın alan ve almayan müşterilerin profilinin çıkarılmasıdır. Bu amaç için Antalya’da faaliyet gösteren A sınıfı bir seyahat acentasının rezervasyon ve günlük tur ile ilgili işlemsel verilerinin kaydedildiği bir veri tabanında tutulan veriler kullanılmıştır. Rezervasyonu yapılan bir turistin günlük tur satın alma durumu “SATIN ALAN” veya “ALMAYAN”dır, bu da bir sınıflandırma problemidir. Dolayısıyla çalışmada veri madenciliğinin sınıflandırma fonksiyonuna ilişkin yöntemler kullanılmıştır.

Araştırma sürecinde seyahat acentası veri tabanındaki farklı tablolarda yer alan rezervasyon, günlük tur bileti, ürünler ve müşteriler hakkındaki veriler bir tabloda birleştirilmiştir. Elde edilen veriler üzerinde 11 farklı sınıflandırma algoritması çalıştırılmıştır. Bu algoritmalarından en iyi performansı ve doğru tahmin oranını veren karar ağacı sınıflandırma tekniklerinden C4.5 algoritması veri içindeki müşteri profillerinin keşfi için kullanılmıştır.

Çalışmada, turizm sektörünün lokomotif firmalarından seyahat acentalarının operasyon kayıtlarının tutulduğu bir veritabanı kullanılmıştır. Bu veri tabanında acentanın vize işlemlerinden rezervasyon, günlük tur ve operasyon gibi günlük faaliyetlerine kadar her türlü veri depolanmaktadır. Çalışma için seyahat acentasının tur satışlarının artırılması çalışmalarına katkı sağlamak amacıyla tur verileri tercih edilmiştir. Günlük tur satışları karşılayıcı olarak hizmet veren bir seyahat acentasının otel satışlarından sonra hatta doğru yapılması durumunda en fazla gelir elde edeceği bir alandır.

Kullanılan veritabanı SQL Server 2012 üzerinde çalışan ilişkisel bir veritabanıdır. Veritabanı içindeki tablolardan, tezin amacına uygun olarak, günlük tur kayıtlarının tutulmasında kullanılan tablolar ve ilgili diğer tablolar Microsoft Access veritabanına aktarılmıştır. Microsoft Access, kurulumunun kolay olması, ekstra bir veri tabanı yönetim sistemi gerektirmemesi, neredeyse Microsoft Office paketinin kurulu olduğu tüm bilgisayarlarda olması, ulaşımının kolay olması nedeniyle tercih edilmiştir. Veriler, Microsoft Access, SQL komutları aracılığıyla süzülerek ve yeniden düzenlenerek analize hazır hale getirilmiş ve RapidMiner 9.1 programında analiz edilmiştir.

Seyahat acentasının yurt dışından getirdiği yolcuların günlük tur satın alma eğilimlerinin belirlenmesi ve turlara göre müşteri profillerinin belirlenmesi iki aşamada gerçekleştirilmiştir. İlk aşamada turistlerin tur satın alma durumlarını tahmin eden sınıflandırma

algoritmalarının performansları değerlendirilmiştir. Kurulan 11 ayrı sınıflandırma modelinden doğruluk ve hız açısından en iyi performansı sergileyen karar ağaçları C4.5 algoritması ile ikinci aşama modeli oluşturulmuştur. Bu model için bağımlı değişken veri setindeki 6 farklı günlük turun satın alınma durumuna göre hazırlanmıştır. Seçilen algoritma farklı tur (Tekne, Alış Veriş, Hamam, Eğlence, Gezi, Rafting) kategorilerinde her seferinde sınıflandırma değişkeni (bağımlı değişken) değiştirilerek çalıştırılmıştır. Bu modelde her tur için ayrı ayrı satın alınma durumları tahmin edilmiştir.

Çalışmaların sonuçları ışığında bu alanda veri madenciliği uygulaması yapmayı düşünen araştırmacılara öneriler ve sonuçların değerlendirilmesi ile çalışma sonlandırılmıştır. Müşteri profillerinin keşfi ile seyahat acentası yöneticileri için potansiyel müşterilerine yönelik promosyon ve fiyatlandırma stratejisi oluşturmalarına katkı sağlamak hedeflenmektedir. Doğru müşteriye doğru günlük turun satılması, günlük tur alma ihtimali yüksek olan müşteriye teşvik verilmesi, çalışmanın amaçları arasındadır. Buradaki temel iki amaç, promosyona ayrılan kaynakları etkin kullanarak müşteri memnuniyetinin en yüksek seviyede gerçekleşmesini sağlamak ve seyahat acentasına görece katkısı az bir turiste gereksiz promosyonu önleyerek maliyetleri azaltmaktır.

5.2. Kapsam ve Yöntem

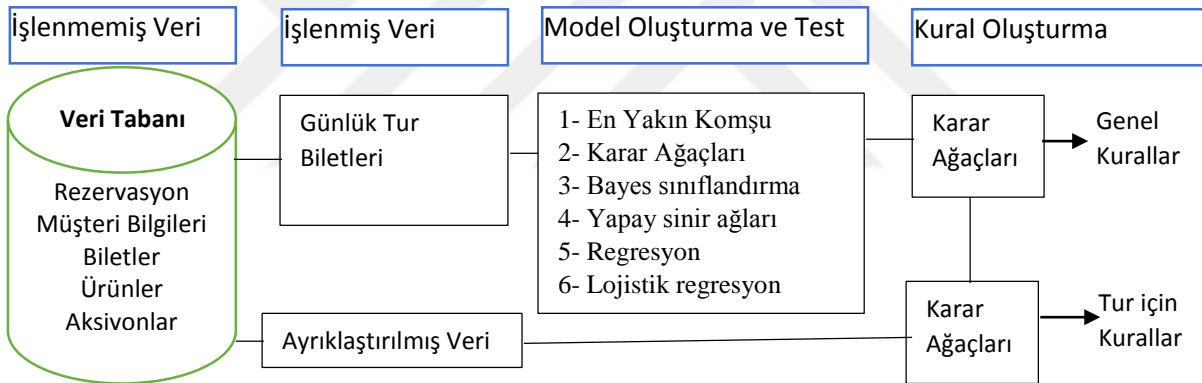
Çalışma, karşılayıcı acentacılık hizmetleri veren bir seyahat acentasının veri tabanından elde edilen örnek veri kümesi üzerinde uygulanmış ve çok geniş bir çalışma alanı olan veri madenciliğinin sınıflandırma fonksiyonuna ilişkin teknikleri ile sınırlandırılmıştır. Bu kapsamda, veri madenciliğinin sınıflandırma teknikleri arasında ölçeklenebilirlikleri, yaygınlıkları ve önceki benzer çalışmaların tercihleri dikkate alınarak, En yakın komşu, Karar ağaçları, Bayes sınıflandırma, Yapay sinir ağları, Regresyon ve Lojistik regresyon analizi kullanılmıştır. Bu algoritmalar arasında yapılan performans değerlendirilmesinin ardından karar ağaçları ile ikinci bir model daha oluşturulmuştur.

Uygulamaya konu olan veri kümesi, ilgili tabloda 30 farklı değişken ile temsil edilen 9.972 adet rezervasyona ait değerleri içermektedir. Uygulamada CRISP–DM konsorsiyumu tarafından önerilen veri madenciliği standart süreci esas alınmıştır. Söz konusu veri kümesi üzerinde Microsoft Access 2016, Microsoft Excel 2016 ve RapidMiner 9.1 yazılımları kullanılarak veri madenciliğinin sınıflandırma problemlerinin çözümüne ilişkin teknikleri için bir örnek uygulama ortaya konmaktadır.

Başarılı bir sınıflandırma uygulaması için, RapidMinerStudio 9.1 tercih edilmiştir. RapidMiner, dünyanın en yaygın ve en çok kullanılan açık kaynaklı veri madenciliği çözümlerinden biridir. 2001 yılında Dortmund Üniversitesi'nde oluşturulan proje, 2007'den bu yana Rapid-I GmbH tarafından geliştirilmektedir. RapidMiner, akademik geçmişi ile sadece veri madenciliği çalışan profesyonelleri değil, aynı zamanda en farklı disiplinlerden gelen üniversiteleri ve araştırmacıları da desteklemeye devam etmektedir.

5.3. Araştırma Süreci

Veri madenciliği projeleri için CRISP-DM metodolojisi, bilginin büyük veri setlerinden çıkarılmasına uygulanan standart bir metodoloji olarak kabul edildiğinden (Nave vd., 2018: 698) bu çalışmada da araştırma süreci olarak tercih edilmiştir. Metodolojide problemin anlaşılması, verinin anlaşılması, verinin hazırlanması, modelleme, değerlendirme ve uygulama gibi 6 aşama vardır (Larose, 2006: 265). Çalışmada bu altı aşama bazen önceki aşamalara tekrar tekrar dönülerek gerçekleştirilmiştir.



Şekil 5.1 Araştırma Süreci

5.4. Araştırmanın Problemleri

Karşılıklı seyahat acentası operasyon bölgesindeki oteller ile anlaşarak tur operatörlerinin gönderdiği yolcuların transferlerini yapan ve günlük turlar düzenleyen bir seyahat acentasıdır. Bu tür seyahat acentalarının en büyük gelir kaynakları yaptıkları transferler ve organize ettikleri veya sattıkları günlük turlardır. Gelirlerini arttırmak isteyen bir acentanın transferler konusunda kendi bünyesinde yapabileceği bir şey yoktur, çünkü transferler yurt dışındaki tur operatörünün satışını yaptığı konaklamalar için yapılmaktadır. Bu durumda acentanın gelirlerini arttırmak için acenta içinde eylem yapabileceği alanların en başında günlük tur satışlarını arttırmak gelmektedir.

Karşılıklı olarak hizmet veren seyahat acentaları günlük tur biletlerini rezervasyon kaydı yapılan müşterilerin otele girişinden sonra yaptıkları (genelde aynı gün içinde) bilgilendirme toplantılarında satmaktadır. Acenta temsilcisinin daha iyi satış yapabilmesi için karşısındaki turisti tanıması gerekmektedir. Bunun için daha önceki satış konusundaki tecrübelerini kullanarak tur satışında hedef kitlesini seçmektedir. Bu tez çalışması ile tur satımı için hedef turist gruplarının belirlenmesi işleminin sınıflandırma yöntemleri ile yapılması amaçlanmıştır. Bu amaç için daha önce tur satın almış müşteriler “SATIN ALAN”, tur almamış müşteriler de “ALMAYAN” olarak sınıflandırılmıştır.

Çalışmanın ana problemi, seyahat acentasında otel rezervasyon kaydı bulunan turistlerin günlük tur satın alma durumlarını tahmin eden sınıflandırma yöntemlerinin performanslarını karşılaştırmaktır.

Karşılaştırma sonucunda en iyi performansı veren algoritma ile aynı veri seti içindeki farklı günlük turların satın alınma durumuna göre ikinci bir sınıflandırma işlemi yapılacaktır. Bağımlı değişken içeriği ana problem için herhangi bir turun satın alması durumunda “SATIN ALAN” iken bu defa özel bir turu (örneğin ALIŞVERİŞ) satın alma durumunda “SATIN ALAN” olarak değiştirilecektir. Bu işlem veri setindeki yedi farklı günlük tur için altı kez tekrarlanacaktır. Bu da çalışmanın alt problemidir. Bu sınıflandırma ile her hangi bir günlük turu hangi turist grubunun tercih ettiği tespit edilecektir.

Günlük tur biletleri otellerde yapılan bilgilendirme toplantılarında satılmaktadır. Oteller operasyon bölgelerine göre dizilmekte ve tur satış ve operasyonları da bölgeler bazında yapılmaktadır. Bu uygulamadan yola çıkılarak çalışmada ikinci alt problemi, tur satın alan müşteri profillerinin bölgelere göre farklılık gösterip göstermediğinin tespit edilmesidir.

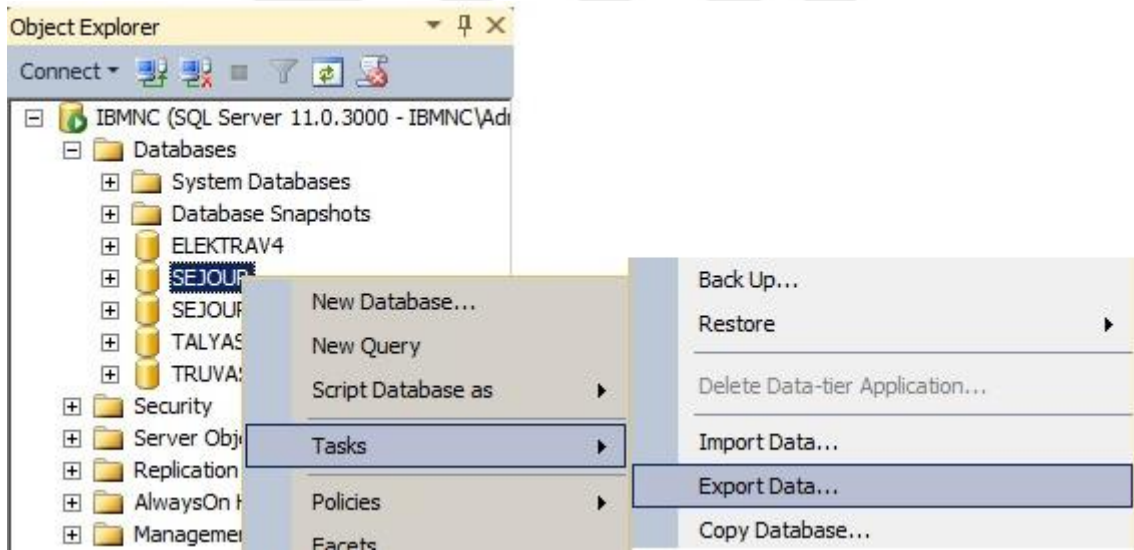
5.5. Verinin Toplanma Süreci

Çalışmanın amacına uygun bir şekilde verilerin elde edilebilmesi için seyahat acentasında daha önce yedek olarak kaydedilmiş bir veri seti kullanılmıştır. Veritabanı Antalya’da faaliyet gösteren ve Avrupa pazarında çalışan A grubu bir seyahat acentasına aittir. Kullanılan veritabanı seyahat acentasının 2016 ve önceki yıllara ait tüm departmanlarının günlük faaliyetlerinin verilerini içeren ve Microsoft SQL Server 2012 üzerinde çalışan bir veri tabanıdır. Microsoft SQL Server 2012 ilişkisel model kullanan bir veri tabanı yönetim sistemidir. Bu sistemin verileri biçimlendiren ve depo edilmesini sağlayan öğeleri tablolardır. Tablolarda bulunan satırlar kayıtların kendisini, sütunlar ise bu kayıtları oluşturan bilgi parçalarının ne türden olduklarını belirtmektedir ve alan olarak adlandırılmaktadırlar.

Veri toplama sürecinde ilk olarak seyahat acentası veri tabanından ilgili tablolar alınarak analize yönelik yeni bir veri tabanı oluşturulmuştur. Oluşturulan veri tabanında analize uygun olmayan veriler temizlenmiş ve farklı tablolarda bulunan veriler yeni bir tabloda birleştirilmiştir. Birleştirme esnasında bazı değişkenler için dönüştürme işlemi de yapılmıştır. Bu işlemler için Microsoft Acces 2016 programı ve SQL sorguları kullanılmıştır.

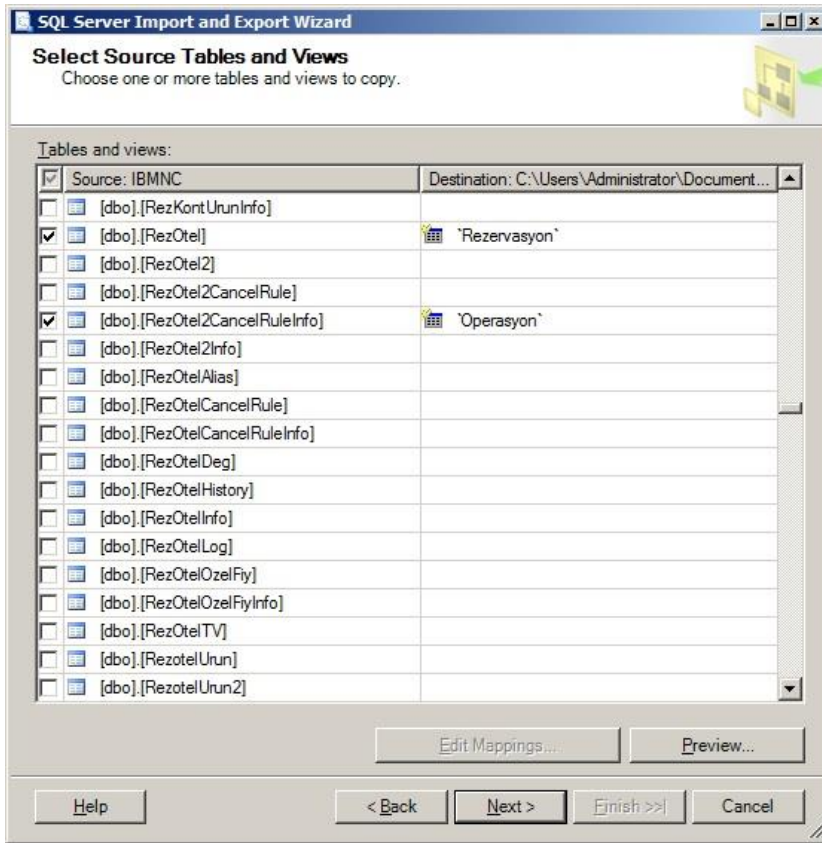
5.5.1. Veri Tabanının Oluşturulması

Seyahat acentasından elde edilen veri tabanı acentanın tüm faaliyetleirnin kaydedildiği Microsoft SQL Server 2012 üzerinde çalışan bir veri tabanıdır. Microsoft SQL Server 2012 veri tabanı yönetim sisteminin her bilgisayarda kurulmamış olması ve veri tabanındaki tüm verilerin tez çalışması kapsamında kullanılmayacak olması nedeni ile çalışma ile ilgisi olan tablolardan yeni bir veri tabanı oluşturulmuştur. Yeni veri tabanı için özellikle lisans sorunu olmadığı ve daha hızlı ve esnek çalışma ortamı sağlayacağı için Microsoft Access veri tabanı yönetim sistemi tercih edilmiştir.



Şekil 5.2 Verinin Microsoft SQL Server'dan Microsoft Access Veritabanına Aktarılması

Verinin Microsoft SQL Server 2012 veri tabanından Microsoft Access 2016 veritabanı olarak aktarılması için SQL Server üzerinde ilgili veritabanı üzerinde farenin sağ tuşuna tıklatarak Tasks - Export Data komutu kullanılmıştır. Daha sonra kaynak veri tabanı ve ardından hedef veri tabanı seçimi yapılmıştır. Son adımda Microsoft Access veritabanına aktarılacak tablolar işaretlenerek Tez_Veri ismi verilen yeni bir veritabanı oluşturulmuştur.



Şekil 5.3 Veritabanından Analiz İçin Seçilen Tablolar

Verileri amaca uygun bir şekilde analize hazırlayabilmek ve veri içindeki ilişkileri ortaya çıkartabilmek için öncelikli olarak veritabanı içindeki veri ilişkilerinin ve alan görevlerinin iyi bilinmesi gerekmektedir. Aksi takdirde veri içerisinde düzenleme yapmak mümkün değildir. Bu amaçla firma yetkilileri tablolarda kayıtlı veriler ile ilgili bilgi paylaşımında bulunmuşlardır. Analiz için verilerin hazırlanmasından önce veri tabanında tablo ve alanlar arasındaki ilişkiler çözülmeye çalışılmış bu amaçla aynı programın kurulu olduğu Akdeniz Üniversitesi Turizm Fakültesi Bilgisayar laboratuvarında program üzerinde rezervasyonlar ve günlük tur biletleri girilerek verilerin hangi tablo ve alanlara kaydedildiği izlenmiştir. Laboratuvara kurulu program ve veri tabanı sayesinde program için kullanılan veri tabının tüm tablo ve alanlarının görevleri, program menü ve alan açıklamaları takip edilerek not edilmiştir. Bunun sonucunda veri tabanı içindeki tablo ilişkileri de ortaya çıkarılmıştır.

Tablo 5.1 Seyahat Acentası Veri Tabanından Alınan Tablolar

Tablo Adı	Tabloda Kaydedilen Veri Özelliği
Rezervasyon	Otel rezervasyonu ile ilgili temel bilgiler (Tur Operatörü, Voucher No, konaklama tarihleri, kişi adedi vb.)
Ürünler	Müşterilerin aldığı otel, uçak gibi ürünler ve bu ürünler ile ilgili detaylar (Oda tipi, pansiyon, uçak, kalkış iniş havalimanı gibi)
Müşteriler	Rezervasyonu yaptıran her bir müşteri için cinsiyet, yaş, milliyet vb bilgiler
Aksiyonlar	Rezervasyon başına müşterilerin yararlandığı indirimler
Biletler	Günlük turlar için satılan biletler

Araştırma problemlerinin çözümüne katkı sağlayacak veriler turistlerin demografik ve satın alma verileridir. Bu veriler veri tabanında farklı tablolara dağıtılmış durumdadır. Günlük tur biletlerine ait veriler Biletler tablosunda yer almaktadır. Bu tabloda sadece günlük tura çıkan müşterilerin grup ismi, müşteri adedi ve bilet fiyatları ile ilgili veriler bulunmaktadır. Müşteriye ait bilgiler farklı bir tabloda tutulmakta ve aralarındaki ilişki voucher no ile sağlanmaktadır. Voucher no ise rezervasyonlara verilen bir numaradır ve müşteri bazlı takibe izin vermektedir. Müşterilere ait tüm bilgiler ise sadece müşteri yaş, unvan, cinsiyet vb. bilgilerin tutulduğu farklı bir tabloda yer almaktadır. Bu durumda tur alan ve almayan müşterilerin bir profilinin çıkartılabilmesi için, Rezervasyon, Ürünler, Aksiyonlar, Müşteriler, Biletler tablolarının alınmasına karar verilmiştir.

Rezervasyon işlemleri seyahat acentasının operasyonel işlemlerinin bel kemiğini oluşturur. Rezervasyon bilgilerine göre operasyon ve fatura bilgileri bazen de günlük tur bilgileri oluşturulur. Rezervasyon işlemlerinin kaydedildiği çok fazla tablo bulunmaktadır. Ana rezervasyon bilgileri için rezervasyon tablosu rezervasyon ile ilgili genel bilgileri içermektedir. Rezervasyon yaptıran müşterilerin bilgileri başka bir tabloda, aldıkları ürünler başka bir tabloda tutulmaktadır.

Tablo 5.2 Rezervasyon Tablosundaki Alanlar ve İçerdiği Veri

Alan Adı	Açıklama	Alan Adı	Açıklama
Sezon	Turizm sezonu	Operatör	Müşterinin tur Operatörü
Voucher	Voucher numarası	Ülke	Müşterinin geldiği ülke
Baş_Tarih	Otele giriş gün	Bit_Tarih	Otelden ayrılış günü
Kodu	Otel / Ürün kodu	Bölge	Otelin bölgesi
Oda Tipi	Oda Özelliği (Standart vb.)	Oda_Kodu	Oteli alan kişi adedi (DBL vb)
Pansiyon	Alınan hizmet türü	Oda_Sayısı	Oda sayısı
Rez_tipi	Rezervasyon Tipi (iptal, yeni)	Statü	Noshow, goshow vb.
Pax	Toplam Misafir Adedi	Yetişkin	Yetişkin Adedi
Çocuk	Çocuk misafir adedi	Bebek	Bebek adedi
Gün	Toplam konaklama günü	Kontrat	Ürün Sözleşmesi
K_Tarih	Kayıt Tarihi	K_User	Kaydeden kullanıcı
Fiyatı	Ürünün hesaplanan Fiyatı	Döviz	Kullanılan para birimi
Konfirme	Konfirme takip alanları	Açıklama	Açıklama alanları

Rezervasyon yaptıran müşterilere ait kimlik ve rezervasyon bilgileri Müşteriler tablosunda tutulmaktadır. Rezervasyon kaç müşteri var ise o kadar müşterinin bu tabloya eklenmesi gerekir. Ayrıca her müşteri için aldıkları ürün ile ilgili olarak Ürünler tablosu ile bağ kurulmaktadır.

Tablo 5.3 Müşteriler Tablosundaki Alanlar ve İçerdiği Veri

Alan Adı	Açıklama	Alan Adı	Açıklama
Sezon	Turizm sezonu	Operatör	Müşterinin tur operatörü
Voucher	Voucher numarası	Rez_SıraNo	Rezervasyondaki sırası
Adı	Müşterinin Adı	Tipi	Müşterinin Tipi (A,C,I)
Cinsiyet	Müşterinin cinsiyeti	Dil	Müşterinin konuştuğu dil
Yaş	Müşterinin Yaşı	Ülke	Müşterinin yaşadığı ülke
K_Tarih	Kayıt tarihi	K_User	Kaydeden kullanıcı

Ürünler tablosu acentanın sattığı ürün bilgilerini içermektedir. Bu tablo rezervasyon işlemlerinin altında kalan bir tablodur ve ürünler voucher no ile rezervasyona bağlıdır. Bazen hem voucher no hem de ürünün alan müşteri sıra numarası ile rezervasyona bağlanabilmektedir. Ayrıca ürünler için yapılan kontratlar ve fiyatlar farklı tablolarda tutulmaktadır. Çalışma için alınan veri setine o tablolar alınmamıştır. Tablo 5.4’de tablo içindeki alanlar listelenmiştir. Ancak, tablo içindeki ürünle ilgili alış ve satış için ayrı ayrı girilen dövizler, fatura tutarı, indirimi, vergi oranları, ödeme tutarları ve yapılan değişikliklerin saklandığı alanlar tabloya alınmamıştır.

Tablo 5.4 Ürünler Tablosundaki Alanlar ve İçerdiği Veri

Alan Adı	Açıklama	Alan Adı	Açıklama
S_Numara	Ürünün rezervasyondaki sırası	R_Numara	Rezervasyonun kayıt numarası
Kodu	Ürün kodu	Ürün_Tipi	Otel, Uçak, Tur vb.
Operatör	Müşterinin tur operatörü	Sezon	Turizm sezonu
Voucher	Voucher numarası	Baş_Tarih	Ürünün alınacağı gün
Fiyatı	Ürünün hesaplanan Fiyatı	Bit_Tarih	Üründen ayrılış günü
Oda Tipi	Ürün özelliği (Standart vb.)	Oda_Kodu	Ürünü alan kişi adedi (DBL vb)
Pansiyon	Alınan hizmet türü	Yetişkin	Yetişkin adedi
Çocuk	Çocuk misafir adedi	Bebek	Bebek adedi
Pax	Toplam misafir adedi	Gün	Toplam gün sayısı
K_Tarih	Kayıt tarihi	K_User	Kaydeden kullanıcı
Kontrat	Ürün sözleşmesi	Açıklama	Açıklama alanları

Turizm sektöründe rezervasyon fiyatlarının hesaplanması otel ile seyahat acentası arasında yapılan kontratlar kullanılarak yapılır. Sezon içerisinde satışları hareketlendirmek için bazen de sezon başında kontratlara ek olarak aksiyonlar kullanılabilir. Erken rezervasyon, yaş indirimi, balayı indirimi, gün promosyonu gibi indirimler ve bazı günlere özel paketler (yılbaşı, düğün vb.) başlıca aksiyonlardır. Müşteri faturalarının hesaplanmasında yararlanacakları indirimler Aksiyonlar tablosunda tutulmaktadır. Çalışmaya bu tablo indirimli olarak tatil yapan müşterilerin tur alma durumlarında bir fark olup olmadığının tespiti için alınmıştır.

Tablo 5.5 Aksiyonlar Tablosundaki Alanlar ve İçerdiği Veri

Alan Adı	Açıklama	Alan Adı	Açıklama
Sezon	Turizm sezonu	Operatör	Müşterinin tur operatörü
Voucher	Voucher numarası	Ürün_SıraNo	Rezervasyondaki sırası
Hesap_Tipi	Fatura hesaplama Tipi	Kontrat	Ürün sözleşmesi
Aksiyon_No	Aksiyon numarası	Aksiyon_Tipi	Balayı, Gün, EB vb.
Art/Ind.	Artış / İndirim Oranı	Hesaplama_Tipi	Aksiyonu Hesaplama Şekli
Gün	Aksiyon geçerli gün sayısı	Satış_Tarihi	Satış Tarihi
Baş_Tarihi	Aksiyon Başlama Tarihi	Bitiş_Tarihi	Aksiyon Bitiş Tarihi
Aksiyon_Fiyat	Aksiyonlu Fiyat	Açıklama	Açıklama alanları

Biletler tablosu seyahat acentasının otel rehberleri ve yurt dışında tur operatörü aracılığıyla sattığı günlük tur biletlerini içermektedir. Tablo içindeki biletle ilgili dövizler, fatura tutarı, indirimi, vergi oranları, tahsilat tutarları, komisyonlar, tur rehberleri ve konaklamalı turlar için otele ait bilgilerin saklandığı alanlar tabloya alınmamıştır.

Bazı seyahat acentaları bir müşteri için bir adet bilet kaydı girerken bazı seyahat acentaları da biletin üzerine kaç kişi olduğu notunu düşerek grup için tek bir bilet kaydı girmektedir. Çalışmadaki veri tabanında da bir grup için bir adet bilet girişi yapıldığı için bireysel olarak (yaş, cinsiyet, meslek) müşterilerin tur alma durumunun tespit edilmesi mümkün değildir.

Tablo 5.6 Biletler Tablosundaki Alanlar ve İçerdiği Veri

Alan Adı	Açıklama	Alan Adı	Açıklama
Sezon	Turizm sezonu	Tipi	Bilet Tipi
Seri	Biletin serisi	Bilet_no	Biletin numarası
Sat_Tarihi	Satış Tarihi	Tur_Tarihi	Tura çıkış Tarihi
Kodu	Turun kodu	Otel_Kodu	Konaklanılan otel kodu
Oda_No	Konaklanmakta olan oda no	Pansiyon	Alınan hizmet türü
Operatör	Müşterinin tur operatörü	Voucher	Voucher numarası
M_Adi	Müşterinin Adı	Unvan	Müşteri Unvanı (mr,mrs, grp)
Fiyat	Bilet Fiyatı	Tutar	Toplam bilet tutarı
Pax	Toplam misafir adedi	Yetişkin	Yetişkin adedi
Çocuk	Çocuk misafir adedi	Bebek	Bebek adedi
K_Tarih	Kayıt tarihi	K_User	Kaydeden kullanıcı
Rep	Bileti satan rehber	Açıklama	Açıklama alanları

İlişkisel veri tabanlarında işin yürütülmesi esnasında tekrarlanan kayıtlar ile ilgili tüm detaylar her tabloda tekrar tekrar kaydedilmez. Bunun yerine örneğin tur operatörü kaydı için, öncelikle tur operatörlerinin kayıtlarının girildiği bir tablo oluşturulmakta ve tur operatörü ile ilgili tüm bilgiler o tabloda kaydedilmektedir. Diğer tablolarla bağ kurmak için bir kod oluşturulmakta ya da kayıt numarası kullanılmaktadır. Çalışma için oluşturulan veri tabanı da ilişkisel bir veri tabanıdır. Dolayısıyla örneğin rezervasyon tablosundaki verilere bakıldığında müşteri isimleri dışında kodlar ve rakamlar görülmektedir. Oteller, tur operatörleri, uçuşlar, günlük

turlar hepsi kodlarla ifade edilmiştir. Bu kodlar farklı bir tabloda tanımlanmış ve kod ile ilgili detaylar ilgili tablolarda verilmiştir. Örneğin, rezervasyon tablosunda sadece uçuşun kodu bulunmaktadır. Uçuş ile ilgili detaylar, firması, koltuk sayısı, kalkış ve iniş saatleri, varsa aktarma bilgileri Uçuşlar ile ilgili diğer tabloda ya da tablolarda tutulmaktadır. Bu tablolar, otel ve operatörlere ait gerçek verileri içerdiğinden firma yetkilileri tarafından silinmiştir. İlgili tablolar silindiği için bölge, havaalanı, otel isimlerine veri tabanı içinde ulaşılamamıştır. Şehir isimleri havaalanı kodları ile milliyetler de milliyet kodları ile internetten araştırılarak veriye eklenmiştir. Bölge, tur ve otel kodları olduğu gibi kullanılmıştır. Tur isimleri daha sonra firma yetkililerinden kod ve isim olarak alınarak veriye eklenmiştir.

5.5.2. Verinin Birleştirilmesi ve Dönüştürülmesi

Veri tabanında yapılan incelemede beş ayrı tabloda kayıtları birbirine bağlayan alanın Voucher No olduğu görülmüştür. Rezervasyonu girilen bir müşterinin aldığı otel bilgileri, uçuşlar, yapılan operasyonlar, günlük tur biletleri tur operatöründen gönderilen voucher no ile tutulmaktadır. Bu da verilerin birleştirilmesi esnasında dayanak noktası olacaktır. Verilerin birleştirilmesinden önce tablolar tek tek incelenmiş ve hatalı olduğu düşünülen kayıtlar temizlenmiştir. Veri tabanında yapılan ilk incelemede 2016 yılı rezervasyonları olmasına rağmen günlük tur verilerinin kayıtlı olmadığı görülmüştür (veri 2016 yılında henüz günlük tur biletlerinin tamamı satılmadan yedeklenmiş olduğu için). Çıkan sonuçların doğru yorumlanabilmesi açısından önceki sezonlara ait veriler tamamen silinmiştir. 2016 yılı için günlük tur verileri eksik olduğu için otel rezervasyonu verileri de analize dâhil edilmemiştir. Sonuç olarak sadece 2015 yılına ait rezervasyon ve günlük tur verileri analiz edilmek üzere seçilmiştir.

Büyük miktarda veri kaydının olduğu ortamlarda, orijinal veri tabanında boş (null) veya tutarsız değerler olması kaçınılmazdır (Feng, 2017: 21). Yapılan incelemede bazı alanlarda boş ya da değeri sıfır olmaması gereken satırlar tespit edilmiştir. Örneğin rezervasyon tablosunda otelsiz rezervasyonlar, yetişkin adedi olmayan rezervasyonlar; Biletler tablosunda 0 “sıfır” ücretle satılmış gibi gözükten biletler bulunmaktaydı. Bu kayıtların ilgili paket programlar aracılığıyla veriye ulaşıldığında listelere yansımadağı görülmüştür. Bu sorunlu kayıtlar ağda ya da kayıt esnasında oluşan bazı hatalardan dolayı olabilmektedir. Bu gibi kayıtların hepsi ilgili tablolardan silinmiştir.

Yukarıda belirtilen sorunların dışında bazı programdan ya da kullanıcıdan kaynaklanan kayıt sorunları da tespit edilmiştir. Karşılıklı acenta programlarında rezervasyon giriş kartlarında yetişkin yaşı girmek zorunlu değildir (Akgün ve Çizel, 2017). Bununla etkisiyle rezervasyonu alınan bazı müşterilerin yaşları acenta çalışanı tarafından sisteme girilmemiştir.

Müşteri yaşları girilmeyen rezervasyonlar da silinmiştir. Ayrıca acenta programı tarafından girilmesi zorunlu tutulmasına rağmen boş kalan müşteri adı, unvanı, otel fiyatı ve uçuş kodu gibi alanlar tespit edilmiş ve ilgili rezervasyonlar silinmiştir. 2015 yılına ait 9972 adet rezervasyon Analiz_Veri ismi verilen yeni bir tabloya analize tabi tutulmak üzere alınmıştır.

