

T.C.  
SAKARYA ÜNİVERSİTESİ  
SOSYAL BİLİMLER ENSTİTÜSÜ

**KAMU HİZMETLERİNDE VERİ MADENCİLİĞİ:  
ÇÖZÜM MASASI VERİLERİ TEMELİNDE  
BİR ARAŞTIRMA**

**DOKTORA TEZİ**

**Yılmaz DEMİRCİ**

**Enstitü Anabilim Dalı : Siyaset Bilimi ve Kamu Yönetimi**

**Tez Danışmanı: Doç. Dr. Özer KÖSEOĞLU**

**EYLÜL-2018**






**KAMU HİZMETLERİNDE VERİ MADENCİLİĞİ:  
ÇÖZÜM MASASI VERİLERİ TEMELİNDE  
BİR ARAŞTIRMA**

**DOKTORA TEZİ**

**Yılmaz DEMİRCİ**

**Enstitü Anabilim Dalı : Siyaset Bilimi ve Kamu Yönetimi**

“Bu tez **21.09/2018** tarihinde aşağıdaki jüri tarafından Oybirliği / ~~Oyçokluğu~~ ile kabul edilmiştir.”

JÜRİ ÜYESİ	KANAATI	İMZA
Prof. Dr. Musa EKEN	Basarılı	
Doç. Dr. Özer KÖRKOĞLU	Basarılı	
Doç. Dr. M. Fatih Bilal AKODAY	Basarılı	
Doç. Dr. Sefer USTA	Basarılı	
Dr.Öpr.Üyesi Adem AKBIYIK	Basarılı	



SAKARYA  
ÜNİVERSİTESİ

T.C.  
SAKARYA ÜNİVERSİTESİ  
SOSYAL BİLİMLER ENSTİTÜSÜ  
TEZ SAVUNULABİLİRLİK VE ORJİNALLİK BEYAN FORMU

Sayfa : 1/1

Öğrencinin

Adı Soyadı	:	YILMAZ DEMİRCİ
Öğrenci Numarası	:	1260D03010
Enstitü Anabilim Dalı	:	SİYASET BİLİMİ VE KAMU YÖNETİMİ
Enstitü Bilim Dalı	:	SİYASET BİLİMİ VE KAMU YÖNETİMİ
Programı	:	<input type="checkbox"/> YÜKSEK LİSANS <input checked="" type="checkbox"/> DOKTORA
Tezin Başlığı	:	KAMU HİZMETLERİNDE VERİ MADENCİLİĞİ: ÇÖZÜM MASASI VERİLERİ TEMELİNDE BİR ARAŞTIRMA
Benzerlik Oranı	:	%12

SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜNE,

Sakarya Üniversitesi Sosyal Bilimler Enstitüsü Enstitüsü Lisansüstü Tez Çalışması Benzerlik Raporu Uygulama Esaslarını inceledim. Enstitünüz tarafından Uygulama Esasları çerçevesinde alınan Benzerlik Raporuna göre yukarıda bilgileri verilen tez çalışmasının benzerlik oranının herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi beyan ederim.

21/09/2018  
Öğrenci İmza

Sakarya Üniversitesi Sosyal Bilimler Enstitüsü Lisansüstü Tez Çalışması Benzerlik Raporu Uygulama Esaslarını inceledim. Enstitünüz tarafından Uygulama Esasları çerçevesinde alınan Benzerlik Raporuna göre yukarıda bilgileri verilen öğrenciye ait tez çalışması ile ilgili gerekli düzenleme tarafımda yapılmış olup, yeniden değerlendirilmek üzere sbetetzler@sakarya.edu.tr adresine yüklenmiştir.

Bilgilerinize arz ederim.

...../...../20.....  
Öğrenci İmza

Uygundur

Danışman  
Unvanı / Adı-Soyadı: Doç.Dr. Özer KÖSEOĞLU

Tarih: 21.09.2018

İmza:

KABUL EDİLMİŞTİR

REDDEDİLMİŞTİR

EYK Tarih ve No:

Enstitü Birim Sorumlusu Onayı

## ÖNSÖZ

Kamu kurumlarında yönetim bilişim sistemlerinin yoğun biçimde kullanılması, veri işleme kapasitesindeki gelişmeler ve vatandaşın değişen talepleri neticesinde sunulan hizmetlerin çeşitlenmesi kamuda büyük verinin oluşmasında etkili olmaktadır. Kamu kurumlarının geleceğe dönük kararlarını doğru ve hızlı biçimde alması, hizmetlerini etkin, hesap verebilir ve şeffaf biçimde gerçekleştirebilmesi için sahip olduğu büyük veriden yararlanması, veri madenciliği uygulamalarını gerçekleştirmesi önemli hale gelmiştir.

Bu çalışmada kamu yönetiminde büyük verinin önemi, veri madenciliğine ilişkin kavramsal çerçeve, Türkiye’de kamu hizmetlerinde büyük verinin durumu ve veri madenciliğinin kullanımı araştırılacaktır. Bu amaçla, büyükşehir belediyeleri ölçeğinde “veri madenciliği” anketi düzenlenecek, Kocaeli Büyükşehir Belediyesi’ne ait veri örneği üzerinden veri madenciliği süreçleri izlenerek uygulama gerçekleştirilecektir.

Çalışmalarım süresince destek, yönlendirme ve yardımlarını esirgemeyen, yüksek lisans ve doktora eğitimim de dâhil olmak üzere her zaman beni motive eden danışman hocam Doç. Dr. Özer KÖSEOĞLU’na saygı ve teşekkürlerimi sunarım. Böyle bir çalışmaya yönelmemi sağlayan ve desteğini esirgemeyen Yönetim Bilişim Sistemleri konusunda akademik dünyada saygın bir yeri olan Prof. Dr. Erman COŞKUN’a müteşekkirim. Tez İzleme Komitesi’nde yer alarak verdiği katkılardan dolayı Doç. Dr. Aziz TUNCER’e teşekkür ederim. Çalışmamın uygulama aşamasında gerekli veriyi sağlayan Kocaeli Büyükşehir Belediyesi’ne, çalışma arkadaşlarım Dr. Emrah ERGÜL ve Mehmet KARAYEL’e, büyük veri ve veri madenciliği konusundaki katkıları için Doç. Dr. Şadi Evren ŞEKER’e, metin madenciliği ve doğal dil işleme konusundaki katkısı için Doç. Dr. M. Fatih AMASYALI’ya sağladıkları katkılardan dolayı minnettarım.

Son olarak, bana her konuda destek olan aileme, hayatıma girdiği andan bu yana beni her anlamda destekleyen sevgili eşim Ebru’ya, mutluluk kaynağımız olan çocuklarımız Mert ve Yiğit’e hayatımdaki bu enerjiyi verdiklerinden dolayı teşekkür ederim.

**Yılmaz DEMİRCİ**

**21.09.2018**

# İÇİNDEKİLER

<b>KISALTMALAR</b> .....	<b>v</b>
<b>TABLO LİSTESİ</b> .....	<b>vi</b>
<b>ŞEKİL LİSTESİ</b> .....	<b>vii</b>
<b>ÖZET</b> .....	<b>x</b>
<b>SUMMARY</b> .....	<b>xi</b>
<b>GİRİŞ</b> .....	<b>1</b>
<b>BÖLÜM 1: BÜYÜK VERİ VE VERİ MADENCİLİĞİ</b> .....	<b>9</b>
1.1. Büyük Veri: Tanımı, Gelişimi ve İşlevi .....	10
1.2. Büyük Veri ile İlişkili Başlıca Bilişim Teknolojileri .....	16
1.2.1. Endüstri 4.0 .....	17
1.2.2. Bulut Bilişim.....	21
1.2.3. Nesnelerin İnterneti.....	26
1.2.4. Yapay Zekâ .....	30
1.3. Veri Madenciliği .....	31
1.3.1. Veri Madenciliğiyle İlişkili Kavram ve Uygulamalar .....	32
1.3.1.1. Veri, Bilgi ve Enformasyon.....	35
1.3.1.2. Veritabanı Yönetim Sistemleri .....	36
1.3.1.3. Veri Ambarları.....	39
1.3.2. Veri Madenciliğinin Tanımlanması .....	41
1.3.3. Veri Madenciliği Türleri .....	43
1.3.3.1. Veri Madenciliği Uygulamaları.....	44
1.3.3.2. Metin Madenciliği .....	45
1.3.3.3. Web Madenciliği .....	46
1.3.4. Veri Madenciliğinin Avantaj ve Dezavantajları .....	47
1.3.4.1. Veri Madenciliğinin Avantajları.....	47
1.3.4.2. Veri Madenciliğinin Dezavantajları .....	48
1.3.5. Veri Madenciliği Modelleri .....	51
1.3.5.1. Sınıflandırma ve Regresyon Modelleri.....	53

1.3.5.2. Kümeleme.....	54
1.3.5.3. Birliktelik Kuralları .....	56
1.3.6. Veri Madenciliği Uygulama Süreci .....	57
1.3.6.1. İşin Tanımlanması (Business Understanding).....	60
1.3.6.2. Veriyi Anlama (Data Understanding).....	60
1.3.6.3. Veriyi Hazırlama (Data Preperation).....	62
1.3.6.4. Modelleme (Modelling).....	74
1.3.6.5. Değerlendirme (Evaluation) .....	75
1.3.6.6. Uygulama (Deployment) .....	75
<b>BÖLÜM 2: KAMU HİZMETLERİNDE BÜYÜK VERİ VE VERİ MADENCİLİĞİNİN GELİŞİMİ.....</b>	<b>78</b>
2.1. Kamuda Büyük Veri ve Veri Madenciliğinin Gelişimini Etkileyen Başlıca Yaklaşımlar .....	79
2.1.1. Küreselleşme ve Yeni Kamu İşletmeciliği .....	80
2.1.2. Bilgi Toplumu ve Kamu Yönetiminin Yenilikçi Teknolojilere Uyumu.....	83
2.1.3. Kanıta Dayalı Politika Yapımı: Kamu Politikaları ve Veri İlişisine Yönelik Bir Çerçeve .....	86
2.1.4. Açık Yönetim.....	88
2.1.5. Dijital Çağ Yönetişimi.....	89
2.2. Kamuda Büyük Veri ve Veri Madenciliğinin Gelişimine Temel Oluşturan Politika ve Uygulamalar .....	92
2.2.1. E-Devlet.....	93
3.2.1.1. E-Devletin Aşamaları ve Boyutları .....	94
3.2.1.2. E-Devlette İlişki Örüntüleri .....	95
2.2.2. Akıllı Şehirler .....	97
2.2.3. Açık Veri Yönetimi .....	100
2.2.4. Web 2.0'a Geçiş.....	102
<b>BÖLÜM 3: TÜRKİYE'DE KAMU HİZMETLERİNDE BÜYÜK VERİ VE VERİ MADENCİLİĞİNİN KULLANIMI.....</b>	<b>105</b>
3.1. Veri Madenciliğinin Kullanım Alanları .....	106

3.2. Büyük Veri ve Veri Madenciliğinin Kamu Hizmetlerinde Başarılı Uygulanması İçin Gerekli Şartlar.....	108
3.3. Büyük Veri ve Veri Madenciliğine Yönelik Yasal Düzenlemeler.....	110
3.3. Merkezi İdare Kuruluşlarında Büyük Veri ve Veri Madenciliğine İlişkin Politika ve Uygulamalar.....	112
3.3.1. Büyük Veri ve Veri Madenciliğine Yönelik Ulusal Politika ve Programlar.....	115
3.3.2. Bakanlıkların Büyük Veri ve Veri Madenciliği Politikaları.....	120
3.3.3. Türkiye’de Yürütülen Başlıca Büyük Veri ve Veri Madenciliği Projeleri.....	124
3.4. Yerel Yönetimlerde Büyük Veri ve Veri Madenciliğine İlişkin Politika ve Uygulamalar.....	127
3.4.1. Yerel Yönetimlerde Büyük Veri ve Veri Madenciliğine Duyulan İhtiyaç.....	128
3.4.2. Yasal Düzenlemeler Açısından Yerel Yönetimlerde Büyük Veri ve Veri Madenciliği.....	130
3.4.3. Yerel Yönetimlerde Büyük Veri ve Veri Madenciliği Uygulamaları.....	132
<b>BÖLÜM 4: KAMUDA VERİ MADENCİLİĞİ UYGULAMASI OLARAK ÇÖZÜM MASASI HİZMETLERİ ÖRNEĞİ.....</b>	<b>136</b>
4.1. İşin Tanımlanması.....	139
4.2. Veriyi Anlama.....	140
4.3. Verinin Hazırlanması.....	148
4.4. Modelin Geliştirilmesi.....	151
4.4.1. WEKA Yazılımı.....	152
4.5. Elde Edilen Bulgular.....	157
4.6. Modelin Sınırlılıkları.....	163
4.7. Tartışma.....	166
<b>SONUÇ VE ÖNERİLER.....</b>	<b>182</b>
<b>KAYNAKÇA.....</b>	<b>190</b>
<b>EKLER.....</b>	<b>214</b>
<b>ÖZGEÇMİŞ.....</b>	<b>217</b>

## KISALTMALAR

<b>AKP</b>	: Adalet ve Kalkınma Partisi
<b>AR-GE</b>	: Araştırma Geliştirme
<b>BİT</b>	: Bilgi ve İletişim Teknolojileri
<b>BTYK</b>	: Bilim Ve Teknoloji Yüksek Kurulu
<b>CHP</b>	: Cumhuriyet Halk Partisi
<b>CRISP-DM</b>	: Veri Madenciliği için Sektörler Arası Standart Süreç Modeli
<b>ETL</b>	: Çıkar, Dönüştür, Yükle
<b>GFS</b>	: Google Dosya Sistemi
<b>HDFS</b>	: Hadoop Dosya Sistemi
<b>IDC</b>	: Uluslararası Veri Şirketi
<b>IOT</b>	: Nesnelerin İnterneti
<b>KBB</b>	: Kocaeli Büyükşehir Belediyesi
<b>KMYKK</b>	: Kamu Mali Yönetimi ve Kontrol Kanunu
<b>MERNİS</b>	: Merkezi Nüfus İdare Sistemi
<b>OGS</b>	: Otomatik Geçiş Sistemi Projesi
<b>RDBMS</b>	: İlişkisel veritabanı yönetim sistemi
<b>SEMMA</b>	: Örnekle, İncele, Düzelt, Modelle, Değerlendir
<b>TAKBİS</b>	: Tapu Kadastro Bilişim Sistemi
<b>TUENA</b>	: Ulusal Enformasyon Altyapısı Ana Planı
<b>VEDOP</b>	: Vergi Daireleri Otomasyonu

## TABLO LİSTESİ

<b>Tablo 1</b> : Bulut Bilişim Üzerinde Sunulan Hizmetler .....	23
<b>Tablo 2</b> : Bulut Bilişim Modelleri .....	25
<b>Tablo 3</b> : Veritabanı ve Veri Ambarı Arasındaki Karşılaştırma.....	40
<b>Tablo 4</b> : Kanıt Türleri ve Kullanım Amaçları .....	87
<b>Tablo 5</b> : Veri Madenciliği Kullanım Alanları .....	106
<b>Tablo 6</b> : Kamu İnternet Siteleri Mevzuat Bilgileri.....	109
<b>Tablo 7</b> : Bakanlıkların Büyük Veri ve Veri Madenciliğine İlişkin Amaç ve Hedefleri .....	121
<b>Tablo 8</b> : Türkiye’de Bazı Kamu Kurumları Tarafından Yürütülen Büyük Veri Projeleri .....	125
<b>Tablo 9</b> : Büyükşehir Belediyelerinin Büyük Veri ve Veri Madenciliğine İlişkin Amaç ve Hedefleri .....	133
<b>Tablo 10</b> : Kocaeli Büyükşehir Belediyesi Çözüm Masası Örnek Başvuruları.....	143
<b>Tablo 11</b> : Kocaeli Büyükşehir Belediyesi Çözüm Masası Başvuru İçeriği .....	144
<b>Tablo 12</b> : Sınıflandırma Algoritmaları Doğruluk Karşılaştırması.....	166
<b>Tablo 13</b> : SMO ve J48 Algoritmalarının Doğru Pozitif Oranları, Yanlış Pozitif Oranları, Kesinlikleri, F-Ölçütleri, ROC Alanları ve Kappa İstatistikleri .	167
<b>Tablo 14</b> : SMO ve J48 Algoritmaları Hata Metrikleri .....	167
<b>Tablo 15</b> : Öznitelik Seçimi Sonucunda Bulunan Kök Kelime Örnekleri.....	169
<b>Tablo 16</b> : Öznitelik Seçimi İşlemi Sonrası Sınıflandırma Algoritmaları Doğruluk Karşılaştırması.....	169

## ŞEKİL LİSTESİ

Şekil 1 : Gelir Bazlı Büyük Veri Pazarı Tahmini (2011-2026) (milyar ABD doları) ..	11
Şekil 2 : Açık Kaynak ve Paralel İşleme Yeteneğine Sahip Veritabanları .....	13
Şekil 3 : Endüstri Devrimleri .....	17
Şekil 4 : Endüstri 4.0 Devriminin Unsurları .....	21
Şekil 5 : Bulut Bilişim Uygulamaları ve Altyapısı .....	22
Şekil 6 : Veri Madenciliği ve İlgili Disiplinler .....	34
Şekil 7 : Veritabanı Teknolojilerinin Gelişimi.....	38
Şekil 8 : Data Mart Veritabanları.....	40
Şekil 9 : Üç Katmanlı Veri Ambarı Mimarisi.....	41
Şekil 10 : Veri Madenciliği Modelleri .....	52
Şekil 11: Veritabanlarında Bilgi Keşfi Aşamaları .....	58
Şekil 12 : SEMMA Aşamaları .....	58
Şekil 13 : CRISP-DM Metodolojisi.....	59
Şekil 14 : ETL (Çıkar, Dönüştür, Yükle) Süreci .....	61
Şekil 15 : Kök Sözcük Ağacı .....	66
Şekil 16: Veri Madenciliği Metodları.....	76
Şekil 17 : Dijital Çağ Yönetişimi Birinci ve İkinci Dalgası .....	90
Şekil 18 : Türkiye Cumhuriyeti cumhurbaşkanlığı Teşkilatı.....	113
Şekil 19 : Metin Madenciliği Süreci ve Çevre Bileşenlerle İlişkisi.....	140
Şekil 20 : Çözüm Masası Başvuru Ekranı .....	142
Şekil 21 : Çözüm Masası Aktivite İşlemleri .....	142
Şekil 22 : Çözüm Masası Verisine Ait Sorgu .....	143
Şekil 23 : Yıllara Göre Başvuru Adedi .....	145
Şekil 24 : Kocaeli Büyükşehir Belediyesi Çözüm Masası Başvuru Tiplerine Göre Başvuru Adedi .....	146
Şekil 25 : Çözüm Masası Uygulamasında Kullanılan Tablolar.....	146
Şekil 26 : Microsoft SQL Server Çözüm Masası Diagramı .....	147
Şekil 27 : Çözüm Masası Tablolarına Ait Örnek Veri.....	147

<b>Şekil 28</b> : Text2arff Yazılımı Grafik Arayüzü.....	150
<b>Şekil 29</b> : Çözüm Masası Veri Kümesinin ARFF Formatındaki Gösterimi .....	151
<b>Şekil 30</b> : WEKA Paket Programı .....	153
<b>Şekil 31</b> : WEKA Gezgini ekranı.....	153
<b>Şekil 32</b> : Weka Gezgin Ekranı-Sınıflandırma Modülü Bölümü.....	154
<b>Şekil 33</b> : Sınıflandırma Algoritması Seçim Ekranı .....	154
<b>Şekil 34</b> : Algoritma Parametreleri Değişim Ekranı .....	155
<b>Şekil 35</b> : Çözüm Masası Veri Madenciliği Süreci.....	156
<b>Şekil 36</b> : Çözüm Masası Süreci Modeli .....	157
<b>Şekil 37</b> : Sınıflandırma Algoritması Seçim Ekranı .....	158
<b>Şekil 38</b> : K Katlamalı Çarpraz Doğrulama Yöntemi.....	160
<b>Şekil 39</b> : Çözüm Masası Örnek Başvurusuna Ait Weka Sınıflandırma Çıktısı .....	160
<b>Şekil 40</b> : WEKA Uygulaması Öznitelik Seçimi Ekranı .....	168
<b>Şekil 41</b> : NaiveBayes Algoritması (TFIDF Ağırlıklandırma) Sonucu .....	170
<b>Şekil 42</b> : NaiveBayes Algoritması (TF Ağırlıklandırma) Sonucu.....	171
<b>Şekil 43</b> : Öznitelik Seçimi Sonrası NaiveBayes Algoritması (TFIDF Ağırlıklandırma) Sonucu .....	171
<b>Şekil 44</b> : Öznitelik Seçimi Sonrası NaiveBayes Algoritması (TF Ağırlıklandırma) Sonucu .....	172
<b>Şekil 45</b> : SMO Algoritması (TFIDF Ağırlıklandırma) Sonucu.....	173
<b>Şekil 46</b> : SMO Algoritması (TF Ağırlıklandırma) Sonucu .....	173
<b>Şekil 47</b> : Öznitelik Seçimi Sonrası SMO Algoritması (TFIDF Ağırlıklandırma) Sonucu .....	174
<b>Şekil 48</b> : Öznitelik Seçimi Sonrası SMO Algoritması (TF Ağırlıklandırma) Sonucu .....	174
<b>Şekil 49</b> : J48 Algoritması (TFIDF Ağırlıklandırma) Sonucu.....	176
<b>Şekil 50</b> : J48 Algoritması (TF Ağırlıklandırma) Sonucu .....	176
<b>Şekil 51</b> : Öznitelik Seçimi Sonrası J48 Algoritması (TFIDF Ağırlıklandırma) Sonucu .....	177
<b>Şekil 52</b> : Öznitelik Seçimi Sonrası J48 Algoritması (TF Ağırlıklandırma) Sonucu ...	177
<b>Şekil 53</b> : Öznitelik Seçimi Sonrası IBk Algoritması (TFIDF Ağırlıklandırma) Sonucu .....	179

<b>Şekil 54 :</b> Öznitelik Seçimi Sonrası IBk Algoritması (TF Ağırlıklandırma) Sonucu... 179
<b>Şekil 55 :</b> Öznitelik Seçimi Sonrası MultiLayer Algoritması (TFIDF Ağırlıklandırma) Sonucu ..... 181
<b>Şekil 56 :</b> Öznitelik Seçimi Sonrası MultiLayer Algoritması (TF Ağırlıklandırma) Sonucu ..... 181



<b>Tezin Başlığı</b> : Kamu Hizmetlerinde Veri Madenciliği: Çözüm Masası Verileri Temelinde Bir Araştırma
<b>Tezin Yazarı</b> : Yılmaz DEMİRCİ Danışman : Doç. Dr. Özer KÖSEOĞLU
<b>Kabul Tarihi</b> : 21.09.2018 <b>Sayfa Sayısı</b> : IX (ön kısım)+217 (tez)+3 (ekler)
<b>Anabilim Dalı</b> : Kamu Yönetimi <b>Bilim Dalı</b> : Kamu Yönetimi
<p>Kurumlar tarafından kullanılan yönetim bilişim sistemleri, gelişen akıllı teknolojilerin etkisiyle oluşan büyük veriden gizli bilgi örüntülerinin çıkarılması ve geleceğe dönük kararlarda kurum yöneticilerine karar desteğinin sağlanması büyük önem arz etmektedir.</p> <p>Kamu yönetimi disiplinde teknoloji odaklı çalışmalar genellikle teorik düzeyde ve ağırlıklı olarak “e-devlet” konusunda yoğunlaşmaktadır. Veri madenciliği uygulamaları ise genellikle yönetim bilişim sistemleri, bilgisayar bilimleri ve işletme gibi disiplinlerde özel sektör verisi ile çalışılmaktadır.</p> <p>Bu çalışma, veri madenciliği konusunu kamu yönetimi ile yönetim bilişim sistemleri disiplinlerine dayalı olarak incelemektedir. Çalışmanın uygulama bölümünde, literatürdeki genel eğilimden farklı olarak, kamu verisiyle veri madenciliği uygulaması gerçekleştirilmiştir. Veri madenciliği için, bir büyükşehir belediyesinden elde edilen çözüm masası verileri Naive Bayes, Destek Vektör Makinesi, K-En Yakın Komşuluk ve Karar Ağaçları gibi makine öğrenmesi algoritmaları kullanılarak analiz edilmiştir. Elde edilen bulguların görsel gösterimi içinse iş zekâsı uygulaması olan “Tableu” kullanılmıştır.</p> <p>Çalışmada, Türkiye’de büyük verinin son yıllarda kamu sektörü kuruluşlarında yaygınlaştığı, kurumların stratejik planlarında yer verildiği, ancak veri madenciliği uygulamalarının çok az kurumda etkin olarak kullanıldığı sonucuna varılmıştır. Uygulama bulguları, yapılandırılmamış veri üzerinde ön işleme aşamasının dikkatli ve doğru şekilde yapılmasının makine öğrenmesinin doğruluk oranlarına doğrudan etki ettiğini göstermesi açısından önemlidir.</p> <p>Büyük veri ve veri madenciliği uygulamalarının, hükümet hizmetlerini, ayrıca devlet operasyonlarını, politika üretme ve yönetimini geliştirmek için kamu sektörü tarafından etkin olarak kullanılabileceği sonucuna ulaşılmıştır. Veri madenciliğinin yalnızca sayısal yöntemleri içeren yazılım aracı değil; çözümüne ihtiyaç duyulan probleme göre tasarlanmış, ilgili yöntem, teknik ve uygulamaları da kapsayan, sonuçları itibariyle probleme ait ilişki, kural ve örüntüyü modelleyen ve gösteren bir süreç olarak kamu hizmetlerinde kullanılabileceğini göstermesi açısından da bu tez önem arz etmektedir.</p>
<b>Anahtar Kelimeler:</b> Büyük Veri, Veri madenciliği, Makine Öğrenmesi, Açık Veri, Endüstri 4.0

<b>Title of the Thesis:</b> Data Mining In Public Services: A Research On Municipality Help Desk Data	
<b>Author :</b> Yılmaz DEMİRCİ <b>Supervisor:</b> Assoc. Prof. Özer KÖSEOĞLU (PhD).	
<b>Date</b> : 21.09.2018	<b>Nu. of Pages:</b> IX (pre text)+ 217 (main body) + 3 (appendices)
<b>Department :</b> Public Administration <b>Subfield</b> : Public Administration	
<p>Management information systems used by the governments and public agencies are crucial in terms of acquiring latent information patterns comprised from big data generated by the developing smart technologies and provision of decision supports on future decisions to policy makers and public managers.</p> <p>Technology-based studies in public administration are generally conducted on the basis of theoretical and practical dimensions of "e-government". Data mining applications are usually studied with focusing on private sector data in disciplines such as management information systems, computer sciences and business administration.</p> <p>This study examines data mining, on the accounts of the disciplines of public administration and management information systems. In the empirical part of the study, fourth chapter, data mining process is implemented with public data, unlike the general tendency in the literature. The help desk data of the Kocaeli Metropolitan Municipality is used in the study. Preprocessing of data and classification methods are implemented via "Weka Machine Learning" tool. The help desk data is analyzed using a number of machine learning algorithms such as Naive Bayes, Support Vector Machine, K-Nearest Neighborhood and Decision Trees. The results were visualized with a business intelligence application called "Tableu".</p> <p>It was concluded that while there is an increasing awareness in recent years on big data technologies and data mining in governments and public agencies in Turkey, the number of applications and projects have still been outnumbered. In essence our study shows that, careful and accurate pre-processing of the raw data, qualitative or quantitative, has a direct impact on the accuracy of machine learning algorithms.</p> <p>Finally, it seems that big data and data mining applications can be effectively used by the public agencies to enhance government operations, to provide effective and efficient public services, and to improve the quality of public policy-making. Data mining is not only a software tool that contains numerical methods; but it includes methods and applications intended to solve the real world problems. This thesis is also important in that it shows data mining can be adopted in public services with a generic model.</p>	
<b>Keywords:</b> Big Data, Data Mining, Machine Learning, Open Data, Industry 4.0	

## GİRİŞ

### Çalışmanın Konusu

Bilişim teknolojilerinin sağladığı imkânlar sayesinde veritabanları daha fazla veriyi saklayabilecek özelliklere kavuşmakta, buna bağlı olarak üretilen sayısal veri miktarı her geçen gün artmakta ve büyük veri oluşmaktadır. Günümüzde dijital verinin hacmindeki artış, dünyanın yanı sıra ülkemizde de devam eden büyük veri ve bulut bilişim gibi projeler, örgütlerin süreçlerinin yeniden yapılandırılmasını gerekli kılmaktadır. Dijital verinin artışıyla birlikte kamu kurumları çok boyutlu ve karmaşık veriyi işlemek için yeni yöntem, altyapı ya da sistemler geliştirme ihtiyacı duymaya başlamıştır. Sayısal olarak artan veriden anlamlı ve kullanılabilir veriye ulaşma ihtiyacı ise yeni bir disiplinin doğmasına yol açmıştır. “Veri madenciliği” adı verilen bu disiplin, verinin çeşitli istatistiksel metotlarla analiz edilmesinin ardından kurumların karar verme sürecine katkıda bulunmaktadır (Kaya ve Köymen, 2008: 159). Veri madenciliği kendi başına bir çözüm olmak yerine, verilecek karar süreçlerini destekleyen ve problemleri çözmek amacıyla gerekli bilgileri sağlamaya yarayan bir çözümdür (Baykal, 2006: 96).

Büyük verinin analizi, yapılan analizler neticesinde anlamlı bilginin elde edilmesi ve yorumlanması süreci insan yeteneği ve ilişkisel veritabanı yöntemlerini aşmaktadır. Yerel yönetimler açısından bakıldığında ise geleneksel analizlerin yetersiz kalmaya başlaması ve geleceğe dönük tahminlere duyulan ihtiyaç veri madenciliği uygulamalarını gündeme getirmiştir.

Bu çalışma özel sektörde olduğu kadar kamu sektöründe de son yıllarda giderek stratejik bir önem kazanan, sağlıktan eğitime, çevre sorunlarından trafiğe kadar pek çok kamusal hizmet alanında oluşan büyük verinin kamu yönetimi için önemi ve veriden geleceğe yönelik tahminlerin oluşturulmasını sağlayan veri madenciliği uygulamalarının kamu hizmetlerinde kullanılmasını konu edinmektedir. Yeni kamu işletmeciliği akımının etkisiyle kamu yönetiminde başlayan piyasalaşma eğilimi, özel sektör yönetim tekniklerinin kamu yönetimine aktarılmasının hızlanması, küreselleşmeyle birlikte bilgi ve iletişim teknolojilerine ilişkin gelişmelerin yayılması, kanıta dayalı kamu politikası yapımı ve bilgi toplumu gibi gelişmelere dayalı olarak büyük veri, veri madenciliği ve

bunlarla ilişkili yeni teknolojilerin kamu örgütlerine ve kamu hizmetlerine entegrasyonu söz konusu olmuştur.

Türkiye’de de söz konusu teknolojik gelişmeler özel sektör kuruluşlarının yanında siyasi partiler ve farklı alanlarda hizmet sunan kamu kurumları tarafından yakından takip edilmektedir. Büyük veri, veri madenciliği ve ilgili diğer teknolojik yenilikler kalkınma planları, hükümet programları ve kurumsal stratejik planlar gibi politika dokümanlarında yer almaktadır. Kamu kurumlarının yenilikçi teknolojilerin sunduğu imkânlardan yararlanmayı amaçladığı anlaşılmaktadır. Bu çalışma, veri madenciliği uygulamalarının kamusal sorunların çözümünde ne gibi alanlarda kullanılabileceğini araştırmaktadır.

### **Çalışmanın Önemi**

Kamu kurumlarının arşivledikleri ve işlenmeye hazır hale getirerek toplumla paylaştıkları veri çok ciddi boyutlara ulaşmıştır. Farklı hizmet alanlarına ilişkin büyük verinin analiz edilerek geleceğe yönelik tahminlerde, verilecek kararlarda, tespit edilen veya şikâyet konusu olan sorunların çözümünde, kamu politikalarının oluşturulmasında ve hizmet kalitesinin iyileştirilmesinde kullanılması için veri madenciliği önemli rol oynamaktadır. Kurumların hali hazırda kullandıkları yönetim bilişim sistemleri sayesinde veri tabanlarında büyük veri oluşmaktadır. Büyük veri hacmi artmaya devam ederken veri alanındaki analiz ve tahminleri geliştirebilecek, veri madenciliği, makine öğrenmesi, model ve algoritmalarını işleyebilecek, anlamlandırabilecek veri bilimcilerine, nitelikli insan gücüne ihtiyaç artarak devam etmektedir.

Türkiye’de büyük veri, son yıllarda giderek daha fazla ilgi gören bir alan haline gelmiştir. Sosyal ağlar, gelecek tahmin sistemleri, akıllı kentler gibi farklı alanlarda uygulamalı ve teorik çalışmalara rastlanmaktadır (Razboyanlı, 2016; Aksu, 2014; Dalgaldere, 2016; Akdamar, 2017a). Öte yandan, çalışmanın temel konusunu oluşturan veri madenciliğine ilişkin Türkçe literatürde yapılan araştırma sayısı her geçen gün artmaktadır. Veri madenciliği çalışmaları sadece yönetim bilişim sistemleri ve bilgisayar bilimleri alanlarında değil işletme ve kamu yönetimi gibi disiplinler altında da çalışılmaktadır. Pazarlama, bankacılık ve finans gibi özel sektör alanlarının yanında eğitim, sağlık, güvenlik gibi kamu sektörü hizmet alanlarında da çalışmalar bulunmaktadır. Bu bağlamda, kullanıcı merkezli interaktif olarak veri tabanlarında

gerçekleştirilecek özbilgi keşfi için kullanılan veri madenciliği çalışması (Karacan ve Yeşilbudak, 2010: 17), yapısal olmayan verinin metin ve web madenciliği yöntemleri ile yapısal hale dönüştürülmesi (Dolgun ve diğerleri, 2009: 48), sağlık sektöründeki veriler kullanılarak gelen hasta profillerinin belirlenmesi (Ertuğrul ve diğerleri, 2013: 97), öğrenci taleplerindeki değişimler (Sevindik ve diğerleri, 2012: 183), sosyal ağların öğrencileri ne derecede etkilediği (Koç ve Karabatak, 2012: 155) gibi araştırmalar farklı disiplinlerde veri madenciliğinin nasıl çalışıldığına örnek oluşturmaktadır.

Yüksek Öğretim Kurumu (YÖK) tez arşivi incelendiğinde veri madenciliğine ilişkin çalışmaların ağırlıklı olarak sağlık, eğitim, pazarlama, işletme ve borsa veri kümeleri üzerinden ilgili sektörlerde çözüm üretmek üzere yoğunlaştığı görülmektedir. Kamu yönetimi alanında teknoloji çalışmaları ağırlıklı olarak e-devlet konusuna yoğunlaşmakta, son yıllarda ise akıllı kentler ilgi çekmektedir. Kamu yönetimi-teknoloji ilişkisi üzerine yapılan çalışmaların ise genelde teorik düzeyde ve gelişmelerin/yeniliklerin tanıtılması şeklinde geliştiği görülmektedir. Uygulama araştırmaları daha çok özel sektör verisi kullanılarak farklı sektörlerde yürütülmektedir.

Literatürde büyük veriye ilişkin çalışmalar son dönemde artmaya başlamakla birlikte Ağustos 2018 itibariyle YÖK tez arşivinde doktora tezi seviyesinde gerçekleştirilmiş dokuz adet çalışma yapıldığı veri madenciliğine ilişkin çalışmaların ise ağırlıklı olarak sağlık, eğitim, pazarlama, işletme ve borsa veri kümeleri üzerinden ilgili sektörlerde çözüm üretmek üzere yoğunlaştığı görülmektedir. Son yıllarda kamu sektöründe ve özelinde yerel yönetimlerde oluşan veri trafiği neticesinde kurumların elinde büyük veri oluşmaktadır. Kamu yatırım programları, stratejik planlar gibi kamu belgeleri incelendiğinde kamununda elindeki veriden beklentisinin değiştiğini büyük veri yaklaşımının ve veri madenciliği uygulamalarının kamunun da dikkatini çektiğini görmek mümkündür. Yapılan çalışmalar, kamu yönetimi-teknoloji ilişkisinin ülkemizde algılandığı üzere sadece “e-devlet” tanımından ibaret olmadığını büyük veri teknolojilerinin ve geleceğe yönelik olmak üzere veri madenciliğinin karar vericilere destek sağlayabilecek kullanışlı bilgiler sağlayabileceğini göstermesi açısından önemlidir. Bu tez, veri madenciliği uygulama sürecini takip ederek veri madenciliğinin kamu hizmetlerinde nasıl kullanılacağını bir örnek uygulama ile ortaya koymaktadır.

Veri madenciliğinin kamu hizmetlerine nasıl entegre edilebileceğine ilişkin ampirik bir örnek geliştirerek bu alandaki eksikliğe dikkat çekmeyi amaçlamaktadır.

Bu çalışmada belediyelerin metinden oluşan çözüm masası veri setleri makine öğrenme ve metin madenciliği kullanılarak analiz edilmekte ve böylece somut bir veri madenciliği girişimi örneği sunulmaktadır. Veri madenciliği sürecine iş zekâsı araçlarının eklenmesiyle çözüm masası hizmetlerinin iyileştirilmesi için bir model geliştirilmesi mümkün hale gelmektedir. Bu nedenle, çalışmanın kamu hizmetlerinde büyük verinin önemine dikkat çekmesinin yanında, gerçekleştirdiği veri madenciliği uygulaması ile kamu kurumları için somut çözümler üretme potansiyeli de bulunmaktadır.

### **Çalışmanın Amacı**

Kamu kurumlarının hali hazırda kullandıkları yönetim bilişim sistemleri, operasyonel işlemlerin yürütülmesi ve sürekli ihtiyaç duyulan standart raporların hızlı bir şekilde oluşturulması amacıyla çalışanlara kolaylık sağlamakta, işlemler neticesinde oluşan bütün veri, veritabanlarında saklanmaktadır. Son dönemde artan işlem hacmi ve kullanılan akıllı teknolojilerin çeşitlenmesi ile birlikte veri boyutu ve işlem hızı artarak kamu verisi büyük veri teknolojileri ile depolanmaya ve işlenmeye başlamıştır.

Yerel yönetimlerin de içinde olduğu kamu örgütleri; karar destek tabanlı bilginin değerinin arttığı günümüzde gelişen teknoloji ve özel sektörün elde ettiği başarılar üzerine hizmet sunumu, karar verme, gelecek planlaması, kaynakların verimli kullanılması ve politika yapımını iyileştirme yönünde büyük veri ve veri madenciliği uygulamalarını içeren projelere yönelmeye başlamıştır. Kamu sağlığı, milli savunma, istihbarat, vergi gelirleri gibi pek çok konuda da büyük veri teknolojileri ile veriler saklanmakta; veri madenciliği, makine öğrenmesi ve iş zekâsı uygulamalarından yararlanılarak veriler işlenmektedir.

Çalışmanın amacı, kamu yönetiminde büyük veri ve yenilikçi teknolojiler ile veri madenciliğinin önemini ortaya koyarak, makine öğrenmesi ve veri madenciliği işlemlerinin yerel kamusal sorunların çözümünde etkili bir araç olarak nasıl kullanılabileceğini göstermek için bir uygulama gerçekleştirmektir. Türkiye’de yerel kamu hizmetlerinin yürütülmesinde büyük veri teknolojileri ve veri madenciliği son

yıllarda ilgi çekmektedir. Özellikle akıllı kentlere yönelik artan teorik ve pratik ilgi nedeniyle, büyük veri, makine öğrenmesi ve veri madenciliği gibi uygulamalar da önem kazanmaktadır. Bu nedenle, çalışmanın sadece teorik-kavramsal düzeyde kalmaması amacıyla, çözüm masasından elde edilen metin verisi kullanılarak bir metin madenciliği uygulaması sayesinde kamusal hizmet kalitesinde iyileşme sağlanabileceği kanıtlanmaya çalışılmaktadır. Bu çalışmanın bir temel ve üç alt araştırma sorusu bulunmaktadır:

1. Büyük veriye dayalı teknolojiler ve özel olarak veri madenciliği uygulamaları kamu hizmetlerinin sunulması ve karar verme süreçlerine nasıl entegre edilebilir?
  - a. Büyük veri teknolojileri ve veri madenciliğinin kamu yönetimi ve hizmet sunumu açısından önemi nedir?
  - b. Türkiye’de büyük veriye dayalı yeni teknolojiler ve özel olarak veri madenciliğine yönelik politika ve projeler nelerdir?
  - c. Veri madenciliği uygulamasına makine öğrenmesi ve iş zekâsı gibi teknolojiler entegre edilerek yerel kamusal sorunların çözümünde yeni araç veya modeller geliştirilebilir mi?

### **Çalışmanın Yöntemi ve Sınırlılıkları**

Veri madenciliği uygulamaları temel olarak istatistik, makine öğrenmesi ve veritabanı teknolojilerinden oluşan bilgisayar bilimlerini ilgilendirmekle birlikte davranış bilimleri, toplum bilimleri ve işletme gibi pek çok farklı disiplini bir araya getirmektedir. Veri madenciliği uygulamaları sayısal (nicel) ve sözel (nitel) bilimleri bir araya getiren disiplinler arası bir alan olması nedeniyle bu çalışmada farklı veri toplama ve analiz araçları birlikte kullanılmıştır.

Çalışmanın araştırma tasarımı için birkaç ön hazırlık yapılmıştır. İlk olarak, kamuda veri madenciliğinin potansiyel uygulama alanları ve ulaşılabilir veri türlerini keşfetmek amacıyla trafik, ulaşım, afet yönetimi ve bütçeleme gibi alanlarda çalışma yapabilmek için farklı kurumlarla iletişime geçilmiştir. Ancak trafik alanında halihazırda bazı projelerin bulunması ve afetle ilgili verilerin ise yerelde değil, kurum merkezinde

tutulduğu bilgisine ulaşılması nedeniyle farklı bir alanda uygulama yapılmasına karar verilmiştir. Sonuçta belediyelerin çoğunda el yordamıyla ve basit ofis uygulamalarıyla toplanan ve değerlendirilen çözüm masası verilerinin bir veri madenciliği türü olan metin madenciliği ile analiz edilebileceği düşünülmüştür. Sözlük uygulamaları, makine öğrenmesi ve veri madenciliği algoritmaları yardımıyla çözüm masası verilerinin işlenmesi ve sınıflandırılarak nitelikli bilgi üretilmesi amaçlanmıştır. Ayrıca iş zekası uygulamalarının sisteme eklenerek tüm belediyeler için kullanılabilir bir model üretilmesi mümkün olacaktır.

Kocaeli Büyükşehir Belediyesi “Çözüm Masası” uygulaması 12 Eylül 2008 tarihinde vatandaşın hizmetine sunulmuş olup veriler 07 Haziran 2016 tarihinde temin edilmiştir. 12 Eylül 2008-07 Haziran 2016 tarihleri arasındaki 100402 adet çözüm masası başvuru verisi excel dosyasında teslim alınmıştır. Elde edilen nitel veri, veri madenciliği prosesinden geçirilmiştir. Verinin hazırlanması aşamasında veri birleştirme, boyut indirgeme, veri sıkıştırma ve kesikli hale getirme gibi tekniklerden yararlanmak suretiyle veri seti anlamlı ve veri madenciliği uygulamalarında kullanılabilir hale getirilmiştir. Verinin modellenmesi aşamasında ise elde edilen veri seti üzerinde makine öğrenmesi algoritmaları yardımıyla en uygun modelin elde edilmesine yönelik çalışmalar gerçekleştirilmiştir. Veri madenciliği uygulamalarında doğruluk değeri yüksek modellerin ve yöntemlerin bulunması için farklı veri madenciliği model ve yöntemleri araştırılmış ve test edilmiştir.

Veri madenciliği uygulamaları geliştirmek üzere tüm algoritmaları, zaman ve maliyet kısıtları nedeniyle örnek veri seti üzerinde denemek oldukça güçtür. Bu nedenle çalışmada model olarak sınıflandırma modeli, algoritma olarak literatürde en çok tercih edilen K-NN, karar ağaçları, yapay sinir ağları ve Naive Bayes algoritmaları çalıştırılarak en yüksek doğruluk oranına erişilmeye çalışılmıştır. Algoritmaların bir kısmı beklenildiği gibi başarılı sonuçlar üretirken, bir kısmı veriyi yeterince iyi sınıflandıramamıştır. Uygulamada sınıflandırma modeli kullanılarak, belirlenen algoritma ile önce çözüm masasına gelen isteklerden oluşan veri setiyle eğitilmesi, ikinci olarak model sınıfı belirli olmayan verilere uygulanarak sınıfların tahmin edilmesi amaçlanmaktadır.

Çalışmada kullanılacak veri kümesinin hacmi de veri madenciliği uygulamalarının farklı bir kısıttır. Kullanılan veri setinin büyük hacimli olması, uygulanacak algoritmalar ile sonuçların uzun süre içinde elde edilmesine neden olmaktadır. Uygulamada veri seti üzerinde veri azaltma teknikleri, öznitelik seçme işlemleri uygulanarak elde edilecek veri setinin hacim olarak daha küçük olmasına ve böylece zaman ve maliyet kısıtlarının azaltılmasına çalışılmıştır.

Çalışma kapsamında gerçekleştirilecek uygulama ile veri madenciliğinin kullanımı, modellenmesi ve ilgili algoritmalar ile veri analizinden elde edilen sonuçların sunulması hedeflenmektedir. Bu sayede önerilecek olan modelin, çözüm masası ve beyaz masa gibi isimlerle adlandırılan ve hizmet kullanıcılarının şikâyet, talep ve bilgi edinme isteklerinin çözüme kavuşturulmasını amaçlayan merkezi ve yerel kamu kuruluşlarının ilgili birimleri tarafından kullanılabilmesi beklenmektedir.

### **Çalışmanın Kapsamı**

Kamu yönetiminde büyük verinin önemi, büyük veriye dayalı nesnelerin interneti, endüstri 4.0, yapay zekâ ve bulut bilişim gibi yeni teknolojiler ile özel olarak veri madenciliği uygulamalarının kullanımına odaklanan bu çalışma dört bölümden oluşmaktadır. Birinci bölümde büyük veri ve veri madenciliğine ilişkin kavramlar ve veri madenciliği prosesi, ikinci bölümde kamu hizmetlerinde büyük veri ve veri madenciliğinin gelişimi, üçüncü bölümde Türkiye’de kamu yönetiminde büyük veri ve veri madenciliğinin kullanımı ve son olarak dördüncü bölümde veri madenciliği uygulaması örneği olarak çözüm masası hizmetleri için bir model geliştirilmesi amaçlanmaktadır.

Kavramsal çerçevenin oluşturulduğu ilk bölümde büyük veri tanımı, gelişimi ve işlevi, büyük veriyle ilişkili bilişim teknolojileri olarak endüstri 4.0, bulut bilişim, nesnelerin interneti ve yapay zekâ araştırılmıştır. Bu bölümde, çalışmanın ana konusunu oluşturan veri madenciliğiyle ilişkili kavram ve uygulamalar, veri madenciliği türleri, avantaj ve dezavantajlar, modeller ve uygulama süreci detaylandırılmıştır.

Kamu hizmetlerinde büyük veri ve veri madenciliğinin gelişimine yönelik ikinci bölümde teorik arka planı oluşturan yeni kamu işletmeciliği ve küreselleşme, politika yapımında verinin kullanımı, kanıta dayalı politika yapımı, açık yönetim ve dijital çağ

yönetişimi tartışılmıştır. Büyük veri ve veri madenciliğine temel oluşturan e-devlet, akıllı şehirler, açık veri ve Web 2.0 uygulamaları da incelenmiştir.

Türkiye’de kamu hizmetlerinde büyük veri ve veri madenciliğinin kullanımı üçüncü bölümün konusunu oluşturmaktadır. Merkezi idare kuruluşlarında ve yerel yönetimlerde veri madenciliğine ilişkin politika ve uygulamalara yer verildikten sonra genel bir değerlendirme yapılmaktadır.

Veri madenciliği uygulaması örneği olarak çözüm masası hizmetleri için bir model geliştirme çalışması dördüncü bölümde gerçekleştirilmektedir. Verilerin toplanması, tasniflenmesi ve diğer ön işleme aşamalarına tabi tutularak elde edilen veri seti üzerinde WEKA programı yardımıyla analizler yapılmaktadır. Weka, Waikato Üniversitesi tarafından geliştirilmiş olan "Waikato Environment for Knowledge Analysis" kelimelerinin baş harflerinden oluşmuş çoğu makine öğrenimi algoritmalarını ve metodlarını destekleyen yazılımdır. Elde edilen bulgulardan yola çıkılarak yerel yönetimler özelinde veri madenciliğine yönelik literatürdeki eksiklik giderilmeye çalışılmakta hem de bir uygulama modelinin geliştirilmesi için ilk adımlar atılmaktadır.

## **BÖLÜM 1: BÜYÜK VERİ VE VERİ MADENCİLİĞİ**

İlk bilgisayarlar matematiksel hesaplamaları yapmanın dışında başka bir amaç için tasarlanmamışlardı. Temel olarak bugünün bilgisayarları ve bilişim sistemleri ise karmaşık matematiksel işlemleri yapmanın yanında kamusal alandaki geniş bir yelpazede karar vericilere karar vermede destek olmaktadır (White, 2007: 21). Bilgi ve iletişim teknolojileri, özel veya kamu sektöründe faaliyet gösteren kuruluşlara yönetim, bütçeleme, sağlık, kent yaşamı, çevre, eğitim, sağlık ve enerji gibi hizmet alanlarında karşılaştıkları sorunlara çözüm bulmak ve kullanıcılara kaliteli mal ve hizmet sunabilmek için veriden elde edilen analizlere ve geleceğe dönük tahminlere dayalı politikalar üretmeye imkân vermektedir. Kamu kurumları internet ve web temelli teknolojileri kullanarak örgütsel etkinlik ve verimliliği sağlamanın yanı sıra vatandaşlara dönük işlemlerin daha hızlı, kaliteli ve müşteri odaklı şekilde gerçekleşmesini sağlamaktadır.

Bilişim teknolojilerinin kullanımının artmasıyla birlikte elde edilen verinin sadece miktarı artmamakta, aynı zamanda niteliği de değişime uğramaktadır. Geleneksel olarak veritabanlarında tutulan sabit verinin yanı sıra günümüzde sensörler, mobil iletişim araçları, sosyal medya gibi pekçok farklı ortamdaki akıcı veri gelmektedir. Farklı kaynaklardan elde edilen ürün değere, anlamlı bilgiye dönüşmediği sürece kurumlar için bir anlam ifade etmemektedir. Büyük veri teknolojilerinin başlıca amacı, niteliği ve niceliği sürekli değişken olan bahse konu verinin analiz edilmesi, geleneksel analiz yöntemleriyle elde edilememiş yeni sonuçlara ulaşılması ve veriye değer katılmasıdır (Çekin, 2016: 630). Büyük veri teknolojileriyle sürekli akan verinin geleneksel veri depolama yöntemleri yerine yeni bir yaklaşımla depolanması, geleneksel veri analiz araçları yerine makine öğrenmesi olarak son dönemde sıkça literatürde yer alan algoritmaların kullanılması ve bilgisayarlara insan gibi düşünme özelliğinin kazandırılarak geleceğe yönelik tahminlerde bulunmasının sağlanması ifade edilmektedir.

Büyük veri başlı başına bir değer olmakla birlikte geleceğe dönük tahminlerin gerçekleştirilmesinde tek başına bir anlam ifade etmemektedir. İş zekâsı uygulamaları, veri madenciliği, istatistik, tahmin edici analiz ve veri modelleme gibi hali hazırda kullanılmakta olan teknolojileri büyük veri şemsiyesi altında düşünmek mümkündür.

Kurumsal kullanıcılara yardımcı olmak, karar destek sistemlerini kullanarak daha iyi iş kararlarının alınmasını sağlamak üzere kullanılan iş zekâsı, verinin farklı makine öğrenmesi algoritmaları ve teknikler ile büyük veri kümelerinden bilgi keşfi süreci olarak veri madenciliği, istatistiki prensiplerle veriye odaklı anketlere yönelik test ve tahmine dayalı istatistik ve veri modelleme büyük veriden yararlanmaktadır (Ohlhorst, 2013: 5). Farklı kanallardan veri toplayan ve nesnelerin interneti (IoT) ve bulut bilişim gibi uygulamalarla veriyi birbirine bağlayan bilişim teknolojileri sayesinde kapasitesi günden güne artan büyük verinin karar vericiler tarafından kullanılması başta olmak üzere kamu ve özel sektör girişimleri tarafından kullanılabilmesi için yardımcı teknolojilere ihtiyaç duyulmaktadır.

Tek başına bir anlam ifade etmeyen veri yığınları veritabanlarından bilgi keşfi araştırmaları neticesinde, belirli bir amaç için kullanıldığında anlamlı hale gelmektedir. Bu sebeple, ham veriden yararlı bilgiye erişmeyi amaçlayan veri madenciliği çalışmaları günümüzde önemli bir çalışma alanı olarak görülmektedir (Gürbüz, 2009: 26). Büyük veri ve veri madenciliğinin ilişkisi bu noktada başlamaktadır. Büyük veri, gelişen bilişim teknolojileri sayesinde artan verinin geleneksel veritabanı sistemlerinden farklı bir anlayışla depolanmasını, veri madenciliği ise bugüne kadar örgütlerin elinde bulundurduğu ham veriden / büyük veriden elde edilecek örnek veri kümesi üzerinde makine öğrenmesi algoritmalarını kullanmak suretiyle geleceğe yönelik kullanışlı bilginin elde edilmesini içermektedir. Büyük veri ve veri madenciliği birbirini tamamlayan ve birbiri ile iç içe geçmiş veri teknolojileridir.

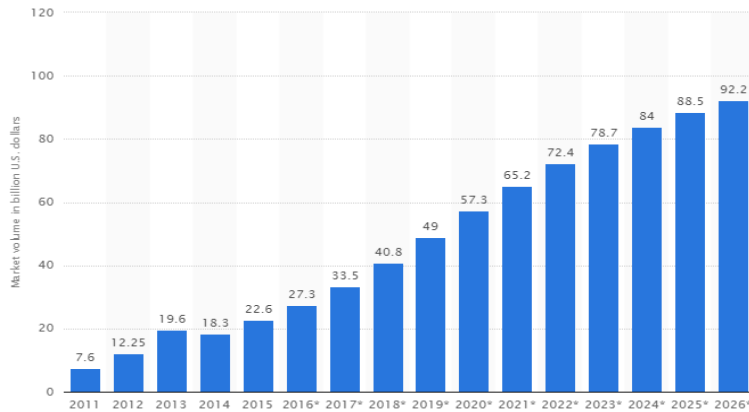
Çok disiplinli bir alan olmasına dayalı olarak özellikle 1980'li yıllardan sonra gündeme gelen bir takım toplumsal, ekonomik, yönetsel ve teknolojik gelişmeler büyük verinin ve veri madenciliğinin gelişmesinde etkili olmuştur. Bu bölümde, aslında kendisi oldukça teknik bir alan olan büyük verinin tanımı, gelişimi ve işlevi, büyük veri ile ilişkili bilişim teknolojileri ve veri madenciliğiyle ilişkili kavram ve uygulamalar, türleri, avantaj ve dezavantajları, modelleri ve son olarak veri madenciliği uygulama süreci incelenmektedir.

### **1.1. Büyük Veri: Tanımı, Gelişimi ve İşlevi**

Bilgi, yönetsel süreçlerde olduğu kadar rekabet üstünlüğü sağlayan bir unsur hale gelmesinin ardından, örgütler açısından daha önemli bir hale gelmiştir. Geleneksel

donanım ve yazılım çözümleriyle saklanması, işlenmesi, paylaşılması ve analiz edilmesi kurumlara yüksek maliyet ve işgücü harcanmasına sebep olan büyük hacimli veri, günümüzde gelişen bilgi teknolojileri ve yatırımları sayesinde işlenebilir hale gelmiştir. Bu kapsamda, sahip olduğu veriden karar alma, politikalar üretme, hizmet sunma ve kâr elde etme gibi sebeplerle yararlanmak isteyen özel sektör kuruluşları ve kamu kurumları bu alana yatırım yapmaya başlamışlardır (Köseoğlu ve Demirci, 2017: 2224).

Dünyada üretilen veri boyutu her yıl katlanarak artmaya devam etmektedir. Uluslararası Veri Şirketi (IDC)'nin yapmış olduğu araştırma sonucuna göre 2013 yılında üretilen veri kapasitesi 4.4 zetabayt iken 2020 yılında veri boyutunun 44 zetabayta ulaşması beklenmektedir (EMC, 2014). Veri madenciliğinin gelişimine etki eden büyük faktör olan veri hacminin 2015 sonrası on yıllık dönemde 44 kat artması, büyük veri pazarının ise büyümeye devam ederek Şekil 1'de görüleceği üzere 2026 yılına gelindiğinde 100 milyar dolar seviyelerine gelmesi beklenmektedir. Veri kapasitesinin hızlı bir şekilde artmasının nedenleri arasında yüksek veri iletişim kapasitelerine ulaşılması, mobil kullanım oranının artması ve sosyal medyanın etkisi gösterilebilir. Bahsi geçen etkenler sayesinde yapılan her işlem neticesinde anlamlı ve işlenebilir halde olmayan "büyük veri" oluşmaktadır. Son zamanlarda büyük veri sadece özel sektörün değil, aynı zamanda kamu kurumlarının da dikkatini çekmektedir.



**Şekil 1:** Gelir Bazlı Büyük Veri Pazarı Tahmini (2011-2026) (milyar ABD doları)

**Kaynak:** <https://www.statista.com/statistics/254266/global-big-data-market-forecast/>,

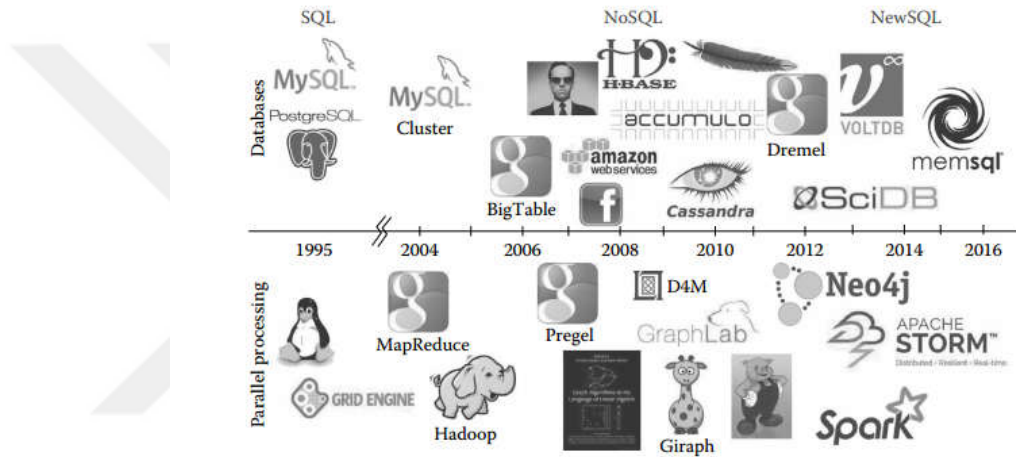
Erişim Tarihi: 10.11.2017.

Yapılandırılmamış metinsel veri; yayın, sosyal medya, haber, blog sitesi, e-posta, kamu resmi yazışmaları ve kamu verisi olmak üzere çeşitli kaynaklarda bulunmaktadır (Korkontzelos, 2014: 231). Kamu sektöründe mevcut kullanım alanları sınırlı olmakla birlikte büyük veri, modern kamu yönetimi uygulamalarında verimlilik ve etkinlik açısından somut sonuçlar doğuran alanlarda dinamik olarak gelişimini devam ettirmektedir. Büyük veri kamu kurumları tarafından kalkınma (Koyuncu, 2016: 2) ve kamu politikası üretme (Altun ve diğerleri, 2017: 3033) gibi farklı amaçlarla kullanılmaktadır. Büyük veri kamuda usulsüzlüklerin belirlenmesi, yasal uyumsuzlukların tespiti amacıyla kamu denetiminde, sosyal davranışı düzenleme ve sosyal ilişkileri şekillendirme (izinler, asaklar veya emirler) amacıyla kamu düzenini sağlamada ve belirli kamu hizmetleri veya ürünleri sağlamak üzere kullanılmaktadır (Maciejewski, 2017: 124).

Geleneksel veri ile büyük veriyi birbirinden ayıran üç temel özellik bulunmakta ve literatürde 3V olarak sembolize edilmektedir. Büyük hacimli (volume), yüksek hızlı (velocity) ve çok çeşitli (variety) veriler “büyük veri” olarak karşımıza çıkmaktadır (Foster ve diğerleri, 2017: 3). Literatürde yer alan 3V tanımının yanına gerçeklik (veracity) kavramı da eklenmektedir. Örgütler terabayt, hatta petabaytlar (1024 terabayt) seviyesinde veriyi günümüzde kolaylıkla toplayabilmekte ve depolayabilmektedir. Büyük veri, metin, ses, video, günlük dosyalar ve sensörlerden gelen akış olmak üzere yapısal olmayan çeşitlilikte veriden oluşmaktadır. Verinin büyüklüğü kadar saflığı ve güvenilirliği de kritik öneme sahiptir. Büyük verinin diğer bir özelliği ise verinin oluştuğu anda, olduğu gibi kullanılmasıdır (Ohlhorst, 2013: 3).

Farklı sektörlerde çalışan örgütlerin ya da bireysel kullanıcıların girdiği veri, özel veya kamusal örgütlerde kurum içi veya kurumlar arası işlemler neticesinde üretilen sağlık, finans, borsa ve diğer hizmetlere ait veri, sosyal medya verisi gibi kaynaklar büyük veri örneği oluşturmaktadır. Söz konusu işlemler neticesinde oluşan büyük veri, gelişen donanım ve yazılım teknolojilerinin yardımıyla veritabanları ve veri ambarlarında kullanılmak ve anlamlı hale getirilmek üzere sayısal ortamda saklanmaktadır. Kurumların sahip olduğu veya ürettiği verinin bu denli büyük hacimlere ulaşması, verinin işlenmesinde geleneksel yöntemleri yetersiz hale getirmiş, ortaya çıkan büyük veriyi analiz edecek yeni teknolojilere ihtiyaç doğmuştur.

Veritabanlarının gelişimine göz atıldığında Şekil 2’de görüleceği üzere SQL, NoSQL ve NewSQL olmak üzere gelişimlerini sürdürdükleri görülmektedir. 1990’ların ortasından itibaren MySQL ve PostgreSQL gibi SQL ilişkisel veritabanları gelişme göstermiş, 2000’li yılların ortasından itibaren ise MapReduce hesaplama yaklaşımı ve ardından Apache Hadoop’un geliştirildiği NoSQL veritabanları olarak adlandırılan Accumulo, Cassandra vb. veritabanları gündeme gelmiştir. Son olarak ilişkisel veritabanının özelliklerini taşımanın yanı sıra veri akışı, makine öğrenmesi ve grafik algoritmaları ile çalışan Apache Storm ve Spark teknolojilerini barındıran NewSQL geliştirilmiştir (Aji ve Wang, 2016: 24).



**Şekil 2:** Açık Kaynak ve Paralel İşleme Yeteneğine Sahip Veritabanları

**Kaynak:** Aji ve Wang, 2016: 24.

Örgütlerin sahip olduğu büyük kapasiteli yapılandırılmamış veri hızla artmakta ve ilişkisel olmayan NoSQL veritabanı adı verilen sistemlerde depolanmaktadır. NoSQL veritabanları, büyük ölçekli veri kümesinin analitik işlemine odaklanarak, veritabanı donanımı üzerinde artan ölçeklenebilirlik sunmaktadır (Moniruzzaman ve Hossain, 2013: 2).

Büyük veri üzerinde uygun yazılımlar, algoritmalar ve teknolojiler kullanılarak toplanan verinin işlenmesi, aralarındaki ilişkinin tespiti, birlikteliklerin belirlenmesi ve örüntülerin çıkarılması mümkün hale gelmiştir. Bu noktada Spark, Storm, FLink, Apache Ranger, Knox Gateway, Kafka, R, Cascading, Elasticsearch, HBase, Hive, Pig Cassandra ve MongoDB gibi programlama dilleri geliştirilmiştir (Pries ve Dunnigan, 2015: 117).

Büyük veri, internet üzerinde yoğun şekilde kullandığımız ve bizlere öneriler sunan arama motorları, Google'ın geliştirdiği şoförsüz gidebilen araç ve birçok farklı kaynaktan elde edilen verinin analiz edilip seçmenlere ulaşma gibi pek çok farklı alanda kullanılmaktadır (Scherer, 2012). Büyük hacimli verinin üretilmesinde en etkili ortamların başında sosyal medya gelmektedir. Tüketici seçimleri, görüşleri ve marka sadakati gibi veriler sosyal medya üzerinde üretilmektedir. Sosyal medya siteleri, aynı zamanda rekabetçi piyasada hizmetlerin pazarlanmasını kişiselleştirmede ve kamu kurumlarının vatandaşa ulaşmasında etkili rol oynamaktadır. Bugün Twitter ve Facebook gibi sosyal medya uygulamalarına yapılan ziyaretler sonucunda oluşan gerçek zamanlı akış verisinin hızı “büyük veri”nin hızı için iyi bir örnektir. İlgili sitelere giren kullanıcılar her 3 dakikada milyonlarca tweet atmaktadır. Toplanan veriler e-posta, tweet, sosyal medya ve sensör verisi olmak üzere çok farklı formatlarda gelmektedir. Gelen veri yapısının formatı hakkında herhangi bir kontrol yoktur. Gerçek verinin ne tür veri barındırdığını belirlemede meta verisinin olması kritik öneme sahiptir. Meta verinin eksik olması verinin incelenmesi ve veri ambarlarına aktarılmasını da geciktirir (Kale, 2018: 174-178).

Büyük veri kurumlara birçok fırsat ve yenilik sunmanın yanı sıra riskleri de beraberinde getirmektedir. Büyük verinin ortaya çıkışı ile birlikte kişi mahremiyetinin korunmasına yönelik yeni zorluklar ortaya çıkmıştır. Geleneksel veritabanları için birçok başarılı kimlik denetimi ve veri güvenliği çözümü bulunmaktadır (Eyüpoğlu ve diğerleri, 2017: 178). Ancak büyük veriye dönük olarak geleneksel çözümlerle gizlilik (mahremiyet), büyük veri analizi, yüksek performanslı kriptografi, güvenlik araştırması için büyük veri kümeleri, veri kaynağı problemi, güvenlik görselleştirme ve nitelikli personel sağlanmasında zorluklar bulunmaktadır (Alguliyev ve Imamverdiyev, 2014: 3). Büyük verinin getirdiği riskleri son dönemde yaşanan örneklerle açıklamak mümkündür. 2012 yılında Cambridge Üniversitesi'nde bir araştırmacı anket yöntemiyle sosyal medya devi Facebook'un üçyüz bin kullanıcılarının verisine ulaşmış, ardından bu veriyi Cambridge Analytica isimli şirketle paylaşmıştır. Kullanıcılara ait verinin 2016 yılında ABD seçimlerini etkilemek amacıyla başkanlık seçimi kampanyalarında kullanıldığı ve 87 milyon Facebook kullanıcılarının etkilendiği iddia edilmiştir (Aziza, 2018).

Büyük veriye ilişkin eleştirilerden birisi de sınırlı erişimin yeni dijital ayrımlar yaratacağı yönündedir. Bazı şirketlerin sosyal medya verisini kendi amaçları doğrultusunda kullanacak araştırmacılara açarken, farklı bir araştırmacıya verisinin belirli bir kısmını açabilmektedir (Boyd ve Crawford, 2015: 210). Bu durumda büyük veride pay sahibi olan üç farklı grup oluşmaktadır. İlk grup büyük veriyi yaratan kesimdir; ikinci grup büyük veriyi depolayan, toplayan kesim iken, üçüncü grup ise büyük veriyi analiz etmede uzmanlaşmış kişileri oluşturmaktadır. Bu durumda, büyük verinin kullanımını ve katılımcılarını belirleyen son grup, en küçük ve en ayrıcalıklı kesimi oluşturmaktadır (Manovich, 2011: 10).

Bilişim altyapısının gelişmesi, Endüstri 4.0 devrimi neticesinde nesnelerin interneti kaynaklarının internet altyapısı üzerinden diğer cihazlarla iletişim kurarak ürettikleri veri miktarındaki artış ve verinin yönetimi amacıyla dağıtık yapıda çalışabilecek yeni modeller üretme çabası sonucunda büyük veri ortaya çıkmıştır. Kaynakların etkin, verimli kullanılması ve uluslararası ortamda stratejik üstünlük elde edilmesine yönelik olarak kendi iş süreçleri ve kaynaklarından elde ettiği veriyi karar verme süreçlerinde ve sunduğu hizmetlerde kullanmak isteyen kamu kurumları için büyük veri önem arz etmektedir.

Kamu kurumları günden güne büyük veri teknolojisinin potansiyelinin farkına varmaktadır. Devletin farklı kurumları emeklilik ve harç ödemeleri yönetimi, günlük cari işlemler, vergi, sağlık sistemleri ve trafik verisi gibi çok farklı alanlarda büyük miktarlarda veri toplamaktadır (Munné, 2016: 195). Dünyanın farklı ülkelerinde kamu hizmetlerinin sunulmasında büyük veri teknolojisine dayanan yüzlerce proje geliştirilmektedir. Kanada'da kamu kurumlarında performans yönetimi projesi (Goh ve diğerleri, 2017: 29), Çin'deki halk otobüs taşımacılığı sistemleri çalışması ile otobüslerin mümkün olan en kısa sürede kullanımı ve hibrid otobüslerin zamanında çalışmasını sağlamak için sensör verisinin kullanımına yönelik büyük veri projesi (Rodriguez, 2017: 93) ve Hindistan'da tedarik zincirinde yapılması gerekenlere yönelik yürütülen büyük veri projesi (Varma ve Tyagi, 2017: 115) gibi örnekler vermek mümkündür. Sadece merkezi kamu hizmetlerinde değil, yerel yönetimler düzeyinde ciddi trafik kazalarının nerede ve ne zaman olacağını tahmin etmesini mümkün

kılabilecek faktörleri arařtırmak gibi alanlarda da büyük veri projeleri geliřtirilmektedir (Geddes ve Lai, 2017: 190).

Veri akıřı yapılandırılmıř halde olduđu gibi yapılandırılmamıř halde web sayfalarından, bloglardan, sosyal medya ve diđer kaynaklardan da sürekli řekilde devam etmektedir. Günümüzde sayısız miktarda sensör yoluyla endüstride kullanılan nesnelere veri akmaktadır. Sensörler bađlı olduđu nesnenin sıcaklık, konum bilgisi, nem ve hatta havadaki kimyasal deđiřim oranı olmak üzere farklı pek çok verisini ölçme kabiliyetine sahiptir. Örgütlerin elde ettiđi büyük veri üzerinde veriyi örnekleyerek analize hazırlamak öncelikli hedefdir (Ohlhorst, 2013: 84). Son yıllarda teknolojiye yeni yatırımlar ve geliřmeler otonom sistemler, sürücüsüz otomobiller, dronlar, yapay zekâ, nesnelere interneti, artırılmıř gerçeklik ve sanal gerçeklik gibi yenilikçi teknolojileri ortaya çıkarmıř ve küreselleřmenin de etkileriyle örgütler buna adapte olmaya bařlamıřtır. Bu yeni nesil teknolojiler, tüketici ve kullanıcı olarak insanların gündelik hayatını ve yařam tarzlarını deđiřtirdiđi gibi, örgütlere yeni fırsatlar sunarak onların da iřleyiřini dönüřtürmektedir.

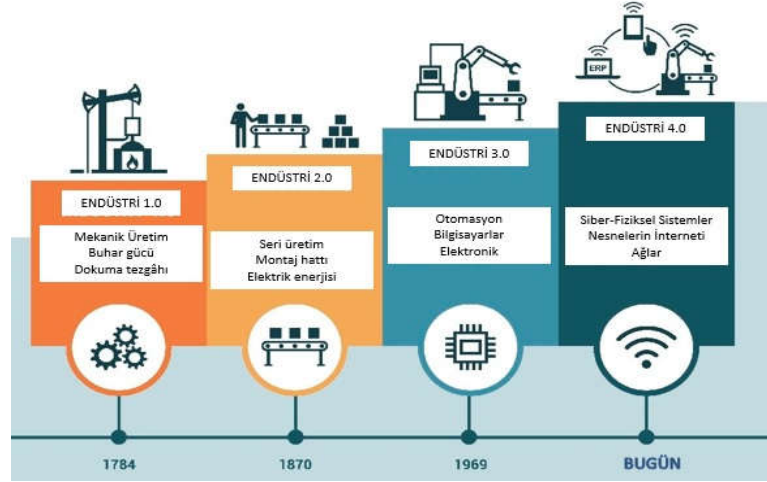
## **1.2. Büyük Veri ile İliřkili Bařlıca Biliřim Teknolojileri**

Biliřim teknolojilerinin kullanılmaya bařladıđı ilk günden bu yana üretilen veri miktarı altyapının geliřmesine paralel olarak artıř göstermiřtir. Veri miktarının artmasında, veri saklama ortamlarının kabiliyetlerinin artması ile birlikte internetin daha fazla kullanıcı tarafından kullanılmasının da payı büyüktür. Devlet kurumları ve özel sektörün elinde kullanıcılar tarafından üretilen verinin yanında kullanılmıř oldukları sistemlerden, sensörlerden ve diđer teknolojik bileřenlerden elde ettikleri iřlemler neticesinde oluřan büyük veri bulunmaktadır.

Biliřim teknolojilerinde sađlanan hızlı ilerlemeler neticesinde kullanılan yenilikçi teknolojiler günden güne çeřitlilik göstermektedir. Endüstri 4.0 devrimi olarak tanımlanan dördüncü sanayi devimi ve internet ađı kapasitesinin de geniřlemesiyle birlikte bulut biliřim, nesnelere interneti (IoT, Internet of Things), giyilebilir teknoloji, dronlar, ses, görüntü tanıma teknolojileri, artırılmıř gerçeklik, sanal gerçeklik ve yapay zekâ teknolojilerinin kullanım alanının artması geleneksel veriye oranla yüksek hacimli, sürekli artan verinin oluřmasını sađlamıřtır.

### 1.2.1. Endüstri 4.0

Buhar sistemlerinin kullanılmaya başlamasıyla gerçekleşen birinci sanayi devrimi, petrolün yaygın kullanımı ve 1920’lerde gelişigüzel yöntemlerin yerine bilimsel yöntemlerin kullanılmasını esas alan “Bilimsel Yönetim Teorisi”nin (Taylor, 2012: 109) ortaya çıkmasıyla üretim bandı sistemlerinin gelişimi sonrası üretim verimliliği artışının yaşandığı ikinci sanayi devrimi, elektrik, elektronik ve internet alanındaki gelişmelerle informatik devrim olarak adlandırılan üçüncü sanayi devriminden bu yana üç sanayi devrimi gerçekleşmiş olup günümüzde dördüncü sanayi devrimi olarak nitelendirilen “Endüstri 4.0” çağı yaşanmaktadır. Makine gücünün insan gücünün yerini almasıyla üretim süreçlerinin kendiliğinden yönetilebilir hale geldiği içinde bulunduğumuz çağ “Endüstri 4.0 devrimi” olarak tanımlanmaktadır (Bulut ve Akçacı, 2017: 50). Endüstri 4.0, makine ve ürünlerin insan kontrolü olmadan birbiriyle iletişime geçtiği akıllı üretim ağını temsil etmektedir (Ivanov ve diğerleri, 2016: 386). Endüstrinin geçmişten günümüze geçirmiş olduğu aşamalar Şekil 3’de görülmektedir.



Şekil 3: Endüstri Devrimleri

**Kaynak:** Cline, 2017: 1.

Tarım ve insan gücüne dayalı ekonomiden makine ve seri üretimin öne çıktığı ekonomiye geçişin yaşandığı insanlık tarihinin en önemli dönüm noktası olarak kabul edilen “Birinci Sanayi Devrimi”, İngiltere’den bütün dünyayı etkilemiştir (Dengiz, 2017: 38). İlk endüstriyel devrim olarak kabul edilen 1750-1890 yıllarında James

Watt'ın buluşu olan buhar motoru kullanılmaya başlanmıştır. Gemicilik ve demiryolları çelik üretiminin artışı ile gelişen sanayiler olmuştur (Ayvaz ve diğerleri, 2010: 2).

Endüstri 2.0, elektriğin kullanılması ve seri üretimin elektrik ile gerçekleştirilmesi ile başlamıştır. Bu dönem aynı zamanda Henry Ford'un araba üretimi için kitlesel üretimin benimsendiği "Fordist dönem"dir (Alçın, 2016: 3). 1960'lı yılların sonlarına kadar geçen bu sürede montaj hattında Taylor'un ilkeleri ve hareketli montaj hattı uygulanarak düşük maliyetli seri üretime geçilmiştir. Araba üretimi ile başlayana kitlesel üretim yöntemleri, üretim bantlarının gelişmesine katkıda bulunmuştur. Elektriğin fabrikalarda kullanılmasıyla rekabet stratejilerinin geliştirilmesi gerekli hale gelmiştir (Macit, 2017: 52).

Üçüncü sanayi devrimi, otomasyon bilgisayarlar ve elektronik altyapının etkin olarak kullanıldığı dönemdir. Dijital teknolojinin üretimde yer almasıyla birlikte programlanabilen cihazlar ve bilişim teknolojileri bu dönemde ilerleme kaydetmiştir. İnternetin hızla yayılması, bu dönemin informatik devrim olarak bilinmesi sağlamıştır (Çelikaş ve diğerleri, 2015: 1).

Dördüncü sanayi devrimi 20. yüzyılın sonlarında başlayarak 21. yüzyılda gelişimini sürdürmüştür. Endüstri 4.0 resmi olarak Almanya'da Hannover Fuarında duyurulmuştur (Özsoylu, 2017: 43). Talebin bireyselleşmesi, kaynakların verimli kullanılması ve kısa ürün geliştirme dönemleri, Web 2.0 teknolojisi, akıllı telefon, laptop ve 3D yazıcıların ortaya çıkmasıyla birlikte bu dönemde radikal değişimler yaşanmış, ekonomilerin gelişiminde büyük potansiyel meydan gelmiştir. Temel olarak Endüstri 4.0, üretim tesisleri, tedarik zincirleri ve servislerin entegrasyonu neticesinde katma değerli ağların kurulmasını sağlayan sistemlerdir. Endüstriyel internetin gelişmesi, kablosuz sistemler, sensör ağları, bulut sistemleri, gömülü sistemler ve özerk robotlar entegre, bilgisayar tabanlı bir ortam sağlamıştır. Her şeyden önce, tüm sistem veri analizini içermeli ve gerçek zamanlı karar verme için çeşitli koordinasyon araçları gereklidir (Salkin ve diğerleri, 2018: 4-6).

Endüstri 4.0 ile birlikte merkezi olmayan kararların verilmesi, fabrikalarda yapılan fiziksel işlemlerin siber fiziksel sistemlerle izlenmesi hedeflenmektedir. "Nesnelerin interneti" (IoT) kavramıyla internet daha etkin olarak kullanılarak nesnelere siber fiziksel sistemler üzerinden insanlarla gerçek zamanlı iletişime geçebilmektedir. Örneğin ev

aletleri, gözetleme kameraları, izleme sensörleri, ekranlar, araçlar vb. birçok cihaz/donanıma kolay erişim ve etkileşim sağlanarak vatandaşa hizmet sağlanması adına çeşitli uygulamalar dizayn edilmektedir. Özel sektör ve kamu kurumları bahsi geçen nesnelerin oluşturduğu potansiyel veriden yararlanmak suretiyle geliştirilen uygulamalardan yararlanarak vatandaşa yeni hizmetler sunma arayışı içerisinde (Zanella ve diğerleri, 2014: 22). Endüstri 4.0 devriminin etkisiyle örgütlerin işleyişinde meydana gelen değişimle hizmet sunumunda kullanılan nesnelerin birbiri ile olan iletişiminin artması ve yönetim bilişim sistemleri uygulamalarının birbiriyle entegre biçimde çalışmaya başlaması, kurumların sahip olduğu verinin kapasitesinin sürekli şekilde artmasını sağlamaktadır. Verideki artış kurumları, büyük veri yaklaşımıyla veriyi saklamaya ve veriden değerli bilgiyi elde edecek veri madenciliği teknikleri kullanmaya yöneltmektedir.

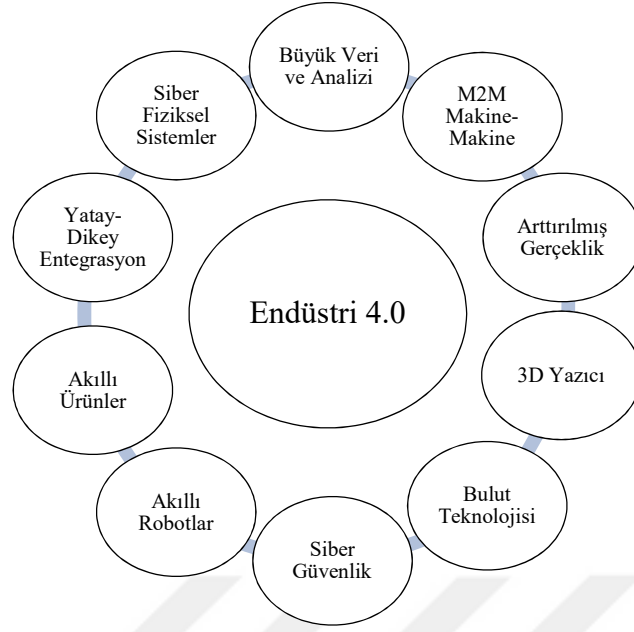
Endüstri 4.0'ın en önemli özelliği, yarı mamül ürünlerin üretim aşamasında makinelerle doğrudan iletişime geçmesi ve bir fabrikanın kendi kendini organize edebilmesi olarak karşımıza çıkmaktadır. Bu sayede ucuz ve verimli müşteri odaklı üretim mümkün olmaktadır. Bununla birlikte, Endüstri 4.0 vizyonu gerçek zamanlı kontrol, entegre bakım, daha iyi uyarlanabilirlik, tedarik zinciri boyunca artan işbirliği ve daha iyi izleme olanaklarına imkân vermektedir (Branke ve diğerleri, 2016: 264). Endüstri 4.0 yöntem ve süreçlerinin çeşitli kamu hizmetlerinde kullanılması, kamu kaynaklarının etkin ve verimli biçimde kullanılmasına katkı sağlamaktadır.

Endüstri 4.0 devrimi ile birlikte ihtiyaç duyulan nitelikli personel de çeşitlilik göstermektedir. Bu kapsamda programlama becerisi olan endüstriyel yazılım programcılarının olan ihtiyaç devam etmektedir. Çağımızda Endüstri 4.0 ile birlikte her cihaz veri üretmekte ve üretilen bu veriler kayıt altına alınmaktadır. Büyük verinin anlamlandırılması, tasnif edilmesi ve bilgi haline getirilmesi ihtiyacı bulunmaktadır. Ancak bu sayede birbirine bağlı cihazlardan elde edilen verinin bir anlamı olacak ve analiz edilecektir. Bu açıdan veri madenciliği, büyük veri, makine öğrenmesi ve yapay zekâ çözümleri için endüstriyel veri analizi uzmanlarına ihtiyaç bulunmaktadır. Akıllı şehir planlayıcıları ve üretim teknolojileri diğer ihtiyaç duyulan uzmanlık alanlarıdır (Sener ve Elevli, 2017: 31). Endüstri 4.0 ile birlikte oluşan büyük verinin karar

vericilere ve üretim süreçlerine katkı sağlamak üzere sunulması büyük önem taşımaktadır.

Sanayi devriminin dördüncü aşaması olan Endüstri 4.0, özel ve kamu sektöründe pek çok farklı alanda toplumun tüm katmanlarını etkileyecek bir süreçtir. Endüstri 4.0 bilişim altyapısının akıllı üretime geçişi neticesinde yeni iş modellerinin ortaya çıkmasını sağlayarak ürünlerin yapısından, tedarik ve satışına kadar süreçlerde ciddi değişikliklerin yaşanmasına neden olmaktadır (Özsoylu, 2017: 60). Belirli kamu hizmetlerinin sunulması kapsamında, fabrikaların üretim süreçlerinin yönetiminde ve bazı özel sektör hizmetlerinin yürütülmesinde Endüstri 4.0 uygulamalarını görmek mümkündür. Makine sensörlerinin entegrasyonu, yazılımlar, bulut bilişim ve depolama sistemleri üzerinden büyük veri kümeleri sorgulanabilmekte; elde edilen sonuçlar firmaların operasyonel süreçlerini yürütmek için bir yöntem olarak kullanabilmektedir (Gilchrist, 2016: 4).

Endüstri 4.0 ve onunla ilişkili teknolojik gelişmelerin geldiği noktada kamu kurumları da bu süreçten etkilenmektedir. Şekil 4'te görüleceği üzere büyük veri analizi, akıllı fabrikalar, yapay zekâ, robotlar, siber güvenlik, nesnelerin interneti, bulut sistemleri, ek üretim ve artırılmış gerçeklik gibi teknolojiler başarılı bir uyum için gerekli teknolojilerdir. Günümüzde bu alandaki araştırmalar verinin toplanması, öğrenilmesi ve örneklenmesine odaklanmaktadır (Kleineidam ve diğerleri, 2016: 339). Endüstri 4.0 unsurlarının sahip olduğu sensörler sayesinde oluşan büyük veri üzerinden anlamlı bilgilerin elde edilmesi için büyük verinin analizi, veri üzerinde veri madenciliği tekniklerinin uygulanması ve makine öğrenmesi algoritmalarının çalıştırılması gerekmektedir. Elde tutulan veri, değerli bilgiye dönüştürülmediği sürece kuruma yük olmaktan öteye geçemeyecektir.



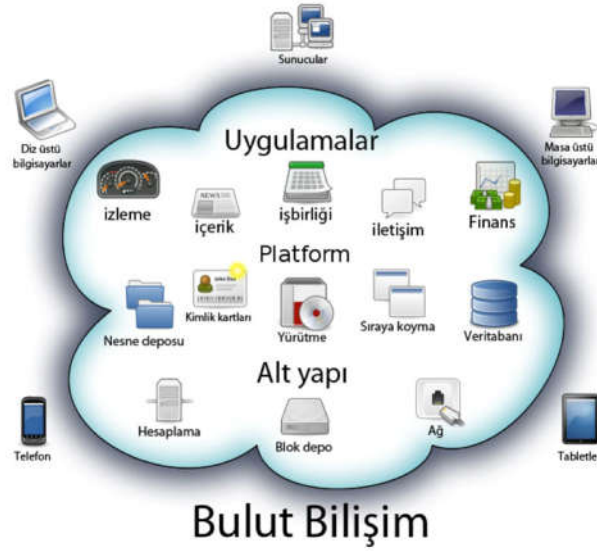
**Şekil 4:** Endüstri 4.0 Devriminin Unsurları

**Kaynak:** Fırat ve Fırat, 2017: 214.

Endüstri 4.0 devriminin dayandığı teknolojik yenilikler ulusal ve yerel düzeyde sunulan kamu hizmetlerine entegre edilmektedir. Örneğin, akıllı ürünlerden akıllı şehirlik yaklaşımında etkin bir şekilde faydalanılması vatandaşa sunulan hizmetlerin kalitesini artırmanın yanı sıra, mali açıdan kamunun tasarruf sağlamasında da etkili olmaktadır. Farklı düzeylerde gerçekleştirilen faaliyetler neticesinde meydana gelen kamu verisinin, Endüstri 4.0 sürecinde büyük veri ve ve madenciligi teknikleri yoluyla geleceğe dönük karar verme ve tahminleme işlemleri için kullanılması da mümkündür.

### 1.2.2. Bulut Bilişim

Bulut bilişim temel olarak, kendi verinizin dışarıda bir veri merkezinde barındırıldığı kullanımı kolay istenildiğinde paylaşılabilirdiği ve en önemli husus olarak internet üzerinden Şekil 5'te görüleceği üzere her türlü donanım (bilgisayar, tablet, cep telefonu vb.) ile erişilebildiği teknolojik bir hizmettir (Marks ve Lozano, 2010: 27). Kurumlar, örgütler ve toplumun ürettiği verinin büyük miktarlara ulaşması neticesinde meydana gelen büyük veriye erişimin sağlanması, gelişen internet ağları üzerinden bulut bilişim teknolojisi ile mümkün hale gelmiş, üretilen büyük veri endüstri ve örgütler tarafından kullanılabilir kılınmıştır.



**Şekil 5:** Bulut Bilişim Uygulamaları ve Altyapısı

**Kaynak:** Wikipedia, 2018: 1

Bulut fikri, başlarda internet üzerinden bulutun diğer tarafına verinin taşınması şeklinde bir metafor olarak doğmuştur. Bu konsept 1960’larda John McCarthy tarafından uygulamaların hizmet tipi bir iş modeli aracılığıyla satılabileceği şeklinde tanımlanmıştır. Amazon şirketi, bulut bilişimin yaygınlaşmasında önemli bir rol üstlenmiştir. 2001 yılında yeni bulut altyapısı hizmeti ile veri merkezine son kullanıcıların erişimini sağlamıştır. 2002 yılına gelindiğinde “Amazon Web Services” hizmetini duyurarak gerçek bulut devrimini başlatmıştır. Amazon Web Services modeli bilişim hizmetlerinin dünyanın herhangi bir yerindeki kullanıcıya, herhangi bir yerinden kiralanarak hizmet verilmesini esas almaktadır (Rittinghouse ve Ransome, 2009: 17).

Bulut bilişim servisleri, hizmet sağlayıcının sunduğu hizmetlerin istenildiği zaman yeniden kullanılabilmesi anlayışına dayanmaktadır. Servis sağlayıcıların sunmuş olduğu bulut hizmetlerinin genel özelliklerini küçük ölçekli işletmelerin (günümüzde bireysel kullanıcılar için) bile hizmetlerden düşük ücretlerle yararlanılabilmesi (Low barriers to entry), geniş ölçeklenebilirlik (Large scalability), kaynakların birçok kullanıcı tarafından paylaşılabilmesini sağlayan çoklu ortam özelliği (Multitenancy), platform ve donanım bağımsız olarak hizmetlere erişilebilmesi (Device independence) şeklinde sıralamak mümkündür (Velte ve diğerleri 2010: 11).

Bulut teknolojisi ile sunulan hizmetleri; SaaS (Software as a Service) yazılımı servis olarak sunma, PaaS (Platform as a Service) platform hizmeti ve IaaS (Infrastructure as a Service) sunucu altyapı hizmeti olmak üzere üç temel hizmet halinde ayırmak mümkündür. Hizmetlerin içeriği ve sağladığı avantajlar Tablo 1’de görülmektedir;

**Tablo 1**  
**Bulut Bilişim Üzerinde Sunulan Hizmetler**

Servis Adı	Açıklama	Avantajlar
<b>Software as a Service (SaaS)</b>	<ul style="list-style-type: none"> <li>• Uygulamaların, internet üzerinden bir hizmet olarak kullanıcılara sunulmasıdır.</li> <li>• Uygulama servis sağlayıcı tarafından kullanıcının sorumluluğu dışında barındırıldığı için kullanıcı uygulamayı korumak (bakım, yedekleme vb.) veya desteklemek (güncellemek, virüs koruması vb.) zorunda değildir.</li> <li>• Kullanıcılar bu hizmeti kullanmak suretiyle uygulamalarını diğer sistemlere entegre etmeye ihtiyaç duymamaktadır.</li> </ul>	<ul style="list-style-type: none"> <li>• Kullanıcıların bir donanım (bilgisayar, tablet vb.) üzerinden hizmete erişebilmesi,</li> <li>• Hizmetlerin servis sağlayıcılar tarafından sağlanması sayesinde kurumda fazla sayıda bilişim personeli çalıştırma ihtiyacının azalmasına paralel olarak maaş, sigorta vb. giderlerin azaltılması, <ul style="list-style-type: none"> <li>• Eski uygulamaların özelleştirilmesindeki zorlukların kolaylaşması,</li> </ul> </li> <li>• Geliştirilen uygulamanın servis üzerinden tüm dünyada kullanılabilir hale gelmesi,</li> <li>• Uygulamaların servis sağlayıcısının sorumluluğunda merkezi olarak güncellenmesi,</li> <li>• Güvenli Ortam Katmanı (Secure Socket Layer SSL) sayesinde kurumların güvenlik için ayrı bir işletme maliyetinden kurtulması ve özel ağlar (VPN) üzerinden uygulamalara erişim,</li> <li>• Son dönemde artan bant genişliği sayesinde düşük gecikmelerle hizmetlere daha hızlı erişim.</li> </ul>
<b>Platform as a Service (PaaS)</b>	<ul style="list-style-type: none"> <li>• Uygulamaların ve hizmetlerin oluşturulması için gerekli tüm kaynakların servis sağlayıcı tarafından sunulmasıdır.</li> <li>• PaaS hizmetleri, uygulama tasarımı, geliştirme, test etme ve barındırmayı içermektedir.</li> </ul>	<ul style="list-style-type: none"> <li>• Coğrafi olarak birbirinden ayrı ekiplerin birlikte çalışabilme kabiliyeti,</li> <li>• Sunulan web hizmetlerinin birleştirilebilmesi,</li> <li>• Yerleşik altyapı hizmetleri kullanılarak maliyet tasarrufu olanağı ile güvenlik ve ölçeklenebilirlik sağlamak zorunda kalınmaması,</li> <li>• Web hizmetlerinin entegre edilmesi, veritabanı entegrasyonu, güvenlik,</li> </ul>

		ölçeklenebilirlik, depolama ve sürüm oluşturulması.
<b>Infrastructure as a Service (IaaS)</b>	<ul style="list-style-type: none"> <li>Sunucu alanı, ağ bileşenleri, bellek, işlemci ve depolama alanı gibi fiziksel donanımların hizmet olarak servis sağlayıcıdan kiralanmasıdır.</li> </ul>	<ul style="list-style-type: none"> <li>Kaynakların ihtiyaç durumuna göre artırılıp azaltılabilmesi,</li> <li>Aynı kaynaklar üzerinde aynı anda birden fazla kiracıya hizmet sunulabilmesi,</li> <li>Kaynakların kullanıldığı kadarının kiracıya fatura edilmesi sayesinde alınan hizmet kadar ücret ödenmesi,</li> <li>Donanım işletme maliyetinin düşmesi.</li> </ul>

**Kaynak:** (Velte ve diğerleri 2010: 11-16)'den tablolaştırılmıştır.

Bulut bilişim, hizmetlerin sağlamış olduğu ya da sağlamayı vaad ettiği avantajların yanında kullanılan internet bağlantısının bant genişliğinden, kişisel verinin korunmasına ve ülkelerin sunmuş olduğu hukuki zeminine kadar birçok dezavantajı da beraberinde getirmektedir (Henkoğlu ve Külcü, 2013: 67-72). Uygulamaların yavaş çalışması ve düşük hızlarda servis sorunları bulut hizmetlerinin kullanımını olumsuz etkilemektedir. Kullanıcılara ait kişisel verinin saklandığı merkezi hizmet sağlayıcıları, uzaktan erişim ve güvenlik sorunlarını çözmek ve veri koruma yükümlülüğünü yerine getirmek için yetkisiz erişimlere karşı önlem almaya zorlanmaktadır.

Bulut hizmeti üzerindeki verinin gizliliği, bütünlüğü ve kullanılabilirliği ile ilgili yasal hakların takip edilmesi ve olası zararların karşılanması için, hizmet sağlayan sunucuların belirli bir bölge içinde bulunması istenmelidir. Aksi takdirde verinin nerede olduğunu bilmeme sorunu ile karşı karşıya kalınmaktadır. Bu konu Avrupa Birliği (AB) başta olmak üzere farklı ülkelerde yasal olarak düzenlenmektedir.

Kişisel verinin gizliliği hususu büyük önem arz etmekte ve her geçen gün daha fazla risk oluşturmaktadır. Kişisel veri ya da kurum verisinin hizmet sağlayıcısının kusuru nedeniyle açığa çıkması büyük problemlere neden olmaktadır. Bu durum hizmet alınan firmaların güvenilirliği, yeterliliği ve denetlenmesi sorunlarını da beraberinde getirmektedir.

Hizmet sağlayıcıların bilgi güvenliği, veri bütünlüğü ve erişim denetimi ile ilgili taahhütte bulunamamaları ve bu konudaki tüm sorumluluğu kullanıcıya bırakması büyük verinin dezavantajlı olduğu diğer bir husustur. Hizmet sağlayıcıların kesintisiz hizmet garantisi verememeleri kullanıcıların/kurumların verisine erişemez duruma

gelmesine ve bazen iş akışının durmasına neden olmaktadır: Servis sağlayıcıları tarafından hazırlanan sözleşmelerde kullanıcı/kurum verisine sadece kendisinin erişebileceğine dair ayrıntı bulunmamaktadır. Bu durum içeriğin kullanımı ve mülkiyet hakkı ile ilgili belirsizliklerin oluşmasına sebep olmaktadır. Siber saldırıların arttığı günümüzde bulut üzerinde tutulan veriler saldırganların hedefi haline gelmiştir. Adli incelemelerin ve dijital delillerin elde edilmesi konusunda belirsizlikler bulunmaktadır.

Bulut bilişim geliştirme modelleri; genel bulut (public cloud), özel bulut (private cloud), melez bulut (hybrid cloud) ve topluluk bulut (community cloud) olmak üzere dörde ayrılmaktadır (bkz. Tablo 2).

**Tablo 2**  
**Bulut Bilişim Modelleri**

<b>Model Adı</b>	<b>Açıklama</b>
<b>Genel Bulut (Public Cloud)</b>	İnternet üzerinden web ara yüzü aracılığıyla genel kullanıma sunulan, hizmet alındığı kadar ödeme yapılan elektronik posta uygulaması gibi hizmetlerdir (Google Apps, Amazon, Windows Azure vb.).
<b>Özel Bulut (Private Cloud)</b>	Kurum ya da kuruluşa özel sunulan hizmetlerin bulut üzerinden verilmesidir. Hizmet sağlayıcı kurumun kendisi olabileceği gibi, üçüncü bir hizmet sağlayıcıdan destek de sağlanabilir. Özel bulut modelinde kurum dışından buluta erişim kapatılarak kurum için hizmet sağlanır. Veri güvenliğine önem veren özel kurumlar ve özellikle de kamu kurumları tarafından tercih edilen modeldir.
<b>Melez Bulut (Hybrid Cloud)</b>	Kurumun bazı hizmetleri genel bulut üzerinden kullanıcıların hizmetine sunulurken, bazı hizmetleri özel bulut üzerinden kendi çalışanlarına kapalı ağ üzerinden sunulmaktadır (IBM, Juniper gibi).
<b>Topluluk Bulut (Community Cloud)</b>	Sunulan hizmetin ortak çalışma gerçekleştiren topluluk ya da gruba özel olmasıdır.

**Kaynak:** (Mell ve Grance 2011: 3)'den tablolaştırılmıştır.

Bulut bilişimin kullanımının geçtiğimiz on yıllık süre içerisinde giderek artması ile birlikte özel sektörden kamu kurumlarına kadar birçok kurum ve kuruluş bulut yatırımlarına ağırlık vermeye başlamıştır. Bu kapsamda, bulut hizmeti veren Amazon, Google, Microsoft ve Oracle başta olmak üzere birçok şirket bulunmaktadır.

Günümüzde özel sektör başta olmak üzere şirketler, sivil toplum kuruluşları ve kamu kurumları kendisine farklı kaynaklardan gelen finans, mobil, sağlık, işlem, müşteri araştırma ve sosyal medya verisi gibi birçok veriyi bulutta saklamaktadır. Bulut bilişimde yaşanan gelişmeler neticesinde çevrimiçi kaynaklara anında erişilebilmesi, büyük veri üzerinde analizlerin gerçekleştirilmesini ve örgütlerin kullanımı için değere

dönüştürülmesini kolaylaştırmıştır (Çelik ve Akdamar, 2018: 262). Bulut bilişim teknolojilerinin kullanımının yaygınlaşmasıyla birlikte örgütler mevcut kaynaklar ve donanım kapasitesi ile sınırlı şekilde hizmet sunmaktan ziyade, ihtiyacı olan kapasiteyi anlık olarak artırma şansına kavuşmaktadır. Bu sayede hizmet kalitesi artmakta, buna paralel olarak ise verinin kapasitesi artarak büyük veri oluşmaktadır.

### **1.2.3. Nesnelerin İnterneti**

İnternetin yaygınlaşması ve toplumun her kesimi tarafından aktif olarak kullanılması, birbirleriyle iletişim ve etkileşim halinde olan donanımların sayısının hızla artması “nesnelerin interneti” (Internet of Things, IoT) kavramını ortaya çıkarmıştır. Nesnelerin interneti temelli servisler, sadece üretim dünyasında değil aynı zamanda ofiste, evde, sokakta kısacası her yerde insanlar ve birbirine bağlı objelerin çeşitli görevlerini daha fazla otomasyona bağlayarak daha akıllı bir dünya yaratmayı hedeflemektedir (Chaouchi, 2010: 2).

Nesnelerin interneti, basit olarak donanımların internete bağlanmasının ötesinde, radyo frekans (RFID) üzerinden bazı cihazlar ile veri üretilmesini de kapsamaktadır. Nesnelerin, algılayıcılar ve elektronik devreler ile donatılması sonucunda, “düşünme” ve “konuşma” gibi duyular üzerinden veri üretmeye başlayarak insanlarla iletişime geçebileceği düşünülmektedir (Kutup, 2011: 154). Sadece kişiler üzerinden değil, aynı zamanda aşağıda daha ayrıntılı olarak incelenecek olan akıllı kentlerde ve otomobiller, su sayaçları, sokak lambaları, ses ve görüntü tanıma gibi yaşamın her alanında veri toplanmaktadır.

Özel sektörde tüketici davranışlarını belirlemek üzere tutulan veriler, akıllı kentlerde sensörler üzerinden toplanan veriler, her sektörde kullanımı artan multimedya içeriği ve sosyal medyanın artan kullanımıyla birlikte üretilen verinin yönetilmesi ve analiz edilmesi ihtiyacı doğrultusunda yeni teknolojilere olan ihtiyaç artmıştır. Nesnelerin interneti kavramı ile akıllı teknolojiler temeline dayanan donanımların dış dünyadan aldıkları veri birbirleriyle paylaşabilmektedir. Nesnelerin interneti uygulamaları kullanımının artmasıyla büyük veri oluşumuna katkı sağlanmaktadır.

İnternet teknolojileri, nesnelerin interneti ve kullanılan akıllı teknolojiler yardımıyla üretilen verinin analiz edilmesinde yararlanılan algoritma ve analizlerin toplumun

gündelik hayatını daha verimli, güvenli ve sağlıklı sürdürmesi için kullanılması amaçlanmaktadır. Nesnelerin interneti fiziksel olarak kullanılan nesnelere internetle, özellikle web servisleriyle entegre etmeyi amaçlamaktadır. Nesnelerin interneti yaklaşımında kullanılan nesnelere birbiri ile internet üzerinden iletişime geçtikçe üretilen veri büyük hızla artmakta ve işlemlerin neticesinde oluşan büyük verinin saklanacağı veritabanı kritik önem kazanmaktadır. Nesnelerin interneti sistemleriyle üretilen büyük verinin esnek ve geçerli veritabanı sistemlerinde depolanması için NoSQL veritabanları tercih edilmektedir (Kalay, 2018: 51). Nesnelerin interneti üzerinden üretilen verinin heterojen ve esnek biçimde ele alınması gerekmektedir. Günümüzde Google, Amazon, Facebook gibi büyük ölçekli teknoloji şirketleri nesnelerin interneti teknolojisi ile üretilen büyük veriyi NoSQL veritabanlarında tutmaktadır.

Nesnelerin interneti şemsiye kavram olarak düşünüldüğünde giyilebilir teknolojik ürünler, dronlar, ses ve görüntü tanıma teknolojileri, artırılmış gerçeklik ve sanal gerçeklik uygulamaları üzerinden sürekli veri üretimi gerçekleştirilmekte, nesnelere üzerinden akan veri “büyük veri” olarak tutulmaktadır. Elde edilen büyük verinin analizi geleneksel yöntemlerle mümkün olmamaktadır. Veri madenciliği teknikleri kullanılabilmesi gibi, büyük veriye özel analiz teknikleri de kullanılabilir.

Giyilebilir teknoloji, farklı aksesuarlar halinde insan bedenine uyum sağlamak suretiyle telefon, tablet ve bilgisayar gibi donanımlarla kablosuz bağlantı kurmak suretiyle veri aktarımı yapabilen araçlardır. Giyilebilir teknolojiler üzerinden toplanan veri ile sağlık, spor, güvenlik ve eğlence alanlarında kullanılmaktadır (Doğan ve diğerleri, 2017: 26). Akıllı saat ve akıllı bileklik tarzındaki ürünler olarak kullanılmaya başlanan giyilebilir teknolojiler, henüz tam anlamıyla yaygınlık kazanmasa da kişinin sağlık verisi, sosyal medya takibi, günlük faaliyetlerinin organizasyonu gibi pek çok farklı veriyi toplayabilmektedir. Firmalar, bu sayede toplanan büyük veri üzerinde gerçekleştireceği analizler yardımıyla alternatif tanıtım, kişinin bulunduğu coğrafi özelliklere uygun pazarlama ve kişiye öze satış yöntemleri üretmektedir. Giyilebilir teknolojik ürünlerin kullanımının yaygınlaşmasıyla birlikte akan verinin miktarında yaşanacak artış, büyük veri teknolojilerinin etkin olarak kullanılması ihtiyacını da beraberinde getirecektir.

Geçmiş uzun yıllar öncesine dayanmakla birlikte dronlar son yıllarda yapay zekanın gelişimi, imaj işleme ve robot teknolojisindeki yeniliklere koşut olarak özel sektörün

ilgisini daha fazla çekmeye başlamıştır. Dronlar 1900'lerin başında askeri hedeflere yönelik keşif amaçlı olarak Amerika Birleşik Devletleri (ABD) tarafından kullanılmaya başlanmıştır. Dronların özel sektör ve kamu kurumları tarafından askeri alan dışında kullanımı son on yılda artmıştır. Elektronik bileşenlerin, kameraların minyatürleşmesi, mikroişlemcilerin veri işleme gücünün artması, veri depolama kapasitelerinin artması dronların fiyatlarının düşmesine ve özel firmalar, endüstri ve bireysel kullanıcılar tarafından kullanılmasında etkili olmuştur. Dronlar, bireysel kullanıcılar tarafından oyun amaçlı tercih edilirken, özel sektör tarafından fotoğraf ve film çekimi, coğrafi haritalama işlemleri ve tarım gibi alanlarda kullanılmaktadır. Endüstride ise gözlem/keşif/ilaçlama işlemleri, kargo ve lojistik firmaları tarafından küçük paket teslimatlarına yönelik profesyonel amaçlarla yararlanılmaktadır. Örneğin Amazon şirketi 2016 yılında kargo teslimatı için insansız hava aracı olarak dronu kullanmış olup, bu teknolojiye yatırım yapmaya devam etmektedir (Giones ve Brem, 2017: 876-881).

Doğal afet izleme, sınır gözetimi, acil durum yardımı, arama ve kurtarma faaliyetleri son zamanlarda dronlar ile gerçekleştirilmektedir (Yanmaz ve diğerleri, 2018: 1) Küresel dron pazarının 2016'da 2 milyar dolardan, 2020'de 127 milyar dolara çıkacağı tahmin edilmektedir (Moskwa, 2016: 1). Kamu kurumlarında milli savunma sistemlerinde, sınır güvenliği ve terörle mücadele kapsamında kullanılan dron sistemleri, emniyet teşkilatı tarafından trafik güvenliği ve faaliyetlerin havadan kayıt altına alınması ve ayrıca tarım alanlarının ilaçlanması, tohumların ekilmesi gibi pek çok farklı amaçla yararlanılmaktadır.

İnternet kullanımının yaygınlaşmasına paralel olarak kullanıcılar tarafından üretilen ses ve görüntülerin artması neticesinde oluşan büyük veri yığınından gerekli analizlerin yapılarak yararlı bilgilerin çıkarılmasında "doğal dil işleme" (Natural Language Processing, NLP) teknolojileri önem kazanmıştır. Özel sektör açısından bakıldığında, ürün fiyatlaması, rekabetçi zekâ, pazar tahmini, risk algılaması ve bankacılık sistemleri için NLP teknolojilerine ihtiyaç duyulmaktadır. Diğer taraftan sosyal medyada üretilen büyük veri haline gelmiş içerik de bu anlamda önem taşımaktadır (Sun ve diğerleri, 2017: 10). Ses ve görüntü tanıma sistemleri sayesinde elde edilen veri yardımıyla büyük veri analizi ve veri madenciliği teknikleri kullanılarak kişilerin cinsiyeti, yaşı gibi

özelliklerine göre kişiye özel kampanyalar düzenlenebilmekte, reklam panolarında kişiye özel reklamlar gösterilebilmektedir.

Bilgisayar ortamında geliştirilen üç boyutlu resim ve animasyonları teknolojik imkanlar kullanmak suretiyle insanlara gerçek ortamda bulunma hissini vermesi ve aynı zamanda oluşturulan ortamdaki nesnelere etkileşimde bulunmalarını sağlayan teknoloji “sanal gerçeklik” (virtual reality) olarak tanımlanmaktadır (Kayabaşı: 2005: 152). Gerçek fiziksel ortam ve sanal ortamın bilgisayar üzerinde sıralı halde gerçek zamanlı olarak hissetmemizi sağlayan geleceğin teknolojisi ise “artırılmış gerçeklik” (augmented reality) olarak tanımlanmaktadır (Raja ve Calvo, 2017: 58). Yaklaşık 50 yıldır akademik araştırma konusu olan artırılmış gerçeklik alanında son on yılda büyük ilerlemeler kaydedilmiş, iş dünyası ve endüstride uygulanmaya başlanmıştır (Palmarini ve diğerleri, 2018: 215).

Sanal gerçeklik ve artırılmış gerçeklik tanımları birbirine benzemekle birlikte farklı iki teknoloji olarak karşımıza çıkmaktadır. Sanal gerçeklik teknolojisinde sanal bir oluşumun sanki içindeymiş gibi hissetmemizi sağlayan nesnelere bütünü varken, artırılmış gerçeklik teknolojisinde gerçek dünya ve bilgisayar ürünü sanal dünya bir araya getirilerek gerçek dünyadan kopmadan sanal dünya ile etkileşime geçilmesi sağlanmaktadır (Ürey ve Akşit, 2018: 1). Sanal gerçeklik ve artırılmış gerçeklik arasındaki fark bahse konu teknolojilerin kullanılmasını sağlayan gözlüklerde ortaya çıkmaktadır. Sanal gerçeklik gözlükleri kullanıcının gerçek dünya ile bağlantısını tamamen keserek bilgisayar ortamında sağlanan görüntüyü gerçek dünyanın konum bilgilerinden yararlanmak suretiyle kullanıcıya sunarken, artırılmış gerçeklik gözlükleri gerçek dünyadaki nesnelere ile iletişim kurmaktadır. Artırılmış gerçeklikte gerçek dünya üzerine sanal bir katman eklenerek gerçeklik zenginleştirilmektedir (Yağcı, 2016: 1). Artırılmış gerçeklik ve sanal gerçeklik (virtual reality) eğitim, eğlence, tasarım ve mobil uygulama gibi farklı alanlarda kullanılmak üzere özel sektörün ilgisini çekmektedir. Örneğin, emlak sektöründe faaliyet gösteren firmalar ürünlerini pazarlamak ve tanıtmak için artırılmış gerçeklikten yararlanmaktadır. Bahse konu teknolojilerin kullanımı için kullanılan gözlükler ve sensörlerin gerçek dünya ile iletişime geçmeleri neticesinde üretilen veri miktarı büyük verinin meydana gelmesinde etkili olmaktadır.

#### 1.2.4. Yapay Zekâ

Yapay zekâ (Artificial Intelligence, AI), bilgisayarların minimum insan müdahalesi ile akıllı davranış gösterebilmesi için modellenmesi olarak tanımlanabilir. Yapay zekâ teknolojileri genel olarak robotların icadıyla birlikte başlamıştır (Hamet ve Tremblay, 2017: 36). Yapay zekâ teknolojisi olarak “makine öğrenmesi”, problemin probleme ait veriye göre modellenmesidir. Mevcut veri kümesi ile kullanılan algoritmalar yardımıyla oluşturulan model en yüksek performans için tasarlanmaktadır (Atalay ve Çelik, 2017: 161). Alanında önde gelen büyük bilişim şirketlerinin yapay zekaya önem göstermesi, yakın gelecekte yapay zekâya dayalı olarak farklı uygulamaların hizmete sunulmaya devam edileceğinin bir göstergesidir. Endüstri 4.0 devrimi için önemli bir faktör olan yapay zekâ sayesinde insanlar tarafından gerçekleştirilen bazı işlemler yapay zekâya sahip robotlar tarafından gerçekleştirilmektedir.

Öğrenen ve kendisini geliştiren yapay zekânın üzerinde çalışmalar yapmak ve en iyi yapay zekâ uygulamalarını geliştirmek için Apple, Google, Facebook, Microsoft ve IBM şirketleri rekabet etmektedir (Tilley, 2017: 1). Microsoft tarafından geliştirilen “Zo” isimli yapay zekâ, sosyal medyanın arka planında yer alan büyük veri kullanılarak tasarlanmıştır. Duygusal ve akıllıca yanıt verebilmek amacıyla duygusal ifadelerle birlikte insan etkileşimini ön planda tutmaktadır (Microsoft, 2017). Google, diyabetik göz hastalığının saptanmasında derin öğrenme kullanmakta (Peng ve Gulshan, 2016), Facebook insanla iletişim kurmanın daha iyi yollarını bulmak için veriden bilgi elde etme üzerine yapay zekâ araştırmalarına devam etmektedir (Facebook Research, 2018).

Yapay zekânın olumlu yanlarının yanı sıra, beraberinde getirdiği riskler de bulunmaktadır. Yapay zekânının kullandığı makine öğrenmesi algoritmaları, öğrenme işlemini gerçekleştirmek amacıyla büyük veri altyapısını kullanmaktadır. Microsoft geliştirdiği yapay zekâ projesinde “gündelik ve eğlenceli sohbet” aracılığıyla “insanları meşgul etmek ve eğlendirmek” için Twitter kullanıcılarının attığı tweetlerden oluşan kamu verisini kullanan “Tay” isimli bot geliştirmiştir. Fakat kullanıcıların attığı tweetler sebebiyle kullanıma açılmasının üzerinden yirmi dört saat geçmeden nefret dolu ve ırkçı ifadeler kullanması sebebiyle kapatılmıştır (Victor, 2016). Büyük veri üzerinde çalıştırılan yapay zekaya ait algoritmalar kullanıcının geçmişini analiz ederek geleceğini yönlendirmeye çalışabilmekte, kullanıcıya e-posta içeriğinden yola çıkmak suretiyle

tercihine göre reklam gösterebilmektedir. Yapay zekânın bu şekilde kullanılması insanın iradesini özgür biçimde kullanamamasını, yapay zekâ tarafından yönlendirilme olasılığını ortaya çıkarmaktadır.

Yapay zekanın, büyük veri desteği ve makine öğrenmesi ile getireceği katkılar ve ortaya çıkaracağı dezavantajlar düşünüldüğünde bu teknolojiye ilişkin standartların belirlenmesi gelecekte önlemez hataların engellenmesini sağlayabilecektir. Bu kapsamda Google yapay zekâ ilkelerini açıklamıştır. Google'ın yayınlamış olduğu yapay zekâ ilkelerine göre yapay zekâ projeleri sağlık, güvenlik, enerji, ulaşım, üretim ve eğlence gibi çok çeşitli alanlarda dönüştürücü etkilere sahip olacağı düşünüldüğünde sosyal açıdan faydalı olmalıdır. Yapay zekayı kullanarak yüksek kalite ve doğrulukta bilgiye erişmeye çalışılırken ülkelerin kültürel, sosyal ve hukuki standartlarına saygı gösterilecektir. Yapay zekâ uygulamalarında önyargılar yaratılmaması veya mevcut olanların güçlendirilmemesi, ırk, etnik köken, cinsiyet, milliyet, gelir, cinsel yönelim, yetenek ve siyasi veya dini inanç gibi hassas konularda insanlar üzerinde haksız etkiler oluşturulmamasına dikkat edilecektir. Yapay zekâ projelerinin büyük veri üzerinde test edilmeden önce güvenlik amacıyla sınırlı test ortamlarında geliştirilmesi sağlanacaktır. İnsanlara hesap verebilen, mahremiyet ilkesine önem veren yapıda yapay zekâ uygulamalarının geliştirilmesi ve multidisipliner yaklaşımla farklı pekçok bilim dalını ilgilendiren paydaşlarla çalışılması sağlanacaktır (Pichai, 2018). Google yapay zekâ ilkeleri doğrultusunda büyük verinin öneminin ve kullanım alanlarının günden güne arttığı, yapay zekâda kullanılan makine öğrenmesi algoritmaları ile geleceğe katkıda bulunmak adına birçok fırsatı içinde barındırdığı görülmektedir. Bu doğrultuda, yapay zekanın ve büyük verinin bir arada sunduğu avantajlardan yararlanırken getirmiş olduğu riskleri de iyi yönetmenin önemi ortaya çıkmaktadır.

### **1.3. Veri Madenciliği**

Bankacılık işlemleri, tıbbi tetkikler, ulaşım, ekonomik faaliyetler, uydular ve bilimsel araştırmalar gibi farklı kaynaklardan elde edilen veriler kurumlar tarafından çok çeşitli amaçlarla depolanmaktadır. Depolanan bu verinin analiz edilmesi için kullanılan yöntemler ve araçlar, artan veri miktarı karşısında yetersiz kaldığından istatistik, makine öğrenmesi, veritabanı teknolojileri ve ilgili diğer disiplinlerdeki teknikleri bir araya getiren veri madenciliği doğmuştur.

Veri madenciliğinin en önemli özelliği farklı disiplinleri bir araya getirmesi ve veri analizinin gerçekleştirilerek kurumların stratejileri doğrultusunda kullanılmasının sağlanmasıdır. Veri yığınları içinden faydalı bilginin elde edilmesi ve geleceğe dönük tahminlerin gerçekleştirilmesi veri madenciliği teknikleri ile mümkün olmaktadır (Çakır, 2008: 6). Verideki yapısal örüntüleri bulmaya çalışmak, veriyi açıklayacak algoritmayı belirleyerek tahmin ve sınıflandırma işlemlerinde kullanmak veri madenciliğinin temel amacıdır.

Özel sektörde yer alan kurumlar birbirlerine üstünlük sağlamak için işletme yönelimlerini (Business Trends) çağımızın bilişim altyapısına göre güncellemektedir. Günümüzün rekabetçi pazar yapısında, birbiriyle rekabet halindeki şirketlerin veri madenciliği uygulamalarından faydalanması verideki artış, iş süreçleri ve organizasyonların yeniden yapılanması, hizmetlerin verilme süresi, küreselleşme ve örgütsel mimarideki değişim nedeniyle zorunlu hale gelmiştir (Kleissner, 1998: 295). Verideki büyük artış neticesinde işletmelerin birbiriyle rekabet edebilmesi ve karar vericilerin bu veri yığınları içerisinde istedikleri bilgiye erişmeleri daha zor hale gelmiştir. Veri madenciliği, büyük veritabanları ve veri ambarlarında bulunan verinin içerisindeki gizli bilgiye erişimi kolaylaştırmaktadır. Kurumlar iş süreçlerini ve organizasyonlarını bilgi toplumunun gereklerine göre yeniden tasarlamaktadır. İş anlayışındaki değişim kurumların daha sade ve yalın bir yapıya kavuşmasını sağlamış, çalışanlar iş süreçlerinin geliştirilmesine ve iyileştirmesine katkıda bulunarak iş süreçlerinde yetki ve sorumluluk üstlenmişlerdir. Bilgi teknolojilerindeki gelişmeye paralel olarak kurumlar tarafından sunulan hizmetlerin vatandaşa/müşteriye ulaştırılma süresi kısalmıştır. Kurumlar, piyasada rekabet avantajını kendi lehlerine çevirmek ve yeni ürünleri belirlemek adına daha hızlı mal ve hizmet üretimi ve dağıtımı için veri madenciliği uygulamalarından yararlanmaktadır. Piyasaların küreselleşmesi, işletmelerin sahip olduğu bilgi teknolojileri tarafından desteklenmektedir. Kurumların ve organizasyonların bilgi teknolojileri altyapıları, veri madenciliği teknolojilerinden yararlanabilecek şekilde entegre edilmektedir.

### **1.3.1. Veri Madenciliğiyle İlişkili Kavram ve Uygulamalar**

Bilgisayarların hayatımızda yer almaya başlamasının ardından günümüzde yaygınlaşan internetle birlikte mobilitenin de artması sonucu sayısal ortamda yapılan her işlem kayıt

altına alınmaya başlamıştır. Sağlık, alışveriş, borsa işlemleri, sosyal medya kullanımı, özel şirketler, sivil toplum ve tüketicilerin faaliyetleri sonucunda ortaya çıkan veriler oluşan veri yığınının ne denli büyük miktarda olduğunu göstermektedir. Ayrıca kamu kurumlarında vatandaşların yaptığı işlemler sonucu meydana gelen veriler de eklendiğinde toplanan ve depolanan verinin büyüklüğü daha iyi anlaşılmaktadır.

Veriler daha önceden, el yordamıyla kâğıt ortamında saklanırken, bilgisayarların hayatımıza girmesiyle birlikte sayısal ortama aktarılmış ve teknolojinin gelişmesi ile birlikte veritabanlarında depolanmaya başlamıştır. Depolanan verinin analizi için insan yeteneği yeterli olmamakta, anlamlı hale gelmesi için bilgiye dönüşmesi gerekmektedir. Uygun yazılımların gelişmesiyle birlikte kurumlar tarafından toplanan verinin işlenmesi, aralarındaki ilişkinin tespiti, birlikteliklerin tespiti ve örüntülerin çıkarılması mümkün hale gelmiştir.