Benzer sorunlar Biletler tablosunda da tespit edilmiştir. Özellikle otel rehberleri tarafından satılan günlük tur verilerinde müşteriye ait adı, kaç kişi oldukları, konakladıkları otel bilgisinin dışında bilgi girilmemiş olduğu gözlemlenmiştir. 2015 yılında çıkışı yapılacak günlük turlara ait 6233 adet biletten, müşteri ismi veya unvanı olmayan, konakladıkları otel ve hatta çıktıkları tur ismi olmayan ve voucher no girilmediği için rezervasyon ve diğer kayıtlarla bağı kurulamayan kayıtlar silinmiş ve geriye analiz için 6114 adet bilet verisi kalmıştır.

Ürünler, Müşteriler ve Aksiyonlar tabloları rezervasyon tablosuna bağımlı olarak çalıştığı için bu tablolarda herhangi bir düzenlemeye gidilmemiştir. Detayları anlatılacağı gibi, rezervasyonda bulunan sadece düzeltilmiş rezervasyon ve günlük tur kayıtlarına ait detaylar bu tablolardan alınmıştır. Bundan dolayı bu tablolarda eksik ya da sorunlu kayıt temizleme ve düzeltme işlemine gerek kalmamıştır.

Veri tabanında hatalı ve eksik kayıtlar temizlendikten sonra farklı tablolarda yer alan alanlar bir tabloda birleştirilerek analiz için kullanılacak değişkenler oluşturulmuştur. Farklı tablolarda farklı biçimlerde kaydedilen veriler SQL sorguları kullanılarak tek bir tabloya aşağıda detayları verildiği şekliyle aktarılmıştır.

Çalışma için kullanılacak asıl veriler tur biletlerinin kayıtlarının tutulduğu Biletler tablosundadır. Müşteriler hakkındaki bilgiler farklı tablolarda olduğu için bir tabloda birleştirilmeleri gerekmektedir. Başlangıçta analiz için kullanılacak değişkenlerin oluşturulacağı tabloda Biletler tablosu referans alınarak, günlük tur bileti satın alan müşterilere ait diğer verilere ilgili tablolardan ulaşılması düşünülmüştür. Yapılan incelemeler sonucunda bilet verisi ile diğer tablolar arasında müşteri bazlı değil rezervasyon bazlı (voucher no ile) ilişki kurulduğu tespit edilmiştir. Müşteri isimleri kullanılarak yapılan incelemede iki tablo arasında sadece bir müşterinin isminin ortak olduğu tespit edilmiştir. Ayrıca bu tabloda sadece satılan tur biletlerine ait bilgiler bulunmaktadır. Çalışmanın amaçlarından olan gereksiz promosyonların yapılmaması düşünülen turist profillerinin tespiti için otel satın almasına rağmen günlük tur satın almayan müşterilerin tespiti için bilet satın almayan müşteri verilerine de ihtiyaç duyulmuştur. Bu nedenlerden dolayı referans tablo olarak Rezervasyon tablosu seçilmiştir.

Rezervasyon tablosundan müşterilerin voucher numaraları, rezervasyonu satın aldıkları tur operatörü bilgileri ile seyahat acentasının otel rezervasyonları alınmıştır. Aynı tablodan müşteri adedi, çocuk sayısı, konaklama süresi bilgileri de alınmıştır. Konaklama süresi, bir

destinasyon içinde belirli bazı bölgelerin ziyaret edilmesi konusunda belirleyici (Tchetchik vd., 2009:225) olabilmektedir. Tabloda konaklama günü müşterinin kaç gün konaklayacağını belirleyen sürekli bir değişkendir. Tek tek rakamların ilişkisinin çıkma ihtimali düşük olduğu için yeni alan kategorik bir alana dönüştürülmüş ve veriler “7 günden az”, “7-13 gün konaklayanlar”, “14-20 gün konaklayanlar” ve “21 gün ve fazla konaklayanlar” şeklinde sınıflandırılmıştır.

Voucher numarası her tur operatöründen benzersiz olarak acentaya gönderilen ve müşteri ile ilgili tüm ürün ve hizmet bilgilerinin kaydında kullanılan bir numaradır. Farklı tur operatörlerinden aynı numaralar gelebilmektedir ve veri içerisinde de farklı operatörlerden gelmiş aynı numaralar mevcuttu. Analiz için yeni tablo oluşturulurken ID isimli sistem tarafından otomatik numaralandırılan bir alan eklenmiş ve analiz için bu değişken rezervasyonları temsil edecek şekilde ayarlanmıştır.

Ürünler tablosundan müşterilerin konakladığı oteller, otellerin bulunduğu bölge, otelde konaklanan pansiyon türü, müşterinin konakladığı oda tipi, uçak bilgilerinin olduğu satırlardan uçağın kalktığı havalimanı alınmıştır. Uçağa binilen havalimanı aynı zamanda müşterinin yaşadığı şehri ifade etmektedir. Turizmin sezona bağlı bir sektör olması nedeniyle, satış ve otele giriş zamanlarına göre tur alma durumunun değişip değişmediğini anlamak için rezervasyon satış tarihi, müşterinin otele giriş ve çıkış günleri alınmıştır. Çünkü günlük tur biletleri ya yurt dışında turizm paketi ile birlikte satılmakta ya da turist otele giriş yaptığı anda acentanın otel temsilcisi tarafından satılmaktadır. Otel satın alma ile otele giriş ve tur satın alma günlerine göre tur satın alma durumlarında farklılık olup olmadığı merak edilmiştir. Sonradan giriş ve çıkış günleri gün yerine sadece yıl ve ayları alınarak düzenlenmiş ve değişkenler Giriş Ayı ve Satış Ayı olarak değiştirilmiştir.

Aksiyonlar tablosunda müşterilerin yararlandıkları indirim türleri ve miktarları bulunmaktadır. Yapılan indirim türü Aksiyon değişkeni olarak yeni tabloya aktarılmıştır. Her rezervasyon için erken rezervasyon, balayı, yaş indirimi ve gün promosyonu gibi indirimler yapılan ön testlerde sonucu etkilemiştir. İndirim isimli yeni bir değişkende sadece indirim VAR/YOK şeklinde düzenleme yapılmıştır.

Müşteriler tablosunda müşteri unvanları, milliyetleri, yaşları tutulmaktadır. Müşteri unvanları M (Bay), F (Bayan), C (çocuk) ve B (Bebek) şeklinde her müşteri bir satır oluşturacak şekilde Şekil 5.4’de görüldüğü gibi kaydedilmiştir. Burada satırlar halinde girilen müşteri bilgisi Biletler tablosunda ise sadece tek satır olarak ve yetişkin, çocuk ve bebek adedi olarak kaydedilmiştir. Yapılan bu düzenlemeden sonraki müşteri verisi Şekil 5.5’de verilmiştir. Verinin veri tabanında kayıtlı olduğu haliyle analiz edilmesi mümkün olmayacaktır ya da

müşteri ile ilgili aile, çift vb. veriler görmezden gelinecektir. Yalnız daha önceki çalışmalarda aile, çift, arkadaş grupları için turizm işletmelerinden genel memnu-niyetin olumsuz belirleyicilerinin farklı olduğu tespit edilmiştir (Xu, 2017: 1676). Müşteri memnuniyeti müşteri sadakatini etkilemektedir (Geetha vd., 2017: 44). Müşteri sadakati de işletmelerin daha çok kazanç elde etmelerine neden olmaktadır. Bu bilgi ışığında veri içindeki grupların tercihlerine de bakma ihtiyacı doğurmuştur. Cinsiyet alanındaki unvanlar (M,F,C I) voucher numarası aynı olan kayıtlar için birleştirilmiştir (Şekil 5.5).

Sezon	Vouche	Rez_S	Adı	Tipi	Cinsiyet	Dil	Yaş	Ülke	MU	MI	MUS_E	IPT
S2015	213266	1	GOEKGUEL, FATİH	A	M	3	30	DE	N	1		N
S2015	213266	2	GOEKGUEL, HILAL	A	F	3	27	DE	N	1		N
S2015	213266	3	GOEKGUEL, BURCU	C	C	3	4	DE	N	1		N
S2015	213266	4	GOEKGUEL, BETUEL	C	C	3	2	DE	N	1		N
S2015	213335	1	KRAUS, VIKTOR	A	M	3	45	AT	N	1		N
S2015	213335	2	KRAUS, DANIEL	C	C	3	11	AT	N	1		N
S2015	213335	3	KRAUS, NATALIE	C	C	3	11	AT	N	1		N

Şekil 5.4 Müşteriler Tablosundaki Ham Veriler

Aynı rezervasyona ait müşterilerin cinsiyetlerinden oluşturulan unvan değişkeninden sonra veri analize tabii tutulmuştur. Sadece unvanların yer aldığı bir değişken içinde neredeyse 100 farklı unvan grubu oluşmuştur. Verinin tamamında çok farklı unvan grupları olduğu için yapılan ön testlerde sonucu etkilememiştir. Benzer unvanlar Müşteri tipi değişkeninde özetlenmiş ve analiz için tüm unvanların kullanılması yerine Müşteri tipi değişkeninin kullanılması tercih edilmiştir.

Tablo 5.7 Müşteri Tipi Değişkeni ve Düzeyleri

Müşteri Tipi	Açıklama
AİLE	Çocuk ve bebekleri ile gelen M ve F ler için
ARKDS	Aynı cinsten birden fazla müşterisi olan rezervasyonlar için 2F -2M gibi
ÇCK-B	Yanında çocukları ve/veya bebekleri olan tek misafirler için
ÇİFT	Sadece M ve F den oluşanlar için
GRUP	Adetleri tutmayan M ve F'den oluşanlar için (örneğin 3M2F)
TEK	Yanında çocukları ve/veya bebekleri olmayan tek misafirleri için

Chiang, 2017: 672), bayan ve bay yaşlarına göre alınan ürünlerin değiştiğini bulmuştur ama turizm yazılımlarında müşteri yaşları yetişkinler için girilmesi zorunlu bir alan değildir. Dolayısıyla acentada rezervasyon girenler yetişkin yaşlarını bazı rezervasyonlar için girilme-mişlerdir. Bu da analizler için kayıp veriye neden olacak ve sonuçların farklı çıkmasına neden olabilecektir. Bundan dolayı yetişkin yaşı olmayan rezervasyonlar silinmiştir. Çocuk ve bebek yaşları girilmesi zorunlu bir alan olduğundan eksik kayıt tespit edilmemiştir. Müşteri yaşlarının her bir müşteri için bağımsız olarak kullanılması bilet bazlı kayıt girildiği için mümkün olma-

diğından gruplanmıştır. Yaş gruplarını oluşturmada Eric Ericson Psikososyal Gelişim Kuramındaki yaşlar kullanılmıştır. Yetişkin yaşları 18-30, 31-59; 60++ şeklinde, çocuk yaşları 0-1,99; 2-3,99; 4-5,99; 6-11,99; 12-17,99 şeklinde gruplandırılmıştır.

Voucher	Milli	Havaala	Bölge	Ünvanla	Yet_Ya	Cocuk	Yetiskin	Cocuk	Bebek	Müşteri	Otel_Ücreti	Konak
213266	DE	DUS	BLK	MF2C	27-30	2-4	18-30	2-11	YOK	AİLE	PAHALI	14-20
213335	AT	LNZ	KMR	M2C	45	11-11	31-59	2-22	YOK	TEK	ORTA	7-13

Şekil 5.5 Analiz İçin Hazırlanan Müşteri Verisi

Yapılan ön testlerde bu şekilde gruplamanın müşterinin tur satın alma durumunu hiç açıklamadığı görülmüştür. Sonra çocuk yaşları turizm sektöründe yaygın olarak kullanılan 0 - 2; 0 - 6 ve 7 - 12 yaş gruplarından esinlenilerek bebek yaşları 0 – 1,99; çocuk yaşları da 2 – 6,99 ve 7-11,99 şeklinde yeniden düzenlenmiştir. Ayrıca çocuk adedinin de tur alımında etkili olabileceği düşünülerek Çocuk Bilgisi değişkeni oluşturulmuştur. Bu değişkende yetişkinlerin yanında bulunan çocuk adedi ve yaşı birlikte takip edilmektedir. Tablo 5.8’de çocuk bilgisi değişkeninin içeriği verilmiştir. Bu değişkende çocuk sayısı ile çocuk yaşları birlikte düzenlenmiştir. Veri seti içinde bir ailenin yanında birden fazla bebek olmadığı (toplam 4 adet rezervasyonda 2 adet bebek var) için Bebek Durumu VAR / YOK şeklinde düzenlenmiştir. Şekil 5.4’de düzenlenmemiş durumdaki 213266 ve 213335 voucher numaralı rezervasyonun müşterileri için müşteri bilgisinin düzenlenmiş hali Şekil 5.5’de görülmektedir.

Tablo 5.8 Çocuk Bilgisi Değişkeninin İçeriğinin Düzenlemesi Şablonu

Kısaltma	Açıklama (Yetişkin Yanında)
YOK	Bebek ve Çocuk YOK
1-0	Bir adet bebek var
1-1	2-6 yaş grubundan bir adet çocuk var
1-2	7-11,99 yaş grubundan bir adet çocuk var
1-3	12-18 yaş grubundan bir adet çocuk var
2-00	İki adet bebek var
2-01	Bir adet bebek ve 2-6 yaş grubundan bir adet çocuk var
2-11	2-6 yaş grubundan iki adet çocuk var
2-12	Birisi 2-6 yaş grubundan, diğeri 7-11,99 yaş grubundan toplam iki çocuk var
2-13	Birisi 2-6 yaş grubundan, diğeri 12-18 yaş grubundan iki çocuk var
2-22	7-11,99 yaş grubundan iki adet çocuk var
2-23	Birisi 7-11,99 yaş grubundan, diğeri 12-18 yaş grubundan iki çocuk var
3++	3 veya daha fazla bebek ya da çocuk var

Günlük Tur tablosunda müşterilerin isim, yaş, milliyet aldıkları tur, konakladıkları otel bilgileri tutulmaktadır. Bu veriler diğer tablolardan elde edildiği için bu tablodan alınmamışlardır. Veri seti içindeki tabloların birleştirilmesi ve bazı verilerin dönüştürülmesi ile aşağıdaki 23 adet bağımsız değişken oluşturulmuştur.

ID, Operatör, Milliyet, Havaalanı, Şehir, Bölge, Otel, Otel Ücreti, Oda Tipi, Pansiyon, Çocuk Sayısı, Satış Ayı, Aksiyon, İndirim, Müşteri Adedi, Giriş Ayı, Çıkış Ayı, Unvanlar, Yetişkin Yaşı, Çocuk Bilgisi, Bebek Durumu, Müşteri Tipi, Konaklama Süresi

23 adet bağımsız değişken otel rezervasyon kaydı girilen turistlerin günlük tur satın alma durumlarını tahmin etmek için kullanılacaktır. Turistlerin rezervasyon ve günlük tur kaydında ortak veri voucher numarasıdır. Bu veri baz alınarak rezervasyon kaydı bulunan bir voucher no, biletler tablosunda da kayıtlı ise, bu ilgili turistin günlük tur satın aldığı, voucher numarası rezervasyon tablosunda olmasına rağmen biletler tablosunda yoksa, tur satın almadığı anlamına gelmektedir. Voucher numaraları hem rezervasyon hem de biletler tablosunda en az bir defa kayıtlı olan turistler “SATIN ALAN”, voucher numarası rezervasyon tablosunda olmasına rağmen biletler tablosunda olmayan turistler de “ALMAYAN” olarak sınıflandırılmıştır. TUR ismi verilen bu değişken araştırmanın ana probleminin çözümü için bağımlı değişken, diğer değişkenler de bağımsız değişkenleridir.

Günlük turlar düzenlenişi, etkinlikleri ve katılımcıları açısından birbirinden farklıdır. Tur özelliğine göre tura katılan turist özellikleri de değişebilmektedir. Hatta turistler katılmayı düşündükleri turların düzenlendiği bölgelerden otel rezervasyonu yaptırmaktadır. Örneğin golf oynamak isteyen bir turist golf oynayabileceği bir bölgede otel rezervasyonu yaptırmaktadır. Bu bilgiler ışığında her günlük tur için satın alınma durumlarının tahmin edilmesi fikri ortaya çıkmıştır. Acenta kayıtlarında Alışveriş turu, Hamam turu, Rafting turu, Gezi ve Kültür turları, Gece / Eğlence turu ve Tekne/Yat turu için satılmış biletler bulunmakta idi. Otel rezervasyonu olan bir turist bu turları satın alma durumunun tahmin edilebilmesi her günlük tur için bir bağımlı değişken oluşturulmuştur.

Tur değişkeninin de herhangi bir turu satın alan turist, bir adet tur satın almışsa da, 6 farklı günlük tur satın almışsa da “SATIN ALAN” olarak sınıflandırılmıştır. Alt problem bağımlı değişkenlerinde ise rezervasyon kaydı girilen turist belirlenen bir günlük turu satın alma durumunun tahmin edileceği için sınıflandırma değişkeni o günlük turu satın alma ya da almama durumuna göre yeniden düzenlenmiştir. Veri setinde kayıtlı 6 farklı günlük tur (Hamam, Alışveriş, Gezi, Tekne, Eğlence ve Rafting) için 6 farklı sınıflandırma değişkeni oluşturulmuştur. Müşteri ilgili turu satın almışsa değişken içeriği “SATIN ALAN”, satın almamış ise “ALMAYAN” olarak düzenlenmiştir. Örneğin Gezi değişkeni içinde bir turist, gezi

ve kültür turlarından birisini satın alması durumunda “SATIN ALAN”, almaması durumunda da “ALMAYAN” olarak sınıflandırılmıştır.

Yapılan temizleme, birleştirme ve dönüştürme işlemleri sonucu oluşan değişkenlerin hepsi kategorik değişken haline gelmiştir. Yaş, konaklama günü gibi değişkenlerde değerlerin gruplandırılması sonucunda kategorik değişken haline dönüşmüşlerdir. Bu durum yapılacak analiz için sorun değil, bilakis daha uygun olmuştur. Çünkü sınıflandırma, sayısal değişkenlerden ziyade kategorik değişken kullanılarak yapılmaktadır (Septiadi vd., 2016: 211).

Çalışmada sınıflandırma algoritmalarının çalıştırılması için Rapidminer 9.1 programı kullanılmıştır. Rapidminer’de arff, clm, aml, bib, cms, cri, csv, mat, log, dat, mod, obf, bar, per, res, sim, thr, wgt, wls, xrff dosyalarından veriler alınarak anlamlı bilgiye ulaşılabilir (Çelik vd., 2017: 23). Microsoft Access veritabanında hazırlanan tablo “CSV” dosyası olarak Rapidminer’de okunabilmesi için Microsoft Excel çalışma kitabı olarak kaydedilmiştir. Dosya işlemlerin daha hızlı olabilmesi için Rapidminer’de Add Data komutu ile Rapidminer deposuna (Repository) yüklenmiştir.

5.5.3. Değişken Seçimi

Değişken seçimi, boyut küçültme ile yakından ilgili bir sorundur. Değişken seçiminin amacı, veri kümesindeki değişkenleri önemli olarak tanımlamak ve diğer tüm değişkenleri ilgisiz ve fazla bilgi olarak atmaktır. Değişken seçimi, verilerin boyutsallığını azalttığı için, veri madenciliği algoritmalarının daha etkili ve hızlı çalışma olasılığını ortaya koymaktadır (örneğin, Veri Madenciliği algoritmaları, değişken seçimi kullanılarak daha hızlı ve daha etkili şekilde çalıştırılabilir) (Maimon ve Rokach, 2010: 84). Değişken seçimi ile bazen yapılacak sınıflandırmadaki doğruluk oranı arttırılabilir, bazen de, daha etkili, özlü ve kolay yorumlanabilen bir sonucun elde edilmesi sağlanabilmektedir

Öte yandan, değişken seçimi zor bir işlemdir ve aynı zamanda, maksimum doğruluk elde etmek için tüm bilgilerin (yani değişkenlerin) gerekli olduğu varsayımına aykırıdır. Bazı değişkenler daha az önemli olsa da, ilgisiz veya gereksiz olanı yoktur. Makine öğrenimi alanında değişkenlerin seçimini sağlayacak algoritmalar paketleyici (Wrappers), Filtreler ve gömülü yöntemler olmak üzere üç ana grupta incelenmektedir (Akpınar, 2014: 140). Makine öğreniminde değişken seçimine ilk yaklaşımlar filtre yöntemleridir. Tüm filtre yöntemleri, özellik altkümelerinin değerini değerlendirmek için bir öğrenme algoritmasından ziyade, verilerin genel özelliklerine dayanan sezgisel taramaları kullanmaktadır. Sonuç olarak, filtre yöntemleri genellikle paketleyici yöntemlerinden çok daha hızlıdır ve bu nedenle yüksek boyut-sallıkta veri kullanımı için daha pratiktir (Maimon ve Rokach, 2010: 85).

Tablo 5.9 Veri Setindeki Değişkenler

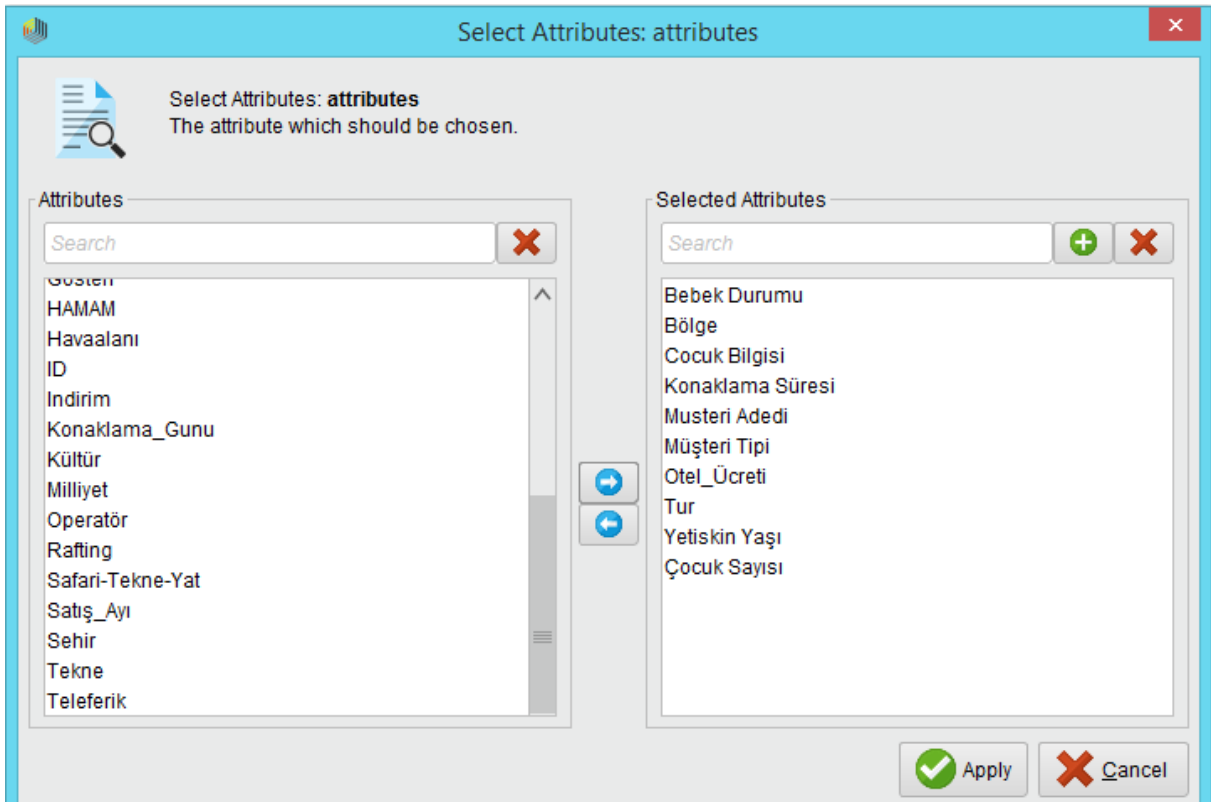
Değişken	Türü	Açıklama
ID	Kayıt No	Verinin kayıt sırası
Operatör	Bağımsız	Turisti yurt dışından gönderen seyahat acantası
Milliyet	Bağımsız	Turistin milliyeti
Havaalanı	Bağımsız	Turistin uçağa bindiği havaalanı
Şehir	Bağımsız	Turistin geldiği şehir
Bölge	Bağımsız	Turistin konaklama yaptığı bölge
Otel	Bağımsız	Turistin konakladığı otel
Otel Ücreti	Bağımsız	Turistin konakladığı otelin fiyatı
Oda Tipi	Bağımsız	Turistin konakladığı oda tipi
Pansiyon	Bağımsız	Turistin konakladığı pansiyon türü
Çocuk Sayısı	Bağımsız	Rezervasyondaki toplam çocuk sayısı
Satış Ayı	Bağımsız	Rezervasyonun satıldığı yıl ve ay
Aksiyon	Bağımsız	Turistin yaralandığı aksiyon
İndirim	Bağımsız	Turistin otel ücreti için indirim alıp almadığı
Müşteri Adedi	Bağımsız	Rezervasyondaki toplam turist adedi
Giriş Ayı	Bağımsız	Turistin otele giriş yaptığı yıl ve ay
Çıkış Ayı	Bağımsız	Turistin otelden çıkış yaptığı yıl ve ay
Unvanlar	Bağımsız	Rezervasyondaki tüm kişilerin ünvanları
Yetişkin Yaşı	Bağımsız	Rezervasyondaki yetişkinlerin yaş grupları
Çocuk Bilgisi	Bağımsız	Rezervasyondaki çocuk adedi ve yaşları
Bebek Durumu	Bağımsız	Rezervasyonda bebek kaydının olup olmadığı
Müşteri Tipi	Bağımsız	Rezervasyon yaptıran turistlerin durumları
Konaklama Süresi	Bağımsız	Turistin otelde konaklama süresi
Alışveriş	Bağımlı	Alış-veriş turunu satın alma durumu
Hamam	Bağımlı	Hamam turunu satın alma durumu
Rafting	Bağımlı	Rafting turunu satın alma durumu
Gezi	Bağımlı	Gezi ve kültür turunu satın alma durumu
Eğlence	Bağımlı	Eğlence turunu satın alma durumu
Tekne	Bağımlı	Tekne/Yat turunu satın alma durumu
Tur	Bağımlı	Herhangi bir turu satın alma durumu

Hazırlanan 23 bağımsız değişkenin tamamının dâhil edildiği bir sınıflandırma modeli büyük bir ihtimalle en iyi sonucu vermeyecektir. Onun için bu değişkenler arasından en iyi sonucu verecek, işletmenin en kısa sürede ilgili müşteriye ulaşmasını sağlayacak model için değişkenlerin tespit edilmesi gerekmektedir. Çünkü veri madenciliği sürecine küçük bilgi içeren veri öğeleri eklemek aslında veri tabanının analiz gücünü azaltacaktır. Alakasız bir veri maddesinin eklenmesi veya aynı öğenin birden fazla ölçümünün yapılması, veri madenciliği sonuçlarının yararlılığını azaltabilmektedir (Law, 1998: 63).

Veri seti incelendiğinde Milliyet gibi bazı değişkenlerin %90'a yakın oranlarda aynı kayıtları içerdiği tespit edilmiştir. Voucher No değişkeni gibi birkaç değişkenin de her satırda farklı bir değer içerdiği tespit edilmiştir. Örneğin, verilerdeki Havaalanı değişkeninden misafirlerin geldiği Şehir değişkeni oluşturulmuştur. Fakat genelde bir şehirde sadece bir havaalanı olduğundan (kayıtlarda sadece Berlin'de iki ayrı havaalanı gözüküyor), bu iki değişken neredeyse birbirinin aynısı durumunda kalmıştır. Bu ve benzeri durumlar değişkenlerin tahmin gücünü düşürmektedir.

Değişken seçimi için literatürde yer alan filtreleme, paketleyici ve gömülü yöntemler kullanılmıştır. İlk önce filtreleme yöntemlerinden Pearson korelasyon katsayısına bakılmıştır. Korelasyon katsayıları incelendiğinde bazı değişkenlerin birbirleri ile ilişkilerinin çok yüksek olduğu görülmüştür. Bu değişkenlerden birisi analize alınmayacaklar arasına alınmıştır ancak, müşteri adedi ile çocuk sayısı değişkenlerinin ilişkileri yüksek olmasına rağmen analize alınmalarına sezgisel olarak karar verilmiştir. Çünkü yetişkinlerin yanında bulunan çocuk sayısının hem olumlu (çocuk gezmek isteyebilir) hem de olumsuz (pahalı olabilir) yönlerden günlük tur satın almaya etkisinin olabileceği düşünülmüştür.

Filtre yöntemi ile adı geçen iki değişken dışında analize alınmayacak değişkenler net olarak kararlaştırılmış olmasına rağmen veri setinde karar almayı az etkileyecek daha başka değişkenlerin de olması olasıdır. Sezgisel yöntem hızlı ve pratik olmasının yanında geride her zaman bir acaba bırakmaktadır. Analizden atılan bir değişkenin değerli bir bilgi içermesi, analize alınan bir değişkenin sonuçları yanlış yönlendirme ihtimali her zaman olabilmektedir. Değişken seçimi için ikinci olarak paketleyici yöntemlere başvurulmuştur. Bu yöntemlerden adım adım geriye doğru eleme yöntemi kullanılmıştır. Sonucu veri madenciliği teknikleri açısından da takip edebilmek için algoritmaların değişkenleri belirlemek için oluşturulması esnasında oluşturulan 23 bağımsız değişken ve Tur bağımlı değişkeni analize katılmıştır.



Şekil 5.6: Adım Adım Geriye Doğru Eleme Yöntemiyle Seçilen Değişkenler

Değişkenleri test edebilmek için sınıflandırma tekniklerinden Karar ağaçları, Bayes sınıflandırma, K-NN sınıflandırma, Yapay sinir ağları, Regresyon ve Lojistik regresyon analizlerinden oluşan 11 adet model kurulmuştur. Her modele hazırlanan verideki tüm değişkenler istisnasız olarak dâhil edilmiştir. Değişkenler sırayla modelden çıkartılarak ve tekrar eklenerek sınıflandırma hatası ve doğruluk oranları kontrol edilmiştir. Değişken modelin doğru sınıflandırma yapmasına katkıda bulunursa modelde tutulmuş ve doğru tahmin yüzdesinin düşmesi durumunda elenmiştir. Bu işlem tüm sınıflandırma algoritmalarında tek tek denenmiş ve tüm algoritmalar için ortak değişkenler tespit edilmiştir. Çalışmada sonucu analize tabi tutulmak üzere seçilen değişkenler Şekil 5.6'de sağ tarafta görülmektedir.

Değişken seçiminde kullanılan yöntemlerden birisi de gömülü yöntemlerdir. Hesaplanabilirlik karmaşıklığı açısından paketleyici ve filtre kullanan yöntemler arasında yer alan gömülü yöntemler, optimal değişken seçiminde doğrudan öğrenme algoritmasına bağlıdır (Akpınar, 2014: 143). Bu amaç için karar ağaçları, ağırlıklaştırılmış naive bayes yaklaşımı, destek vektör makineleri kullanılabilir.

Değişken seçimi yapabilmek için ikinci bir yöntem olarak karar ağaçlarından yararlanılmıştır. Karar ağacı algoritması içeren bir model kurularak ve 24 değişkenin sınıflandırma değişkenini tahmin etmedeki performansına bakılmıştır. Karar ağaçları, çok sayıda aday giriş değişkeninin bir hedef değişkeni ile olan ilişkilerini anlamak için verileri araştırmak için de faydalıdır. Karar ağaçları hem veri keşiflerini hem de modellemeyi bir araya getirdiği için, diğer bir tekniği kullanarak nihai modeli oluştururken bile modelleme sürecinde güçlü bir ilk adımdır (Berry ve Linoff, 2004: 165).

Karar ağacı modeli ile 23 değişken için ayrı ayrı ve tek olarak karar ağacı modeli çıkartılmış ve her bir değişkenin bağımlı değişkeni (Tur) tahmin oranları not edilmiştir. Tablo 5.10'da değişkenlerin tek başına alındıkları modelin tahmin performansları verilmiştir.

Tablo 5.10 Değişkenlerin Teker Teker Analize Alınması Durumundaki Model Performansı

Değişken	Doğruluk	Sınıf Kesinliği(%)		Sınıf Hassasiyeti(%)	
		SATIN ALAN	ALMAYAN	SATIN ALAN	ALMAYAN
Operatör	70,78	0	70,78	0	100
Milliyet	71,18	56,25	71,67	6,17	98,02
Şehir	70,68	47,73	71,19	3,60	98,37
Bölge	82,51	67,73	89,81	76,67	84,92
Otel	70,78	0	70,78	0	100
Otel Odası	70,58	48,00	71,77	8,23	96,32
Otel Ücreti	71,13	57,78	71,44	4,46	98,65
Pansiyon	70,78	0	70,78	0	100
Çocuk Sayısı	70,78	50,00	70,86	0,69	99,72
Satış Ayı	70,83	100	70,81	0,17	100

Aksiyon	70,83	100	70,81	0,17	100
Müşteri Sayısı	70,83	52	71,07	2,23	99,15
Giriş	70,78	0	70,78	0	100
Unvanlar	70,78	0	70,78	0	100
Yetişkin Yaşı	85,96	66,72	87,24	66,72	93,91
Çocuk	71,03	61,90	70,12	2,23	99,43
Bebek Durumu	70,78	0	78,78	0	100
Müşteri Tipi	70,78	0	70,78	0	100
Konaklama Süresi	70,78	0	78,78	0	100

Tablo 5.10’da da görüldüğü üzere kontrollü olarak oluşturulan tüm modellerde modelin doğruluk oranına katkı sağlayan Çocuk Sayısı, Müşteri Sayısı, Bebek Durumu, Müşteri Tipi ve Konaklama Süresi karar ağaçları ile yapılan ön testte etkili olmamışlardır. Bu değişkenler tek tek test edilmek yerine ikişer ikişer modele dâhil edilerek doğru tahmin oranları kontrol edilmiştir. Tek başına etkisi olmayan bir değişken, diğer değişkenler ile birlikte modele dâhil edildiğinde sınıf değişkenini tahmin performanslarında belirgin bir yükseliş olmuştur. Tablo 5.11’de bu değişkenlerden doğru tahmin oranı %80’nin üzerinde olanlar verilmiştir.

Sezgisel yöntemle yapılan azaltma işleminde analize alınmasına karar verilen Bebek, Çocuk sayısı ve Müşteri adedi karar ağaçları ile yapılan çalışmaların sonucunda analiz dışına alınmıştır. Aslında Çocuk bilgisi değişkeni bebek bilgisini, ayrıca çocuk sayısı bilgisini yaşları ile birlikte içermektedir. Dolayısıyla neredeyse aynı veriyi içeren farklı değişkenlerin ikisinin de analize girmesi sonucu olumsuz etkileyebilecektir. Müşteri adedi değişkeni ile yine çocuk bilgisinde yer alan çocuk adetleri ile doğru orantılı bir değişkendir. Çocuk adedi arttıkça, doğal olarak müşteri adedi de artmaktadır.

Tablo 5.11 Değişkenler İkişer - İkişer Analize Alındığında Model Performansı

Değişken	Doğruluk	Sınıf Kesinliği(%)		Sınıf Hassasiyeti(%)	
		SATIN ALAN	ALMAYAN	SATIN ALAN	ALMAYAN
Bölge + Çocuk Bilgisi	82,91	64,90	95,27	90,39	79,82
Müşteri Tipi +Yetişkin Yaşı	85,91	85,12	86,13	62,78	95,47
Bölge +Yetişkin Yaşı	89,97	84,50	92,08	80,45	93,91
Bölge + Müşteri Tipi	85,06	68,29	91,81	91,25	82,51
Çocuk Sayısı+ Bölge	82,51	64,23	95,31	90,57	79,18
Müşteri Sayısı + Bölge	81,85	62,75	96,54	93,31	72,12
Satış Ayı + Bölge	81,60	63,37	94,02	87,82	79,04
Bebek Durumu + Bölge	81,50	61,68	98,33	96,91	75,14
Konaklama A. +Yetişkin Yaşı	85,96	81,89	87,24	66,72	93,91
Yetişkin Yaşı + Otel Ücreti	86,07	83,22	86,91	65,52	94,55
Yetişkin Yaşı + Çocuk Bilgisi	86,57	82,61	87,83	68,44	94,05

Bu çalışmalar sonucu aşağıdaki değişkenler ile model kurulmasına karar verilmiştir. Diğer değişkenler sonuçlar için belirleyici olmadıkları için analiz dışında tutulmuşlardır. Bebek

Durumu, Çocuk Sayısı değişkenlerinin yerine, bu iki değişkene ait verileri de içinde barındıran Çocuk bilgisi değişkeninin analizlerde kullanılmasına karar verilmiştir. Sonuç olarak rezervasyon kaydı bulunan turistlerin tur alma durumunu en iyi açıklayan aşağıdaki değişkenler analize dâhil edilmek üzere seçilmiştir.

<u>Bağımsız Değişkenler</u>	<u>Bağımlı Değişken</u>
Müşteri Tipi	Tur
Yetişkin Yaşı	
Çocuk Bilgisi	
Konaklama Süresi	
Otel Ücreti	
Bölge	

5.7. Verilerin Analiz Süreci

Çalışmanın bu aşaması iki adımda gerçekleştirilmiştir. İlk adımda rezervasyon yaptıran turistlerin tur satın alıp almadığının tahminini yapan sınıflandırma yöntemlerinin performanslarının değerlendirilebilmesi için 11 ayrı tahmin modeli oluşturulmuştur.

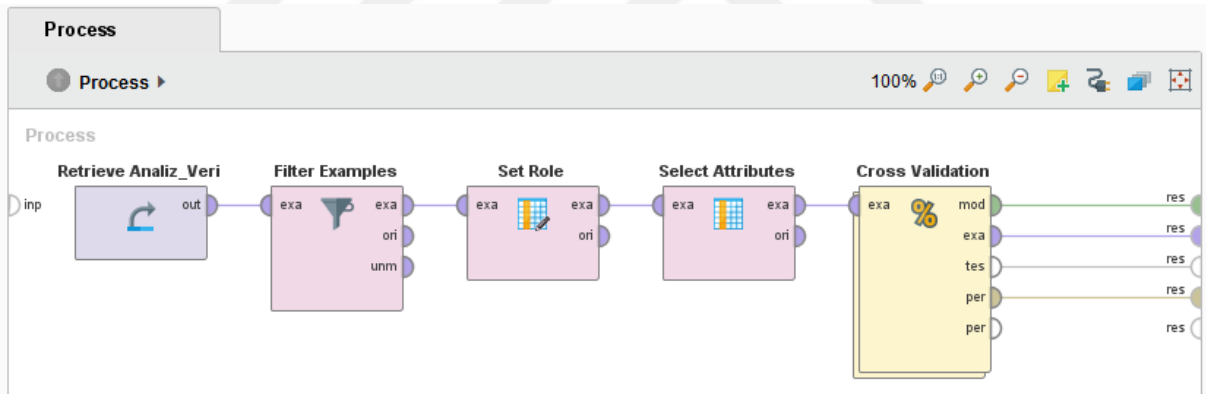
11 modelin performansları değerlendirildiğinde en iyi performansın C4.5 karar ağacı sınıflandırma algoritmasının olduğu görülmüştür. İkinci adımda bu algortiritmanın kullanıldığı yeni bir model oluşturulmuştur. Bu modelde aynı veri seti farklı sınıflandırma değişkenleri ile kullanılmış ve altı farklı sınıflandırma değişkeni (günlük tur) için altı defa çalıştırılmış ve sonuçları not edilmiştir.

Çalışmada sınıflandırma algoritmalarının çalıştırılması için Rapidminer 9.1 programı kullanılmıştır. RapidMiner, Almanya'nın Dortmund kentindeki Rapid -I, GmbH tarafından 2001 yılında geliştirilen ticari bir popüler veri madenciliği açık kaynak aracıdır. RapidMiner, makine öğrenmesi, veri madenciliği, metin madenciliği, tahmine dayalı analitik ve iş analitiği için entegre bir ortam sağlayan bir yazılım platformudur. Araştırma, eğitim, öğretim, prototip ve uygulama geliştirme ile iş ve endüstriyel uygulamalar için kullanılmaktadır. Sonuç görselleştirme, doğrulama ve optimizasyon dahil olmak üzere veri madenciliği sürecinin tüm adımlarını birleştirmektedir. Tahmine dayalı analitik için kullanımı kolay bir görsel ortamdır. Analiz işlemlerinin tasarımı için en güçlü ve sezgisel grafik kullanıcı arayüzüdür. Güçlü ve açık uzantı API'lerini, 3-D grafiklerini, dağılım matrislerini, kendi kendini düzenleyen haritalarını kullanarak uzman algoritmalar kolayca entegre edebilmektedir. RapidMiner'in gelişmiş motoru, maksimum görsel etki için yakınlaştırma, kaydırma ve yeniden ölçeklendirme desteğiyle verilerin tamamen özelleştirilebilir, dışa aktarılabilir tablolara dönüştürülmesini sağlamaktadır.

5.7.1. Modellerin Performans Analizleri

Çalışmanın ana problemi, seyahat acentasında otel rezervasyon kaydı bulunan turistlerin günlük tur satın alma durumlarını tahmin eden sınıflandırma yöntemlerinin performanslarını karşılaştırmaktır. Bunun için K- en yakın komşu, ID3, CHAID, C4.5, Rastgele Orman, Gradyan Arttırılmış Karar ağaçları, Rastgele Ağaç, Naive Bayes, Derin öğrenme, Genelleştirilmiş Doğrusal Model ve Lojistik regresyon algoritmalarının ayrı ayrı koşturulduğu 11 adet model oluşturulmuştur.

Rapidminer’de model “Design” ekranında kurulmakta ve bir süreç (process) şeklinde sırayla yürütülmektedir. Şekil 5.7’de görülen model için de Design ekranı kullanılmıştır. Retrieve operatörü ile Rapidminer deposuna (Repository) yüklenmiş olan verinin sürece katılması sağlanmıştır. Set Role operatörü içinde TUR değişkeninin LABEL olacağı yani sınıflandırmaya baz alınacak sınıfları içerdiği belirlenmiştir. Select Attributes operatörü ile veri seti içinden değişkenlerin analize dâhil edilip edilmeyecekleri seçilmiştir. Cross Validation operatörü ile veri seti eğitim ve test olmak üzere iki kısma ayrılmaktadır. Kullanılan algoritma eğitim kısmında öğrenecek ve test kısmında ise öğrenmeye ilişkin performans ölçülecektir.



Şekil 5.7 Sınıflandırma Modelleri Süreci

Çapraz geçerlilik, bir makine öğrenmesi modelinde yapılan testin hatasını daha iyi tahmin edebilmek için model seçiminde kullanılan bir tekniktir. Çapraz doğrulamanın arkasındaki fikir, eğitim verileri setinden doğrulama kümeleri olarak bilinen örnek gözlem bölümlerini oluşturmaktır. Bir modeli eğitim verilerine yerleştirdikten sonra, performansı, her yeni doğrulama kümesine karşı ölçülmekte ve daha sonra, yeni gözlemleri öngörmek istenildiğinde modelin nasıl performans göstereceğine ilişkin iyi bir değerlendirme elde edilmektedir.

Modellerde veri kümesi 10 adet alt kümeye bölünmüştür. Her alt kümede eşit sayıda örnek vardır. Ayrıca gerçekleşecek olan yineleme sayısı alt küme sayısı ile aynıdır. Her defasında, 10 alt kümeden biri test kümesi olarak kullanılırken diğer 9 alt küme bir eğitim

kümesi oluşturmak üzere bir araya getirilmiştir. Ardından, 10 denemenin ortalama hatası hesaplanmıştır. Bu yöntemin avantajı, verilerin nasıl bölündüğünü daha az önemsemektir. Her veri noktası tam olarak bir kez test kümesine ve 9 kez bir eğitim setine girmektedir. Sonuç tahmini varyansı, bölünme (alt küme) adedi arttıkça azaltılır. Bu yöntemin dezavantajı, değerlendirme yapmak için 10 kere hesaplama gerektirmektedir.

Çapraz Geçerlilik Operatörü, alt kümeleri oluşturmak için birkaç örnekleme türü kullanılabilir. Kullanılan programda doğrusal örnekleme, rastgele örnekleme ve tabakalı örnekleme seçenekleri mevcuttur. Doğrusal örneklemede örneklerin sırası değiştirilmeden veri seti bölümlere ayrılır. Ardışık örneklerle alt kümeler oluşturulur. Rastgele örneklemede altkümeler yapmak için rastgele örnekler seçilir. Tabakalı örnekleme rasgele altkümeler oluşturur. Alt gruplardaki sınıf dağılımının (Öznitelik etiketi ile tanımlanmıştır) tüm bölümlerde aynı olmasını sağlar. Örneğin, bir binominal sınıflandırma durumunda, tabakalı örnekleme, her bir alt kümenin aşağı yukarı özellik etiketinin iki değerinin aynı oranlarını içerecek şekilde rasgele altkümeler oluşturmaktadır.

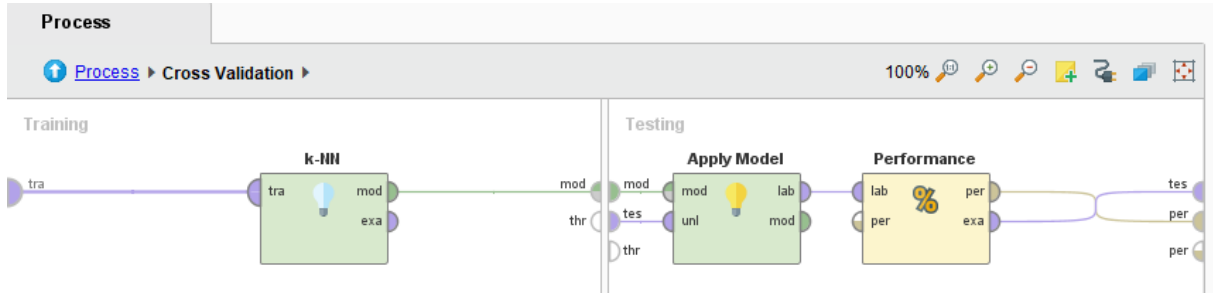
Model kurulumu esnasında kullanılacak algoritma Çapraz Geçerlilik (Cross Validation) operatörü içine yerleştirilmektedir. Şekil 5.7’de verilen model tüm modellerde kopyalanarak kullanılmış, dolayısıyla tüm algoritmaların aynı koşullarda performanslarına bakılmıştır. Ancak, her algoritmanın yapısına göre farklı parametreler alabilmektedir. Algoritmalara ait parametrelerde yapılan değişiklikler ilgili modelin anlatımında açıklanmıştır.

5.7.1.1. En Yakın Komşu

En yakın komşu, işlevin sadece yerel olarak yaklaştırıldığı ve tüm hesaplamaların sınıflandırmaya kadar ertelendiği bir örnek tabanlı öğrenme veya tembel öğrenmedir. K-en yakın komşu algoritması, tüm makine öğrenimi algoritmalarının en basitlerinden biridir. Sınıflandırılacak verinin komşularıyla olan uzaklığına bakmaktadır ve en uygun etiket ile sınıflandırma işlemini yapmaktadır. Sınıflandırmada aktif rol alan komşuların sayısı k ile gösterilmektedir. Algoritma son derece basit olmasına rağmen k değerinin belirlenmesi algoritmanın etkinliği açısından önemlidir (Akpınar, 2014:232). En yakın komşu algoritması için oluşturulan modelde k değeri 1 iken % 88,29 oranında doğru sınıflandırma yapmış ve en yüksek oranda doğru sınıflandırmayı $k=10$ iken % 92,14 oranında yapmıştır. Doğru sınıflandırma oranları $k=1$ den $k=10$ kadar artarken, $k=11$ den sonra doğru tahmin oranları tekrar düşmeye başlamıştır.

K en yakın komşu algoritması, tüm (veya seçilen) eğitim örneklerini saklar ve test örneğinin sınıfını tahmin ederken saklanan eğitim örnekleriyle bir test örneğinin benzerliğini

ölçmek için mesafe işlevini kullanmaktadır. Mesafeyi ölçmek için programda nominal, sayısal, bunların karışımı ve Bregmann sapmaları seçenekleri mevcuttur. Veri türünden dolayı mesafe ölçümü karışım ya da nominal seçeneklerden Öklid uzaklığı ile ölçülmüştür. Veri setinde sayısal ve Bregmann sapmaları ölçme seçenekleri ile mesafe ölçümü yapılamamaktadır.

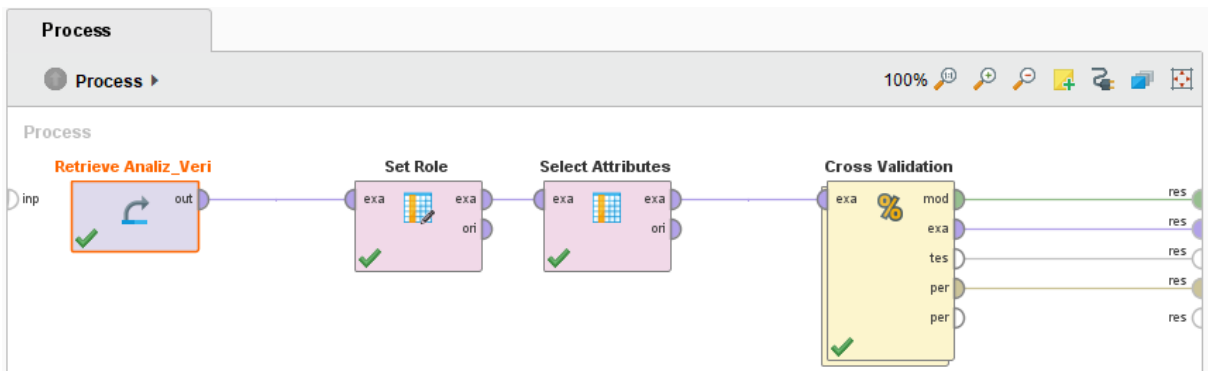


Şekil 5.8 K- En Yakın Komşu Algoritmasının Performansının Değerlendirilmesi

Çapraz geçerlilik için tabakalı örnekleme yapıldığında turistlerin tur satın alma tahmin doğruluğu % 92,11, (4 saniyede) rastgele örnekleme yapıldığında ise % 92,10 (4 saniyede) olmaktadır. En yakın komşu algoritması en iyi performansı 10 katlı doğrusal çapraz geçerlilikte % 92,14 (5 saniye) olarak vermiştir. K- en yakın komşu algoritması veri setinde bulunan toplam 2913 adet SATIN ALAN sınıfının 2436 adedini doğru olarak, 477 adedini de yanlış olarak tahmin etmiştir. Toplam 7059 adet olan ALMAYAN sınıfının da 6752 adedini doğru olarak ve 307 adedini de yanlış olarak tahmin etmiştir.

5.7.1.2. Karar Ağaçları

RapidMiner yazılımında 9 adet karar ağacı algoritması bulunmaktadır. Bu algoritmalar ID3, CHAID, C4.5, Rastgele Orman, Gradyan Arttırılmış Karar ağaçları, Rastgele Ağaç algortimalarının kullanıldığı altı adet model oluşturulmuş ve performansları aşağıda verilmiştir. Modellerde çapraz geçerlilik operatörünün eğitim kısmına ilgili algoritma operatörü yerleştirilerek model çalıştırılmıştır.



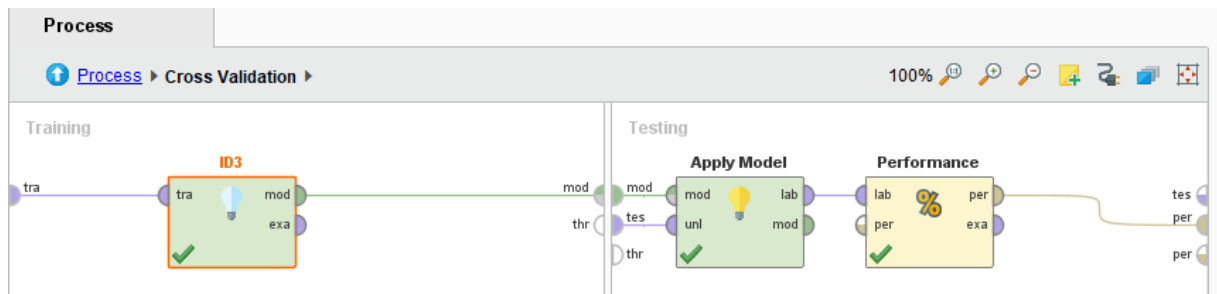
Şekil 5.9 Karar Ağaçları Modeli

Bir karar ağacı, bir dizi basit karar kuralını uygulayarak, geniş bir kayıt koleksiyonunu birbirini izleyen küçük kayıt kümelerine bölmek için kullanılabilecek bir yapıdır. Her ardışık bölünme ile sonuç kümelerinin üyeleri birbirine daha fazla benzer hale gelmektedir (Berry ve Linoff, 2004:166). Bölme için hangi değişkenin seçileceği kararını vermek için tüm karar ağacı modellerinde bilgi kazanımı (Information gain) kullanılmıştır. Bilgi kazanımı (Information gain) için tüm değişkenlerin entropileri hesaplanmakta ve bölünme için en az entropiye sahip olan seçilmektedir.

Bir ağacın derinliği, veri setinin boyutuna ve özelliklerine bağlı olarak değişmektedir. Modellerde karar ağacının derinliği 5 düğüm ile sınırlandırılmıştır. İstenirse bu değer '-1' olarak ayarlanarak diğer durma kriterleri yerine getirilinceye kadar düğüm oluşumuna izin verilebilmektedir. Bir düğümün kazancı bölünmeden önce hesaplanmaktadır. Düğümün kazancı minimum kazançtan büyük olduğunda düğüm bölünmektedir. Modellerde minimum kazanç %10 seçilmiştir. Daha yüksek bir minimum kazanç değeri, daha az bölünme ve dolayısıyla daha küçük bir ağaçla sonuçlanabilir. Çok yüksek bir değer, bölünmeyi tamamen önler ve tek düğümlü bir ağaç oluşturulur. Modelde %20'den sonra tek bir düğüm oluşmaktadır. Bir yaprağın boyutu, alt kümesindeki örneklerin sayısıdır. Ağaç, her yaprağın en azından minimum yaprak büyüklüğü sayısına sahip olmasını sağlayacak şekilde üretilir. Modellerde minimum yaprak büyüklüğü 3 olarak seçilmiştir.

ID3 Algoritması

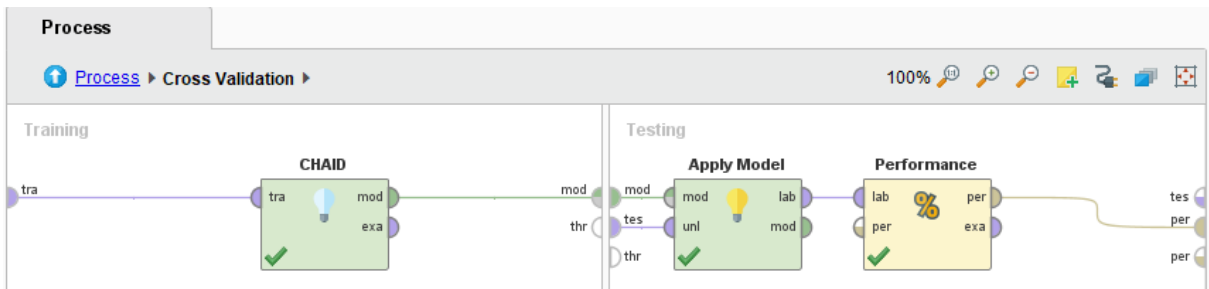
ID3 algoritmasının kullanıldığı karar ağacı modelinde en yüksek doğruluk oranı % 92,15 (10 kat tabakalı geçerlilikte) olmaktadır. 10 kat rastgele geçerlilikte ise % 91,99'a ve doğrusal örnekleme % 91,88'e düşmektedir. Bu durumda ID3 algoritması veri setinde bulunan toplam 2913 adet SATIN ALAN sınıfının 2445 adedini doğru olarak, 468 adedini de yanlış olarak tahmin etmiştir. Toplam 7059 adet olan ALMAYAN sınıfının da 6744 adedini doğru olarak ve 315 adedini de yanlış olarak tahmin etmiştir.



Şekil 5.10 ID3 Algoritmasının Eğitim ve Test Süreci

CHAID Algoritması

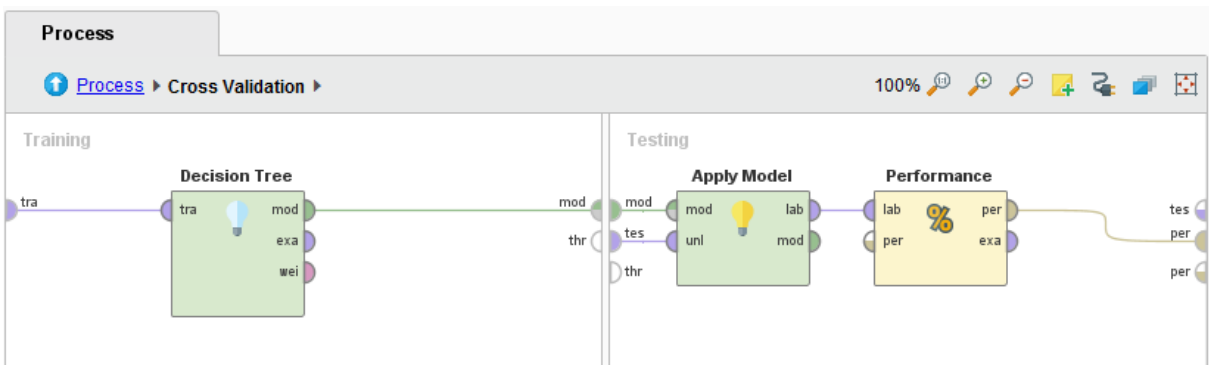
CHAID algoritmasının temel parametreleri diğer algoritmalar ile eşit olarak ayarlandıktan sonra çapraz geçerlilik için doğrusal örnekleme yapıldığında doğruluk oranı % 91,26, tabakalı örnekleme yapıldığında ise % 91,25 olmaktadır. CHAID algoritması en iyi performansı 10 katlı rastgele çapraz geçerlilikte % 91,59 olarak vermiştir. Bu durumda CHAID algoritması veri setinde bulunan toplam 2913 adet SATIN ALAN sınıfının 2407 adedini doğru olarak, 506 adedini de yanlış olarak tahmin etmiştir. Toplam 7059 adet olan ALMAYAN sınıfının da 6726 adedini doğru olarak ve 333 adedini de yanlış olarak tahmin etmiştir.



Şekil 5.11 CHAID Algoritmasının Eğitim ve Test Süreci

C4.5 Algoritması

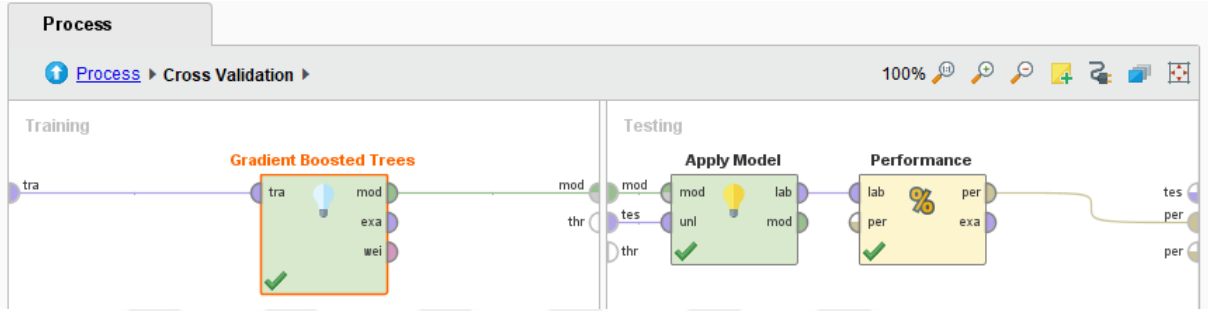
C4.5 algoritması ile budama işlemi yapılmadan çalıştırılan modelde çapraz geçerlilik için doğrusal örnekleme yapıldığında doğru tahmin oranı % 92,56 ve rastgele örnekleme yapıldığında ise % 92,63 olmaktadır. C4.5 algoritması en iyi performansı 9 katlı tabakalı çapraz geçerlilikte % 92,73 olarak vermiştir. Bu durumda C4.5 algoritması veri setinde bulunan toplam 2913 adet SATIN ALAN sınıfının 2473 adedini doğru olarak, 440 adedini de yanlış olarak tahmin etmiştir. Toplam 7059 adet olan ALMAYAN sınıfının da 6774 adedini doğru olarak ve 285 adedini de yanlış olarak tahmin etmiştir.



Şekil 5.12 C4.5 Algoritmasının Eğitim ve Test Süreci

Gradyan Arttırılmış Ağaçlar Algoritması

Gradyan Arttırılmış Ağaçlar algoritması için ortak olan temel parametreleri diğer algoritmalar ile eşit olarak ayarlanmış ve farklı olanlar da programın verdiği varsayılan değerde bırakılmıştır. Çapraz geçerlilik için doğrusal örnekleme yapıldığında doğru tahmin oranı % 92,06 ve tabakalı örnekleme yapıldığında ise % 92,07 olmaktadır. Gradyan Arttırılmış Ağaçlar algoritması en iyi performansı 10 katlı rastgele çapraz geçerlilikte % 92,11 olarak vermiştir.

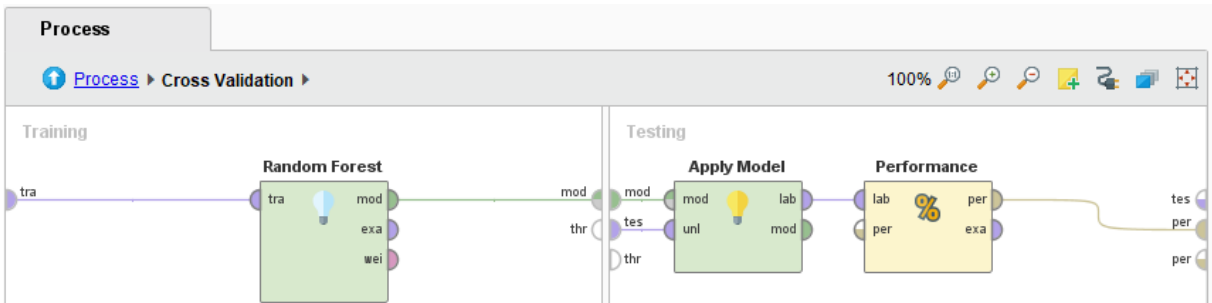


Şekil 5.13 Gradyan Arttırılmış Ağaçlar Algoritmasının Eğitim ve Test Süreci

Gradyan Arttırılmış Ağaçlar algoritması veri setinde bulunan toplam 2913 adet SATIN ALAN sınıfının 2468 adedini doğru olarak, 445 adedini de yanlış olarak tahmin etmiştir. Toplam 7059 adet olan ALMAYAN sınıfının da 6713 adedini doğru olarak ve 346 adedini de yanlış olarak tahmin etmiştir.

Rastgele Orman Algoritması

Rastgele Orman algoritması için ortak olan temel parametreleri diğer algoritmalar ile eşit olarak ayarlanmış ve farklı olanlar da programın verdiği varsayılan değerde bırakılmıştır. Budama işlemi yapılmamıştır. Çapraz geçerlilik için doğrusal örnekleme yapıldığında doğru tahmin oranı % 92,40 ve rastgele örnekleme yapıldığında ise % 92,41 olmaktadır. Rastgele Orman algoritması en iyi performansı 10 katlı tabakalı çapraz geçerlilikte % 92,67 olarak vermiştir.

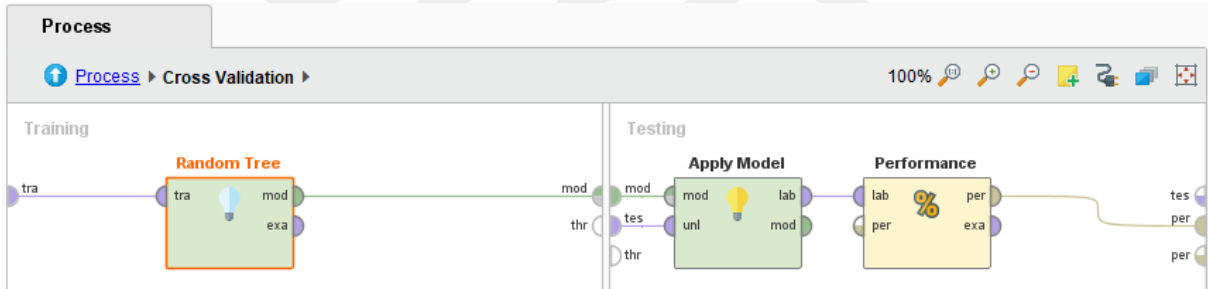


Şekil 5.14 Rastgele Orman Algoritmasının Eğitim ve Test Süreci

Rastgele Orman algoritması veri setinde bulunan toplam 2913 adet SATIN ALAN sınıfının 2444 adedini doğru olarak, 469 adedini de yanlış olarak tahmin etmiştir. Toplam 7059 adet olan ALMAYAN sınıfının da 6797 adedini doğru olarak ve 262 adedini de yanlış olarak tahmin etmiştir.

Rastgele Ağaç Algoritması

Rastgele Ağaç algoritması için ortak olan temel parametreleri diğer algoritmalar ile eşit olarak ayarlanmış ve farklı olanlar da progamın verdiği varsayılan değerde bırakılmıştır. Çapraz geçerlilik için doğrusal örnekleme yapıldığında doğru tahmin oranı % 75,90 ve tabakalı örnekleme yapıldığında ise % 78,75 olmaktadır. Rastgele Ağaç algoritması en iyi performansı 10 katlı rastgele çapraz geçerlilikte % 78,80 olarak vermiştir. Rastgele Ağaç algoritması veri setinde bulunan toplam 2913 adet SATIN ALAN sınıfının 1697 adedini doğru olarak, 1216 adedini de yanlış olarak tahmin etmiştir. Toplam 7059 adet olan ALMAYAN sınıfının da 6642 adedini doğru olarak ve 417 adedini de yanlış olarak tahmin etmiştir.

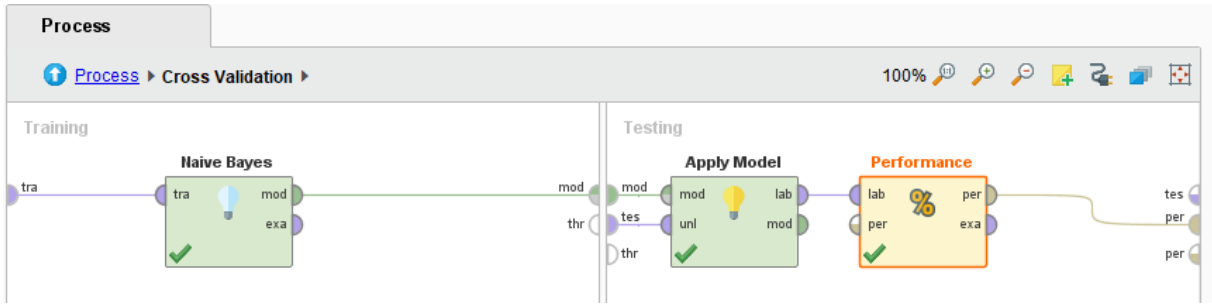


Şekil 5.15 Rastgele Ağaç Algoritmasının Eğitim ve Test Süreci

5.7.1.3. Bayes Sınıflandırma

Hesaplama açısından oldukça zor bir yöntem olsa da, veri kümesini bir kere eğittikten sonra oldukça hızlı çalışan, bir durumun olma ihtimalinin en yüksek olma koşuluna göre hareket eden bir tür sınıflandırma algoritmasıdır. Büyük veri kümeleri için oldukça kullanışlıdır. Naive Bayes dinamik sistemler için, yani verinin sürekli olarak değiştiği durumlar için uygun değildir. Çünkü yeni veri geldikçe eğitim sürecini ve test sürecini tekrarlamak gerekmektedir. Modelde Naive Bayes algoritması kullanılmıştır.

Diğer algoritmalarda olduğu gibi Naive Bayes algoritmasında da 10 katlı çapraz geçerlilik kullanılmıştır. Naive Bayes algoritması tabakalı örnekleme oluşturulduğunda % 90,30; doğrusal örnekleme ile % 90,12 oranında doğruluk gösterirken rastgele örnekleme yapıldığında ise % 90,18 doğruluk oranına ulaşmaktadır. Modelin çalışması 1 saniyeden daha az sürmüştür.



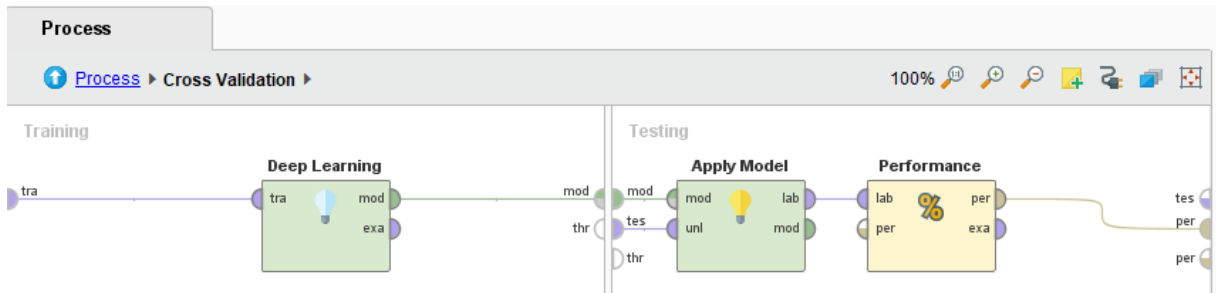
Şekil 5.16 Naive Bayes Algoritması İçin Eğitim ve Test Verilerinin Düzenlenmesi

Naive Bayes algoritması veri setinde bulunan toplam 2913 adet SATIN ALAN sınıfının 2434 adedini doğru olarak, 479 adedini de yanlış olarak tahmin etmiştir. Toplam 7059 adet olan ALMAYAN sınıfının da 6559 adedini doğru olarak ve 500 adedini de yanlış olarak tahmin etmiştir.

5.7.1.4. Yapay Sinir Ağları

Çalışmamızdaki veri seti içerdiği çoklu nominal değişkenler nedeniyle, RapidMiner yazılımı bünyesinde bulunan yapay sinir ağları algoritmalarından sadece Deep Learning tarafından desteklenmiştir. Dolayısıyla bu modelde Deep Learning algoritması kullanılmıştır.

Deep Learning modelinde gizli katmanlardaki nöronlar tarafından kullanılacak olan harekete geçirme fonksiyonu olarak Üstel Doğrultucu Doğrusal Birim işlevi ve veri setinin 20 kez yinelenmesi seçilmiştir.