İstatistik, veri analizi ve yönetim bilişim sistemleri gibi bilim dalları tarafından sıklıkla kullanılan veri madenciliği kavramı, ilk olarak geçerliliği araştırılacak bir hipotez olmaksızın verinin araştırılmasını hedefleyerek istatistikçiler tarafından kullanılmıştır. 1978 yılından Leamer tarafından klasik istatistik yöntemlerle çözülmesi zor problemlerin doğrusal regresyon kullanılarak çözülebileceğini göstermek için kullanılmıştır (Usama ve Paul, 1997: 102).

1990'lı yıllardan itibaren veri depolama araçları, barkod ve sensör teknolojilerinin de gelişmeye başlamasıyla birlikte kullanımı artarak literatürde daha sık kendinden söz ettirmeye başlamıştır. Eldeki verilerden yola çıkılarak, önceden bilinmeyen geleceğe dönük tahminde bulunma olarak tanımlanabilecek veri madenciliğine, müşterilerin satınalma davranışlarının belirlenebilmesi için yapılan sepet analizi (Chen ve diğerleri, 2005: 339, Ozyirmidokuz ve diğerleri, 2015: 77) klasikleşen bir örnek olarak verilebilir. Veri madenciliği, bazen veri tabanından bilgi keşfi ile aynı anlamda kullanılmakla birlikte, bazı veri madenciliği algoritmaları ile verinin analiz edilmesini esas almaktadır.

Literatürde, farklı disiplinlerde çalışmalar yapan araştırmacılar tarafından yapılan veri madenciliği tanımları bulunmaktadır. Veri madenciliği; veritabanında bulunan büyük miktardaki bilinmeyen veriler üzerinde yapılan kompleks işlemlerle değerli bilginin keşfidir (He ve diğerleri, 2013: 123). Bir diğer tanıma göre veri madenciliği daha

önceden bilinmeyen, geçerli ve uygulanabilir bilgilerin geniş veritabanlarından elde edilmesi ve bu bilgilerin karar verirken kullanılmasıdır” (Silahtaroglu, 2013: 12). Diğer bir tanımda ise kurumsal verideki gizli bilgiye erişerek karar vericilere sunulmasını sağlayan karar destek analiz sürecidir (Kleissner, 1998: 296). Büyük miktardaki veri setlerindeki kullanışlı desen (pattern) ve tercihlerin (trends) keşfedilmesi, ortaya çıkarılması sürecidir (Larose, 2006: 2). Veri setlerindeki desenlerin, ilişkilerin, kuralların ve istatistiksel olarak önem arz eden yapıların keşfedilmesi, daha önce tespit edilmemiş veri desenlerinin tespit edilmesi sürecidir (Dener ve diğerleri, 2009: 788). Veri madenciliği, genel olarak veritabanlarında bilgi keşfi (knowledge discovery in databases, KDD) sürecinin en önemli adımı olarak tanımlandığından, veri madenciliği ile veritabanlarında bilgi keşfini birbirinin yerine kullanmaktadır (Rokach ve Maimon, 2015: 4). Bilişim teknolojilerindeki ilerleme ile birlikte farklı disiplinlerin kullandığı araçlardan faydalanmak suretiyle karar vericilerin etkin ve daha hızlı karar verebilmesi için kurumların sahip olduğu yüksek hacimli veriden önceden bilinmeyen, gizli, örtük gerekli olan bilgiye ulaşılmasıdır (Şentürk, 2006: 3). Tanımlarda geçen ortak terimler göz önünde bulundurulduğunda veri madenciliği; karar vericilerin etkin ve verimli karar verebilmesi amacıyla bilgisayar teknolojilerinin sağlamış olduğu imkânlarla yüksek hacimli verinin analizi ve farklı disiplinlerin yöntemleri vasıtasıyla veri içerisindeki yeni, gizli kalmış veya beklenmeyen örüntüleri bulmak için kullanılan uygulamalardır.

Veri madenciliği Şekil 6’da görüldüğü üzere istatistik, makine öğrenimi, yapay zekâ ve veritabanı teknolojilerini birleştiren disiplinler arası bir alandır. Birçok kurum ve işletmenin sahip olduğu veri miktarı kurulduğu dönemden itibaren geçen süre göz önüne alındığında büyük kapasitelere ulaşmakta olup, bu veriler üzerinden yapılacak veri madenciliği uygulamaları ile çok değerli bilgiler elde edilebilir (Sayad, 2010: 1).



**Şekil 6:** Veri Madenciliği ve İlgili Disiplinler

**Kaynak:** Han ve diğerleri, 2012: 23.

Veri madenciliği veriyi toplama, sınıflandırma, özetleme, organize etme, analiz etme ve yorumlama bilimi olarak “istatistik” bilimi, bir bilgisayarın veya bilgisayar kontrolündeki bir robotun çeşitli faaliyetleri zeki canlılara benzer şekilde yerine getirme kabiliyeti olarak tanımlayabileceğimiz “yapay zeka”, bilgisayarların algılayıcı verisi ya da veritabanları gibi veri türlerine dayalı öğrenimini olanaklı kılan algoritmaların tasarım ve geliştirme süreçlerini konu edinen bir bilim dalı olarak “makine öğrenmesi” ve birbiriyle ilişkili alanların depolandığı “veritabanı” yazılımı olmak üzere farklı disiplinleri bir araya getirmektedir. Veri madenciliği ve makine öğrenmesinin çıktısı tahmin/sınıflandırma üzerinden elde edilen kurallardır. Karar destek sistemlerinde makine öğrenmesinin avantajlarından faydalanabilmesi için kurum çalışanlarının tecrübesi büyük önem arz etmektedir.

Veri madenciliği 1990 ve 2000’li yıllar arasında gelişimini devam ettirmiş özel sektörde farklı alanlarda uygulanmaya başlanmıştır. Google, Facebook, LinkedIn gibi ticari bilişim firmalarının yanı sıra kamu kurumları tarafından da veri madenciliği uygulamaları üzerinden verilere dayalı olarak çeşitli çalışmalar gerçekleştirilmeye başlanmıştır.

Veri madenciliğine yönelik Amerika’da “Knowledge Discovery in Databases”, Avrupa’da farklı ülkelerde “Principle of Knowledge Discovery in Databases” ve Avusturalya’da “Pasific Asia Knowledge Discovery and Data Mining” gibi uluslararası konferanslar düzenlenmektedir. “Data Mining and Knowledge Discovery” ve “Statistical Analysis and Data Mining” gibi akademik yayın yapan dergilerin yanı sıra “SIGKDD: The community for data mining, data science and analytics” disiplinlerarası olarak yapılan veri madenciliği ve veri bilimine yönelik yapılan çalışmalarla da akademik literatüre katkıda bulunmaktadır.

### **1.3.1.1. Veri, Bilgi ve Enformasyon**

Hedeflenen uygulama için anlamlı olduğu sürece herhangi bir veri seti üzerinde veri madenciliği gerçekleştirilebilir. Veritabanı verisi (database data), veri ambarı verisi (datawarehouse data) ve günlük işlem gören veri (transactional data) veri madenciliği uygulamaları için temel veri kaynaklarıdır. Veri madenciliği çalışmaları gerçekleştirilecek “veri” kavramı, literatürde bilgi hiyerarşisi içerisinde yer almaktadır.

Her ne kadar birbiri yerine kullanılsa da veri (data), enformasyon (information) ve bilgi (knowledge) kavramları arasında ciddi farklar bulunmaktadır (Bensghir, 1996: 14).

Kurumsal amaçlara bağılı olarak işlemlerin yapılandırılmamış halde, kavramsal bir çerçeve ve yapı içerisinde bulunduğu enformasyona “veri” denir (Aktan ve Vural, 2005: 6; Barutçugil, 2002: 57). Kurumda elde edilmek istenen bilgi ve ürünler için kurum çalışanlarının sahip olduğu veri yeterli olmamakla birlikte, karar verme aşamasına bir temel oluşturur (Davenport ve Prusak, 2001: 22).

Belirli bir amaca hizmet etmek üzere biçimlendirilmiş ve düzenlenmiş halde bulunan, kurumdaki çalışanların doğru olup olmadığına kendisinin karar vermesi suretiyle bakışında ve konuyu algılayışında fark yaratan dağınık haldeki veriler “enformasyon” olarak tanımlanır. Enformasyon kurum içinde bilgi teknolojileri ya da geleneksel yöntemlerle yer değiştirmektedir. Verilere insanlar tarafından anlam katılarak amaca uygun hale getirilmesi ve düzenlenmesi neticesinde veri, enformasyon halini almaktadır (Davenport ve Prusak, 2001: 25; Laudon ve Laudon, 1998: 8). “Bilgi” ise öğrenme, gözlem ve araştırma yoluyla elde edilen gerçek ve kavrayışın tümü, diğer bir ifadeyle doğruluğu ispatlanmış inançlardır (Nonaka ve Takeuchi, 1995: 58).

Yukarıda verilen tanımlardan yola çıkılacak olursa veri, işlenmemiş enformasyon parçacıkları iken, enformasyon organize edilmiş veri seti, bilgi ise anlamlı olan enformasyonları ifade etmektedir (Bhatt, 2001: 68-75).

Veri madenciliği ile ilişkisi bakımından “veritabanı verisi”; birbiriyle ilişkilendirilmiş veriyi bir arada barındıran, yapılandırılmış veritabanı sorgulama dili (SQL) veya grafik arayüzler vasıtasıyla erişilebilen veridir. Verinin tek bir veritabanında barındırılması, okuma, yazma ve bunların senkronizasyonu gibi birçok faaliyeti içeren “eş zamanlı (concurrently)” kavramını ortaya koyarak, veriye birden fazla kullanıcı ve uygulama tarafından aynı anda erişilmesine imkân sağlamıştır.

### **1.3.1.2. Veritabanı Yönetim Sistemleri**

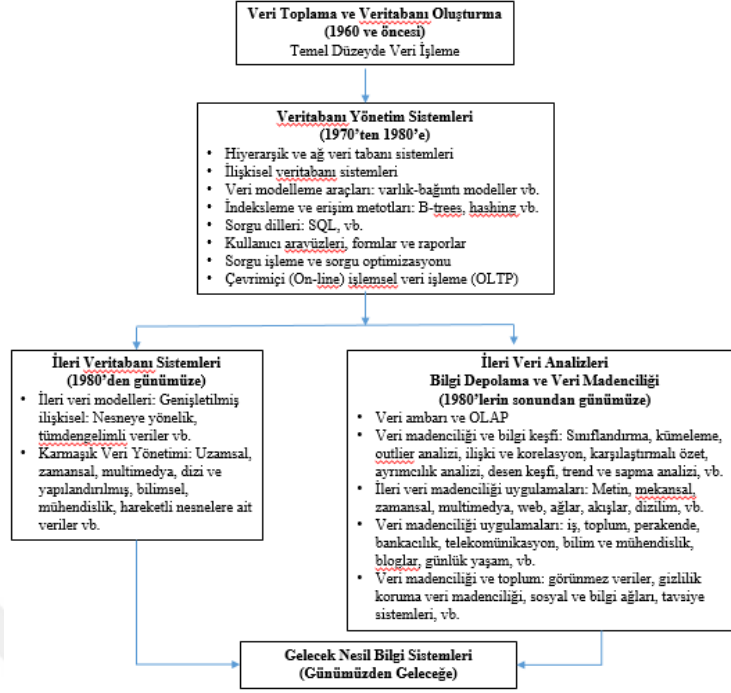
Veritabanı Yönetim Sistemleri (VTYS), verileri tek bir yapısal veritabanında birleştirerek bu verinin gereksiz tekrarını önlemeye yardımcı olur. Veritabanları yaratmak ve yönetmek, kullanıcıları yetkilendirmek, yedekleme işlemlerini gerçekleştirmek, verinin fiziksel ve bellek anlamda düzenli bir şekilde kullanılmasını

sağlamak için kullanılan sistemler veritabanı yönetim sistemleri olarak adlandırılır. 1970'lerde geliştirilen VTYS, 1980'lerin sonunda bilgi depolama ve veri madenciliği gibi ileri veri analizleri aşamasına ulaşmış, günümüzde ise gelecek nesil bilgi sistemleri tartışılır hale gelmiştir (bkz. Şekil 7)

İlişkisel veritabanlarındaki verilerle yapılacak veri madenciliği çalışmalarıyla ileri düzeyde veri modelleri üzerinde çalışılabilir. 1970 yılında E. F. Codd tarafından (Codd, 1970: 377) yayınlanan ilişkisel veritabanı hakkındaki makale ile bilim dünyası ayrı tablolara yerleştirilmiş veriler belirli alanlar üzerinden birbiri ile ilişkilendirildiği veritabanı ile tanışmıştır. İlişkisel veritabanlarının en önemli nesnesi satır ve sütunlardan oluşan tablolardır. Tablo içindeki her satır bir kayda, her sütun bir özelliğe denk gelmektedir. Örneğin, veri madenciliği sistemleri, yeni müşterilerin gelir, yaş ve önceki kredi bilgilerini esas alarak yapacağı analizlerle müşterinin kredi riskini hesaplayabilir.

Günümüzde açık kaynak ve ücretli ilişkisel veritabanı yazılımları mevcuttur. Oracle ve Microsoft SQL Server ücretli yazılımlar olarak en çok tercih edilen yazılımlardır. MySQL Server ve Postgre SQL Server ise ücretsiz olarak kullanılan yazılımlar arasında çoğunlukla kurumlar/geliştiriciler tarafından tercih edilmektedir. Ücretsiz yazılımların kurumlar ve şirketlere işletme/lisans maliyeti getirmemesi artı bir avantaj sağlarken, kurumsal desteğinin sınırlı kalması dezavantajı olarak görünmektedir.

Ücretli yazılımları ücretsiz yazılımlardan ayıran husus ise kurumsal desteğinin olması ve lisans unsurudur. Ücretli yazılımların lisans maliyetleri kurumlara ciddi anlamda yük oluşturabilmektedir. Kurumsal desteğin önemi, veritabanı yazılımlarında herhangi bir problemle karşılaşıldığında başvurulup destek alınabilecek bir muhatabın olması durumunda ortaya çıkmaktadır. Ayrıca veritabanının kurumun verisini barındırdığı göz önünde bulundurulduğundan bu konunun bilgi güvenliği bakımında önemi de daha iyi anlaşılmaktadır.



**Şekil 7: Veritabanı Teknolojilerinin Gelişimi**

**Kaynak:** Han ve diğerleri, 2012: 3.

Veritabanları saklanacak verinin kullanım amacına göre, verinin sürekli olarak değiştiği, üzerinde ekleme, silme ve güncelleme işlemlerinin gerçekleştirildiği “OLTP (Online Transaction Processing) veritabanları” ve verinin geçmişe dönük raporlama ve geleceğe dönük karar destek amaçlı tahminlerde kullanılmak üzere saklandığı, içeriği çok sık değişmeyen verilerden oluşan “OLAP (Online Analytical Processing) veritabanları” olmak üzere ikiye ayrılmaktadır. Veri ambarları üzerinde karar vermeye yönelik olarak yapılan veri analizi ve sorgulama işlemlerine OLAP (On-Line Analytical Processing) denir. Operasyonel veritabanlarında günlük işlemlere ait kayıtlar üzerinde işlemler yapılırken, veri ambarlarında işlemler üzerinden geleceğe dönük tahminler yapmak üzere işlemler gerçekleştirilir.

“İşlem verisi”, müşteri siparişi bilgisi veya uçuş rezervasyon kaydı verisi gibi anlık olarak veritabanına işlenen verilerdir. Günümüzde birçok kurum ve kuruluşun operasyonlarıyla veritabanları üzerinde ekleme, silme ve değiştirme neticesinde oluşan ve sürekli değişen verilerdir.

### 1.3.1.3. Veri Ambarları

Veri madenciliği çalışmalarını gerçekleştirmek üzere veri ve veritabanı gerekli olmakla birlikte kurumlarda işlemsel veritabanları üzerinde veri madenciliği uygulamaları yapılamaz. Mevcut verinin veri madenciliği uygulamalarında kullanmak üzere uygun hale getirilmesi gerekmektedir. Veriler tarihsel biçimde enformasyon sağlamak için saklanırlar. Örneğin, her satışın detayını tutmak yerine, her ürün tipinin işlem özetini ya da bölgeye göre satışını saklar. Bu sayede karar vermede kullanılmak üzere veri analizleri ve raporların oluşturulması için veri üzerinde çeşitli görüntüler oluşturulur. Belirli döneme ait, konu odaklı düzenlenmiş, birleştirilmiş ve sabitlenmiş kurumlara ait veritabanlarına “veri ambarları” denilmektedir (Silahtaroglu, 2013: 17).

Veri ambarları mimarisi “yıldız”, “kartanesi” ve “anatablo” olmak üzere üç değişik veri ambarı şemasını kullanmaktadır. *Yıldız şema* türünde ortada bir ana tablo ve çevresinde veri ambarının boyutlarını oluşturan boyut tabloları bulunur. *Kartanesi şema* türünde ise, yıldız şemadan farklı şekilde normalize edilmiş diğer boyut tablolarına bağlanmıştır. *Anatablolar* birliğinde ise, birden fazla anatablo mevcut boyut tablolarını ortak olarak kullanır (Silahtaroglu, 2013: 19-20).

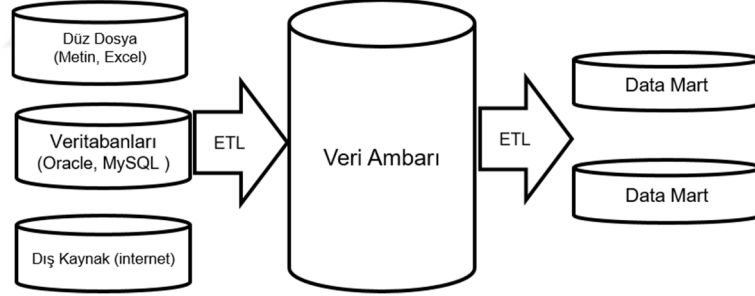
Veri ambarları hareketsetel ve organizasyonel görevlere yönelik oluşan verinin depolanarak sorgulanması ve analiz edilmesi, pazarda yeni fırsatlar bulunması, rekabete katkı, iş, envanter ve ürün maliyetlerinin azaltılması gibi amaçlar için oluşturulmaktadır (Şentürk, 2006: 8-10). Veritabanları ile veri ambarları arasındaki temel fark, veritabanlarının operasyonel işlem yapan kullanıcılar ve veritabanı yöneticileri, veri ambarlarının ise bilgi üretmeyi amaçlayan analistler ve uzmanlar tarafından kullanılmasıdır (Silahtaroglu, 2013: 20-21). Veritabanı ve veri ambarı arasında daha kapsamlı bir karşılaştırma Tablo 3’te gösterilmektedir.

**Tablo 3**  
**Veritabanı ve Veri Ambarı Arasındaki Karşılaştırma**

Veritabanı	Veri Ambarı
Verilerden oluşur.	Tarihsel verilerden oluşur.
Organizasyonun her aşamasında veriye ulaşılır.	Üst yönetime hitap eden karar destek sistemleridir.
Son kullanıcı sayısı fazladır.	Son kullanıcı sayısı (ortalama olarak 100'den) azdır.
Çevrimiçi çalışır.	Çevrimdışı çalışır.
Sorgularla istenilen sonuca anında ulaşılabilir.	Veri madenciliği gibi uzun süreçlerle analiz edilebilmektedir.
Güncel verilerden oluşur.	Geçmiş (5-10 yıl) hakkında uzun süreli veri saklanmaktadır.

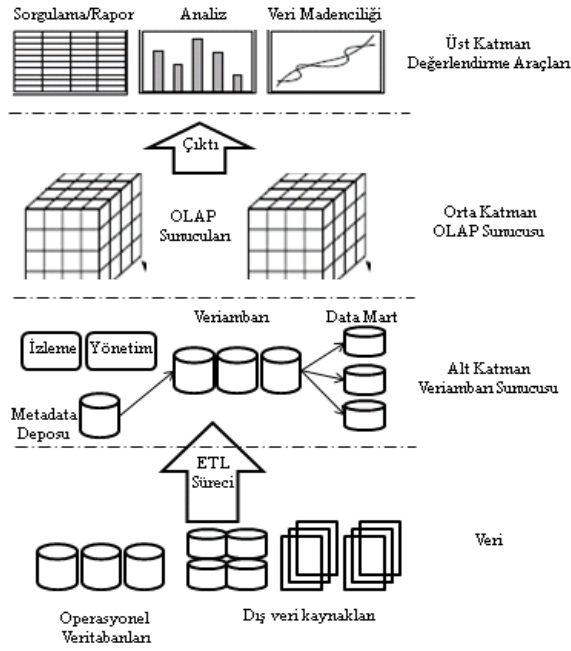
**Kaynak:** Şentürk, 2006: 10.

Kurumların sahip olduğu veriler Şekil 8'de görüleceği üzere "Data Mart" olarak isimlendirilen işleve özel veritabanlarına aktarılarak veri madenciliği tekniklerinin doğrudan bu veriler üzerine uygulanabilmesi sağlanmaktadır. Bu sayede karar vericiler, veri ambarında yer alan bütün veriler üzerinde işlem yapmadan sadece ihtiyaca dönük veriler üzerinden karar verebilmektedir.



**Şekil 8:** Data Mart Veritabanları

Veri madenciliği aynı zamanda, diğer formatlardaki veri akışı (data streams), grafik ve ağ verisi (graph or networked data), uzaysal veri (spatial data), metin verisi (text data) gibi veri setlerine de uygulanabilir. Borsa verisi, tarihsel kayıtlar, video kayıtları, sensör verisi, harita verisi, mühendislik tasarım verisi, sosyal medya verisi gibi zamana dayalı veriler bu tarz verilere örnek olarak gösterilebilir (Han ve diğerleri, 2012: 9-10).



**Şekil 9:** Üç Katmanlı Veri Ambarı Mimarisi

**Kaynak:** Han ve diğerleri, 2012: 131.

Veri madenciliği uygulamalarında veri ambarı ve OLAP teknolojisi yapılandırması Şekil 9'da gösterilmiştir. Veri ambarları mimarisinde temelde operasyonel veritabanı ve dış veri kaynakları bulunmaktadır. Alt katmanda ise metadata deposu, izleme ve yönetim işlemlerinin yapıldığı veritabanları ile birlikte belirli bir amaca yönelik data mart veritabanları bulunmaktadır. Orta katmanda geçmişe dönük raporlama ve geleceğe dönük karar destek amaçlı tahminlerde kullanılmak üzere saklandığı, içeriği çok sık değişmeyen verilerden oluşan OLAP veritabanları bulunmaktadır. Son katmanda ise raporlama araçları, analiz araçları ve veri madenciliği uygulamalarını barındıran değerlendirme araçları bulunmaktadır. Veri madenciliği uygulamalarında veri ambarlarında toplanan veri kullanılabilir.

### 1.3.2. Veri Madenciliğinin Tanımlanması

Günümüzde “bilgi çağında yaşıyoruz” ifadesi popüler bir terim olsa da aslında veri çağında yaşıyoruz. İş, toplum, bilim, mühendislik ve sağlık başta olmak üzere hayatın her alanında terabayt hatta petabaytlarca veri bilgisayar ağında toplanmaktadır. Günlük olarak çok fazla miktarda verinin toplandığı bir dünyada yaşamaktayız. Böylesine büyük veri yığınlarının analiz edilmesi önemli bir ihtiyaçtır (Han ve diğerleri, 2012: 1). Veri madenciliği terimi popüler bir seçim olarak karşımıza çıkmakla birlikte “veriden bilgi

madenciliği”, “bilgi çıkarımı”, “veri analizi” ve “veri taraması” gibi veri madenciliğine benzer deyimler de bulunmaktadır. Birçoğu ise veriden bilgi keşfini (knowledge discovery from data, KDD) tercih ederken, diğerleri ise veri madenciliğini bilgi keşfi sürecinin önemli bir adımı olarak görmektedir. Bilgi keşfinin adımları şunlardır (Han ve diğerleri, 2012: 8);

1. Veri Temizleme (Data cleaning): Tutarsız ve parazitli verinin temizlenmesi,
2. Veri Entegrasyonu (Data Integration): Çeşitli veri kaynaklarındaki verinin bir araya getirilmesi,
3. Veri Seçimi (Data Selection): Veri analizine yönelik verinin seçimi,
4. Veri Dönüşümü (Data Transformation): Seçilen verinin uygun formatta birleştirilerek veri madenciliği için uygun hale getirilmesi,
5. Veri Madenciliği (Data Mining): Algoritmalar uygulanarak veri modellerinin (paternleri) ortaya çıkarılması,
6. Model Değerlendirme (Pattern Evaluation): Veri modelindeki dikkat çeken hususların değerlendirilmesi,
7. Bilgi Sunumu (Knowledge presentation): Görselleştirme ve bilgi sunumu teknikleri kullanılarak veri madenciliği sonucunda elde edilen bilginin sunumu.

Yukarıda sunulan adımlar veri madenciliğinin bilgi keşfi süreçlerinden birisi olduğunu, gizli kalmış bilgiyi açığa çıkardığı için de en temeli olduğunu göstermektedir. Hâlbuki endüstride, medyada ve araştırma literatüründe veri madenciliği bütün bilgi keşfi süreçlerini kapsayacak şekilde kullanılmaktadır. Bu nedenle, veri madenciliği büyük miktarlardaki veriden gizli modelleri, desenleri keşfetmek için kullanılan bir süreç olarak kapsamlı bir şekilde tanımlanabilir (Han ve diğerleri, 2012: 8). Veri madenciliği uygulamasının hangi probleme çözüm üretmek amacıyla yapılacağına karar vermek sürecin ilk adımıdır. Tanımlanan problemle ilgili farklı kaynaklardan verinin toplanarak veritabanının oluşturulması ikinci aşamada uygulanır. Toplanan verinin veri madenciliğine uygun hale getirilmesi, temizlenmesi, bütünleştirilmesi, dönüşümü ve azaltılması gibi işlemler verinin ön işlenmesi sürecinde uygulanır. Ön hazırlık süreci veri madenciliğinin en önemli ve uzun zaman alan kısmıdır. Elde edilen temiz ve azaltılmış

veri üzerinde farklı amaçlar için geliştirilmiş algoritmalarından yararlanmak suretiyle verinin özelliklerine en uygun model belirlenir.

### **1.3.3. Veri Madenciliği Türleri**

Veri madenciliğinin tahminleyici ve tanımlayıcı olmak üzere iki esas hedefi vardır. Tahmin, veritabanındaki bazı değişkenleri kullanarak bilinmeyeni tahmin etmek veya gelecekteki değerlerini öngörmeyi içerir. Tanımlama, veri içindeki örüntüleri meydan çıkarmayı amaçlar. Temelde veri madenciliği birliktelik, kümeleme ve sınıflandırma şeklinde üç işlemi içerir.

Tahmin edici veri madenciliği (predictive data mining), tarihsel verilerden yola çıkarak belirlenmiş hedef değişkenin değerini tahmin etmeyi gerektirir. Bu hedef ayrık olduğunda, işlemin sınıflandırma ile yapılması tercih edilir. Kredi başvurularında ödeme koşullarının tahmin edilmesi, müşterilerin sigorta üzerinden talep ettikleri alacakların birer dolandırıcılık işlemi olup olmadığı, bu işlemlere örnek olarak verilebilir. Aynı zamanda, devamlı değişen hedefin tahmin edilmesi regresyonun görevidir (Baesens ve diğerleri, 2009: 16).

Tanımlayıcı veri madenciliği (descriptive data mining), birliktelik kuralları ve kümeleme gibi veri içindeki belli başlı ilişkileri bulmayı hedefler. Birliktelik kuralları, sıklıkla market sepeti analizi için veri içindeki örüntüleri arar. Market alışverişi yapan insanların kola yanında cips satın alması tanımlayıcı veri madenciliğidir.

Veri madenciliğinin yanında, birbiriyle ilişkili olan “metin madenciliği” ve “web madenciliği” de bulunmaktadır. Metin madenciliği, büyük kapasiteli dokümanların analiz edilerek, metin tabanlı verilerden gizli bilgiye ulaşılmasını hedeflemektedir. Metin madenciliği e-postalar, raporlar ve makaleler gibi metin içerikli veriler üzerinde işlemler gerçekleştirilmesini, web madenciliği ise yapılandırılmamış web içerikleri, linkler ve web erişim istatistikleri gibi internet materyallerine dayanılarak analizlerin gerçekleştirilmesini hedeflemektedir (Tan ve Yu, 2003: 239).

Tanımlanabilen, içerikteki veri tipine göre organize edilebilen ve arama yapılabilen veri tipine yapısal veri, tanımlanabilir bir yapısı olmayan resim dosyaları, pdf, word, metin dosyaları, web üzerinde tutulan log (kayıt) verisine yapısal olmayan veri adı verilir. Veri madenciliği ile yapısal veri analiz edilmekte, metin ve web madenciliği ile yapısal

olmayan verinin veri madenciliğinde kullanılmak üzere yapısal hale dönüştürülmesi işlemi gerçekleştirilmektedir (Dolgun ve diğerleri, 2009: 49). Kurumların sahip olduğu verinin çoğunun yapısal olmayan veri olduğu düşünüldüğünde, metin ve web madenciliğinin ne kadar önemli olduğu daha iyi anlaşılacaktır.

### **1.3.3.1. Veri Madenciliği Uygulamaları**

Verideki artış neticesinde kullanışlı bilgiye erişmek için ihtiyaç duyulan veri madenciliği uygulamaları için geliştirilen yazılımlar sayesinde, veri madenciliğinde kullanılan veri sınıflandırma, kümeleme, kural çıkarma yöntemi gibi birçok algoritma veriler üzerinde çalıştırılabilmekte, bu algoritmalar sayesinde işlenen ham verilerden, istenilen ve amaçlanan bilgi elde edilebilmektedir (Kaya ve Özel, 2014: 7).

Bu bağlamda ticari yazılımlara bakıldığında IBM SPSS Modeler, SAS Enterprise Miner, Microsoft Analysis Services, KXEN, Angoss ve Oracle Data Mining gibi ürünler bulunmaktadır. Açık kaynak kodlu veri madenciliği yazılımlarından Keel, Knime, Orange, R, Rapid Miner (Yale) ve Weka en çok kullanılanlardır. Sık kullanılan veri madenciliği yazılımları şu şekilde açıklanabilir:

1. Intelligent Miner: İlişkilendirme, sınıflama, tahmin modelleme, gruplama, sıralı desen analizi, regresyon analizi gibi fonksiyonları içeren IBM ürünü bir sistemdir. Ayrıca yapay sinir ağları algoritmaları, istatistik metotları, veri hazırlama ve görsel gösterim araçları gibi ek özellikler içerir. Ölçeklendirilebilir olması ve IBM'in DB2 ilişkisel veritabanı sistemi ile sıkı birleşik olması onu diğerlerinden ayıran özelliklerdir.
2. Enterprise Miner: Regresyon, sınıflama ve istatistik analiz gibi fonksiyonları içeren SAS ürünü bir sistemdir. İstatistik analiz araçlarının çeşitliliği en önemli özelliğidir.
3. Mineset: İlişkilendirme, sınıflama gibi fonksiyonları ile istatistik ve görsel araçlar gibi özelliklere sahip Silicon Graphics ürünü bir sistemdir. Veriyi pek çok formda gösterebilen grafiksel ara yüzü en önemli özelliğidir.
4. Clementine: Kural indirgeme, yapay sinir ağları, sınıflama ve görsel araçlar gibi özellikleri destekleyen, Integral Solutions ürünü bir sistemdir. Clementine tarihi

bilgi içeren veriler için idealdir. Kredi kartı dolandırıcılığı tespiti, müşteri davranışı analizi, kimyasal maddelerin aşındırıcılık tahmini gibi pek çok farklı alanda uygulanabilir.

5. DBMiner: OLAP analizi, ilişkilendirme, sınıflama ve gruplama algoritmalarını içeren veri küp tabanlı çevrim içi çözümsel madencilik en önemli özelliğidir.
6. Weka: GNU lisansı altında açık kod bir yazılımdır ve bütünüyle ücretsizdir. %100 Java'da yazıldığı için platform bağımsızdır. Gelişmiş kullanıcı arabirimleri ile hızlı ve rahat uygulama geliştirmeye imkân verir. Barındırdığı geniş algoritma spektrumuyla kıyaslamalı uygulama geliştirmeye olanak sağlar.

### **1.3.3.2. Metin Madenciliği**

Bir sınıfa ait bir metni başka bir sınıfa ait metinden hangi özellikleriyle ayrıldığına dair metin dosyalarının otomatik sınıflandırılmasına yönelik çalışmalar uzun zamandır devam etmektedir (Can ve Amasyalı, 2016: 197). Yapısal olmayan veriler, standart olmayan metinlerden oluşmasına rağmen insanlar tarafından dilsel kalıplara bölünebilirken, bilgisayar bunu analiz edememektedir. Metin madenciliğinde amaçlanan unsur, insanların yapılandırılmamış veriyi analiz etme kabiliyeti ile bilgisayarın yüksek kapasiteli veriyi hızlı ve doğru bir şekilde analiz etme kabiliyetini birleştirmektir (Fan ve diğerleri, 2006: 78).

Yapısal olmayan veriden içerik çıkarma yöntemleri olan anahtar kelimeler, mantıksal aramalar, istatistiksel veya olasılıksal algoritmalar ve sinir ağları metinde geçen kelimenin özelliğini ortaya çıkarma temeline dayanmakta, içeriği açıklayıcı sonuçlar elde edilememektedir. Metnin içeriğini anlamak için doğal dil işleme (Natural Language Processing, NLP) sistemleri kullanılmakta, bu sayede metin içinde geçen terimler sınıflanabilmektedir. Metin madenciliği, anahtar ifadelerin metinlerden elde edilmesi, elde edilen bu ifadelerin kategorilere ayrılması şeklinde iki aşamadan oluşmaktadır (Dolgun ve diğerleri, 2009: 50).

Metin madenciliği, telekomünikasyon, bilimsel/medikal araştırmalar, iş dünyası, eğitim ve kamu kurumları olmak üzere farklı sektörlerde kullanılmaktadır. Bilimsel / medikal araştırmalarda yeni ilaç ve tedavilerin geliştirilmesi; iş dünyası/özel sektörde rekabet analizi, medya etkisi analizi, mevcut farkındalık, içerik kişiselleştirme, fikri mülkiyet

ihlali, müşteri desteği; kamu sektöründe milli güvenlik, terörist bağlantıların tespiti, kolluk görevlileri için suç ve suçluların önlenmesi; eğitimde ise araştırma ve atıf analizi gibi konularda uygulanmaktadır (Fan ve diğerleri, 2006: 81).

Merrill Lynch ve Gartner (aktaran Sharda ve diğerleri, 2014: 229) tarafından yapılan araştırma sonuçlarına göre yapılandırılmamış veri her 18 ayda iki katına çıkmaktadır. Günümüzde kaliteli hizmet sunumu ve rekabet avantajını elinde bulundurmaya isteyen kurumlar, veriyi etkili ve verimli bir şekilde kullanarak daha iyi kararlar vermek adına gerekli bilgiye sahip olmak istemektedir. Metin madenciliği çalışmaları büyük resmi görmek isteyen kurumlar için tam da bu noktada kendisini göstermektedir.

### **1.3.3.3. Web Madenciliği**

İnternetin kullanımına paralel olarak teknolojik imkanlar ve altyapının gelişmesi ile birlikte bilgi paylaşımı giderek artmaktadır. Kullanıcılar webdeki bilgiler üzerinden istedikleri analizleri yaparak ilginç, potansiyel ve kullanışlı bilgiyi keşfedebilmek için çeşitli araçlara ihtiyaç duymaktadır. Bu ihtiyaç web madenciliği çalışmalarının gelişmesini sağlamıştır.

Web madenciliği ile yapısal olmayan web verisi yapısal veriye dönüştürülmektedir. Web madenciliği, web içeriklerinin incelenerek kullanışlı bilginin çıkarılması, web site/sayfaları arasındaki bağlantının incelenmesi, web sitelerine erişim kayıtları üzerinden ziyaretçilerin davranış ve tutumlarının belirlenmesi olmak üzere üç farklı kategoride kullanılmaktadır (Dolgun ve diğerleri, 2009: 51):

1. Web İçeriği Madenciliği: Doğal dil metinleri, resimler, ses ve videolar gibi yapılanmamış ve anlamsal olarak bulanık verilerden oluşan web siteleri içeriğinde bilgi araştırma web içeriği madenciliği olarak tanımlanmaktadır. Web içeriği madenciliğinde kullanılan teknikler arama motorlarının ilk versiyonlarında kullanılan kelime çıkarmanın ötesine geçmiştir.
2. Web Yapısı Madenciliği: Webin link yapısı üzerinden web sayfaları hakkında bilgi çıkarma, yapısal bilginin keşfedilme süreci olarak tanımlanabilir. Konu çeşitliliği açısından çok sayıda linke sahip web sayfalarını sınıflandırmak veya dokümanlar arası benzerlik ölçümlerini yapmak için kullanılmaktadır.

3. Web Kullanım Madenciliği: Web sunucusuna yapılan erişim kayıtları üzerinden kullanıcılara ait anlamlı örüntülerin keşfedilmesi web kullanım madenciliği olarak adlandırılmaktadır.

#### **1.3.4. Veri Madenciliğinin Avantaj ve Dezavantajları**

Örgütler stratejik olarak rekabette avantaj sağlamak amacıyla değişim ve yenilik yapmaya ihtiyaç duymaktadır. Fark yaratmak içinse örgütün sahip olduğu ham veriler üzerinden anlamlı bilgiyi elde ederek kullanması adeta bir zorunluluk haline gelmiştir. Yönetim bilişim sistemleri üzerinden veri madenciliği uygulamaları yoluyla elde edilen amaca yönelik bilgi, karar vericilere tarafından kurumun nihai hedefleri doğrultusunda kullanılmaktadır.

Bilişim çağında bulunduğumuz dönemde verinin bu denli hızlı bir şekilde artması verinin kurumların ihtiyaçları doğrultusunda analiz edilmesi ihtiyacını doğurmuştur. Bu sebeple veri, veri madenciliğini etkileyen en önemli faktör olarak karşımıza çıkmıştır. Veriden kurumun geleceği adına verilecek kararlarda karar destek faaliyetlerinde yararlanabilmek için veri madenciliğinde farklı analiz teknikleri ve yöntemlerine ihtiyaç duyulmuştur. Bu teknikleri ayrıntılı olarak incelemeden önce, veri madenciliğinin avantaj ve dezavantajları değerlendirilmektedir.

##### **1.3.4.1. Veri Madenciliğinin Avantajları**

Kurumların sahip olduğu ham veriler veritabanlarında tutulmakta ve bu veriler üzerinde sorgulamalar, standart sorgulama dili (SQL) ile gerçekleştirilmektedir. Ham verinin kullanışlı bilgiye dönüştürülmesi için geliştirilmiş algoritmaların çalıştırılması veri madenciliği ile mümkündür. Veri madenciliği konusunda kullanılan yazılımlar, veri madenciliği süreçlerini kolaylaştırmaktadır.

Büyük kapasiteli verinin depolanabileceği, hızlı erişim sağlanabileceği donanımlara sahip olma maliyetleri, bu alandaki teknolojik gelişmelerle birlikte, son yıllarda giderek düşmektedir. Yüksek kapasiteli veri depolama birimleri ve hızlı işlemcilerle sahip donanımlarla veri madenciliği uygulamalarını geliştirmek mümkün hale gelmiştir.

Bilişim altyapısına yapılan yatırımların artması, internete erişimin kolaylaşması ve mobilite verinin hızla büyümesini sağlamıştır. Bilişim altyapısının gelişimi dağıtık şekilde bulunan verinin analiz edilmesini kolaylaştırmıştır.

Kamuda vatandaşa sunulan hizmetin etkin verimli ve etkili bir şekilde sunulması, özel sektörde sunulan hizmetlerin kaliteli bir şekilde sunulması için veri madenciliği tekniklerinin kullanılmasını etkilemektedir. Bu sayede veriler üzerinden analizler gerçekleştirilmekte, ileriye dönük tahminler yapılmakta ve karar destek faaliyetleri desteklenmektedir.

Veri madenciliğinde kullanılan araçlar problemlere kendiliğinden çözüm üretmemektedir. Her aşamasında dikkatli çalışma yapılması gerekmektedir. Veri madenciliği aşamalarına insan tecrübesi dâhil edilmeli ve maliyetler iyi hesaplanmalıdır. Veri madenciliği için verinin ön hazırlığı büyük önem taşımaktadır (Çakır, 2008: 7).

#### **1.3.4.2. Veri Madenciliğinin Dezavantajları**

Küçük ve tutarlı veri kümelerine hızlı ve doğru şekilde veri madenciliği teknikleri uygulanabilirken büyük ve karmaşık veri kümelerine bu teknikler uygulandığında aynı doğrulukta sonuçlar elde edilemeyebilir.

Veri madenciliğinin kamu sektöründe kullanımı veri madenciliği alanında uzman ihtiyacı, karar vericilerin veri madenciliğinin kabiliyetleri konusunda eğitilmesi ve bilgilendirilmesi ihtiyacı nedeniyle zor olabilir. İstenilen miktarda veri madenciliği uzmanının çalıştırılmayacağı ve kamudaki her karar vericiye veri madenciliği konusunda eğitim verilemeyeceği göz önünde bulundurulduğunda, veri madenciliği süreçlerinin otomatik hale getirilerek modellenmesi ve konun uzmanı olmayan personel tarafından da kullanılabilir hale gelmesi bu dezavantajı ortadan kaldıracığı gibi, farklı veri setleri için de kullanılabilir hale gelecektir.

Teknolojideki gelişmelerin hızına bağlı olarak vatandaşların gerek internet gerekse kurumlar üzerinden yaptığı işlemler neticesinde, oluşan kişisel verinin büyüklüğü, bu veriler üzerinden yapılan işlemler kişisel verinin korunması sorununu da beraberinde getirmektedir. Kişisel veriler işlenerek bilgi haline getirilmekte, kurumlar tarafından saklanmakta belirli durumlarda üçüncü kişilere/kurumlara devredilebilmektedir. Kişisel verinin dolandırıcılık amacıyla elde edilmesi (kimlik hırsızlığı), ticari ya da meslek

sırlarının kişi/kurumlara çıkar sağlanması amacıyla kullanılması (kişisel verinin suistimali), güvenliğin sağlanması, suçun önceden tespiti ve yargılanma esnasında delil olarak kullanılabilmesi için iletişim ortamlarının izlenmesi, denetlenmesi ve kayıt altına alınması, kişisel veriler üzerinde uygulanan veri madenciliği algoritmaları ve uygulamaları neticesinde gizli örüntülerin tespit edilmesi sonucu elde edilen bilgiler iyi niyetli amaçlarla kullanılabilmesi gibi, kötü niyetli amaçlar için de kullanılabilir (Karimi ve Korkmaz, 2013: 194-196). Veritabanlarında kamu ve özel sektör başta olmak üzere kişisel veriler saklanmaktadır. Verinin gizliliğine önem gösterecek şekilde bilgi keşfi sürecinde verilere erişim sağlanmalıdır.

Tıbbi araştırmaların yayınlanması, paylaşılmasına yönelik olarak doktorla hasta arasındaki ahlaki ilişkiye dikkat edilerek “zarar vermeme” ilkesine uyulması gerekmektedir. Etik olarak hasta verisi, veritabanlarına doğru yüklenmeli, veritabanlarındaki hasta verisine sadece doktor erişebilmeli, hasta verisi her isteyen kişi veya kurumla paylaşılmamalıdır (Can ve diğerleri, 2012: 4). Buradaki problem, veri madenciliği çalışmalarıyla hassas veri üzerinde yapılan çalışmalar neticesinde kişisel verinin izinsiz olarak yayınlanmasıdır. Kamu kurumları ve özel sektör kuruluşları, sıklıkla kamu fonları ve tıbbi araştırmalar, analizler gibi kişilere ait bilgileri içeren hassas veriyi yayınlamaya ihtiyaç duymaktadır. Esas problem, kullanıcılara ait olan verinin kullanıcıların bilgisi haricinde yayınlanmasıdır. Hassas veya kişisel veri, tıbbi veri, sayım verisi, seçmen kayıt verisi ve sosyal medya verisi ve müşteri verisinden oluşabilmektedir. Veri madenciliğinde gizliliği korumak için sayısız araştırmalar yapılmıştır. Genel olarak kullanılan çeşitli yaklaşımlar beş aşamalı olarak gizlilik çerçevesi olarak sınıflandırılabilir (Prakash ve Singaravel, 2015: 134):

1. Verinin dağıtımı: Veriler dağıtılmış veya merkezi olarak kategorize edilebilir. Merkezi veritabanlarında veri tek merkezde saklanırken, dağıtık veritabanlarında veri farklı veritabanlarında saklanmaktadır.
2. Veri üzerinde yapılan değişiklikler: Verinin değiştirilmesi, birleştirilmesi, örneklenmesi, gizlenmesi ve parazitli veri eklenmesi gibi farklı veri değiştirme yaklaşımları bulunmaktadır.
3. Veri madenciliği algoritmaları: Karar ağaçları, kümeleme, birliktelik kuralları ve regresyon gibi farklı algoritmalar kullanılmaktadır.

4. Verinin gizlenmesi: Ham veya birleştirilmiş, işlem görmüş verinin gizlenmesidir. Verinin saklanması, orjinal veritabanındaki kimlik bilgileri ve adres gibi doğrudan veya dolaylı olarak kişi ile bağlantısı olabilecek kişisel verinin gizlenmesidir.
5. Gizliliği koruma teknikleri: Genelleştirme, veriyi bozma, veriyi engelleme ve kriptolama gibi bazı teknikler kullanılarak kişilerin hassas verisine erişimin engellenmesidir.

Büyük hacimli, eksik gürültülü, aykırı veya belirsiz veriler üzerinde yapılan uygulamalarda yanlış sonuçlar üretilmesinin önüne geçmek için problemlere çözüm üretilmesi gerekmektedir (Savaş ve diğerleri, 2012: 6). Gürültülü veri hatalı veri toplama araçlarının kullanılması, veri giriş problemleri, teknolojik sınırlamalar veya veri isimlendirmede uyumsuzluk neticesinde oluşabilmektedir. Bunun neticesinde veri seti içinde tekrarlanan, çelişkili ve yetersiz veri oluşabilmektedir.

Veritabanı boyutları hızla artmakta iken, veri madenciliği uygulamaları örnek veri kümeleri üzerinde çalıştırılan algoritmalar ile test edilerek oluşturulmaktadır. Bu algoritmalar küçük örneklemeler üzerinden çalışacak şekilde tasarlanmıştır. Büyük hacimli veriler örüntülerin gerçekliğini göstermesi açısından avantaj gibi görünse de elde edilecek örneklem sayısının fazla olması en önemli problemlerden birisidir. Problemin giderilmesinde kullanılacak örnek veri kümesinde, çözüme yönelik olmayan gereksiz, niteliksiz artık verinin veri madenciliği teknikleri uygulanmadan önce temizlenmesi gerekmektedir.

Kamu kurumları coğrafi konum verisi gibi veriyi vatandaşların kullanımına sunduğunda bu verinin doğruluğu ve eksiksiz olduğunu taahhüt etmeyebilir. Bu veriyi kullanmak kullanıcının tercihine bırakılabilir (Rocheleau, 2006: 260). Vatandaşlar tarafından oluşturulan her bir verinin kontrollü bir şekilde veritabanlarına girişinin sağlanmadığı göz önünde bulundurulduğunda verinin hatalı olma ihtimali yüksektir. Veri girişi sırasında yapılan insan hataları veya oluşan sistem dışı hatalar gürültülü veri olarak adlandırılmaktadır. Günümüzde ilişkisel veritabanları gürültülü verinin önlenmesine yönelik sınırlı olsa destek sağlamaktadır.

Veri yığını içinde, anlam ifade etmeyen boş verinin örnek veri seti hazırlığında temizlenmesi gerekir. Boş verilere ilişkisel veritabanlarında sıklıkla rastlanılmaktadır. Kurumlara ait operasyonel veritabanları 7/24 esasına göre hizmet vermekte ve sürekli olarak veri giriş-çıkışı olmaktadır. Veri madenciliğinin dinamik veri setleri üzerinden uygulanması halinde tutarlı sonuçlara erişilmesi zor olacaktır. Veri madenciliği tekniklerinin, veri ambarı tarzındaki tutarlı veriler üzerinden gerçekleştirilmesi, sonuçların güvenilirliğini artıracaktır.

Veri setinin hacimli olması dolayısıyla, veri setinde eksik veri olmaktadır. Veri setindeki her özellik eksiksiz bir şekilde tanımlanmış olsaydı, sınıflandırma işlemi bu özellikler yardımıyla yapılabilirdi. Kurumların ihtiyaçlarına yönelik veri toplanması dolayısıyla veri madenciliği çalışması yapılacak veri seti içinde eksik veri bulunabilmektedir. Eksik veri içeren kayıtlar veri setinden çıkarılarak, değişken veri seti ortalaması veya eldeki veri setine dayalı en uygun değer kullanılarak veri madenciliğinde karşılaşılan bu problemlere çözüm üretilmektedir.

Gerçek veriler, ortam verisi ve coğrafi veri gibi farklı tipte verilere işlem yapılmasını gerektirebilir. Kullanılan veriler ilişkisel veritabanlarında saklanabileceği gibi düz metin olarak da saklanıyor olabilir. Farklı ortamlarda saklanan, farklı veri tiplerine uygulanacak şekilde bir veri madenciliği algoritmasının tüm verilere uygulanması mümkün değildir.

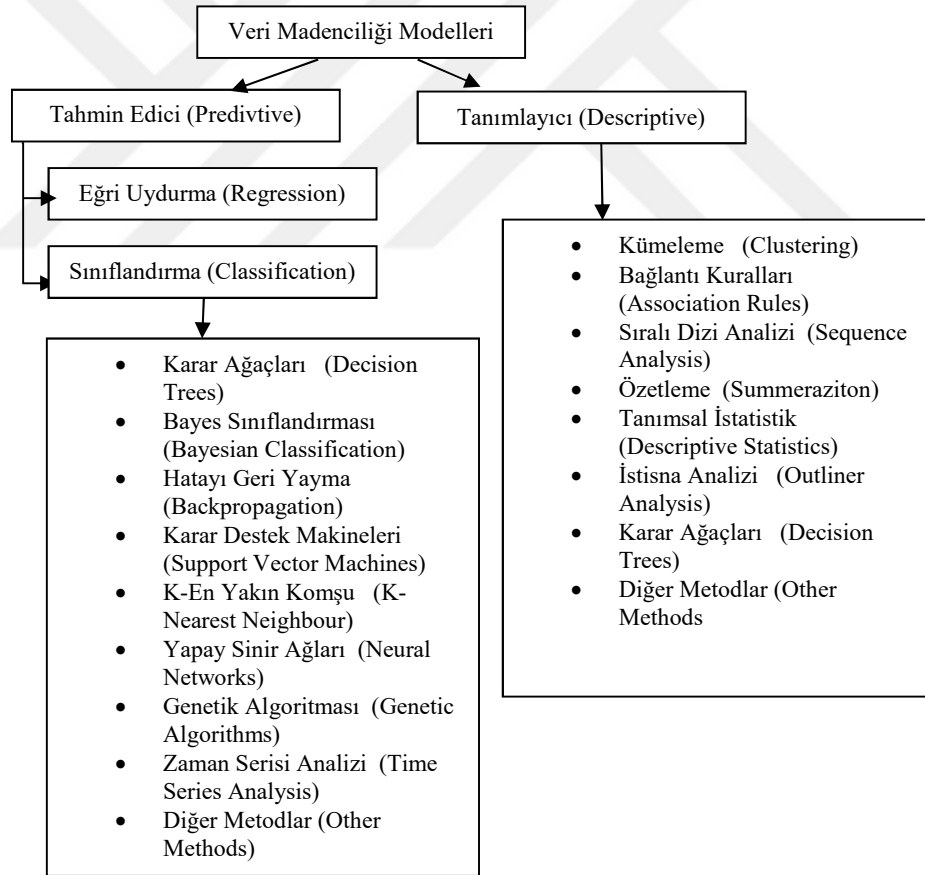
### **1.3.5. Veri Madenciliği Modelleri**

Veri madenciliği modelleri uzun araştırma ve geliştirme aşamaları neticesinde meydana gelmiş olup teknolojik gelişmeler, değişen ihtiyaçlar ve gerçek zamanlı büyük veriden etkilenmek suretiyle gelişmeye devam etmektedir. Veri madenciliği modelleri, uzun bir araştırma ve ürün geliştirme sürecinin sonucunda ortaya çıkmaktadır. Veri madenciliğinin menşei, verinin bilgisayarlarda ilk depolanmasıyla, kullanıcılara gerçek zamanlı olarak veri dolaşımı yapmasına izin verilen günümüze kadar veri erişimindeki gelişmelerle devam etmektedir. İş verisinden yararlı bilgilere geçiş evresinde, her adım önceki adımlara dayanmaktadır.

Geçmişte elde edilen veriler üzerinden çıkarılan kurallar gelecekle ilgili yapılacak tahminler için önemlidir. Uygulamada, tahminleyici (predictive) ve tanımlayıcı

(descriptive) olmak üzere iki farklı yaklaşımdan bahsetmek mümkündür. Tahminleyici modeller, “bir işlemde dolandırıcılık olup olmadığı” veya “bir müşteriden ne kadar gelir elde edilebileceği” gibi soruların yanıtlarını ararken, tanımlayıcı modeller ise veriler arasında gizli kalmış örüntülerden “çocuk bezi alan bir müşterinin mama alma olasılığı diğer müşterilerden üç kat fazladır” gibi sonuçlar çıkarmayı hedefler (Silahtaroglu, 2013: 49). Veri madenciliği modellerini işlevlerine göre Şekil 10’da detayları görüleceği üzere üç başlık altında incelemek mümkündür.

1. Sınıflama (Classification) ve Regresyon (Regression)
2. Kümeleme (Clustering)
3. Birliklik Kuralları (Association Rules) ve Ardışık Zamanlı Örüntüler (Sequential Patterns)



**Şekil 10:** Veri Madenciliği Modelleri

**Kaynak:** Sayad, 2010: 1.

Tahmin edici yaklaşımlar; sınıflama ve regresyon modelleri, kümeleme; birliklik kuralları ve ardışık zamanlı örüntü modelleri ise tanımlayıcı modellerdir. Kümeleme ise

tahmin ve tanımlayıcı modellerde kullanılabilir (Rygielski ve diğeri, 2002, 488). Bu modeller, neyin tahmin edilmesinin istendiğine dayalı olarak farklılaşırlar. Çıktı niteliğinin sürekli değerleri için tahmin istenir ise regresyon analizi, zamanın ayırt edici özellikleri ile ilgileniliyor ise zaman serileri, iyi veya kötü gibi az sayıdaki ayrık kategoriye sahip bir özel veri ögesi için bir kategorikal değer tahmini yapılmak isteniyor ise sınıflama gerekir. Eldeki verinin gruplarını bulan kümeleme, birliktelik ve ardışıklık kurallarını elde etmeyi kapsayan birliktelik analizi ve ardışıklık keşfi davranışı ise tanımlama amaçlı kullanılır.

Veri madenciliğinde örüntü tanıma faaliyetleri ise keşif (discovery), tahmin edici modelleme (predictive modelling) ve vaka analizi (forensic analysis) olmak üzere üç temel sınıfta toplanabilir (Rygielski ve diğeri, 2002: 487-488). Bir veri yığınındaki saklı örüntülerin önceden belirlenmiş bir fikir veya hipotez olmadan ortaya çıkarma sürecine keşif, sonuçları bilinen verilerden örneklem almak suretiyle elde edilen modelden bilinmeyen veri kümelerinin tahmin edilmesi tahmin edici modelleme (Köktürk ve diğeri, 2009: 22), anormal veya olağandışı veri öğelerini bulmak için çıkarılan kalıpları uygulama sürecini ise vaka analizi olarak tanımlananabilir.

Veri madenciliğinde kullanılan algoritma, teknik ve modeller kullanım amaçlarına göre birbirinden farklılaşsa da bazı ortak özellikleri bulunmaktadır. Örnek veri setinin bir bölümü üzerinde çalışan algoritmalar, bu veri setleri üzerinden bazı kurallar ve sonuçlar elde ederler. Bu sürece “öğrenme işlemi” adı verilmektedir. Öğrenme süreci neticesinde elde edilen bu kurallar ve sonuçlar veri setinin diğer bölümü üzerinde uygulanır. Bu şekilde çıkarılan kural ve sonuçların geçerliliği test edilir, gerekirse öğrenme işlemi tekrarlanarak kurallar yenilenir. Bu sürece ise doğrulama denir. Son olarak farklı veri setleri üzerinde kurallar denenerek kuralların doğruluğu test edilir. Kuralların geçersiz olması halinde “aşırı öğrenme” durumu söz konusudur.

#### **1.3.5.1. Sınıflandırma ve Regresyon Modelleri**

Sınıflandırma, makine öğrenmesinin önemli araçları arasında yer almaktadır. Tümevarımsal öğrenmenin temel çerçevesi olarak eğitim kümesi ve test örneklerinden oluşan bir test veri seti ile kullanılmaktadır (Emel ve Taşkın, 2005: 224). Sınıflandırma, verisi tasniflemek için kullanılan veri analizi modelidir. Hastanın hangi tedaviyi alması gerektiğinin tahmin edilmesi amacıyla kanser teşhisi verisinin analiz edilmesi, kredi

başvurularının “güvenli veya “riskli” olarak tahmin edilmesi gibi ihtiyaçlarda kullanılan veri madenciliği modelidir.

Sınıflama daha çok kategorik değerleri tahmin etmede kullanılırken, regresyon süreklilik gösteren değerlerin tahmin edilmesinde kullanılır (Han ve diğerleri, 2012: 328-329). Sınıflandırma sayesinde yeni öğelerin özelliklerine bakılarak veri sınıfının tahmin edilmesi mümkündür. Sınıfı tanımlanmış mevcut verilerden yararlanılarak herhangi bir sınıfa dâhil edilmemiş verinin sınıfının belirlenmesi için kullanılan veri madenciliği modelidir. Bu süreçte, ilk olarak tahmin işleminde kullanılacak model belirlenir. İkinci olarak, belirlenen bu model sınıfı belirli olmayan verilere uygulanarak sınıfları tahmin edilir. Sınıflandırma teknikleri geniş bir alanda problemlere çözüm üretmek için aynı alanın henüz bilinmeyen nesnelere davranışlarını tahmin etmek için kullanılabilir.

Sınıflandırmaya çok benzer olan regresyon (kestirim), hedef niteliğin sınıflandırmadaki gibi niteliksel bir vasıf değil, sürekli olmasıdır. Regresyonun amacı gözle görülmeyen nesnelere için hedef niteliklerin sayısal değerlerinin bulunmasıdır. Regresyon, zaman serisi verisi ile ilgiliyse “öngörü” olarak adlandırılmaktadır. Regresyonda kullanılan teknikler regresyon analizi, regresyon ağaçları, sinir ağları (artificial neural networks), k-en yakın komşuluk (k-nearest neighbour), box-jenkins, naive-bayes sınıflayıcısı, lojistik regresyon, karar ağaçları (decision trees) ve genetik algoritmalarıdır (genetic algorithms) (Tüzüntürk: 2010: 76; Çalış ve diğerleri, 2014: 5). Sınıflandırma ve regresyon teknikleri arasındaki temel fark, tahmin edilen bağımlı değişkenin kategorik veya süreklilik gösteren bir değere sahip olmasıdır (Akpınar, 2000: 6).

### **1.3.5.2. Kümeleme**

Kümeleme, birbiri ile yüksek benzerlik gösteren nesnelere birden çok grup veya kümede toplanması, benzerlik göstermeyen nesnelere ise diğer kümelerde toplanması işlemidir (Vinh ve Houle, 2010: 4). Aynı kümedeki elemanlar birbirleriyle benzerlik gösterirken, başka kümelerin elemanlarından farklıdır. Market verisi üzerinden ürünlerin raflardaki konumlarının belirlenmesi, hasta verisi üzerinden hastaların kümelenecek benzer tedavi yöntemlerinin hastalara gruplar halinde uygulanması bilindik kümeleme teknikleri ile çözüm bulunan uygulamalardır (Ramkumar ve Swami, 1998: 1). Verinin sınıflandırma modelinde olduğu gibi, kümeleme modelinde veri

sınıfları yoktur. Sınıflandırma modelinde verinin sınıfları bilinmekte, yeni bir veri geldiğinde hangi sınıftan olabileceği tahmin edilmektedir. Kümeleme modelinde ise tanımlayıcı modelleme olarak, sınıfları bulunmayan veriler gruplar halinde kümelere ayrılmaktadır. Kümeleme algoritmaları biyolojide bitki ve hayvan sınıflandırmaları üzerinden benzer genlerin sınıflandırılması, şehir planlanmasında evlerin tiplerine, değerlerine ve coğrafi konumlarına göre gruplara ayrılması ve web üzerinde bilgi keşfi için dokümanların sınıflanması amacıyla da kullanılabilir.