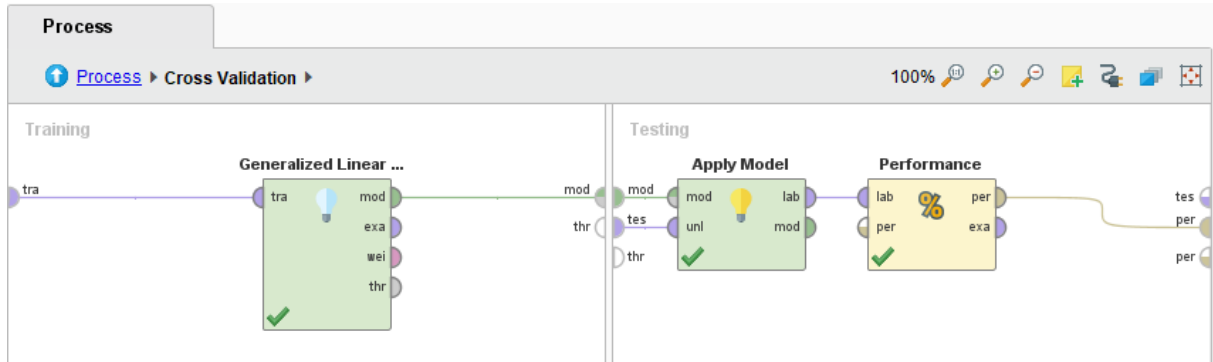


Şekil 5.17 Deep Learning Algoritması İçin Eğitim ve Test Verilerinin Düzenlenmesi

Modelde çapraz geçerlilik için doğrusal örnekleme yapıldığında doğru tahmin oranı % 92, 52 ve rastgele örnekleme yapıldığında ise % 92,55 olmaktadır. Deep Learning algoritması en iyi performansı 10 katlı tabakalı çapraz geçerlilikte % 92,57 olarak vermiştir. Deep Learning algoritması veri setinde bulunan toplam 2913 adet SATIN ALAN sınıfının 2501 adedini doğru olarak, 412 adedini de yanlış olarak tahmin etmiştir. Toplam 7059 adet olan ALMAYAN sınıfının da 6730 adedini doğru olarak ve 329 adedini de yanlış olarak tahmin etmiştir.

5.7.1.5. Regresyon

Rapidminer yazılımında 8 farklı model desteklenmektedir. Bu modellerden Genelleştirilmiş Doğrusal Model kullanılmıştır. Bu model veri setindeki çoklu nominal değişkenler üzerinde analiz yapmayı desteklemektedir. Bağımlı değişken normal dağılımlı değilse; örneğin Bernoulli, Poisson, Gamma veya başka bir dağılım olabilir; sonuç çıkarım Genelleştirilmiş Doğrusal Model (Generalized Linear Model, GLM) teorisi altında yapılmaktadır.

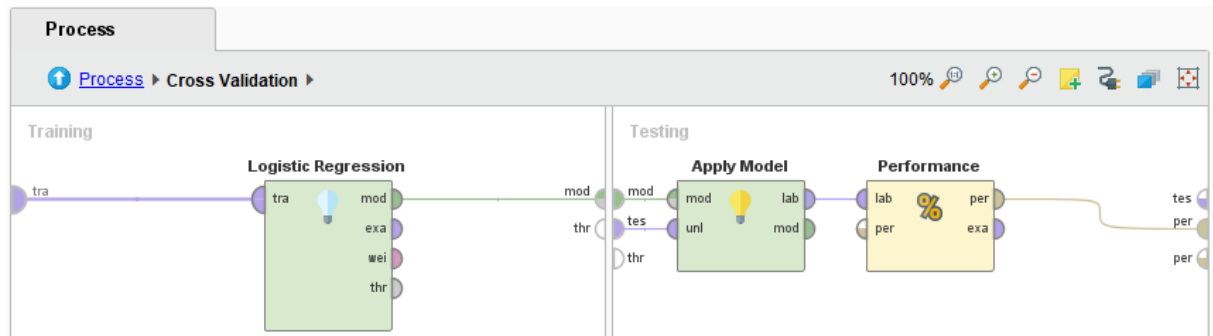


Şekil 5.18 Genelleştirilmiş Doğrusal Model İçin Eğitim ve Test Verilerinin Düzenlenmesi

Çapraz geçerlilik için doğrusal örnekleme yapıldığında doğruluk oranı % 92,29 ve rastgele örnekleme yapıldığında ise % 92,37 olmaktadır. Genelleştirilmiş Doğrusal Model algoritması en iyi performansı 10 katlı tabakalı çapraz geçerlilikte % 92,40 olarak vermiştir. Genelleştirilmiş Doğrusal Model veri setinde bulunan toplam 2913 adet SATIN ALAN sınıfının 2428 adedini doğru olarak, 485 adedini de yanlış olarak tahmin etmiştir. Toplam 7059 adet olan ALMAYAN sınıfının da 6786 adedini doğru olarak ve 273 adedini de yanlış olarak tahmin etmiştir.

5.7.1.6. Lojistik Regresyon

Rapidminer yazılımında 3 farklı lojistik regresyon modeli bulunmaktadır. Bunlardan lojistik regresyon kullanılmıştır.



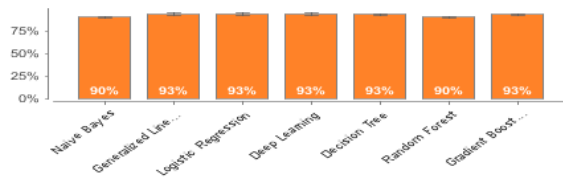
Şekil 5.19 Lojistik Regresyon İçin Eğitim ve Test Verilerinin Düzenlenmesi

Çapraz geçerlilik için doğrusal örnekleme yapıldığında doğruluk oranı % 92,27 ve rastgele örnekleme yapıldığında ise % 92,29 olmaktadır. Lojistik regresyon algoritması en iyi performansı 10 katlı tabakalı çapraz geçerlilikte % 92,29 olarak vermiştir. Lojistik regresyon veri setinde bulunan toplam 2913 adet SATIN ALAN sınıfının 2425 adedini doğru olarak, 488 adedini de yanlış olarak tahmin etmiştir. Toplam 7059 adet olan ALMAYAN sınıfının da 6778 adedini doğru olarak ve 281 adedini de yanlış olarak tahmin etmiştir.

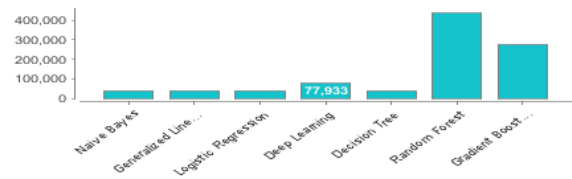
5.7.1.7. RapidMiner Otomatik Model

RapidminerStudio 9.1 analistin müdahalesinin olmadığı ve her koşulu aynı olan tek bir model aracılığıyla tahmin modelleri oluşturabilmektedir. Program Auto Model adını verdiği model ile veri hazırlama ve model optimizasyonu dâhil, veriler için mükemmel modeli oluşturmaya yardımcı olmaktadır. Bu model incelenip değiştirilecek, hatta optimize edebilecek ya da uygulanabilecek bir RapidMiner sürecidir. Süreç beş adımdan oluşmaktadır. İlk önce ilgili veri programa yüklenir. Sonra hangi veri madenciliği görevinin kullanılacağı seçilir. Tahmin modelleri seçildiğinde hangi sınıfın tahminleri yapılacağı modellere bildirilir. Dördüncü adımda hangi değişkenlerin analize alınacağı seçilir. Model kendince uygun gördüğü değişkenleri otomatik olarak seçmektedir. Burada farklı seçimler yapılabilir. Girişe alınacak değişkenler belirlendikten sonraki süreçte program veriye uygun modelleri otomatik olarak listelemektedir. Çalışma için gerekli görülmeyen model analizden çıkartılabilir. Model çalıştırdıktan sonra örneği Şekil 5.20’de görülen bir rapor sunmaktadır. Analist bu rapora göre hangi modeli kullanacağına karar vermektedir.

Accuracy



Runtime (ms)



Model	Accuracy	Standard Deviation	Runtime
Naive Bayes	89.9%	± 0.8%	38 s
Generalized Linear Model	92.9%	± 1.2%	39 s
Logistic Regression	93.1%	± 1.3%	38 s
Deep Learning	93.2%	± 0.9%	1 min 17 s
Decision Tree	93.3%	± 0.7%	38 s
Random Forest	90.0%	± 0.6%	7 min 17 s
Gradient Boosted Trees	92.8%	± 0.7%	4 min 36 s

Şekil 5.20 Rapidminer Programının Auto Model Sonucu

Çalışmada da hangi modelin kullanılacağına karar vermek için ikinci bir alternatif olarak RapidMiner 9.1'in Auto Model seçeneği değerlendirilmiştir. Bu model ile ilk önce değişken seçimi işlemi tekrarlanmıştır. Modele veri içindeki değişkenler eklenerek ve çıkartılarak denemeler defalarca yapılmıştır. Denemelerin çoğunda modeller açısından benzer sonuçlar çıkmış, bazı değişkenler dâhil edildiğinde tahmin oranları yüksek, bazı değişkenler dâhil edildiğinde de düşük çıkmıştır. Tablo 5.11'da verilmeyen ve sınıfların tahmin edilmesinde çok fazla etkisi olmayan değişkenler analiz listesinden çıkartılmıştır. Buradan elde edilen değişkenler ile manuel olarak tek tek kurulan modellerden elde edilen değişkenler karşılaştırılarak bu aşamada aynı zamanda son modelde kullanılacak değişkenler de tespit edilmiştir. Auto model veriye uygun 7 modeli otomatik olarak seçmiştir. En son elde edilen sonuçlara ait ekran görüntüsü Şekil 5.20'da verilmiştir.

5.7.2. Tur Bazlı Analizler

Çalışmanın alt problemi veri setinde bulunan altı farklı günlük tur için satın alma durumlarının tespit edilmesidir. Çünkü günlük turlar düzenlenişi, etkinlikleri ve katılımcıları açısından birbirinden farklıdır. Tur özelliğine göre tura katılan turist özellikleri de değişebilmektedir. Bunun için rezervasyon kaydı bulunan turistlerin veri seti içerisindeki farklı günlük turları satın alma durumlarının tahmin edilmesi için karşılaştırma sonucunda en iyi performansı veren C4.5 algoritması kullanılmıştır.

Bu model kurulmadan önce sınıflandırma değişkeni (bağımlı değişken) günlük turlara göre yeniden düzenlenmiş ve değişken ayrıklaştırılmıştır. Ana problem için oluşturulan modellerde turist en az bir adet tur satın alması durumunda "SATIN ALAN" olarak sınıflandırılmıştır. Bu modelde rezervasyon kaydı girilen turistin belirli bir günlük turu satın alma durumunun tahmin edileceği için sınıflandırma değişkeni o günlük turu satın alma ya da almama durumuna göre yeniden düzenlenmiştir. Veri setinde kayıtlı 6 farklı günlük tur (Hamam, Alışveriş, Gezi, Tekne, Rafting ve Eğlence) için 6 farklı sınıflandırma değişkeni oluşturulmuştur. Müşteri ilgili turu satın almışsa değişken içeriği "SATIN ALAN", satın almamış ise "ALMADI" olarak düzenlenmiştir. Alt problem için kullanılan bağımlı ve bağımsız değişkenler aşağıda verilmiştir.

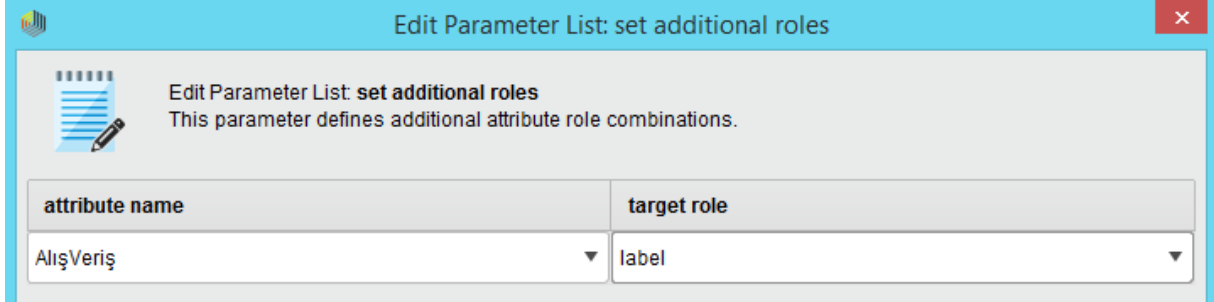
Bağımsız Değişkenler

Müşteri Tipi
Yetişkin Yaşı
Çocuk Bilgisi
Konaklama Süresi
Otel Ücreti
Bölge

Bağımlı Değişken

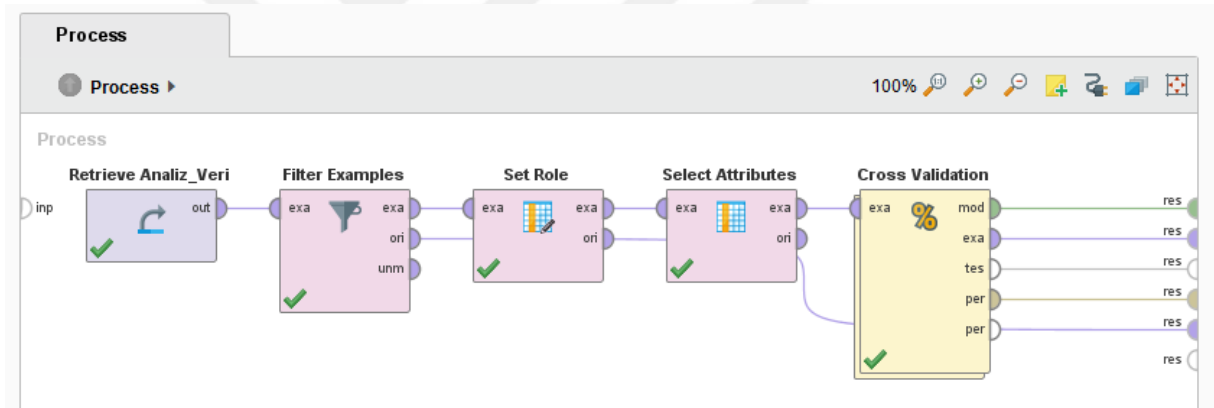
Alışveriş
Hamam
Rafting
Gezi
Tekne
Eğlence

Şekil 5.22’de görülen modele Set Role operatörü ile Şekil 5.21’de görüldüğü gibi sınıflandırma değişkeni her günlük tur için değiştirilerek modelin performansı incelenmiş ve elde edilen karar ağacı kurallarına bakılmıştır.



Şekil 5.21 Günlük Turların Sınıf Değişkeni Olarak Belirlenmesi

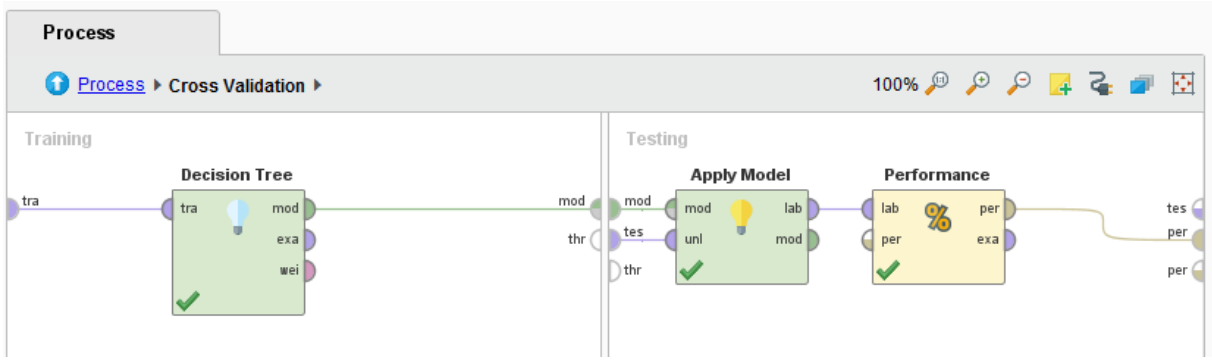
Select Attributes operatörü ile analize dâhil edilecek bağımlı ve bağımsız değişkenler belirlenmektedir. Yazılım burada belirlenen değişkenleri analize almaktadır.



Şekil 5.22 Günlük Turlara Göre Profil Belirleme Modelinin Süreci

Çapraz geçerlilik operatörü içinde öğrenme bölümünde karar ağacı operatörü bulunmaktadır (Şekil 5.22). Bir karar ağacı, bir dizi basit karar kuralını uygulayarak, geniş bir kayıt koleksiyonunu birbirini izleyen küçük kayıt kümelerine bölmek için kullanılabilir bir yapıdır. Her ardışık bölünme ile, sonuç kümelerinin üyeleri birbirine daha fazla benzer hale gelmektedir (Berry ve Linoff, 2004:166). Bölme için hangi değişkenin seçileceği kararını vermek için Bilgi kazanımı (Information gain) kullanılmıştır. Bilgi kazanımı (Information gain) için tüm değişkenlerin entropileri hesaplanır ve bölünme için en az entropiye sahip olan seçilir.

Bir ağacın derinliği, veri setinin boyutuna ve özelliklerine bağlı olarak değişmektedir. Modelde karar ağacının derinliği 10 düğüm ile sınırlandırılmıştır. Bu default değerdir. İstenirse Bu değer '-1' olarak ayarlanarak ağaç diğer durma kriterleri yerine getirilinceye kadar yapılabilir. Karar ağacı modeli oluşturulduktan sonra budanabilir. Modelde budanmadan önce fazla ağaç düğümü oluştuğu için %50 hata oranında kötümser budama yöntemi ile budanmıştır



Şekil 5.23 Günlük Turlara Göre Profil Belirleme Modelinin Eğitim ve Test Bölümü

Bir düğümün kazancı bölünmeden önce hesaplanır. Eğer kazancı minimum kazancından büyükse, düğüm bölünür. Modelde minimum kazanç %8 seçilmiştir. Daha yüksek bir minimum kazanç değeri, daha az bölünme ve dolayısıyla daha küçük bir ağaçla sonuçlanır. Çok yüksek bir değer, bölünmeyi tamamen önler ve tek düğümlü bir ağaç oluşturulur. Modelde %20'den sonra tek bir düğüm oluşmaktadır. Bir yaprağın boyutu, alt kümesindeki örneklerin sayısıdır. Ağaç, her yaprağın en azından minimum yaprak büyüklüğü sayısına sahip olmasını sağlayacak şekilde üretilir. Modelde minimum yaprak büyüklüğü 3 olarak seçilmiştir. Aynı şekilde bir düğümün boyutu da alt kümesindeki örneklerin sayısıdır. Yalnızca, bölünmüş parametre için minimum boyuttan büyük (Modelde 4'tür) veya eşit büyüklükteki düğümler bölünür. Belli bir düğümde ön çalışma yapılarak bölünme önlendiğinde, bölme için test edilen alternatif düğüm sayısına ihtiyaç olabilir, modelde 2 olarak ayarlanmıştır.

Modelin çalıştırılması sonucunda elde edilen karar ağacı kuralları Araştırmanın Bulguları kısmında verilmiştir.

5.8. Araştırmanın Bulguları

Çalışmanın başlıca iki farklı bulgusu bulunmaktadır. İlk bulgu kullanılan 11 adet veri madenciliği sınıflandırma algoritmasının performans sonuçlarına yönelik bulgudur. Elde edilen bulgular algoritmaların seyahat acentası veri tabanından elde edilmiş bir veri setindeki tahmin performansları konusunda fikirler vermektedir. İkinci bulgu ise en iyi performansı gösteren C4.5 algoritmasının sunduğu karar ağaçlarıdır. Karar ağaçlarından elde edilen bulgular ise seyahat acentasının günlük tur satın alan müşterilerini tanımasında ve onlara yönelik kampanya ve fiyat düzenlemesinde karar vericilere yol gösterecektir.

Araştırma bulguları ana ve alt problemler için ayrı ayrı kurulan modeller için ayrı ayrı verilmiştir.

5.4.1. Algoritmaların Performansları

Çalışmanın ana problemi, seyahat acentasında otel rezervasyon kaydı bulunan turistlerin günlük tur satın alma durumlarını tahmin eden sınıflandırma yöntemlerinin performanslarını karşılaştırmaktır. Yöntemleri karşılaştırmak için iç ve dış geçerlilik ölçeklerine bakılabilmektedir. Dış geçerlilik, önceki veri etiketine, yani önceden sınıflandırılmış etiketlere dayanmaktadır. İç geçerlilik verinin dahili karakterine dayanır ve önceden bir etiket gerektirmez (Puri vd., 2017: 241). Önceden belirlenen sınıflara göre gerçekliğin test edilebilmesi için veri setinin bir kısmı öğrenme bir kısmı da modelin geçerliğinin test edilmesi için ayrılmaktadır. Verilerin eğitim ve test setlerine bölünmesinden kaynaklanan yanlılığı en aza indirmek ve böylece farklı tahmin yöntemlerin tarafsız bir karşılaştırmasını sağlamak için tabakalı 10-kat çapraz geçerlilik testi kullanılmıştır. Veri seti 10 ayrık alt gruba bölünmüştür ve bu 10 alt kümeden biri, 10 deneyin her biri için test kısmı olarak kullanılmıştır. Bu değerlendirme prosedüründe sınıflandırıcı eğitimi ve sınıflandırıcı testi ayrı gözlemler üzerinde yapıldığından fazla zaman harcatmadığı için avantajlıdır. Ayrıca, önceki araştırmalar, bu çalışmada olduğu gibi işlemsel veri setleri ile eğitilmiş modelleri değerlendirmek için 10-katlı tabakalı çapraz geçerlilik testinin en iyi performansı gösterdiğini ortaya koymuştur (Siering vd., 2018: 58).

Çapraz geçerlilik testinden sonra farklı performans ölçüleri de hesaplanabilmektedir (Siering vd., 2018: 58). Doğruluk, Kesinlik, Hassasiyet, F1 ölçümü, gibi harici küme geçerliliği ölçümleri kümelenmenin ne kadar iyi olduğunu değerlendirmek için bilinen sınıflarla üretilen sınıfları karşılaştırarak kullanılmaktadır (Puri vd., 2017: 241).

Doğruluk, doğru sınıflandırılmış toplam gözlem sayısı (TP + TN) toplam gözlem sayısına bölünerek (TP + FP + TN + FN) hesaplanmaktadır (Siering vd., 2018: 58). (TP: Doğru pozitif; TN: Doğru negatif; FP: Yanlış pozitif; FN: Yanlış negatif).

$$\text{Doğruluk} = \frac{TP+TN}{TP+FP+TN+FN}$$

Kesinlik, bir sınıflandırıcının ne kadar kesin olduğunu ölçmektedir, yani belirli bir sınıfa ait olarak sınıflandırılan bir örneğin gerçekten o sınıfa ait olup olmadığıdır (Siering vd., 2018: 58). Kesinlik, genelde p harfi ile gösterilir ve doğru pozitif tahminlerin sayısı ile pozitif tahminlerin sayısı arasındaki orandır (Puri vd., 2017: 241).

$$\text{Kesinlik} = \frac{TP}{TP+FP}$$

Hassasiyet doğru pozitif tahminlerin sayısı ile pozitif örneklerin sayısı arasındaki oran olarak tanımlanmaktadır (Puri vd., 2017: 241). Hassasiyet genelde r harfi ile gösterilir ve getirilen doğru sonuçların, getirilmesi gereken doğru sonuçlara oranı ile hesaplanmaktadır.

$$\text{Hassasiyet} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

F1 ölçümü, hassasiyet ve kesinliğin harmonik ortalaması olarak tanımlanmaktadır (Puri vd., 2017). F1 ölçümü harmonik ortalamalarını kullanarak kesinlik ve hassasiyeti bir araya getirmektedir. Bu, sıklıkla, kesinlik ve hassasiyetin birbiriyle ilişkili olduğunu ifade etmekte ve yüksek bir kesinlik düşük bir hassasiyete ve bunun tersine eşlik etmektedir (Siering vd., 2018).

$$F_1 = 2 \cdot \frac{\text{Kesinlik} \cdot \text{Hassasiyet}}{\text{Kesinlik} + \text{Hassasiyet}}$$

Yukarıdaki değerlendirme ölçümleri çerçevesinde sınıflandırma algoritmaları için tek tek model oluşturularak modellerin performansları değerlendirilmiştir. Yukarıdaki değerlendirme ölçümleri çerçevesinde sınıflandırma algoritmaları için tek tek model oluşturularak modellerin performansları değerlendirilmiştir. Modeller oluşturulurken bir modelde en iyi tahmin sonucunu almak için parametreler ile oynanmış ve en iyi sonucu verdiği anda model performansı Tablo 5.12'de verilen tabloya işlenmiştir. Doğruluk ölçütü baz alındığında Desicion Tree (C4.5) doğru tahmin ve hız konusunda diğer algoritmalarından daha iyi sonucu vermiştir. Algoritmaların çalıştırılması esnasında sadece Rapidminer programı çalıştırılmış, bilgisayarın internet bağlantısı kesilerek, anti virüs devredışı bırakılmıştır. Bu koşullarda algoritmanın sorunu çözme hızına bakılmıştır. Hız ne kadar algoritma seçimi için belirleyici olmasa da her algoritmanın aynı koşullarda yarışmasına özen gösterilmiştir.

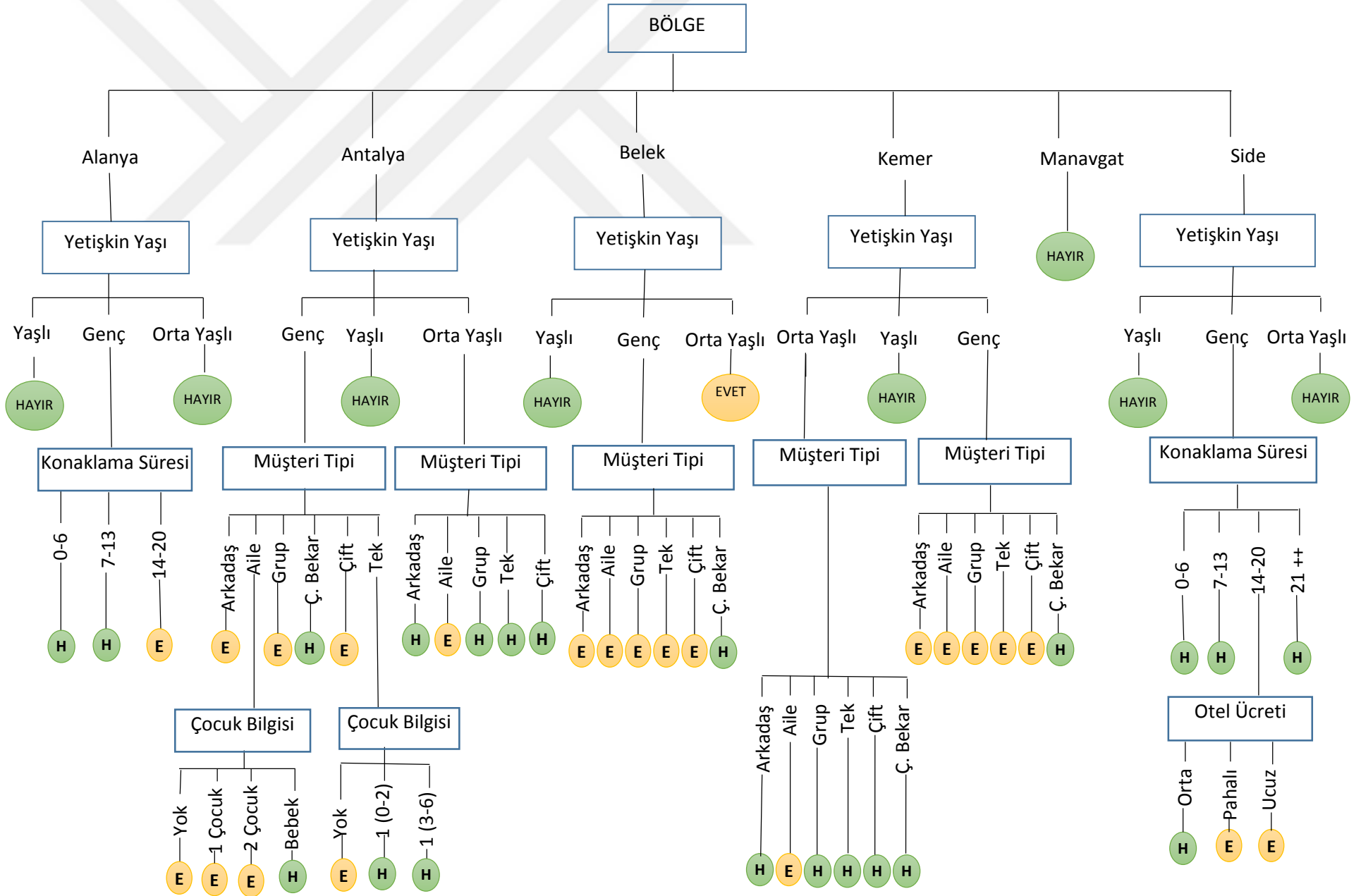
Tablo 5.12 Sınıflandırma Algoritmaları Karşılaştırması

Algoritma	Doğruluk	Sınıflandırma Hatası (%)	İşlem Süresi	Sınıf Kesinliği (%)	
				SATIN ALAN	ALMAYAN
K-NN Clasification	92,14	7,86	6 saniye	88.81	93.40
Desicion Tree (C4.5)	92,73	7,27	1 saniye	89.67	93.90
ID3	92,15	7,85	1 saniye	88.59	93.51
CHAID	91,59	8,41	23 saniye	87.85	93.00
RandomForest	92,67	7,33	8 saniye	90.32	93.55
GradientBoostedTrees	92,07	7,93	45 saniye	87.70	93,78
RandomTree	78,80	21,20	1 saniye	74.46	79.65
NaiveBayes	90,18	9,82	1 saniye	82.96	93.19
Deep Learning	92,72	7,28	42 saniye	89.26	94.06
GeneralizedLinear	92,29	7,71	10 saniye	89.59	93.30
Logistic regresyon	92,29	7,71	2 saniye	89.62	93.28

Auto Model en iyi tahmin yapan modeli tespit etmektedir. Modeller tek tek çalıştırıldıktan sonra programın Auto Model seçeneği değerlendirilmiş ve çıkan sonuçlar modelleri tek tek çalıştırarak elde edilen sonuçlara benzerlik göstermiştir. Hem modellerin tek tek çalıştırılması sonucu elde edilen sonuçlar hem de Auto Model sonuçları C4.5 algoritmasının en iyi performansı verdiğini göstermektedir. Ayrıca algoritma seçiminde çözülecek problem ve çalışma amacı da önemli rol oynamaktadır. Çalışmanın amaçlarında birisi de daha önceki müşterilere bakarak işletmenin müşteri profilinin çıkartılması ve teşvik verilecek müşterilerin tespit edilmesi ve satış esnasında uygun günlük turların önerilmesidir. Deep learning bazı model girişlerinde en yüksek tahmin oranını yakalamış olsa da (en fazla % 4) amaca daha uygun sonuçlar vereceği için müşteri profillerinin belirlenmesinde C4.5 algoritmasının sonuçları kullanılmıştır.

5.4.2. Müşteri Profilleri

Performansları değerlendirilen 11 adet algoritmadan C4.5 algoritmasının çalıştırılması sonucu elde edilen karar ağacı (kötümser budama yöntemi ile budandıktan sonra) Şekil 5.24'de verilmiştir. Karar ağacı, veri madenciliğinde kullanılan en popüler ve güçlü tekniklerden biridir (Berry ve Linoff, 2004). Anlaşılması ve yorumlanması kolaydır. Kullanılan ağaç yapılar görselleştirilebilir. Analizlerde kullanılan program ağaç yapısını metin halinde de sunmaktadır. Ağaç yapının metin hali budama işlemi gerçekleştirilmeden Ek-1 Karar Ağacı Kuralları kısmında verilmiştir. Burada çok fazla ve bazen sadece tek bir gözleme göre karar verilmesine neden olabilecek karar kuralları olduğu için %50 oranından kötümser budama gerçekleştirilmiştir. Budamadan sonra oluşan karar ağacı metin kuralları Ek-2 Budanmış Karar Ağacı Kuralları kısmında verilmiştir. Tur satın alan müşterilere yönelik profil belirlemede de budanmış karar ağacı sonuçları kullanılmış ve ilgili karar ağacı Şekil 5.24'de verilmiştir.



Şekil 5.24 Karar Ağacı

Çıkan kurallar bir seyahat acentasının günlük tur satın alan müşteri profilinin konaklama bölgesine göre değiştiğini, müşteri yaşının da etkili bir belirleyici olduğunu göstermektedir. Antalya bölgesinde yetişkinlerin yanında bulunan çocuk, Kemer ve Belek bölgesinde tur satın almada karar vermede etkili bir rol oynamamaktadır. Tablo 5.13’de acentanın tur satın alan müşterilerine ait profil bilgileri verilmektedir.

Tablo 5.13 Tur Satın Alan Müşteri Grupları

Bölge	Tur satın alan müşteri grupları
ALANYA	<ul style="list-style-type: none"> ▪ En az 14 gün konaklayan genç yetişkinler
ANTALYA	<ul style="list-style-type: none"> ▪ Pahalı otellerde konaklayan genç arkadaş grupları ▪ Yanında bebek olmayan genç aileler ▪ Gençlerden oluşan gruplar ▪ Yanında çocuk olmayan tek konaklayan genç misafirler ▪ Genç çiftler ▪ Yanında 1 veya 2 çocuk bulunan orta yaşlı aileler
BELEK	<ul style="list-style-type: none"> ▪ Genç arkadaş grupları ▪ Genç aileler ▪ Gençlerden oluşan gruplar ▪ Tek konaklayan genç misafirler ▪ Genç çiftler
KEMER	<ul style="list-style-type: none"> ▪ Genç arkadaş grupları ▪ Genç aileler ▪ Gençlerden oluşan gruplar ▪ Tek konaklayan genç misafirler ▪ Genç çiftler
SİDE	<ul style="list-style-type: none"> ▪ Pahalı otellerde en az 14 gün konaklayan Gençler

Seyahat acentasının gereksiz promosyon yapmasını önlemek için tur alanların önemli olduğu kadar tur satın almayan müşteri profilinin de belirlenmesi gerekmektedir. Tablo 5.14’de tur satın almayan müşteri profilleri verilmiştir. Orta yaşlı, yaşlı yetişkinlerin ve yanlarında çocuk ile gelen bekarların genel olarak tur satın almadığı tespit edilmiştir.

Tablo 5.14 Tur Satın Almayan Müşteri Grupları

Bölge	Tur satın almayan müşteri grupları
ALANYA	<ul style="list-style-type: none"> ▪ Orta yaşlı misafirler ▪ Yaşlı misafirler ▪ 14 günden daha az konaklayan genç misafirler
ANTALYA	<ul style="list-style-type: none"> ▪ Yanında 1 adet çocuk olan tek konaklayan genç misafirler ▪ Bebekleri ile gelen genç aileler ▪ Yanında çocuk bulunan bekâr/dul genç misafirler ▪ Aileler dışındaki orta yaşlı misafirler ▪ Yaşlı misafirler
BELEK	<ul style="list-style-type: none"> ▪ Yanında çocuk bulunan bekâr/dul genç misafirler ▪ Yaşlı misafirler
KEMER	<ul style="list-style-type: none"> ▪ Yanında çocuk bulunan bekâr/dul genç misafirler ▪ Aileler dışındaki orta yaşlı misafirler ▪ Orta yaşlı çiftler ▪ Yaşlı misafirler
SİDE	<ul style="list-style-type: none"> ▪ Pahalı ve ucuz bir otelde 14-20 gün konaklayanlar dışındaki genç misafirler ▪ Orta yaşlı misafirler ▪ Yaşlı misafirler

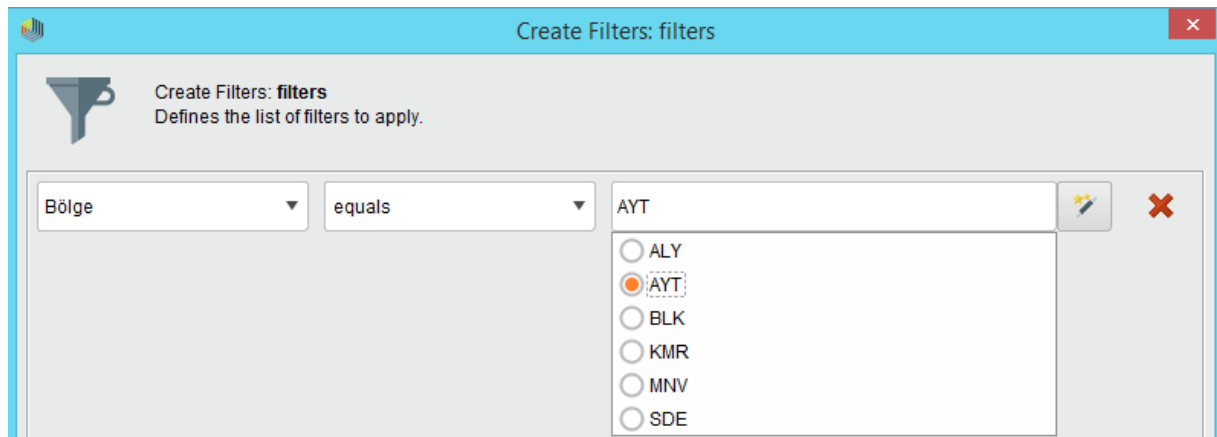
Hazırlanan bağımsız değişkenlerden birisi de Pansiyon değişkeniydi. Pansiyon değişkeni tur satın alma ya da almama ile hiç ilgili olmadığı için sonradan analizden çıkartılmıştır. Veri içinde yapılan frekans analizinde aslında her şey dâhil rezervasyon alanlarla yarım pansiyon ya da tam pansiyon rezervasyon yaptıranların tur alımları arasında bir farklılık gözlemlenmemiştir. “Her şey dâhil sistemden dolayı tur satışı yapamıyoruz” ifadesini yalancılayan bir bulgu mevcut değil ama tur satışı yapılamamasının başka nedenlerinin olduğunu düşünmek gerektiğinin kanıtıdır.

5.4.3. Bölgelere Göre Müşteri Profilleri

Modele veri setinin tamamının katıldığında elde edilen karar ağaçlarında ilk düğüm bölge olarak karşımıza çıkmıştır. Daha sonra aynı modele sadece ilgili bölgede konaklayanların dâhil edilmesiyle bölge bölge tur satın alma kuralları ortaya çıkarılmıştır.

Günlük turlar bölgenin özelliğine ve olanaklarına göre düzenlenmektedir. Turistler de destinasyon seçimi yaparken o destinasyonda görmek istediği yerler, katılmak istediği etkinlikler vb. dikkate almaktadır. Günlük tur biletleri de otelde acenta temsilcisi tarafından bu turistlere satılmaktadır. Çalışmanın bu aşamasında her bölgede günlük tur satın alan müşteri profillerinin belirlenmesi, sonucunda da o bölgede tur bileti satın temsilcilerin info kokteyl esnasında potansiyel müşterilerini daha iyi tanıması amaçlanmıştır.

Tur satın alan müşterilerin profillerinin belirlendiği ana modele Filter Examples operatörü eklenerek her bölgenin tur satın alma kurallarına bakılmıştır.



Şekil 5.25 Bölgelere Göre Veri Filtreleme

Ek-1’de verilen karar ağacı kuralları incelendiğinde Alanya bölgesinde konaklayan genç turistlerden 14 – 20 gün arası konaklayan çiftlerin günlük tur satın turist olarak sınıflandırıldığı görülmektedir. Alanya bölgesi süzülerek yapılan bir analizde ise 14 – 20 gün arası konaklayan genç turistler tur satın alan sınıfında yer almaktadır.

Genel müşteri profilleri incelendiğinde Antalya bölgesinde genç misafirlerin tur satın aldığı görülmektedir. Sadece Antalya’da konaklayanlar için model çalıştırıldığında yine aynı sonuçlar biraz daha özet olarak karşımıza çıkmaktadır. Kurallara göre Antalya’da konaklama satın alan genç misafirlerden ailer, çiftler, gruplar günlük tur satın alırken çocuğu ile beraber gelen tek yetişkinler (dul) ve bebeği olan aileler tur satın almamaktadır. Orta yaşlı misafirlerden aileler dışında günlük tur satın alımı yapılmamaktadır. Antalya bölgesinde yanında bebeği olan genç aileler, çocukla tatile gelen tek yetişkinler, orta yaşlı misafirler (aileler hariç) ve yaşlı misafirler tur satın almayan turist grubunda sınıflandırılmaktadır.

Belek bölgesinde konaklayan genç misafir gruplarından aileler, arkadaş grupları ve çiftler tur satın alırken Antalya bölgesinde olduğu gibi çocuğu ile beraber gelen tek yetişkinler (dul) tur satın almamaktadır. Orta yaşlı ailelerin Belek bölgesinde günlük tur satın almada yoğun ilgilerinin olduğu görülmektedir. Bunun bölgede düzenlenen turlarla özellikle golf turu ile ilgisinin olabileceği varsayılabilir. Belek bazlı yapılan süzme sonrası elde edilen karar ağacında ise Belek’te konaklayan orta yaşlı misafirler tur satın alan sınıfına yerleştirilmiş görülmektedir. Yaşlı misafirler diğer bölgelerde olduğu gibi Belek’te de tur satın almayan turist grubunda sınıflandırılmaktadır.

Kemer bölgesinde konaklayan genç misafir gruplarından Antalya ve Belek bölgesinde konaklayanlara ek olarak tek tatile gelen turistlerin de günlük tur satın aldıkları görülmektedir. Belek bölgesinde olduğu orta yaşlı ailelerin Kemer’de de günlük tur satın almada yoğun ilgilerinin olduğu görülmektedir. Çocukları ile tatile gelen genç bekârlar, aileler dışındaki orta yaşlı ve yaşlı misafirler günlük tur satın almamaktadır.

Genel müşteri profilleri incelendiğinde Manavgat bölgesinde konaklayan misafirler ile ilgili hiçbir karar ağacı kuralı oluşmamıştır. Ancak sadece Manavgat’ta konaklayanlar için model çalıştırıldığında genç çiftlerin tur satın aldığı görülmektedir.

Side bölgesinde üç hafta konaklayan gençlerin tur aldığı genel müşteri profilleri incelendiğinde ortaya çıkmaktadır. Sadece Side’de konaklayanlar için model çalıştırıldığında genç grupların tur satın aldığına dair kural oluşmaktadır.

5.4.4. Turlara Göre Müşteri Profilleri

Çalışmanın ikinci aşaması için hazırlanan karar ağacı modelinde tekrar tabakalı 10-katlı çapraz geçerlilik testi uygulanmıştır. Hazırlanan model, seyahat acentası müşterilerini günlük tur satın alanlar (SATIN ALAN) ya da almayanlar olarak sınıflandırarak, belirlenen amaçlara uygun kurallar çıkarmıştır. İşletmenin eski tur kayıtlarından yararlanarak, müşteri hakkında yeterli detay veri olmamasına rağmen (meslek, seyahat nedeni vb.) günlük tur satın alan

müşterilerinin profillerine ulaşılmıştır. Müşteri profillerinin müşterilerin konakladığı bölgelere göre değiştiği saptanmıştır. Müşterinin yaşı, müşteri tipi, yanlarında bulunan çocuk bilgisi, profiller için temel belirleyiciler olurken, konaklanan oteldeki kişi başı konaklama ücreti ve konaklama süresi de tur satın almayı etkilemektedir.

Seyahat acentalarında günlük tur bileti satışlarını otellerde acenta temsilcileri satmaktadır. Onların hangi tür günlük turu hangi müşteri grubuna satması durumunda daha çok satış yapabilmesi için tur gruplarına özel müşteri profillerinin çıkartılması için model güncellenmiştir. Burada profillere bakarken tur alanlar arasından ilgili turu alan müşterilerin özelliklerine bakılmıştır. Turlar için model tarafından verilen karar ağaçları eklerde verilmiştir. Tablo 5.15’de turlara göre müşteri profilleri verilmiştir.

Tablo 5.15 Turlara Göre Müşteri Profilleri

Bölge	Tur satın alan müşteri grupları
Alış Veriş Turu	<ul style="list-style-type: none"> ▪ Antalya’da 14-20 gün konaklayan yaşlı arkadaş grupları ▪ Antalya’da 7-13 gün konaklayan yaşlı gruplar ▪ Antalya’da 7-13 gün ucuz bir otelde konaklayan tek misafirler ▪ Antalya’da 7-13 gün konaklayan orta yaşlı ve yaşlı misafirler ▪ Belek’de 7-13 gün ucuz bir otelde konaklayan orta yaşlı misafirler ▪ Kemer’de 7 günden az konaklayan genç arkadaş grupları ▪ Kemer’de 7-13 gün ucuz bir otelde konaklayan orta yaşlı tek misafirler
Gece (Eğlence) Turları	<ul style="list-style-type: none"> ▪ Antalya’da 7 günden daha az orta ücretli bir otelde konaklayanlar ▪ Antalya’da pahalı otellerde konaklayan 7 yaşından büyük çocuğu olan genç misafirler. ▪ Antalya’da 7-13 gün orta ücretli bir otelde konaklayan arkadaş grupları ▪ Belek’de pahalı bir otelde 7 günden az konaklayanlar ▪ Kemer’de yanında 12-18 yaş grubundan çocuk bulunan tek misafirler ▪ Kemer’de yanında 7 yaşından büyük çocuk bulunan, pahalı otellerde konaklayan aileler ▪ Belek’de pahalı bir otelde konaklayan yaşlı aileler ▪ Belek’de pahalı bir otelde 7 günden az konaklayan çiftler
Gezi –kültür Turları	<ul style="list-style-type: none"> ▪ Alanya’da 14-20 gün konaklayanlar ▪ Alanya’da 7-13 gün ucuz ve orta fiyatlı otellerde konaklayanlar ▪ Antalya’da orta ücretli otellerde konaklayan gruplar ▪ Antalya’da orta ücretli otellerde tek konaklayan genç misafirler ▪ Belek’de yanında 7-12 yaş grubundan bir çocuk olan Bekar/dul misafirler ▪ Kemer’de 7-13 gün arası yanında 12-18 yaş arası bir çocuk bulunan orta yaşlı tek (bay ya da bayan) misafirler ▪ Manavgat’ta konaklayanlar ▪ Side’de en az 14 gün konaklayanlar ▪ Side’de 7-13 gün orta ve pahalı bir otelde konaklayan genç misafirler ▪ Side’de 7-13 gün orta ve ucuz bir otelde konaklayan genç çiftler ▪ Side’de 7-13 gün konaklayan orta yaşlı misafirler
Hamam Turu	<ul style="list-style-type: none"> ▪ Belek’te orta ücretli otellerde konaklayan genç arkadaş grupları ▪ Belek’te pahalı otellerde yanında 7 yaşından büyük çocuk bulunanlar
Tekne – Yat Turları	<ul style="list-style-type: none"> ▪ Antalya’da orta ücretli otellerde yanında çocuk olmayan tek misafirler ▪ Antalya’da pahalı otellerde konaklayan yanında 7 yaşından büyük çocuk bulunan orta yaşlı misafirler ▪ Antalya’da pahalı otellerde konaklayan aileler ve gruplar ▪ Belek’te orta ücretli otelde konaklayan yanında çocuk bulunan gençler ▪ Kemer’de orta ücretli otelde konaklayan yanında çocuk bulunan misafirler ▪ Kemer’de yanında çocuk bulunmayan orta yaşlı misafirler
Rafting	<ul style="list-style-type: none"> ▪ Turu satın alan müşteri profiline ulaşamadı

Alış veriş turları seyahat acentalarının en az masrafla düzenlediği turlardandır. Elde edilen sonuçlara göre alış veriş turlarına orta yaşlı ve yaşlı turistler ilgi göstermektedir. Genel olarak müşteri profillerine bakıldığında yaşlı, hatta orta yaşlı misafirlerin hiçbir şekilde tur almadıkları söylenebilir, ancak sadece alış veriş turuna göre sınıflandırılma yapıldığı genel müşteri profilinden çok dah farklı bir müşteri grubu karşımıza çıkmaktadır. Yaşlı grupların alış veriş turlarına olan katılımları onların tanıdıklarına hediyelik eşya götürme isteklerinin sonucu olabilir. Bu da alış veriş turlarına bilet satan oteldeki acenta temsilcisinin yaşlı gruplara biraz daha ilgi göstermesi ve onları alış verişe teşvik edecek kampanyalar düzenlemesi ile seyahat acentası için çok iyi bir gelir kaynağına dönüştürülebilir.

Eğlence turları seyahat acentasının kayıtlarında gece vb. ifadeler ile kaydedilmiştir. Diğer turlara katılanların genel özelliklerinden birisi de uzun süreli konaklama yapıyor olmalarıydı ama eğlence turuna genelde kısa süreli konaklama yapanların katıldığı tespit edilmiştir. Kemer ve kısmen Belek bölgesinde yetişkinlerin yanında çocukların da özellikle çocukları ile gelen tek misafirlerin (dul) eğlence turlarına çocuklarıyla birlikte katıldıkları görülmektedir. Antalya bölgesinde ise ucuz otellerde, Belek bölgesinde pahalı otellerde konaklayan yaşlı misafirlerden de eğlence turlarına katılım olmaktadır. Gösteri turlarına ise Antalya, Belek ve Kemer'de genelde pahalı otellerde konaklayan genç arkadaş gruplarının ilgi gösterdiği tespit edilmiştir. Kemer bölgesinde Antalya ve Belekten farklı olarak iki haftadan fazla konaklayanların ve iki haftadan az konaklayan yaşlı misafirlerin de gösteri turlarına katıldıkları görülmüştür.

Manavgat ve Side bölgesinde konaklayan turistler gezi turlarına katılan turist grupları arasında karşımıza çıkmaktadır. Diğer modellerde hiç karar kurallarında tur satın alan sınıfta geçmeyen bu iki bölgede konaklayanlar gezi turlarının dâhil edildiği modelde tur satın alan sınıfta karşımıza çıkmıştır. Aynı şekilde Alanya bölgesinde konaklayanlar da diğer iki bölge gibi gezi turlarına ilgi göstermektedir.

Tekne turlarına genelde çocuklarıyla birlikte gelen misafirlerin katıldığı izlenmektedir. Diğer turlarda ve genel müşteri profillerinde hep tur almayan sınıfta çıkan yanında çocuğu olan tek misafirlerin de Kemer ve Belek bölgelerinde tekne turlarında genelde tur alan sınıfta oldukları ortaya çıkmıştır. Tekne turlarına katılımda genç, orta yaşlı ve yaşlı misafirlerin aralarında belirgin bir fark gözlemlenmemiştir.

Kısaca özetlenirse, orta yaşlı ve yaşlı misafirler için alış veriş ve gösteri turları, yanında çocuğu bulunan misafirler için tekne ve gösteri turları, konaklama süresi az olan genç misafirler için eğlence turları, gezilecek bölgelere yakın otellerde konaklayanlar için gezi turları, günlük tur bileti satışı esnasında önerilecek turlar arasındadır.

Tablo 5.16 Tur Satışlarını Arttırmak İçin Yapılabilecekler

Bölge	Tur Satışlarını Arttırmak İçin Yapılabilecekler
ALANYA	<ul style="list-style-type: none"> ▪ Alanya bölgesine yapılan rezervasyon istekleri diğer bölgelere yönlendirilebilir. ▪ 14 günden fazla konaklayan genç yetişkinlere teşvikler sunulabilir. ▪ Bölge konaklayanlar gezi ve kültür turlarına yönlendirilebilir.
ANTALYA	<ul style="list-style-type: none"> ▪ Grup rezervasyonu yaptıranlara teşvik verilebilir. Müşteri adedi arttıkça tur satın alma ihtimali de artmaktadır. ▪ Kısa süreli gelen turistler için eğlence ve gece turları önerilebilir. ▪ Yanında çocuğu bulunan aileler ve genç misafirlere teşvik verilebilir. ▪ Alışveriş turu bileti özellikle orta yaşlı ve yaşlı turistlere satılabilir. ▪ Alışveriş turları yerel ürünler ile daha da zenginleştirilebilir.
BELEK	<ul style="list-style-type: none"> ▪ Zengin misafirlere Golf turu gibi turlara yönlendirilebilir. ▪ Orta yaşlı misafirlere pahalı otelleri tercih etmeleri durumunda Belek bölgesine yönlendirilebilir. ▪ Kısa süreli gelen turistler için eğlence ve gece turları önerilebilir. ▪ Yanında çocuğu olan zengin misafirlere gösteri turları tanıtımı yapılabilir.
KEMER	<ul style="list-style-type: none"> ▪ Çocukla gelen tek misafirlere çocuklarını emanet edebilecekleri ortamların oluşturulması konusunda yerel yetkililer uyarılabilir. ▪ Yanında 12 yaşından büyük çocuğu olanlara eğlence turlarının tanıtımı yapılabilir.

5.9. Tartışma ve Sonuç

Veri madenciliği, bilgisayarlarda depolanan büyük miktardaki verilerin analizini ifade etmektedir (Olson ve Delen, 2008: 3) ve büyük veri tabanlarından geçerli, önceden bilinmeyen, anlaşılabilir ve eyleme geçirilebilir bilgilerin çıkarılması ve kritik iş kararları verilebilmesi için kullanılmasıdır (Ha ve Park, 1998: 1). Birçok sektör, e-ticaret ve perakende satışta müşteri tercihlerini ölçmek, sağlık hizmetlerinde tıbbi geçmişi belirlemek, sigortacılıkta risk faktörlerini değerlendirmek ve bankacılıkta finansal verileri toplamak gibi çeşitli amaçlarla veri madenciliği teknolojisini benimseyerek rekabet güçlerini arttırmaktadır (Shapoval vd., 2018 : 311). Turistlere ilişkin bilgiler artan bir hızla toplanmakta ve destinasyonların rekabet edebilmeleri ve pazar paylarını arttırması giderek daha zor hale gelmektedir (Pyo vd., 2002). Destinasyon yönetim örgütleri, Seyahat acentaları ve oteller rekabetçi kalmak için veri madenciliği tekniklerine her geçen gün daha çok ihtiyaç duyacaklardır.

Veri madenciliği projelerinin başarılı bir şekilde uygulanması, çeşitli algoritmaların dikkatli bir şekilde değerlendirilmesini gerektirir, çünkü kısmen herhangi bir durumda diğerlerinden daha iyi performans gösteren tek bir algoritma yoktur (Delen ve Sirakaya, 2006: 329). Her tekniğin kendine özgü güç ve zayıflığı vardır, bu da onu sadece belirli problem alanlarına uygulanabilir kılmaktadır (Cho ve Leung, 2002: 117). Farklı veri kombinasyonları ve ele alınmakta olan problem, üstün tahmin ediciler olmak için farklı veri madenciliği algoritmalarını gerektirebilmektedir (Delen ve Sirakaya, 2006: 316).

Çalışmada veri madenciliği tekniklerinden sınıflandırma yöntemi kullanılmıştır. 11 adet sınıflandırma algoritması veri setinde çalıştırılmış ve doğruluk seviyesi ve amaca en uygun olan

karar ağacı algoritması ile seyahat acentasının tur satın alan / almayan misafirlerinin profillerinin keşfi yapılmıştır. Burada öncelikli olarak tercih edilen yöntem daha sonra da elde edilen sonuçlar tartışılacaktır.

Günlük tur alan müşteri profillerinin ortaya çıkartılması için kullanılacak veriye ve amaca en uygun veri madenciliği algoritmasının tespiti için sınıflandırma algoritmalarından literatürde en yaygın olarak kullanılanlardan analiz için kullanılan paket programın destek verdikleri veri setinde koşturulmuştur. Turizmde ya da diğer alanlarda üzerinde anlaşılmış ve en iyi olduğu söylenen bir algoritma mevcut değildir (Delen ve Sirakaya, 2006: 329). Algoritmaların performansı kullanılan veri setine, hatta aynı veri setinde kullanılan parametre ve özelliklere göre değişebilmektedir. Bu gözönünde bulundurularak tüm modellerde aynı değişkenler (Müşteri Tipi, Yetişkin Yaşı, Çocuk Bilgisi, Konaklama Süresi, Otel Ücreti, Bölge ve Tur) modele dâhil edilmiş ve tüm modellerde 10 katlı çapraz geçerlilik testi uygulanmıştır.

Modellerin performansları karşılaştırıldığında aralarında çok önemli bir fark bulunmamaktadır. Hespinin performansı kabul edilebilir düzeydedir. Doğruluk ve hız kriterleri baz alınarak çalışmanın ikinci aşaması için karar ağaçları modeli tercih edilmiştir. Çalışmamızda en iyi performansı veren karar ağaçları Delen ve Sirakaya (2006)'nın çalışmasında karşılaştırılan üç yönteminden en az tahmin gücü olan yöntem olmuştur. Delen ve Sirakaya (2006), yapay sinir ağları, karar ağaçları ve kaba kümeler olmak üzere üç popüler veri madenciliği yöntemini test etmişler ve kaba kümeler algoritmasının üçü arasında en iyi tahmin aracı olduğunu bulmuşlardır. Xiang vd. (2018) çalışmalarında en iyi performansı sergileyen bayes sınıflandırıcı çalışmamızda 11 algoritma arasında orta düzeyde bir performans sergilemiştir. Xiang vd. (2018), metin sınıflandırması için aday algoritmalar olarak SGD Sınıflandırıcı, Rasgele Orman, Adaboost, Doğrusal SVC, Lojistik Regresyon ve Multinomial Naïve Bayes dahil olmak üzere altı klasik makine öğrenme algoritması kullanmışlar. Bu algoritmaların performansını değerlendirmek için, hassasiyet, kesinlik, doğruluk ve F-1 puanı dâhil olmak üzere bir dizi standart metrikten elde edilen sonuçlar, Multinomial Naïve Bayes'in en iyi performansı ürettiğini göstermiştir. Benzer şekilde Bjørkelund vd. (2012) duygu analizi skorlarını 5 kategoride toplamış ve kategorileri sınıflandırmak için kullanılan bayes ağları çok iyi performans göstermiştir. Siering vd. (2018) Yapay Sinir Ağı (YSA) ve Destek Vektör Makinesi (DVM) gibi karmaşık öğrenme algoritmaları ile birlikte Naïve Bayes algoritmasını da kullanmışlardır. Çalışmada Naïve Bayes sınıflandırıcılarının genel olarak iyi performans gösterdikleri kanıtlanmıştır (Siering vd., 2018: 57).

Algoritmaların performansını değerlendirmek için, hassasiyet, kesinlik, doğruluk ve F-1 puanı dahil olmak üzere bir dizi standart metrik kullanılabilir (Xiang vd., 2018). Veri

madenciliği tekniklerinin performansının değerlendirilmesinde, kestirim doğruluğuna ek olarak, bazı araştırmacılar modellerin açıklayıcı niteliğinin önemini ve geçerli, yeni, kullanışlı ve en önemlisi anlaşılabilir ve açıklanabilir kalıpları ortaya çıkarmanın gerekliliğini vurgulamıştır (Delen ve Sırakaya. 2006: 328). Ancak algoritma seçiminde algoritmanın performansına bakıldığı kadar çözülmek istenen probleme uygunluğuna da bakılmalıdır. Örneğin, Naive Bayesin diğer sınıflandırma methodlarına göre daha iyi sonuç verdiği tespit edilmiştir (Xiang vd., 2018; Björkelund vd., 2012) ama Naive Bayes dinamik sistemler için, yani verinin sürekli olarak değiştiği durumlar için uygun değildir. Çünkü yeni veri geldikçe eğitim sürecini ve test sürecini tekrarlamak gerekmektedir.

Modelleme gücüne ve karmaşık tahmin-modelleme görevlerinde yaygın kullanımına rağmen, YSA'ları açıklayıcı güce sahip olmadıkları için eleştirilmiştir. Başka bir deyişle, algoritma içinde yürütülen veri dönüşümünün karmaşıklığı ve doğrusal olmayan doğası nedeniyle, mantığın giriş değişken-lerinden nasıl çıkarıldığını izah etmek ve açıklamak zordur. Sonuç olarak, tahmin alanında çalışan araştırmacılar tarafından YSA'nın tahmin modellemesi kara kutu yaklaşımı olarak adlandırılmaktadır. Bu eleştiriye yanıt olarak, araştırma topluluğu, YSA'larından açık bilgi elde etme yollarını geliştirmiştir. Her ne kadar bu sezgisel yöntemlerin geliştirilmesinde bir ölçüde başarılı olsa da, teorik bir temele sahip olmayan YSA'ların temel eleştirisi hala devam etmektedir. YSA'ların aksine, karar ağaçları ve kaba kümeler, modellerini sezgisel olarak açıklanabilen iş kuralları biçiminde sunmaktadır. Giriş değişkenlerini (faktörleri) IF <koşul (lar)> THEN <eylem (ler)> yapısını kullanarak çıktı değişkenlerine (sonuçlara) bağlamaktadırlar (Delen ve Sırakaya. 2006: 328).

Çalışmada müşteri profillerinin çıkartılması için karar ağaçları seçilmiştir. Karar ağaçları modelinin verdiği karar ağacı kuralları Ek-1'de verilmiştir. Çıkan sonuçlara göre; seyahat acentasının müşteri profilleri bölge bölge değişmektedir. Bunda da her bölgede satılan günlük turların bölgenin özelliğine göre düzenlenmesi ve her bölgeyi tercih edenlerin de bölge özelliğine göre tercih etmesinin etkisi olduğu düşünülenebilir. Antalya bölgesinde konaklayan gençlerin günlük tur satın aldıkları görülmektedir. Genç çiftler, genç arkadaş grupları, yanında çocuk olmayan genç tek başında konaklayan gençler turlara ilgi göstermektedir. Diğer bölgelerden farklı olarak yanında çocuk bulunan genç ailelerinde tur satın aldığı görülmektedir. Bu durumda çocuklu ailelere rezervasyon yaptırırken Antalya bölgesindeki oteller tavsiye edilebilir. Bununla birlikte orta yaşlı çiftler, yaşlı misafirler ve yanında çocuk ile gelen tek genç ve orta yaşlı misafirler ve tek başına konaklayan orta yaşlı misafirler de turlardan uzak durmaktadırlar. Belek bölgesinin turist profili Antalya bölgesi ile yakınlık göstermektedir ancak, Belek bölgesinde çocuksuz genç aileler turlara ilgi göstermektedir. Tüm yaşlı misafirler

ve yanında çocuk bulunan bekâr/dul genç misafirler turlardan uzak durmaktadırlar. Kemer bölgesinde diğer bölgelerde olduğu gibi gençler turlara ilgi göstermektedir. Yanında çocuk bulunan bekâr/dul genç misafirler ile orta yaşlı ve yaşlı misafirler turlardan uzak durmaktadırlar. Alanya bölgesinde en az 14 gün konaklayan genç misafirler ve Side bölgesinden pahalı otellerde en az 14 gün konaklayan genç misafirler turlara ilgi göstermektedir. Her iki bölgede de orta yaşlı ve yaşlı misafirler turlardan uzak durmaktadırlar.

Genel olarak tur satın alan misafirler için müşteri profilleri araştırmasından sonra tek tek günlük turlar için müşteri profili araştırması yapılmıştır. Benzer özellikteki turlar gruplandırılmıştır. Örneğin alış veriş turları tek isim altında toplanmış, farklı bölgelerdeki hamam turları hamam turu olarak tek tur halinde düzenlenmiştir. Her bir tur grubu için yeniden model oluşturulmuştur. Bu modellerde elde edilen sonuçlar genel tur satın alan müşteri profillerinden çok daha farklı çıkmıştır.

Genel olarak hiçbir bölgede tur almadığı düşünülen orta yaşlı ve yaşlı misafirlerin alış veriş turlarına ilgi gösterdikleri ortaya çıkmıştır. Özellikle Antalya bölgesinde 7 günden daha fazla konaklayan yaşlı grupların ve 7 -13 gün konaklayan orta yaşlı ve yaşlı misafirler ile ucuz bir otelde konaklayan tek misafirlerin alış veriş turlarına ilgi gösterdikleri ortaya çıkmıştır. Belek ve Kemer’de 7-13 gün ucuz bir otelde konaklayan orta yaşlı misafirler ile Kemerde 7 günden az konaklayan genç arkadaş gruplarının alışveriş turuna katıldıkları tespit edilmiştir.

Eğlence turlarına ise 7 günden daha az konaklayan misafirler ilgi göstermektedir. Antalya’da 7 günden daha az orta ücretli bir otelde konaklayanlar, Belek’de pahalı bir otelde 7 günden az konaklayanlar ve Antalya’da 7-13 gün orta ücretli bir otelde konaklayan arkadaş grupları eğlence turuna ilgi göstermektedir. Belek bölgesinde ise genellikle pahalı otellerde konaklayanlar eğlence turunu tercih etmektedirler. Kemer’de farklı olarak yanında 12-18 yaş grubundan çocuk bulunan tek misafirler de eğlence turu satın almışlardır. Kemer bölgesinde ayrıca gece turlarına bilet alan yetişkinlerin yanında biletsiz olarak küçük çocukların bulunduğu veri setinden izlenmiştir. Bunun nedeninin yetişkinlerin küçük çocuklarını bırakacak bir yer ya da kişi bulamamalarından kaynaklandığı düşünülmektedir.

Tur satın alan müşteri profilleri incelemesinde belirgin turist profili çıkmayan Alanya, Manavgat ve Side bölgesi Gezi – Kültür turları için yapılan araştırmada bu iki bölgenin gezi ve kültür turlarını tercih ettiğini ortaya çıkarmıştır. Alanya’da 7 günden daha fazla, Antalya’da orta ücretli otellerde konaklayanlar, Manavgat’ta konaklayanlar, Side’de konaklayanlar Gezi-kültür turlarına ilgi duymaktadırlar. Yanında 7 yaşından büyük çocuğu olan misafirlerin de gezi-kültür turlarını tercih ettikleri görülmektedir.

Bu çalışmanın amacı, veri madenciliği yöntem ve teknikleri ile seyahat sektörü özellikle seyahat acentası veritabanlarında bulunan işlemsel verileri kullanarak yönetim karar desteği sağlayabilecek bilgilerin çıkartılması süreçlerini gerçekleştirmek, bu süreçlerde geçilen aşamaları ortaya koymak ve kullanılan yöntem ve tekniklerin etkinliğini araştırmaktır. Ayrıca, uygulamada elde edilen sonuçlarla tekniklere ilişkin farklılıkları tartışmak ve hangi tekniğin hangi koşullarda uygulanmasının daha uygun olacağına yönelik önerilerde bulunmaktır. Bu amacı gerçekleştirebilmek için veri seti seyahat acantasından alınmıştır.

Veri setinde ilk önce farklı tablolarda bulunan ve çalışma amacına hizmet eden veriler tek bir tabloda birleştirilmiştir. Birleştirilen tabloda 24 farklı değişken oluşturulmuştur. Daha iyi bir model oluşturmak ve daha anlaşılabilir sonuçlar elde etmek için bu değişkenlerde azaltma işlemi gerçekleştirilmiştir. Bu işlemin sonunda sınıf değişkenini belirlemede en fazla katkısı olan 6 değişken kullanılarak 11 farklı sınıflandırma modeli oluşturulmuştur.

Modeller performans değerlendirme metriklerinden doğruluk, kesinlik, hassasiyet ve hız ölçütleri ve çalışma amacı göz önünde bulundurularak değerlendirilmiştir. Bu değerlendirme sonunda çalışmanın ikinci aşamasına karar ağaçları ile devam edilmesine karar verilmiştir. Bu aşamada ilk önce genel olarak günlük tur satın alan müşteri profilleri keşfedilmiş, daha sonra hangi turları hangi müşteri gruplarının satın aldığına bakılmıştır.

Çalışmada elde edilen veri seti turizm sektöründe faal durumdaki bir operasyonel sistemin yedeklerinden elde edilmiştir. Günümüzde milyonlarca insanın turist olarak seyahat ettiği ve bu insanlara ait verilerin işletmelerin operasyonel sistemlerinde kaydedildiği düşünülürse, karşımıza gerçekten muazzam bir veri kaynağı çıkmaktadır. Üstelik bu veri kaynağı birleri tarafından manipüle edilebileceği, olmayan bir işlemin olmuş gibi gösterilebileceği, sırf bir yer edinebilmek adına gerekli gereksiz paylaşımların yapılabildiği sosyal medya gibi kontrolsüz değildir. Bu veri kaynaklarındaki tüm veriler gerçek turist hareketlerini, satın alma davranışlarını içermektedir. Kısacası operasyonel sistemlerde işletmeler için kıymetli bilgiler ortaya çıkarılmayı beklemektedir.

Ne kadar zor olsa da, işletmelerin günlük operasyonlar sırasında oluşturdukları ve topladıkları veri yığınlarından yararlanılmalı ve işlenmiş verileri yararlı enformasyon ve bilgiye dönüştürülmelidir. Yine de, bilginin değerini ayırt etme zorluğu, birçok işletmenin elindeki veri zenginliğinden tam olarak yararlanılmasını engellemektedir (Ha ve Park, 1998: 1). Turizm yöneticileri ve araştırmacılarının bu konudaki zorlukları aşmalarına veri madenciliği teknikleri çözüm getirmektedir. Veri madenciliğinin tanımlayıcı ve tahmin edici modelleri veri tabanlarında ve diğer ortamlarda bulunan veriler içindeki yararlı bilgiyi keşfetme konusunda iyi performans gösterdiklerini kanıtlamışlardır.

Tüm bunlara rağmen alandaki çalışmalar incelendiğinde operasyonel verilerden müşteri bilgisi çıkartan çalışma sayısı yok denecek kadar azdır. Bu açıdan bu çalışma aslında var olan ama araştırmacılar tarafından ihmal edilmiş olan bu veri kaynağına dikkat çekmesi açısından önemlidir. Daha önceki çalışmalarda veri tabanlarından elde edilen bilgilerin işletme için hayati olduğunu savunan birçok çalışma da (örneğin: Cho ve Leung, 2002) veri tabanlarından elde edilmiş veri üzerinde veri analizi yapmamıştır.

Operasyonel sistemlerden veri elde etmenin bir başka sorunu da seyahat acentalarının operasyonlarını yaptıkları müşterileri hakkında demografik verileri sistemlerine kaydetmiyor olmalarıdır. Örneğin J. Wong (2006) Tayvan'daki seyahat acenteleri genellikle bankacılık endüstrisi kadar müşteri verisi kaydetmediğinden benzer bir çalışma için verileri seyahat acentalarından almak yerine anket ile toplamayı tercih etmiştir. Bu tez çalışmasında bu sorun göze alınarak veriler seyahat acentasının veri tabanından alınmıştır. Dolayısıyla, çalışmadaki değişkenler alanda daha önce yapılan çalışmalarda kullanılan değişkenlere göre seçilememiş, seyahat acentasının veri tabanında başka bir deyişle, kullanılan programda girilmesi istenen kayıtlardan elde edilen değişkenler kullanılmak zorunda kalmıştır. Daha önceki müşteri profili tanımlayan çalışmalarda sınıf değişkenini belirgin bir şekilde etkileyen bilgilerin karşılayıcı acenta programlarında sisteme kaydedilmemektedir. Örneğin turizmin farklı alanlarında daha önce yapılan çalışmalarda önem teşkil eden medeni hal (Emel vd., 2007; Liao, 2010 - evli çocuklu olması), yetişkinler için yaş (Chiang, 2017); Emel vd., 2007 - genç tek bayan), çocuklar (Tchetchik vd., 2009 - küçük çocukların olmadığı gruplar), meslek (Min vd., 2002; Wong vd., 2006), gelir (Emel vd., 2007, Chiang, 2017; Wong vd., 2006), eğitim (Wong vd., 2006; Liao, 2010 - bayan öğrenci; Emel vd., 2007 - genç ve eğitimli grup), yaşadığı bölge (Liao vd., 2010; J. Wong vd., 2006) ve seyahat amacı (Min vd., 2002) gibi bilgilerin hiç birisinin Türkiye'de kullanılan karşılayıcı acenta programlarında sisteme girilmesi zorunlu değildir (Akgün ve Çizel, 2017). Bu da literatüre uygun bir şekilde profil çıkarma işlemini engellemiştir. Bu soruna rağmen yapılan çalışmada acentanın tur satın alan müşterilerini tanıyabileceği ve promosyon verebileceği müşteri gruplarının tespiti mümkün olmuştur.