Kümeleme problemlerinde birbirine benzer veriyi gruplandırmak üzere çok boyutlu veri setleri veritabanında toplanır. Uygulama ve problemin içeriğine göre veri setlerinden oluşan gruplar, kümeler farklı şekillerde kullanılabilir (Aggarwal ve Yu, 1999: 13). Kümeleme metodunun veri madenciliği aracı olarak kullanılması için ölçeklenebilir, farklı özelliklere sahip veri tiplerine uygulanabilir, alan bilgisine gereksinim duyma, gürültülü veri setleri ile çalışabilir (Gürsoy, 2012: 4) ve sürekli değişen veri setlerine karşı gerektiğinde duyarsız olmasına ihtiyaç vardır (Han ve diğerleri, 2012: 446).

1. Ölçeklenebilirlik: Veritabanlarında, özellikle web içeriklerinde büyük veri setleri bulunmasına rağmen kümeleme algoritması veritabanındaki büyük veri setleri yerine daha küçük ölçekli, örnek veri setleri ile daha iyi çalışır. Büyük veri setlerinden alınan veri kümesi örneği ise önyargılı sonuçlar elde edilmesine neden olabilir. Bu sebeple ölçeklenebilir kümeleme algoritmalarına ihtiyaç vardır.
2. Farklı özelliklere sahip veri tiplerine uygulanabilmesi: Birçok algoritma sayısal verilerle çalışmaktadır. Bununla birlikte, bazı uygulamalar ikili (binary), kategorik (nominal) ya da bunların karışımı veri tiplerinden meydana gelen veri setleri üzerinden kümeleme işlemlerinin yapılmasını gerektirir. Artan bir şekilde daha fazla uygulama kümeleme algoritmalarına ihtiyaç duymaktadır.
3. Giriş parametrelerinin belirlenmesi için alan bilgisi gereksinimi: Birçok kümeleme algoritması, istenen küme sayısı gibi parametreleri belirlemesini istemektedir. Özellikle yüksek boyutlu veri setleri üzerinde kullanıcının veri seti ile ilgili bilgi sahibi olmaması durumunda bu tip parametreleri belirlemesi zor olacak, kümeleme algoritması neticesinde elde edilecek sonucun kalitesini de olumsuz etkileyecektir.

4. Gürültülü veri setleri ile çalışabilirlik: Veri setleri içerisinde gürültülü veri olarak adlandırılan eksik, bilinmeyen ve hatalı veriler bulunmaktadır. Bu tür gürültülü verilere duyarlı güçlü kümeleme algoritmalarına ihtiyaç bulunmaktadır.
5. Sürekli değişen veri setlerine karşı duyarsızlık: Yeni veriler geldikçe veri setlerinde meydana gelen değişimler nedeniyle bazı algoritmalar yeniden kümeleme yapmak zorunda kalır. Tutarlı sonuçlara ulaşmak için ihtiyaç duyulduğunda verinin değişimini hesaba katmayacak kümeleme algoritmalarına gereksinim duyulmaktadır.
6. Büyük hacimli verinin kümelenmesi: Küçük hacimli veri setleri iki-üç boyut içerirken, büyük hacimli veri setleri daha büyük boyutlar içerebilir.
7. Kısıtlama gerektiren veri setleri: Veri setleri üzerinde kısıtlama yapılarak kümeleme yapılması gerekebilir. Örneğin, belirli sayıdaki para çekme makinesini (ATM) şehrin hangi noktalarına yerleştireceğinize karar vermeniz istendiğinde, nüfusun yoğunlaştığı noktalar, yolların durumu gibi kısıtlamalara gitmeniz gerekecektir.
8. Sonuçların yorumlanabilir ve kullanılabilir olması: Kümeleme algoritmalarından elde edilen sonuçların kullanışlı olması gerekmektedir.

### **1.3.5.3. Birliktelik Kuralları**

Eldeki veri setleri üzerinden olayların birlikte gerçekleşme durumlarını inceleyen teknik birliktelik kuralları (Association rules) olarak literatürde yer almaktadır. Birliktelik kuralları ile veriler arasındaki ilginç ilişkilerin tespit edilmesi amaçlanır. Elde edilen verinin her geçen gün artması neticesinde kurumlar bu verilerden yararlanmak suretiyle birliktelik ilişkilerini keşfetmek ve karar alma süreçlerinde bu bilgiden yararlanmak istemektedir (Özekes, 2003: 76).

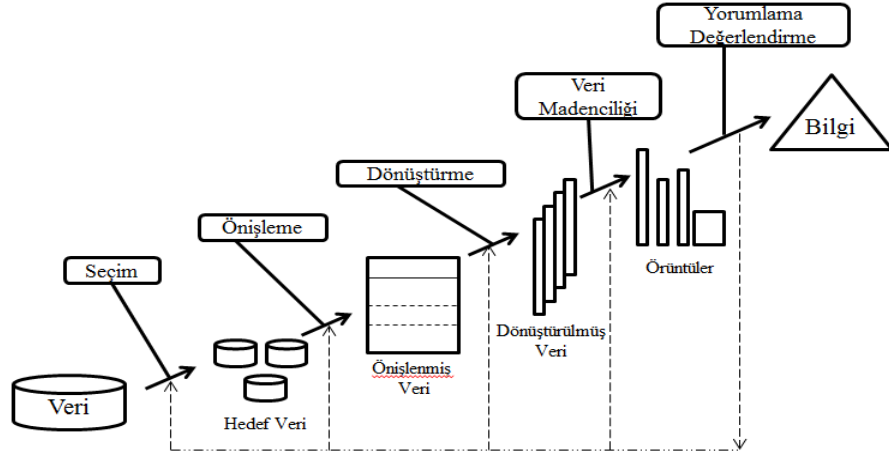
Farklı ürün gruplarının satışları yapılan bir süpermarkette, hangi ürünlerde indirimle gidileceği, satış miktarını arttırmak amacıyla ürünlerin raflara ne şekilde yerleştirilmesi gibi kararların alınması gerekmektedir. Bu tarz kararların doğruluk oranının artırılması için geçmiş satış işlemlerinin analiz edilmesi bilindik bir yaklaşımdır. Teknolojideki

ilerlemeler sayesinde eldeki veriler üzerinden sepet analizi (basket analysis) yapılarak belirli zaman dilimlerinde yapılan satış miktarlarının hesaplanması mümkündür. Veri madenciliği uygulamalarının farkı burada ortaya çıkmaktadır. Birliktelik kurallarının uygulanması ile “ekmek ve tereyağı” satın alan bir müşterinin süt alma olasılığının tahmin edilmesi, diet kola satışlarını artırılması için marketin ne kadar depolama yapması gerektiği ve önceki simit satış verisi üzerinden satışların durdurulması gibi kararların alınmasında etkili analiz sonuçlarına ulaşılması mümkündür (Agrawal ve diğerleri 1993: 1). Birliktelik kuralları uygulamalarında problem çözmek için korelasyon analizi, regresyon analizi, bayezyan ağlar ve görselleştirme gibi teknikler kullanılmaktadır.

### **1.3.6. Veri Madenciliği Uygulama Süreci**

Veri madenciliği, farklı beceri ve bilgi gerektiren yaratıcı bir süreçtir. Veri madenciliği projelerini yürütmek için izlenmesi gereken standart bir süreç yoktur. Bu durumda, veri madenciliği projelerinin başarı ya da başarısızlığı projeyi yürüten kişi ve ekibe bağımlı hale gelmekte, aynı başarı kurumların genelinde uygulanan projelerde tekrar edilememektedir. Bu kapsamda, veri madenciliği çözüm üretmek istenilen problemin uygun veriye dönüştürülmesi, veri madenciliği modellerinin önerilmesi ve sağlanması, elde edilen sonuçların etkinliğinin değerlendirilmesi ve belgelenmesi aşamalarında yardımcı olmak üzere standart bir yaklaşıma ihtiyaç duymaktadır (Wirth ve Hipp, 1998: 29).

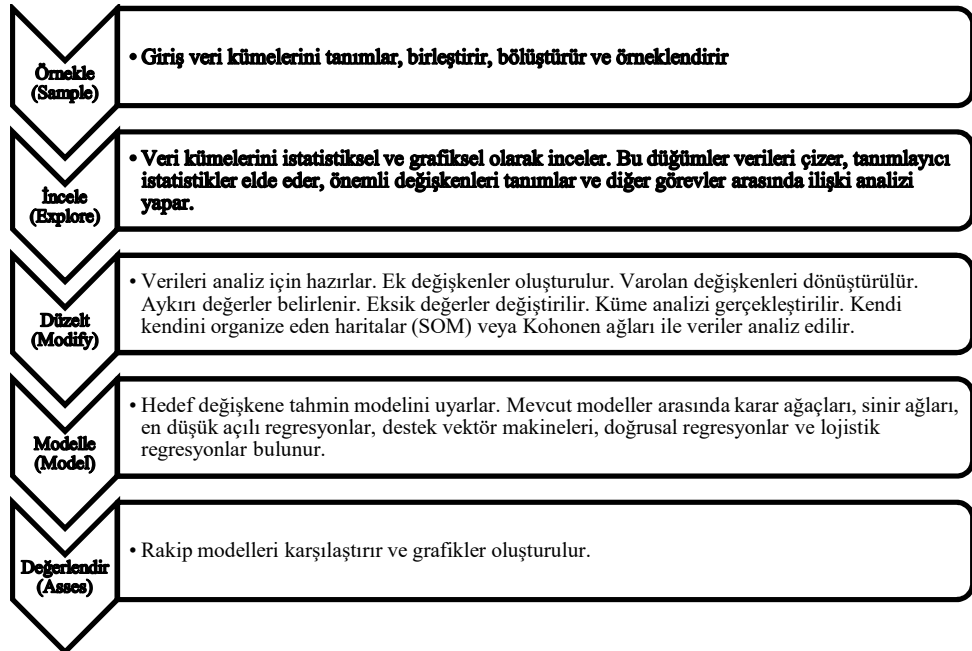
Veri madenciliği uygulamalarında sürecin tanımlanması noktasında ortak bir anlayışa ulaşamadığı (Geman ve diğerleri, 2016: 93) görülmektedir. Veri madenciliği, bazı kaynaklarda veritabanlarında bilgi keşfi süreci (knowledge discovery from data - KDD) şeklinde (Han ve diğerleri, 2012: 6; Kendall ve Kendall, 2010: 426; Fayyad, 1996: 21) bazı kaynaklarda veriden kullanışlı bilginin keşfi sürecinde bir adım (Brohman, 2006: 2; Zaitseva ve diğerleri, 2015: 312) olarak kullanılmaktadır. Birinci yaklaşımda veritabanlarında bilgi keşfi süreci Şekil 11’de görüldüğü gibi tanımlanmıştır. İkinci yaklaşımda veri madenciliğine veritabanlarında bilgi keşfi süreci içerisinde yer verilmektedir. Geline aşamada veri madenciliği, metodolojik olarak standart hale getirmeyi amaçlayan endüstri standartları (SEMMA, CRISP-DM, vb.) tarafından veri madenciliği süreci olarak tanımlanmaktadır.



**Şekil 11: Veritabanlarında Bilgi Keşfi Aşamaları**

**Kaynak:** Fayyad ve diğerleri, 1996: 41.

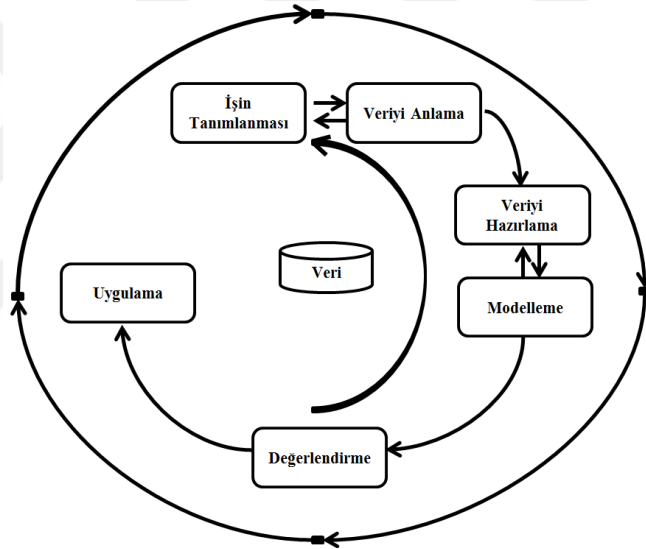
Veri madenciliği süreci SAS firması tarafından SEMMA ile standartlaştırılmıştır. SEMMA (Sample, Explore, Modify, Model, Assess) süreçlerin baş harflerinden oluşmaktadır. SEMMA süreci, seçilen veri madenciliği aracından bağımsız olarak veri madenciliği uygulaması sürecinde kullanıcılar rehberlik etmek üzere tasarlanmıştır. SEMMA, Şekil 12’de görüleceği üzere anlaşılması kolay bir süreç sunarak veri madenciliğinin organize edilmesine yardımcı olmaktadır (Azevedo ve Santos, 2008: 3).



**Şekil 12: SEMMA Aşamaları**

**Kaynak:** SAS, 2018: 1.

Bu çalışmada, NCR Systems Engineering Copenhagen (ABD ve Danimarka), Daimler Chrysler AG (Almanya), SPSS Inc. (ABD) ve OHRA Verzekerings en Bank Groep B.V. (Hollanda) şirketleri tarafından geliştirilmiş konsorsiyuma ait olan CRISP-DM (Cross Industry Standard Process Model for Data Mining) metodolojisi esas alınmıştır. Bahse konu süreç modeli, büyük veri madenciliği projelerini, daha az maliyetli, daha güvenilir, daha tekrarlanabilir, daha kolay yönetilebilir ve daha hızlı hale getirmeyi amaçlaması ve literatürde yer alan veri madenciliği projelerinde en yaygın tercih edilen süreç modeli (Nadali ve diğerleri, 2011: 161; Aytaç ve Bilge, 2013: 28) olması sebebiyle tercih edilmiştir. CRISP-DM metodolojisi, Şekil 13'te görüldüğü üzere işin tanımlanması, verinin tanımlanması, verinin hazırlanması, modelleme, değerlendirme ve uygulama olmak üzere 6 aşamadan meydana gelmektedir.



**Şekil 13:** CRISP-DM Metodolojisi

**Kaynak:** Chapman ve diğerleri, 2000: 10.

Veri madenciliği teknolojilerinin hazır çözüm şeklinde kurumlara uygulanabileceğini beklememek gerekmektedir. Veri madenciliği, çeşitli araçlar ve farklı yetkinliklere sahip teknik personel gerektiren karmaşık bir süreçtir. Kurumlarda uygulanacak veri madenciliği projesi iyi seçilen araç ve teknik personele bağlıdır. Teknik personel tarafından seçilen araçların belirlenen süreç modeli doğrultusunda uygulanması veri madenciliği projelerinin başarısında önemli rol oynamaktadır.

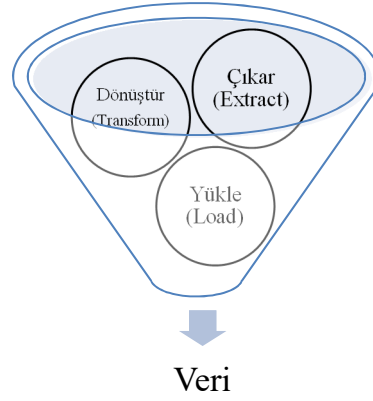
### **1.3.6.1. İşin Tanımlanması (Business Understanding)**

İşin tanımlanması (başlangıç) aşaması, proje hedefleri ve gereksinimlerin iş perspektifinden anlaşılması, daha sonra bu bilginin veri madenciliği problemi tanımına dönüştürülmesi ve hedeflere ulaşılması için tasarlanmış bir ön proje planına odaklanır (Wirth ve Hipp, 1998: 33). Kurumun hedeflerine yönelik alınacak kararlarda ulaşılması istenen hedeflerin iyi anlaşılması, belirlenen hedeflere ulaşmak için ise problemlerli sahaların tespiti önemlidir. Neyi ölçmeye veya öngörmeye çalıştığımıza karar vermek, amacı açıkça belirlemek ve amaca giden süreci geliştirmek gerekir.

### **1.3.6.2. Veriyi Anlama (Data Understanding)**

Veri madenciliğinde problem tespiti sonrası, çözümde kullanılmak üzere problemlerli sahaya ait verinin hazırlanması gerekmektedir. Veriyi anlama aşaması, ilk olarak verinin toplanması ile başlamaktadır. Teknik personelin, toplanan veri üzerinde bilgi sahibi olması, veri kalitesi problemlerinin tespiti ile süreç devam etmektedir (Shearer, 2000: 15). Veri anlama aşamasında veriyi toplama işlemleri için genel olarak aşağıda bahsi geçen ETL (Extract, Transform, Load) süreci takip edilmektedir.

Veri madenciliğinde kullanılacak veriler ilişkisel veritabanları (RDBMS-Relational Database Management System), veri ambarları (Datawarehouses), işlemsel veritabanları (Transactional Databases), uzaysal veritabanları (Spatial Databases), metin veritabanları, multimedya veritabanları ve internet olmak üzere farklı kaynaklardan elde edilebilir. ETL süreci olarak adlandırılan veri çıkarma (toplama)-taşıma-dönüştürme işlemleri ile Şekil 14'te görüleceği üzere bahsi geçen farklı kaynaklardaki ham verinin bir araya getirilerek kullanılabilir hale gelmesi sağlanır.



**Şekil 14:**ETL (Çıkar, Dönüştür, Yükle) Süreci

**Kaynak:** Han ve diğerleri, 2012: 93.

Çıkarma (Extract) süreciyle veri birçok farklı kaynak sistemden yardımcı araçlar vasıtasıyla alınmaktadır. Bu esnada kayıp ve gürültülü verilere yönelik işlemler gerçekleştirilir. Kayıp verinin ortadan kaldırılması için, kayıp verinin bulunduğu kaydı veritabanından çıkarmak ya da bu şekildeki kayıtları iptal etmek bir çözüm olabileceği gibi elle tamamlama, aynı bilgiyi girme, ortalama değer girme veya diğer değişkenler vasıtasıyla kayıp olan veriyi tahmin etme gibi çözümler de ihtiyaca göre tercih edilebilir. Gürültülü veriler ise yanlış girilme bile uygulamanın sağlığı açısından temizlenmesinde fayda olan verilerdir. Gürültülü veriler kutulama, sınırlar yardımıyla düzgülleştirme ya da kümeleme yöntemiyle temizlenebilir.

Dönüştürme (Transform) sürecinde farklı sistemlerden alınan veriler, kendi sistem yapımıza uygun hale getirilir. Bu işlem esnasında verinin temizlenerek kalitesinin artırılması, kayıp verinin ortadan kaldırılması, istatistiksel olarak sağlıklı hale getirilmesi ve birleştirilmesi amaçlanır. Veritabanının iyi dizayn edilmemesi aynı kayıtların farklı değişken isimleriyle girilmesine neden olabilir.

Yükleme (Load) aşamasında, hazırlanan veriler kendi sistemimize yüklenir. ETL süreci ile hedeflenen nihai amaç verinin en etkin şekilde kullanılmasıdır. Bu durumlarda eldeki verinin kullanımına karar verilen algoritmaya göre yeniden yapılandırılması gerekmektedir.

### 1.3.6.3. Veriyi Hazırlama (Data Preperation)

Eldeki veri yığınının küçük hacimli bir örnek elde edilerek veri madenciliği uygulamalarından etkili sonuçlar elde edilmeye çalışılır (Oğuzlar, 2003: 74). Veri madenciliği uygulamalarında elimizdeki veriler, kullanacağımız algoritma ile uyumlu olmayabilir. Bazı algoritmalar sadece sayısal değerler ile çalışırken, bazıları kategorik değer kullanabilir. Veri ön işleme aşaması, veri madenciliğinin en önemli aşamasını oluşturmaktadır. Modelleme aşamasında verinin yeniden düzenlenmesi ihtiyacı doğabilir. Veri ön işleme (indirgeme) yöntemleri aşağıda olduğu gibidir:

- Veri birleştirme veya veri küpü (Data Aggregation or Data Cube)
- Boyut indirgeme (Dimension Reduction)
- Veri Sıkıştırma (Data Compression)
- Kesikli hale getirme (Discretization)

Veri birleştirme veya veri küpü ile birleştirilmiş ya da çok değişkenli verinin farklı kaynaklardan elde edilmesinin ardından yapılacak analizlerde kullanılmak üzere birleştirilmiş hali kullanılır. Örneğin, kurumun farklı lokasyonlarda bulunan birimlerinden alınan harcama detaylarının tek bir tabloda toplanması neticesinde veri kümesinin hacmi küçülmekte, yapılacak analiz için veri kaybının önüne geçilmektedir. Veri küpleriyle çok değişkenli birleştirilmiş veri saklanır. Bahsi geçen satış tutarlarının satış yaptığı ürünler ve farklı satış noktaları verisi aynı veri küpü içinde saklanabilir. Verinin farklı kaynaklardan toplanması, veri uyumsuzluklarına neden olabilmektedir. Uyumsuzlukların giderilmesi amacıyla eldeki verinin 0.0-1.0 gibi aralıklara indirilmesi işleme normalizasyon işlemi adı verilir.

Normalizasyon dışında verinin miktarının azaltılması işleminin uygulanması gerekebilir. Bu işleme verilerde boyut indirgeme işlemi denir. Boyut indirgeme ile elde edilmek istenen bilgiye dönük olmayan gereksiz değişkenlerin veri kümesinden çıkarılması amaçlanır. Ürün satışına yönelik yapılacak veri madenciliği uygulaması için gereken veri kümesinden tüketicilerin telefon numaralarının çıkarılması boyut indirgemeye örnek olarak gösterilebilir.

Veri sıkıştırma, temel veri kümesinin veri şifreleme veya dönüşümü ile elde edilmesidir. Veri madenciliği çalışması yapılacak olan verinin büyük miktarda olması durumunda verinin kesikli ve kategorik veri haline dönüştürme işlemine tabii tutulması gerekebilir. Örneğin maaşların düşük, orta ve yüksek şeklinde dönüştürülmesi, yaş aralıklarının genç, orta ve yaşlı olarak kategorize edilmesi dönüştürme işlemine örnek olarak gösterilebilir. Kesikleştirme ile sürekli verinin kesikli değer aralıklarına dönüştürülmesiyle elde edilen kategorik değerler, orijinal veri değerlerinin yerine kullanılırlar. Veri kümesinde yer alan düşük düzeyli kavramların yüksek düzeyli kavramlarla değiştirilmesiyle veri indirgenmiş olur. Bu şekilde veri indirgemede detay bilgiler kayboluyorsa da genelleştirilmiş veriler daha anlamlı olacak, daha kolay yorumlanabilecek ve orijinal verilerden daha düşük hacim kaplayacaktır.

Verinin hazırlanması aşamasında, korpus olarak adlandırılan kitaplık oluşturma süreci, terim-doküman matrisinin oluşturulması ve bilginin çıkarımı esas süreçlerdir. Metin madenciliği sürecinde korpusun oluşturulması amacıyla verinin toplanması, sınıflandırma sürecinin ilk adımını oluşturmaktadır (Korde, 2012: 86). Bu aşamanın temel amacı çalışılan içerik ile ilgili hedeflenen tüm verinin toplanmasıdır. Metin dokümanları, ses kayıtları, XML dosyaları, e-postalar, web sayfaları vb. tüm dokümanlar toplandıktan sonra bilgisayarın işleyebileceği ortak bir forma yazılım araçları yardımıyla dönüştürülürler (Sharda ve diğerleri, 2014: 245). İnternet teknolojileri sayesinde pek çok süreç artık elektronik ortamına taşındığı ve erişimi kolay olduğu için, verinin bulunması kapsamında ciddi problemler yaşanmayacağı söylenebilir. Ancak bu noktada önemli olan veriyi model girişine uygun formata çevirmek, dolayısıyla bir takım ön işleme yöntemlerini (normalizasyon, temizleme, öznitelik çıkartımı, veri entegrasyonu gibi) uygulamaya özgü konuşlandırmaktır.

#### **1.3.6.3.1. Ön İşleme**

Metin madenciliği çalışmaları yapmak üzere toplanarak oluşturulan veri setinin (korpus) üzerinde ön işleme sürecinin uygulanması, hedeflenen bilginin elde edilmesi noktasında verinin iyi temsil edilmesi açısından en önemli aşamadır. Ön işleme ile kasıt verinin model girişine uygun hale getirilmesini sağlayan “öznitelik çıkartımı” işlemidir. Ön işleme aşaması uygulanmadan ham veri üzerinde ilgili algoritmaların çalıştırılması hacimli veri üzerinde çalışılmasından dolayı hem zaman kaybına hem de hatalı

sonuçların elde edilmesine sebep olmaktadır. İlâveten, kullanıcı giriş hataları ile gereksiz özniteliklerin elenmesini kapsayan öznitelik seçimi (feature selection) ön işleme kapsamında değerlendirilebilir.

Etkili bir metin madenciliği yapabilmek için detaylı bir ön işleme aşamasının uygulanması gerekir. Metin madenciliğinde yapılandırılmamış ham veri başta olmak üzere yapılandırılmış veriye de uygulanabilecek farklı ön işleme teknikleri bulunmaktadır. Yapılandırılmamış ham veri üzerinde metin madenciliği işlemleri gerçekleştirmek için yapılandırılmış veriden bilgi keşfi amacıyla yapılan klasik veri madenciliğinde kullanılan ön işleme süreçlerinden farklı ön işleme tekniklerinin uygulanması zorunludur. Eldeki belgeleri, verileri yapılandırmaya yarayan çok çeşitli metin madenciliği ön işleme teknikleri mevcuttur.

Sonuç olarak, farklı tekniklerin uygulanarak ön işleme yapılmasının amacı sınıflandırma işlemi öncesi verinin hazır hale getirilmesidir. Bu işlem için iki farklı yol bulunmaktadır. Bunlardan birisi makalelerin yazarlarının tespiti gibi yapılacak sınıflandırma işlemi öncesinde kullanılacak algoritmalara ve biçimsel çerçevelere göre ön işleme gerçekleştirmektir. İkinci yöntem ise doğal dil metinlerine de uygulanabilen karmaşık olguları analiz etme için kullanılan sınıflandırma şemaları, olasılık modelleri ve kural tabanlı sistem yaklaşımların ön işleme için kullanılmasıdır (Feldman ve Sanger, 2007: 57). Metin madenciliğinde ön işleme aşamalarını aşağıdaki şekilde sıralamak mümkündür (Kaşıkçı ve Gökçen, 2014: 26)

- Dönüştürme: Metin madenciliğinde birinci adım olarak ham verinin uygun formata dönüştürülmesi gerekmektedir. Bu amaçla, metinde olması halinde XML (Extensible Markup Language) ve HTML (Hyper Text Markup Language) gibi her türlü taglar (etiket kelimelerin) başta olmak üzere, amaç için kullanılmayacak diğer yazınsal veri çıkartılır.
- Tarama ve İşaretleme: İkinci aşamada harf olmayan karakterler boşlukla yer değiştirilerek bütün karakterler küçük harfe çevrilir.
- Durak Kelimelerin Çıkarılması: Veri setleri içerisinde çok geçen fakat sınıflandırma işlemlerinde ayırt edici özelliği bulunmayan edat, bağlaç ve zamir gibi kelimelerin veri setleri içerisinde çıkarılması işlemidir.

- Kök Bulma: Ön işlemenin son aşaması olarak aynı kökten gelen farklı ek almış kelimelerin veri seti içinde kelime sıklıklarına bakılarak aynı kelime olarak algılanması amacıyla kelime köklerinin bulunması gerekir.

Ayrıca kelimelerin sadece fiziksel yazınsal özelliklerinin dikkate alınmasının getireceği zorluklar da bulunmaktadır. Aynı kelimenin cümle içerisinde (context) farklı anlamlarda kullanılabilmesi (çok anlamlılık), farklı kelimelerin ise aynı anlamda (eş anlamlılık) kullanılabilmesi mümkündür (Feldman ve Sanger, 2007: 57).

Zemberek kütüphanesi; yazım denetimi, biçimbirimsel çözümlene, gövdeleme, sözcük üretimi, sözcük önerme, sadece ASCII karakterle yazılmış sözcükleri Türkçe karakterli hale çevirme ve heceleme gibi doğal dil işleme yöntemlerini sağlamaktadır. Zemberek kütüphanesinin yapısı dil yapı bilgisi ve doğal dil işleme-DDİ (natural language processing-NLP) olmak üzere iki ana bölümden oluşmaktadır. Kütüphanenin çekirdeğinde doğal dil işleme amaçlı özel algoritmalar bulunmaktadır. Çekirdek özellikle Türki dillere yönelik tasarlanmış olup, birkaç yardımcı mekanizma ve soyutlama sayesinde özel bir dil uygulaması içermemesi sağlanmıştır. Diller, gerçekleşmesi önceden belirlenmiş dilbilgisi kurallarına göre gerekli veriyi sağlamak zorundadır. Zemberek kütüphanesi de bu kuralları genel olarak kullanarak, son kullanıcıya kullanımı kolay yazılım erişim mekanizması üzerinden hizmetler sunmaktadır. Kullanılacak dile ait harfler ve alfabe bilgisi doğal dil işleme adımları öncesi zemberek kütüphanesi tarafından kontrol edilebilmesi amacıyla basit bir metin dosyasında, Türki dillerin temel yapıtaşları olan ekler kolay kullanılabilmesi amacıyla kümelenmiş (çoğul ekleri, sahiplik ekleri gibi) halde bir XML dosyasında tanımlanmış durumdadır (Akın ve Akın, 2007:1-2).

Zemberek kütüphanesinde her dil için ek içermeyen kök sözcükleri metin dosyasında tutulmaktadır. Aşağıdaki örneklerde görüleceği üzere Türkçeye yabancı dilden geçmiş olan kelimeler biçimindeki özel durumlarda tutulmaktadır.

*Aile → aileler → ailelar değil (Arapça kökenli)*

*Belde → beldeyi → beldei değil (Arapça kökenli)*

*Emir → emri → emiri değil (Ünlü düşmesi, Türkçe kuralı)*



tarih, sayı vb olma durumuna göre biçimbirimsel olarak modelleme işlemi bu bölümde gerçekleştirilmektedir.

3. Yapım eki almış sözcüklerin çözümlenmesi: Kök kelimelerin sonlarına gelerek sözcüğe yeni bir anlam katan yapım eki almış sözcüklerin çözümlenmesi işlemidir. Örneğin; bil kökünden türetilmiş bildiri (bil-dir-i) sözcüğü ile yeni bir sözcük türetilmiştir.
4. Ekfiillerin çözümlenmesi: İsim ve isim soylu sözcüklerin sonuna gelmek suretiyle cümlede yüklem görevinde kullanılmasını sağlayan ekfiillerin çözümlenmesi işlemidir. Örneğin “Baharda havalar güzeldir.” cümlesinde yer alan “dir” eki ekfiil görevini yerine getirerek isim soylu “güzel” sözcüğünün cümlede fiil olarak kullanılmasını sağlamıştır.
5. Aykırı durumların çözümlenmesi: Kısaltma ve özel isim türündeki sözcüklere ait bir liste hazırlanarak aykırı durumların çözümlenmektedir.
6. Sözlükte olmayan sözcüklerin çözümlenmesi: Biçimbirimsel çözümleme amacıyla elde edilmeye çalışılan sözcüğün kökünün belirlenen sözlükte olmaması durumunda ayrı bir yöntem geliştirme ihtiyacı bulunmaktadır.
7. Etiketleme: Biçimbirimsel çözümleme sonucunda elde edilen sonuçların düzgün bir şekilde isim-sıfat-fiil-ilgi eki-çoğul eki-mastar eki vb. şeklinde anlaşılacak şekilde etiketlenmesi gerekmektedir.
8. Sözcük üretimi: Biçimbirimsel çözümleme neticesinde elde edilen etiketleme dizisinin kullanılması suretiyle kökte belirtilen etiketlerin ek olarak uygun şekilde elde edilmesiyle sözcüğün türetilmesidir.

#### **1.3.6.3.2. Özellik Çıkarımı**

Özellik çıkarımı (tokenization) olarak adlandırılan bu süreçte, verinin sınıflandırma ve kümeleme algoritmalarınca işlenebilecek özellikleri içerecek küçük parçacıklara ayrılması gerekmektedir. Veri setinden tahmin edilmek ve elde edilmek istenen sonuca göre özellik çıkarma işlem gövdeleme, kelime, n-gram, hece ya da harf şekline gerçekleştirilebilir. Sınıflandırma ve kümeleme algoritmalarında özellik çıkarımı için veri kümelerinde yer alan sözcükler genellikle kelime kökleri ve ngramlar halinde

sayısallaştırıldıktan sonra terim olarak kullanılmaktadır (Biricik, 2011: 49). Literatürde geçen özellik çıkarımı teknikleri aşağıda açıklanmaktadır.

Birçok dilde kelimelere bağlı olarak farklı sözdizimsel formları bulunmaktadır. Örneğin, İngilizce’de genel olarak kelimelerin sonuna eklenen –s ekiyle kelimelerin çoğul halleri elde edilmektedir. Bu kelimelerin hepsi sözdizimsel olarak aynı kök formundaki kelimenin çeşitleridir. Bu tür aynı kökten türetilen kelimeler, metin madenciliğinde ayırt ediciliğin düşmesine sebep olmaktadır. Bu sebeple kelimenin köküyle uğraşmak gerekmektedir. Sözcüklerin köklerine indirgenmesine gövdeleme (stemming) denir. Kök, sözcüğün ekleri kaldırıldıktan sonra kalan kısımdır (Liu, 2007: 200). Türkçe sondan eklemeli bir dildir. Sözcüğün sonuna getirilen ekler çekim eki olarak adlandırılır ve getirildikleri kelimenin türüne göre, isim (ad) çekim ekleri (ev-de) ve fiil (eylem) çekim ekleri (ev-len) olmak üzere ikiye ayrılır.

Kelimelerin bütünü yerine, sadece kökünün alınması sayesinde aynı anlama gelen kelimelerinde ele alınması neticesinde metinler arasındaki benzerliğin daha iyi ifade edilmesi sağlanabilir. Tüm metinlerde en az bir kere geçen farklı kelime sayısı ile oluşan metinleri gösteren vektörün boyutu her kelimenin en az bir kere vektörde bulunması nedeniyle artmaktadır. Literatürde vektör boyutuyla ilgili problemin, kelimelerin kendileri yerine kelime kökleri kullanılarak giderilebileceği belirtilmektedir. Bu sayede aynı anlama gelen sözcüklerin vektörde yer alması sağlanarak metinler arası benzerlik daha iyi ifade edilebilir (Amasyalı ve diğerleri, 2010: 630).

N-gram, sözcüğün N-karakter dilimidir. Literatürde, bir dizide yer alan herhangi N-karakter kümesinin dilimlenmesidir. Sözcüğün bir dizi örtüşen N-Gramına bölünmesi olarak tanımlanır (Cavnar ve Trenkle, 1994: 162). N-Gramın 2-, 3-, 4- şeklinde farklı birkaç uzunluğu alınarak N-Gram tabanlı sınıflandırma işlemi gerçekleştirilebilir. Örneğin, “veri madenciliği” kelimelerinin N-gramlarını aşağıdaki şekilde çıkarmak mümkündür.

2-gramlar: ve-er-ri ma-ad-de-en-nc-ci-il-li-iğ-ği

3-gramlar: ver-eri mad-ade-den-enc-nci-cil-ili-liğ-iği

Daha büyük boyutlu 4-, 5- şeklinde N-gramlar çok sayıda frekans matrisleri üretmesi sebebiyle tercih edilmemektedir (Amasyalı ve diğerleri, 2010: 630). N-gram yönetimi

dileden bağımsız çalışmakta, dile ait dilbilgisi kurallarına ve sözcüğün yapısına ihtiyaç duymamaktadır.

Bir konuşma örneğinde, metinde yer alan farklı sözcüklerin sayısı, kendi sayısı üzerinden ampririk olarak belirlenebildiğinden yapılacak çalışmanın sonraki aşamasında ilgili sözcüklerin geçme sayısı ve sıklığının belirlenmesi gerektiği gözlemlenmiştir (Zipf, 1949: 23). Metin içerisindeki kelime, en sık rastlananan sözcük ise metinde P(1) frekansıyla görüntülenir, sonraki en sıklıkla oluşan kelime frekansı P(2) ile görüntülenir. Bu durum kelime frekansı (P), geçme sıklığı (r) olacak şekilde aşağıdaki şekilde formülleştirilerek Zipf kanunu olarak litartüre geçmiştir (Li, 1992: 1842).

$$P(r) = \frac{c}{r^\alpha}$$

Kullanım sıklıkları diğer söcüklere göre daha fazla olan “ve, veya” gibi kelimeler mecuttur. Doğal dillerin hepsinde var olan bu kelimeler, kullanıldığı metnin konusunun ve türünün belirlenmesinde ayırt edici olabilmektedir.

Metinler içerisinde sıklıkla geçen “daha, gibi, kendi, kadar, gerek, ile, sonra” vb. sözcükler ayırt ediciliğinin az olması dolayısıyla boyut sayısını artırabilir (Amasyalı ve diğerleri, 2010: 631). Bu tarzdaki kelimelerin metin içerisindeki geçme sıklığına göre belirlenerek filtrelenmesi, uygulanacak makine öğrenmesi algoritmasının daha doğru ve gerçekçi sonuçlar üretmesini sağlamaktadır.

Doküman matrisi (A) üzerinde Tekil Değer Ayırıştırma (Singular Value Decompositon) uygulanması suretiyle boyut sayısının azaltılması amaçlanmaktadır. Matrisin öz vektör çarpanlarına ayrılması amacıyla (m\*n’lik bir matris, m→metin sayısı, n→farklı kelime sayısı)  $A_{m \times n} = U_{m \times k} \cdot S_{k \times k} \cdot V^T_{n \times k}$  formülü uygulanır (Amasyalı ve diğerleri 2013: 10). Formülün uygulanması neticesinde S matrisinde, özvektörlerin özdeğerleri diagonalde büyükten küçüğe sıralanarak saklı anlam indeksleme uygulanmış olur (Amasyalı ve diğerleri, 2012: 98).

### 1.3.6.3.3. Yapısal Forma Dönüştürme

Bilgisayarların sözcükleri anlayabilmesi için metinlerin uygun bir biçime dönüştürülmesi gerekmektedir. Metinler, sayısal bir veri içermediği için ilk aşamada

bilgisayarlar tarafından işlenememektedir. Makine öğrenmesi algoritmalarının metinlere uygulanabilmesi için metinlerin vektörel uzayda ifade edilmesi gerekmektedir.

#### 1.3.6.3.4. Metinlerin İndekslenmesi (Ağırlıklandırma)

Yapısal forma dönüştürme aşamasında oluşturulacak vektör uzayı oluşturma öncesi indeksleme aşamasında metinde yer alan tüm kelimeler seçilen ağırlıklandırma yöntemi ile frekanslarına göre indekslenerek öznitelik vektörü oluşturulmaktadır. Ağırlıklandırma yöntemleri aşağıda olduğu gibidir. Bu teknikler içinde en çok tercih edilen teknik TF-IDF ağırlıklandırma yöntemidir (Liu ve diğerleri, 2012: 44). Sözcük metin içerisinde ne kadar sıklıkla geçmesi ilgili metnin belirlenen kategoriye atanmasını kolaylaştıracaktır. Tam tersi durumda, sözcük ne kadar fazla metinde geçerse ilgili sözcüğün ayırt etme özelliği düşecektir. Örnek metin üzerinden ağırlıklandırma yöntemleri açıklanmıştır.

Örnek Metin (M1): “Gölcük’e giden otobüs seferlerinin her saat yerine yarım saatte bir olmasını belediyeden talep ediyorum. Büyükşehir belediyesi olarak otobüs seferlerinde vatandaşın isteklerine önem verdiğiniz için teşekkür ederim”.

Sözlük={ çözüm\*, otobüs\*, belediye\*, öneri\*, sefer\*, talep\* }

Sözcüğün metin içerisinde geçmesi halinde ağırlık değeri 1, geçmemesi halinde 0 olacak şekilde gerçekleştirilen ağırlıklandırma yöntemi; “Bitsel (Boolean) ağırlıklandırma yöntemidir. Örnek metin (M1) içerisinde geçen sözcüklere ait oluşacak vektörün bitsel ağırlıklandırması aşağıdaki şekilde olacaktır.

M1=( 0, 1, 1, 0, 1, 1)

Metin içerisinde ilgili sözcüğün kaç defa geçtiğinin belirlenmesine dayanan yöntem . Frekans Ağırlıklandırma, TF (Term Frequency) ağırlıklandırma yöntemidir. Frekans ağırlıklandırma işleminde  $a_{ik} = f_{ik}$  formülü kullanılır. Örnek metne ait oluşan vektörün frekans ağırlıklandırması şu şekildedir:

M1= ( 0, 2, 2, 0, 2, 1)

Sözcüğün ilgili metinlerde geçme sayısının çok olması halinde belirlenen kategoriye tespit etmesinde önemli olduğu, az olması halinde ayırt ediciliğinin düşmesi prensibine

dayanan yöntem Terim frekansı x Ters Doküman frekansı ağırlıklandırma yöntemidir (TF-IDF, Term Frequency-Inverse Document Frequency). Frekans ağırlıklandırma işlemi  $a_{ik} = f_{ik} * \log\left(\frac{N}{n_i}\right)$  formülü ile hesaplanmaktadır. Terim frekansı x Ters Doküman Frekansı ağırlıklandırmada  $t$  özelliğinin  $d$  metnindeki TF-IDF değerini bulmak için aşağıdaki eşitlik kullanılmaktadır (Amasyalı ve diğerleri, 2010: 630).

$$TFIDF(t,d) = TF(t,d) \cdot \log\left(\frac{Tr}{Tr(t)}\right)$$

Geleneksel yöntemlerde kullanılan terim frekansı (TF) terimin ilgili metinde kaç kez geçtiğini belirtirken, “Terim Frekansı X Ters Doküman Frekansı” çok sıklıkla kullanılan terimlerin etkisinin azaltılması kullanılmaktadır (Çetin ve Amasyalı, 2013: 1).

Metinlerin uzunluğunun değişik olması, uzun ve kısa metinlerde geçen sözcük sayısının farklı olmasına, ayırt ediciliğin bu durumda düzgün tespit edilememesine sebep olmaktadır. Bu durumun önünde geçmek amacıyla sözcüklerin frekans değerlerinin yanı sıra metinlerin toplam sözcük sayısının da göz önünde bulundurulması esasıyla yapılan indeksleme işlemi Normalizasyon (TFC) olarak tanımlanmaktadır. Normalizasyon, ağırlıklandırmada kullanılan formül şu şekildedir;

$$a_{ik} = \frac{f_{ik} * \log\left(\frac{N}{n_i}\right)}{\sqrt{\sum_{j=1}^M \left[ \log(f_{ik}+1) * \log\left(\frac{N}{n_j}\right) \right]^2}}$$

Normalizasyon ağırlıklandırma yönteminde kullanılan frekans yerinde logaritma kullanılarak elde edilen aşağıdaki formül üzerinden hesaplanan ağırlıklandırma yöntemi Logaritmik Terim Bileşeni (LTC) Ağırlıklandırma yöntemidir.

$$a_{ik} = \frac{\log(f_{ik}+1) * \log\left(\frac{N}{n_i}\right)}{\sqrt{\sum_{j=1}^M \left[ \log(f_{ik}+1) * \log\left(\frac{N}{n_j}\right) \right]^2}}$$

### 1.3.6.3.5. Öznitelik Seçme

Veri kümesinin büyük olması, büyük miktarda karmaşık veri analizi gerektirmektedir. Bu tür analizlerin uzun zaman alması veri madenciliği analizlerini pratik ve

uygulanabilir olmaktan uzaklaştırabilir. İlaveten, eğitim setinde yer alan örnek sayısının fazla olması her zaman daha iyi sonuçlar elde edilebileceği anlamına gelmemektedir. Gerçek dünyayı (test uzayını) temsil etmeyen, olasılık dağılımı uyuşmayan ve birbirine benzeyen çok sayıda verinin modelin eğitim seviyesini arttırmayacaktır. Bu kapsamda niteliğin yanında nicel açıdan da kaliteli örnekleme yöntemlerinin eğitim aşamasında tercih edilmesi, model performansını test aşamasında arttıracaktır. Veri azaltma teknikleri sayesinde elde edilecek veri kümesi hacim olarak daha küçük olmasına rağmen, orijinal veri kümesinin tamamına yakınının özelliklerini korur. Bu sayede yapılacak analizler daha verimli olacaktır (Han ve diğerleri, 2012: 99). Veri kümeleri içinde yer alan tahmin gücü yüksek özelliklerin belirlenmesi ve karşılaştırmalı olarak kullanılabilmesi için, altyapısında veri kümesindeki ayırt ediciliği yüksek özellikleri tahmin eden yazılımlara ihtiyaç vardır. Öznitelik seçme yöntemleri, entropi yöntemiyle belirsizliği ölçen “Bilgi Kazancı”, “Kazanım Oranı Yöntemi”, “Bilgi Değeri Yöntemi” ve Ki-kare ( $\chi^2$ ) istatistiğini kullanarak özelliklerin taşıdığı bilginin birbirinden farklılığını ölçen “Ki-kare Özellik Seçimi Yöntemi” yaklaşımları olarak sıralanmaktadır.

Entropiye dayalı özellik seçimi bilgi kazancı yöntemidir. Entropi belirsizliğin ölçütü olarak tanımlanmaktadır. Olasılık-istatistik yöntemlerinde (Naive-Bayesian gibi) sıklıkla başvurulan bir metriktir. Bu kapsamda entropi değerinin yüksek olması belirsizliği, bir sistemdeki düzensizliğin ya da belirsizliğin ölçüsü ise entropiyi göstermektedir. Entropi 0 ve 1 aralığında değerler alırken, 1 değerine yaklaştıkça belirsizlik artmaktadır. Yüksek entropiye sahip veri daha çok bilgi içermektedir.  $D$  veri kümesinde yer alan “ $i$ ” sınıfının olasılığı  $p_i$  değeridir. Bu değer “ $i$ ” sınıfına düşen örnek sayısının tüm veri kümesindeki toplam örnek sayısına bölünmesiyle elde edilmektedir (Aktaş ve Kalıpsız, 2015: 74). Entropi, aşağıdaki formülle ifade edilmektedir:

$$E(D) = - \sum_{k=1}^m p_i \log_i(p_i)$$

Bilgi Kazancı yöntemi, en ayırt edici özelliği belirlemek amacıyla veri kümesindeki her bir özellik için ölçülmekte, karar ağaçlarında dallanmaların optimize edilmesinde kullanılmaktadır.  $D$  veri kümesi,  $n$  tane alt bölüme  $X$  özelliğinden bölünecekse  $X$ 'e ait bilgi kazancı aşağıdaki formülle hesaplanmaktadır. Formülde yer alan  $E(D)$  veri kümesinin  $X$  üzerinden bölünmeden önceki entropisini,  $E(D_i)$ ;  $i$  alt bölümünün  $X$

üzerinden bölünme olduktan sonraki entropisini ve  $p(D_i)$  ise  $i$  alt bölümünün  $X$  üzerinden bölünme olduktan sonraki olasılığını göstermektedir (Aktaş ve Kalıpsız, 2015: 74). Buna ilişkin formül ise şöyledir:

$$\text{Bilgi Kazancı } (D, X) = E(D) - \sum_{k=1}^n p(D_i)E(D_i)$$

Kazanım oranı yöntemi, bilgi kazancı yönteminin çok çeşitli değerlere sahip özellikleri seçme eğilimi nedeniyle sonuçlarının da sapmalı olacağından yola çıkılarak sapmayı azaltmak üzere geliştirilmiş bir yöntemdir. Sapmayı azaltmak için aşağıdaki formül kullanılarak bulunan bölünme bilgisinin elde edilmesinin ardından kazanım oranı bulunarak en yüksek kazanım oranına sahip özellikler belirlenmektedir.

$$\text{Bölünme Bilgisi } (S) = - \sum_{i=1}^v \left( \frac{|S_i|}{|S|} \right) \log_2 \left( \frac{|S_i|}{|S|} \right)$$

$$\text{Kazanım Oranı } (A) = \text{Bilgi Kazancı } (A) / \text{Bölünme Bilgisi } (S)$$

Veri kümesindeki özelliklerin tahminleyici gücünü hesaplamak üzere kullanılan istatistiksel yöntem bilgi değeri yöntemi adı verilmektedir. Bilgi değerini ölçmek için bir hipotezi desteklemek üzere kanıtları birleştirmek amacıyla kullanılan ve niceliksel bir yöntem olan “Kanıtsal Ağırlık” yönteminden yararlanılmaktadır. Bu yöntemle özellikler tahmin edilmek istenen sınıra göre analiz edilerek, konu olumlu ve olumsuz olmak üzere ikiye ayrılmaktadır (Aktaş ve Kalıpsız, 2015: 75).

Ki-kare testi ( $\chi^2$ ) Pearson’ın iki değişken arasındaki ilişkinin bağımlı veya bağımsız olduğunu belirlemeye yarayan ayırık veriler için kullanılan bir hipotez test yöntemidir. Sınıflandırma niteliği ile seçimde değerlendirilecek öznitelik arasındaki ilişkinin ki-kare istatistikleri hesaplanmaktadır. Ölçüm sonucu ki-kare dağılımına benzer olacaktır. Ki-kare testi sadece kesikli öznitelikler için uygulanabilmekte, gözlemlenen dağılım ile beklenen dağılım arasındaki mesafeye dayanmaktadır  $E_i$  beklenti,  $O_i$  gözlem sayısı olmak üzere aşağıdaki formül ile ifade edilmektedir. Öznitelik seçimi belirli bir öznitelik uzayı içerisinde yapılacak analiz ile ilgili en yüksek faydayı sağlayacak öznitelik alt kümesinin belirlenmesini sağlayan yöntemlerin genel adıdır (Var ve İnan, 2018: 327).

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

#### 1.3.6.4. Modelleme (Modelling)

Veri madenciliği sınıflandırma ve kümeleme gibi teknikleri iyi yapılandırılmış bir teknoloji alanı haline gelmiştir. Bu alanda, yeni algoritmalar üretilmeye, var olan algoritmaların geliştirilmesine yönelik fikirler üretilmeye devam edilmektedir. Veri madenciliği, veri setlerinden kullanacağı modele karar verme ya da modeli veriye uydurmayla ilgilenir. Kullanımına karar verilen modeller, tüm interaktif veriden bilgi elde etme süreci neticesinde elde edilecek kullanışlı bilgiyi etkiler ve kişinin karar vermesini gerektirir (Ralha ve Silva, 2012: 11645). Uygun modelleme tekniği seçilerek uygulanması sonrası sonuçları optimize etmek için model ayarları gözden geçirilir. Veri madenciliğinde problemi çözmek için farklı teknikler kullanılabilir. Bazı tekniklerin kullanılması, veri tipi için uygun olmayabilir. Bu durumda veri hazırlama aşamasına geri dönülerek gereksinimlere uygun hale getirilir (Larose, 2005: 7).

Modelleme işlemi esnasında başka algoritmalar denenerek farklı modeller kurulabilir ve zaman içerisinde veride meydana gelen değişiklikler neticesinde kurulan modelde değişiklik yapılması ihtiyacı doğabilir. Bu sebeple modelin yeni veriler göz önünde bulundurularak değerlendirilmesi gerekmektedir. Modelin değerlendirilmesi sürecinde doğrulanması ve anlaşılabilir olması önemlidir. Modelin doğruluğu basit geçerlilik (simple validation) testi ile değerlendirilebilir. Basit geçerlilik testinde verinin %5-%33'lük kısmı test verisi olarak belirlenerek diğer verinin üzerinde modelin öğrenimi gerçekleştirilir. Modelin doğruluğu yanında anlaşılabilir olması ve kararın ne için verildiği kurum açısından önemlidir (Gürbüz, 2009: 36).

Modelleme süreci, örnekten öğrenme şeklinde gerçekleştirilen denetimli (supervised) ve örneklerin gözlenmesi yoluyla aralarındaki benzerliklerden hareketle sınıfların tanımlanması denetimsiz (unsupervised) öğrenimin kullanıldığı modellere göre farklılık gösterir. Modeli doğrulamak için yapılan testler esnasında, modelin başlangıcında varsayılan verinin doğruluğunu kaybetme olasılığı bulunduğu için kurulan modelin doğruluk derecesi ne kadar yüksek olsa da gerçek dünyanın tam anlamıyla modellendiğini söylemek zordur (Akpınar, 2000: 13).

Modelleme tekniği seçilirken işin tanımlanmasından farklı olarak kullanılacak teknik net olarak belirlenir. Birden fazla tekniğin uygulanacak olması halinde varsayımlara göre model, her teknik için ayrı ayrı oluşturulmalıdır. Model kurulmadan önce kaliteli

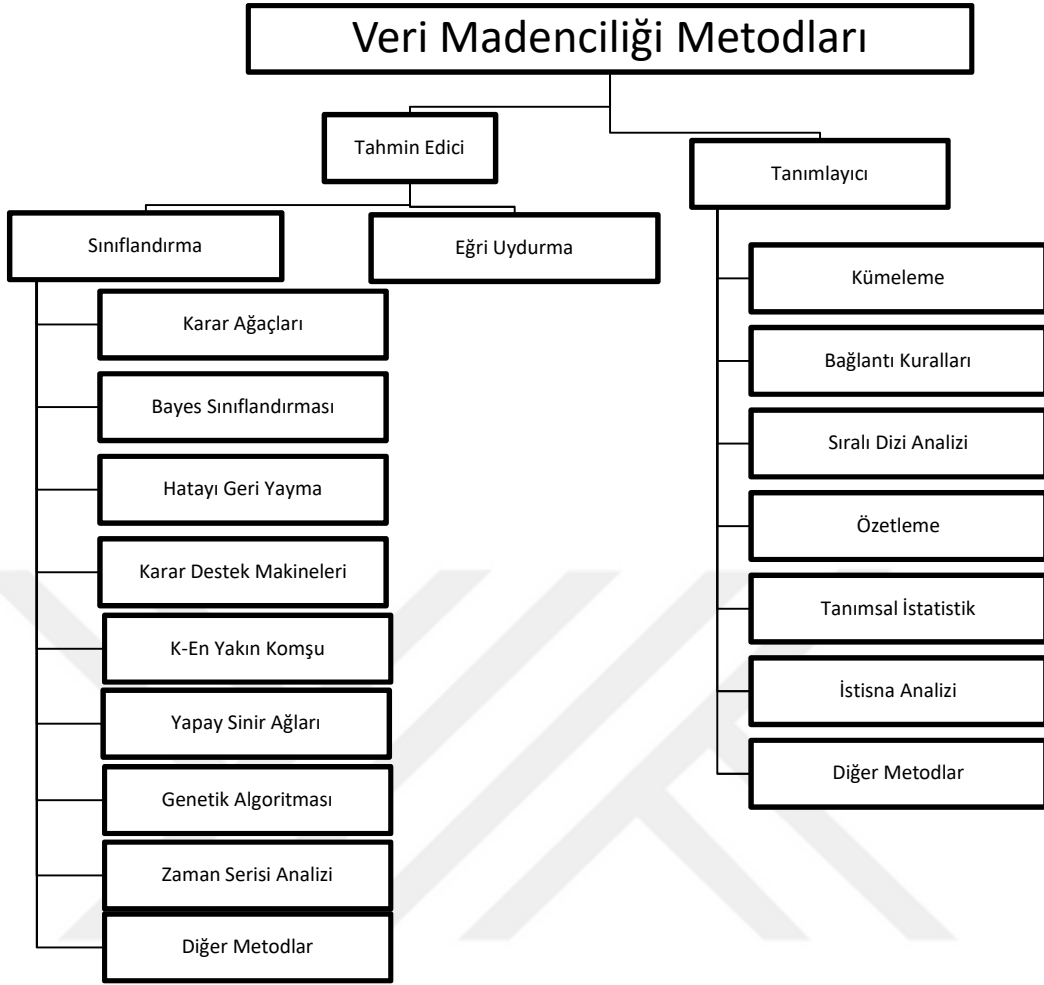
ve geçerli olabilmesi için birbirinden farklı veriler üzerinde sınama işlemine tabi tutulmalıdır. Sınama işlemi için oluşturulacak farklı veri kümelerinin hangi kriterlere göre yapılacağı önemlidir. Öğrenme kümesinin yanıltıcı sonuçlar üretebilmesi nedeniyle modelin kesinliğinin belirlenmesi için verinin modelleme öncesinde bölünerek bir deneme kümesi oluşturulması gerekir. Modelleme sürecinin sonunda karar verilen modelleme tekniği için parametreleri, deneme ve öğrenme kümeleri belirlenerek model ortaya çıkarılır (Çakır, 2008: 25).

#### **1.3.6.5. Değerlendirme (Evaluation)**

Kullanıma sunulmadan önce belirlenen hedeflere ulaşp ulaşılmadığı, bir veya daha fazla model denemek suretiyle değerlendirilerek karar verilir (Larose, 2005: 7). Modelin uygulama aşamasına geçmeden önce modelin eksiksiz olarak değerlendirilmesi ve iş amaçlarına uyup uymadığına emin olmak için model oluşturulana kadar yürütülen adımların tekrar gözden geçirilmesi büyük önem taşımaktadır. Değerlendirme aşamasında belirlenen hedeflere ulaşmak için oluşturulan model ve bu modelin elde edilmesi süreci detaylı bir şekilde gözden geçirilir. Buradaki temel amaç, analiz süresince gözden kaçan önemli bir noktanın var olup olmadığını belirlemektir. Değerlendirme aşaması sonucunda veri madenciliğinde uygulanacak adımların yürütülmesi ve belgelenmesi sağlanır.

#### **1.3.6.6. Uygulama (Deployment)**

Yaratılan modellerin kullanılması bu aşamada gerçekleştirilir. Modelin oluşturulması, veri madenciliği uygulamasının tamamlandığı anlamına gelmemektedir. Kurumun farklı bölümlerindeki süreçler üzerinde modelin çalıştırılması modelin doğruluğunu görmek açısından önemlidir (Larose, 2005: 7). Tahmin edici yaklaşımlar; sınıflama ve regresyon modelleri, tanımlayıcı modeller ise kümeleme, birliktelik kuralları ve ardışık zamanlı örüntü modelleridir. Veri madenciliğinde kullanılan modeller Şekil-16'da görülmektedir.



**Şekil 16:** Veri Madenciliği Metodları

Denetimsiz öğrenim denetimli öğrenimden eğitim verisinin sınıf etiketleri hazırda bulunmaması gerçeği ile farklılık göstermektedir. Denetimsiz öğrenim metotları hangi objeler bir sınıf olarak gruplandırılması gerektiğine dair karar vermektedir. K en yakın komşu, kendinden düzenlenen haritalar (SOM) ve veri kümeleme algoritmaları öngörü olarak adlandırılmaktadır. Regresyonda kullanılan teknikler; regresyon analizi, regresyon ağaçları, Sinir Ağları, k-en yakın komşuluk, box-jenkins ve genetik algoritmalarıdır (Tüzüntürk: 2010: 76). Sınıflandırma ve regresyon teknikleri arasındaki temel fark, tahmin edilen bağımlı değişkenin kategorik veya süreklilik gösteren bir değere sahip olmasıdır (Akpınar, 2000: 6).

Bankada kredi görevlisinin, krediye başvuran müşterilerden hangisinin güvenilir, hangisinin riskli olduğunu, tıbbi araştırmacının bir kanser hastasının üç adet tedaviden hangi en iyi tedaviyi seçmesi gerektiğini öğrenmek için veri setlerinde yapılacak

analizlere ihtiyacı vardır. Sınıflama daha çok bu şekildeki kategorik değerleri tahmin etmede kullanılırken, regresyon süreklilik gösteren değerlerin tahmin edilmesinde kullanılır (Han ve diğerleri, 2012: 329).

Sınıflama ve regresyon modellerinde kullanılan başlıca teknikler (Çalış ve diğerleri, 2014: 5); yapay sinir ağları (artificial neural networks), genetik algoritmalar (genetic algorithms), k-en yakın komşu (k-nearest neighbour), naive-bayes sınıflayıcısı, lojistik regresyon ve karar ağaçlarıdır (decision trees).



## **BÖLÜM 2: KAMU HİZMETLERİNDE BÜYÜK VERİ VE VERİ MADENCİLİĞİNİN GELİŞİMİ**

Üretilen büyük miktardaki sayısal veri önümüzdeki dönemde stratejik önemini koruyacaktır. Büyük verinin nitelikli kullanımı ve üretilen nitelikli bilgiye erişimin kolaylaşması, günümüzde 7 milyar civarında olan dünya nüfusunun yüzde 52'sinin kentlerde yaşaması sebebiyle önem arz etmektedir. Dünya genelinde şehirleşme oranının yükselme eğiliminde olması, konut, altyapı, ulaşım, eğitim, sağlık, güvenlik gibi hizmetlerin kaliteli ve sürdürülebilir biçimde sunulmasını giderek zorlaştırmaktadır (Kalkınma Bakanlığı, 2014: 30). Büyük veri ve analizi bağlamında veri madenciliği süreçlerinin sağlıklı bir şekilde kamu hizmetleri ve kamusal karar verme aşamalarına uygulanmasının kamu yöneticileri ve politika yapıcılara önemli avantajlar sunmaktadır.