Bu çalışma, gelişmiş veri madenciliği yöntemlerinin, bir seyahat acentası veri tabanında kayıtlı olan müşteri verilerinden nispeten yüksek bir doğruluk derecesiyle tur satın alan müşteri profillerini tahmin etmek için modeller geliştirmek için başarılı bir şekilde kullanılabilceğini göstermiştir. Bununla birlikte, kestirimci modelleme bağlamında, veri madenciliği modellerinin uygulanabilirliğini ve tahmin edilebilir doğruluğunu sınırlandıran birkaç konunun farkında olmalıdır Delen ve Sirakaya, 2006: 328): verilerin niteliği (zenginlik, doğruluk, eksiksizlik ve temsili de dâhil olmak üzere); veri madenciliği metotları (farklı türlerin ve veri

kombinasyonlarının ele alınabilmesi için kapasiteleri ve sınırlamaları ile bunların bağımsız ve bağımlı değişkenler arasındaki doğrusal olmayan ilişkileri yakalayabilme yetenekleri dâhil) ve uygulama alanı (Doğru öngörü modelleri geliştirmek için gerekli olan anlaşılabilirlik ve ilgili faktörlerin mevcudiyeti dâhil olmak üzere). Modeller, yalnızca onları oluşturmak için kullanılan veriler kadar iyi tahmin yapabilir. Tahmine dayalı doğruluğun temel belirleyicilerinden biri, bir çalışma için kullanılan verinin miktarı ve kalitesidir. Bu çalışmada oluşturulan tüm modeller, eldeki değişkenlere dayanan kabul edilebilir bir tahmin doğruluğu seviyesi sağlamış olsa da, daha uygun değişkenler dâhil edilerek performans seviyeleri geliştirilebilir. Doğruluk seviyelerini arttırmak için mümkün olduğunda modele daha fazla (ve ilgili) değişken eklenmelidir. Bunu da seyahat acentalarına yazılım yapan firma ve seyahat acentası yöneticilerinin planlaması gerekmektedir. Turizm ürünlerinin satışında henüz müşteri satın almadan önce genellikle müşteriye sorulan, “meslek, bütçe, nereyi ister” vb. soruların cevapları programlara kaydedilememektedir. Bu da seyahat işletmelerinin müşterilerini daha yakından tanımasını ve müşteriye uygun ürünler ile karşılarını çıkmalarını engellemektedir.

Hazırlanan model işletmenin pazarlama biriminin kullandığı yazılımına entegre edilebilir, böylece rezervasyon esnasında ve tur bileti satışı esnasında buradan elde edilen profillere göre günlük tur önerileri yapılabilir. Hatta bazı günlük turlarda ya da bazı bölgelerde bazı günlük turlar için indirimler yapılabilir ya da tur paketleri oluşturulabilir. Tur satın alma ihtimali olan ya da belli bir tur grubunun potansiyel müşterisi olan müşteri grupları için reklam ve promosyon kampanyaları yürütülebilir.

Çalışmanın amaçları arasında elde edilen modelin verilerin aldığı sisteme geri bildirimde bulunması olmadığı için modelin verilerin alındığı sisteme entegrasyonu yapılmamıştır. Çalışma kapsamında model, genel olarak günlük tur satın alan müşteri alt gruplarının tespiti ve aynı zamanda farklı günlük turları satın alan müşteri alt gruplarının tespiti için kullanılmıştır. Aynı model ile rezervasyon yaptıran bir müşterinin günlük tur satın alıp almama ihtimali de tahmin edilebilmektedir. Örneğin rezervasyon yaptıran yaşlı bir çiftin günlük tur satın almayacağı, ama genç bir çiftin günlük tur satın alacağı öngörülmektedir.

Veri madenciliği, müşterilerin yaşam döngüsü boyunca müşterilerin ihtiyaçlarını belirleyerek ve tahmin ederek şirketlerin kişisel ve karlı müşteri ilişkileri kurmasına yardımcı olmaktadır. Bundan dolayı veri madenciliğinin işletmeler tarafından kullanma şekillerinden birisi de kârlılığı artırmak için işletme için en karlı müşteri alt gruplarını belirlemedir (Olson ve Delen, 2008:4). Danubianu ve Hapenciuc (2008: 3263)’e göre veri madenciliği, firmaların hangi müşterilerin belirli ürünleri ve hizmetleri satın alma olasılığının yüksek olduğunu anlamalarına yardımcı olabilmektedir. Bu da işletmelerin daha yüksek yanıt oranları ve daha iyi

yatırım getirileri için hedeflenmiş pazarlama programları geliřtirmelerini saęlamaktadır. Ayrıca veri madencilięi, veritabanı pazarlamacılarının, müřterilerinin hedeflenen ihtiyaçlarına, isteklerine ve tutumlarına daha yakın kampanyalar geliřtirmesine yardımcı olmaktadır. Gereкли bilgiler bir veritabanında bulunduęu takdirde, veri madencilięi mevcut iř sorunları ile ilgili kalıpları tanımlamak için geniř bir müřteri aktivitesini modelleyebilmektedir. Örneęin, veri madencilięi “Hangi müřterilerin belirli bir hizmeti alma olasılıęı en yüksek?” gibi soruları yanıtlayabilmektedir. Bu türden soruların yanıtlanması, müřteri kazanımı ve kampanya yanıt oranlarını; bu da sonuçta satışları ve yatırım getirilerini artırmaktadır.



KAYNAKÇA

- Akçetin, E. ve Çelik, U. (2014). İstenmeyen Elektronik Posta (Spam) Tespitinde Karar Ağacı Algoritmalarının Performans Kıyaslaması. *Journal of Internet Applications & Management / İnternet Uygulamaları ve Yönetimi Dergisi*, 5(2):43-56.
- Akgün, A., ve Çizel, B. (2017). "Günlük Tur Programları Oluşturmada Veri Madenciliği: A Grubu Seyahat Acentası Örneği". *Turar Turizm & Araştırma Dergisi*, 6(1): 73-87.
- Akpınar, H. (2014). *DATA Veri Madenciliği Veri Analizi*. Papatya Yayıncılık, İstanbul.
- Alexander, K. A., Ramotadima, M., ve Sanderson, C. E. (2018). "The power of consensus: Developing a community voice in land use planning and tourism development in biodiversity hotspots". *Ecosystem Services*, 30: 350-361.
- Al Shehhi, M., ve Karathanasopoulos, A. (2018). "Forecasting Hotel Prices in Selected Middle East and North Africa Region (MENA) Cities with New Forecasting Tools". *Theoretical Economics Letters*, 8(09): 1623-1638.
- Al-Salim, B. (2007). "Mass customization of travel packages: data mining approach". *International Journal of Flexible Manufacturing Systems*, 19(4): 612-624.
- An, J., Zhao, S., Lu, X., ve Liu, N. (2018). "A two-stage multiple-factor aware method for travel product recommendation". *Multimedia Tools and Applications*, 77(21): 28991-29012.
- Arabi, M., Dehshiri, A. M., ve Shokrgozar, M. (2018). "Modeling transportation supply and demand forecasting using artificial intelligence parameters (Bayesian model)". *Journal of Applied Engineering Science*, 16(1): 43-49.
- Bach, M. P., Schatten, M., ve Marušić, Z. (2013). "Data mining applications in tourism: A keyword analysis". *Central European Conference on Information and Intelligent Systems. 18-20 Eylül 2013, Varazdin, Croatia*, 26-32.
- F. Balyemez, E. İ. Develi, D. Z. Bayazıt Şahinoğlu. "Ürün Yaşamı Sürecinde Üretimin Yeniden Yapılandırılması". *V. Ulusal Üretim Araştırmaları Sempozyumu. 25-27 Kasım 2005, İstanbul*, 83-87.
- Belcastro, L., Marozzo, F., Talia, D., ve Trunfio, P. (2018). "G-RoI: Automatic Region-of-Interest detection driven by geotagged social media data". *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 12(3): 27.
- Bell, D., ve Mgbemena, C. (2018). "Data-driven agent-based exploration of customer behavior". *Simulation*, 94(3): 195-212.
- Bendler, J., Wagner, S., Brandt, T., & Neumann, D. (2014). "Taming uncertainty in big data". *Business & Information Systems Engineering*, 6(5): 279-288.

- Berezina, K., Bilgihan, A., Cobanoglu, C., ve Okumus, F. (2016). "Understanding satisfied and dissatisfied hotel customers: text mining of online hotel reviews". *Journal of Hospitality Marketing ve Management*, 25(1): 1-24.
- Berry, M. J., ve Linoff, G. S. (2004). *Data mining techniques: for marketing, sales, and customer relationship management*. Wiley Publishing, Inc., Indianapolis, Indiana.
- Bevanda, V., Grzanic, J., ve Cervar, E. (2008). "Analysing The Users Perception Of Web Design Quality By Data Mining Tools". *Tourism and hospitality management*, 14(2): 251-262.
- Bjørkelund, E., Burnett, T. H. ve Nørvåg, K. (2012). "A study of opinion mining and visualization of hotel reviews". *Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services. 03 - 05 Aralık 2012, Bali, Endonezya*, 229-238.
- Bloom, J. Z. (2005). "Market segmentation: A neural network application". *Annals of Tourism Research*, 32(1): 93-111.
- Bogicevic, V., Yang, W., Bujisic, M. ve Bilgihan, A. (2017). "Visual data mining: analysis of airline service quality attributes". *Journal of Quality Assurance in Hospitality ve Tourism*, 18(4): 509-530.
- Bose, I. (2009). *Data mining in tourism*. Encyclopedia of Information science and technology (2. baskı., s. 936-940). Hershey, PA: IGI Global.
- Boz, M., Canbazoglu, E., Özen, Z. ve Gülseçen, S. (2018). "Otel Rezervasyon İptallerinin Makine Öğrenmesi Yöntemleri Tahmin Edilmesi". *Veri Bilimi*, 1(1): 7-14.
- Brandt, T., Bendler, J. ve Neumann, D. (2017). "Social media analytics and value creation in urban smart tourism ecosystems". *Information ve Management*, 54(6): 703-713.
- Birenboim, A., Anton-Clavé, S., Russo, A. P. ve Shoval, N. (2013). "Temporal activity patterns of theme park visitors". *Tourism Geographies*, 15(4): 601-619.
- Brida, J. G., Disegna, M. ve Osti, L. (2012). "Segmenting Visitors of Cultural Events by Motivation: A sequential Non-linear Clustering Analysis of Italian Christmas Market Visitors", *Expert Systems with Applications*, 39(13): 11349-11356.
- Brunk, J., Riehle, D. M. ve Delfmann, P. (2018). "Prediction of customer movements in large tourism industries by the means of process mining". *European Conference on Information Systems. 23 - 28 Haziran 2018, Portsmouth, İngiltere*, 1-16.
- Buhalis, D. (1998). "Strategic Use of Information Technologies in the Tourism Industry." *Tourism Management*, 19 (5): 409-21.

- Byrd, E. T., ve Gustke, L. (2007). "Using decision trees to identify tourism stakeholders: The case of two Eastern North Carolina counties". *Tourism and Hospitality Research*, 7(3-4): 176-193.
- Cai, G., Lee, K., & Lee, I. (2018). "Mining mobility patterns from geotagged photos through semantic trajectory clustering". *Cybernetics and Systems*, 49(4): 234-256.
- Camacho, L. A. G. ve Alves-Souza, S. N. (2018). "Social network data to alleviate cold-start in recommender system: A systematic review". *Information Processing ve Management*, 54(4): 529-544.
- Cankurt, S., & Subaşı, A. (2016). "Tourism demand modelling and forecasting using data mining techniques in multivariate time series: a case study in Turkey". *Turkish Journal of Electrical Engineering & Computer Sciences*, 24(5): 3388-3404.
- Cantalops, A. S. ve Salvi, F. (2014). "New consumer behavior: A review of research on eWOM and hotels". *International Journal of Hospitality Management*, 36: 41-51.
- Cantarella, G.E. ve De Luca, S., (2003). "Modeling transportation mode choice through artificial neural networks". *Fourth International Symposium on Uncertainty Modeling and Analysis*. 24 Eylül 2003, ABD, 84 - 90.
- Čech, P. ve Bureš, V. (2009). "Advanced technologies in e-tourism". *The 9th WSEAS International Conference on Applied Computer Science*, WSEAS Press. 17-19 Ekim 2009, Genova, 85-92.
- Chakrabarti, S., Cox, E., Frank, E., Güting, R. H., Han, J., Jiang, X. ve Pyle, D. (2008). *Data mining: know it all*. Morgan Kaufmann Publishers. Burlington.
- Chen, L. F. ve Tsai, C. T. (2016). "Data mining framework based on rough set theory to improve location selection decisions: A case study of a restaurant chain". *Tourism Management*, 53: 197-206.
- Chiang, W. Y. (2018). "Applying data mining for online CRM marketing strategy: An empirical case of coffee shop industry in Taiwan". *British Food Journal*, 120(3): 665-675.
- Cho, V. (2003). "A comparison of three different approaches to tourist arrival forecasting". *Tourism management*, 24(3): 323-330.
- Cho, V. ve Leung, P. (2002). "Towards using knowledge discovery techniques in database marketing for the tourism industry". *Journal of quality assurance in hospitality ve tourism*, 3(3-4): 109-131.
- Chua, A., Servillo, L., Marcheggiani, E. ve Moere, A. V. (2016). "Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy". *Tourism Management*, 57: 295-310.

- Clavería González, Ó., Monte Moreno, E. ve Torra Porras, S. (2016). "Modelling tourism demand to Spain with machine learning techniques. The impact of forecast horizon on model selection". *Revista de Economía Aplicada*, XXIV (72): 109-132.
- Çelik, U., Akçetin, E. ve Gök, M. (2017). *RapidMiner İle Veri Madenciliği*. Pusula Yayıncılık. İstanbul.
- Çizel, B. (2005). "İşletme stratejileri ile enformasyon sistem stratejilerinin uyumu (stratejik uyum) ve stratejik uyumun Antalya yöresindeki konaklama işletmelerinde araştırılması" *Doktora Tezi*. Antalya
- Çuhadar, M., Güngör İ. ve Göksu A. (2009). "Turizm Talebinin Yapay Sinir Ağları ile Tahmini ve Zaman Serisi Yöntemleri ile Karşılaştırmalı Analizi: Antalya İline Yönelik Bir Uygulama". *Süleyman Demirel Üniversitesi, İktisadi ve İdari Bilimler Fakültesi Dergisi*, 14(1): 99-114.
- Danubianu, M., Socaci, T. ve Barila, A. (2009). "Some aspects of data warehousing in tourism industry". *The USV Annals of Economics and Public Administration*, 9(1): 290-296.
- Danubianu, M. ve Hapenciuc, V. C. (2008). "Improving Customer Relationship Management In hotel industry by Data Mining techniques". *Annals of the University of Craiova, Economic Sciences Series*, 7(36): 3261-3268.
- Delen, D. ve Sirakaya, E. (2006). "Determining the efficacy of data-mining methods in predicting gaming ballot outcomes". *Journal of Hospitality ve Tourism Research*, 30 (3): 313-332.
- Deng, N. ve Li, X. R. (2018). "Feeling a destination through the "right" photos: A machine learning model for DMOs' photo selection". *Tourism Management*, 65: 267-278.
- Dietz, L. W., Herzog, D. ve Wörndl, W. (2018). Deriving tourist mobility patterns from check-in data. *Proceedings of the WSDM 2018 Workshop on Learning from User Interactions. Learn-IR'18*, 9 Şubat 2018, Los Angeles, California, USA
- Dirsehan, T. (2016). "Creating Competitive Advantage from Non Structured Data: A Text Mining Approach in Turkish Hospitality Sector". *Eurasian Business & Economics Journal*, S2: 78-87.
- Dolničar, S. (2004). "Beyond "commonsense segmentation": A systematics of segmentation approaches in tourism". *Journal of Travel Research*, 42(3): 244-250.
- Duan, W., Yu, Y., Cao, Q. ve Levy, S. (2016). "Exploring the impact of social media on hotel service performance: a sentimental analysis approach". *Cornell Hospitality Quarterly*, 57(3): 282-296.

- Dursun, A. ve Caber, M. (2016). "Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis". *Tourism management perspectives*, 18: 153-160.
- Eirinaki, M., Gao, J., Varlamis, I. ve Tserpes, K. (2018). "Recommender systems for large-scale social networks: A review of challenges and solutions". *Future Generation Computer Systems*. 18: 153–160
- Emel, G.G., Taskin, C. ve Akat, O. (2007). "Profiling a Domestic Tourism Market by Means of Association Rule Mining". *Anatolia*, 18(2): 334-342.
- Fang, G. S., Kamei, S. ve Satoshi, F. (2017). "A Japanese tourism recommender system with automatic generation of seasonal feature vectors". *Int. J. Adv. Comput. Sci. Appl*, 8(6): 347-354.
- Fayyad, U., Piatetsky-Shapiro, G. ve Smyth, P. (1996). "From data mining to knowledge discovery in databases". *AI magazine*, 17(3): 37-54.
- Feng, Y. (2017). "Research on the Construction of Characteristic Town and Lijiang Tourism Development Based on Site Selection Analysis". *Boletín Técnico*, 55(14): 17-23.
- Feşel, N. N. (2013). "Product Mix Determination under Uncertainty within a Framework Proposed for Effective Product Management", Ph. D. Dissertation, Middle East Technical University.
- Figueredo, M., Ribeiro, J., Cacho, N., Thome, A., Cacho, A., Lopes, F. ve Araujo, V. (2018, March). "From Photos to Travel Itinerary: A Tourism Recommender System for Smart Tourism Destination". *2018 IEEE Fourth International Conference on Big Data Computing Service and Applications*. 4 - 9 Nisan 2019, Kaliforniya, ABD, 85-92
- Fu, X., Budzik, J. ve Hammond, K. J. (2000). "Mining navigation history for recommendation". *Proceedings of 2000 International Conference on Intelligent User Interfaces*. 9–12 Ocak 2000, New Orleans, 106-112.
- Gal-Tzur, A., Rechavi, A., Beimel, D. ve Freund, S. (2018). "An improved methodology for extracting information required for transport-related decisions from Q&A forums: A case study of TripAdvisor". *Travel Behaviour and Society*, 10: 1-9.
- Geetha, M., Singha, P. ve Sinha, S. (2017). "Relationship between customer sentiment and online customer ratings for hotels-An empirical analysis". *Tourism Management*. 61: 43–54.
- Giudici, P. ve Figini, S. (2009). *Applied Data Mining for Business and Industry*. John Wiley & Sons, Ltd. Cornwall.

- Gong, Y., Fan, Y., Bai, B., Zhang, J., Gao, Z., ve Wu, X. (2018, March). "HTARF: A Hybrid Tourist Attraction Recommendation Framework for Trip Scheduling". *IEEE Fourth International Conference on Big Data Computing Service and Applications*. 10-13 Aralık 2018, Seattle, WA, ABD, 93-98.
- Goh, C., Law, R. ve Mok, H. M. (2008). "Analyzing and forecasting tourism demand: A rough sets approach". *Journal of Travel Research*, 46(3): 327-338.
- Goh, C. ve Law, R. (2003). "Incorporating the rough sets theory into travel demand analysis". *Tourism Management*, 24(5): 511-517.
- Goodchild, M. F. (2009). "Geographic information system". S. Aitken ve G. Valentine (Ed.), *Approaches to Human Geography*. Springer, ABD, 251-262.
- Govers, R., Go, F.M. ve Kumar, K. (2007). "Virtual Destination Image: A New Measurement Approach". *Annals of Tourism Research*, 34: 977-997.
- Grinberger, A. Y. ve Shoval, N. (2019). "Spatiotemporal Contingencies in Tourists' Intradaily Mobility Patterns". *Journal of Travel Research*, 58(3): 512-530.
- Guo, Y., Barnes, S. J. ve Jia, Q. (2017). "Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation". *Tourism Management*, 59: 467-483.
- Guo, T., Guo, B., Ouyang, Y., Yu, Z., Lam, J. C. ve Li, V. O. (2018). "CrowdTravel: Scenic spot profiling by using heterogeneous crowdsourced data". *Journal of Ambient Intelligence and Humanized Computing*, 9(6): 2051-2060.
- Güleş, H. K. (2004). "Bilişim Teknolojilerinin Müşteri İlişkileri Yönetimine Katkıları". *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (12): 231-243.
- Gürsoy, U.T.Ş. (2009). *Veri Madenciliği ve Bilgi Keşfi*. Pegem Akademi, Ankara.
- Ha, S. H. ve Park, S. C. (1998). "Application of data mining tools to hotel data mart on the Intranet for database marketing". *Expert Systems with Applications*, 15(1): 1-31.
- Haldrup, M. (2004). "Laid-back mobilities: Second-home holidays in time and space". *Tourism Geographies*, 6(4): 434-454.
- Han, J. ve Kamber M. (2006). *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, San Francisco.
- Hargreaves, C. A. (2015). "Analysis of hotel guest satisfaction ratings and reviews: An application in Singapore". *American Journal of Marketing Research*, 1(4): 208-214.
- Hassani, H., Webster, A., Silva, E. S. ve Heravi, S. (2015). "Forecasting US tourist arrivals using optimal singular spectrum analysis". *Tourism Management*, 46: 322-335.

- Hassani, H., Silva, E. S., Antonakakis, N., Filis, G. ve Gupta, R. (2017). "Forecasting accuracy evaluation of tourist arrivals". *Annals of Tourism Research*, 63: 112-127.
- Hausmann, A., Toivonen, T., Slotow, R., Tenkanen, H., Moilanen, A., Heikinheimo, V. ve Di Minin, E. (2018). "Social media data can be used to understand tourists' preferences for nature- based experiences in protected areas". *Conservation Letters*, 11(1): 1-10.
- He, W., Tian, X., Tao, R., Zhang, W., Yan, G. ve Akula, V. (2017). "Application of social media analytics: a case of analyzing online hotel reviews". *Online Inf. Rev.* 41 (7): 921–935.
- Heckerman, D. (1997). "Bayesian networks for data mining". *Data mining and knowledge discovery*, 1(1): 79-119.
- Hensher, D. A., ve Ton, T. T. (2000). "A comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice". *Transportation Research Part E*, 36(3): 155-172.
- Hoontrakul, P. ve Sahadev, S. (2008). "Application of data mining techniques in the on-line travel industry: A case study from Thailand". *Marketing Intelligence & Planning*, 26(1): 60-76.
- Höpken, W., Ernesti, D., Fuchs, M., Kronenberg, K. ve Lexhagen, M. (2017). "Big data as input for predicting tourist arrivals". *Springer International Publishing. 24-26 Ocak 2017, Roma*, 187-199.
- Hu, F., Li, Z., Yang, C. Ve Jiang, Y. (2019). "A graph-based approach to detecting tourist movement patterns using social media data". *Cartography and Geographic Information Science*, 46(4): 368-382.
- Hu, Y. H. ve Chen, K. (2016). "Predicting hotel review helpfulness: The impact of review visibility, and interaction between hotel stars and review ratings". *International Journal of Information Management*, 36(6): 929-944.
- Hui, F. A. N. G., Chongcheng, C. H. E. N., Jiexiang, L. I. N., Xianfeng, L. I. U. ve Dong, F. A. N. G. (2017). "Association Rule Analysis For Tour Route Recommendation And Application To WCTSNOP". *International Archives of the Photogrammetry, Remote Sensing ve Spatial Information Sciences*, 18–22 September 2017, Wuhan, China, 1121-1126.
- Ivanovic, S., Mikinac, K. ve Perman, L. (2011). "CRM development in hospitality companies for the purpose of increasing the competitiveness in the tourist market". *UTMS journal of economics*, 2(1): 59-68.
- Jiang, J., Gretzel, U. ve Law, R. (2010). "Do negative experiences always lead to dissatisfaction?—testing attribution theory in the context of online travel reviews". *Information*

- and communication technologies in tourism.10–12 Şubat 2010*, Springer, Vienna, 297-308.
- John, G. H. ve Langley, P. (1995). "Estimating continuous distributions in Bayesian classifiers". *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc,18-20 Ağustos 1995, Montreal*, 338-345.
- Jun, W., Yuyan, L., Lingyu, T., & Peng, G. (2018). "Modeling a combined forecast algorithm based on sequence patterns and near characteristics: An application for tourism demand forecasting". *Chaos, Solitons & Fractals*, 108: 136-147.
- Junping, D., Min, Z. ve Xuyan, T. (2008, June). "The realization of distributed sampling association rule mining algorithm in tourism". In *2008 7th World Congress on Intelligent Control and Automation. 25-27 Haziran 2008, Chongqing, Çin*, 183-187
- Karathiya, M. B., Sakshi, R. S., Sakshi, D. S. ve Kathiriya, D. R. (2012). "Data mining for travels and tourism". *Journal of Information and Operations Management*, 3(1): 114-118.
- Karl, M., Reintinger, C. ve Schmude, J. (2015). "Reject or select: Mapping destination choice". *Annals of Tourism Research*, 54: 48-64.
- Kasavana, M. L.(2011). *Management Technology in The Hospitality Industry*. Sixth Edition, AHLA. Michigan.
- Keçeci, S., Özbek, E. E., Türkel, M. S., Düzağaç, R. ve Yıldız, O. T. (2018). "Doğrudan pazarlama amaçlı hedef kitle analizi". *SİU 2018*, 02-05 Mayıs 2018, İzmir.
- Kesorn, K., Juraphanthong, W. ve Salaiwarakul, A. (2017). "Personalized attraction recommendation system for tourists through check-in data". *IEEE Access*, 5: 26703-26721.
- Kim, W.G., Lim, H. ve Brymer, R.A. (2015). "The effectiveness of managing social media on hotel performance". *Int. J. Hospitality Manag.* 44: 165–171.
- Kim, K., Park, O. J., Yun, S., ve Yun, H. (2017). "What makes tourists feel negatively about tourism destinations? Application of hybrid text mining methodology to smart destination management". *Technological Forecasting and Social Change*, 123: 362-369.
- Kim, D. ve Yu, S. J. (2017). "Hotel Review Mining For Targeting Strategy: Focusing On Chinese Free Independent Traveler". *Journal of Theoretical ve Applied Information Technology*, 95(18):4436-4445.
- Knoblock, C. A. (2004). "Building software agents for planning, monitoring, and optimizing travel". *Proceedings of the Eleventh International Conference on Information Technology and Travel and Tourism. 26–28 Ocak 2004, Kahire, Mısır*,1–15.

- Koyun, A. ve Yangeç, D. (2018). "Veri Madenciliği Teknikleri Yardımıyla Otel Yorumlarından Anahtar Kelimeler Keşfi". *Avrupa Bilim ve Teknoloji Dergisi*, (14): 261-268.
- Kon, S. C. ve Turner, L. W. (2005). "Neural network forecasting of tourism demand". *Tourism Economics*, 11(3): 301-328.
- Küçük Yılmaz, A., Durak, M. Ş., Özdemir, E., Kiracı, K. ve Tanrıverid, G. (2017). "Eskişehir İlinde Havayolu Yolcu Potansiyelinin Belirlenmesi Üzerine Bir Karar Ağacı Uygulaması". *Anadolu University of Sciences & Technology-B: Theoretical Sciences*, 5(1): 22-33.
- Larose, D. T. (2006). *Data Mining Methods and Models*. John Wiley & Sons, Inc., Hoboken, New Jersey.
- Lau, K. N., Lee, K. H., Lam, P. Y. ve Ho, Y. (2001). "Web-site marketing for the travel-and-tourism industry". *Cornell Hotel and Restaurant Administration Quarterly*, 42(6): 55-62.
- Lau, K.-N., Lee, K.-H. ve Ho, Y. (2005). "Text mining for the hotel industry". *Cornell Hotel Restaur. Admin. Q.* 46: 344-362.
- Law, R., Bauer, T., Weber, K. ve Tse, T. (2006). "Towards a rough classification of business travelers". *International Conference on Advanced Data Mining and Applications*. 14-16 Ağustos 2006, Xi'an, China, 135-142.
- Law, R., Leung, R. ve Buhalis, D. (2009). "Information technology applications in hospitality and tourism: a review of publications from 2005 to 2007". *Journal of travel ve tourism marketing*, 26(5-6): 599-623.
- Law, R., Rong, J., Vu, H. Q., Li, G. ve Lee, H. A. (2011). "Identifying changes and trends in Hong Kong outbound tourism". *Tourism Management*, 32(5): 1106-1114.
- Law, R. ve Au, N. (2000). "Relationship modeling in tourism shopping: a decision rules induction approach". *Tourism Management*, 21(3): 241-249.
- Law, R. (1998). "Hospitality Data Mining Myths". *Hospitality Review*. 16(1):59-66.
- Leal, F., González-Vélez, H., Malheiro, B. ve Burguillo, J. C. (2017). "Semantic profiling and destination recommendation based on crowd-sourced tourist reviews". *In International Symposium on Distributed Computing and Artificial Intelligence*. 21-23 Haziran 2017, Porto, Portekiz, 140-147.
- Leung, X. Y., Wang, F., Wu, B., Bai, B., Stahura, K. A., ve Xie, Z. (2012). "A social network analysis of overseas tourist movement patterns in Beijing: the Impact of the Olympic Games". *International Journal of Tourism Research* 14 (5):469-484.

- Lew, A., ve McKercher, B. (2006). "Modeling tourist movements: A local destination analysis". *Annals of Tourism Research*, 33(2): 403-423.
- Li, G., Law, R., Vu, H. Q. ve Rong, J. (2013). "Discovering the hotel selection preferences of Hong Kong inbound travelers using the Choquet Integral". *Tourism Management*, 36: 321-330.
- Li, G., Law, R., Vu, H. Q., Rong, J. ve Zhao, X. R. (2015). "Identifying emerging hotel preferences using emerging pattern mining technique". *Tourism management*, 46: 311-321.
- Li, Y., Xiao, L., Ye, Y., Xu, W., ve Law, A. (2016). "Understanding tourist space at a historic site through space syntax analysis: The case of Gulangyu, China". *Tourism Management*, 52: 30-43.
- Li, X., Pan, B., Law, R. ve Huang, X. (2017). "Forecasting tourism demand with composite search index". *Tourism management*, 59: 57-66.
- Li, J., Xu, L., Tang, L., Wang, S. ve Li, L. (2018). "Big data in tourism research: A literature review". *Tourism Management*, 68: 301-323.
- Li, Y., ve Kwortnik, R. (2017). "Categorizing cruise lines by passenger perceived experience". *Journal of Travel Research*, 56(7): 941-956.
- Li, Xinxin, ve Lorin M. Hitt. (2008). "Self-selection and information role of online product reviews". *Information Systems Research* 19 (4): 456-74.
- Liao, S., Chen Y. ve Deng M., "Mining customer knowledge for tourism new product development and customer relationship management". *Expert Systems with Applications* (2010): 4212–4223.
- Line, N. D. ve Runyan, R. C. (2012). "Hospitality marketing research: Recent trends and future directions". *International Journal of Hospitality Management*, 31(2): 477-488.
- Lim, G.G., et al., "An Exploratory Study of The Weather and Calendar Effects on Tourism Web Site Usage". *Emerald Group Publishing Limited*, 34(1): 127-144.
- Liou, J. J., Tang, C. H., Yeh, W. C. ve Tsai, C. Y. (2011). "A decision rules approach for improvement of airport service quality". *Expert Systems with Applications*, 38(11): 13723-13730.
- Litvin, S. W., Goldsmith, R. E. ve Pan, B. (2008). "Electronic word-of-mouth in hospitality and tourism management". *Tourism management*, 29(3): 458-468.
- Liu, Q., Ge, Y., Li, Z., Chen, E. ve Xiong, H. (2011). "Personalized travel package recommendation". *11th International Conference on Data Mining*. 11-14 Aralık, Vancouver, Kanada, 407-416.

- Liu, S., Law, R., Rong, J., Li, G. ve Hall, J. (2013), "Analyzing changes in hotel customers' expectations by trip mode", *International Journal of Hospitality Management*, 34: 359-371.
- Liu, Y., Teichert, T., Rossi, M., Li, H. ve Hu, F. (2017), "Big data for big insights: investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews", *Tourism Management*, 59: 554-563.
- Logesh, R., Subramaniaswamy, V., Vijayakumar, V. ve Li, X. (2018). "Efficient user profiling based intelligent travel recommender system for individual and group of users". *Mobile Networks and Applications*, 1-16.
- Loveman, G. (2003). "Diamonds in the data mine". *Harvard business review*, 81(5): 109-113.
- Lucas, J. P., Luz, N., Moreno, M. N., Anacleto, R., Figueiredo, A. A. ve Martins, C. (2013). "A hybrid recommendation approach for a tourism system". *Expert Systems with Applications*, 40(9): 3532-3550.
- Magnini, V. P., Honeycutt, E. D. ve Hodge, S. K. (2003). "Data mining for hotel firms: use and limitations". *Cornell Hotel ve Restaurant Administration Quarterly*. 94-105.
- Maimon, O. ve Rokach L. (2010). *Data Mining and Knowledge Discovery Handbook*. Springer Science+Business Media, New York.
- Majewska, J. Ve Truskolaski, S. (2018). "Cluster-mapping procedure for tourism regions based on geostatistics and fuzzy clustering: example of Polish districts". *Current Issues in Tourism*, 1-21.
- Mckercher, B. ve Lau, G. (2008). "Movement patterns of tourists within a destination". *Tourism geographies*, 10(3): 355-374.
- Miah, M. S. U., Masuduzzaman, M., Sarkar, W., Islam, H. M., Porag, F. ve Hossain, S. (2018). "Intelligent Tour Planning System using Crowd Sourced Data". *International Journal of Education and Management Engineering (IJEME)*, 8(1): 22-29.
- Mills, J. E., Kook Lee, J. ve Douglas, A. C. (2007). "Exploring Perceptions of US State Tourism Organizations' Web Advertising Effectiveness". *Asia Pacific Journal of Tourism Research*, 12(3): 245-266.
- Min, H., Min, H. ve Emam, A. (2002). "A data mining approach to developing the profiles of hotel customers". *International Journal of Contemporary Hospitality Management*, 14(6): 274-285.
- Mirela, D., Tiberiu, S. ve Adina, B. (2009). "Toward a distributed data mining system for tourism industry". *Annals of Faculty of Economics*, 4(1): 922-926.