Kamu kurumları ve özel sektörün karar almasını gerektiren durumlarda paydaşların en uygun kararı alabilmek için büyük veriden yararlanmak suretiyle veri madenciliğine benzer süreçleri işletmesi, kararları alması ve politikaları belirlemesi ihtiyacı bulunmaktadır (Bai ve Sarkis, 2016: 94). Kurumlar, ilk başta anlam yüklenemeyecek büyük veri yığınlarından doğru zamanda doğru kararlar almak suretiyle başarıya ulaşabilecek stratejiler üretmeyi istemektedir. Doğru kararların alınmasını sağlayacak nitelikli bilgi ise kurumların sahip olduğu işlenmemiş ham veride saklıdır. Ancak üretilen veri miktarı göz önünde bulundurulduğunda, nitelikli bilgiyi elde etmek gün geçtikçe zorlaşmaktadır.

Büyük veri ve buna dayalı teknolojilerin gelişmesi, sadece özel firmaların yatırımıyla değil, fakat devletin bu alanlara ilişkin düzenlemeleri ve kamu hizmetlerini üretmede söz konusu teknolojileri entegre etme çabası ile mümkün olabilir. Kamu kurumlarının elindeki veriyi paylaşımına açması, açık veri kültürünün gelişmesi, kamu hizmetlerinde büyük verinin kullanımı, veri madenciliği teknikleri ile geleceğe yönelik tahminlerin gerçekleştirilmesi ve karar vericilere yararlı bilgilerin sunulması ve bu doğrultudaki projelerin kamu ve özel sektör örgütleri tarafından işbirliğine dayalı olarak yürütülmesi gerekmektedir.

## **2.1. Kamuda Büyük Veri ve Veri Madenciliğinin Gelişimini Etkileyen Başlıca Yaklaşımlar**

1980'li yıllardan itibaren yönetsel reform dalgasının etkisiyle başlayıp günümüzde sosyal ağlar, büyük veri, bulut bilişim ve endüstri 4.0 gibi teknolojik gelişmelerle devam eden değişim dalgası kamu yönetimini etkilemeye devam etmektedir. Ülkelerin sürdürülebilir kalkınma hedefleri ve küresel rekabette güçlü olabilmesi ülkeyi oluşturan kamu, yerel, özel ve sivil toplum kuruluşları olmak üzere tüm paydaşların işbirliği yapması, aralarındaki ilişkileri bilişim çağının gereklerine göre yeniden yapılandırması halinde mümkündür. İnternet erişiminin artmasıyla dijital ağlar üzerinden üretilen bilgi, ekonomik ve toplumsal faydaya dönüşmektedir. Günümüzde sunulan ürünler ve hizmetler, toplanan veri üzerinde gerekli analiz ve tahminlerin gerçekleştirilmesi sayesinde vatandaş ve toplumun ihtiyacına özel bir şekilde sunulabilir hale gelmiştir.

Büyük veri ve veri madenciliği, kamu hizmetlerinde kurumlara önemli fırsatlar sunmaktadır. Kamu kurumları büyük veri sayesinde geleneksel veritabanı sistemleri ile depolayamayacağı büyüklükte yapılandırılmamış veriyi depolama imkânına sahip olmaktadır. Kamu kurumlarının hedeflerine ulaşabilmesi ve vatandaş memnuniyetini sağlayabilmesi, etkili kamu politikaları geliştirebilmesi ve yönetsel kararlar alabilmesi, sahip oldukları büyük verinin güvenilir, güncel ve doğru olmasıyla mümkündür.

Veri madenciliği uygulamaları sayesinde kamu örgütleri, elindeki veriyi değere dönüştürmek suretiyle işlemsel ve kurumsal maliyetlerini düşürme olanağına sahiptir. Veriden elde edeceği kazanımla karar vericiler bilgiye dayalı politikalar üretme, riskleri daha iyi yönetme ve kurumun geleceği adına daha isabetli kararlar alma fırsatı yakalamaktadır. Bu kapsamda içinde bulunduğumuz bilgi çağında geniş anlamda devlet organlarında ve dar anlamda kamu yönetiminde büyük veri teknolojisi ve veri madenciliği tekniklerinden faydalanılmasına katkı sağlayan yaklaşımlar değerlendirildiğinde yeni kamu işletmeciliği ve küreselleşme, bilgi toplumu, kanıta dayalı politika yapımı, açık yönetim ve dijital çağ yönetişimi yaklaşımları ön plana çıkmaktadır.

### 2.1.1. Küreselleşme ve Yeni Kamu İşletmeciliği

Küreselleşme, 20. yüzyılın sonu ve 21. yüzyılın başlarında hem ulusal ekonomiler üzerinde hem de ulus-devlet üzerindeki artan bir dış baskıyı beraberinde getirmiştir. Küreselleşme neticesinde ortaya çıkan değişim, kamu kurum ve kuruluşlarını yeniden yapılanmaya zorlamıştır. 1970’li yıllarda yaşanan refah devleti krizi başta İngiltere ve ABD olmak üzere bir dizi ülkenin yeni sağ olarak tanımlanan politikalara yönelmesine neden olmuştur.

Gelişmiş ekonomiye ve büyük kamu kuruluşlarına sahip ülkeler yeni sağ politikaların etkisiyle kamu harcamaları ve kırtasiyeciliğin azalacağı sisteme geçmektedir (Özer, 2012: 20). Küresel dönemin özellikleri ile örtüşen yeni anlayışta devlet piyasaya müdahale etmeyerek sosyal alanlardan çekilmelidir (Çımrın, 2009: 201). Yeni sağ düşüncesiyle birlikte devlet ve kamu yönetiminin özellikle ekonomik anlamdaki faaliyetlerinin daraltılarak kamu işletmeciliği mantığı çerçevesinde sunulması savunulmaktadır (Arslan, 2010: 25). Yeni sağ politikalarla serbest piyasa anlayışı öne çıkarak ekonomik olarak faaliyet alanı daralan devletin yerine piyasanın güçlenmesi hedeflenmiştir. Kaynakları verimsiz kullanan devlet, özelleştirme vasıtasıyla ekonomik faaliyetlerden çekilme yoluna gitmiştir. Yeni sağ politikalarla birlikte piyasa ekonomisinin kaynak kullanımındaki verimliliği ve etkinliği ön plana çıkartılmış, bu durum kamu yönetimi disiplininde ve pratiğinde önemli değişimleri tetiklemiştir.

Yeni sağ politikaların etkileriyle kamu yönetimi alanında 1980’li yıllardan itibaren yeni bir yönetsel reform dalgası başlamıştır. İzlenen reformların önceki dönemden farklı olduğunu vurgulamak için 1990’lı yılların başında “yeni kamu işletmeciliği” (new public management, NPM) ismi verilmiştir. Yeni kamu işletmeciliği (YKİ), örgütsel yapı ve süreçlerde özel sektör/işletme yönetim ve modellerinin kullanmasını esas alan bir yaklaşımdır. Dünyada YKİ fikirlerinin uygulanmasındaki ortak tema; kamu kurumlarının vatandaş bir müşteri olarak görmesi ve vatandaşla arasında pazarda yer alan mekanizmaya benzer bir ilişki kurmasıdır. Kamu yöneticileri hesap verebilirlik ve yüksek performans üzerine yoğunlaşmış ve bürokratik yeniden yapılandırmaya, kamu kurumlarını daha önce yerine getirdiği hizmetleri özelleştirmeye, üretkenliğini arttırmaya, alternatif ve yenilikçi hizmet yöntemleri bulmaya yönlendirmiştir. Aynı zamanda karar alma süreçlerini hızlandırmak adına yerele güç/yetki aktarımı

(decentralisation), yöneticilerin kurumun performans hedeflerinden sorumlu tutulması ve kurumların bu yönde yeniden yapılandırılması gündeme gelmiştir (Denhardt ve Denhardt, 2000: 550).

YKİ'nin isim babası olarak anılan Hood (1991: 4-5) kamu sektöründe büyük oranda rekabete geçiş, özel sektör yönetim tekniklerinin kamu yönetiminde de uygulaması, kaynak kullanımında verimlilik ve disiplin gibi özelliklerin altını çizmektedir. ABD'de işletmeciler reformlarının savunucusu olan Osborne ve Gaebler (1992) "Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector" isimli kitabıyla devletin tekele karşı rekabet etmesi, kamu kaynaklarının daha verimli kullanılması, karar alma ile mal ve hizmet sunumunun birbirinden ayrılarak siyasal karar alma görevinin devlet tarafından hizmet sunumunun ise özel sektör tarafından yerine getirilmesi, tek gelir kaynağı olarak vergileri gören devletin gelir elde etme amacıyla diğer kaynaklara da başvurması, girişimci olması, piyasadaki değişimden yararlanılması gibi hususları vurgulamaktadır.

Refah devleti anlayışında devletin ekonomiye doğrudan müdahalesi ile istenilen sonuçları doğurmamış, bu dönemde meydana gelen yeni problemler ve yeni toplumsal hareketler devlet tarafından yeterince çözümlenememiştir. Refah devleti, kıt kaynakların kötü kullanılmasından dolayı ekonomik sorunlar yaşamaya başlayarak toplumsal hizmetlerde kaliteyi temin edememiştir. YKİ, refah devletinin sorunlarına çözüm üretmek için özelleştirme yoluyla devletin ekonomik ve toplumsal alandan çekilmesini savunmaktadır. YKİ, devletin boşaltacağı ekonomik ve toplumsal alanların ülkeler tarafından piyasa mekanizmasının çarklarına göre piyasa odaklı bir şekilde ve piyasa tekniklerinden yararlanmak suretiyle işleyen kuruluşlar tarafından yerine getirilmesi gerektiğini savunmaktadır (Baltacı, 2004: 362-363).

YKİ, küreselleşme ve neo-liberal politikalara ait deregülasyon, ticarileşme ve özelleştirme ile birlikte somut olarak uygulama alanı bulmuştur. Deregülasyon ile devletin kural koruyucu rolünün azalması sağlanmıştır. Ticarete, finansal alanda ve dış yatırımlarda mümkün olduğunca serbestlik sağlanarak uluslararası alanda liberalleşme temin edilmiştir. Deregülasyon devletin kural koymadaki egemenliğini ortadan kaldırırken serbestleşme ise yabancı sermayenin etkinliğini artırmaktadır (Sönmez, 2002: 406). Bahse konu bu süreçler İkinci Dünya Savaşı sonrası dönemde ABD'de

başlayarak uygulamaya konulan Fordist sermaye birikimi ile Batı Avrupa'ya ihraç edilmiş ve yetmişli yıllara kadar yaygınlık kazandırılarak uygulanmaya devam edilmiştir.

YKİ yaklaşımı çerçevesinde, kamu yönetiminde işletme yönetimi değerlerinin egemen kılınmasıyla kamu kuruluşları verimsiz ve hantal yapılarından kurtarılması ve kamu örgütlerinde verimliliğin sağlanması yönünde reformlar amaçlanmıştır (Karcı, 2008: 41). Etkin ve verimli hizmet sunumunu sağlamak üzere piyasa sistemine göre çalışan kamu kurumlarının daha başarılı olacağı düşüncesi, YKİ'nin kamu yönetiminde piyasa mekanizmalarının kullanılmasını yönündeki temel ilkedir (Eren, 2003: 83). YKİ, 21. yüzyılda hemen hemen bütün dünya ülkelerini etkisi altına alarak özellikle ekonomik ve teknolojik alanda hızlı dönüşümlerin gerçekleşmesine kapı aralamıştır.

Teknolojik ve bilimsel anlamda yaşanan gelişmeler vatandaşın ihtiyaçlarında değişimi de beraberinde getirmiştir. İletişim alanındaki hızlı ilerleme, internetin yaptığı etkiyle birlikte kitle iletişim araçlarının hızla yaygınlaşmasına vatandaşın haber alma ve kamuoyu oluşturmaya katkıda bulunmuştur. Hizmet noktasında oluşan kamuoyu talebi iletişim araçları vasıtasıyla tüm dünya genelinde diğer ülke vatandaşları tarafından teknoloji ve bilimdeki gelişmeler, toplumun ihtiyaçlarının çeşitlenmesine ve değişmesine yol açmıştır. Kitle iletişim araçlarının hızla yaygınlaşması insanların haber alma ve kamuoyu oluşturma imkânlarını arttırmıştır.

Ekonomik anlamda verimliliğe yönelik sosyal gelişme, verimliliği artırmaya yönelik teknolojilerin kullanımı ve bu yönde motive işgücü, verimliliği sağlamaya inanmış yönetici ve yönetsel esneklik işletmecilik anlayışının temelini oluşturmaktadır. Değişim toplumun diğer alanlarını etkilemekle kalmamış, devletin yönetim yapısını da etkisi altına almıştır. Dünya çapında yaşanan ekonomik krizler ve yeni teknolojik gelişmeler devletin yapısı ve rolleri üzerinde etkili olmuştur (Arslan, 2010: 21). Bu bağlamda YKİ, piyasa benzeri mekanizmaların kamu yönetiminde kurulmasına, piyasada kullanılan yönetim araçlarının kamu kurumlarına adapte edilmesine ve böylece özel firmaların yatırım yaptığı yeni teknolojilerin hem kullanıcısı hem de rakibi olmasına neden olmuştur.

YKİ'nin arkasındaki piyasaya dayalı, işletmeci fikri destekleyen reform unsurları, teknoloji ve internete devletin adaptasyonunu ifade eden "e-devlet" yaklaşımı tarafından

desteklenmektedir. E-Devlet bilgi teknolojileri ile modernleşmeyi bir bütün şeklinde yorumlayarak YKİ ilkelerinin kamuda uygulanabilirliğini kolaylaştırmaktadır (Schedler ve Scharf, 2002: 775). YKİ vatandaş odaklı yaklaşımı ön plana çıkararak, hizmette belirli kalite standartlarının yakalanması ve vatandaşın taleplerine daha duyarlı yaklaşılarak kamu hizmetinde vatandaş memnuniyetinin geliştirilmesini istemektedir. Küreselleşmenin etkisiyle daha çok ön plana çıkan bilgi ve iletişim teknolojileri, YKİ'nin açtığı kapıdan girerek devletin işleyişini değiştirmiştir. Kamu yönetiminin işleyişi ve kamu hizmetlerini sunma yöntemlerinde meydana gelen değişim e-devletin gelişmesini sağlamıştır (Göküş, 2010: 209).

### **2.1.2. Bilgi Toplumu ve Kamu Yönetiminin Yenilikçi Teknolojilere Uyumu**

Günümüzde bilgi teknolojilerinde meydana gelen devrimi diğerlerinden ayıran husus bilgiyi temel unsur olarak belirlemesidir (Al, 2002: 69). Toplumun bilgi ve iletişim teknolojileri ile olan bağı çağımızda yadsınamaz bir olgu halini almış, içinde bulunduğumuz çağ “bilgi çağı”, bilgi çağının toplumsal düzeni ise “bilgi toplumu” olarak adlandırılmaktadır.

Teknoloji ile toplumun birbiri arasındaki ilişki açısından devlet teknolojik yenilikleri istediğinde başlatarak, istediğinde yasaklayarak, gerektiğinde ise yeniliklerin öncülüğünü üstlenerek belirleyici rol üstlenmektedir (Castells, 2013: 15). 1970’lerdeki kriz sonrası küresel bir ekonominin doğuşu, yeni enformasyon ve iletişim teknolojileri temelinde gerçekleşmiş olup küreselleşmenin kurumlarının ve kurallarının yerleşmesi 1990’larda mümkün olmuştur (Castells, 2013: 171-175). Bilgi toplumunda kamu yönetiminin yenilikçi teknolojilere uyum çabası da bu yıllardan itibaren yoğunluk kazanmıştır.

İnternet teknolojilerinin yaygınlaşmasıyla birlikte bireylerin siyasi, ekonomik kültürel veya yerel amaçlar için birbiriyle iletişim kurmaları ve kişilerin belirli toplumsal sorunlara çözüm üretmek üzere oluşturdukları örgütler “Ağ Toplumu” olarak tanımlanmaktadır (Göker ve Doğan, 2011: 175). Ağ “birbiriyle bağlantılı düğümler dizisi” olarak tanımlayan Castells’e göre ağlar sınırsız şekilde genişleyebilen, aynı iletişim altyapısını kullanmaları halinde yeni düğümlerle bütünleşebilen açık yapılardır (2013: 622-623). Bilgi ve iletişim teknolojilerinde 21.yüzyılın ilk çeyreğinden itibaren meydana gelen hızlı gelişmeler, yeni toplumsal yapıların ortaya çıkmasını sağlamıştır.

Yaşanan değişimle birlikte, vatandaşların birey olarak iletişim kurma ve enformasyon edinme biçimleri şekli ve niteliği değişmiştir (Doğan, 2011: 171). Bilgi toplumunu oluşturan bireyler, kamu kurumları, özel sektör ve sivil toplum örgütleri yaşanan dijital dönüşüme uygun niteliklere sahip olmak için kendilerini çağımızın gereklerine uygun şekilde yenilemeye çalışmaktadır. Bu kapsamda bilgiyi üreterek ihtiyaçlara cevap vermek adına karar verme süreçlerinde etkin şekilde kullanabilen kurumlar ön plana çıkmaktadır.

Bulduğumuz çağda meydana gelen hızlı değişim, kamu yönetiminde örgütlenme ve hizmet verme şekillerinin değişmesine ve kurumların yeniden yapılanmasına neden olmaktadır (Göküş, 2010: 584). Bu süreçte, yönetme ve yönetilme olguları arasında bulunan doğal engellerin, yoğun bir şekilde kullanılmaya başlayan bilgi ve bilişim teknolojilerinin aktif olarak kullanılmasıyla aşılabileceği düşünülmektedir (Kösecik ve Karkın, 2004: 98). Farklı medya platformlarının kullanımının artması yaygınlaşan veri analitiği, veri madenciliği sistemleri, vatandaşların proaktif bir şekilde daha fazla süreçlerin içinde yer almasına, kamu kurumları ve yöneticilerinden beklentilerini daha hızlı ifade edebilmelerine imkân sağlamaktadır.

Enformasyonel küresel ekonomide özel sektör ve kamu sektörünün birbirine bağlanması sebebiyle bu alandaki faaliyetlerin koordine edilmesine, yönetilmesine ihtiyaç bulunmaktadır. Finans, sigortacılık, emlakçılık, danışmanlık, güvenlik, bilgi toplama ve enformasyon hizmetleri ekonomik süreçlerin temelinde yer almakta olup sayılan hizmetler ülkelerin gelirlerinde önemli bir yer tutmaktadır (Castells, 2013: 508). Bilişim teknolojilerinin sahip olduğu potansiyel, kamu kurumlarının iş yapma şekillerini radikal biçimde değiştirmiştir. Kamuda yenileşme yolundaki arayışlar İngiltere ve ABD’de Thatcher ve Reagan dönemine kadar gitmektedir. Bu dönemlerden günümüze kadar çeşitli yönetim yaklaşımları ve yöntemleri kamu kurumları tarafından denenmiştir (Milner, 2000: 70).

Bilgi toplumu yaklaşımında önemli bir yer tutan Castells’in ağ toplumu paradigmasında dört temel özellik bulunmaktadır. Ağ toplumu paradigmasının ilk özelliği hammaddesinin enformasyon olmasıdır. İkinci özelliği yeni teknolojilerin yayılması, üçüncü özelliği enformasyon teknolojilerini kullanan sistemin ağ kurma yaklaşımını sergilemesi, dördüncü özelliği teknolojik paradigmanın yeniliği üretmeyi temel

almasıdır (Castells, 2013: 88-96). Bu özelliklerden yola çıkarak ağ toplumu, enformasyon ağlarından oluşan sosyal bir örüntü olarak karşımıza çıkmaktadır (Çımrın, 2011: 70).

Küresel sosyal hareketler sürecinde çok çeşitli hareketler elektronik iletişim sayesinde bir ağ yapısı oluşturmaktadır. İnternet, sosyal hareketler arasında ilişkilerin kurulması, çeşitli tartışmaların mekândan ve zamandan bağımsız bir biçimde gerçekleştirilmesi ve protestoların düzenlenmesi gibi birçok önemli organizasyonel öğeye kaynaklık etmektedir. Castells'e göre internete dayalı ağlar oluşturma süreci yalnızca bir örgütlenme ve mücadele aygıtı değil, siyasi bir varolma biçimi, yeni bir toplumsal etkileşim, harekete geçme ve kollektif karar alma biçimi anlamına da gelmektedir (Castells, 2008: 200-201).

Küreselleşme kamu kurumlarını, kültürleri ve kurumsal yapıları birbirine bağlayan enformasyon ağını farklı şekillerde harekete geçirmektedir. Bu noktadan harekete siyaset ve yönetim enformasyonel hale gelmektedir. Bilgi teknolojilerinin kullanımı ile birlikte zaman önemini yitirmekte, zamandan kaynaklanan sınırlamalar teknolojinin imkânları ile aşılabilmektedir (Castells, 2000 aktaran Henden ve Henden, 2005: 52). Kamuda zamana yayılan işlem ve prosüderler bilgi teknolojilerinin gelişmesi ve kamunun teknolojik olanaklardan yararlanmaya başlamasıyla birlikte daha hızlı yapılabilmektedir. Bilgi toplumuna doğru yaşanan dönüşümle birlikte bireyler, kamu kurumlarından talep ettiği hizmetlere bilgi teknolojileri altyapısı ve araçlarıyla ulaşabilmektedir.

Devletin iktidarı açısından bakıldığında teknolojinin güdümünde iletişim biçimlerinin çeşitlenmesi bilişim teknolojilerine dayalı iletişimi kontrol etmeyi imkânsız hale getirmiştir. Bilgi çağında devletler medya kontrolünü kaybetmiş durumdadır. Bilgi, uydular üzerinden internetin sunduğu imkânlarla kolay bir şekilde yayılmaktadır (Castells, 2008: 401-403). Kamu kurumlarının yenilikçi teknolojilere uyum sağlamak adına şeffaf, katılımcı ve hesap verebilir bir yönetim anlayışını benimsemeleri kaçınılmaz hale gelmiştir. Bilgi toplumunun getirdiği imkânlar sayesinde bireylerin kamu kurumundan beklentileri değişmiştir. Bu imkânlar, kamusal kararlarda söz sahibi olmak ve kamu politikalarına katılmak konusunda bireyler üzerinde motivasyon sağlamaktadır

### **2.1.3. Kanıta Dayalı Politika Yapımı: Kamu Politikaları ve Veri İlişkisine Yönelik Bir Çerçeve**

Dış dünyaya ilişkin tüm gerçekliğin deney ve gözleme dayalı sosyal bilim araştırmaları yoluyla keşfedilebileceği ve toplumun dönüştürülebileceği fikri pozitivist sosyal araştırmanın temelini oluşturmaktadır. 20. yüzyılda bilimsel bilgi kamusal sorunları çözmek amacıyla farklı sosyal bilim disiplinleri tarafından kullanılmıştır. Bu sayede karar vericiler ihtiyaç duydukları bilgiye erişebilmiştir.

Kanıta dayalı politika yapımı İngiltere’de 1997 yılındaki seçimlerde başa gelen İşçi Partisi hükümeti döneminde başta sağlık sektörü olmak üzere sosyal hizmetler, suçun önlenmesi ve eğitim gibi hizmet alanlarına dönük kamu politikalarının bilimsel verilere dayalı kanıtlar temel alınarak yeniden yapılandırılması sonucunda ortaya çıkmıştır. Kanıta dayalı politikalar kısa süre içinde ABD, Avustralya, Kanada ve Yeni Zelanda gibi Anglosakson ülkeleri, AB ülkeleri ve diğer gelişmiş ve gelişmekte olan ülkelerde başta sağlık olmak üzere çeşitli politika alanlarında kullanılmıştır (Köktaş ve Köseoğlu, 2015: 33).

Kanıta dayalı politika üretimi için kullanılacak bilginin kaliteli ve anlamsız veriden temizlenmiş olması sonuçlara doğrudan etki etmektedir. Veri analizi ve üretilen politikaların değerlendirilmesi için yetkin personel gerekmektedir. Kanıta dayalı olarak üretilen analiz ve politika tavsiyelerini hayata geçirecek siyasi otoritenin varlığı, kurum kültürünün bu konuda gerekli olgunluğa erişmiş olması önem arz etmektedir (Brian, 2010, 81-82). Kanıta dayalı politika üretimi, büyük miktarda verinin depolanması ve işlenmesini sağlayacak teknolojik araçların gelişmesine koşut olarak önem kazanmıştır. Özellikle veri analizi ve verinin temizlenmesine yönelik yapılacak çalışmalarda veri madenciliği uygulamaları üzerinden veri madenciliği süreçlerinin belirlenmiş kriterlere göre uygulanması kanıta dayalı politikaların üretilmesi için gerekli bilginin elde edilmesi açısından önemlidir.

Kanıta dayalı politika üretiminde iki hususa dikkat edilmesi gerekmektedir. Birinci husus; tecrübe, gelenek, uzman görüşü ve bilimsel bilgi gibi farklı temellere dayanan kanıtlardan hangisi/hangilerinin tercih edileceğidir. İkinci husus ise bu kanıtların ne şekilde elde edileceğidir (Köktaş ve Köseoğlu, 2015: 39). Politika yapım sürecinin iyileştirilmesi ve etkin bir şekilde işleyebilmesi için çeşitli kaynaklardan toplanmış

nitelikli bilgiye ihtiyaç vardır. Ulusal ve uluslararası arařtırmalar, mevcut istatistikler, önceki politikaların deęerlendirilmesinden elde edilen sonuçlar ve maliyet analizleri nitelikli bilgi kaynakları arasında sayılmaktadır (Strategic Policy Making Team Cabinet Office, 1999: 33). Günümüzde bu gibi veri, bilgi ve istatistikler, büyük veri olarak biliřim ortamlarında depolanabilmektedir. Veri analizi teknikleriyle gerekli ayıklamalar yapılarak uzmanların politika yapımında kullanabileceęi nitelikli bilgi üretilebilmektedir. Bu kapsamdaki teknolojilerin geliřmesi, kanıta dayalı politika üretimini de kolaylařtırmaktadır.

Kanıta dayalı politika üretiminde hangi kanıtların kullanılabilmesine iliřkin Weiss (2001: 288) kapsamlı bir açıklama yapmaktadır. Weiss'e göre tanımlayıcı, analitik, deęerlendirmeye dayalı ve ileriye dönük olmak üzere dört farklı kanıt türü bulunmaktadır. Tablo 4'te başlıca kanıt türleri ve bu verinin kullanım amacına yönelik örnekler verilmektedir.

**Tablo 4**  
**Kanıt Türleri ve Kullanım Amaçları**

<b>Kanıt Türü</b>	<b>Kullanım Amacı</b>
Tanımlayıcı veriler (descriptive data)	Mevcut olan durumun resmini vermeye yönelik göstergeleri içerir (iřsizlik oranı, ticaret hacmi, kiři başına düşen gelir, okullařma oranı vb.)
Analitik bulgular (analytic findings)	Suç ya da okullařma oranına etki eden dinamikleri belirlemeye yönelik çalıřmalar gibi mevcut duruma etki eden faktörleri arařtırmaya iliřkin bulgular
Deęerlendirmeye dayalı kanıtlar (evaluative evidence)	Yürürlükte olan mevcut program ve politikaların etkinlięine yönelik analizler
İleriye dönük politika analizleri (policy analytic forecasts)	Deęerlendirmeye dayalı kanıtları tamamlayıcı nitelikte olmakla birlikte geleceęe dönük politika tavsiyeleri geliřtirme

**Kaynak:** (Weiss, 2001: 288-289)'den tablolařtırılmıřtır.

Tanımlayıcı veriler ve analitik bulgular günümüzde büyük veri teknolojileri ile depolanabilmektedir. Politikaların etkinlięinin ölçülmesi ve ileriye dönük politika tahminlerinin geliřtirilebilmesi ise veri madencilięi teknikleri ile mümkündür. Kanıta dayalı kamu politikası yapımında elde edilen verinin depolanması büyük veri teknolojileri ile saęlanırken politikaya dayanak saęlayacak bilimsel bilgi veri madencilięi teknikleri ile elde edilecektir.

#### 2.1.4. Açık Yönetim

ABD’de Başkan Obama tarafından 2009 yılında yayımlanan memorandumun ardından şeffaflık, katılım ve işbirliğine dayanan, dijital dönüşümün getirdiği olanakların yönetim süreçlerine eklenmesini savunan açık yönetim (*open government*) politikaları gelişmiştir. Yönetimde açıklık yeni bir kavram olmamakla birlikte, günümüz açık yönetim politikalarının öncekilerden farkı şeffaflık, katılım ve işbirliğini içerecek biçimde daha bütüncül olması ve ayrıca bilişim teknolojileri, internet ve sosyal medya gibi araçların entegre edilmesiyle birlikte daha etkili bir biçimde kullanılmasıdır. Açık yönetimin bu hedefi, onu çevrimiçi işlem ve hizmet sunumunu önceleyen klasik e-devlet uygulamalarından farklılaştırmaktadır (Harrison ve diğerleri, 2012: 84).

Açık yönetim; yönetimin vatandaşa, medyaya ve özel sektöre açık hale getirilmesinin yanı sıra şeffaflığın, katılımın ve işbirliğinin sağlanabilmesi amacıyla kurumun içinin de açık hale getirilmesidir. Web 2.0 teknolojileri ile bahse konu açık yönetim hususlarının uygulanması daha kolay hale gelmiştir. Açık yönetim siyaset ve yönetimi vatandaşa ve özel sektöre açık hale getirmesi yönüyle e-devletten farklılaşmaktadır. E-devlet, farklı hizmet kullanıcıları ve paydaşlara aynı platform üzerinden kamu hizmetlerinin elektronik ortam vasıtasıyla sunulmasıdır.

Dünyada açık veri ve açık yönetime ilişkin ülkeler tarafından oluşturulmuş “Açık Yönetim Ortaklığı” bulunmaktadır. Açık Yönetim Ortaklığına katılım için ilk aşmada gerekli olan isteklerin katılımcı ülke tarafından karşılanması ve Açık Yönetim Ortaklığı Bildirisinin kabul edilmesi gerekmektedir. Şeffaf, hesap verebilir kamusal kararların alınması, vatandaşları ve sivil toplum kuruluşlarının karar alma süreçlerine katılımının sağlanması, yolsuzlukla mücadele edilmesi ve yenilikçi teknolojiler ile yönetişimin sağlanması için 2011 yılında kurulan “Açık Yönetim Ortaklığı’nın, 2017 itibariyle 70 üyesi bulunmaktadır. Ortaklık için ilk adım olan asgari yeterliliklerin karşılanması ile kastedilen, hükümetlerin dört temel alanda (Mali Şeffaflık, Bilgi Edinme Hakkı, Gelir ve Mal Bildirimi, Vatandaş Katılımı) belirlenmiş olan kriterlere uyum sağlamalarıdır (Ak, 2014: 51). Açık Yönetim Ortaklığı, kamu kurumları, özel sektör ve sivil toplum kuruluşları olarak paydaşları arasındaki bilgi paylaşımı ve işbirliğini kolaylaştıran bir ağdır. Bu alanlarda reform ve politika yapma süreçlerinden vatandaş etkinliğinin sağlanması ortaklığın temel şartları arasındadır. Bahse konu ortaklık için devletlerin

ulusal eylem planı geliştirme aşamalarında sivil toplum ile birlikte çalışması önem kazanmaktadır (TUSEV, 2015: 2). Açık yönetim ortaklığına katılım için asgari yeterliliklerin sağlanması durumunda kamu kurumları, özel sektör ve sivil toplum kuruluşları arasında açık veri paylaşımı sağlanmış olmaktadır. Aynı zamanda belirlenmiş olan mali şeffaflık, bilgi edinme hakkı, gelir ve mal bildirim, vatandaş katılımı unsurları bahse konu kamu, özel sektör, sivil toplum kuruluşları ile vatandaş arasında veri paylaşımına ve açık verinin kullanımına doğrudan etki etmektedir.

Ülkelerin geliştirdiği stratejiler ve eylem planlarına dayalı olarak uyguladığı açık yönetim politikaları, kamunun elindeki verinin farklı formatlarda özel ve bireysel kullanıcılarla paylaşılmasını sağlamaktadır. Farklı düzeylerde ve değişik hizmetleri üreten, işlemleri yapan ve kararlar veren devlet kurumlarının elindeki veriyi açması, büyük veri havuzunun daha da büyümesine imkân vermekte, veri madenciliği için ihtiyaç duyulan formata uygun nitelikli verinin kolayca ulaşılabilir olmasını sağlamaktadır.

#### **2.1.5. Dijital Çağ Yönetişimi**

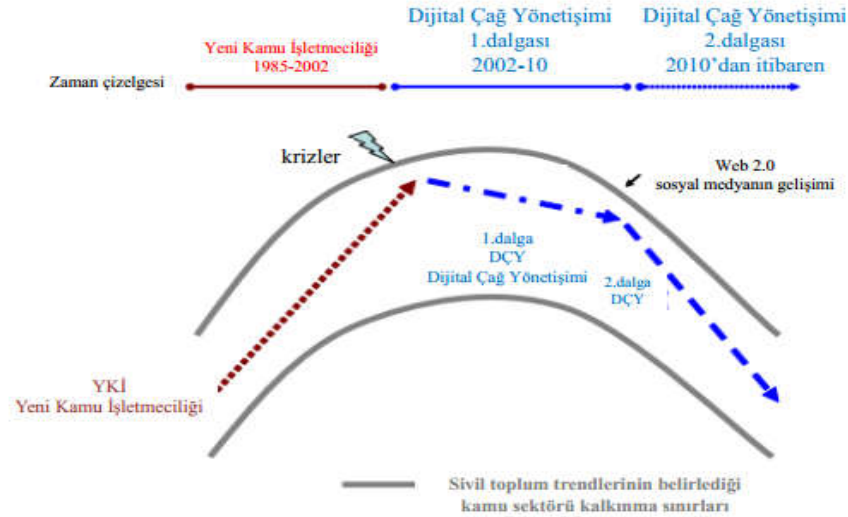
1970’li yıllarda görülen petrol krizindeki etkisiyle gündeme gelen ve 1990’lı yılların başında YKİ olarak adlandırılan hareket; ölçülebilir sonuçlar elde etmeyi, yetki devri ile yöneticilere daha fazla özgürlük sağlamayı, özelleştirme ve performans ölçümü gibi uygulamalarda etkinlik ve verimliliği artırmayı hedeflemiştir (Tuncer ve Usta, 2013: 181). Bununla birlikte YKİ, sadece işletmecilik yöntemlerinin kamu yönetimine uyarlanması anlamına gelmemektedir. Genel anlamda yönetim anlayışının kapsadığı alanın, dayandığı felsefenin ve işlevlerinin gözden geçirilmesi anlamına gelmektedir.

YKİ 1980’li yılların başında karşılaşılan sorunlara çözüm olarak devletin yeniden yapılanması, düzenleyici rol üstlenmesi, devletin tek karar verici olması yerine özel sektör ve sivil toplumun da sürecin içinde olduğu katılımcı yönetim anlayışını gündeme getirmiştir (Kalfa ve Ataay, 2008: 229). Global anlamda “yönetişim” olarak adlandırılan bu yaklaşım devletin küçültülmesi, şirket yönetimi yaklaşımının benimsenmesi, yeni kamu işletmeciliği, iyi yönetişim, toplumsal ve teknik yönetim ve kendi kendini yöneten yönetim ağı şeklinde altı ayrı anlam taşımaktadır (Rhodes, 1996: 652). YKİ ve yönetişim anlayışının etkileriyle kamu yöneticileri bürokrasiyi kontrol etmeye

odaklanmaktan ziyade hizmetlerin daha yalın ve giderek daha özelleştirilen bir yapıya evrilmesine odaklanmıştır (Denhardt ve Denhardt, 2000: 549).

Günümüzde YKİ ve yönetişimin içinde bulunduğu yeni sorunlar karşısında “Yeni Kamu Yönetişimi” (New Public Governance), “Yeni Kamu Hizmeti” (New Public Service) ve “Kamu Değeri Yönetimi” (Public Value Management) gibi alternatifler veya öneriler geliştirilmektedir (Sobacı ve Köseoğlu, 2015: 231; Dunleavy ve diğerleri, 2006: 468; Köseoğlu ve Tuncer, 2014: 145). “Dijital Çağ Yönetişimi” (Digital Era Governance) de YKİ sonrası dönemde bilişim teknolojilerinin e-devleti açacak biçimde kamu yönetiminde oluşturduğu dönüşüme işaret eden akımlardan biridir (Yavuz, 2015: 273).

Rekabet, ayrışma ve teşvik etme temaları üzerine kurulu olan YKİ dalgasının etkisi gelişmiş ülkelerde devam etmekle birlikte gelişmiş ülkelerde büyük ölçüde etkisini kaybetmiş durumdadır. YKİ sonrası yaşanan bilgi teknolojileri merkezli değişim, bir sonraki değişim dalgası için kritik öneme sahiptir. Bilgi teknolojileri ekseninde yaşanan bu değişim kamu örgütlerinin bütünsel ve ihtiyaç temelli olarak bütünleşmesini ve idari süreçlerin dijitalleşmesini içeren “dijital çağ yönetişimi” olarak adlandırılmaktadır (Dunleavy ve diğerleri, 2006: 467).



**Şekil 17:** Dijital Çağ Yönetişimi Birinci ve İkinci Dalgası

**Kaynak:** Dunleavy ve Margetts, 2010: 3.

Dijital çağ yönetişiminin birinci dalgası Şekil 17’de görüleceği üzere 2002-2010 yılları arasında gerçekleşmiştir. Dijital çağ yönetişimi birinci dalgasında YKİ anlayışında *parçalanmış süreçlerin yeniden entegrasyonu, hizmetlerin müşteri bakış açısıyla yeniden dizaynı ve kamu hizmetlerinin elektronik ortamda dijitalleştirilmesi* olmak üzere üç alana odaklanmıştır. Dijital çağ yönetişimi ikinci dalgası ise 2010 yılında Web 2.0’ın kamu yönetimi üzerinde etkisini göstermesiyle başlamıştır (Dunleavy ve Margetts, 2010: 3-5). İnternet kullanımının yaygınlaşmasıyla birlikte Web 2.0’ın etkisiyle sosyal medya gibi kullanıcılar tarafından değiştirilebilen içerik sayesinde kamu hizmetleri sunumu vatandaşın katılımıyla ortaklaşa etkileşim halinde yapılmaya başlanmıştır. Aynı zamanda teknolojik gelişmelerin etkisiyle vatandaş ve kamu kurumları arasındaki iletişim sürekli hale gelmiştir (Dunleavy ve Margetts, 2010: 13-15).

Dijital platformlar ve çevrimiçi içerikler internetin bulunduğu 1960’lı yılların sonundan itibaren kullanılmaya yaygınlaşmıştır. İnternet öncelikli olarak “Web 1.0” şeklinde tanımlanan, kullanıcıların sadece okuyucu olduğu tek taraflı iletişime imkân sağlayacak şekilde kullanılmaya başlanmıştır. Kullanıcının içerikte söz sahibi olduğu, kendi içeriğini oluşturabildiği karşılıklı iletişimin kurulmaya başlandığı 2000’li yıllarda “Web 2.0” teknolojileri kullanılarak dijital çağ yeni bir aşamaya geçmiştir. Son olarak içinde bulunduğumuz “Endüstri 4.0” ile birlikte dijitalleşen dünyada gündemde olan nesnelerin interneti, makine öğrenmesi, büyük veri, yapay zekâ, veri madenciliği gibi yenilikçi teknolojiler kamu yönetiminin dijitalleşmesine farklı bir boyut kazandırmıştır.

Dijital çağ yönetişiminin ikinci dalgası ile birlikte yönetişimin elektronik boyutu olarak e-yönetişim kavramı ile devlet dışı aktörlerin kamu politikaları süreçlerine katılması mümkün hale gelmiş ve toplumun tüm kesimleri devlet yönetiminde paydaş konumuna gelmiştir (Doğan ve Ustakara, 2013; 2). Bu sayede kamu politikalarının şekillendirilmesi, uygulanması ve denetlenmesinde birçok aktör kamu politikası yapımına daha kolay katılabilmıştır (Yavuz, 2015: 279).

Dijital çağ yönetişiminin YKİ yaklaşımı ile benzeştiğini öne süren “Girişimci (Entrepreneurship) Yaklaşımı Kuramı” ve farklı yönlerinin olduğuna yönelik “Katılımcı Yaklaşım” şeklinde literatürde farklı çalışmalar bulunmaktadır (Yavuz, 2015: 282-284). Yapılan çalışmaların bir kısmı YKİ paradigmasının eski gücünü kaybettiğini ve yerini dijital çağ yönetişimine bıraktığını savunmaktadır (Dunleavy ve diğerleri, 2006). Diğer

taraf ise dijital çağ yönetişiminin YKİ'nin bir uzantısı olduğunu, hizmet sunumunun zenginleştirilmesine, devlet politikasının geliştirilmesine yönelik şeffaflık, hesapverebilirlik ve katılım gibi hedeflerin yanında vatandaş devlet karşı daha güçlü konuma getirdiğini savunmaktadır (Longo, 2011: 42).

Bireylerin bilgi teknolojileri üzerinden demokratik katılımını sağlayan dijital çağ yönetişimine yukarıda bahsedildiği üzere olumlu yaklaşımlar bulunmakla birlikte, yönetimle devletin rolünün azaltılması neticesinde yönetsel işlemlerden kimin sorumlu olacağını belirsizleştireceği (Demirel, 2010: 69), devletin yerel, bölgesel ve ulusüstü kurumlarla yetki paylaşması nedeniyle otoritesini dağıtmak zorunda kalması, yönetime tüm paydaşların katılımı sonucunda yetki paylaşımı nedeniyle vatandaşın kamu yararının gözetilmesi konusunda devleti tam sorumlu tutamaması ve son olarak yeniden bütünleşme neticesinde kamu kurumlarının bilgi ve iletişim teknolojilerini kullanarak yönetim ve işleyiş yapısını değiştirmesi gerekeceğinden gerekli kurumsal dönüşümü başarılı bir şekilde gerçekleştiremeyeceğine dair olumsuz eleştiriler de bulunmaktadır (Yavuz, 2015: 285-287).

Sonuç olarak, kamu yönetiminde paradigma arayışlarının bir sonucu olarak YKİ'nin yetersiz kaldığı noktalarda bilgi teknolojilerinden yararlanmak üzerine kurulmuş olan dijital çağ yönetişimi yaklaşımı kamunun dönüşümüne etki etmeye devam etmektedir. Sayısal ortamda sunulan hizmetlere vatandaşlar tarafından herhangi bir kısıt olmaksızın erişilmesi neticesinde oluşan ve açık veri olarak adlandırılan kamu verisine kamu yöneticilerinin kamu politikaları üretme noktasında sıradan bir veri yığını şeklinde yaklaşmayıp dijital çağ yönetişimi ilkelerine uygun olarak büyük veri şeklinde yaklaşması, bu veri üzerinde gerekli veri madenciliği uygulamalarını yapması halinde mümkündür.

## **2.2. Kamuda Büyük Veri ve Veri Madenciliğinin Gelişimine Temel Oluşturan Politika ve Uygulamalar**

Kamu politikalarında ve hizmetlerinde 1980'li yıllardan itibaren idari reform dalgasının etkisiyle örgütsel yapı ve süreçlerde özel sektör model alınarak kamu ve vatandaş arasında yenilikçi hizmet yöntemleri uygulanmaya başlanmıştır. Kamu yöneticilerinin hesap verebilir ve kurumun performans hedeflerinden sorumlu hale

gelmesi, 1990'lı yıllarda internet ve veritabanları gibi yaşanan teknolojik yeniliklerin yanı sıra yeni kamu işletmeciliğinin de etkisiyle 2000'li yıllarda kamu kurumlarında verimliliğe yönelik çalışmaların hız kazanması kurumların veriden beklentisini değiştirmiştir. Kamu hizmetlerindeki veri akışının hızlanması kamu kurumlarını veriyi saklamak, işlemek ve geleceğe dönük veriden değerli bilgi etmek üzere arayışa sokmuştur.

Günümüzde verimsiz, sorumsuz, bürokratik ve hantal klasik kamu yönetiminden içsel verimliliğe, bölümlendirmeye, işlevselliğe, yukarıdan aşağıya yönetime, hiyerarşik geleneksel yapıdan örgütsel esnekliğe geçiş doğrultusunda yönetim anlayışında değişim söz konusudur (Senyucel, 2007: 132). Yönetim anlayışındaki değişim ihtiyacı doğrultusunda e-devlet, akıllı şehirler, açık veri ve Web 2.0 gibi büyük veri ve veri madenciliğine temel oluşturan temel politika ve uygulamalar geliştirilmiştir.

### **2.2.1. E-Devlet**

Sanayi toplumundan bilgi toplumuna geçiş, kamu yönetiminde e-devlet çalışmalarıyla kendisini göstermiştir. Özel sektör tarafından internet ve e-ticaretin kullanımındaki hızlı büyüme karşısında vatandaşlar, özel sektör kuruluşları ve sivil toplumun elektronik ortamda hizmet sunumu için kamu kurumları üzerinde bir baskı oluşturmalarına neden olmuştur. Bu girişim başta internet olmak üzere, bilişim teknolojileri üzerinden kamu hizmetini sunarak vatandaşı ve toplumu güçlendirmeyi amaçlamaktadır.

E-devlet en yalın haliyle bilgi ve hizmet alışverişinde bulunularak devlet hizmetlerinin elektronik ortamda sunulması için bilgi iletişim teknolojilerinin kullanılması olarak tanımlanabilir (Birleşmiş Milletler, 2017: 1). Küreselleşmeyle birlikte oluşan rekabet ortamında ulusal kaynakların vatandaşların değişen taleplerine uygun bir şekilde verimli olarak kullanılması ihtiyacı, özel sektör karşısında kamunun ağır işleyen mekanizmaları, e-devletin kamu yönetimi tarafından kabul görmesinde etkili olmuştur (Sobacı, 2012: 4). E-devlet, kamu hizmetlerinin elektronik ortam üzerinden sunulması, farklı hizmet kullanıcıları ve paydaşların aynı platformda buluşmasına olanak sağlamaktadır.

Geleneksel olarak kâğıt ortamında tutulan kayıtların elektronik ortama aktarılması ile hizmetlerin daha verimli ve hızlı bir şekilde vatandaşa ulaştırılması amaçlanmaktadır. E-devlet kapsamında verilen hizmetler neticesinde üretilen veriler günümüzde büyük

kapasitelere ulaşmış durumda olup bilgi teknolojilerinin bu şekilde kullanılmasının yanı sıra, yönetimde yenilikçi adımlar atmak adına açık veri politikalarının oluşturulması ile e-devletin kullanımı daha da genişlemiş durumdadır.

1990'lı yılların başında yerel yönetimler e-posta, web sayfası gibi hizmetler üzerinden bilgi ve hizmetlerini sunmaya başlamış, 1990'ların sonlarına gelindiğinde ise web tabanlı hizmetler e-devletin ayrılmaz bir parçası haline gelmiştir (Ho, 2002: 434). E-devlete geçişle birlikte kâğıda dayalı iletişimin yerini elektronik iletişim almaya başlamış, kamu yönetiminin veri yüklemesi yerine hizmet kullanıcılarının veri yüklemesi söz konusu olmuş, vatandaş temelli denetimin yerine otomatik veri güncellemesiyle denetime geçilmiş, tek tip hizmet yerini kişiselleştirilmiş/farklılaştırılmış hizmete bırakmış, kesintili hizmetin yerine bütünsel/kesintisiz hizmet sunmak mümkün hale gelmiş, işlem maliyetleri azalmış ve kapalı yönetimlerin yerin açık yönetimler almaya başlamıştır (Şener ve Paşayığıt; 2006: 3).

### **3.2.1.1. E-Devletin Aşamaları ve Boyutları**

E-devletin, bilgi teknolojilerini kullanarak veriyi yaymak, devlet ve vatandaş arasında iki yönlü iletişimi sağlamak, hizmetlerin sunumunda internetten yararlanmak ve devlet hizmetlerini dikey ve yatay olarak bütünleştirmek olmak üzere dört temel aşamada geliştiği söylenebilir. Web siteleri üzerinden devlet tarafından sağlanan tek yönlü veri akışı, zaman içinde vatandaşların kamu kurumlarına talep ve isteklerini e-posta ve faks gibi iletişim araçlarıyla iletmesine evrilmiştir. Kamu kurumlarının kendi aralarında ve hizmet kullanıcılarıyla olan iletişimi ve aynı hizmetlerin tekrarlanması yoluyla dikey bütünleşme, kamu hizmetlerinin bir e-devlet portalı aracılığıyla vatandaşlara sunulmasıyla da yatay bütünleşme gerçekleşmektedir (Chatfield ve Alhujran, 2009: 153-154).

Vatandaş odaklı hizmet sunumuna dayanan e-devletin “e-yönetim”, “e-hizmet”, “e-ticaret” ve “e-demokrasi” olmak üzere dört boyutu bulunmaktadır (Cook ve diğerleri, 2002: 3). Örgütlerin idari işlem ve süreçlerini kolaylaştırmak maksadıyla bilişim teknolojilerinden faydalanmasına “e-yönetim” adı verilir. E-yönetim, hem kamu kurumlarının kendi arasındaki hem de özel sektör ve vatandaşın devletle ilişkisine yeni bir boyut kazandırmaktadır (Durna ve Özel, 2008: 5). İnternet teknolojilerinin

vatandaşın günlük hayattaki yerini alması ve mobilitenin artması ile birlikte zaman ve mekân bağımlı gerçekleştirilen işlemler her an her yerden yapılabilir duruma gelmiştir. Yönetimsel süreçleri hızlandırmak, etkinliği arttırmak, kayıtları elektronik ortamda tutmak, farklı kurumlardaki bilgileri bütünleştirmek ve kurumlar arasındaki bilgi akışını sağlamak önem arz etmektedir. Kamu kurumları tarafından sunulan bilgi ve hizmetlerin elektronik ortamda sunulması “E-Hizmet” ile amaçlanmaktadır. Özel sektör ve kamu yönetimi tarafından gerçekleştirilen mal ve hizmet alımı, ihale, vergi ve faturalandırma gibi işlemleri için bilgi işlem teknolojilerinin kullanılması “E-Ticaret” olarak karşımıza çıkmaktadır. Vatandaşın yönetime katılarak alınacak kararlarda, seçim süreçlerinde ve politika yapımında aktif rol alabilmesini sağlamak amacıyla e-posta ve internet gibi bilişim teknolojilerini kullanması “E-Demokrasi” olarak tanımlanmaktadır.

Kamu kurumlarının vatandaştan bilgi alması, gerektiğinde vatandaşa bilgiyi sunması, kamusal faydanın, hesap verebilirliğin sağlanması ve birlikte karar almayı geliştirmek için e-yönetim, e-hizmet, e-demokrasi ve e-ticaret boyutlarından yararlanılmaktadır (Buss ve Redburn, 2006: 114-115).

### **3.2.1.2. E-Devlette İlişki Örüntüleri**

E-devlet projelerinin hizmetlerin çevrimiçi verilmesi, kâğıt bazlı olmayan yönetim, bilgi esaslı yönetim ve şeffaf yönetim olmak üzere en az dört hedefi bulunmaktadır. Bu dört hedef için merkezi ve yerel düzeyde, devlet-vatandaş (Government to Citizen-G2C), devlet-iş dünyası (Government to Business-G2B) ve kamu kurumu-kamu kurumu (Government to Government-G2G) olmak üzere üç temel ilişki örüntüsü ortaya çıkmaktadır (Sobacı, 2012: 9). Ayrıca devlet-kâr amacı gütmeyen kuruluş (Government to NGO-G2N) ile devlet-çalışan (Government to employee-G2E) ilişkilerini de bu örüntüye eklemek gerekir (Trajkovik, 2013: 21).

Devlet ve vatandaşın birbiri ile bilişim teknolojileri üzerinden elektronik ortamda iletişime geçtikleri, hizmetlerin elektronik ortamda sunulduğu e-devletin hizmet ve e-demokrasi boyutunu ilgilendiren model devlet-vatandaş modeli olarak tanımlanabilir. Günümüzde kamu kurumlarının web sitesi üzerinden bilgi vermesi, vatandaşın resmi form ya da taleplerini kurumu elektronik ortamda iletmesi, hizmet kalitesi ve vatandaşın tercihlerinin alınmasında e-oy yönteminin web üzerinden gerçekleştirilmesi devlet-vatandaş modeli çerçevesinde gerçekleştirilen hizmetlerdir (Sobacı, 2012: 10).

Kamu kurumları ile özel sektör arasındaki ilişki sadece devletin elindeki verinin internet üzerinden paylaşılması değil aynı zamanda devletin bilgi sunumu ve hizmet çeşitliliğini arttırmak maksadıyla hizmetlerini web temelli bilgi teknolojileri ile sunmasıdır (Guo ve Lu, 2007: 403). Devlet ve iş dünyası arasındaki bilgi ve verinin elektronik ortamda paylaşılması esasına dayanan devlet-iş modelinde ihale sistemleri üzerinden kamu alımlarının gerçekleştirilmesi, ödemelerin gerçekleştirilmesi, vergilendirme ve gelir beyanları gibi hizmetler sunulmaktadır.

Kamu kurumu-kamu kurumu modelinde, kamu kurumlarının birbirleri arasındaki hizmetleri elektronik ortamda gerçekleştirerek etkin, verimli ve en önemlisi zamandan tasarruf ederek ortak kamu kaynağının boşa harcanmasının önüne geçilmektedir.

E-devlet faaliyetleri, sadece kurum içindeki aktivitelerden ibaret olmayıp, elektronik posta web sayfaları ve sosyal medya gibi farklı ortamlarda yürütülen geniş bir yelpazeyi kapsamaktadır. Bu ortamlarda yoğun bir veri üretimi gerçekleşmektedir. (Yılmaz, 2005: 313). Sunulan hizmetler neticesinde toplanan veriler üzerinden gerekli analizler yapılarak yöneticilerin karar alma süreçlerinde yararlanılmaktadır. Bu süreçlerin gerçekleştirilmesinde veri madenciliği uygulamalarından da yararlanılmaktadır.

E-devlet üzerinden toplanan verinin analizi neticesinde, kolluk kuruluşlarının verimliliğini ve etkinliğini artırmak, karar vericilere zamanında destekleyici bilgi sağlamak için otomatik metin raporlarını analiz ederek metin analizi ve sınıflandırma tekniklerinin kullanılması (Ku ve Leroy, 2014: 534), kaçan suçluların yakalama alanlarında sınıflandırma modeli ile siber izleme ve özellikle kaçan suçluların sorgularının karşılaştırılmasında derinlemesine araştırmalar yapmak üzere öneriler sunulması (He ve diğerleri, 2013: 123), başvurular arasından kartel oluşumunun tespit edilerek önlenmesi (Ralha ve diğerleri, 2012: 11642), e-devletin yapılanması için arazi kayıtlarının sistematik ve güvenli bir şekilde güncellenerek devletin birimlerinin daha iyi kamu hizmeti vermesi (Hanumanthappa ve diğerleri, 2012: 208), kişisel vergilendirme sistemi (Brown ve diğerleri, 2006: 666) ve polislerin düz metin olarak hikaye formatında hazırladıkları raporlardan önemli bulguları çıkarmak için veri madenciliği temelli yaklaşımların kullanımı önerilmiştir (Keyvanpour ve diğerleri, 2011: 872). Tüm bu örnekler, e-devletin büyük veri ve veri madenciliği ile olan yakın ilişkisini göstermesi bakımından önemlidir.

### 2.2.2. Akıllı Şehirler

Bilgi teknolojilerinin sağladığı çözümleri odağına insanı alarak, paydaş kurumlarla birlikte katılımcı ve şeffaf olarak kişisel verinin gizliliğine uygun bir biçimde uygulayan şehirler “akıllı şehir” olarak tanımlanmaktadır. Akıllı şehirlerin akıllı insan, akıllı yaşam, akıllı ulaşım, akıllı ekonomi ve akıllı yönetim olmak üzere beş temel bileşeni bulunmaktadır (Deloitte, 2016: 12).

Literatürde akıllı şehirlere ait üzerinde anlaşılmış bir tanım bulunmamakla birlikte bütünsel bir bakış açısına ihtiyaç bulunmaktadır. Bu ihtiyaçtan hareketle akıllı şehir kavramını akıllı ekonomi, akıllı hareketlilik, akıllı yönetim, akıllı ortam, akıllı yaşam kalitesi ve akıllı insan olmak üzere altı kavramsal özellik üzerinden tanımlamak mümkündür (Vanolo, 2014: 887). Girişimcilik, esneklik, işgücü piyasası, uluslararası entegrasyon, pazar yeteneği ve dönüşümü “Akıllı Ekonomi”, yerel ve yerel erişilebilirlik, kullanılabilirlik, bilgi iletişim teknolojileri, modern ve sürdürülebilir taşıma sistemleri “Akıllı Hareketlilik”, katılımı ilgili karar verme süreçlerinde yönetim sistemlerinin şeffaflığı, kamu hizmetlerinin kaliteli ve kullanılabilir olması “Akıllı Yönetişim”, kaynakların yönetimi, doğal yaşam çevriminin korunması “Akıllı Ortam”, kültürel etkinlikler, eğitim imkânları, sağlıklı çevre, barınma, güvenlik ve turistik açıdan yaşam standartlarının oluşturulmuş olması “Akıllı Yaşam Kalitesi”, vatandaşa esneklik, yaratıcılık ve katılım hakkının sağlanması “Akıllı İnsan” kavramı olarak literatürde yer almaktadır. Kavramların detayına bakıldığında akıllı şehir kavramının farklı ihtiyaçlara çözüm üretmek üzere ortaya çıkan bir üst kavram olduğu görülmektedir.

“Akıllı Şehir” ifadesi günümüzde sıkça karşımıza çıkmakta ve dünyadaki şehirler tarafından içinde “akıllı” ifadesi geçen yeşil enerji, yapay zekâ ve nesnelerin interneti gibi teknoloji odaklı pekçok proje başlatılmaktadır. Bahse konu projelerin hemen hepsinde ciddi anlamda veri ihtiyacı bulunmaktadır. Akıllı şehirler sadece yirmibirinci yüzyılda karşımıza çıkan teknolojileri kullanan bir şehir olmanın ötesinde hizmet sunduğu toplumu geliştirmek üzere yeni teknolojilerden yararlanmak suretiyle iyi yönetim, ekonomik ve eğitim fırsatları ve toplum için sosyal eşitlik sağlamayı hedeflemektedir. Yüksek katma değerli bilişim teknolojilerin birbirine entegrasyonu sağlanmış olsa bile toplumun gizlilik, mahremiyet, sosyal medya ve veriler üzerinde

politikalar geliřtirmek üzere analitik iřlemler gerekleřtirilmesi ihtiyaı bulunmaktadır (Lehr, 2018: 3). Akıllı Őehir hedefine ulařmada kullanılacak biliřim teknolojileriyle elde edilecek verinin gerekli politikaların oluřturulmasında kullanılmasında veri madencilięi sũreci ve uygulamalarının etkili Őekilde kullanılması ihtiyaı bulunmaktadır.

Kullanıcıların tercihini yansıtan ve kiřisel bilgilerini tařıyan mobil cihazlar, konum tabanlı hizmetler ve ierięi zenginleřtirilmiř servisler ile akıllı Őehir özũmlerinde etkin bir rol üstlenmektedir. Yerel yönetimlerin sahip olduęu büyük hacimli veriler (büyük veri) üzerinden yapılan analizler sayesinde geleceęe dönük tahminler yapılabilmektedir. Akıllı Őehir projelerinde büyük verinin etkileri ve kullanımına yönelik alıřmalar incelendięinde (Akdamar, 2017b: 210-212); toplu tařıma sistemlerinin iyileřtirilmesi (Batty: 2013: 274 ), kaynakların kullanımında ve kentsel sorunlara özũm üretmek için kent planlamasına yönelik politikaların geliřtirilmesi (Bettencourt, 2014: 12), akıllı Őehirlerde karar destek sũrelerinde yararlanmak üzere bulut biliřim teknolojileri ile büyük veri iřleme ve analizi (Khan ve dięerleri, 2013: 381), kent altyapısına yönelik veri toplama, iřlemede nesnelerin interneti teknolojisinin katkıları (Liu ve dięerleri, 2017: 975; Zanella ve dięerleri, 2014: 22) Őeklinde alıřmalar yapıldıęı görũlmektedir. Akıllı Őehirlerde endũstri 4.0 teknolojilerinden elde edilecek büyük veri üzerinden yapılacak veri madencilięi uygulamaları ile vatandařlara hizmete yönelik alıřmalar gerekleřtirilebilmektedir.

Avustralya Fremantle Belediyesi akıllı Őehir planlaması alıřmaları kapsamında enerji ve su sistemleri entegrasyonunu gerekleřtirmek ve üzere veri madencilięi uygulamalarını kullanarak proje yürütmektedir (Kesayak, 2017: 1). Akıllı teknolojilerden faydalanma oranlarının giderek arttıęı günümüzde telefondan, televizyona, evlerden otomobillere kadar hayatımızın her alanında akıllı sistemlerin etkisi görũlmektedir. Akıllı sistemlerin ürettięi verinin iřlenerek karar alma sũrelerinde etkin olarak kullanılması için veri madencilięine ihtiya duyulmaktadır. Karar vericilere etkin karar vermelerini saęlayacak sonuçların sunulabilmesi veri madencilięi algoritmaları ile mümkündür.

Kentlerin yönetimi günümüzde hızla artan nüfus ve büyüyen yerleřim alanlarının ihtiyalarını karřılamak, yönetmek için önemlidir. Saęlıklı bir kent yönetimi bahse konu ihtiyaları karřılamak adına belirlenen hedeflerin etkili ve verimli bir Őekilde saęlanması

ile mümkündür (Yaşamış, 1991: 178). Faaliyet alanları çeşitli olmak üzere merkezinde insan olan akıllı şehirler, belediyeler ve telekom şirketleri gibi özel firmalar temelinde farklı paydaşlardan oluşmaktadır. İnsanlar akıllı şehrin hedef kitlesi, aynı zamanda kullanıcısı iken; belediyeler şehrin yöneticisi ve hizmetin vatandaşa sunucusu ve özel girişimler ise paydaşları tek çatı altında toplamak için gerekli altyapı ve bilgi iletişim teknolojilerinin sağlayıcısıdır.

Dünyadaki kaynakların hızlı bir şekilde tükendiği, mevcut şehir altyapılarının yetersiz kaldığı çağımızda yerel yönetimler tarafından vatandaşa sunulan hizmetlerin daha az maliyetle, daha verimli bir şekilde gerçekleştirilmesinde akıllı teknolojilerden faydalanılması önem arz etmektedir. Akıllı şehirler, doğal kaynakların, insan kaynağının ve finansal kaynakların etkin ve verimli bir şekilde kullanılmasını sağlayarak maliyet tasarrufu sağlamakta, belediyeçilik uygulamaları başta olmak üzere, elektrik, doğalgaz ve su alanlarında verimlilik artışını sağlayarak şehirlerde yaşam kalite ve standartlarını yükseltmektedir.

Los Angeles'ta akıllı ulaşım sistemleri ile duraklamalarda %35, kavşaklardaki beklemeelerde %20, seyahat süresinde %13 azalma ve bunlara bağlı olarak yakıt tüketiminde %12,5 azalma sağlanmıştır. Önceden Suç Tespiti Sistemi sayesinde hırsızlık oranlarında %19 düşüş sağlanabilmektedir. Akıllı park sistemiyle, 200.000 insanın yaşadığı bir Avrupa şehrinde, park yeri bulma zamanı %50 azaltılarak hem yaşam kalitesine katkıda bulunulabilmekte hem de yakıt tasarrufu sağlanabilmektedir. Park gelirleri artarken personel giderleri %5-10 arası düşürülebilmektedir. Akıllı sokak aydınlatması sistemiyle Oslo'daki elektrik tüketimi tasarruf oranı %70 olarak gerçekleşmiştir (Deloitte, 2016: 10). Dünyada verilen örneklerden de görüleceği üzere akıllı şehir uygulamaları ile kent yönetiminde tasarruf ve etkinlik sağlandığını ortaya koyan çok sayıda örnek bulunmaktadır.

Geleceğin şehirleri olarak görülen akıllı şehirler endüstri 4.0 devriminin getirdiği bulut bilişim, yeni nesil mobil internet altyapısı, nesnelerin interneti, RFID sensör teknolojilerinden aktif olarak yararlanmaktadır. Şehrin coğrafi planlanma ve tasarımı, ulaşım ihtiyaçları, mimari tasarımı ve enerji kullanımı gibi pek çok ihtiyacı bahse konu teknolojiler yardımıyla tasarlanmaktadır (Alkan, 2015: 74). Nesnelerin interneti akıllı şehirlerin yapılandırılmasına yönelik erişilebilir açık veriye ihtiyaç duymaktadır. Açık

sistemler sayesinde Endüstri 4.0 teknolojileri ve vatandaşlara yönelik yeni hizmetler geliştirilmesi ihtiyaç duyulan açık verinin üretilmesini sağlamaktadır (Ahlgren ve diğerleri, 2016: 52). Bilgisayarca okunabilen, herkesin kullanımına ve yeniden yayımlayabilmesine olanak sağlayan açık veri, kamu dahil farklı sektörlerdeki paydaşların verisini bireysel kullanıma açık hale getirerek rekabeti ve hizmet kalitesini artırmaktadır. Akıllı şehirlerin oluşturulması sürecinde açık veriden yararlanılmaktadır.

Akıllı şehirler konusunun pek çok yerel yönetim birimi tarafından teknoloji merkezli olarak ele alınması insan odağından uzaklaşılmasına sebep olmaktadır. Yaşanan dijital dönüşüm hem üretim süreçlerinin yeniden tanımlanması hem de ileri teknolojiyi kullanan ve toplumun rutinini değiştiren bir kavram olarak akıllı şehirlerin en önemli unsurlarından birisidir. Fakat akıllı şehirlerin dijital dönüşümden beklentisi endüstriyel anlamda özel sektörün beklentisinden farklıdır. Bu kapsamda özel sektör maliyet azalımı ve daha hızlı üretimi amaçlarken yerel yönetimler akıllı kentlerde büyük veri analitiği yoluyla sorunların tespit edilerek çözülmesi, daha güvenli bir toplum hayatı için sayısal görüntü yöntemlerinin kullanılması, dijital hasta kayıtları sayesinde her vatandaşa önleyici sağlık çözümlerinin sunulması ve sayısal bina modelleme ile enerji-verimli konutlar tasarlanmasında kullanılmaktadır. Geline nokta özel sektör ile yerel ihtiyaçların buluşması anlamında “faktör odaklı değil, inovasyon odaklı” dijital dönüşüm projelerinin hayata geçirilmesi önem arz etmektedir (Pehlivan, 2017: 1).

### **2.2.3. Açık Veri Yönetimi**

Dünyada yaşanan siyasal, ekonomik ve teknolojik gelişmelerin etkisiyle vatandaşın kamu kurumlarından olan beklentisi arttı. Vatandaşlar, kamudan aldığı hizmetlerin eskisine oranla daha kaliteli, hızlı ve herkes tarafından ulaşılabilir olmasını talep etmektedir. Vatandaşın beklentilerinin yanına teknolojik gelişmeler de eklendiğinde, kamu yönetiminde paradigmatik bir değişime gidilmesi kamunun daha katılımcı, saydam ve hesap verebilir olması zorunlu hale gelmiştir. Kamunun bu talepleri yerine getirebilmesi için “açık veri”, “açık erişim” ve “açık kaynak” kavramları gündeme gelmiştir (Bozkurt, 2014: 87). Bu değişim ihtiyacı doğrultusunda kamunun sahip olduğu verinin paydaşlar arasında teknolojinin imkânları kullanılarak paylaşılmaktadır.

Kanunların kendisine çizmiş olduğu görev tanımları çerçevesinde sunmuş olduğu hizmetler doğrultusunda veri üreten ve aynı zamanda bu hizmetler neticesinde

bireylerden veri toplayan kamu kurumları, sahip olduğu veriyi depolamakta, kullanmakta ve ihtiyaçlar doğrultusunda yetkisi dâhilinde paylaşmaktadır. Oluşan coğrafi veri, ulaşım ve trafik verisi, mali veri, adalet sistemi verisi ve akademik veri kamu kurumları, özel sektör, sivil toplum kuruluşları veya vatandaşlar tarafından ticari ya da ticari olmayan amaçlar için kullanılmakta ve “açık kamu verisi” (open data) olarak tanımlanmaktadır (Türkiye Bilişim Derneği, 2016: 3).

Açık veri, geniş kitlelerin erişebildiği, teknolojik ve kullanılabilirlik engeli olmayan kamu verisi olarak tanımlanmaktadır (Akdamar, 2017b: 45). Bir diğer tanıma göre açık veri, herkes tarafından serbestçe kullanılabilen, yeniden kullanılabilen ve yeniden dağıtılabilen veridir. Açık veriyi tanımlamak için “kullanılabilirlik ve erişim”, “yeniden kullanım ve yeniden dağıtım” ve “evrensel katılım” olmak üzere 3 temel özellik söz konusudur (Open Knowledge International, 2018: 1). Kullanılabilirlik ve erişim, verinin internet üzerinden uygun ve değiştirilebilir biçimde indirilebilmesini ifade etmektedir. Diğer veri kümeleriyle yeniden kullanım ve yeniden dağıtımına izin veren şartların sağlanması verinin yeniden kullanım ve yeniden dağıtım özelliğinin olmasına vurgu yapmaktadır. Veriyi herkesin yeniden kullanabilmesi ve yeniden dağıtabilmesi özelliği ise evrensel katılım ilkesine işaret etmektedir.

Özel sektör verisi, suç ve adalet istatistikleri, gözlem, eğitim, sağlık, ulaşım ve altyapı, finans ve sözleşmeler, yer/mekân verisi ve küresel kalkınma bilim ve araştırma ile toplumsal hareketlilik ve kalkınma verisi açık veri olarak tanımlanabilir (Bozkurt, 2014: 89). Bahse konu veri kümesi kamu kurumları başta olmak üzere özel sektör ve sivil toplum örgütlerinde bulunan ve internet üzerinden bireylerin ulaşabileceği türden veridir. Açık verinin bireylerin kullanımına sunulması, kamu kurumlarının şeffaflık ve hesap verebilirlik ilkelerinin yerine getirmesi açısından önemlidir.

Devlet kamu verisini açık ve ulaşılabilir hale getirerek kamu kurumları ve özel sektör için yeni fırsatlar yaratırken, diğer taraftan açık veriyi politika yapımında kullanmak için zorluklar yaşayabilmektedir (Zhang ve diğerleri, 2016: 6). Açık veriden bahsedebilmek için devletin veriyi internet üzerinden erişilebilir kılması, hem ham verinin hem de işlenmiş verinin bireyler ve kurumlar tarafından istenildiğinde erişilmek üzere gerekli lisansların sağlanmış olması, verinin temizlenmiş, birleştirilmiş ve görselleştirilmiş sunulması ve son olarak verinin elde edildiği birime geri bildirim

yapılarak veri kümelerinin gerektiğinde tartışılabilir olmasının sağlanması olmak üzere dört temel özelliğinin bulunması gerekmektedir (Zuiderwijk ve diğerleri, 2014: 30). Bu sayede toplanan açık veri, kamunun elinde büyük veri halini alarak stratejik bir kaynak haline gelmektedir.

#### **2.2.4. Web 2.0'a Geçiş**

Küreselleşmenin etkisinin tüm dünyada hissedildiği dönemde iletişim teknolojileri gelişerek yeni iletişim kanalları vatandaşın kullanımına sunulmuştur. İnternet sayesinde bilginin hızla yayılması önemli bir etki oluşturmuş asıl dönüşüm Web 2.0 teknolojilerinin kullanılmaya başlamasıyla yaşanmıştır (Karagöz, 2013: 131). İnternetin kullanılmaya başlandığı 1960'lı yıllardan bu yana yeni çevrimiçi teknolojiler ve yeni dijital platformlar hayatımıza girmeye devam etmektedir. 2000'li yıllardan itibaren internetin kullanıcı içerikli bir boyuta geçişiyle Web 2.0 teknolojisi kullanılmaya başlanmıştır (Demirtaş, 2012: 299).

Sosyal medya, insanların daha önce iletişim amacıyla kullandığı teknolojilerin sahip olmadığı etkileşim potansiyelini barındıran teknolojik altyapı olarak tanımlanabilir. Öncelikle bu potansiyelin ne olduğunu ve dünyanın sosyal medya üzerinden nelere sahip olduğunu açıklamak için sosyal medyanın olmadığı zamana gitmek gerekmektedir. İlk toplu iletişim araçlarının radyo, televizyon ve gazete olduğu dönemde yayıncı dinleyici kitlesinin kim ya da kimlerden oluşacağı konusunda bir etkiye sahip değildir. Telefonda bire bir etkileşim olanağı olmakla birlikte diğer iletişim araçlarıyla interaktif bir şekilde etkileşim kurmak mümkün değildir. İnternetin gelişimi ile birlikte kamu ve özel medya da değişmeye başlamıştır. Özelleştirilmiş forum sayfaları, sohbet odaları, gruba e-posta gönderme gibi teknolojiler toplu ve özel iletişim imkanlarını artırmıştır (Miller ve diğerleri, 2016: 2). İnternete bağlı bir şekilde ortalama bir kullanıcının katılım ve bilgisini paylaşabileceği yeni nesil internet uygulamalarına verilen bu yeni teknolojiler "Web 2.0" olarak tanımlanmaktadır. Web 2.0'la birlikte kullanıcılar herhangi bir içeriğe katkıda bulunma ve mevcut içeriği güncelleyerek yeniden yayınlama imkanına kavuşmuş, internet statik yapısından evrilerek dinamik bir yapıya dönüşmüştür (Thompson, 2008: 711).

Yeni iletişim teknolojileri sayesinde insanlar düşüncelerini tartışma kültürü çerçevesinde paylaşmaktadır. Sosyal medya, Web 2.0'ın etkileşimli yapısı üzerinden

kullanıcıların ürettiği içeriği değiştirmesine olanak veren uygulamalardır (Kaplan ve Haenlein, 2010: 61). Diğer bir tanımla kullanıcıları bir araya getirerek aralarındaki etkileşimi artıran internet üzerindeki platformların ortak adıdır (Vural ve diğerleri, 2010: 3348). Literatür incelendiğinde sosyal medya, çift taraflı ve eş zamanlı paylaşım sağlama özelliğiyle diğer iletişim kavramlarından ayrılmaktadır (Erkek, 2016: 142).

“Web 1.0” ve “Web 2.0” kullanım amacı, kullandıkları teknolojiler ve içerik açısından birbirinden ayrılmaktadır. Bununla birlikte Web 1.0 ve Web 2.0 arasındaki temel fark içerik oluşturma noktasındadır. Kullanıcılar Web 1.0'da az sayıda ve sadece tüketici/okuyucu olarak hareket ederken, Web 2.0'da içerik oluşturucu olabilmekte ve çok sayıda teknolojik yeniliği kullanabilmektedir. Web 2.0 kullanıcıların demokratik bir ortamda her türden içerik (metin, ses, video), etiket, yorum ve bağlantı paylaşabileceği etkileşimli bir ortamdır. Web 2.0 ile birlikte web sitelerinin kullanımında büyük bir artış yaşanmıştır. Günümüzde Web 1.0 siteleri kullanılmaya devam edilmekle birlikte kullanıcılar arkadaşları ile iletişim kurma/gruplara ayırma/paylaşım vb. özelliklerinden dolayı Web 2.0 sitelerini daha çok tercih etmektedir (Cormode ve Krishnamurthy, 2008: 1).

Web 2.0 ve sosyal medya uygulamaları kurumlarda politika üretme, karar alma ve sosyal kararlar alma noktasında yekili durumda bulunanların günlük hayatlarının bir parçası haline gelmiştir. Web 2.0 ve sosyal ağ uygulamaları sayesinde vatandaşlar eğitim, sağlık çevre ve diğer konulardaki kamu politikaları hakkındaki görüşlerini kurumlara iletme olanağı bularak politikaların belirlenmesinde aktif rol almaya başlamıştır. Merkezi yönetim kurumlarından yerel yönetimlere kadar kamu kurumları vatandaşla iletişime geçmek için Web 2.0 teknolojilerini aktif olarak kullanmaktadır. Daha önce masaüstü bilgisayarlar üzerinden sağlanan internet erişimi kablosuz ve akıllı cep telefonu teknolojileri sayesinde anlık olarak kullanılmaya başlanmıştır. Vatandaş-kamu arasında anlık iletişim imkanı sayesinde sosyal medyanın artan kullanımı da karşılıklı etkileşimi olumlu yönde etkilemiştir.

Vatandaşların iletişim kültürü üzerinde köklü değişimlere neden olan sosyal medya, kamu kurumlarının işleyişini ve paydaşları ile aralarındaki ilişkileri de etkilemektedir. Kamu kurumları temel görevlerini geleneksel yöntemlerle yerine getirmenin yanı sıra şeffaflık, hesap verebilirlik, kamu hizmetlerini geliştirmek, katılımı teşvik etmek ve

kurumlar arası işbirliğini güçlendirmek için sosyal medyanın potansiyelinden yararlanmak istemektedir (Sobacı ve diğerleri, 2015: 13). Bu sebeple, sadece günlük yaşam ve özel sektörde değil aynı zamanda yönetsel ve siyasal hayatta da yoğun bir şekilde tercih edilmeye başlanan sosyal medya zaman ve mekân sınırlaması olmaksızın kişi ve kurumlar arasında paylaşım ve etkileşimi esas alan bir iletişim aracı olarak dikkat çekmektedir.

Günümüzde görüldüğü üzere merkezi ve yerel yönetim birimleri karar alma ve toplumu ikna etmek üzere vatandaşla iletişim kurmak için sosyal medyayı etkin bir şekilde kullanmaktadır. Vatandaşlar etkin bir şekilde kullandığı sosyal medya araçları üzerinden kamu kurumlarının daha şeffaf, hesap verebilir, demokratik ve yönetim ilkeleri doğrultusunda hizmet vermesine katkıda bulunmaktadır (Erkek, 2016: 141).

Kamu kurumları her geçen gün daha fazla web hizmeti sunmaktadır. Gerçekleştirilen akademik çalışmalardan da anlaşılacağı üzere Web 2.0 teknolojileri ile birlikte kamusal politikalardan hizmetlerin sunulmasına kadar farklı pek çok alanda vatandaşlara yönelik hizmetler sunmaktadır.

### **BÖLÜM 3: TÜRKİYE’DE KAMU HİZMETLERİNDE BÜYÜK VERİ VE VERİ MADENCİLİĞİNİN KULLANIMI**

Devletler ve daha özel olarak kamu kurumları sahip olduğu “büyük veriyi” her geçen gün yeni bir alanda kullanmaktadır. Günümüzde hilekârlık tespiti ve engelleme, uygunsuz ödemelerin tespiti, gelir ve vergi toplama, kamu güvenliği, asayiş ve ülke güvenliği ve halk sağlığı ve toplum refahı kamu kurumlarının önem verdiği konuların başında gelmektedir. Veri madenciliği uygulamaları yardımıyla uygunsuz ödemeleri tespit etmek ve hilekarlık, israf ve suistimal ile mücadele etmek mümkündür. Aynı zamanda kullanılacak modeller yardımıyla kamu kaynağının boşa harcanmasına neden olan usülsüz ödemelerin tespit edilmesi mümkündür (SPSS, 2015: 8). Kamu kurumlarının kararlarını doğru, akıllı ve hızlı yapabilmesi, gerekli politikaları üretmesi sahip olduğu veriyi doğru amaçlara yönelik olarak değerlendirmesiyle doğru orantılıdır. Türkiye’de kamu hizmetlerinin çeşitliliği, farklı kamusal işlemler ve hizmetlerden üretilen verinin miktarı, veri işleme kapasitesindeki gelişmeler ve vatandaşların değişen talepleri birlikte düşünüldüğünde büyük veri üzerinden veri madenciliği uygulamalarına duyulan ihtiyaç da ortaya çıkmaktadır.

Kamu kurumları, özel sektörün aksine rekabet ortamından uzak ve çoğunlukla tekel niteliğinde hizmet sunmaları sebebiyle sahip oldukları veriyi kullanarak farklı kurumlara karşı üstünlük kurma gereksinimi duymamışlardır. Bilişim teknolojilerinin gelişmesi, veritabanı sistemlerinin artan kullanımıyla birlikte veri kullanım hacimlerinde meydana gelen artış, özel sektör benzeri rekabet anlayışının kamu kurumları arasında gelişmesi, şeffaflık, hesap verebilirlik ve daha nitelikli kamu hizmeti taleplerinin yükselmesi, müşteri anlayışının kamu sektörüne de yansması gibi gelişmeler kamu kurumlarını farklı arayışlar içerisine sokmuştur. Giderek büyüyen kamu kurumları ve karmaşıklaşan kamu hizmetleri göz önüne alındığında, yöneticilerin başarı için neleri bilmesi gerektiği konusunda sorunlar yaşanmaktadır. Başka bir ifadeyle, kamu kurumları geçmişten günümüze nasıl ve niçin çalıştığına dair bir yığın veriye sahip olmasına rağmen, yönetim eldeki bilgilerin kurum için ne kadar önemli olduğunun çoğu zaman farkında değildir (White, 2007: 256).

Bu bölümde genel olarak büyük veri ve veri madenciliğinin kamusal hizmetler ve işlemlerde nasıl ve hangi amaçlarla kullanıldığına ilişkin farklı örnekler

değerlendirildikten sonra, Türkiye’de merkezi idare ve yerel yönetimler açısından büyük veri ve veri madenciliği politikaları ile ilgili temel uygulamalar yasal metinler, politika dokümanları ve stratejik planlar incelenmek suretiyle analiz edilmektedir.

### 3.1. Veri Madenciliğinin Kullanım Alanları

Veri madenciliği kamu ve özel sektörde başta bankacılık, sigortacılık, telekomünikasyon, borsa, finans ve sağlık olmak üzere farklı alanlarda kullanılmaktadır (Çankırı ve diğerleri, 2009: 150). Perakendecilik sektöründe kredi kartları ve satış sistemlerinin kayıtları tutularak müşteri tercihleri, sepet analizi ile müşterilerin satın alma eğitimi, satış tahminleri ile stok, mağaza yerleşimi ve promosyonların planlanması için kullanılmaktadır. Bankacılık sektöründe müşterilerin sınıflandırılması sayesinde hangi müşteriye hangi ürünlerin sunulacağı tahmini neticesinde bankanın kârlılığının artırılması, kart kullanımına yönelik düzenlenecek bağlılık promosyonlarının hedef kitlesinin tahmini ile sunulacak özel hizmetler ve indirimler sayesinde müşteri kaybının en aza indirilmesi ve kazancın maksimize edilmesi, geçmişteki yapılan işlemlerin incelenmesi sayesinde dolandırıcılığın önüne geçilmesi için veri madenciliği uygulamalarından faydalanılmaktadır. Telekomünikasyon sektöründe artan rekabetle başa çıkmak, mevcut müşterileri korumak ve yenilerini elde etmek, müşteri sadakatini belirlemek amacıyla çağrı detayları veri madenciliği algoritmaları ile analiz edilerek müşterilere cazip fiyat ve kampanyalar sunmaya çalışılmaktadır (Rygielski ve diğerleri, 2002: 488). Veri madenciliğinin başlıca hangi sektörlerde ve ne amaçla kullanıldığı Tablo 5’te gösterilmektedir.

**Tablo 5**  
**Veri Madenciliği Kullanım Alanları**

<b>Kullanım Alanları</b>	<b>Kullanım Amacı</b>
<b>Bankacılık</b>	Risk analizi ve usulsüzlük tespiti, müşteri kazanma, kredi kartı dolandırıcılıklarının tespiti, risk analizi ve yönetimi, usulsüzlük tespiti, poliçe fiyatı tespiti.
<b>Pazarlama</b>	Çapraz satış analizleri, müşteri segmentasyonu, müşteri değerlendirme, müşteri ilişkileri yönetimi, satış tahminleri, pazarlama kampanyaları.
<b>Bilişim</b>	Görüntü ve metin işleme, web siteleri analizi, sosyal medya analizi.
<b>Sigortacılık</b>	Müşteri kaybı sebeplerinin belirlenmesi, riskli müşteri profilinin belirlenmesi.
<b>Telekomünikasyon</b>	Hile tespiti, hatların yoğunluk tahminleri, kalite iyileştirme

	analizleri, hatların kullanım oranlarının tahmini.
<b>Borsa</b>	Hisse senedi fiyat tahmini, genel piyasa analizleri, Pazar segmentasyonu.
<b>Tıp</b>	Tıbbi teşhis, uygun tedavi süreçleri, ürün geliştirme, sonuçların tahmini.
<b>Bilim ve mühendislik</b>	Bilimsel ve teknik problemlerin çözülmesi.
<b>Endüstri</b>	Kalite, kontrol, lojistik.
<b>Perakendecilik</b>	Satış veri analizi, alış-veriş sepeti analizi, tedarik ve mağaza yerleşimi.

**Kaynak:** Tekerek, 2011: 2.

Bankaların sahip olduğu merkezi veritabanı ve uygulamaları üzerinden elde ettikleri verilere dayalı olarak iş ve stratejik kararların oluşturulması, müşteri profillerinin çıkarılması, ihtiyaçlarının anlaşılıp kişiye özel hizmetlerin sunulmasında “Veri Ambarı”, “Veri Madenciliği” ve “Müşteri İlişki Yönetimi” uygulamalarını içeren sistemlerden faydalanılmaktadır (Halkbank, 2013: 13).

Kamudaki altyapı yatırımları ile ülkelerin gelir düzeyleri arasındaki ilişki, ülkelerin yatırım eğilimleri ve yatırım büyüklükleri veri madenciliği yöntemleri ile değerlendirmeye tabi tutulmaktadır (Koyuncugil, 2007: 2). Sosyal ağlar üzerindeki verinin etkili bir şekilde işlenerek takip edilmesi, veri madenciliği çalışmalarının yapılması ile mümkün olmaktadır. Veri madenciliği çalışmaları istatistik ve uygulama alanı olarak bilgisayar bilimlerini kapsamakla birlikte, sosyal ağlar üzerinde oluşan veriler, davranış bilimleri, toplum bilimleri veya işletme gibi çok farklı disiplinlerin de ilgisini çekmektedir (Şeker, 2015: 38).

Tıp alanındaki verinin büyüklüğü ve hayati önem taşıması, veri madenciliği uygulamalarından elde edilecek sonuçları önemli hale getirmektedir. Kronik hastalıkların hem tıbbi açıdan hem de hastane kaynak ve maliyetleri açısından ele alınarak doğru yönetilmesi için bilgi sistemleri üzerinde çalıştırılabilecek klasik sorgulama yöntemleri yetersiz kalmaktadır. Tetkik ve tedavilerdeki hilelerin tespit edilmesi, hasta durumunun takibi ve tahmini, hastane maliyetlerinin tahmini, genetik bozuklukların tespiti, ölüm oranları ve salgın hastalıkların tahmin edilmesinde veri madenciliği uygulamalarına ihtiyaç duyulmaktadır (Köktürk ve diğerleri, 2009: 23).

Dolgun’un da ifade ettiği gibi, günümüzde ham veri zengini nitelikli bilgi fakiri olan kurumların başarılı olabilmesi ve başarılarını sürdürmesi giderek zorlaşmaktadır (2014: 6). Kamu kurum ve kuruluşlarının vatandaşa sunmuş olduğu hizmetler neticesinde elde

ettiği veriler üzerinde yapılan analizler ve tahminlerle çeşitli avantajlar sağlanmakta bu konuda projeler geliştirilmektedir. Vergi, sosyal güvenlik, sağlık ve emniyet gibi bazı alanlarda büyük miktardaki veri üzerinde yapılan analizler yoluyla hizmet kalitesini iyileştirmek için çalışmalar gerçekleştirilmektedir (Kalkınma Bakanlığı, 2015: 56). Projeler çeşitlendirilerek kamu emniyeti ve ülke güvenliği, tarım arazileri üzerinde sulanabilir arazi kullanımının planlaması, toplam kamu harcamaları ve üretim ilişkisi, sağlık harcamalarına ilişkin denetimlerin daha etkin bir şekilde yapılması, kayıp-kaçakların, usulsüzlüklerin daha hızlı tespiti, gelir ve vergi toplamada etkinliğin artırılmasına yönelik vergi mükelleflerinden denetime tabi tutulacak olanların belirlenmesi, sağlık ve yoksullara yardım hizmetlerini geliştirmeye yönelik salgın hastalık belirtilerinin görüntülenmesi (Kalkınma Bakanlığı, 2010: 2-5) gibi farklı alanlarda makine öğrenmesine ait algoritmalar kullanılması suretiyle veri madenciliği uygulamaları modelleri geliştirilebilir.

### **3.2. Büyük Veri ve Veri Madenciliğinin Kamu Hizmetlerinde Başarılı Uygulanması İçin Gerekli Şartlar**

Büyük hacimli veri sayesinde kurumlar verimlilik, maliyet, hizmet sunumu ve ürün geliştirme gibi önemli avantajlar elde etmektedir. Günümüzde kamu hizmetlerinde yaşanan e-dönüşüm, kurumsal bilginin elektronik ortamdaki yönetilmesi ve kurumlar arası veri paylaşımının artması neticesinde devletin veriyi işleme kapasitesi gelişmiştir. Sayısal ortamdaki veri hacminin önümüzdeki on yıl içinde 44 kat artması beklenmektedir. 2016 yılındaki büyük veri pazarının 53,4 milyar olduğu tahmin edilmektedir (Kalkınma Bakanlığı, 2014: 21).

Geleneksel kamu yönetimi anlayışının yerini YKİ'ye bıraktığı, teknolojinin yardımıyla yönetim bilişim sistemlerinden faydalanmanın zorunlu hale geldiği çağımızda vatandaş odaklı, hesap verilebilir, şeffaf, açık, değişen ihtiyaçlara yanıt verebilecek şekilde dinamik olarak kamu hizmetinde alternatif mekanizmalar geliştirmek isteyen kamu kurumları sahip olduğu veri üzerinde büyük veri ve veri madenciliği uygulamalarını kullanmaktadır.

Kamu kurumları elinde geçmişe yönelik elektronik ortama aktarılmamış, atıl durumda bekletilen, çeşitli amaçlarla talep edilse de kişilerin kullanımına sunulmayan

ve günümüzde farklı kaynaklardan beslenerek büyük kapasitelere ulaşan veri bulunmaktadır. Verinin çeşitli kaynaklardan oluşacak şekilde hızlı biçimde artması kamu kurumlarının da “büyük veri” olarak adlandırılan veriye sahip olması sonucunu doğurmuştur.

Kamunun elinde büyük veri üretimi ve paylaşımını kolaylaştıran temel araçlardan birisi kurumsal internet siteleridir. Kamu kurumlarına ait internet sitelerinin fazla sayıda ve çeşitlilikte olması, sunulan bilgi ve hizmetlerin kapsamı ve içeriğinin farklılık göstermesi kamu internet sitelerinin verimli, etkin ve uyumlu olarak hizmet vermesini zorlaştırmaktadır. Kamu internet siteleri üzerinden sunulan hizmetlerin kalitesinin artırılarak birbirleriyle uyumlu şekilde çalışmalarına yönelik 2005 yılında Kalkınma Bakanlığı tarafından hazırlanan “Birlikte Çalışabilirlik Esasları Rehberi”, 2006 yılında TÜBİTAK’ın yayınladığı “Kamu Kurumları İnternet Kılavuzu” ve 2009 yılında hazırladığı “Kamu Kurumları İnternet Siteleri Standartları ve Önerileri Rehberi” bulunmaktadır. Bürokrasinin azaltılması ve idari basitleştirme çalışmaları kapsamında etkin, verimli, hesap verebilir, vatandaş beyanına güvenen ve şeffaf bir kamu yönetimi oluşturmak; kamu hizmetlerinin hızlı, kaliteli, basitleştirilmiş ve düşük maliyetli bir şekilde yerine getirilmesini sağlamak üzere, idarelerin uyması gereken usul ve esasları düzenlemek amacıyla “Kamu Hizmetlerinin Sunumunda Uyulacak Usul ve Esaslara İlişkin Yönetmelik” 31 Temmuz 2009 Tarihli ve 27305 Sayılı Resmî Gazetede yayınlanmıştır. Rehberlik faaliyetlerinin yanında Türkiye’de kamu internet sitelerinin oluşturulması ve kullanılmasına ilişkin hukuki altyapıyı oluşturan bazı hukuki düzenlemelere Tablo 6’da yer verilmektedir.

**Tablo 6**  
**Kamu İnternet Siteleri Mevzuat Bilgileri**

No	Mevzuat	Kamu İnternet Siteleri İle İlgili Dikkat Edilmesi Gerekenler
1	Kamu Hizmetlerinin Sunumunda Uyulacak Usul Ve Esaslara İlişkin Yönetmelik	Vatandaşın bilgilendirilmesi, Hizmet standartları oluşturma, Özürlülerle ilgili tedbirler
2	Bilgi Edinme Hakkı Kanununun Uygulanmasına İlişkin Esas Ve Usuller Hakkında Yönetmelik	Bilgi verme yükümlülüğü kapsamında alınacak tedbirler,
3	Kamu İdarelerince Hazırlanacak Faaliyet Raporları Hakkında Yönetmelik	Faaliyet raporlarının kamuoyuna açıklanması

4	Birlikte Çalışabilirlik Esasları Rehberi Konulu 2009/4 Sayılı Başbakanlık Genelgesi	Dosya Sunumu ve Değişimi, Kullanılacak Standartlar, İçerik Yönetimi
---	---	---

**Kaynak:** TÜBİTAK, 2018 Kamu İnternet Siteleri Mevzuat Bilgileri kaynağından tablolatırılmıştır (Erişim Tarihi: 28.06.2018).

Yapılan çalışmalar, kamu internet sitelerinde barındırılan verinin anlamlı hale getirilmesi açısından büyük önem taşımaktadır. Büyük veri ve veri madenciliğinin en önemli aşamalarından birisini oluşturan “Verinin Hazırlanması” sürecinin anlamlı veri olmadan gerçekleştirilmesi, “Modelleme” aşamasında makine öğrenmesi algoritmalarının veriye uygulanması neticesinde doğruluğu yüksek sonuçların elde edilmesini zorlaştırarak büyük veri ve veri madenciliği projelerinin başarıya ulaşmasının önünde engel oluşturmaktadır.

Kamu kurumlarında dijitalleşme yönünde atılan adımlar, günümüzde internet siteleri ve sosyal medya gibi araçların kullanımının ötesine geçmiş durumdadır. Bilgi ve iletişim teknolojilerindeki gelişmeler neticesinde nesnelerin interneti, giyilebilir teknoloji, ses ve görüntü tanıma teknolojileri, artırılmış gerçeklik ve yapay zekâ gibi yenilikçi teknolojiler kamu hizmetlerinin yürütülmesinde kullanılmaktadır. Büyük veri teknolojileri ve veri madenciliği uygulamalarının kamu hizmetlerinde uygulanabilmesi, bu teknolojilere kamu kurumlarının ciddi yatırımlar yapması, bu alanda uzun vadeli politika ve stratejiler ile amaçlar geliştirmesine bağlıdır. Bu nedenle, Türkiye’de merkezi idare ve yerel yönetimler düzeyinde büyük veri ve veri madenciliği teknolojilerine ne düzeyde önem verildiğini anlamak üzere kurumsal stratejik planlar üzerinden bir analiz yapılmasına ihtiyaç bulunmaktadır.

### **3.3. Büyük Veri ve Veri Madenciliğine Yönelik Yasal Düzenlemeler**

Türkiye’de kamuda büyük veri ve veri madenciliği uygulamaları açısından bakıldığında, 2000’li yıllara kadar açık kamu verisinin paylaşımı ile ilgili oluşmuş güçlü bir siyasi irade veya bu konuya yönelik kapsamlı yasal düzenlemelerin olmadığı görülmektedir. Ayrıca geleneksel olarak bürokrasi, gizlilik ve kapalılık ilkelerine uygun olarak çalışma eğilimindedir. Ancak son dönemde çıkarılan 4982 sayılı Bilgi Edinme Hakkı Kanunu, Kişisel Verilerin Korunması Kanunu, eylem planları ve strateji belgelerinde şeffaf, açık ve hesap verebilir kamu yönetimi olma adına yapılan düzenlemeler bürokraside hâkim

olan ilke ve değerlerin değişmesini sağlamıştır. Söz konusu kanunlar, idari düzenlemeler ve bürokrasideki davranış değişikliğiyle birlikte kamu verisinin vatandaş, özel sektör ve sivil toplum örgütleriyle paylaşımında büyük artış yaşanmıştır. Kamu kurumları ilgili Bilgi Edinme Hakkı Kanunu, 5018 sayılı Kamu Mali Yönetimi ve Kontrol Kanunu gibi düzenlemelerin etkisiyle kurumsal internet sitesi üzerinden çeşitli bilgi ve belgeleri paylaşımına açmıştır (Yağmurlu, 2011: 8).

Ayrıca kişisel verilerin işlenmesinde başta özel hayatın gizliliği olmak üzere kişilerin temel hak ve özgürlüklerini korumak, kişisel verileri işleyen gerçek ve tüzel kişilerin uyacakları esasları belirlemek üzere 24 Mart 2016 tarihli 6698 sayılı “Kişisel Verilerin Korunması Kanunu” yürürlüğe girmiştir. Kanun gereğince kişisel verilerin tamamen veya kısmen otomatik olmayan yollarla herhangi bir veri kayıt sisteminin bir parçası olarak elde edilmesi, işlenmesi, kaydedilmesi, depolanması ve muhafaza edilmesi kişinin açık rızası olmadan mümkün değildir. Kanun kapsamında açık kamu verisi haline getirilecek verinin kişisel veriden arındırılmış olmasına özen gösterilmesi gerekmektedir.

“Açık Yönetim Ortaklığı”na kurulduğu 2011 yılında katılan Türkiye 2012-2013 yılı için sunduğu “Ulusal Eylem Planında” (Open Government Partnership, 2011) vatandaşla daha fazla bilgi paylaşımı sağlanması, politika yapımı ve karar alma süreçlerine vatandaş ve sivil toplum örgütleri katılımının artırılması ve taraflar arasında diyalogun artırılması şeklinde hedeflere yer verilmiştir. Bu kapsamda bütünlük, şeffaflık, hesap verebilirlik ve bürokrasi ve yolsuzlukla mücadele doğrultusunda “www.transparency.gov.tr” saydamlık/şeffaflık portalının hizmete sunulacağı, kamu hizmetlerinin kalitesini iyileştirmek amacıyla kamu harcamalarında şeffaflığı sağlamak üzere “www.spending.gov.tr” harcama portalının hizmete sunulacağı, politika oluşturma sürecinde vatandaş ve diğer ilgili tarafların katılımını sağlamak üzere “www.regulation.gov.tr” düzenleme portalının hizmete sunulacağı, ihale ve değerlendirmelerin elektronik ortamda yapılmasını sağlayacak “Elektronik Kamu İhale Platformu”nun hizmete alınacağı taahhüt edilmiştir. 2012-2013 yılında sunulan ilk eylem planından 2014 yılına kadar bahse konu taahhütlerle ilgili somut adımların atılmadığı (Open Government Partnership, 2014) “Açık Yönetim Ortaklığı” tarafından bildirilmiştir. 2016 yılında yapılan gözden geçirme (Open Government Partnership,

2016a) neticesinde 1 Eylül 2016'ya kadar taahhütleri yerine getirmesinden sorumlu bakanlığın belirlenmesi ve eylem planı sunulması kararına istinaden 2016-2019 yıllarını kapsayan "Saydamlığın Artırılması ve Yolsuzlukla Mücadelenin Güçlendirilmesi" Eylem Planı (Başbakanlık, 2016) Nisan 2016'da Resmî Gazete'de yayımlanarak yürürlüğe girmiştir.

Bahse konu yapılan karşılıklı çalışmalara rağmen 2011 yılından bu yana ilk eylem planı doğrultusunda bir ilerleme kaydedilemediği için Türkiye'nin üyeliği 21 Eylül 2016 yılında askıya (Open Government Partnership, 2016b) alınmıştır. Bununla birlikte 2016-2019 Ulusal e-Devlet Stratejisi ve Eylem Planında "Kullanım, Katılım ve Şeffaflığın Artırılması"na (Stratejik Amaç 4) yönelik belirlenen stratejik amaç doğrultusunda açık verinin kullanım alanının yaygınlaştırılması (Hedef 4.2) ve e-Katılım mekanizmalarının güçlendirilmesi (Hedef 4.3) hedefleri bulunmaktadır. Geline son noktada Türkiye 2016 yılında üyeliğinin askıya alınmasının ardından taahhüt ettiği standart ve kriterlere uygun faaliyetleri gerçekleştirmediği, bilgilendirme ve sonuç alınmadığı için 21 Eylül 2017'de üyelikten çıkarılmıştır (Open Government Partnership, 2017). Buna rağmen, açık yönetim ve açık veriye ilişkin politika ve stratejiler, sistematik ve bütüncül olmamakla birlikte ilgili kurumlar tarafından uygulanmaya devam etmektedir. Türkiye'nin açık veri yönetimine ilişkin kapsamlı politikaları hayata geçirmesi durumunda, kamu verisinin paylaşımının hızlanacağı, farklı formatlarda hazırlanan verilerden ihtiyaç odaklı veri madenciliği çalışmalarının yapılabileceği söylenebilir.

### **3.3. Merkezi İdare Kuruluşlarında Büyük Veri ve Veri Madenciliğine İlişkin Politika ve Uygulamalar**

Veri madenciliği uygulamaları özel sektöre rekabet avantajı sağlarken, sağlık, eğitim, güvenlik, savunma, enerji ve ulaşım gibi kamusal hizmetlerin sunumu noktasında etkinliği ve verimliliği artırmak, daha kaliteli kamu hizmetini vatandaşa sunmak ve kamu yöneticilerinin daha etkili karar almalarını sağlamak gibi amaçları gerçekleştirmektedir.

Kamu verisinin kullanımına ilişkin politika ve hedeflere bakıldığında 10. Kalkınma Planında (2014-2018) "*katma değerli yeni hizmetlerin üretimi amacıyla ticari değeri olan veriler başta olmak üzere kamu sektörü bilgisinin paylaşımı ve yeniden kullanımının sağlanacağı*" (md. 410) ifade edilirken Bilgi Toplumu Stratejisi ve Eylem



getirmektedir. Ofisler, verilen görevleri yerine getirmek üzere Cumhurbaşkanlığına bağlı, özel bütçeli, kamu tüzel kişiliğini haiz, idari ve mali özerkliğe sahip yapılardır.

Cumhurbaşkanlığı Teşkilatı Hakkında Resmî Gazete’de yayımlanan 10 Temmuz tarihli 1 No’lu Cumhurbaşkanlığı Kararnamesi’ne göre, Bilim, Teknoloji ve Yenilik Politikaları Kurulu görev ve yetkileri şu şekildedir;

- Bölgesel ve küresel rekabette mukayeseli üstünlük kazandıracak alanları tespit edip bu alanlarda tavsiyelerde bulunmak,
- Bilgi toplumuna ilişkin hedef ve strateji tavsiyelerinde bulunmak,
- Ar-Ge çalışmalarının özendirilmesi, teşvik edilmesi ve sonuçlarının ürünleştirilmesi için politika önerilerinde bulunmak,
- Bilim ve teknoloji alanındaki araştırma ve geliştirmek politikalarının ekonomik kalkınma, sosyal gelişme ve millî güvenlik hedefleri doğrultusunda tespit edilmesi, yönlendirilmesi ve koordinasyonunun sağlanması amacıyla altyapı, insan kaynağı ve diğer tüm kaynakların geliştirilmesine yönelik araştırmalar yaparak öneriler oluşturmak,
- Stratejik önemi haiz teknoloji alanlarını belirlemek için araştırmalar yapmak ve bu alanlara ilişkin öneriler oluşturmak,
- Teknolojik dönüşüm alanlarını tespit ederek orta ve uzun vadeli millî teknoloji politika önerileri oluşturmak,
- Girişimcilik kültürünün yaygınlaştırılması için araştırmalar yapmak ve öneriler oluşturmak,
- Sanayimizin ithal bağımlılığını azaltacak, rekabet gücüne artıracak programların hayata geçirilmesi ile sanayi ürünleri güvenliği ve denetimi politikasının ve endüstri bölgesi kurulacak alanların belirlenmesi amacıyla çalışmalar yürütmek,
- Serbest bölgelerde yapılan faaliyetleri değerlendirmek, bu bölgelerin geliştirilmesine ve sorunların çözümüne ilişkin stratejileri belirlemek ve önerilerde bulunmak,

- Araştırma altyapılarıyla ilgili genel strateji çalışmaları yapmak ve bu konuda öncelikleri değerlendirmek,
- İnternet ortamının ekonomik, ticarî ve sosyal hayat ile bilim, eğitim ve kültür alanında etkin, yaygın, kolay erişilebilir olarak kullanımı teşvik edecek politika ve strateji önerileri hazırlamak.

10 Temmuz 2018 tarihli 1 No'lu Cumhurbaşkanlığı Kararnamesine göre, Dijital Dönüşüm Ofisi'nin görevleri ise şu şekildedir;

- Kamunun dijital dönüşümünü (e-devlet dönüşümü) koordine etmek,
- Millî teknolojinin geliştirilmesi ve bu kapsamda farkındalık oluşturmak amacıyla gerekli olan projeler geliştirmek,
- Büyük veri analizi yapmak, öncelikli proje alanlarında yapay zekâ uygulamalarına öncülük etmek,
- Siber güvenlik ve bilgi güvenliğini artıracak projeler geliştirmek.

Üniter ve merkezîyetçi bir devlet olan Türkiye'de, kamu politikaları da büyük ölçüde merkezi düzeyde belirlenmektedir. Bu bağlamda, büyük veri ve veri madenciliğine ilişkin hedef ve politikalar ulusal düzeyde hükümet programları, kalkınma planları gibi politika belgeleri ile kurumsal stratejik planlar, eylem planları ve performans programları gibi araçlarla oluşturulmaktadır. Bakanlıklar ve diğer merkezi idare kuruluşları son yıllarda içinde bulunduğu projelere, hizmete sunduğu uygulamalara ve belirledikleri hedeflere ulaşmak üzere büyük veri ve veri madenciliği uygulamalarından faydalanmaktadır.

### **3.3.1. Büyük Veri ve Veri Madenciliğine Yönelik Ulusal Politika ve Programlar**

Yeni iletişim teknolojilerine yönelik ulusal politika ve planlar incelendiğinde özellikle 1980'li yıllardan itibaren e-Avrupa projesine bağlı olarak enformasyon toplumu kazanımlarının elde edilmesi amaçlanmıştır (Kaymas, 2007: 117-118). Büyük veri ve veri madenciliği uygulamalarının yakın zamanda teknolojik açıdan olgunlaşmasına dayanak oluşturan geçmişteki üst politika belgelerine bakıldığında 1999 yılında yayınlanan Ulusal Enformasyon Altyapısı Ana Planı (TUENA), e-Türkiye Girişimi

Eylem Planı (Başbakanlık, 2002) ve e-Dönüşüm Türkiye Projesi Kısa Dönem Eylem Planı (2003-2004) gibi politika belgeleri dikkati çekmektedir.

TUENA raporu (Ulaştırma Denizcilik ve Haberleşme Bakanlığı, 1999); kamu hizmetlerinin elektronik ortamlarda yürütülmesi hususunda çalışmaların bulunduğu ilk resmi rapor olması açısından önemlidir (İnce, 2001: 74). 1999 yılında enformasyon teknik altyapısıyla ilgili saptamalar ve ilerye dönük 10 yıllık perspektif doğrultusunda AB içindeki 6. Çerçeve Programında belirlenmiş hedeflerle büyük oranda örtüşmüş bir plandır (Arkun, 2003: 185). TUENA projesinin temel vizyonları sürdürülebilir gelişme için genel sosyo-ekonomik yararın en çoğa çıkarılması, enformatik sektörünün donanım, iletişim hizmetleri, yazılım ve içerik sanayi alt dallarında ulusal katkı payını yukarı doğru sıçratması, küresel enformatik pazarından pay alabilmek için bölgesinde öncülük yapması ve bu noktaları gerçekleştirecek politikalar ve yapılanmaların üretilmesi olarak ortaya konmuştur.

TUENA raporu kaynaklarından derlenen MERNİS Projesi, MEB-SİS (İLSİS, OKULSİS, PERSİS Projeleri), Coğrafi Bilgi Sistemleri Projesi, VEDOP Projesi (Vergi Daireleri Otomasyonu), TAKBİS Projesi (Tapu Kadastro Bilişim Sistemi), Otomatik Geçiş Sistemi Projesi (OGS), POL-NET 2000 gibi projeler (Canpolat, 2001: 76-78) göz önünde bulundurulduğunda günümüze kadar geçen zaman dilimi içerisinde projelerin farklı zamanlarda hayata geçtiği, kurumlar ve vatandaşlar tarafından aktif olarak kullanıldığı görülmektedir. Bilgi toplumuna dönüşüm için toplumun tüm kesimlerine eşit ölçüde ve eşit erişim gücüyle katılımı öngörülmektedir. Türkiye’de Türk Telekom’un özelleştiği 2005 yılından itibaren enformasyon ve iletişim teknolojileri alanında kayda değer yeni yatırımların gerçekleştirilmediği görülmekle birlikte (Kaymas: 2012: 28) 3. Nesil (3G) hizmeti 30 Temmuz 2009 tarihinde, 4.Nesil (4G) iletişim altyapısı 01 Ocak 2016 tarihinde hizmete girmiştir.

TUENA raporunun tamamı uygulamaya konmamış olmasına rağmen bilgi toplumuna geçiş çalışmaları devam etmiş, e-devlet ve e-vatandaş kavramlarının ortaya çıkmasıyla geleneksel devlet yapısında elle yapılan işlemler yönetim bilişim sistemleri uygulamalarıyla kayıt altına alınmaya başlanmıştır. Geçiş işlemlerine yönelik çıkarılan politika ve kanunların AB mevzuatına uygun şekilde olması amacıyla eAvrupa+ Eylem

Planına uygun şekilde e-Türkiye Girişimi Eylem Planı ve e-Dönüşüm Türkiye Projesi başlatılmıştır.

Türkiye'nin bilgi toplumuna dönüşümündeki bakanlık düzeyinde ilk stratejik planı 2006-2010 yılında Kalkınma Bakanlığı tarafından hazırlanmıştır. "Bilgi Toplumu Stratejisi ve Eylem Planında" sosyal dönüşüm, bilgi ve iletişim teknolojilerinin iş dünyasına nüfuzu, küresel rekabet, AR-GE ve yenilikçiliğin geliştirilmesi, yaygın ve ucuz altyapı hizmetleri, vatandaş odaklı hizmet dönüşüm ve kamu yönetiminde modernizasyonu ele alınmıştır (Kalkınma Bakanlığı, 2014: 10-11). 2006-2010 yıllarını kapsayan eylem planına yönelik 2013 yılında yayınlanan nihai raporda %64 oranında başarıya ulaşılarak vatandaşa yönelik hizmetlerin dönüşümü ve altyapı hizmetlerinin iyileştirilmesinde %65-70 oranlarının yakalandığı belirtilmektedir (Kalkınma Bakanlığı, 2014: 12).

Dijitalleşme ve kamu kurumlarının bilgi ve iletişim teknolojilerinden yararlanmalarına yönelik yürütülen strateji ve eylem planları çerçevesinde Avrupa Komisyonu tarafından sunulan 2012 yılı e-Devlet Kıyaslama Raporuna göre elektronik ortamda hizmet alternatiflerinin sunulması açısından Türkiye, AB ortalamasının üzerinde olmasına rağmen vatandaşlara gerekli ve yeterli bilgilendirme yapılmaması, internet ağının yaygınlaşmakta olması, kullanım oranlarının çeşitli sebeplerle düşük olması ve hizmetlere ilişkin her halükarda bir bürokratik işlemin gerekliliği hizmetlerin vatandaşlar tarafından kullanımının düşük olmasına neden olmuştur (Dijital Türkiye Platformu, 2014: 45-46).

İkinci Bilgi Toplumu Stratejisi ve Eylem Planı (2015-2018) planının temel hedefi, büyüme ve istihdam alanlarında iyileştirmeler yapılması olarak belirlenmiştir. Türkiye'nin geldiği nokta, karşılaştığı sorunlar, ihtiyaç duyulan konular, 10. Kalkınma Planı ve "Avrupa İçin Sayısal Gündem" gibi plan ve programların nitelikleri, 2015-2018 Eylem Planı'nın hazırlanmasında etkili olmuştur (Kalkınma Bakanlığı, 2014: 14).

2015-2018 Bilgi Toplumu Stratejisi ve Eylem Planı kapsamında 2016 yılında e-devlet ekosisteminin etkinliğinin ve sürdürülebilirliğinin sağlanması, altyapı ve idari hizmetlere yönelik ortak sistemlerin hayata geçirilmesi, kamu hizmetlerinde e-dönüşümün sağlanması, kullanım, katılım ve şeffaflığın artırılması olmak üzere dört

adet stratejik hedefi içerecek şekilde “2016-2019 Ulusal e-Devlet Stratejisi ve Eylem Planı” hazırlanmıştır (Ulaştırma, Denizcilik ve Haberleşme Bakanlığı, 2016).

Cari eylem planı olarak halen yürürlükte olan Ulusal e-Devlet Stratejisi ve Eylem Planı (2016-2019) etkin (entegre-teknolojik-katılımcı-inovatif-nitelikli) e-devlet ile toplumun yaşam kalitesinin artırılmasını vizyon olarak belirlemiştir. E-devlet hizmet sunumunda etkinlik, e-devlet hizmet kullanımı, e-devlet hizmet memnuniyeti, e-devletten sağlanan kamu yararı ve Türkiye'nin uluslararası e-devlet endekslerindeki (Birleşmiş Milletler, 2017) konumunun yükseltilmesi eylem planının temel hedefleridir (Ulaştırma Denizcilik ve Haberleşme Bakanlığı, 2016: 7).

Kalkınma ve strateji planlarının yanında son yıllarda hükümet programlarında büyük veri ve veri madenciliğine odaklanan politika ve amaçlara yer verilmektedir. 65. Hükümet Programı'nda (Başbakanlık, 2016) Kamu Bulut Projesi ve Kamu Entegre Veri Merkezi Projesi, Dijital Türkiye Projesi ve Entegre Kent Yönetimi Bilgi Sistemi Projesi dikkati çekmektedir. Ayrıca kamu kurumlarının topladıkları ve ürettikleri verilerin, bilgi güvenliği tedbirlerini almak suretiyle paylaşımının sağlanması ve kamuda büyük veri uygulamalarının hayata geçirilmesi hususu hükümet programında yer almıştır. 3 Kasım 2002 yılından günümüze kadar tüm hükümet programları incelendiğinde “büyük verinin” ilk kez 64. Hükümet Programında yer aldığı görülmektedir. Yönetişim başlığı altında yöneten ve yönetilenlerin birlikte tanımlanması ve birlikte çözümler araması gayretleri kapsamında kamuda büyük veri uygulamalarının hayata geçirilmesi hedeflenmiştir (64. Hükümet Programı, 2015). 65. Hükümetin bu kapsamdaki hedefleri devam ettirdiği söylenebilir.

Kalkınma planları incelendiğinde ise büyük veri ve veri madenciliğine yönelik hedeflere ilk olarak 2014-2018 yıllarını kapsayan 10. Kalkınma Planında (Kalkınma Bakanlığı, 2013) yer verildiği anlaşılmaktadır. Önceki kalkınma planlarının içeriğinde bilişim teknolojilerine yönelik ağırlıklı olarak e-devlet kapsamında politika ve hedefler belirlenmiştir. Katma değerli hizmet üretimi için kamu sektörü bilgisinin paylaşımı ve yeniden kullanımının sağlanması; açık kaynak kodlu yazılımlar, büyük veri, bulut bilişim, yeşil bilişim, mobil platform, nesnelerin interneti gibi ürün, hizmet ve yönelimlerin değerlendirilerek kamu için uygun olabilecek çözümlerin hayata geçirilmesi (s.55); bulut bilişim hizmetlerinin gelişebilmesi ve yaygınlaşması için

gerekli yasal ve idari düzenlemelerin yapılması ve akıllı uygulamaların sağlık, ulaştırma, bina, enerji ile afet ve su yönetimi gibi alanlar başta olmak üzere kullanımı (s.97) bunlar arasında öne çıkmaktadır (Kalkınma Bakanlığı, 2013).

2016-2019 Ulusal e-Devlet Stratejisi ve Eylem Planı e-Devlet politikasının şekillendirilmesi, politikalara uygun bütüncül ve sürdürülebilir stratejik bakış açısının benimsenmesi amacıyla hazırlanmıştır. Ulusal E-Devlet Stratejisi ve Eylem Planı (2016-2019), doğrudan büyük veri ve veri madenciliğine ilişkin politika ve hedefler içermesi, 65. Hükümet Programı ve 10. Kalkınma Planına uygun olarak hazırlanmış olması bakımından önemli bir politika dokümanıdır. Ulusal E-Devlet Stratejisi; e-devlet ekosisteminin etkinliği ve sürdürülebilirliğinin sağlanması, altyapı ve idari hizmetlere yönelik ortak sistemlerin hayata geçirilmesi, kamu hizmetlerinde e-dönüşümün sağlanması ile kullanım, katılım ve şeffaflığın artırılması şeklinde dört stratejik amaç üzerine inşa edilmiştir. Belirlenen stratejik amaçlar büyük veri ve veri madenciliği ile yakından ilişkili olup, Kamu Entegre Veri Merkezlerinin kurulması, kamu bulut bilişim altyapısının oluşturulması, akıllı kentler programı geliştirilmesi, kent yönetimi bilgi sistemi geliştirilmesi, kamu verisinin paylaşılması, açık veri paylaşım portalının oluşturulması, kamu verilerinin açık veriye dönüştürülmesi ve paylaşılması, kamuda büyük veri ve nesnelerin interneti politikalarının geliştirilmesi şeklindeki eylem ve uygulamalar bu kapsamda değerlendirilebilir. Politika ve stratejilerin geliştirilerek projeler üretilmesi, ihtiyaç duyulan konularda mevzuat çalışmalarının yapılarak uygulanması, kamu çalışanlarının eğitimi ve farkındalık artırıcı çalışmaların gerçekleştirilmesine yönelik yurtiçi ve yurtdışı örneklerinin incelenerek hangi hizmetlerin büyük veri ve nesnelerin interneti kullanılarak sağlanabileceğinin tespit edilmesi öngörülmektedir (Ulaştırma, Denizcilik ve Haberleşme Bakanlığı, 2016).

Belirtilmesi gerekir ki, önemli politika dokümanlarından biri olan “Bilgi Toplumu Stratejisi ve Eylem Planında” (2015-2018) büyük verinin ekonomik değere dönüşmesini sağlamak amacıyla sağlık, sosyal güvenlik, vergi, güvenlik ve istatistik gibi alanlar başta olmak üzere kamuda büyük veri uygulamaları geliştirmek, kamuda büyük veri pilot uygulaması gerçekleştirmek şeklinde eylem ve politikalar belirlenmiştir (Kalkınma Bakanlığı, 2015).

Büyük veri teknolojileri ve veri madenciliği Türkiye’de 24 Haziran 2018 tarihinde gerçekleştirilen seçimlerde de siyasi partilerin seçim beyannamelerinde yer bulmuştur. Bu kapsamda milletvekili seçimlerinden birinci parti olarak çıkan Adalet ve Kalkınma Partisi (AK Parti) seçim beyannamesinde büyük verinin oluşturulup işlenerek ticarileştirileceği, yazılım ve donanım endüstrisinin geliştirilerek stratejik sektörler ve yenilikçi üretim kabiliyetinin güçlendirileceği, dijitalleşmenin ürünü olan nesnelerin interneti, sensör teknolojileri, akıllı otomasyon teknolojileri, robotik, veri analitiği ve yapay zekâ, artırılmış gerçeklik, bulut teknolojisi ve eklemeli üretim vb. yenilikçi teknolojilerle toplumsal hayatın her aşamasında eğitimden sağlığa, adaletten, kamu yönetimine üretimden ticarete kadar her alanda dijital dönüşümün gerçekleştirileceği belirtilmektedir (AK Parti, 2018: 195, 199-200). Ayrıca dijital dönüşümün hızlandırılması amacıyla özel kesim, sivil toplum temsilcileri, kamu ve üniversite işbirliğinin geliştirileceği, yapa zekâ, robot teknolojileri, bulut bilişim ve büyük veri, nesnelerin interneti gibi yenilikçi teknolojilere öncelik verileceği, bilgilerin elektronik ortama aktarılarak veri altyapılarının kurulmasının destekleneceği, dijital dönüşümü gerçekleştirecek ve kullanacak insan kaynağına yatırım yapılacağı, kamu kurumlarındaki verinin daha güvenli bir ortamda saklanması ve mükerrer veri merkezi yatırımlarının engellenmesi amacıyla, stratejik öneme sahip “Ulusal Kamu Entegre Veri Merkezi Projesi”nin hayata geçirileceği (AKP, 2018: 241, 201-203) ifade edilmektedir.

### **3.3.2. Bakanlıkların Büyük Veri ve Veri Madenciliği Politikaları**

Stratejik planlar, üst politika belgelerinde geliştirilen strateji, amaç ve politikaların somut hedeflere, faaliyet ve projelere dönüştürüldüğü, gerçekleşme süreçlerinin takip edildiği ve ulaşılan başarının değerlendirildiği belgelerdir. Bakanlıkların güncel stratejik planlarının, geliştirdikleri büyük veri ve veri madenciliği politikaları ve projelerini anlamak açısından önemli belgelerdir. Cumhurbaşkanlığı hükümet sistemi öncesi yapıda yer alan bakanlıkların stratejik planları incelenmiş olup 11 bakanlığın stratejik planında konuyla ilgili hedefler bulunduğu tespit edilmiştir. Yeni hükümet sistemi ile uyumlu bakanlıkların stratejik planları hazırlık aşamasında olduğundan çalışma kapsamı dışında tutulmuştur. Daha sonra söz konusu bakanlıkların stratejik planları büyük veri, açık veri, veri ambarı, Endüstri 4.0 ve veri madenciliği gibi anahtar kelimeler üzerinden metin içerik taraması yapılmış, aynı zamanda anahtar kelime taramasından

ulaşılamayacak örtük hedef ve projelerin tespiti için planlar tek tek okunmuş ve değerlendirilmiştir. İçerik analizi sonucunda incelenen planlarda, büyük veri ve veri madenciliği ile ilişkisi bulunan strateji ve faaliyetler ile bunların ilişkili olduğu amaç ve hedefler Tablo 7’de gösterilmektedir.