- Mor, M. ve Dalyot, S. (2018). "Computing touristic walking routes using geotagged photographs from Flickr". *14th International Conference on Location Based Services. 15-17 Ocak 2018, Zürich, İsviçre*, 63-68.
- Nagao, M., Kawamura, H., Yamamoto, M. ve Ohuchi, A. (2004). "Acquisition of personal tourism activity information based on GPS log mining method". *Fifth Asia Pacific Industrial Engineering and Management Systems Conference*. 37:1-37.
- Nave, M., Rita, P. ve Guerreiro, J. (2018). "A decision support system framework to track consumer sentiments in social media". *Journal of Hospitality Marketing ve Management*, 27(6): 693-710.
- Neidhardt, J., Rümmele, N., ve Werthner, H. (2017). "Predicting happiness: user interactions and sentiment analysis in an online travel forum". *Information Technology ve Tourism*, 17(1): 101-119.
- Nguyen, T. T., Camacho, D. ve Jung, J. E. (2017). "Identifying and ranking cultural heritage resources on geotagged social media for smart cultural tourism services". *Personal and Ubiquitous Computing*, 21(2): 267-279.
- Nguyen, T. T., Camacho, D. ve Jung, J. E. (2017). "Identifying and ranking cultural heritage resources on geotagged social media for smart cultural tourism services". *Personal and Ubiquitous Computing*, 21(2): 267-279.
- Nijkamp, P.; Reggiani, A.; Tritapepe, T. (1996). "Modeling inter-urban transport flows in Italy: A comparison between Neural Network analysis and logit analysis", *Transportation Research Part C: Emerging Technologies*, 4(6): 323-338.
- Nilashi, M., Esfahani, M. D., Roudbaraki, M. Z., Ramayah, T. ve Ibrahim, O. (2016). "A multi-criteria collaborative filtering recommender system using clustering and regression techniques". *Journal of Soft Computing and Decision Support Systems*, 3(5): 24-30.
- North, M. (2012). *Data Mining For the Masses*. Global Text Project. Atina.
- Nuankaew, W., Nuankaew, P., Phanniphong, K. ve Bussaman, S. (2017). "For Discovery: Significant Factors for the Promotion of Tourist Attractions based on Individual Behaviour through Data-mining Techniques". *Chophayom Journal*, 28(3): 88-102.
- O'Connor, Peter. (2010). "Managing a hotel's image on TripAdvisor". *Journal of Hospitality Marketing ve Management* 19 (7): 754-72
- Olmeda, I. ve Sheldon, P. J. (2002). "Data mining techniques and applications for tourism internet marketing". *Journal of Travel & Tourism Marketing*, 11(2-3): 1-20.
- Olson D.L. Ve D.Delen. (2008). *Advances Data Mining Techniques*. Springer, Berlin Almanya.

- Orellana, D., Bregt, A. K., Ligtenberg, A. ve Wachowicz, M. (2012). "Exploring visitor movement patterns in natural recreational areas". *Tourism Management*, 33(3): 672-682.
- Özdemir, A. ve Orçanlı, K. (2012). "İki Aşamalı Kümeleme Algoritması İle Pazar Bölümlemesi, Müşteri Profillerinin Belirlenmesi ve Niş Pazarların Tespiti". *Uşak Üniversitesi Sosyal Bilimler Dergisi*, (5/3): 1-27.
- Özkan, Y. (2008). *Veri Madenciliği Yöntemleri*. Papatya Yayıncılık, İstanbul.
- Öztürk, G. ve Tanrısevdi, A. (2017). "Uluslararası Kruvaziyer Ziyaretçilerine Ait Özelliklerin Birliktelik Kuralı Modeli İle Analizi". *Uluslararası İktisadi ve İdari Bilimler Dergisi*, 3(1): 131-148.
- Pan, S., Lee, J. ve Tsai, H. (2014). "Travel photos: Motivations, image dimensions, and affective qualities of places". *Tourism Management*, 40: 59-69.
- Pan, B., Yang, Y., 2017. Pan, B. ve Yang, Y. (2017). "Forecasting destination weekly hotel occupancy with big data". *Journal of Travel Research*, 56(7): 957-970.
- Pantano, E., Priporas, C.-V. ve Stylos, N. (2017). "You Will Like It!" Using Open Data to Predict Tourists' Response to a Tourist Attraction". *Tourism Management*, 60: 430-438.
- Park, D.-H., Lee, J. ve Han, I. (2007). "The effect of on-line consumer reviews on consumer purchasing intention: The moderating role of involvement". *International Journal of Electronic Commerce*, 11(4): 125-148.
- Patel, K. (2013). "Incremental journey for World Wide Web: Introduced with Web 1.0 to recent Web 5.0 – A survey paper". *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(10): 410-417.
- Patil, P. ve Kolhe, V. (2014). "Survey of travel package recommendation system". *International Journal of Science and Research* 3(12): 1557-1561.
- Pattie, D.C. ve Snyder, J. (1996). "Using a Neural Network to Forecast Visitor Behavior". *Annals of Tourism Research*, 23: 151-164.
- Peng, B., Song, H. ve Crouch, G. I. (2014). "A meta-analysis of international tourism demand forecasting and implications for practice". *Tourism Management*, 45: 181-193.
- Petković, G., Pindžo, R. ve Agić-Molnar, M. (2018). "Digital economy and (non) incremental changes in tourism and retail business model". *Ekonomika preduzeća*, 66(1-2): 151-165.
- Petropoulos, C., Nikolopoulos*, K., Patelis, A. ve Assimakopoulos, V. (2005). "A technical analysis approach to tourism demand forecasting". *Applied Economics Letters*, 12(6): 327-333.

- Phillips, P., Zigan, K., Silva, M. M. S. ve Schegg, R. (2015). "The interactive effects of online reviews on the determinants of Swiss hotel performance: A neural network analysis". *Tourism Management*, 50: 130-141.
- Pilepic, L. ve Simunic, M. (2009). "Applying Information Technology To Business Decision-Making In The Hotel Enterprises". *Ekonomiska misao i praksa*, 18(2): 411.
- Praničević, D. G.(2006). "Application of Information and Communication Technologies (ICT) in Tourism" *3rd International Conference:An Enterprise Odyssey: Integration or Disintegration*. Zagreb: University of Zagreb, Faculty of Economics and Business, 925-932.
- Prideaux, B. (2000). "The role of the transport system in destination development". *Tourism management*, 21(1): 53-63.
- Puri, C., ve Kumar, N. (2017). "Opinion Ensembling for Improving Economic Growth through Tourism". *Procedia computer science*, 122: 237-244.
- Pyo, S., Uysal, M. ve Chang, H. (2002). "Knowledge discovery in database for tourist destinations". *Journal of Travel Research*, 40(4): 374-384.
- Qu, Z., Zhang, H. ve Li, H. (2008). "Determinants of online merchant rating: Content analysis of consumer comments about Yahoo merchants". *Decision Support Systems*, 46(1): 440-449.
- Raun, J., Ahas, R. ve Tiru, M. (2016). "Measuring tourism destinations using mobile tracking data". *Tourism Management*, 57: 202-212.
- Rong, J., Vu, H. Q., Law, R. ve Li, G. (2012). "A behavioral analysis of web sharers and browsers in Hong Kong using targeted association rule mining". *Tourism Management*, 33(4): 731-740.
- Rud, O.P. (2001). *Data Mining Cookbook; Modeling Data for Marketing, Risk, and Customer Relationship Management*. John Wiley & Sons, Inc, New York.
- Sağır, M. (2010). "Stratejik yönetim sürecinde ürün stratejilerinin kullanımı ve önemi". *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (24): 311-321.
- Scherrer, L., Tomko, M., Ranacher, P. ve Weibel, R. (2018). "Travelers or locals? Identifying meaningful sub-populations from human movement data in the absence of ground truth". *EPJ Data Science*, 7(1): 19-40.
- Schuckert, M., Liu, X. ve Law, R. (2015). "A segmentation of online reviews by language groups: How English and non-English speakers rate hotels differently". *International Journal of Hospitality Management*, 48: 143-149.

- Septiadi, H. M., Ariandika, C., ve Alamsyah, A. (2018). "Prediction Models Based on Flight Tickets and Hotel Rooms Data Sales for Recommendation System in Online Travel Agent Business". *Sustainable Collaboration in Business, Technology, Information and Innovation (SCBTII)*. 15-16 Agustus 2016, Batı Cava, Endonezya, 210-215.
- Shao, H., Zhang, Y. ve Li, W. (2017). "Extraction and analysis of city's tourism districts based on social media data". *Computers, Environment and Urban Systems*, 65: 66-78.
- Shapoval, V., Wang, M. C., Hara, T. ve Shioya, H. (2018). "Data Mining in Tourism Data Analysis: Inbound Visitors to Japan". *Journal of Travel Research*, 57(3): 310-323.
- Shelar, S., Kamat, P., Varpe, A. ve Birajdar, A. (2018). "TRAVELMATE Travel Package Recommendation System". *International Research Journal of Engineering and Technology*, 5(3): 4095-4097.
- Shi, B., Zhao, J. ve Chen, P. J. (2017). "Exploring urban tourism crowding in Shanghai via crowdsourcing geospatial data". *Current Issues in Tourism*, 20(11): 1186-1209.
- Shmueli, D., Salomon, I. ve Shefer, D. (1996). "Neural network analysis of travel behavior: evaluating tools for prediction". *Transportation Research Part C: Emerging Technologies*, 4(3): 151-166.
- Siering, M., Deokar, A. V. ve Janze, C. (2018). "Disentangling consumer recommendations: Explaining and predicting airline recommendations based on online reviews". *Decision Support Systems*, 107: 52-63.
- Singh, N., Hu, C. ve Roehl, W. S. (2007). "Text mining a decade of progress in hospitality human resource management research: Identifying emerging thematic development". *International Journal of Hospitality Management*, 26(1): 131-147.
- Soava, G. (2015). "Development prospects of the tourism industry in the digital age". *Young Economists Journal / Revista Tinerilor Economiști*, 12(25): 101-116.
- Song, H. ve Li, G. (2008). "Tourism demand modelling and forecasting—A review of recent research". *Tourism management*, 29(2): 203-220.
- Sun, X., Huang, Z., Peng, X., Chen, Y. ve Liu, Y. (2018). "Building a model-based personalised recommendation approach for tourist attractions from geotagged social media data". *International Journal of Digital Earth*, 12(6): 661-678.
- Sutjiadi, R., Trianto, E. M. ve Budihardjo, A. G. (2018). "Surabaya Tourism Destination Recommendation Using Fuzzy C-Means Algorithm". *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 10(2-3): 177-181.
- Talón-Ballesteros, P., González-Serrano, L., Soguero-Ruiz, C., Muñoz-Romero, S. ve Rojo-Álvarez, J. L. (2018). "Using big data from customer relationship management

- information systems to determine the client profile in the hotel sector". *Tourism Management*, 68: 187-197.
- Tavakoli, R. ve Mura, P. (2018). "Netnography in tourism–Beyond Web 2.0". *Annals of Tourism Research*, 73 (C): 190-192.
- Tchetchik, A., Fleischer, A. ve Shoval, N. (2009). "Segmentation of visitors to a heritage site using high-resolution time-space data". *Journal of Travel Research*, 48(2): 216-229.
- Thomaz, G. M., Biz, A. A., Bettoni, E. M., Mendes-Filho, L. ve Buhalis, D. (2017). "Content mining framework in social media: A FIFA world cup 2014 case analysis". *Information & Management*, 54(6): 786-801.
- Tilly, R., Fischbach, K. ve Schoder, D. (2015). "Mineable or messy? Assessing the quality of macro-level tourism information derived from social media". *Electronic Markets*, 25(3): 227-241.
- Ting, P. H., Pan, S. ve Chou, S. S. (2010). "Finding ideal menu items assortments: An empirical application of market basket analysis". *Cornell Hospitality Quarterly*, 51(4): 492-501.
- Toral, S. L., Martínez-Torres, M. R. ve Gonzalez-Rodriguez, M. R. (2018). "Identification of the unique attributes of tourist destinations from online reviews. *Journal of Travel Research*, 57(7): 908-919.
- Uysal, M. (2004). "Advancement in Computing: Implications for Tourism and Hospitality". *Scandinavian Journal of Hospitality and Tourism*, 4 (3): 208-224.
- Üngören, E. ve Doğan, H. (2010), "Beş Yıldızlı Konaklama İşletmelerinde Çalışanların İş Tatmin Düzeylerinin Chaid Analiz Yöntemiyle Değerlendirilmesi" *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 11(2): 39-52
- Vajirakachorn, T. ve Chongwatpol, J. (2017). "Application of business intelligence in the tourism industry: A case study of a local food festival in Thailand". *Tourism Management Perspectives*, 23: 75-86.
- Valls, A., Gibert, K., Orellana, A. ve Antón-Clavé, S. (2018). "Using ontology-based clustering to understand the push and pull factors for British tourists visiting a Mediterranean coastal destination". *Information ve Management*, 55(2): 145-159.
- Veloso, B., Leal, F., González-Vélez, H., Malheiro, B. ve Burguillo, J. C. (2018). "Scalable data analytics using crowdsourced repositories and streams". *Journal of Parallel and Distributed Computing*, 122: 1-10.
- Verma, R. (2010). "Customer choice modeling in hospitality services: A review of past research and discussion of some new applications". *Cornell Hospitality Quarterly*, 51(4): 470-478.

- Versichele, M., De Groote, L., Bouuaert, M. C., Neutens, T., Moerman, I. ve Van de Weghe, N. (2014). "Pattern mining in tourist attraction visits through association rule learning on Bluetooth tracking data: A case study of Ghent, Belgium". *Tourism Management*, 44: 67-81.
- Vu, H. Q., Li, G., Law, R. ve Zhang, Y. (2017). "Tourist Activity Analysis by Leveraging Mobile Social Media Data". *Journal of Travel Research*, 57(7): 883-898.
- Vu, H. Q., Li, G., Law, R. ve Ye, B. H. (2015). "Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos". *Tourism Management*, 46: 222-232.
- Vu, H. Q., Li, G., Law, R. ve Zhang, Y. (2018). "Travel diaries analysis by sequential rule mining". *Journal of travel research*, 57(3): 399-413.
- Wan, L., Hong, Y., Huang, Z., Peng, X. ve Li, R. (2018). "A hybrid ensemble learning method for tourist route recommendations based on geo-tagged social networks". *International Journal of Geographical Information Science*, 32(11): 2225-2246.
- Wang, Z., He, S. Y. ve Leung, Y. (2018). "Applying mobile phone data to travel behaviour research: A literature review". *Travel Behaviour and Society*, 11: 141-155.
- Werthner, H. ve Ricci, F. (2004). "E-commerce and tourism". *Communications of the ACM*, 47(12): 101-105.
- Wong, J., Chen, H., Chung, P. ve Kao, N. (2006). "Identifying Valuable Travelers and their Next Foreign Destination by the Application of Data Mining Techniques". *Asia Pacific Journal of Tourism Research*, 11(4): 355-373.
- Wong, K. K., Song, H. ve Chon, K. S. (2006). "Bayesian models for tourism demand forecasting". *Tourism Management*, 27(5): 773-780.
- Wu, E. H., Law, R. ve Jiang, B. (2010). "Data mining for hotel occupancy rate: an independent component analysis approach". *Journal of travel & tourism marketing*, 27(4): 426-438.
- Wu, E. H., Law, R. ve Jiang, B. (2013). "Predicting browsers and purchasers of hotel websites: A weight-of-evidence grouping approach". *Cornell Hospitality Quarterly*, 54(1): 38-48.
- Wu, L., Shen, H., Fan, A. ve Mattila, A. S. (2017). "The impact of language style on consumers' reactions to online reviews". *Tourism Management*, 59: 590-596.
- Xiang, Z. ve Gretzel, U. (2010). "Role of social media in online travel information search". *Tourism management*, 31(2): 179-188.
- Xiang, Z., Du, Q., Ma, Y. ve Fan, W. (2018). "Assessing reliability of social media data: lessons from mining TripAdvisor hotel reviews". *Information Technology ve Tourism*, 18(1-4): 43-59.

- Xie, C., Lu, J. ve Parkany, E. (2003). "Work travel mode choice modeling with data mining: decision trees and neural networks". *Transportation Research Record*, 1854(1): 50-61.
- Xie, K. L., Zhang, Z., ve Hang, Z. (2014). "The business value of online consumer reviews and management response to hotel performance". *International Journal of Hospitality Management*, 43: 1-12.
- Xie, K., Kwok, L. ve Wang, W. (2017). "Monetizing managerial responses on TripAdvisor: Performance implications across hotel classes". *Cornell Hospitality Quarterly*, 58(3): 240-252.
- Xu, X. (2018). "Does traveler satisfaction differ in various travel group compositions? Evidence from online reviews". *International Journal of Contemporary Hospitality Management*, 30 (3): 1663-1685.
- Xu, X. ve Li, Y. (2016). "The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach". *International journal of hospitality management*, 55: 57-69.
- Sun, Yao., Ma, H. ve Chan, E. (2017). "A Model to Measure Tourist Preference toward Scenic Spots Based on Social Media Data: A Case of Dapeng in China". *Sustainability*, 10(1): 1-13.
- Ye, Q., Law, R., Gu, B. ve Chen, W. (2011). "The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings". *Computers in Human Behavior*, 27(2): 634-639.
- Yotsawat, W. ve Srivihok, A. (2015). "Rules Mining Based on Clustering of Inbound Tourists in Thailand". In *Advanced Computer and Communication Engineering Technology* . Springer, Cham, 693-705.
- Van der Zee, E., Bertocchi, D. ve Vanneste, D. (2018). "Distribution of tourists within urban heritage destinations: a hot spot/cold spot analysis of TripAdvisor data as support for destination management". *Current Issues in Tourism*, 1-22.
- Zhang, Z., Ye, Q., Law, R. ve Li, Y. (2010). "The impact of e-word-of-mouth on the online popularity of restaurants: a comparison of consumer reviews and editor reviews". *Int. J. Hospitality Manag.* 29 (4): 694–700.
- Zhang, X., Yu, Y., Li, H. ve Lin, Z. (2016). "Sentimental interplay between structured and unstructured user-generated contents: an empirical study on online hotel reviews". *Online Inf. Rev.* 40 (1): 119–145.
- Zhang, J. ve Piramuthu, S. (2018). "Product recommendation with latent review topics". *Information Systems Frontiers*, 20(3): 617-625.

- Zhao, Y., Xu, X. ve Wang, M. (2019). "Predicting overall customer satisfaction: big data evidence from hotel online textual reviews". *International Journal of Hospitality Management*, 76: 111-121.
- Zheng, W., Huang, X. ve Li, Y. (2017). "Understanding the tourist mobility using GPS: Where is the next place?". *Tourism Management*, 59: 267-280.
- Zheng, X., Luo, Y., Sun, L., Zhang, J. ve Chen, F. (2018). "A tourism destination recommender system using users' sentiment and temporal dynamics". *Journal of Intelligent Information Systems*, 51(3): 557-578.
- Zhou, L., Ye, S., Pearce, P. L. ve Wu, M.-Y. (2014). "Refreshing hotel satisfaction studies by reconfiguring customer review data". *International Journal of Hospitality Management*, 38: 1-10.
- Zhou, X., Wang, M. ve Li, D. (2017). "From stay to play—A travel planning tool based on crowdsourcing user-generated contents". *Applied geography*, 78: 1-11.
- Zhu, G., Cao, J., Li, C. ve Wu, Z. (2017). "A recommendation engine for travel products based on topic sequential patterns". *Multimedia Tools and Applications*, 76(16): 17595-17612.
- Zisos, I., Matsatsinis, N. ve Tsotsolas, N. (2018). "A Hybrid Multi-Criteria Hotel Recommender System Using Online Ratings and Reviews". *7th International Symposium and 29th National Conference on Operational Research The contribution of Operational Research, new technologies and innovation in agriculture and tourism*. June 14-16, 2018, Hanya, Girit, 88-93.

Ek-1. Karar Ağacı Kuralları

Bölge = ALY

- | Yetiskin Yaşı = 18-30
 - | | Konaklama Süresi = --7: ALMAYAN {ALMAYAN=8, SATIN ALAN=0}
 - | | Konaklama Süresi = 14-20
 - | | | Müşteri Tipi = ARKDS: SATIN ALAN {ALMAYAN=1, SATIN ALAN=2}
 - | | | Müşteri Tipi = AİLE: ALMAYAN {ALMAYAN=2, SATIN ALAN=1}
 - | | | Müşteri Tipi = GRUP: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 - | | | Müşteri Tipi = TEK: ALMAYAN {ALMAYAN=3, SATIN ALAN=0}
 - | | | Müşteri Tipi = ÇİFT: SATIN ALAN {ALMAYAN=5, SATIN ALAN=11}
 - | | Konaklama Süresi = 7-13
 - | | | Müşteri Tipi = ARKDS: ALMAYAN {ALMAYAN=17, SATIN ALAN=0}
 - | | | Müşteri Tipi = AİLE: ALMAYAN {ALMAYAN=8, SATIN ALAN=0}
 - | | | Müşteri Tipi = GRUP: ALMAYAN {ALMAYAN=1, SATIN ALAN=1}
 - | | | Müşteri Tipi = TEK: ALMAYAN {ALMAYAN=2, SATIN ALAN=1}
 - | | | Müşteri Tipi = ÇİFT: ALMAYAN {ALMAYAN=63, SATIN ALAN=7}
- | Yetiskin Yaşı = 31-59
 - | | Çocuk Bilgisi = 1-0: ALMAYAN {ALMAYAN=4, SATIN ALAN=0}
 - | | Çocuk Bilgisi = 1-1: ALMAYAN {ALMAYAN=19, SATIN ALAN=0}
 - | | Çocuk Bilgisi = 1-2: ALMAYAN {ALMAYAN=79, SATIN ALAN=0}
 - | | Çocuk Bilgisi = 1-3
 - | | | Müşteri Tipi = AİLE: ALMAYAN {ALMAYAN=60, SATIN ALAN=1}
 - | | | Müşteri Tipi = TEK: ALMAYAN {ALMAYAN=3, SATIN ALAN=0}
 - | | | Müşteri Tipi = ÇCK-B: ALMAYAN {ALMAYAN=8, SATIN ALAN=0}
 - | | | Müşteri Tipi = ÇİFT: SATIN ALAN {ALMAYAN=0, SATIN ALAN=1}
 - | | Çocuk Bilgisi = 2-01: ALMAYAN {ALMAYAN=3, SATIN ALAN=0}
 - | | Çocuk Bilgisi = 2-02: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 - | | Çocuk Bilgisi = 2-03: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 - | | Çocuk Bilgisi = 2-11: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 - | | Çocuk Bilgisi = 2-12: ALMAYAN {ALMAYAN=11, SATIN ALAN=0}
 - | | Çocuk Bilgisi = 2-22: ALMAYAN {ALMAYAN=21, SATIN ALAN=0}
 - | | Çocuk Bilgisi = 2-23: ALMAYAN {ALMAYAN=18, SATIN ALAN=0}
 - | | Çocuk Bilgisi = 2-33: ALMAYAN {ALMAYAN=8, SATIN ALAN=0}
 - | | Çocuk Bilgisi = 3++: ALMAYAN {ALMAYAN=21, SATIN ALAN=0}
 - | | Çocuk Bilgisi = YOK: ALMAYAN {ALMAYAN=267, SATIN ALAN=0}
- | Yetiskin Yaşı = 60++
 - | | Otel_Ücreti = ORTA
 - | | | Konaklama Süresi = 14-20: ALMAYAN {ALMAYAN=120, SATIN ALAN=2}
 - | | | Konaklama Süresi = 21++: ALMAYAN {ALMAYAN=31, SATIN ALAN=0}
 - | | | Konaklama Süresi = 7-13: ALMAYAN {ALMAYAN=66, SATIN ALAN=0}
 - | | Otel_Ücreti = PAHALI: ALMAYAN {ALMAYAN=132, SATIN ALAN=0}
 - | | Otel_Ücreti = UCUZ: ALMAYAN {ALMAYAN=88, SATIN ALAN=0}
 - | | Otel_Ücreti = ÇOK PAHALI: ALMAYAN {ALMAYAN=7, SATIN ALAN=0}
 - | | Otel_Ücreti = ÇOK UCUZ: ALMAYAN {ALMAYAN=11, SATIN ALAN=0}

Bölge = AYT

- | Yetiskin Yaşı = 18-30
 - | | Müşteri Tipi = ARKDS
 - | | | Otel_Ücreti = ORTA: SATIN ALAN {ALMAYAN=0, SATIN ALAN=5}
 - | | | Otel_Ücreti = PAHALI: SATIN ALAN {ALMAYAN=1, SATIN ALAN=9}
 - | | | Otel_Ücreti = UCUZ: SATIN ALAN {ALMAYAN=0, SATIN ALAN=2}

| | | Otel_Ücreti = ÇOK PAHALI: SATIN ALAN {ALMAYAN=0, SATIN ALAN=29}
 | | Müşteri Tipi = AİLE
 | | | Çocuk Bilgisi = 1-0: ALMAYAN {ALMAYAN=3, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 1-1: SATIN ALAN {ALMAYAN=1, SATIN ALAN=19}
 | | | Çocuk Bilgisi = 1-2: SATIN ALAN {ALMAYAN=2, SATIN ALAN=35}
 | | | Çocuk Bilgisi = 2-12: SATIN ALAN {ALMAYAN=1, SATIN ALAN=10}
 | | | Çocuk Bilgisi = 2-22: SATIN ALAN {ALMAYAN=0, SATIN ALAN=8}
 | | | Çocuk Bilgisi = YOK: SATIN ALAN {ALMAYAN=0, SATIN ALAN=6}
 | | Müşteri Tipi = GRUP
 | | | Otel_Ücreti = ORTA: SATIN ALAN {ALMAYAN=0, SATIN ALAN=1}
 | | | Otel_Ücreti = PAHALI: SATIN ALAN {ALMAYAN=0, SATIN ALAN=4}
 | | | Otel_Ücreti = ÇOK PAHALI: SATIN ALAN {ALMAYAN=1, SATIN ALAN=13}
 | | Müşteri Tipi = TEK
 | | | Çocuk Bilgisi = 1-1: ALMAYAN {ALMAYAN=5, SATIN ALAN=1}
 | | | Çocuk Bilgisi = 1-2: ALMAYAN {ALMAYAN=9, SATIN ALAN=1}
 | | | Çocuk Bilgisi = YOK: SATIN ALAN {ALMAYAN=0, SATIN ALAN=11}
 | | Müşteri Tipi = ÇCK-B
 | | | Çocuk Bilgisi = 1-0: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 1-1: ALMAYAN {ALMAYAN=6, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 1-2: ALMAYAN {ALMAYAN=6, SATIN ALAN=1}
 | | | Çocuk Bilgisi = 2-01: ALMAYAN {ALMAYAN=3, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-02: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-11: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-12: ALMAYAN {ALMAYAN=1, SATIN ALAN=1}
 | | Müşteri Tipi = ÇİFT
 | | | Konaklama Süresi = --7: SATIN ALAN {ALMAYAN=0, SATIN ALAN=3}
 | | | Konaklama Süresi = 14-20: SATIN ALAN {ALMAYAN=1, SATIN ALAN=94}
 | | | Konaklama Süresi = 21++: SATIN ALAN {ALMAYAN=0, SATIN ALAN=3}
 | | | Konaklama Süresi = 7-13: SATIN ALAN {ALMAYAN=8, SATIN ALAN=207}
 | Yetiskin Yaşı = 31-59
 | | Müşteri Tipi = ARKDS
 | | | Otel_Ücreti = ORTA: ALMAYAN {ALMAYAN=4, SATIN ALAN=0}
 | | | Otel_Ücreti = PAHALI: ALMAYAN {ALMAYAN=6, SATIN ALAN=0}
 | | | Otel_Ücreti = ÇOK PAHALI: ALMAYAN {ALMAYAN=5, SATIN ALAN=1}
 | | Müşteri Tipi = AİLE
 | | | Çocuk Bilgisi = 1-0: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 1-1: ALMAYAN {ALMAYAN=8, SATIN ALAN=1}
 | | | Çocuk Bilgisi = 1-2: SATIN ALAN {ALMAYAN=20, SATIN ALAN=23}
 | | | Çocuk Bilgisi = 1-3: SATIN ALAN {ALMAYAN=26, SATIN ALAN=34}
 | | | Çocuk Bilgisi = 2-00: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-01: ALMAYAN {ALMAYAN=3, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-02: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-11: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-12: ALMAYAN {ALMAYAN=8, SATIN ALAN=3}
 | | | Çocuk Bilgisi = 2-13: SATIN ALAN {ALMAYAN=0, SATIN ALAN=1}
 | | | Çocuk Bilgisi = 2-22: SATIN ALAN {ALMAYAN=7, SATIN ALAN=10}
 | | | Çocuk Bilgisi = 2-23: SATIN ALAN {ALMAYAN=5, SATIN ALAN=7}
 | | | Çocuk Bilgisi = 2-33: SATIN ALAN {ALMAYAN=0, SATIN ALAN=6}
 | | | Çocuk Bilgisi = 3++: ALMAYAN {ALMAYAN=3, SATIN ALAN=1}
 | | Müşteri Tipi = GRUP
 | | | Konaklama Süresi = 14-20: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}

| | | Konaklama Süresi = 7-13: SATIN ALAN {ALMAYAN=1, SATIN ALAN=2}
 | | Müşteri Tipi = TEK
 | | | Çocuk Bilgisi = 1-1: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 1-2: ALMAYAN {ALMAYAN=47, SATIN ALAN=3}
 | | | Çocuk Bilgisi = 1-3: ALMAYAN {ALMAYAN=29, SATIN ALAN=5}
 | | | Çocuk Bilgisi = 2-22: ALMAYAN {ALMAYAN=11, SATIN ALAN=1}
 | | | Çocuk Bilgisi = 2-23: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-33: SATIN ALAN {ALMAYAN=0, SATIN ALAN=1}
 | | | Çocuk Bilgisi = YOK: ALMAYAN {ALMAYAN=2, SATIN ALAN=1}
 | | Müşteri Tipi = ÇCK-B
 | | | Çocuk Bilgisi = 1-1: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 1-2: ALMAYAN {ALMAYAN=22, SATIN ALAN=1}
 | | | Çocuk Bilgisi = 1-3: ALMAYAN {ALMAYAN=11, SATIN ALAN=4}
 | | | Çocuk Bilgisi = 2-02: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-12: ALMAYAN {ALMAYAN=3, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-22: ALMAYAN {ALMAYAN=8, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-23: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-33: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 3++: ALMAYAN {ALMAYAN=3, SATIN ALAN=0}
 | | | Çocuk Bilgisi = YOK: SATIN ALAN {ALMAYAN=0, SATIN ALAN=1}
 | | Müşteri Tipi = ÇİFT
 | | | Otel_Ücreti = ORTA: ALMAYAN {ALMAYAN=8, SATIN ALAN=0}
 | | | Otel_Ücreti = PAHALI: ALMAYAN {ALMAYAN=13, SATIN ALAN=3}
 | | | Otel_Ücreti = UCUZ: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Otel_Ücreti = ÇOK PAHALI: ALMAYAN {ALMAYAN=39, SATIN ALAN=4}
 | | Yetiskin Yaşı = 60++
 | | Konaklama Süresi = --7: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | Konaklama Süresi = 14-20
 | | | Çocuk Bilgisi = 1-1: ALMAYAN {ALMAYAN=3, SATIN ALAN=1}
 | | | Çocuk Bilgisi = 1-2: ALMAYAN {ALMAYAN=3, SATIN ALAN=3}
 | | | Çocuk Bilgisi = 1-3: ALMAYAN {ALMAYAN=2, SATIN ALAN=1}
 | | | Çocuk Bilgisi = 2-01: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-02: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 3++: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = YOK: ALMAYAN {ALMAYAN=131, SATIN ALAN=33}
 | | Konaklama Süresi = 21++
 | | | Otel_Ücreti = ORTA: ALMAYAN {ALMAYAN=3, SATIN ALAN=0}
 | | | Otel_Ücreti = PAHALI: ALMAYAN {ALMAYAN=15, SATIN ALAN=1}
 | | | Otel_Ücreti = UCUZ: ALMAYAN {ALMAYAN=3, SATIN ALAN=0}
 | | | Otel_Ücreti = ÇOK PAHALI: ALMAYAN {ALMAYAN=5, SATIN ALAN=0}
 | | Konaklama Süresi = 7-13
 | | | Çocuk Bilgisi = 1-1: ALMAYAN {ALMAYAN=5, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 1-2: ALMAYAN {ALMAYAN=7, SATIN ALAN=1}
 | | | Çocuk Bilgisi = 1-3: ALMAYAN {ALMAYAN=5, SATIN ALAN=2}
 | | | Çocuk Bilgisi = 2-11: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-12: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-22: ALMAYAN {ALMAYAN=5, SATIN ALAN=2}
 | | | Çocuk Bilgisi = 2-23: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-33: ALMAYAN {ALMAYAN=1, SATIN ALAN=1}
 | | | Çocuk Bilgisi = YOK: ALMAYAN {ALMAYAN=114, SATIN ALAN=20}
 Bölge = BLK

| Yetiskin Yaşı = 18-30
 | | Müşteri Tipi = ARKDS
 | | | Konaklama Süresi = --7: ALMAYAN {ALMAYAN=1, SATIN ALAN=1}
 | | | Konaklama Süresi = 14-20: SATIN ALAN {ALMAYAN=0, SATIN ALAN=16}
 | | | Konaklama Süresi = 7-13: SATIN ALAN {ALMAYAN=1, SATIN ALAN=69}
 | | Müşteri Tipi = AİLE
 | | | Çocuk Bilgisi = 1-0: ALMAYAN {ALMAYAN=3, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 1-1: SATIN ALAN {ALMAYAN=1, SATIN ALAN=51}
 | | | Çocuk Bilgisi = 1-2: SATIN ALAN {ALMAYAN=1, SATIN ALAN=63}
 | | | Çocuk Bilgisi = 1-3: SATIN ALAN {ALMAYAN=0, SATIN ALAN=1}
 | | | Çocuk Bilgisi = 2-01: SATIN ALAN {ALMAYAN=2, SATIN ALAN=3}
 | | | Çocuk Bilgisi = 2-11: SATIN ALAN {ALMAYAN=1, SATIN ALAN=17}
 | | | Çocuk Bilgisi = 2-12: SATIN ALAN {ALMAYAN=0, SATIN ALAN=34}
 | | | Çocuk Bilgisi = 2-22: SATIN ALAN {ALMAYAN=0, SATIN ALAN=16}
 | | | Çocuk Bilgisi = 2-23: SATIN ALAN {ALMAYAN=0, SATIN ALAN=2}
 | | | Çocuk Bilgisi = 2-33: SATIN ALAN {ALMAYAN=0, SATIN ALAN=1}
 | | | Çocuk Bilgisi = 3++: SATIN ALAN {ALMAYAN=1, SATIN ALAN=2}
 | | | Çocuk Bilgisi = YOK: SATIN ALAN {ALMAYAN=0, SATIN ALAN=6}
 | | Müşteri Tipi = GRUP: SATIN ALAN {ALMAYAN=0, SATIN ALAN=48}
 | | Müşteri Tipi = TEK
 | | | Çocuk Bilgisi = 1-1: ALMAYAN {ALMAYAN=20, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 1-2: ALMAYAN {ALMAYAN=4, SATIN ALAN=2}
 | | | Çocuk Bilgisi = 2-22: SATIN ALAN {ALMAYAN=0, SATIN ALAN=1}
 | | | Çocuk Bilgisi = YOK: SATIN ALAN {ALMAYAN=1, SATIN ALAN=25}
 | | Müşteri Tipi = ÇCK-B
 | | | Çocuk Bilgisi = 1-0: ALMAYAN {ALMAYAN=4, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 1-1: ALMAYAN {ALMAYAN=20, SATIN ALAN=1}
 | | | Çocuk Bilgisi = 1-2: ALMAYAN {ALMAYAN=5, SATIN ALAN=2}
 | | | Çocuk Bilgisi = 2-00: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-01: ALMAYAN {ALMAYAN=4, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-11: ALMAYAN {ALMAYAN=3, SATIN ALAN=3}
 | | | Çocuk Bilgisi = 2-12: ALMAYAN {ALMAYAN=1, SATIN ALAN=1}
 | | | Çocuk Bilgisi = 2-22: SATIN ALAN {ALMAYAN=0, SATIN ALAN=1}
 | | | Çocuk Bilgisi = 2-23: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 3++: SATIN ALAN {ALMAYAN=0, SATIN ALAN=1}
 | | Müşteri Tipi = ÇİFT
 | | | Otel_Ücreti = ORTA: SATIN ALAN {ALMAYAN=1, SATIN ALAN=140}
 | | | Otel_Ücreti = PAHALI: SATIN ALAN {ALMAYAN=7, SATIN ALAN=320}
 | | | Otel_Ücreti = UCUZ: SATIN ALAN {ALMAYAN=0, SATIN ALAN=1}
 | | | Otel_Ücreti = ÇOK PAHALI: SATIN ALAN {ALMAYAN=0, SATIN ALAN=18}
 | | | Otel_Ücreti = ÇOK UCUZ: SATIN ALAN {ALMAYAN=0, SATIN ALAN=47}
 | Yetiskin Yaşı = 31-59
 | | Müşteri Tipi = ARKDS
 | | | Konaklama Süresi = --7: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Konaklama Süresi = 14-20: SATIN ALAN {ALMAYAN=0, SATIN ALAN=1}
 | | | Konaklama Süresi = 7-13: ALMAYAN {ALMAYAN=6, SATIN ALAN=1}
 | | Müşteri Tipi = AİLE
 | | | Çocuk Bilgisi = 1-0: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 1-1: ALMAYAN {ALMAYAN=6, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 1-2: SATIN ALAN {ALMAYAN=31, SATIN ALAN=61}
 | | | Çocuk Bilgisi = 1-3: SATIN ALAN {ALMAYAN=20, SATIN ALAN=141}