**Tablo 7**  
**Bakanlıkların Büyük Veri ve Veri Madenciliğine İlişkin Amaç ve Hedefleri**

<b>Bakanlık Stratejik Planı</b>	<b>Stratejik Amaç</b>	<b>Stratejik Hedef</b>	<b>Hedefe Yönelik Strateji ve Faaliyetler</b>
<b>Bilim, Sanayi ve Teknoloji Bakanlığı (2013-2017)</b>	“Ekonominin bütün kesimlerinde verimlilik ve temiz üretim/eko-verimlilik ile ilgili gösterge ve istatistikler oluşturmak, ölçümler yapmak”	“Politika belirleme ve araştırma çalışmaları için temiz üretim/eko-verimlilikle ilgili makro düzeyde gösterge ve istatistikler oluşturulacak”	“Temiz üretim/eko-verimlilik ile ilgili istatistiklerin hesaplanmasına yönelik veri altyapısı çalışmalarının yapılması ve güncellenmesi”
<b>Ekonomi Bakanlığı (2013-2017)</b>	“İkili, bölgesel ve çok taraflı ticaret ve yatırım ilişkilerini ülkemizin çıkarları doğrultusunda yürütmek”	“İkili, bölgesel ve çok taraflı stratejilerimizin oluşturulmasında nicel ve nitel analiz yöntemlerinin etkin kullanımı sağlanacaktır.”	“Ekonomi ve dış ticaret konusunda modelleme çalışmaları ... ekonomi politika uygulamaları ile ticaret politikası araçlarının etkinliğinin geliştirilmesi için nicel ve nitel analiz yöntemleri ile etki analizi mekanizmaları tesis edilecektir.”
<b>Enerji ve Tabii Kaynaklar Bakanlığı (2015-2019)</b>	“Enerji dışı hammadde tedarik güvenliği”	“Ülkemiz maden kaynaklarının aranmasının uluslararası standartlarda yapılması ve raporlandırılmasına ilişkin altyapı çalışmaları tamamlanacaktır.”	“Maden kaynaklarıyla ilgili] Kapsayıcı bir veri envanterinin diğer paydaşların katılımı ile oluşturulması, bu kapsamda tüm jeolojik bulguların yer aldığı bir veri bankasının oluşturulması amacıyla yasal ve

			fiziki altyapının geliştirilmesi (fiziki ve elektronik veri tabanı şeklinde)”
<b>Gıda, Tarım ve Hayvancılık Bakanlığı (2013-2017)</b>	“Üretimden tüketime kadar, uluslararası standartlara uygun gıda güvenilirliğini sağlamak.”	“Güvenilir gıda konusunda toplumda duyarlılığı artırmak”	“Alo Gıda Hattına gelen başvurular doğrultusunda elde edilen bilgiler veri havuzunda toplanarak risk haritası çıkarılmakta ve hangi ürünlerin hangi bölgelerde riskli olduğu saptanarak denetim ve izleme programlarının daha etkin olarak yapılması sağlanmaktadır.”
<b>Sağlık Bakanlığı (2013-2017)</b>	“Birey ve topluma erişilebilir, uygun, etkili ve etkin sağlık hizmetleri sunmak”	“Sağlık hizmet sunumunun izlenmesi, değerlendirilmesi ve kanıta dayalı karar almak için sağlık bilgi sistemlerini geliştirmek”	“Sağlık hizmetlerini planlamak amacıyla oluşturulan “Karar Destek Sistemi” için veri ambarları kurmak ve veri madenciliği uygulamalarını iyileştirmek”
<b>Ulaştırma Denizcilik ve Haberleşme Bakanlığı (2017-2021)</b>	“Kurumsal kapasiteyi geliştirmek”	“Kurumsal ihtiyaçları etkin, verimli ve güvenli bir şekilde karşılamak amacıyla bilişim alt yapısını güçlendirmek”	“Büyük verinin toplanması, organize edilmesi ve analizi ile birlikte, stratejik kararlarının alınabilmesi, ihlal olayları olmadan önce tahmin edilebilmesi amacıyla, yapısal verilerin yanı sıra yapısal olmayan verilerin de kurumların karar destek sürecinde kullanılması”

<b>İçişleri Bakanlığı (2015-2019)</b>	“Hizmet sunumunda bilgi teknolojilerini etkin, verimli ve güvenli olarak kullanmak”	“Nüfus ve vatandaşlık hizmetlerinin sunumunda bilgi teknolojilerinin kullanımını yaygınlaştırmak”	“Açık kaynak kodlu yazılımlar, büyük veri, bulut bilişim, yeşil bilişim, mobil platform, nesnelerin interneti gibi ürün, hizmet ve yönelimler değerlendirilerek kamu için uygun olabilecek çözümler hayata geçirilecektir.”
<b>Çevre ve Şehircilik Bakanlığı (2015-2017)</b>	“Sağlıklı, güvenli ve mekân kalitesi yüksek yerleşmeleri oluşturmak için; mekânsal planlama, kentsel dönüşüm proje ve uygulamaları ile altyapı hizmetlerini yürütmek, ulusal coğrafi bilgi sistemini kurmak “	“Yatırım ortamını iyileştirmek, kamu hizmetlerini etkinleştirmek ve verimliliği arttırmak için Coğrafi Bilgi Sistemleri çalışmaları yürütülecektir.”	“Kent Bilgi Sisteminin yaygın kullanımını sağlamak ve coğrafi bilginin üretimi ve paylaşımında çağdaş bilgi teknolojilerini kullanmak amacıyla yerel yönetimlere hizmet veren bulut bilişim altyapısı kurulması ve idame ettirilmesi hedeflenmektedir.”
<b>Gümrük ve Ticaret Bakanlığı (2013-2017)</b>	“Gümrük ve ticarete ilişkin yönlendirici politikaları ilgili kurum ve kuruluşlarla işbirliği ve eşgüdüm halinde belirlemek”	“Politika geliştirme ile hızlı ve nitelikli karar almaya yardımcı olacak veri, enformasyon ve bilgi yönetimini geliştirmek ve kurumsallaştırmak”	“Veri, enformasyon ve bilgi ile desteklenen politikalar geliştirmek”
<b>Kalkınma Bakanlığı (2014-2018)</b>	“Küresel ve ulusal gelişmeler ile temel eğilimleri yakından izleyerek, ... kısa, orta ve uzun vadeli tahmin, analiz ve araştırmalar yapmak.”	“Mikro veri temelli politika ve etki analizi çalışmaları yapılacaktır”	“Kamu politikalarının etkilerini ölçmek amaçlı geliştirilmiş ve uluslararası bilimsel standartlara uygun niceliksel metodolojilerle mikro veriye dayalı analizler yapılacaktır”.

<b>Millî Eğitim Bakanlığı (2015-2019)</b>	“Bütün bireylere çağın gerektirdiği bilgi, beceri ve davranışın kazandırılması ile girişimci, yenilikçi bireylerin yetişmesine imkân sağlamak”	“Bütün bireylerin bedensel ve zihinsel gelişimlerine yönelik faaliyetlere katılım oranını ve öğrencilerin akademik başarı düzeylerini artırmak”	“Öğrencilerin açık öğretimde ortalama kayıtlı kalma süresi gibi önemli verilerin toplanmasına yönelik izleme sistemiyle veriye dayalı açık öğretim sisteminin niteliği geliştirilecek”
---	--	---	--

**Kaynak:** Köseoğlu ve Demirci, 2017: 2234-2236.

Bakanlıkların büyük veri ve veri madenciliğine ilişkin amaç ve hedefleri incelendiğinde, Çalışma ve Sosyal Güvenlik Bakanlığı, Gençlik ve Spor Bakanlığı, Kültür ve Turizm Bakanlığı, Maliye Bakanlığı, Orman ve Su İşleri Bakanlığı dışında kalan tüm bakanlıklar doğrudan ya da dolaylı strateji ve/veya faaliyet belirlemiştir. Bakanlıkların, büyük veri ve veri madenciliğine ilişkin projeleri olmakla birlikte, büyük veri ve veri madenciliğini stratejik amaç ve hedeften ziyade, projelerinde faaliyet olarak belirtmişlerdir. Sağlık Bakanlığı, Ulaştırma, Denizcilik ve Haberleşme Bakanlığı ile İçişleri Bakanlığı ise doğrudan büyük veri ve veri madenciliğine atıf yapmaktadır (Köseoğlu ve Demirci, 2017: 2236). Stratejik planlarında büyük veri ve veri madenciliğine doğrudan yer veren bakanlıklar, stratejik hedeflerine ulaşmak ve karar vericilerine destek sağlamak amacıyla geleceğe dönük tahminleri elde etmek üzere büyük veriden yararlanmak istemektedir. Tüm bakanlıkların sayısı ve hizmet alanlarının çeşitliliği düşünüldüğünde, bakanlıkların stratejik planlarında büyük veri teknolojileri ve veri madenciliğine daha çok örtük olarak değindikleri, doğrudan hedef ve projelerin ise sayısal olarak az kaldığı söylenebilir. Ancak Cumhurbaşkanlığı sisteminde hem Cumhurbaşkanlığı ofis ve kurullarında hem de bakanlıkların hazırlayacağı yeni stratejik planlarda bu alanda somut hedef ve politikalar üretmeleri beklenmektedir.

### **3.3.3. Türkiye’de Yürütülen Başlıca Büyük Veri ve Veri Madenciliği Projeleri**

Türkiye’de veri madenciliğine yönelik yürütülen projeler son yıllarda büyük verinin gündeme gelmesinin ardından artış göstermekle birlikte, kamu kurumlarının sahip olduğu veriden faydalanmasını sağlayacak büyük veri ve veri madenciliği projelerinde istenen ivme yakalanamamıştır. Teknolojik yenilikleri örgütün ihtiyaçları doğrultusunda

kullanmakta daha esnek ve başarılı olan özel sektör kuruluşları büyük veri teknolojileri ve veri madenciliğine dönük çalışmaları daha yoğun biçimde yürütmektedir.

Türkiye’de telekomünikasyon, bankacılık ve pazarlama sektörleri başta olmak üzere müşteri davranışlarının önemli olduğu alanlarda hizmet veren şirketler müşterileriyle ilgili topladıkları büyük hacimli veriyi kullanarak müşteri davranışlarını tahmin etme, hizmet kalitesini artırma, müşteri memnuniyetini ve bağlılığını artırma, karar alma mekanizmalarına yardımcı olacak öngörüler üretme, detaylı müşteri segmentasyonu yapma, yeni sunulacak ürünleri müşteri beklentilerine göre belirleme ve ürün fiyatlandırmasını optimize etme gibi alanlarda “büyük veri” ve “veri madenciliği” çalışmaları yapmaktadır (Kalkınma Bakanlığı, 2015: 132).

Merkezi idare özelinde kamu kurumlarına bakıldığında ise vergi, sosyal güvenlik, sağlık ve emniyet gibi alanlarda ilgili bakanlıklar ve kamu kuruluşları tarafından yürütülen “büyük veri” ve veri madenciliği uygulamaları ile veri analiz edilerek verimliliği artırmak, kayıp kaçak oranlarını düşürmek, hizmet kalitesini artırmak ve karar vermede üst yönetime yardımcı olmak üzere projeler gerçekleştirilmektedir (Kalkınma Bakanlığı, 2015: 56). Bununla birlikte, Türkiye’de kamu kurumları tarafından yürütülen büyük veri ve veri madenciliği projelerinin bir envanteri bulunmamaktadır. Türkiye Bilişim Derneği (TBD) Tablo 8’de görüleceği üzere yakın dönemde Milli Eğitim ve Sağlık Bakanlığı ile Sosyal Güvenlik Kurumu tarafından yürütülen büyük veri projelerini içeren bir rapor yayımlamıştır.

**Tablo 8**  
**Türkiye’de Bazı Kamu Kurumları Tarafından Yürütülen Büyük Veri Projeleri**

<b>Kurum</b>	<b>Proje ve Faaliyet Adı</b>	<b>Proje ve Faaliyet Açıklaması</b>
<b>Sosyal Güvenlik Kurumu</b>	e-Bildirge Sistemi	Tüm sigorta prim tahsillerinin ve işyeri tescil kayıtlarının takibinin gerçekleştirildiği sosyal güvenlik projesi
	MEDULA	Vatandaşlara ait tüm sağlık ödemelerinin belirlenen kurallar çerçevesinde yürütülmesine yönelik bir proje
	Aylık Tahsis ve Diğer Uygulamalar	Emeklilere yönelik Aylık Tahsis Uygulaması başta olmak üzere 1500 adet olgunlaşmış eski (legacy) uygulamadan oluşan bir uygulama grubu

<b>Kurum</b>	<b>Proje ve Faaliyet Adı</b>	<b>Proje ve Faaliyet Açıklaması</b>
	ALO 170 Projesi	Çalışma ve Sosyal Güvenlik Bakanlığı, Sosyal Güvenlik Kurumu ve Türkiye İş Kurumu Genel Müdürlüğü tarafından sunulan tüm hizmetlerle ilgili olarak bilgilendirme yapan ve çözüm üreten bir iletişim merkezi
	Veri Ambarı Projesi	Sağlık, Sigortalama, Tahsis (Emeklilik) rapor grupları ile ortalama 500 standart rapordan on binlerce farklı tertipte (kombinasyonda) sorgu yapılarak rapor üretilebilen bir veri ambarı
<b>Milli Eğitim Bakanlığı</b>	MEBBİS	Öğretmenler, okul yöneticileri gibi Milli Eğitim Bakanlığı'na bağlı personelden oluşan kullanıcı gruplarının giriş yapabildiği modüllerden oluşan bir yazılım
	e-Okul	Bir öğrencinin okula kaydından başlayıp, mezuniyetine kadar olan tüm sürece ilişkin veriyi içeren bir sistem
	ALO 147 Projesi	Bakanlığın görev ve sorumluluklarıyla ilgili her türlü talep, şikâyet, görüş ve öneri, ihbar, bilgi edinme soruları ile bakanlığın vermiş olduğu tüm hizmetler hakkında bilgiyi etkin ve hızlı bir biçimde sunabilmek ve sorunların çözüme kavuşturulabilmesi amacıyla kurulan iletişim merkezi
	e-Nabız Projesi	Muayene, tetkik ve tedavi bilgileri başta olmak üzere tüm sağlık verilerinin yönetebildiği, tıbbi özgeçmişe tek bir yerden ulaşabilen bir kişisel sağlık kaydı sistemi
<b>Sağlık Bakanlığı</b>	UBYS Projesi (Sağlık.NET)	Her bireyin kendi bilgilerine erişebildiği, bireyin doğumundan önce başlayıp tüm yaşamı boyunca sağlığıyla ilgili verilerden oluşan işlevsel bir veri tabanının, yüksek bant genişlikli ve tüm ülkeyi kapsayan bir iletişim omurgasında paylaşılması ve tele-tıp uygulamalarına varan teknolojilerin mesleki pratikte kullanılmasını temel alan elektronik kayıt sistemi
	MHRS Projesi	Vatandaşların Merkezi Hekim Randevu Sistemi'ni arayarak canlı operatörlerden veya web üzerinden kendilerine istedikleri hastane ve hekimden randevu alabilecekleri bir uygulama
	Aşı Takip Sistemi	Tüm ülke genelindeki aşıların stok kontrolünden, taşınmasının yapıldığı ve barındırıldığı her noktadaki ısı maruziyetine, uygulanan kişi ve yerine kadar her türlü veri ve bilginin tutulduğu sistem

**Kaynak:** Türkiye Bilişim Derneği, 2016: 19-23 kaynağından tablolaştırılmıştır.

### **3.4. Yerel Yönetimlerde Büyük Veri ve Veri Madenciliğine İlişkin Politika ve Uygulamalar**

Türkiye’de yerel yönetimler, sorumluluk sahasındaki kentsel problemlere çözüm üretme noktasında yetersiz kalmaya başlamıştır. Modern devlet yapılanmasının sonucu olarak yerel yönetimler birçok kamusal hizmeti yerine getirmek durumundadır. Ancak yerel yönetimler sahip oldukları yönetim araçları ile gerekli kamusal hizmeti hızlı, etkin ve verimli bir şekilde yerine getirememektedir (Durna ve Özel, 2008: 3). Sanayileşmenin etkisiyle göçün kırsal alandan kente plansız gerçekleşmesiyle nüfus belli bölgelerde toplanmıştır. Nüfus artışı kent yönetimlerinin yerel sorunları çözebilme kapasitesini zorlamaktadır. Karar vericiler yerel sorunları çözmek, kamusal görevlerini etkin ve verimli bir şekilde gerçekleştirmek için büyük verinin depolanması, işlenmesi ve analizi için yeni bilişim teknolojilerine ihtiyaç duymaktadır.

Yerel yönetimler, kamu hizmetlerinin vatandaşlara ve diğer paydaşlara sunumunda gelişen bilişim teknolojilerinden yoğun şekilde yararlanmaktadır. Kaynaklarını etkin, verimli, hesap verebilir ve şeffaf bir biçimde kullanmak durumunda olan belediyeler e-belediye uygulamaları sayesinde kâğıt ve matbaa masraflarından, personel giderlerine ve zaman tasarrufuna kadar kullanıcılara işlem kolaylıkları sağlamıştır. Bununla birlikte, son yıllarda mobil cihazlar, sosyal medya, akıllı kentler, akıllı kentler, giyilebilir teknolojiler ve nesnelerin interneti gibi bilişim teknolojilerinde yaşanan farklı teknoloji ve çözümler yerel yönetimleri e-belediye hizmetlerinin ötesine geçmeye zorlamaktadır.

E-belediye, belediye kayıtları ve işlemler neticesinde yerel yönetimlerin elinde kentsel hizmetler ve hemşehrilerle ilgili büyük hacimli veri oluşmaktadır. Büyük verinin değerinin farkına varan yerel yönetim kuruluşları büyük veri analiz ve veri madenciliği uygulamalarını sunduğu hizmetlere entegre etmektedir. Hali hazırda operasyonel işlemlerin yürütülmesi ve raporların elde edilmesi imkânına sahip olan yerel yönetimler, geçmişe dönük verileri yorumlayarak geleceğe dönük verilecek kararlarda gerekli olan analizler için veri madenciliği uygulamalarını kullanabilmektedir.

### 3.4.1. Yerel Yönetimlerde Büyük Veri ve Veri Madenciliğine Duyulan İhtiyaç

Gelişen teknoloji ve internet sayesinde vatandaş kamu ve özel sektörle ilgili bilgi alma ve işlem yapmanın yanı sıra işlemlerini yalnızca mesai saatlerinde değil, günün her saati sıra beklemeden yapma imkânına kavuşmuştur. Kurumlar ve vatandaşlar arasında internet vasıtasıyla karşılıklı kurulan iletişim neticesinde, kurumlar vatandaşa ait istek, şikâyet, ödeme bilgisi gibi verilere sahip olmaktadır. Vatandaş, kendisine sunulan kamu hizmetlerinin ihtiyaçlarını daha fazla karşılayacak şekilde kolaylaşmasını ve çeşitlenmesini beklemektedir. Vatandaşa en yakın yönetim birimi olarak yerel yönetimler, vatandaşın bu beklentisine karşılık verebilmek için elindeki verileri kullanmak suretiyle sunduğu hizmetin niteliğini arttırmayı hedeflemektedir. Kullanışlı bilgiyi keşfetmek için, veri madenciliği uygulamalarına ihtiyaç duyulmaktadır.

Yerel yönetimlerdeki veri çeşitliliğine bakıldığında, şehir ile ilgili alt yapıdan üst yapıya, planlamadan sağlığa, güvenlikten ulaşım, eğitimden turizme kadar çok farklı alanlarda veri söz konusudur. Kentteki kurumların vermiş olduğu hizmetler sonucu elde ettikleri verilerin toplanması, saklanması; gerektiğinde vatandaş, özel ve resmî kurumlarla paylaşılması yerel yönetimler tarafından kullanılacak kent bilgi sistemleri ile mümkündür. Bir şehri iyi yönetebilmek, doğru kararlar alabilmek şehirle ilgili doğru bilgiye sahip olmayı ve doğru analiz etmeyi gerektirir. Bahsi geçen analiz işlemleri ise kent bilgi sistemleri altyapısında kullanılacak veri madenciliği uygulamaları ile mümkündür.

Yerel yönetimlerde bilişim teknolojilerinden yararlanarak, kent yaşamından ulaşım hizmetlerine kadar yaşam kalitesi ve zamandan tasarruf edilmesi sağlanabilir. Yerel yönetimlerin ulaşımındaki çeşitli unsurlarının birbirine entegre edilerek akıllı teknolojilerden yararlanılması verilecek hizmet kalitesinin artmasına, enerji tüketiminin azalmasına, çevre kirliliğinin önlenmesine, kaynakların kullanımında tasarrufa gidilmesine ve kazaların önlenmesine olumlu yönde etkisi olacaktır. Kent içindeki trafik probleminin de kurulacak bütünleşik yapı sayesinde hızlı, sağlıklı ve verimli bir şekilde çözülebileceği değerlendirilmektedir. Kent bilgi sistemleri vasıtasıyla toplanan verilerin veri madenciliği uygulamaları yardımıyla analiz edilmesi, karar vericiler tarafından değerlendirilmesi ve kıymetlendirilmesi sonucu bahsi geçen faydaların elde edilmesi mümkün olacaktır.

Mobil teknolojisinin gelişmesiyle birlikte yerel yöneticiler ve vatandaş daha fazla etkileşimde bulunmaktadır. Bu durum şeffaflığın artmasına, vatandaşların daha fazla söz hakkına sahip olmasına, şehir yaşamının gelişmesine katkı sağlamaktadır. Örneğin New York şehrindeki VisionZero insiyatifi, sürücülerin kural ihlallerini ve şehrin zayıf altyapıya sahip noktalarını, şehirde yaşayanlar tarafından harita üzerine işaretlemesine imkân sağlayarak trafik kazalarını azaltmayı amaçlamaktadır (Deloitte, 2016: 83).

Yerel yönetimler, kendilerine yasal olarak verilen görev ve hizmetleri yürütmek için uzun vadeli politikalar ve hedefler oluşturmaktadır. Bu amaçlar yerel yönetim kuruluşları da stratejik plan hazırlamakta ve uygulamaktadır. Stratejik planlamada ihtiyaç duyulan fayda-maliyet analizi, plan bütçe analizi, performans denetimi, mevcut sorunlar ve kurumsal kapasitenin tespiti gibi hususlar karar alma süreçleri yerel yönetimlerin elinde bulunan veriler üzerinden yönetim bilişim sistemlerinin kullanılmasıyla daha etkili bir şekilde yerine getirilebilir.

Geçmişte verilere ulaşım daha uzun zaman alıp, bu verilerden yararlanılarak yapılan analizler ve alınan kararlar daha kısa sürede tamamlanıyordu. Bugün ise, verilere ulaşım hızlanmış ve kolaylaşmış, ancak bu verilere dayalı yapılan analizler de o derece karmaşıklaşmış ve zorlaşmıştır. Bundan dolayı analiz yapma, sonuca ulaşma ve karar verme süreçlerini de bilgisayarlara yaptırabilme fikri doğmuştur. Bunun sonucunda, çeşitli matematiksel ve istatistiksel hesaplamalara dayanan algoritmalar geliştirilmiş ve veri madenciliği kavramı doğmuştur. Hali hazırda birçok büyükşehirde sensörler, kameralar ve GPS cihazları vasıtasıyla veri toplanmaktadır. Bu verileri işleyerek kullanışlı hale getirip, karar alma mekanizmasına eklemleyerek daha etkili ve daha az maliyetli hizmet sunmak mümkündür.

Etkili ve şeffaf bir bütçe ve finans kontrolü, yerel yönetimler için büyük önem arz etmektedir. Açık veri ve şehirle ilgili uygulamalar vatandaşlara ve diğer paydaşlara kamu harcamalarını değerlendirme ve diğer şehir yönetimleriyle karşılaştırma olanağı sunmaktadır. Bütçe dengesinin ve yolsuzluğa karşı önlemlerin şeffaf bir şekilde diğer paydaşlar tarafından takip edilebilmesi şehirleri ekonomik açıdan güçlendirmektedir. Akıllı ödeme sistemleri, kolay, pratik ve verimli olması sebebiyle vatandaşlar ve kamu kurumları için büyük önem arz etmektedir. Çevrimiçi veya mobil kanallar üzerinden çalışan sistem, ödeme işlemleri kayıt altına alarak kurumun şeffaflık ilkelerini

desteklemektedir. Kamu kurumları bu yöntem sayesinde alacaklarını daha kolay ve hızlı bir şekilde tahsil ederken, vezneler önünde biriken kalabalıklar engellenmekte, kişilerin ödeme işlemleri için kayb ettikleri zaman telafi edilmektedir (Deloitte, 2016: 85-86).

Bilişim teknolojilerinin yaygınlaşmaya başlaması ile birlikte pek çok merkezi yönetim, kamu işlemlerinin yönetiminde “nesnelerin interneti” kavramını ilgi çekici bulmaya başlamıştır. Bu kapsamda yerel yönetimlerin ilgisini çeken “akıllı şehir” yönetiminin yaygınlaşması sayesinde “nesnelerin interneti” üzerinden geliştirilen uygulamalar ile ulaşım ve park etme, bakım, aydınlatma ve gözetleme, kültürel mirasın korunması, temizlik işlemleri gibi geleneksel kamu hizmetlerinin optimizasyonu mümkün olabilmektedir (Zanella ve diğerleri, 2014: 22). Toplanan veri türlerinin veri madenciliği uygulamaları ile kullanılması sayesinde vatandaşların şehirlerinin durumu hakkındaki farkındalığının ve şeffaflığının artması sağlanacaktır.

Sonuç olarak; veri madenciliği ile yerel yönetimler, vatandaşın ihtiyaçlarını ve davranışlarını anlama, olayları önceden tahmin etme, bütçelerini ve mevcut insan kaynaklarını en verimli şekilde kullanma gibi olanaklara kavuşabilir.

#### **3.4.2. Yasal Düzenlemeler Açısından Yerel Yönetimlerde Büyük Veri ve Veri Madenciliği**

2004 ve 2005 yıllarında gerçekleştirilen yerel yönetimler reformlarıyla birlikte, yerel yönetimlerin görev ve sorumlulukları yeniden düzenlenmiştir. Reform kapsamında 5216 sayılı Büyükşehir Belediye Kanunu ve 5393 sayılı Belediye Kanunu kabul edilmiştir. Yeni belediye kanunları değerlendirildiğinde belediyelerde yönetim bilgi sistemleri, büyük veri ve açık veri kapsamında görev ve fonksiyonlar bulunmaktadır. İlk olarak, belediyelere verilen stratejik plan hazırlama göreviyle ilgili yasal düzenlemeler incelendiğinde, büyük veri analizi ve veri madenciliği tekniklerinin kullanılmasını teşvik eden düzenlemeler bulunmaktadır.

5393 sayılı Kanununun hemşehri hukuku (md. 13) bölümünde vatandaşların belediyelerin karar ve hizmetlerine katılma, belediye faaliyetleri hakkında bilgilendirme hakkı bulunmaktadır. Bilgi paylaşımının internet üzerinden yapılmasıyla büyük veriye temel teşkil eden açık veri politikalarının desteklendiği söylenebilir.

5393 sayılı Belediye Kanunu'nun 14. maddesi, 5216 sayılı Büyükşehir Belediye Kanunu'nun ise 7. maddesi "coğrafi ve kent bilgi sistemlerini kurma" görevini belediyelere vermiştir. İlgili maddeler kapsamında belediyeler tarafından kurulan coğrafi ve kent bilgi sistemleri çağımızda bilgi toplumuna dönüşüm doğrultusunda kullanımı giderek artan nesnelerin interneti ve akıllı kent teknolojileriyle bağlantılı ve etkileşimli olarak sürekli akan ve değişen büyük veri olarak sayılabilecek veriyi üretmektedir. Coğrafi ve kent bilgi sistemleriyle oluşan veriyle ilgili sıradan analizler ve tahminler günümüzde yetersiz kalmakta, büyük veri ve veri madenciliği metodlarından yararlanılması gerekmektedir.

5393 sayılı Kanun'un 54. maddesine göre belediyelerde denetim amacı, "hizmetlerin süreç ve sonuçlarını mevzuata, önceden belirlenmiş amaç ve hedeflere, performans ölçütlerine ve kalite standartlarına göre tarafsız olarak analiz etmek, karşılaştırmak ve ölçmek; kanıtlara dayalı olarak değerlendirmek, elde edilen sonuçları rapor haline getirerek ilgililere duyurmaktır". Belediyelerde denetim süreçleri, denetime esas teşkil eden veri ve kanıtların bilgi sistemleri kullanılarak analiz edilmesine imkân vermektedir. Buna uygun bilişim altyapısının kurulması durumunda elde edilen büyük verinin analiz edilerek karar vericilere nitelikli bilgi sunulması mümkün olabilir.

5393 sayılı Belediye Kanunu Ek Madde 3'e göre belediyelerin, mevzuatla kendilerine verilen görev ve hizmetlerin yürütülmesi ve vatandaşlar tarafından yapılan başvuruların sonuçlandırılması amacıyla her türlü idari iş ve işlemin yürütüldüğü "e-Belediye bilgi sistemini" kullanmalarını zorunlu hale getirilmiştir. E-Belediye bilgi sistemini kurmaya, işletmeye, veri saklama, veri iletimi ve veri paylaşımı ile ilgili politikaları tespit etmeye, çalışma usul ve esaslarını belirlemeye ve bu sistem ile ilgili merkezî bir hizmet standardizasyonu oluşturmaya İçişleri Bakanlığı yetkili kılınmıştır.

Aynı kanunun Geçici 10. Maddesine göre ise, e-Belediye bilgi sisteminin kurulduğuna dair bildirim İçişleri Bakanlığı tarafından yapılmasından itibaren, e-Belediye bilgi sistemi ile ilgili çalışmaların bir yıl içinde tamamlanması yükümlülüğü getirilmiştir. Bu kanunla birlikte belediyelerin "E-Belediye bilgi sistemini" kullanmaları sağlanarak vatandaşlara dönük hizmetlerin basitleştirilmesi ve standart hale getirilmesi, kamu kurumları ve yerel yönetimler arasında anlık veri paylaşımının sağlanması, vatandaşın işlemlerini elektronik ortamda yapabilmesi, bürokratik işlemlerin azaltılması,

hizmetlerin şeffaf ve denetlenebilir olması hedeflenmektedir. Bahse konu hedeflerin gerçekleştirilmesi için geliştirilecek olan belediyeçilik yazılımları ile vatandaş odaklı bir yerel yönetim yapısının oluşturulması için bilgi ve iletişim teknolojileri etkin şekilde kullanılması öngörülmektedir. Geliştirilmesi planlanan belediyeçilik yazılımlarının oluşacak büyük veriden faydalanması ve veri madenciliği yazılımlarının kullanılması halinde belirlenen hedeflere ulaşılması mümkün olabilecektir.

Yerel yönetimlerin de içinde bulunduğu kamu kurum ve kuruluşlarında teknolojik altyapının büyümesine paralel olarak gerekli teknik desteği sağlayacak nitelikli personeli bulmanın zorlaşması üzerine çıkarılan yönetmelik (Başbakanlık, 2008) sayesinde 2008 yılından bu yana büyük ölçekli bilgi işlem birimi olan kamu kurum/kuruluşlarında sözleşmeli bilişim personeli çalıştırılabilme imkânı bulunmaktadır. Kamu kurumlarında nitelikli personel istihdamının sürekliliğinin zor olduğu görülmektedir. Bu nedenle personel yetkinliğinin azaltılmasına yönelik olarak kurumsal hafızanın artırılması ve kuruma ait teknik yeterliliğin artırılması gerekmektedir (Ulaştırma Denizcilik ve Haberleşme Bakanlığı, 2016).

Yasal düzenlemeler incelendiğinde Türkiye’de büyük veriyi ve veri madenciliğini doğrudan düzenleyen bir mevzuatın bulunmadığı görülmektedir. Bununla birlikte, 6698 sayılı “Kişisel Verilerin Korunması Kanunu” açısından kişisel verilerin korunması anlamında büyük verinin kullanılması ve uygulanmasında yasal engel görülmemektedir (Altun ve diğerleri, 2017: 2040). Sonuç itibarıyla büyük veri, açık veri ve veri madenciliği uygulamalarıyla nitelikli nüfus yapısına yönelik etkili kentsel hizmet ve politikaların sunulması mümkündür. Ancak Türkiye’de yerel yönetimler alanında büyük veri teknolojileri ve veri madenciliği tekniklerinin kullanılmasına ilişkin hukuki boşluğun doldurulması için gerekli düzenlemelerin yapılmasına duyulan ihtiyaç da devam etmektedir.

### **3.4.3. Yerel Yönetimlerde Büyük Veri ve Veri Madenciliği Uygulamaları**

Toplumsal yapı ve ilişkilerin hızlı değişim ve dönüşüme uğradığı günümüzde kamu yönetimi düşüncesi, yapısı ve fonksiyonları da değişime uğramakta, devletin görevleri ve iş yapma şekli de değişmektedir. Geleneksel kamu yönetiminin yeterli olmadığı günümüzde yeni kamu yönetimi anlayışının da etkisiyle ülkemizde de çağdaş yönetim anlayışına doğru geçiş söz konusudur. Bu kapsamda hazırlanan stratejik planlar

belediyelere deęişim için yol haritası, stratejilerin oluşturulması, başarılı yönetim için gerekli bilginin elde edilmesine katkı ve stratejik hedeflere ulaşmak için gerekli bakış açısının kazanılmasına yardımcı olmaktadır.

Türkiye’deki belediyelerin tamamının stratejik planını incelemek bu çalışmanın sınırlarını aşacağından ve ayrıca ölçek ve kapasite olarak büyük şehir belediyelerinin büyük veri, yeni teknolojiler ve veri madencilięi alanlarında strateji, politika ve proje geliştirdikleri düşünöldüğünden, çalışmanın bu bölümünde 30 büyükşehir belediyesinin stratejik planları içerik analiz yöntemiyle deęerlendirilmiştir. Bu bağlamda, stratejik planlarında büyük veri ve veri madencilięine ilişkin doğrudan ve dolaylı amaç ve hedefleri bulunan 7 büyükşehir belediyesi Tablo 9’da verilmektedir.

**Tablo 9**  
**Büyükşehir Belediyelerinin Büyük Veri ve Veri Madencilięine İlişkin Amaç ve Hedefleri**

<b>Büyükşehir Stratejik Planı</b>	<b>Stratejik Amaç</b>	<b>Stratejik Hedef</b>
<b>Gaziantep Büyükşehir Belediyesi</b>	“Kurumsal kapasitenin geliştirilmesi”	“Bilgi sistemlerinin kurulması ve kararların doğru veri üzerinden alınmasını sağlamak”
	“Sosyal hizmet uygulamalarının etkinleştirilmesi”	“Veriye dayalı bütünleşik sosyal yardım sistemi kurmak”
<b>Hatay Büyükşehir Belediyesi</b>	“Yönetim bilgi sisteminin kurulumu ve geliştirilmesine yönelik bilişim altyapısının oluşturulması”	“Arşiv tasnifinin yapılarak dijital ortama aktarılması, e-belediye hizmetleri ile kent bilgi sisteminin kapasitesinin artırılması”
<b>İstanbul Büyükşehir Belediyesi</b>	“Kentın doğal tarihi ve kültürel mirasına ...yön vererek yaşanabilir bir kent oluşumuna katkı sağlamak”	“Standart ve nitelikli coęrafi bilginin üretilmesini ve paylaşımını ... planlama, yatırım ve denetim faaliyetlerinin ..katkı sağlamak”
	“Hizmet Alanlarıyla İlgili Doğru Bilgiye Hızlı Erişim Sağlamak”	“Kurumsal Karar Verme Sürecini Etkin Yönetmek ve Kararlara İnternet Üzerinden Erişimi Sağlamak”
<b>Kocaeli Büyükşehir Belediyesi</b>	“Sunulan hizmetlerde bilgi teknolojisinin kullanımını yaygınlaştırmak”	“Belediye yazılım hizmetleri ve e-Belediyecilik uygulamaları geliştirilecektir”

<b>Büyükşehir Stratejik Planı</b>	<b>Stratejik Amaç</b>	<b>Stratejik Hedef</b>
<b>Konya Büyükşehir Belediyesi</b>	“Teknolojinin İmar ve Planlama Süreçlerinde Kullanılmasını Sağlamak”	“Veri Paylaşım Protokollerinin Kurumlar arası Sayısını Arttırmak Kent Bilgi Sistemi Altyapısını Diğer Kamu Kurumlarına Yaygınlaştırmak, İş Zekâsı Projelerini Geliştirmek, Belediyeciliğe Ait Her Türlü Verinin Tek Merkezde Toplanacağı ve Yönetileceği CBS ni Kullanan Merkez Oluşturmak, Belediye Yönetim ve Karar Aşamalarında Teknolojinin Etkin Biçimde Kullanmak”
<b>Malatya Büyükşehir Belediyesi</b>	“Kurum tarafından yürütülen işlemlerin dijital ortamda takip edilmesi, birim ve kullanıcı performanslarının ölçülmesi, verilerin tek bir kaynakta güncel tutulmasını sağlamak”	“Online Başvuru Sistemi Uygulamasının geliştirilmesi ve devreye alınması”
	“Tüm belgelerin dijital ortamda üretilerek zaman, kâğıt ve enerji tasarrufu sağlanması ile dijital ortamda tutulması”	“Kurum tarafından üretilen resmi yazışmalar dışındaki diğer belgelerin dijital ve e-imzalı olarak üretilmesi”
	“Kurumlar arası ortak veri havuzunu oluşturmak”	“Açık veriler ile İlgili Malatya genelinde çalışma yapmak”
<b>Ordu Büyükşehir Belediyesi</b>	“Belediyemizin bilgi sistemlerini geliştirerek kurumsal verimlilik, hizmet kalitesinin yükseltilmesi ve bilgiyi etkin kullanan bir yapıya kavuşturmak”	“İnternet Belediyeciliği hizmetlerinin iyileştirilmesi ve yaygınlaştırılarak bilgi ve haberleşmenin sürekliliğinin sağlanması”

Büyükşehir belediyelerine ait stratejik planlar incelendiğinde kurumsal kapasiteyi geliştirme amacıyla özellikle yeni kurulan büyükşehir belediyelerinde yönetim bilişim sistemlerinin kurulması ya da yaygınlaştırılmasının ortak stratejik hedef olarak belirlendiği görülmektedir. Bahse konu yönetim bilişim sistemleri, yerel yönetim birimleri tarafından kurum içi hizmetler, vatandaşa hizmet ve kurumlar arası paylaşım başta olmak üzere farklı amaçlarla kullanılmaktadır. Büyük veri, bulut bilişim, artırılmış gerçeklik ve akıllı kent gibi yenilikçi teknolojiler yerel yönetimlerin coğrafi bilgi

sistemleri, e-belediyecilik ve çevre hizmetleri gibi farklı alanlarda vatandaşa hizmet sunmak, etkin, verimli, şeffaf ve hesapverebilir bir yönetim birimi olmak amacıyla yoğun şekilde kullanılmaya başlanmıştır.

Stratejik planlar incelendiğinde büyükşehir belediyeleri tarafından büyük veri ve veri madenciliğine yönelik açık şekilde atıfta bulunulmadığı görülmektedir. Diğer taraftan, stratejik planlarda yer verilen e-belediye uygulamaları olarak tanımlanan yönetim bilişim sistemleri, akıllı kent uygulamaları ve yenilikçi bilgi teknolojileri yardımıyla sunulan hizmetlerin neredeyse tamamında veri kullanılmaktadır. Bahse konu hizmetler üzerinden karar vericilere destek sağlanması, gerekli analizlerin gerçekleştirilmesi, raporlamaların yapılması ve kentin geleceğinin planlanmasında büyük veri ve veri madenciliği uygulamalarından yararlanılması ihtiyacının olduğu değerlendirilmektedir. Nitekim web siteleri üzerinde derinlemesine araştırma gerçekleştirildiğinde veri tekilleştirme, veri madenciliği, iş zekâsı ürünleri ile süreç analizi ve otomasyon sistemleriyle yorumlanması ve elektronik çağın kullandığı teknik altyapılar ile yönetimin güçlendirilmesi (Demir, 2017: 1), veri merkezi ile çok fazla teknoloji barındıran hizmetin bulut servisleriyle birlikte hayata geçirilmesi (Kuru, 2017: 1) gibi hizmetlerin yerel yönetim birimleri tarafından kullanıldığı görülmektedir.

## **BÖLÜM 4: KAMUDA VERİ MADENCİLİĞİ UYGULAMASI OLARAK ÇÖZÜM MASASI HİZMETLERİ ÖRNEĞİ**

Yönetim bilişim sistemlerinin kamu kurumlarında yaygın bir şekilde kullanılması, internetin bireyler tarafından başta sosyal medya ve kamu hizmetlerine ilişkin işlemler olmak üzere artan kullanımı sayesinde kamu yönetiminin elinde yüksek hacimli veri depolanmakta, bu veri işlenmekte ve yönetilmektedir. Kamu kurumlarında hizmetin gereği olarak saklanan verinin yanı sıra yapısal olmayan çok sayıda veri depolanmaktadır. Daha önce yapısal olmadığı için kurumlar tarafından kullanılamayacağı düşünülen veri üzerinde doğru analiz yöntemlerinin uygulanmasıyla, kurumların politika üretiminden stratejik kararlar almasına kadar farklı pek çok alanda katkı sağlayacağı düşünülmektedir. Kamu kurumlarının sahip olduğu verinin ciddi bir kısmı yapısal olmayan çok çeşitli (variety, çeşitlilik), üretilme hızı çok yüksek olan (velocity, hız), büyüklüğü (volume, hacim) her geçen gün artan, gerekli güvenlik ve gizlilik seviyesinde (verification, doğrulama) ve değer yaratan (value, değer) özelliktedir. Yani söz konusu veri büyük verinin özelliklerini taşımaktadır. Özel sektör bu veri sayesinde kullanıcılarına bireysel ürünler sunabilmekte, yeni pazar stratejileri üretebilmektedir. Kamu kurumları da sahip olduğu verinin farkına varmaya başlayarak elindeki veri sayesinde kritik noktalarda karar vericilerin doğru kararlar almasına yardımcı olabilmek, vatandaşa daha iyi hizmet sunabilmek, kamu kaynaklarını daha verimli kullanabilmek, şeffaf ve hesapverebilir bir kamu yönetimini oluşturabilmek amacıyla büyük veriden yararlanmayı hedeflemektedir.

Veri madenciliği sayesinde kamu kaynakları optimum şekilde paylaşılabilir, yöneticilere geleceğe yönelik karar almada kullanabilecekleri bilimsel bilgi sağlanmakta, bütçe ve insan kaynakları verimli şekilde kullanılabilir. Aynı zamanda iş zekâsı uygulamalarından yararlanmak suretiyle veri madenciliğinden elde edilen bilgi kolay anlaşılabilir grafiklere dönüştürülmek suretiyle kamu yöneticileri ve vatandaşların sayısal bilgiyi yorumlaması mümkün hale gelmektedir.

Çalışmanın bu bölümünde, Türkiye’de nispeten yeni yeni keşfedilmeye başlanan büyük verinin önemini ve veri madenciliği yöntemiyle sadece sayısal verinin değil metinlerin de analiz edilerek kamusal kararların alınmasında etkili çözümler üretilebileceğini göstermek amacıyla Kocaeli Büyükşehir Belediyesi’nin çözüm masası verilerine dayalı

olarak veri madenciliği uygulaması gerçekleştirilmektedir. Ayrıca veri madenciliği uygulaması sonucunda elde edilen verinin iş zekâsı ürünleri yoluyla karar destek sistemi olarak tasarlanma potansiyeline sahip olduğu gösterilmektedir.

Bu bağlamda, yapılandırılmamış çözüm masası verisi üzerinde veri madenciliği uygulaması için metin madenciliği süreçleri takip edilmektedir. Metin madenciliği, veri madenciliği sürecinde yer alan işin tanımlanması, verinin anlamlandırılması, verinin hazırlanması, modellenmesi, değerlendirilmesi ve uygulanması aşamalarını içermektedir. Metin madenciliği çalışmalarında başarıya ulaşabilmek için iyi kurgulanmış uygulama temelli bir metodoloji izlenmelidir. Metin üzerinde veri madenciliği uygulamaları için endüstri standardı olarak kabul edilen ve birinci bölümde detayları açıklanan CRISP-DM'ye benzer standartlaştırılmış bir süreç modeli kullanılmaktadır. Çözüm masasına yapılan başvurulardan elde edilen veri seti üzerinden, veri madenciliği uygulama sürecine uygun geliştirme adımları modelin geliştirilmesi için uygulanmıştır.

Çözüm masası verilerinin tercih edilmesinde belediyelerin çoğunda el yordamıyla ve basit ofis uygulamalarıyla toplanan ve değerlendirilen çözüm masası verilerinin bir veri madenciliği türü olan metin madenciliği ile analiz edilebileceği düşünülmüştür. Bu sonuca Türkiye’de merkezi kamu kurumları ile özel olarak yerel yönetimlerde veri madenciliği uygulamalarının ne düzeyde uygulandığı büyükşehir belediyeleri ölçeğinde yapılan anketle (Ek 1.1 ve Ek 1.2) ulaşılmıştır. Bu doğrultuda; büyükşehir belediyelerinin “iletişim”, “çözüm masası”, “beyaz masa” ve “bize ulaşın” şeklindeki iletişim kanalları üzerinden elde edilen e-posta adreslerine 04 Mayıs 2018 tarihinde e-posta ile düzenlenen anket formu gönderilmiştir. 16 Mayıs 2018 tarihinde büyükşehir belediyelerinin bilgi işlem şubelerinde ilgili personele telefonla ulaşılarak, büyük veri ve veri madenciliği konusunda anket soruları doğrultusunda mülakat ile bilgi alınmaya çalışılmıştır.

Veri madenciliğine yönelik gerçekleştirilen anket ve yapılan mülakatlar, veri madenciliğinin büyükşehir belediyelerinde henüz yoğun şekilde kullanılmaya başlanmadığını göstermektedir. Fakat yönetim bilişim sistemlerinin büyükşehir belediyelerinin süreçlerini otomasyon sistemleri üzerinden kullanılacak şekilde dönüştürmesi, nüfus bilgileri, konum bilgileri gibi verinin farklı kurumlardan

paylaşılarak alınması neticesinde yerel yönetimlerin elinde büyük verinin yakın dönemde ciddi boyutlara ulaşması beklenebilir. Akıllı kent uygulamalarının hayata geçirilmesi ve yenilikçi teknolojilerin kullanımının artması ile birlikte merkezi olarak belirlenen ulusal politika ve programlara uygun şekilde yerel yönetimler tarafından da büyük veri ve veri madenciliği uygulamalarına daha yoğun şekilde yer verilecektir.

Çözüm Masasına gelen şikâyet/istek/taleplerin ilgili belediye birimlerine iletilmesi, çözüm sürecinin takip edilmesi ve ilgili vatandaşa geri bildirim/bilgilendirme yapılmasında 30 büyükşehir belediyesine gönderilen ankete yanıt veren 3 adet büyükşehir belediyesinde veri madenciliği ya da metin madenciliği yöntemlerinden Eskişehir ve Tekirdağ Büyükşehir Belediyesi tarafından yararlanılmadığı, Hatay Büyükşehir Belediyesi tarafından ise yararlanıldığı bildirilmiştir. İstanbul Büyükşehir Belediyesi'nin "Beyaz Masa - Alo153" hizmetini telefon, sosyal medya, mobil, yüz yüze ve faksla olmak üzere 700 temsilci üzerinden 7x24 esasına göre Türkçe dışında Afrika Dilinden Zulu diline olmak üzere 103 farklı dilde verdiği web sitesinde yapılan derinlemesine inceleme ile tespit edilmiştir. Aynı zamanda trafik ağı, nüfus kütükleri, vatandaşlık bilgileri, afet koordinasyon merkezi veritabanı, harita uygulamaları ile entegrasyon ve başvuruya göre lokasyon tespiti ile farklı kanallardan elde edilen veri ile çevrimiçi altyapı ve ileri veri bütünlüğü sağlanmaktadır.

Çözüm masası faaliyetlerinin (gelen başvuruların sayısı, türü, çözülen başvuru oranı vs.) genellikle kullanılan yönetim bilişim sistemleri uygulamaları üzerinden, çoğunlukla ise word veya excel tablosu şeklinde otomasyon kullanılmadan raporlandığı anlaşılmaktadır. İstanbul Büyükşehir Belediyesi'nin başvuruları gerçek zamanlı, zaman çizelgesiyle tarihsel dokümantasyon ve veri madenciliği uygulamalarından yararlanmak suretiyle raporlanmaktadır. Büyükşehir belediyelerinde veri madenciliği uygulaması olarak değerlendirilen diğer çözümlerin olup olmadığı, varsa kurumun hangi bölümünde ne amaçla (politika üretimi/çözümü vb.) kullanıldığı sorusuna yapılan mülakatlar neticesinde tam olarak tespit edilememiştir. Kurum içindeki birimlerin birbiri ile entegre ve uyumlu şekilde çalışmaması, birbirlerinden habersiz şekilde uygulamalar kullanıyor olması, diğer birimlerin sorumluluk alanlarındaki uygulamalar hakkında bilgi sahibi olmadıkları anlaşılmaktadır. Fakat yerel yönetimler tarafından kullanılmaya başlanan e-belediyeçilik çözümleri adı altında yönetim bilişim sistemleri

projeleri yöneticilerin iş süreçlerini ve performans izleyebileceği, kurumun stratejik hedefleri doğrultusunda yönlendirilmesi, yönetim bilişim sistemleri üzerinden elde edilen verinin analiz edilebildiği, raporların oluşturulabildiği, bu sayede yöneticilerin karar verme sürecine katkı sağlandığı değerlendirilmektedir.

#### **4.1. İşin Tanımlanması**

Veri madenciliği uygulamalarının ilk aşamasında işin tanımlanmasına yönelik olarak sürecin ana hatlarıyla belirlenmesi ve ulaşılmak istenen iş hedeflerinin belirlenmesi, mevcut durumun değerlendirilmesi, veri madenciliği amaçlarının oluşturulması ve uygulama planının hazırlanması üzerinde durulmaktadır.

Bilgi teknolojilerinde yaşanan gelişmeler ve kamu kurumlarının sahip olduğu verinin günden güne artması, kamunun veriden geleceğe dönük yararlı tahminler üretme ihtiyacını da beraberinde getirmiştir. Kamu kurumlarının vatandaşın ihtiyaçlarını ve davranışını anlayarak olayları önceden tahmin etmesine yardımcı olmak, bütçe ve mevcut insan kaynaklarını verimli şekilde kullanmasını sağlamak üzere veri madenciliği uygulaması kullanılmaktadır.

Yapılan çalışma, herhangi bir ekip çalışması olmadan kısıtlı olanaklarla geliştirilmiş olup, ideal bir veri madenciliği projesinde izlenmesi gereken adımlar izlenmek suretiyle kamuda veri madenciliğinin bütçe, finans ve sağlık gibi bilindik uygulamaların ötesinde farklı hizmet alanlarında da uygulanabileceğini göstermek adına örnek bir uygulama niteliğindedir. Uygulamayı gerçekleştirmek üzere afet verileri, trafik verileri ve çözüm masası verileri ile çalışılabileceği değerlendirilmiştir. Çalışmayla ilgili Kocaeli ili ulaşım kolaylığı ve ilgili kamu kurumlarının veri paylaşımı ile ilgili talebimize olumlu yanıt vermesi nedeniyle tercih edilmiştir. Kocaeli İl Afet ve Acil Durum Müdürlüğü ile gerçekleştirilen görüşmede afet verilerinin merkezi olarak tutulduğu, Kocaeli Büyükşehir Belediyesi Ulaşım Dairesi Başkanlığı ile gerçekleştirilen görüşmede trafik ve ulaşım verileri kullanılarak özel sektör desteğiyle proje geliştirildiği bilgisi edinilmiştir. Çalışmanın özgün olması amacıyla, Kocaeli Büyükşehir Belediyesi'nin çözüm masası verileri kullanılmıştır. Farklı belediyelerin çözüm masası verilerinin birleştirilerek daha büyük bir veri seti elde edilebilirdi. Ancak belediyelerden veri elde etmedeki güçlükler nedeniyle ve bu tez kapsamında veri madenciliği uygulamasını

gerçekleştirmek için yeterli büyüklüğe sahip olduğu düşünüldüğünden uygulamada aşağıda açıklanan veri seti kullanılmıştır.

Bu bağlamda, veri madenciliğinin sınıflandırma modeline ilişkin teknikler Kocaeli Büyükşehir Belediyesi tarafından sunulan “Çözüm Masası” hizmetlerine ait veri üzerinde uygulanmaktadır. Uygulamada kullanılacak veri seti büyükşehir belediyesi çözüm masasına vatandaşların yaptığı başvurular ve vatandaşlara ait veri içermesi sebebiyle gizlilik içermektedir. Bu açıdan vatandaşlara ilişkin gizli verinin paylaşılmasına özen gösterilmiştir. İlk adım olarak, “kamu kurumlarının elinde bulunan veriyi karar verme amaçlı geleceğe dönük tahminlerde kullanmak üzere modelin geliştirilmesi” şeklinde iş tanımlanmıştır.

#### 4.2. Veriyi Anlama

Veri madenciliği uygulaması için belirlenen iş ve veri madenciliği hedefleri doğrultusunda veri setinin incelenmesi veriyi anlama aşamasını oluşturmaktadır. Bu aşamada analizde yararlanılacak olan Kocaeli Büyükşehir Belediyesi’ne ait “Çözüm Masası” verisi hazırlanmıştır. Verinin toplanması ile başlayan veriyi anlama aşaması, teknik personelin toplanan veri üzerinde bilgi sahibi olması ve veri kalitesi problemlerinin tespitiyle devam etmiştir.

CRISP-DM modelinin büyük kısmı metin madenciliği projelerindeki süreçlerde de geçerli olmakla birlikte çok daha ayrıntılı ve detaylı bir veri ön işleme süreci gerektirmektedir (Sharda ve diğerleri, 2014: 244). Metin madenciliği süreci ve çevre bileşenlerle ilişkisi Şekil 19’da görülmektedir.



Şekil 19: Metin Madenciliği Süreci ve Çevre Bileşenlerle İlişkisi

Metin olarak yapılandırılmamış ve yapılandırılmış veri halinde bulunan veritabanlarının veri kaynağı olarak kullanılması neticesinde bilgi elde edilmektedir. Metin madenciliği sürecinin çıktısı olarak Şekil 19'un sağ tarafında görüleceği üzere karar vermede kullanılacak amaca yönelik bilgi elde edilmektedir. Metin madenciliği süreçlerinde Şekil 19'un üst kısmında belirtilen yazılım/donanım sınırlamaları, gizlilik sorunları ve dilbilimsel sorunlar metin işlemede karşımıza çıkan problemlerdir. Teknolojik altyapıda meydana gelen ilerlemeler neticesinde yazılım/donanım sınırlamalarının olumsuz etkisi en az düzeye inmiş durumdadır. Metin madenciliğinde veriden bilgiyi elde etmek üzere Şekil 19'un alt kısmında gösterilen ilgili yazılımlara, uygun tekniklere/algoritmalara ve bu işlemleri gerçekleştirebilecek alan uzmanlarına ihtiyaç vardır. Metin madenciliğinde ulaşılmak istenen birincil hedef, yapılandırılmamış veriyi işlemek suretiyle daha iyi karar vermede kullanmak için anlamlı ve uygulanabilir kalıplar çıkarmaktır. Bu noktada en önemli problem, metinlerin semantik özelliklerine uygun şekilde sayısal öznitelik vektörlerine çevrilmesi işlemidir.

Bu çalışmada, çözüm masasına başvuruda bulunan vatandaşların girmiş oldukları veri üzerinden bir veri seti (korpus) oluşturmak amacıyla Kocaeli Büyükşehir Belediyesi (KBB) Yönetim Bilişim Sistemleri Şube Müdürlüğü personeli ile görüşme gerçekleştirilmiştir. Yapılan görüşme neticesinde Kocaeli Büyükşehir Belediyesi'nden çözüm masasına ait ham veriler talep edilmiştir. Elde edilen ham verinin sorgulanabilir hale getirilmesi amacıyla sanal sunucu uygulaması olarak açık kaynak (ücretsiz) kodlu VirtualBox yazılımı üzerine Windows 10 sanal işletim sistemi kurulmuştur. Çözüm masası verisinin "Microsoft SQL Server" veritabanı formatında olması nedeniyle ilgili sanal bilgisayara "Microsoft SQL Server"ın ücretsiz sürümü olan "Express" sürümü kurularak alınan ham veri, veritabanına taşınmıştır.

**Veri Seti:** Bu çalışmada, Kocaeli Büyükşehir Belediyesi Çözüm Masasına ulaşan talep/istek/şikâyet gibi başvurulardan oluşan veri seti kullanılmıştır. Vatandaş başvurusunu çözüm masasına yüz yüze, web veya mobil erişim gibi çeşitli yöntemlerle gerçekleştirebilmektedir. Şekil 20'de web üzerinden gerçekleştirilebilecek "Çözüm Masası Başvuru Ekranı" görülmektedir.

**Çözüm Masası Başvuru**

Formdaki zorunlu alanları doldurmayı unutmayınız.

**Lütfen Dikkat!**  
Sistem üzerinden yapılan başvurular, birimlerimizce mesai saatleri içerisinde yönlendirilmektedir.  
Mesai saatleri dışındaki **acil durumlarda**, 444 11 41 Çağrı Merkezimizi arayarak, operatörlerimizden bilgi alabilirsiniz.

Tc Kimlik No \*

Ad \*

Soyad \*

Başvurunuz Gizli mi ?

**Başvurunuz ile ilgili bilgilerinizin gizli tutulmasını istiyorsanız işaretleyiniz.**

Cep Telefonu\* (Lütfen Başında 0 Olmadan Giriniz)

Email (Size daha hızlı geri dönüş yapabilmemiz için email adresinizi giriniz.)

Dilek / İstek / Şikayeti\*

Eklenecek Dosya (Birden fazla dosya eklemek istiyorsanız .rar veya .zip formatında sıkıştırarak gönderebilirsiniz)  
Dosya Seç Dosya seçilmedi

Adres Tarifi (Şikayete Konu Adres)\*

**Şekil 20: Çözüm Masası Başvuru Ekranı**

Vatandaşın yaptığı başvuruların yönetimi, Kocaeli Büyükşehir Belediyesi ilgili personeli tarafından Şekil 21’de görülen “Çözüm Masası Aktivite İşlemleri” ekranından gerçekleştirilmektedir. Çözüm masası aktivite işlemleri yönetim ekranı üzerinden arama işlemleri çözüm masası aktivite işlemleri, çözüm masası raporları, kişisel bilgiler ve kullanıcı yönetimi işlemleri gerçekleştirilmektedir.

KOCAELİ BÜYÜKŞEHİR BELEDİYESİ

KOCAELİ BEY HİZMETLERİ BAĞLI BİRİMLERİN LİSTESİ (KURUMSAL YAPISI) - İZMİR - KOCALIZ / BEŞİRTİMDEN ÇIKIŞ

ARAMA MOTORU | YENİ DOSYA AÇ | ÇÖZÜM MASASI AKTİVİTE İŞLEMLERİ | ÇÖZÜM MASASI RAPORLARI | BİTİŞSEL BİLGİLER | KULLANICI YÖNETİMİ

Dosya No: 8108 | Gözlemler | Çözüm Masası Başvuruları | Çözüm Masası Aktivite İşlemleri | Belge

Çözüm Masası Aktivite İşlemleri

Başvuru No	Başvuru Şekli	Personel	Departman	Başvuru Tipi	Başvuru Türü	İşlem	Başvuru Tarihi
99019	MOBİL	ibb.park.burolar@ibb.kocaeli.tr	Park ve Bahçeler Saha Müdürlüğü	TALEP	Sekizparç	BİLDİRİLDİ	19.08.2016
87043	MOBİL	ibb.cocum.mellie	Çocuk Masası	BİLGİ	Çözüm Başvuru	REDDEDİLDİ	19.01.2016
80412	EVROMOBİL	ibb.cocum.mellie	Çocuk Masası	BİLGİ	Çözüm Başvuru	REDDEDİLDİ	18.01.2016
87339	WEB	ibb.cocum.mellie	İhtifa	TALEP	Paç Keşni Kağıt Sıkıştı	TANIMLANMIŞ	21.12.2015
85028	WEB	ibb.cocum.mellie	Çocuk Masası	BİLGİ	Çözüm Başvuru	REDDEDİLDİ	17.11.2015
85026	WEB	ibb.cocum.mellie	Çocuk Masası	BİLGİ	Çözüm Başvuru	REDDEDİLDİ	17.11.2015
82476	WEB	ibb.cocum.mellie	Çocuk Masası	BİLGİ	Çözüm Başvuru	REDDEDİLDİ	08.10.2015
79189	MOBİL	ibb.cocum.cocuk	Çocuk Masası	BİLGİ	Bilgi	REDDEDİLDİ	20.08.2015
77390	WEB	ibb.cocum.cocuk	Çocuk Masası	BİLGİ	Bilgi	REDDEDİLDİ	13.07.2015
76787	MOBİL	ibb.cocum.mellie	Çocuk Masası	BİLGİ	Çözüm Başvuru	REDDEDİLDİ	02.07.2015
76073	MOBİL	ibb.cocum.mellie	Çocuk Masası	BİLGİ	Çözüm Başvuru	REDDEDİLDİ	18.06.2015
76030	MOBİL	ibb.cocum.mellie	Çocuk Masası	BİLGİ	Çözüm Başvuru	REDDEDİLDİ	19.06.2015
76048	MOBİL	ibb.cocum.mellie	Çocuk Masası	BİLGİ	Çözüm Başvuru	REDDEDİLDİ	19.06.2015
72043	MOBİL	ibb.cocum.cocuk	Çocuk Masası	BİLGİ	Bilgi	REDDEDİLDİ	17.04.2015
82967	MOBİL	ibb.cocum.cocuk	Park ve Bahçeler Saha Müdürlüğü	TALEP	Çöpün Yatırma ve Sıkıştırma	TANIMLANMIŞ	21.09.2014

Arama Detayları

BİTİŞSEL | Tanımlandı | İşlem Durumu | Tanımlanmış Kapasite | Reddedilmiş Kapasite | Onaylandı

Başvuru No: 99019 | Başvuru Tarihi: 19.08.2016 | Başvuru Şekli: MOBİL | Adı Soyadı: FATMA ACAR | Adres: | Şikayete Konu Adres: | Telefon1: 0537783380 | Telefon2: 0537783380 | EPOSTA: |

Bellekme Durumu

Verilen Departman \* | Park ve Bahçeler Saha Müdürlüğü

Başvuru Tipi \* | TALEP

Başvuru Türü \* | Sekizparç

Verilen Personel \* | ibb.parkbahcesaha@ibb.kocaeli.tr

Aktivite \* | ÇÖZÜM MASASI BAŞVURUSU

Yapılanlar Gösteriler

Sekizparç için gizli faaliyet sonuçları için çok tehlikeli. Deniz kenarından 3 metre yükseklikte bulunuyor için yer ver faaliyet karar kısıtlılığı için. Kazara oluşan bir ota çıkılarak bir menden bile yok. Biraz önce devam alınması. (25 tane fotoğrafın olduğu yer.)