| | | Çocuk Bilgisi = 2-01: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-03: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-11: SATIN ALAN {ALMAYAN=0, SATIN ALAN=1}
 | | | Çocuk Bilgisi = 2-12: SATIN ALAN {ALMAYAN=6, SATIN ALAN=12}
 | | | Çocuk Bilgisi = 2-13: SATIN ALAN {ALMAYAN=0, SATIN ALAN=4}
 | | | Çocuk Bilgisi = 2-22: SATIN ALAN {ALMAYAN=7, SATIN ALAN=31}
 | | | Çocuk Bilgisi = 2-23: SATIN ALAN {ALMAYAN=7, SATIN ALAN=65}
 | | | Çocuk Bilgisi = 2-33: SATIN ALAN {ALMAYAN=4, SATIN ALAN=31}
 | | | Çocuk Bilgisi = 3++: SATIN ALAN {ALMAYAN=2, SATIN ALAN=16}
 | | | Çocuk Bilgisi = YOK: SATIN ALAN {ALMAYAN=0, SATIN ALAN=1}
 | | Müşteri Tipi = GRUP: SATIN ALAN {ALMAYAN=0, SATIN ALAN=1}
 | | Müşteri Tipi = TEK
 | | | Çocuk Bilgisi = 1-2: ALMAYAN {ALMAYAN=43, SATIN ALAN=13}
 | | | Çocuk Bilgisi = 1-3: ALMAYAN {ALMAYAN=33, SATIN ALAN=10}
 | | | Çocuk Bilgisi = 2-12: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-22: ALMAYAN {ALMAYAN=21, SATIN ALAN=1}
 | | | Çocuk Bilgisi = 2-23: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-33: ALMAYAN {ALMAYAN=5, SATIN ALAN=1}
 | | | Çocuk Bilgisi = 3++: ALMAYAN {ALMAYAN=1, SATIN ALAN=1}
 | | | Çocuk Bilgisi = YOK: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | Müşteri Tipi = ÇCK-B
 | | | Çocuk Bilgisi = 1-2: ALMAYAN {ALMAYAN=14, SATIN ALAN=3}
 | | | Çocuk Bilgisi = 1-3: ALMAYAN {ALMAYAN=17, SATIN ALAN=8}
 | | | Çocuk Bilgisi = 2-02: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-12: ALMAYAN {ALMAYAN=4, SATIN ALAN=2}
 | | | Çocuk Bilgisi = 2-13: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-22: ALMAYAN {ALMAYAN=6, SATIN ALAN=5}
 | | | Çocuk Bilgisi = 2-23: SATIN ALAN {ALMAYAN=1, SATIN ALAN=5}
 | | | Çocuk Bilgisi = 2-33: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 3++: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | Müşteri Tipi = ÇİFT
 | | | Otel_Ücreti = ORTA: ALMAYAN {ALMAYAN=6, SATIN ALAN=2}
 | | | Otel_Ücreti = PAHALI: ALMAYAN {ALMAYAN=36, SATIN ALAN=4}
 | | | Otel_Ücreti = ÇOK PAHALI: ALMAYAN {ALMAYAN=4, SATIN ALAN=1}
 | | | Otel_Ücreti = ÇOK UCUZ: SATIN ALAN {ALMAYAN=0, SATIN ALAN=2}
 | | Yetişkin Yaşı = 60++
 | | Çocuk Bilgisi = 1-1
 | | | Otel_Ücreti = ORTA: ALMAYAN {ALMAYAN=1, SATIN ALAN=1}
 | | | Otel_Ücreti = PAHALI: ALMAYAN {ALMAYAN=5, SATIN ALAN=0}
 | | | Otel_Ücreti = ÇOK PAHALI: ALMAYAN {ALMAYAN=1, SATIN ALAN=1}
 | | Çocuk Bilgisi = 1-2
 | | | Otel_Ücreti = ORTA: ALMAYAN {ALMAYAN=2, SATIN ALAN=1}
 | | | Otel_Ücreti = PAHALI: ALMAYAN {ALMAYAN=6, SATIN ALAN=4}
 | | | Otel_Ücreti = ÇOK PAHALI: ALMAYAN {ALMAYAN=6, SATIN ALAN=0}
 | | Çocuk Bilgisi = 1-3
 | | | Otel_Ücreti = PAHALI: ALMAYAN {ALMAYAN=7, SATIN ALAN=0}
 | | | Otel_Ücreti = ÇOK PAHALI: ALMAYAN {ALMAYAN=2, SATIN ALAN=2}
 | | Çocuk Bilgisi = 2-02: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | Çocuk Bilgisi = 2-11: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | Çocuk Bilgisi = 2-12
 | | | Otel_Ücreti = PAHALI: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}

| | | Otel_Ücreti = ÇOK PAHALI: ALMAYAN {ALMAYAN=1, SATIN ALAN=1}
 | | Çocuk Bilgisi = 2-22: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | Çocuk Bilgisi = 2-23
 | | | Konaklama Süresi = 14-20: ALMAYAN {ALMAYAN=1, SATIN ALAN=1}
 | | | Konaklama Süresi = 7-13: ALMAYAN {ALMAYAN=3, SATIN ALAN=0}
 | | Çocuk Bilgisi = 2-33: SATIN ALAN {ALMAYAN=0, SATIN ALAN=1}
 | | Çocuk Bilgisi = 3++: SATIN ALAN {ALMAYAN=0, SATIN ALAN=4}
 | | Çocuk Bilgisi = YOK
 | | | Otel_Ücreti = ORTA: ALMAYAN {ALMAYAN=41, SATIN ALAN=20}
 | | | Otel_Ücreti = PAHALI: ALMAYAN {ALMAYAN=127, SATIN ALAN=31}
 | | | Otel_Ücreti = UCUZ: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Otel_Ücreti = ÇOK PAHALI: ALMAYAN {ALMAYAN=26, SATIN ALAN=10}

Bölge = KMR

| Yetiskin Yaşı = 18-30
 | | Müşteri Tipi = ARKDS
 | | | Otel_Ücreti = ORTA: SATIN ALAN {ALMAYAN=0, SATIN ALAN=33}
 | | | Otel_Ücreti = PAHALI: SATIN ALAN {ALMAYAN=1, SATIN ALAN=7}
 | | | Otel_Ücreti = UCUZ: SATIN ALAN {ALMAYAN=0, SATIN ALAN=3}
 | | Müşteri Tipi = AİLE
 | | | Çocuk Bilgisi = 1-1: SATIN ALAN {ALMAYAN=0, SATIN ALAN=18}
 | | | Çocuk Bilgisi = 1-2: SATIN ALAN {ALMAYAN=1, SATIN ALAN=35}
 | | | Çocuk Bilgisi = 1-3: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-01: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-11: SATIN ALAN {ALMAYAN=0, SATIN ALAN=3}
 | | | Çocuk Bilgisi = 2-12: SATIN ALAN {ALMAYAN=0, SATIN ALAN=5}
 | | | Çocuk Bilgisi = 2-22: SATIN ALAN {ALMAYAN=0, SATIN ALAN=5}
 | | | Çocuk Bilgisi = 3++: SATIN ALAN {ALMAYAN=0, SATIN ALAN=1}
 | | | Çocuk Bilgisi = YOK: SATIN ALAN {ALMAYAN=0, SATIN ALAN=3}
 | | Müşteri Tipi = GRUP
 | | | Otel_Ücreti = ORTA: SATIN ALAN {ALMAYAN=0, SATIN ALAN=20}
 | | | Otel_Ücreti = PAHALI: SATIN ALAN {ALMAYAN=1, SATIN ALAN=19}
 | | | Otel_Ücreti = UCUZ: SATIN ALAN {ALMAYAN=0, SATIN ALAN=6}
 | | | Otel_Ücreti = ÇOK PAHALI: SATIN ALAN {ALMAYAN=0, SATIN ALAN=5}
 | | | Otel_Ücreti = ÇOK UCUZ: SATIN ALAN {ALMAYAN=0, SATIN ALAN=2}
 | | Müşteri Tipi = TEK
 | | | Çocuk Bilgisi = 1-1: ALMAYAN {ALMAYAN=4, SATIN ALAN=1}
 | | | Çocuk Bilgisi = 1-2: ALMAYAN {ALMAYAN=6, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-12: SATIN ALAN {ALMAYAN=0, SATIN ALAN=1}
 | | | Çocuk Bilgisi = 2-22: SATIN ALAN {ALMAYAN=0, SATIN ALAN=1}
 | | | Çocuk Bilgisi = YOK: SATIN ALAN {ALMAYAN=0, SATIN ALAN=23}
 | | Müşteri Tipi = ÇCK-B: ALMAYAN {ALMAYAN=12, SATIN ALAN=0}
 | | Müşteri Tipi = ÇİFT
 | | | Konaklama Süresi = --7: SATIN ALAN {ALMAYAN=0, SATIN ALAN=3}
 | | | Konaklama Süresi = 14-20: SATIN ALAN {ALMAYAN=0, SATIN ALAN=148}
 | | | Konaklama Süresi = 21++: SATIN ALAN {ALMAYAN=0, SATIN ALAN=7}
 | | | Konaklama Süresi = 7-13: SATIN ALAN {ALMAYAN=9, SATIN ALAN=192}
 | Yetiskin Yaşı = 31-59
 | | Müşteri Tipi = ARKDS
 | | | Otel_Ücreti = ORTA: ALMAYAN {ALMAYAN=5, SATIN ALAN=3}
 | | | Otel_Ücreti = PAHALI: ALMAYAN {ALMAYAN=3, SATIN ALAN=0}
 | | Müşteri Tipi = AİLE

| | | Çocuk Bilgisi = 1-0: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 1-1: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 1-2: SATIN ALAN {ALMAYAN=20, SATIN ALAN=30}
 | | | Çocuk Bilgisi = 1-3: SATIN ALAN {ALMAYAN=10, SATIN ALAN=47}
 | | | Çocuk Bilgisi = 2-01: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-02: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-12: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-13: SATIN ALAN {ALMAYAN=0, SATIN ALAN=1}
 | | | Çocuk Bilgisi = 2-22: SATIN ALAN {ALMAYAN=4, SATIN ALAN=6}
 | | | Çocuk Bilgisi = 2-23: SATIN ALAN {ALMAYAN=8, SATIN ALAN=14}
 | | | Çocuk Bilgisi = 2-33: SATIN ALAN {ALMAYAN=0, SATIN ALAN=3}
 | | | Çocuk Bilgisi = 3++: ALMAYAN {ALMAYAN=3, SATIN ALAN=2}
 | | Müşteri Tipi = GRUP: ALMAYAN {ALMAYAN=8, SATIN ALAN=0}
 | | Müşteri Tipi = TEK
 | | | Çocuk Bilgisi = 1-2: ALMAYAN {ALMAYAN=31, SATIN ALAN=7}
 | | | Çocuk Bilgisi = 1-3: ALMAYAN {ALMAYAN=12, SATIN ALAN=4}
 | | | Çocuk Bilgisi = 2-12: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-22: ALMAYAN {ALMAYAN=7, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-23: SATIN ALAN {ALMAYAN=0, SATIN ALAN=2}
 | | | Çocuk Bilgisi = 2-33: SATIN ALAN {ALMAYAN=0, SATIN ALAN=2}
 | | | Çocuk Bilgisi = YOK: ALMAYAN {ALMAYAN=2, SATIN ALAN=2}
 | | Müşteri Tipi = ÇCK-B
 | | | Çocuk Bilgisi = 1-2: ALMAYAN {ALMAYAN=10, SATIN ALAN=2}
 | | | Çocuk Bilgisi = 1-3: ALMAYAN {ALMAYAN=14, SATIN ALAN=2}
 | | | Çocuk Bilgisi = 2-11: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-12: ALMAYAN {ALMAYAN=3, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-22: ALMAYAN {ALMAYAN=3, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-23: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-33: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 3++: ALMAYAN {ALMAYAN=3, SATIN ALAN=1}
 | | Müşteri Tipi = ÇİFT
 | | | Otel_Ücreti = ORTA: ALMAYAN {ALMAYAN=42, SATIN ALAN=7}
 | | | Otel_Ücreti = PAHALI: ALMAYAN {ALMAYAN=14, SATIN ALAN=1}
 | | | Otel_Ücreti = UCUZ: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | Yetişkin Yaşı = 60++
 | | Konaklama Süresi = --7: ALMAYAN {ALMAYAN=4, SATIN ALAN=0}
 | | Konaklama Süresi = 14-20
 | | | Çocuk Bilgisi = 1-2: SATIN ALAN {ALMAYAN=1, SATIN ALAN=2}
 | | | Çocuk Bilgisi = 1-3: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-12: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-22: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | | Çocuk Bilgisi = YOK: ALMAYAN {ALMAYAN=162, SATIN ALAN=47}
 | | Konaklama Süresi = 21++
 | | | Müşteri Tipi = ARKDS: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Müşteri Tipi = AİLE: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Müşteri Tipi = TEK: ALMAYAN {ALMAYAN=3, SATIN ALAN=1}
 | | | Müşteri Tipi = ÇİFT: ALMAYAN {ALMAYAN=25, SATIN ALAN=0}
 | | Konaklama Süresi = 7-13
 | | | Müşteri Tipi = ARKDS: ALMAYAN {ALMAYAN=4, SATIN ALAN=3}
 | | | Müşteri Tipi = AİLE: ALMAYAN {ALMAYAN=11, SATIN ALAN=1}
 | | | Müşteri Tipi = GRUP: ALMAYAN {ALMAYAN=6, SATIN ALAN=2}

| | | Müşteri Tipi = TEK: ALMAYAN {ALMAYAN=6, SATIN ALAN=0}
 | | | Müşteri Tipi = ÇCK-B: ALMAYAN {ALMAYAN=2, SATIN ALAN=1}
 | | | Müşteri Tipi = ÇİFT: ALMAYAN {ALMAYAN=55, SATIN ALAN=15}

Bölge = MNV

| Yetiskin Yaşı = 18-30
 | | Çocuk Bilgisi = 1-0: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | Çocuk Bilgisi = 1-1: ALMAYAN {ALMAYAN=18, SATIN ALAN=0}
 | | Çocuk Bilgisi = 1-2
 | | | Otel_Ücreti = ORTA: ALMAYAN {ALMAYAN=8, SATIN ALAN=0}
 | | | Otel_Ücreti = PAHALI: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | | Otel_Ücreti = UCUZ: SATIN ALAN {ALMAYAN=0, SATIN ALAN=1}
 | | | Otel_Ücreti = ÇOK PAHALI: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | Çocuk Bilgisi = 2-01: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | Çocuk Bilgisi = 2-02: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | Çocuk Bilgisi = 2-11: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | Çocuk Bilgisi = 2-12: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | Çocuk Bilgisi = YOK
 | | | Konaklama Süresi = 14-20: SATIN ALAN {ALMAYAN=0, SATIN ALAN=5}
 | | | Konaklama Süresi = 7-13: ALMAYAN {ALMAYAN=6, SATIN ALAN=2}

| Yetiskin Yaşı = 31-59

| | Konaklama Süresi = --7: ALMAYAN {ALMAYAN=4, SATIN ALAN=0}
 | | Konaklama Süresi = 14-20
 | | | Çocuk Bilgisi = 1-1: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 1-2: ALMAYAN {ALMAYAN=12, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 1-3: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-01: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-12: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-22: ALMAYAN {ALMAYAN=3, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 2-23: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | | Çocuk Bilgisi = 3++: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | | Çocuk Bilgisi = YOK: ALMAYAN {ALMAYAN=11, SATIN ALAN=1}
 | | | Konaklama Süresi = 21++: ALMAYAN {ALMAYAN=3, SATIN ALAN=0}
 | | | Konaklama Süresi = 7-13: ALMAYAN {ALMAYAN=167, SATIN ALAN=0}
 | | Yetiskin Yaşı = 60++: ALMAYAN {ALMAYAN=195, SATIN ALAN=0}

Bölge = SDE

| Yetiskin Yaşı = 18-30
 | | Konaklama Süresi = --7: ALMAYAN {ALMAYAN=8, SATIN ALAN=0}
 | | Konaklama Süresi = 14-20
 | | | Otel_Ücreti = ORTA: ALMAYAN {ALMAYAN=22, SATIN ALAN=19}
 | | | Otel_Ücreti = PAHALI: SATIN ALAN {ALMAYAN=7, SATIN ALAN=11}
 | | | Otel_Ücreti = UCUZ: SATIN ALAN {ALMAYAN=5, SATIN ALAN=7}
 | | | Otel_Ücreti = ÇOK PAHALI: ALMAYAN {ALMAYAN=3, SATIN ALAN=2}
 | | | Otel_Ücreti = ÇOK UCUZ: SATIN ALAN {ALMAYAN=0, SATIN ALAN=8}
 | | Konaklama Süresi = 21++
 | | | Otel_Ücreti = ORTA: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | | Otel_Ücreti = ÇOK PAHALI: SATIN ALAN {ALMAYAN=0, SATIN ALAN=1}
 | | | Otel_Ücreti = ÇOK UCUZ: ALMAYAN {ALMAYAN=1, SATIN ALAN=1}
 | | Konaklama Süresi = 7-13
 | | | Müşteri Tipi = ARKDS: ALMAYAN {ALMAYAN=30, SATIN ALAN=0}
 | | | Müşteri Tipi = AİLE: ALMAYAN {ALMAYAN=15, SATIN ALAN=0}
 | | | Müşteri Tipi = GRUP: ALMAYAN {ALMAYAN=5, SATIN ALAN=1}

| | | Müşteri Tipi = TEK: ALMAYAN {ALMAYAN=6, SATIN ALAN=2}
 | | | Müşteri Tipi = ÇCK-B: ALMAYAN {ALMAYAN=1, SATIN ALAN=1}
 | | | Müşteri Tipi = ÇİFT: ALMAYAN {ALMAYAN=65, SATIN ALAN=23}
 | Yetiskin Yaşı = 31-59
 | | Çocuk Bilgisi = 1-0: ALMAYAN {ALMAYAN=8, SATIN ALAN=0}
 | | Çocuk Bilgisi = 1-1: ALMAYAN {ALMAYAN=43, SATIN ALAN=0}
 | | Çocuk Bilgisi = 1-2
 | | | Konaklama Süresi = --7: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | | Konaklama Süresi = 14-20: ALMAYAN {ALMAYAN=95, SATIN ALAN=3}
 | | | Konaklama Süresi = 21++: ALMAYAN {ALMAYAN=1, SATIN ALAN=1}
 | | | Konaklama Süresi = 7-13: ALMAYAN {ALMAYAN=154, SATIN ALAN=0}
 | | Çocuk Bilgisi = 1-3
 | | | Konaklama Süresi = --7: ALMAYAN {ALMAYAN=5, SATIN ALAN=0}
 | | | Konaklama Süresi = 14-20: ALMAYAN {ALMAYAN=113, SATIN ALAN=4}
 | | | Konaklama Süresi = 21++: SATIN ALAN {ALMAYAN=0, SATIN ALAN=1}
 | | | Konaklama Süresi = 7-13: ALMAYAN {ALMAYAN=166, SATIN ALAN=1}
 | | Çocuk Bilgisi = 2-00: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | Çocuk Bilgisi = 2-01: ALMAYAN {ALMAYAN=9, SATIN ALAN=0}
 | | Çocuk Bilgisi = 2-02: ALMAYAN {ALMAYAN=7, SATIN ALAN=0}
 | | Çocuk Bilgisi = 2-03: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
 | | Çocuk Bilgisi = 2-11: ALMAYAN {ALMAYAN=8, SATIN ALAN=0}
 | | Çocuk Bilgisi = 2-12: ALMAYAN {ALMAYAN=20, SATIN ALAN=0}
 | | Çocuk Bilgisi = 2-22: ALMAYAN {ALMAYAN=46, SATIN ALAN=0}
 | | Çocuk Bilgisi = 2-23
 | | | Otel_Ücreti = ORTA: ALMAYAN {ALMAYAN=19, SATIN ALAN=0}
 | | | Otel_Ücreti = PAHALI: ALMAYAN {ALMAYAN=16, SATIN ALAN=1}
 | | | Otel_Ücreti = UCUZ: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | | Otel_Ücreti = ÇOK PAHALI: ALMAYAN {ALMAYAN=15, SATIN ALAN=0}
 | | | Otel_Ücreti = ÇOK UCUZ: ALMAYAN {ALMAYAN=6, SATIN ALAN=0}
 | | Çocuk Bilgisi = 2-33: ALMAYAN {ALMAYAN=15, SATIN ALAN=0}
 | | Çocuk Bilgisi = 3++: ALMAYAN {ALMAYAN=28, SATIN ALAN=0}
 | | Çocuk Bilgisi = YOK
 | | | Otel_Ücreti = ORTA: ALMAYAN {ALMAYAN=388, SATIN ALAN=2}
 | | | Otel_Ücreti = PAHALI: ALMAYAN {ALMAYAN=183, SATIN ALAN=0}
 | | | Otel_Ücreti = UCUZ: ALMAYAN {ALMAYAN=69, SATIN ALAN=0}
 | | | Otel_Ücreti = ÇOK PAHALI: ALMAYAN {ALMAYAN=11, SATIN ALAN=0}
 | | | Otel_Ücreti = ÇOK UCUZ: ALMAYAN {ALMAYAN=11, SATIN ALAN=0}
 | Yetiskin Yaşı = 60++
 | | Otel_Ücreti = ORTA
 | | | Müşteri Tipi = ARKDS: ALMAYAN {ALMAYAN=87, SATIN ALAN=2}
 | | | Müşteri Tipi = AİLE: ALMAYAN {ALMAYAN=68, SATIN ALAN=0}
 | | | Müşteri Tipi = GRUP: ALMAYAN {ALMAYAN=19, SATIN ALAN=0}
 | | | Müşteri Tipi = TEK: ALMAYAN {ALMAYAN=49, SATIN ALAN=0}
 | | | Müşteri Tipi = ÇCK-B: ALMAYAN {ALMAYAN=14, SATIN ALAN=0}
 | | | Müşteri Tipi = ÇİFT: ALMAYAN {ALMAYAN=803, SATIN ALAN=1}
 | | Otel_Ücreti = PAHALI: ALMAYAN {ALMAYAN=726, SATIN ALAN=0}
 | | Otel_Ücreti = UCUZ: ALMAYAN {ALMAYAN=233, SATIN ALAN=0}
 | | Otel_Ücreti = ÇOK PAHALI: ALMAYAN {ALMAYAN=52, SATIN ALAN=0}
 | | Otel_Ücreti = ÇOK UCUZ
 | | | Müşteri Tipi = ARKDS: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | | Müşteri Tipi = AİLE: ALMAYAN {ALMAYAN=6, SATIN ALAN=0}

| | | Müşteri Tipi = TEK: ALMAYAN {ALMAYAN=4, SATIN ALAN=0}
| | | Müşteri Tipi = ÇCK-B: ALMAYAN {ALMAYAN=1, SATIN ALAN=0}
| | | Müşteri Tipi = ÇİFT: ALMAYAN {ALMAYAN=24, SATIN ALAN=1}



Ek-2. Budanmış Karar Ağacı Kuralları

Bölge = ALY

- | Yetiskin Yaşı = 18-30
 - | | Konaklama Süresi = --7: ALMAYAN {ALMAYAN=8, SATIN ALAN=0}
 - | | Konaklama Süresi = 14-20
 - | | | Müşteri Tipi = ARKDS: SATIN ALAN {ALMAYAN=1, SATIN ALAN=2}
 - | | | Müşteri Tipi = AİLE: ALMAYAN {ALMAYAN=2, SATIN ALAN=1}
 - | | | Müşteri Tipi = GRUP: ALMAYAN {ALMAYAN=3, SATIN ALAN=1}
 - | | | Müşteri Tipi = TEK: ALMAYAN {ALMAYAN=3, SATIN ALAN=0}
 - | | | Müşteri Tipi = ÇİFT: SATIN ALAN {ALMAYAN=3, SATIN ALAN=10}
 - | | Konaklama Süresi = 7-13: ALMAYAN {ALMAYAN=91, SATIN ALAN=9}
- | Yetiskin Yaşı = 31-59: ALMAYAN {ALMAYAN=526, SATIN ALAN=2}
- | Yetiskin Yaşı = 60++: ALMAYAN {ALMAYAN=455, SATIN ALAN=2}

Bölge = AYT

- | Yetiskin Yaşı = 18-30
 - | | Müşteri Tipi = ARKDS: SATIN ALAN {ALMAYAN=1, SATIN ALAN=45}
 - | | Müşteri Tipi = AİLE: SATIN ALAN {ALMAYAN=7, SATIN ALAN=78}
 - | | Müşteri Tipi = GRUP: SATIN ALAN {ALMAYAN=1, SATIN ALAN=43}
 - | | Müşteri Tipi = TEK
 - | | | Çocuk Bilgisi = 1-1: ALMAYAN {ALMAYAN=5, SATIN ALAN=1}
 - | | | Çocuk Bilgisi = 1-2: ALMAYAN {ALMAYAN=9, SATIN ALAN=1}
 - | | | Çocuk Bilgisi = YOK: SATIN ALAN {ALMAYAN=0, SATIN ALAN=11}
 - | | Müşteri Tipi = ÇCK-B: ALMAYAN {ALMAYAN=20, SATIN ALAN=2}
 - | | Müşteri Tipi = ÇİFT: SATIN ALAN {ALMAYAN=9, SATIN ALAN=282}
- | Yetiskin Yaşı = 31-59
 - | | Müşteri Tipi = ARKDS: ALMAYAN {ALMAYAN=15, SATIN ALAN=1}
 - | | Müşteri Tipi = AİLE: SATIN ALAN {ALMAYAN=85, SATIN ALAN=86}
 - | | Müşteri Tipi = GRUP: ALMAYAN {ALMAYAN=8, SATIN ALAN=2}
 - | | Müşteri Tipi = TEK: ALMAYAN {ALMAYAN=92, SATIN ALAN=11}
 - | | Müşteri Tipi = ÇCK-B: ALMAYAN {ALMAYAN=52, SATIN ALAN=6}
 - | | Müşteri Tipi = ÇİFT: ALMAYAN {ALMAYAN=56, SATIN ALAN=7}
- | Yetiskin Yaşı = 60++: ALMAYAN {ALMAYAN=311, SATIN ALAN=65}

Bölge = BLK

- | Yetiskin Yaşı = 18-30
 - | | Müşteri Tipi = ARKDS: SATIN ALAN {ALMAYAN=2, SATIN ALAN=86}
 - | | Müşteri Tipi = AİLE: SATIN ALAN {ALMAYAN=9, SATIN ALAN=196}
 - | | Müşteri Tipi = GRUP: SATIN ALAN {ALMAYAN=0, SATIN ALAN=101}
 - | | Müşteri Tipi = TEK: SATIN ALAN {ALMAYAN=25, SATIN ALAN=28}
 - | | Müşteri Tipi = ÇCK-B: ALMAYAN {ALMAYAN=39, SATIN ALAN=9}
 - | | Müşteri Tipi = ÇİFT: SATIN ALAN {ALMAYAN=8, SATIN ALAN=473}
- | Yetiskin Yaşı = 31-59
 - | | Müşteri Tipi = ARKDS: ALMAYAN {ALMAYAN=7, SATIN ALAN=2}
 - | | Müşteri Tipi = AİLE: SATIN ALAN {ALMAYAN=87, SATIN ALAN=363}
 - | | Müşteri Tipi = GRUP: ALMAYAN {ALMAYAN=1, SATIN ALAN=1}
 - | | Müşteri Tipi = TEK: ALMAYAN {ALMAYAN=108, SATIN ALAN=26}
 - | | Müşteri Tipi = ÇCK-B: ALMAYAN {ALMAYAN=46, SATIN ALAN=23}
 - | | Müşteri Tipi = ÇİFT
 - | | | Otel_Ücreti = ORTA: ALMAYAN {ALMAYAN=5, SATIN ALAN=2}
 - | | | Otel_Ücreti = PAHALI: ALMAYAN {ALMAYAN=40, SATIN ALAN=5}
 - | | | Otel_Ücreti = UCUZ: SATIN ALAN {ALMAYAN=0, SATIN ALAN=2}
- | Yetiskin Yaşı = 60++: ALMAYAN {ALMAYAN=235, SATIN ALAN=77}

Bölge = KMR

| Yetiskin Yaşı = 18-30
 | | Müşteri Tipi = ARKDS: SATIN ALAN {ALMAYAN=1, SATIN ALAN=43}
 | | Müşteri Tipi = AİLE: SATIN ALAN {ALMAYAN=3, SATIN ALAN=70}
 | | Müşteri Tipi = GRUP: SATIN ALAN {ALMAYAN=3, SATIN ALAN=79}
 | | Müşteri Tipi = TEK
 | | | Konaklama Süresi = --7: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | | Konaklama Süresi = 14-20
 | | | | Otel_Ücreti = ORTA: SATIN ALAN {ALMAYAN=0, SATIN ALAN=3}
 | | | | Otel_Ücreti = UCUZ
 | | | | Çocuk Bilgisi = 1-1: ALMAYAN {ALMAYAN=2, SATIN ALAN=0}
 | | | | Çocuk Bilgisi = YOK: SATIN ALAN {ALMAYAN=0, SATIN ALAN=8}
 | | | Konaklama Süresi = 7-13
 | | | | Otel_Ücreti = ORTA: SATIN ALAN {ALMAYAN=1, SATIN ALAN=8}
 | | | | Otel_Ücreti = PAHALI: ALMAYAN {ALMAYAN=1, SATIN ALAN=1}
 | | | | Otel_Ücreti = UCUZ
 | | | | Çocuk Bilgisi = 1-2: ALMAYAN {ALMAYAN=4, SATIN ALAN=0}
 | | | | Çocuk Bilgisi = YOK: SATIN ALAN {ALMAYAN=0, SATIN ALAN=6}
 | | Müşteri Tipi = ÇCK-B: ALMAYAN {ALMAYAN=12, SATIN ALAN=0}
 | | Müşteri Tipi = ÇİFT: SATIN ALAN {ALMAYAN=7, SATIN ALAN=323}
 | Yetiskin Yaşı = 31-59
 | | Müşteri Tipi = ARKDS: ALMAYAN {ALMAYAN=8, SATIN ALAN=3}
 | | Müşteri Tipi = AİLE: SATIN ALAN {ALMAYAN=52, SATIN ALAN=103}
 | | Müşteri Tipi = GRUP: ALMAYAN {ALMAYAN=14, SATIN ALAN=0}
 | | Müşteri Tipi = TEK: ALMAYAN {ALMAYAN=53, SATIN ALAN=17}
 | | Müşteri Tipi = ÇCK-B: ALMAYAN {ALMAYAN=36, SATIN ALAN=5}
 | | Müşteri Tipi = ÇİFT: ALMAYAN {ALMAYAN=52, SATIN ALAN=8}
 | Yetiskin Yaşı = 60++: ALMAYAN {ALMAYAN=287, SATIN ALAN=72}
 Bölge = MNV: ALMAYAN {ALMAYAN=447, SATIN ALAN=9}

Bölge = SDE

| Yetiskin Yaşı = 18-30
 | | Konaklama Süresi = --7: ALMAYAN {ALMAYAN=8, SATIN ALAN=0}
 | | Konaklama Süresi = 14-20: SATIN ALAN {ALMAYAN=37, SATIN ALAN=47}
 | | Konaklama Süresi = 21++: ALMAYAN {ALMAYAN=2, SATIN ALAN=2}
 | | Konaklama Süresi = 7-13: ALMAYAN {ALMAYAN=122, SATIN ALAN=27}
 | Yetiskin Yaşı = 31-59: ALMAYAN {ALMAYAN=1442, SATIN ALAN=13}
 | Yetiskin Yaşı = 60++: ALMAYAN {ALMAYAN=2088, SATIN ALAN=4}

ÖZGEÇMİŞ

Adı ve SOYADI	Abdullah AĞÜN
Doğum Yeri - Tarihi	Korkuteli - 06.03.1973
EĞİTİM DURUMU	
Mezun Olduğu Lise	Antalya Endüstri Meslek Lisesi
Lisans Diploması	Marmara Üniversitesi, Teknik Eğitim Fakültesi
Yüksek Lisans Diploması	Akdeniz Üniversitesi, Sosyal Bilimler Enstitüsü, Turizm İşletmeciliği ve Otelcilik Anabilim Dalı
Tez/ Dönem Projesi Konusu	SEYAHAT ACENTALARINDA VERİ MADENCİLİĞİ: ANTALYA BÖLGESİNDE BİR UYGULAMA
Yabancı Dil / Diller	İngilizce
BİLİMSEL FAALİYETLER	
<p><u>Makaleler</u> Ehtiyar V.R., Ersoy A., Akgün A., Karapınar E., "Üniversite Öğrencilerinin Geleceğe Yönelik Tutum ve Olumlu Gelecek Beklentilerinin Psikolojik İyilik Halleri Üzerindeki Etkisi", Akdeniz İnsani Bilimler Dergisi, Vol.VII, pp.251-262, 2017 Akgün A., Çizel B., "Günlük Tur Programları Oluşturmada Veri Madenciliği: A Grubu Seyahat Acentası Örneği", Turizm Araştırma Dergisi (TURAR), cilt.6, ss.1-1, 2017 Çetin H., Akgün A., "Yeşil Bilişim Teknolojileri Bağlamında Sanallaştırılmış ve Klasik Sistemlerin Karşılaştırılması", Uluslararası Alanya İşletme Fakültesi Dergisi, cilt.7, ss.131-142, 2015</p> <p><u>Bildiriler</u> Çizel B., Akgün A., "Turizm Öğrencilerinin Mesleklerinde İletişim ve Enformasyon Teknolojilerini Kabul ve Kullanım Niyetini Etkileyen Faktörlerin Bir Model içerisinde İncelenmesi", 13. Ulusal Turizm Kongresi, ANTALYA, TÜRKİYE, 6 Aralık 2012 - 9 Aralık 12, ss.343-352 Kinay H., Akgün A., Çetin H., Yılmaz Y., Akinci Z., "The impacts of using additional teaching materials on students' performance in package program education:the case of fidelio and sejour", ISSD 2012, Sarajevo, BOSNA HERSEK, 31 Mayıs - 1 Haziran 2012, pp.137-137</p>	
İŞ DENEYİMİ	
Çalıştığı Kurumlar	Atatürk Anadolu Teknik ve Endüstri Meslek Lisesi, İzmir Akdeniz Koleji, Antalya Akdeniz Üniversitesi, Turizm Fakültesi
E-Posta	akgun@akdeniz.edu.tr