Yapılanlar

Sekizparç için gizli faaliyet sonuçları için çok tehlikeli. Deniz kenarından 3 metre yükseklikte bulunuyor için yer ver faaliyet karar kısıtlılığı için. Kazara oluşan bir ota çıkılarak bir menden bile yok. Biraz önce devam alınması. (25 tane fotoğrafın olduğu yer.)

Tanımlanmış Güncelleme

Personel / İşlem Tarihi | ibb.park.burolar / 20.07.2016 14:49:00

BİTİŞSEL | Tanımlandı | İşlem Durumu | Tanımlanmış Kapasite | Reddedilmiş Kapasite | Onaylandı

Yazdır

**Şekil 21: Çözüm Masası Aktivite İşlemleri**

“Çözüm Masası” verisi Şekil 22’de görülen sorgu ile çözüm masası veritabanından çekilmiştir. Çözüm masası başvuru tipine göre 1\_Talep, 2\_Şikâyet, 3\_Öneri, 4\_Teşekkür, 5\_Bilgi, 6\_Tüketici\_Hakları ve 7\_İştalebi şeklinde veri çekilmiştir.

- 1 Talep
- 2 Şikayet
- 3 Öneri
- 4 Teşekkür
- 5 Bilgi
- 6 Tüketici Hakları
- 7 İş Talebi

```

select CM.BASVURUTAR,CM.BASVURUNO,cm.DEPARTMAN,DEPT.DAIRE,
CM.DEPARTMANANASORUN,DEPT_ANA_SRN.ICERIK AS DEPARTMANANASORUNTEXT,
CM.DEPARTMANSORUN,DEPT_SRN.ICERIK AS DEPARTMANSORUNTEXT,
CM.BASVURUDURUM,BSVDRM.ICERIK AS BASVURUDURUMTEXT,
cm.BASVURUTIP,BSVTIP.ICERIK AS BASVURUTIPITEXT, LOWER(REPLACE(CM.ACIKLAMA,',',' ')) AS BASVURUTEXT
FROM [COZUMMASASI].[dbo].[COZUMMASASI_BASVURU] as CM
INNER JOIN [COZUMMASASI].[dbo].[T_COZUMMASASI_DEPARTMAN] AS DEPT ON DEPT.TUR=CM.DEPARTMAN
INNER JOIN [COZUMMASASI].[dbo].[T_COZUMMASASI_DEPARTMAN_ANA_SORUN] AS DEPT_ANA_SRN ON DEPT_ANA_SRN.TUR=CM.[DEPARTMANANASORU]
INNER JOIN [COZUMMASASI].[dbo].[T_COZUMMASASI_DEPARTMAN_SORUN] AS DEPT_SRN ON DEPT_SRN.TUR=CM.[DEPARTMANSORUN]
INNER JOIN [COZUMMASASI].[dbo].[T_COZUMMASASI_BASVURUDURUM] AS BSVDRM ON BSVDRM.TUR=CM.BASVURUDURUM
INNER JOIN [COZUMMASASI].[dbo].[T_COZUMMASASI_BASVURUTIPI] AS BSVTIP ON BSVTIP.TUR=CM.BASVURUDURUM
WHERE CM.BASVURUDURUM=4 and cm.DEPARTMANANASORUN=2

```

### Şekil 22: Çözüm Masası Verisine Ait Sorgu

Kocaeli Büyükşehir Belediyesi çözüm masası başvurularına ilişkin örnek kayıtlar Tablo 10'da görülmektedir.

**Tablo 10**  
**Kocaeli Büyükşehir Belediyesi Çözüm Masası Örnek Başvuruları**

BASVURU TARİHİ	NO	DAIRE	DEPARTMAN ANA SORUN	DEPARTMAN SORUN	BASVURU	BASVURU TIPI
09.12.2009	1	Sağlık ve Sosyal Hizmetler Dairesi Başkanlığı	TALEP	Aynı ve Nakdi Yardımlar	TAMAMLANDI KAPATILDI	MOBİL
10.12.2009	2	Sağlık ve Sosyal Hizmetler Dairesi Başkanlığı	TALEP	Sosyal Yardım (Gıda)	TAMAMLANDI KAPATILDI	MOBİL
17.12.2009	3	Sağlık ve Sosyal Hizmetler Dairesi Başkanlığı	TALEP	Aynı ve Nakdi Yardımlar	TAMAMLANDI KAPATILDI	MOBİL
30.12.2009	4	Toplu Taşıma Daire Başkanlığı	TALEP	Otobüs Hattı Talebi-Kara Ulaşım Şube Müdürlüğü	TAMAMLANDI KAPATILDI	MOBİL
30.12.2009	6	Ulaşım Dairesi Başkanlığı	TALEP	Yol Bakım ve Onarım	TAMAMLANDI KAPATILDI	MOBİL
30.12.2009	7	Toplu Taşıma Daire Başkanlığı	TALEP	Şoför Şikâyeti (Belediye Otobüsleri)	TAMAMLANDI KAPATILDI	MOBİL

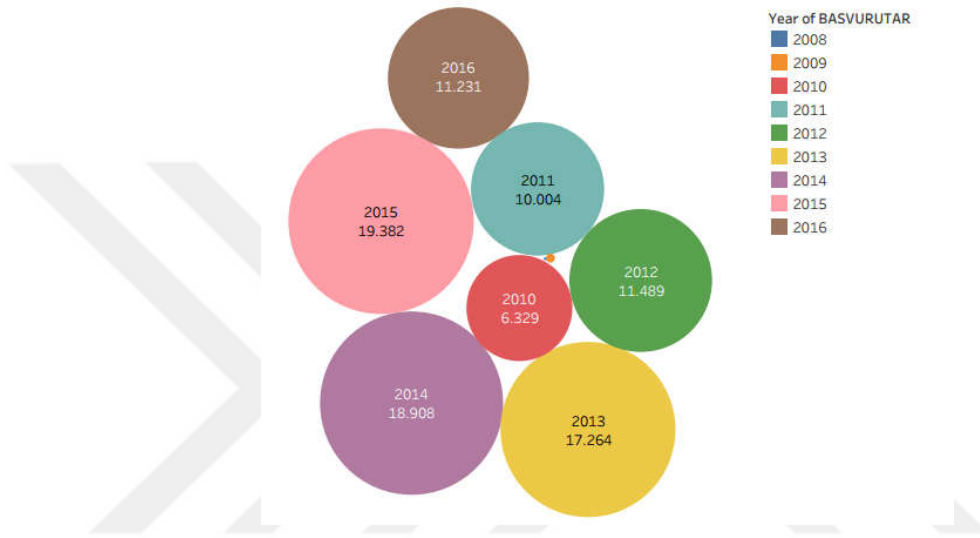
Kocaeli Büyükşehir Belediyesi çözüm masası başvuru içeriği örnek olması açısından Tablo 11’de gösterilmiştir. Tablodan görüleceği üzere vatandaşın girmiş olduğu talepler serbest metin şeklinde olduğu için imla hataları, devrik cümle ve harf hataları içermektedir.

**Tablo 11**  
**Kocaeli Büyükşehir Belediyesi Çözüm Masası Başvuru İçeriği**

BASVURUTEXT
merhaba,ben bir arkadaşım için yardım isteyecektim gölcük siretiye köyünde çadırda yaşayan 5 çocuklu bir anne çaresiz gülcan arkadaşımın sağlığıda yerinde değil eşini 2 yıl önce kaybetmiş orman kesim işlerinde çalışıyorlar çalışırsa parası var ciğerlerinden rahatsız siretiye muhtarıyla irtibata geçerseniz gerekli bilgiyi alırsınız lütfen yardım edin
benim eşim çalışıyo ssk var fakat evimiz kira gecinemiyoruz sizden gıda ve yakacak yardımı istiyorum size kocaeli büyükşehir belediyesine müracat yapmıştım fakat bir sonuc alamadım eyerki yakacak ve gıda yarmı yapmıyosanız iş verin cok maduruz allah aşkına yardım edein 1tane kızım var elerinizden öper ne olur yardım edin hayırlı günler
merhaba sayın büyükşehir çalışmanı yukarıdaki yazılı adrese dolgu malzemesi için gelmişsiniz ama sorunuzun cevabını birinci kişiden alamadığınız herhangi bir işlem yapılamadı sanırım ben sizden bu adrese yine dolgu toprağı istiyorum çok teşekkür ederim ilginize başarılar
be hereke kışladüzü mahalesin de ikamet etmek değım ulaşım poroplemimiz sıkıntılı izmiten kalkan otobüsleriniz yollarda oyalanıyorlar artı saatleri cok uzun kış günlerin de soyuk bukonuda sizlerden ricamız gebzeye saat başında giden otobüsler hereke yolcusunuda alsın lütfen bukonuyu işi düşünün sağgılarımla iyi çalış malar
belirttiğim adresin ön tarafında bulunan caddedeki elektrik direkleri 15 gündür yanmamaktadır ve akşamları zifiri karanlık olmaktadır mağdur olmaktayız konuyu ilgili birçok yere bildirmemize rağmen sonuç alamamaktayız ve konu aciliyet arz etmektedir gereğinin yapılmasını ve tarafıma konuyla alakalı bilgi verilmesini arz ederim

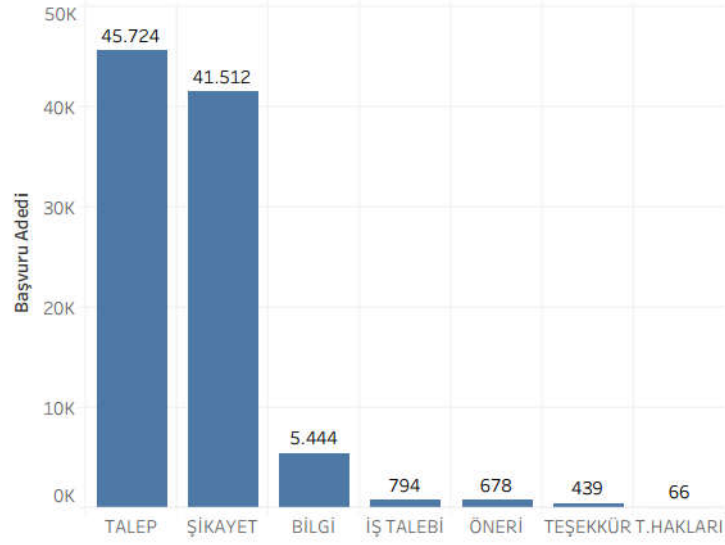
Kocaeli Büyükşehir Belediyesi Yönetim Bilişim Sistemleri Şube Müdürlüğü tarafından işletilen büyükşehir belediyesi veritabanlarından 12 Eylül 2008 – 07 Haziran 2016 tarihleri arasında Şekil 23’te detayları görüleceği üzere 100.402 adet başvuru verisi teslim alınmıştır. Çözüm masasına gerçekleştirilen başvuruların her yıl artarak devam ettiği görülmektedir. Kocaeli Büyükşehir Belediyesi “Çözüm Masası” uygulaması 12 Eylül 2008 tarihinde vatandaşın hizmetine sunulmuş olup uygulamanın devreye alındığı 2008-2009 yıllarında az sayıda başvuru alırken, 2010 yılında 6.329, 2011 yılında 10.004 şeklinde artmaya devam ederek 2015 yılında 19.382 başvuru alınmıştır. Veri setinin

alındığı 07 Haziran 2016 tarihinde başvuru sayısı 11.231'e ulaşmış durumdadır. Başvuru sayısındaki artış vatandaşların büyükşehir belediyesine ileterek çözüm bulmak istediği istek ve şikâyetleri için bilişim sistemlerini her geçen yıl daha aktif olarak yoğun şekilde kullandığını göstermektedir. Yıllara göre başvuru sayısı, kullanım durumunu karar vericilere görsel sonuçlar halinde sunarak daha anlaşılır olmasını sağlayan iş zekâsı (Business Intelligence, BI) uygulaması Tableau ile Şekil 23'te gösterilmektedir.



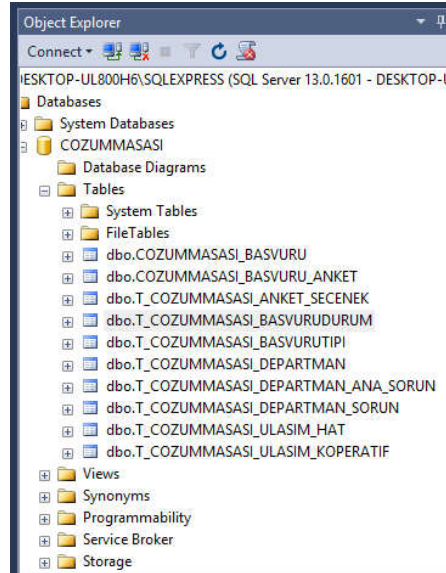
**Şekil 23:** Yıllara Göre Başvuru Adedi

Veri seti içinde başvuru durumu TAMAMLANDI-KAPATILDI olarak 94.657 başvuru bulunmaktadır. Başvuru tiplerine ait detaylar, iş zekâsı (Business Intelligence, BI) uygulaması Tableau uygulaması kullanılarak Şekil 24'te görsel ve kolay anlaşılır hale getirilmiştir. Başvuruların 45.724 adedi talep, 41.512 adedi şikâyet, 5.444 adedi başvuru, 794 adedi iş talebi, 678 adedi öneri, 439 adedi teşekkür ve 66 adedi tüketici hakları oluşturmaktadır. Vatandaşlar Kocaeli Büyükşehir Belediyesi Çözüm Masasına talep ve şikâyet öncelikli olmak üzere sırasıyla bilgi, iş talebi, öneri, şikâyet ve tüketici hakları için başvuruda bulunmaktadır.



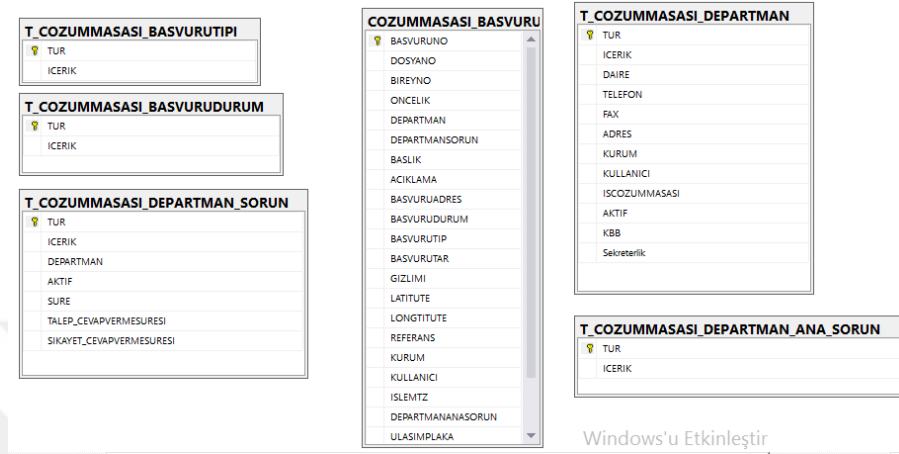
**Şekil 24:** Kocaeli Büyükşehir Belediyesi Çözüm Masası Başvuru Tiplerine Göre Başvuru Adedi

Çözüm masası uygulaması Microsoft SQL Server veritabanını kullanmaktadır. “COZUMMASASI” veritabanında, Şekil 25’te görüleceği üzere “BASVURU”, “BASVURU\_ANKET”, “ANKET\_SECENEK”, “BASVURUDURUM”, “BASVURUTIPI”, “DEPARTMAN”, “DEPARTMAN\_ANA\_SORUN”, ”ULASIM\_HAT” ve “ULASIM\_KOOPERATIF” tabloları yer almaktadır. Uygulamada kullanılan tablolara ait görüntü Şekil 23’te gösterilmektedir.



**Şekil 25:** Çözüm Masası Uygulamasında Kullanılan Tablolar

Çözüm masası uygulaması başvuru tipi, başvuru durumu, departman ana sorun, departman sorun, departman ve başvuru tabloları kullanılarak yapılandırılmıştır. Kullanılan tablolara ait Microsoft SQL Server üzerinde kullanılan diagram Şekil 26’da olduğu gibidir.



Şekil 26: Microsoft SQL Server Çözüm Masası Diagramı

Çözüm masası uygulamasında kullanılan tabloların içeriğine göz atıldığında, başvuru tablosunda vatandaşlar tarafından gerçekleştirilen başvuruların “içeriği”, departman tablosunda “daire başkanlıkları” tanımları, departman ana sorun tablosunda “şikayet”, “istek”, “başvuru” şeklinde başvuru tipleri, departman sorun tablosunda “çevre düzenlemesi”, “asfalt ve yolların yenilenmesi” gibi sorun tipleri başvuru durum tablosunda “başvuru alındı”, “birim tarafından tamamlandı”, “tamamlandı kapatıldı” şeklinde başvurunun durumu, başvuru tipi tablosunda “çözüm masası”, “çağrı masası”, “web”, “mobil” şeklinde başvurunun tipi bulunmaktadır. İlgili tabloların içeriğine ait veritabanından örnek olarak seçilen kayıtlara ait veri Şekil 27’de olduğu gibidir.

The screenshot shows the results of a query in Microsoft SQL Server Management Studio. The query is as follows:

```

select top 2 * from [COZUMMASASI].[dbo].[COZUMMASASI_BASVURU]
select top 2 * from [COZUMMASASI].[dbo].[T_COZUMMASASI_DEPARTMAN]
select top 2 * from [COZUMMASASI].[dbo].[T_COZUMMASASI_DEPARTMAN_ANA_SORUN]
select top 2 * from [COZUMMASASI].[dbo].[T_COZUMMASASI_DEPARTMAN_SORUN]
select top 2 * from [COZUMMASASI].[dbo].[T_COZUMMASASI_BASVURUDURUM]
select top 2 * from [COZUMMASASI].[dbo].[T_COZUMMASASI_BASVURUTUPI]

```

The results are displayed in a grid format with the following columns:

BASVURUNO	DOSYANO	BIREYNO	ONCELIK	DEPARTMAN	DEPARTMANSORUN	BASLIK	ACIKLAMA	BASVURUADRES
1	1	30516	0	68	604	NULL	merhaba.ben bir arkadaşım ıgı yardım isteyecekti...	
2	2	30733	0	68	605	NULL	benim eđim çağıyo esk var fakat evimiz kira geçi...	

The grid also shows data for the other tables mentioned in the query, including department information and complaint details.

Şekil 27: Çözüm Masası Tablolarına Ait Örnek Veri

Çözüm masası uygulamasında tutulan verinin anlaşılmasından sonraki aşama ilgili tablolarda kullanılan verinin, veri madenciliği algoritmalarını uygulamak üzere hazırlanması sürecidir.

### 4.3. Verinin Hazırlanması

Metin işleme süreçlerinin dezavantajlı noktalarından birisi olan gizlilik sorunları paralelinde veritabanından çekilen verinin içerisinde vatandaşa ait şahsi verinin gizlenmesi amacıyla filtreleme işlemi uygulanmıştır. Sonuç olarak Kocaeli Büyükşehir Belediyesi'nden alınan veri içinden, vatandaşın TC kimlik numarası ve adresi şeklindeki özel veri yerine, sadece çözüm bulunmasına yönelik iletmış oldukları veri toplanarak bir veri kütüphanesi oluşturulmuştur. Sonrasında elde edilen veri kümesine metin işleme teknikleri ve sınıflandırma yöntemleri uygulanmıştır.

Bu çalışmada doğal dil metinlerine uygulanabilecek ön işleme yöntemi için Türkçe ve diğer Türkî diller için yazılmış, biçim birimsel çözümleme, yazım denetimi, sözcük üretme gibi temel doğal dil işleme adımlarını gerçekleştirebilen açık kaynak kodlu doğal dil işleme kütüphanesi olarak “Zemberek” kullanılmıştır. Açık kaynak kodlu olması, kabul edilebilir doğruluk seviyesinde çalışıyor olması, Türkçe ve diğer Türki diller için yazılmış olmasından dolayı çalışmada gövdeleyici olarak Zemberek tercih edilmiştir.

Metin içerisindeki kelimelerin ayrıştırılması işlemi oldukça zahmetli bir iştir. Dilin sahip olduğu yapısal zorluklar probleme ilave yük getirmektedir. Zemberek kütüphanesi hâlihazırda Türkçe için kullanılan en kapsamlı ve uygun uygulama olarak gözükse de özellikle doğal (biçimsel olmayan) metin öbeklerinden kelime çıkartımı konusunda çoğu zaman hata yapabilmektedir. Bu nedenle, Zemberek uygulaması üzerinden oluşturulan kelime sözlüğünün, hata analizleri yöntemiyle manuel olarak kontrol edilmesi ve göreve uygun ilave düzeltici tedbirlerin ayrıca alınması gerektiği anlaşılmıştır.

Yapılacak olan uygulamada metinleri ifade etmek amacıyla özellik çıkarımında yer alan biçimbirimsel çözümleyici işlemleri aşamasında kelime kökleri (Yıldız ve diğerleri, 2007: 2, Amasyalı ve Yıldırım, 2004: 225) tercih edilmiştir. Özellik çıkarımı işlemi, ön işleme aşamasına tabi tutulan “Çözüm Masası” veri kütühanesine uygulanmaktadır.

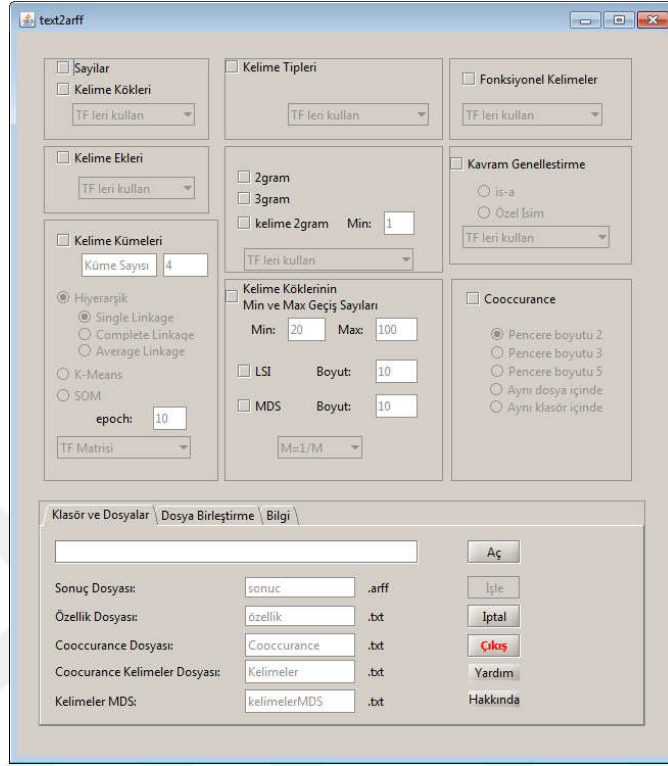
“Çözüm Masası” verisi vatandaşın Kocaeli Büyükşehir Belediyesi’ne yaptığı taleplerden oluşmakta; telefon, internet, elden başvuru alınabilmekte ve vatandaşın belediyeye direkt olarak ulaşma imkânı sağlaması nedeniyle “Çözüm Masası” verisi hızla artmakta ve veri yığını oluşmaktadır. Büyük veri yığını içinde saklı kalan bilginin ortaya çıkarılması, yöneticilere karar vermede yardımcı olabilecek geleceğe dönük tahminlerin gerçekleştirilebilmesi güçleşmektedir. Bahse konu veri yığını içersinden istenilen bilgiye erişebilmek amacıyla ham, işlenmemiş verilerden ziyade ön işleme aşamasından geçirilmiş ve özel yöntemler uygulanmak suretiyle sayısallaştırılmış veriye ihtiyaç duyulmaktadır.

Yapısal forma dönüştürme aşamasında metinlerin sayısallaştırma işlemleri için literatürde yaygın olarak kullanılan vektör uzayı modeli (Salton ve diğerleri, 1975: 613) tercih edilmiştir. Vektör uzay modelinin çalışma prensibine göre her metin anlamını ifade eden anahtar kelimelerden meydana gelen  $M \times N$  boyutlu bir vektörle ifade edilmekte,  $M$  tüm metindeki her farklı kelimeyi,  $N$  ise metinlerin toplam sayısını belirtmektedir. Vektörün her elemanının değeri, metinde kelimenin ne sıklıkta geçtiğini göstermektedir.

Bu tez çalışmasında vektör uzayı modellenmesinde TF (Term Frequency) ve TF-IDF indeksleme yöntemleri uygulanmıştır. TF-IDF yöntemi özellikle öznitelik vektörünün normalizasyonunu sağladığı için tercih edilen bir yöntemdir. Verinin hazırlanması amacıyla bahse konu adımları gerçekleştirmek üzere, “WEKA paket programı” ve “text2arff yazılımı” kullanılmıştır.

Amasyalı ve arkadaşları (2010) tarafından geliştirilen “Text2arfff” adlı yazılıma yönelik kullanıcılardan gelen geribildirimler ve son dönemde ortaya çıkan yeni metin temsil yöntemleri sebebiyle söz konusu yazılım güncellenmiştir. Yeni sürüm ile birlikte Türkçe metinlerin yanında İngilizce metinler de işlenebilir hale gelmiş, kullanıcılar bir arayüzün yanı sıra java kütüphanesi üzerinden kendi projelerinde kullanabilir hale gelmiş, veri yapılarında meydana gelen değişiklikler sebebiyle işlem hızı artmış ve yeni metin temsil yöntemleri eklenmiştir. “Text2arff” metinlerin özellik çıkarımını çeşitli yöntemlerle yapan, ağırlıklandırma metodları yardımıyla metinleri sayısal hale getiren, makine öğrenmesi algoritmalarını uygulamak üzere WEKA yazılımının giriş dosyası

olan ARFF formatına çeviren bir yazılımdır. Yazılıma ait grafik arayüzü Şekil 28’de görüldüğü gibidir.



Şekil 28: Text2arff Yazılımı Grafik Arayüzü

Yazılım özellik çıkarımı için, metinde kullanılan kelimelerin köklerinin bulunması, metinlerdeki kelimelerin türlerinin bulunması, metinlerin ngramlar (2gram, 3gram) türünden ifade edilmesi, fonksiyonel kelimelerin bulunması, kelimelerin eklerinin çıkarılması, birbirine yakın anlamlı kelimelerin kümelenecek tek bir özellik haline getirilmesi, birlikte geçme (cooccurrence) matrisinin oluşturulması ve metinlerin boyut sayısının azaltılması gibi çeşitli seçeneklere sahiptir. Metinler üzerinde analiz yapılabilmesi amacıyla metinlerin vektör uzay modeline dönüştürülmesi işlemi frekansa göre ağırlıklandırma (TF), terim frekansı x ters doküman frekansı (TFXIDF), normalizasyon (TFC), logaritmik terim bileşeni (LTC) gibi ağırlıklandırma yöntemleri ile gerçekleştirilir.

Bu çalışmada Can ve Amasyalı (2016) tarafından güncellenen Text2arff yazılımı kullanılarak metinler sayısallaştırılmıştır. Metinlerin sayısallaştırma işlemi için, çözüm masasına gelen başvuru tiplerine göre özellik vektörü elde edilmiştir. Özellik çıkarımı tekniği olarak gövdeleme tekniği seçilmiş, ağırlıklandırma yöntemi olarak



EM, PageRank, AdaBoost, kNN, Naive Bayes, ve CART algoritmalarının kullanıldığı belirtilmiştir (Wu ve diğerleri, 2008: 1). Bu kapsamda çalışmada, makine öğrenmesi yöntemlerinden açık kaynak kodlu WEKA yazılımında kullanılan Naive Bayes, Destek Vektör Makinesi, K-En Yakın Komşuluk Algoritması ve Karar Ağaçları sınıflandırma algoritmaları kullanılmıştır.

Çözüm masası verisi kullanılarak elde edilen özellik vektörlerinden gövdeleme yöntemi ile elde edilen özellik vektörü ile sınıflandırma işlemi gerçekleştirilmiştir. Makine öğrenmesi yöntemlerinden olan Naive Bayes, Destek Vektör Makinesi, Karar Ağacı, K-En Yakın Komşuluk Algoritması ve Yapay Sinir Ağları algoritmalarından yararlanılması suretiyle geliştirilen model kullanılarak yapılan başvurunun istek/öneri/şikâyet vb. tahmini gerçekleştirilmiştir.

Metin madenciliği, çoğunlukla, önceden tanımlanmış sınıf etiketlerinin de kullanılması vasıtasıyla, bir belgeye ya da dokümana bir veya daha fazla sınıf etiketinin atanmasını amaçlamaktadır. Denetimli makine öğrenmesi yaklaşımı, sınıf etiketleri önceden belirlenmiş bir korpus kümesine dayanmaktadır. Bu korpus, model öğrenmesi aşamasında parametre öğrenmek ve sistemin getirdiği hiperparametreleri optimize etmek üzere eğitim ve validasyon (geliştirme seti adıyla da anılır); öğrenilen modelin performansını değerlendirmek için ise öğrenme aşamasında kullanılmayan test setlerine bölünmektedir. Sınıflandırma işlemi eğitim veri setlerinin farklı makine öğrenme algoritmalarına nasıl tepki verdiğini, nasıl bir karakteristik ortaya koyduğunu belirlemek üzere gerçekleştirilir (Goh ve Ubeyena, 2017: 123).

#### **4.4.1. WEKA Yazılımı**

Veri madenciliği adımlarının uygulanabileceği, verinin ön işleminin gerçekleştirilebileceği, sınıflandırma, kümeleme ve birliktelik analizi için kullanılacak makine öğrenmesi algoritmalarının çalıştırılabileceği Waikato Environment for Knowledge Analysis (WEKA) yazılımı java uygulama dili ile geliştirilmiş açık kaynak kodlu (GNU lisanslı) bir yazılımdır.

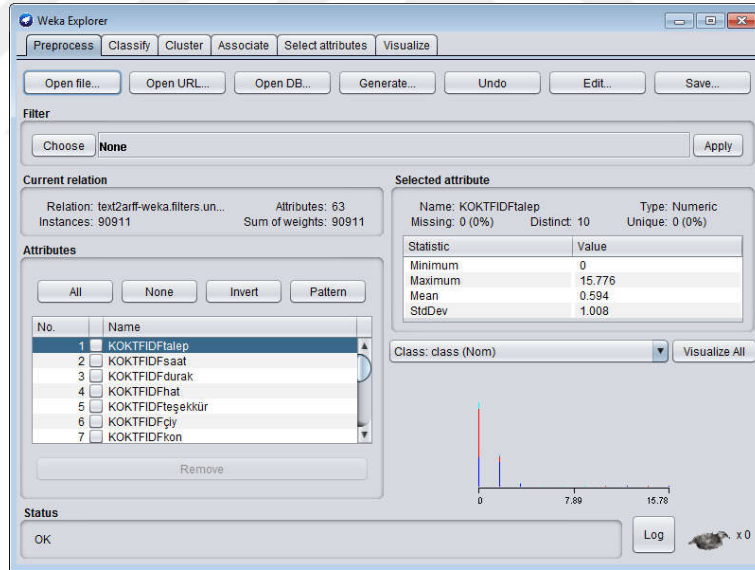
WEKA yazılımında verinin işlenebilmesi için uygulamaya özel “\*.arff” uzantılı dosyalar kullanılabilmesi gibi \*.cvs formatındaki dosyalar da WEKA’da uygulamaya tanıtılarak veriler işlenebilir. Aynı zamanda veritabanı bağlantısı kurulmak suretiyle

direk veritabanından veriler çekilerek de işlem yapılması mümkündür. WEKA yazılımına ait arayüz Şekil 30’da bulunmaktadır.



Şekil 30: WEKA Paket Programı

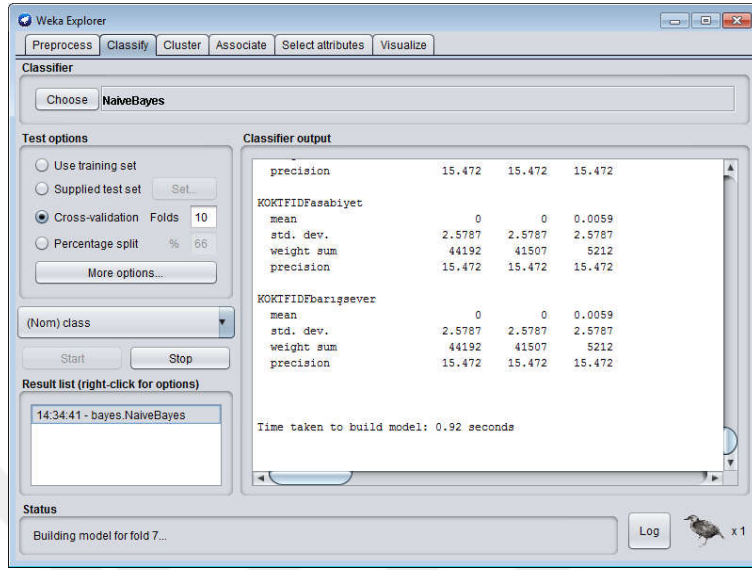
Modeli geliştirmeye devam etmek amacıyla WEKA GUI Seçim sayfasından “Explorer” butonuna basılarak Şekil 31’de yer alan WEKA gezgin ekranı açılır.



Şekil 31: WEKA Gezgin ekranı

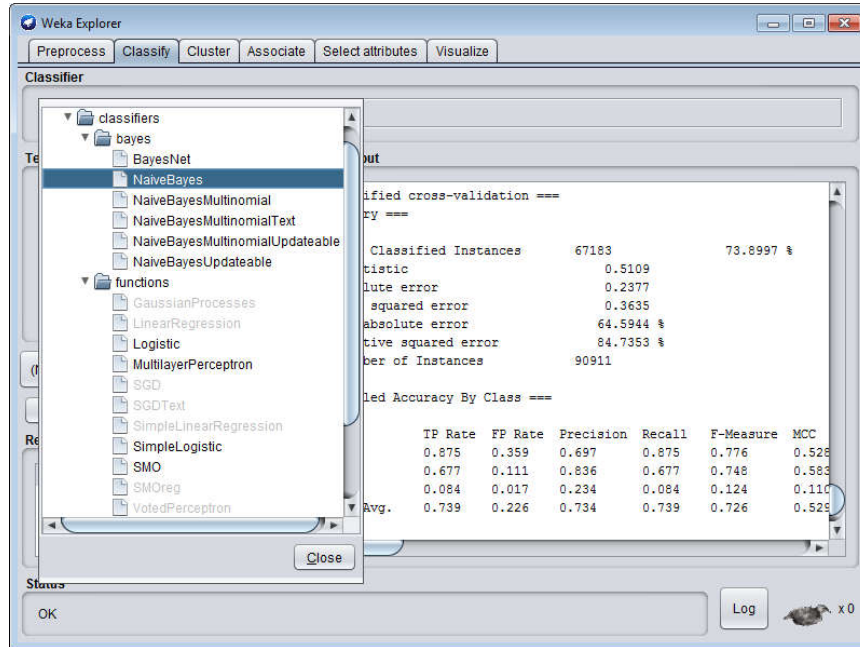
Weka Gezgin ekranında, verinin hazırlanması aşaması neticesinde elde edilen “arff” uzantılı veri dosyası “Open File” butonuna basılarak yüklenir. Bu işlemin sonunda, WEKA Gezgin ekranında yer alan nitelikler (Attributes) bölümüne “arff” veri dosyasından sorgulanan veri matrisinde yer alan veri gelmektedir. Veri madenciliği çalışmasında kullanılmak istenmeyen veri “Remove” butonu ile kaldırabilir.

Veri madenciliği sürecinde istenilen modeli seçmek için WEKA gezgin ekranından “Classify” butonuna tıklanarak Şekil 32’deki sınıflandırma modeli bölümü açılır.



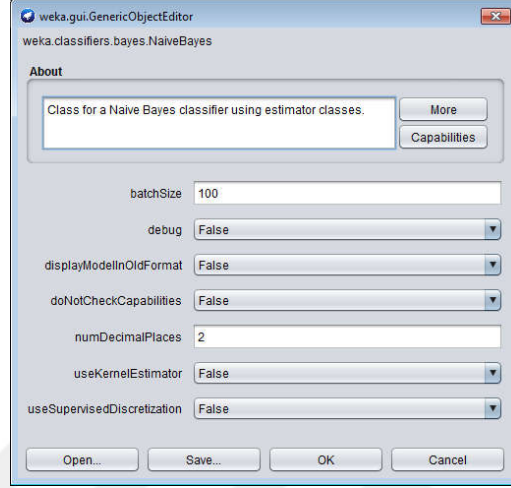
Şekil 32: Weka Gezgin Ekranı-Sınıflandırma Modülü Bölümü

Sınıflandırma modülü bölümünde “Choose” butonuna tıkladığında açılan Şekil 33’teki sınıflandırma algoritması seçim ekranından sınıflandırma işlemi için kullanılmak istenilen algoritma seçilir.



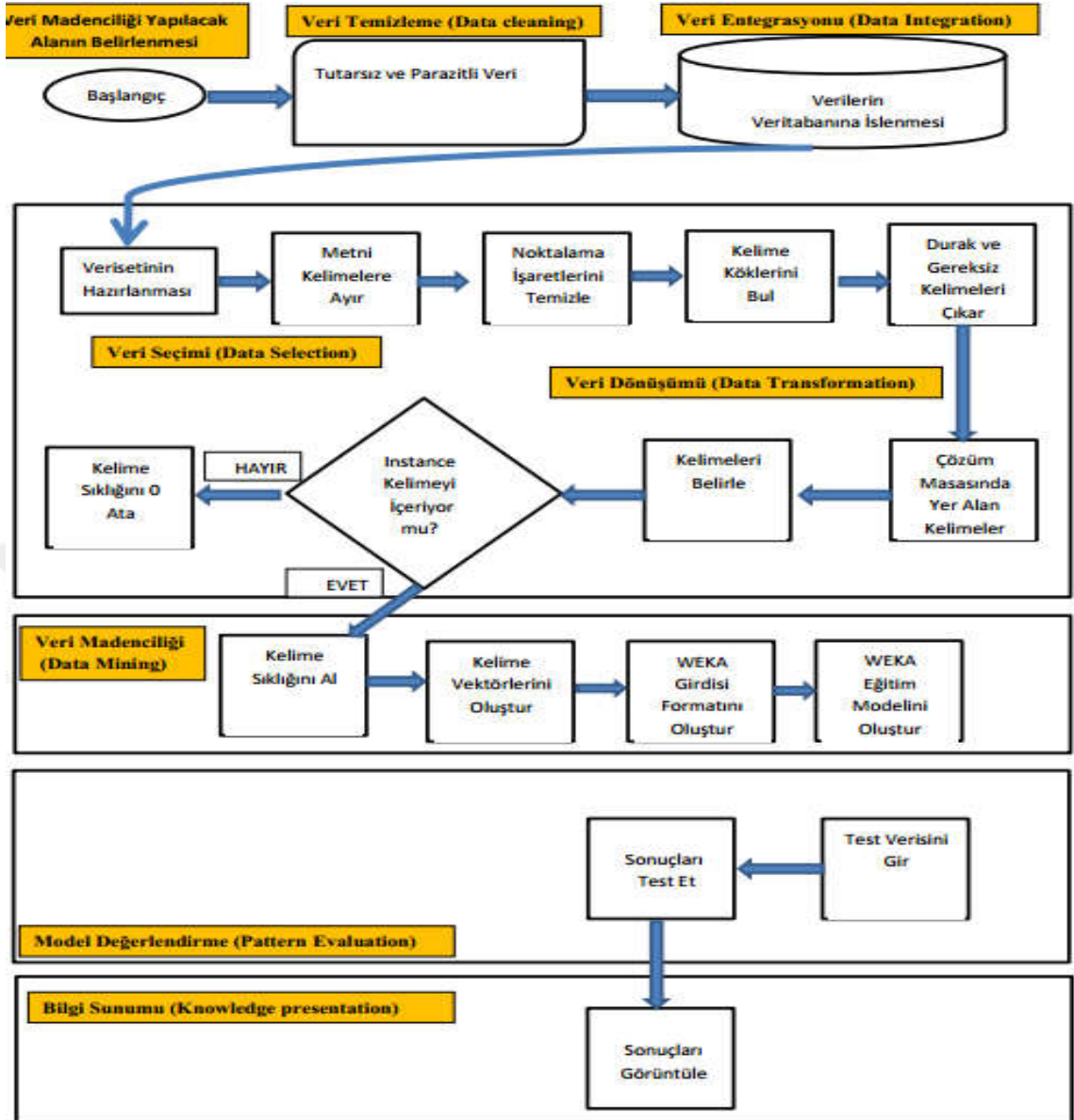
Şekil 33: Sınıflandırma Algoritması Seçim Ekranı

Algoritma seçimi yapılmasının ardından algoritma varsayılan parametreleri ile aktif olmaktadır. İstenildiğinde algoritma üzerine çift tıklanarak ilgili algoritma parametrelerinde değişiklik yapılabilir. Şekil 34’te Naive Bayes sınıflandırma algoritması parametrelerinin değiştirilebileceği ekran görünmektedir.



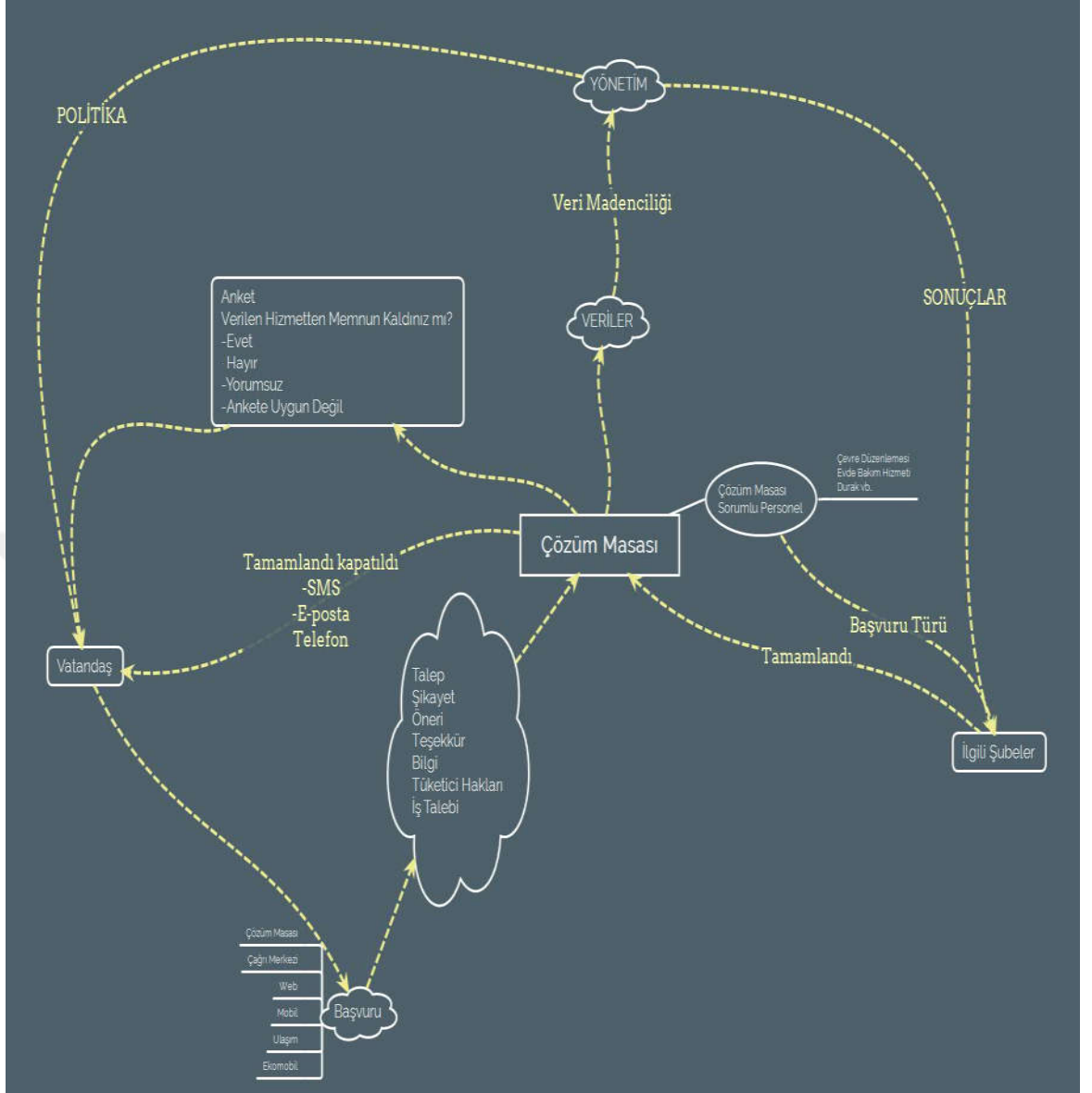
**Şekil 34:** Algoritma Parametreleri Değişim Ekranı

WEKA'nın veri sınıflandırma seçeneğini kullanarak, çözüm masasına gelen başvurular, istek, şikâyet ve başvuru olmak üzere “Naive Bayes, K-En Yakın Komşuluk Algoritması, Destek Vektör Makinesi, Karar Ağaçları ve Yapay Sinir Ağları” algoritmaları yardımıyla sınıflandırılmıştır. Veri Madenciliği süreci, Şekil 35’te görüldüğü şekilde çözüm masası verisine uygulanmıştır.



**Şekil 35:** Çözüm Masası Veri Madenciliği Süreci

Çözüm masası süreci Şekil 36’da görüleceği şekilde oluşturulmuştur. Bahse konu süreç içinde vatandaşın yapmış olduğu talep çeşidinin tahmin edilmesine yönelik olarak veri madenciliği algoritmaları çalıştırılmıştır.



**Şekil 36: Çözüm Masası Süreci Modeli**

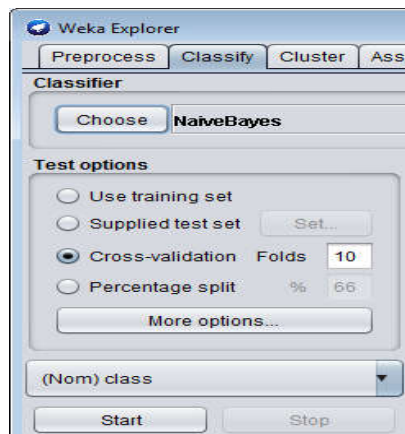
#### 4.5. Elde Edilen Bulgular

Çözüm masası verisi üzerinde vatandaşın yapmış olduğu talep çeşidini tahmin etmek amacıyla WEKA yazılımında yer alan 5 adet sınıflandırma algoritması kullanılmıştır. Araştırma kapsamında Bayes algoritması olarak NaiveBayes, Destek vektör makinesi algoritması olarak SMO, karar ağacı algoritmalarından J48, K-en yakın komşuluk algoritması K-NN olarak IBk ve yapay sinir ağlarına ait MultiLayer Perceptron algoritmaları kullanılmıştır. Algoritmalar varsayılan parametreleri seçilerek çalıştırılmıştır. Sınıflandırma algoritmaları çözüm masası verisiyle oluşturulmuş olan WEKA yazılımı formatlı “arff” dosyası üzerinde çalıştırılmıştır.

Sınıflandırma işlemleri için eğitim (training) ve test (test) olmak üzere iki farklı veri setinden yararlanılabilmektedir. Bu sayede eldeki veri setinin bir kısmı makine öğrenmesi için kullanılırken, diğer kısmı da testi gerçekleştirmek üzere kullanılmaktadır. WEKA yazılımında sınıflandırma algoritmalarını kullanmak üzere Şekil 37’de Sınıflandırma Algoritması Seçim Ekranında “Test Options” bölümünde de görüleceği üzere 4 yöntem mevcuttur.

- Use training set: Eğitim veri setini test işlemi içinde kullanma
- Supplied test set: Test işlemi için farklı veri seti kullanma
- Percentage split: Eğitim veri setinin belirlenen yüzdelik kısmını test işlemi için kullanma
- K-fold cross validation: Eğitim veri setinin k kadar eğitilmesi ve test edilmesi

Eğitim veri setinde yüksek doğruluğa sahip olma eğiliminde olan çok spesifik kurallar veri üzerinde iyi performans göstermeyerek güvenilirliğini kaybetmektedir. Bu durum aşırı uyum (overfitting) olarak adlandırılmaktadır. Algoritmanın veri seti üzerinde testi algoritmanın eğitildiği veri seti üzerinde yapılması durumunda doğruluk oranı yüksek çıkacaktır. Eğitim veri seti üzerinde çalıştırılan algoritma neticesinde oluşan model eğitim verisini birebir içerdiğinden dolayı yüksek performans gösterecektir. Algoritmanın testi farklı bir eğitim veri seti üzerinde yapılmaması nedeniyle aşırı uyum durumu tespit edilemeyecektir. Bu sebeple test yöntemi olarak “Cross-validation” seçeneği tercih edilmemiştir.



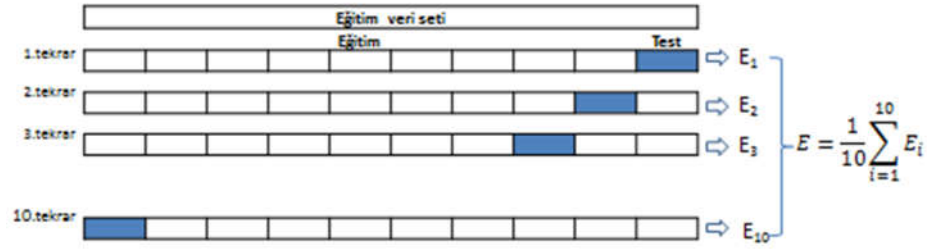
Şekil 37: Sınıflandırma Algoritması Seçim Ekranı

Bahse konu çalışma kapsamında çözüm masası veri seti üzerinden eğitim ve test işlemlerinde kullanmak üzere ayrı ayrı veri setlerinden oluşturulan “arff” dosyaları oluşturulmuş ve “Supplied test set” işlemi denenmiştir. Fakat algoritmaların çalıştırılması neticesinde elde edilen sonuç yüzdeleri kullanıcı tercihinin göre değişiklik gösterebildiği için “Supplied test set” seçeneği tercih edilmemiştir.

WEKA yazılımı üzerinde test seçeneği olarak “Percentage split” şeklinde veri setinin istenilen yüzdelik bölümünün eğitim veri seti istenilen yüzdelik bölümünün test veri seti seçilebildiği seçenek bulunmaktadır. Varsayılan değer olarak eğitim veri seti %66, test veri seti yüzde %34 olarak seçilmiş durumdadır. Çözüm masası veri setini istenilen yüzdelik bölümlere ayırabilecek büyüklükte veri seti mevcut olmakla birlikte veri setini etkin şekilde kullanmak ve sonuçlarda aşırı uyumla karşı karşıya kalmamak için bu seçenek tercih edilmemiştir.

Bu çalışmada, veri madenciliği analizlerinde sınıflandırma modellerini değerlendirmek ve modeli eğitmek üzere veri setini parçalara ayırmak için k katlamalı çarpaz doğrulama (k-fold cross validation) yöntemi kullanılmaktadır. Veriseti eğitim ve test olmak üzere ayrılırken verinin dağılımına bağlı olarak modelin eğitim ve testinde bazı sapmalar (bias) ve hatalar meydana gelebilir. K-fold cross validation, veriyi belirlenen k sayısına göre eşit parçalara bölerek ayrılan her bir parçanın hem eğitim hem de test için kullanılmasını sağlayarak meydana gelebilecek sapma ve hataları en aza indirmeyi hedefler.

Literatürde en çok tercih edilen k değeri 10’dur. K katlamalı çarpaz doğrulama yönetimiyle k-kere yöntem çalıştırılır. Şekil 38’de görüleceği üzere her aşamada veri kümesi k parçaya bölünür. Veri setinin k-1 parçası eğitim için kullanılırken, kalan 1 parça test için kullanılır ve k defa bu işlem tekrar eder. Her tekrarda elde edilen değerler toplanır, ortalaması alınır ve modelin performansı değerlendirilir.



**Şekil 38:** K Katlamalı Çarpaz Doğrulama Yöntemi

Modelin belirlenen k kadar eğitilmesi ve test edilmesi ilave olarak veri işleme yükü ve zamanı gerektirmektedir. Küçük hacimli veriler için modelin eğitilmesi ve test edilmesi kısa sürerken büyük hacimli veri setlerinde modelin hesaplanması gereken zaman açısından maliyetli olacaktır. Bu çalışmada, sınıflandırma başarısının tesadüfi olmadığını göstermek için tüm analizlerde 10 kat çarpaz geçerlilik (10-fold cross validation) veri eğitim seçeneği kullanılmıştır.

Sınıflandırma algoritmalarının performansları genellikle sınıflandırma doğrulukları (Correctly Classified Instances) yüzdesi ile ölçülmekte ve algoritmaların başarısı bu yüzdeler üzerinden birbirleri ile karşılaştırılmaktadır. Çözüm masası veri seti üzerinde çalıştırılan sınıflandırma algoritmalarının doğruluğu ilgili algoritmanın çalıştırılması neticesinde doğru olarak sınıflandırılan örneklerin toplam örnek sayısına oranıdır. Tezde çalıştırılan WEKA yazılımı algoritmalarının yorumlanma biçimini açıklamak için Şekil 39'da örnek WEKA sınıflandırma çıktısı verilmiştir.

```

=== Summary ===

Correctly Classified Instances      73370           80.7053 %
Kappa statistic                    0.6385
Mean absolute error                 0.2751
Root mean squared error            0.3559
Relative absolute error             74.7675 %
Root relative squared error        82.9631 %
Total Number of Instances          90911

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.882	0.222	0.790	0.882	0.833	0.662	0.833	0.758	1_Talep
	0.813	0.128	0.842	0.813	0.827	0.687	0.847	0.775	2_sikayet
	0.120	0.009	0.438	0.120	0.189	0.207	0.660	0.143	5_Bilgi
Weighted Avg.	0.807	0.167	0.793	0.807	0.794	0.647	0.830	0.731	

```

=== Confusion Matrix ===
  a   b   c  <-- classified as
38991 4771 430 | a = 1_Talep
 7380 33751 376 | b = 2_sikayet
 3010 1574 628 | c = 5_Bilgi

```

**Şekil 39:** Çözüm Masası Örnek Başvurusuna Ait Weka Sınıflandırma Çıktısı

Bu örnekte 90.911 (Total Number of Instances) adet çözüm masası verisi, veri seti olarak sınıflandırma modeli algoritmasında kullanılmıştır. “1\_Talep” ve “2\_sikayet” ve “3\_Bilgi” olmak üzere toplam 3 adet sınıf değeri bulunmaktadır. Test verisi üzerinde algoritmanın çalıştırılması neticesinde hata matrisi (confusion matrix) üzerinde gerçekte belirli sınıflara ait verinin hangi sınıfa atandığı atandıkları ile ilgili ayrıntılı bilgileri vermektedir. “1\_Talep” sınıfına ait veri seti vatandaş tarafından oluşturulan talep başvurusunu göstermekte olup veriden 38.991 tanesi doğru olarak “1\_Talep” sınıfına atanmıştır (“1\_Talep” için doğru pozitif). “1\_Talep” sınıfına ait veriden 4.771 adedi “2\_sikayet”, 430 adedi “5\_Bilgi” sınıfına yanlışlıkla atanmıştır (1\_Talep için yanlış pozitif). “2\_sikayet” sınıfına ait veri seti vatandaş tarafından oluşturulan şikayet başvurusunu göstermekte olup veriden 33.751 tanesi doğru olarak “2\_sikayet” sınıfına atanmıştır (“2\_sikayet” için doğru pozitif). “2\_sikayet” sınıfına ait veriden 7.380 adedi “1\_talep”, 376 adedi “5\_Bilgi” sınıfına yanlışlıkla atanmıştır (2\_sikayet için yanlış pozitif). “5\_bilgi” sınıfına ait veri seti vatandaş tarafından oluşturulan bilgi başvurusunu göstermekte olup veriden 628 tanesi doğru olarak “5\_bilgi” sınıfına atanmıştır (“5\_bilgi” için doğru pozitif). “5\_bilgi” sınıfına ait veriden 3.010 adedi “1\_talep”, 1574 adedi “2\_sikayet” sınıfına yanlışlıkla atanmıştır (5\_bilgi için yanlış pozitif).

Çözüm masası veri setinden vatandaşın yaptığı başvuruların doğru pozitif oranı (true positive rate, TP Rate) ve yanlış pozitif oranı (false positive rate, FP Rate) hesaplanmıştır. Talep sınıfı için gerçekleşen formülasyonu şu şekilde açıklamak mümkündür. Çözüm masası veri setinde bulunan talep başvuruları için doğru pozitif oranı, gerçekten talep başvuru sınıfına ait olup algoritma tarafından talep sınıfına atanan verinin talep sınıfındaki toplan kayıt sayısına oranıdır. Doğru pozitif oranı, geri çağırma (recall) değerine eşittir. Hata matrisinde (confusion matrix) talep sınıfının köşegen değerinin (matristeki talep ve talebin kesişim değeri) talpe sınıfının satır değerlerinin toplamına eşittir. “1\_Talep” sınıfı için doğru pozitif oranı  $38991/(38991+4771+430)=0.882$ 'dir.

Çözüm masası veri setinden vatandaşın yaptığı talep sınıfı için başvuruların yanlış pozitif oranı, diğer başvuru tipi sınıfına ait olup algoritma tarafından talep sınıfına atanan kayıtların talep sınıfına ait olmayan kayıtlara oranıdır. Hata matrisinde talep sınıfının kolon değerleri toplamından köşegen değerinin çıkarılması ile elde edilen

sayının diğer sınıfların satır toplamlarının toplamına bölünmesiyle bulunmuştur. “1\_Talep” sınıfı için yanlış pozitif oranı  $(7380+3010) / ((7380+3010)+(33751+376+1574+628))=0.222$ 'dir.

Çözüm masası veri seti üzerinden elde edilen kesinlik (precision) değeri, belirli bir sınıfa ait olup, algoritma tarafından belirlenen sınıfa atanan veri seti sayısının belirlenmiş sınıfa atanan tüm veri sayısına oranıdır. Hata matrisinde köşegen değerinin belirlenen sınıfın kolon değerleribib toplamına bölünmesi ile bulunur. “1\_Talep” sınıfı için kesinlik değeri:  $38991 / (38991+7380+3010)=0.790$ 'dır. Kesinlik değerinin büyük değerler alması algoritmanın doğru sınıflandırma işlemini yaptığını göstermesi açısından önemlidir.

Sınıflandırma algoritmalarının veriyi ne kadar başarılı sınıflandırdığını belirlemek için model başarı ölçütleri bulunmaktadır. Verinin iki alt örnekleme ayrılması (Hold-Out Yöntemi), Genel Doğruluk (Overall Accuracy), Dengeli Doğruluk (Balanced Accuracy), Duyarlılık (Sensitivity), Seçicilik (Specificity), Matthews Korelasyon Katsayısı (Matthews Correlation Coefficient) ve F-Ölçütü (F-Measure) bu ölçütlerden en sık kullanılanlardır (Dolgun, 2014: 28). F-Ölçütü (F-measure), kesinlik (precision) ve geri çağırma (recall) için birleştirilmiş bir ölçüttür. F-Ölçütü aşağıdaki formül ile hesaplanmaktadır. F-Ölçütü değerinin büyük değerler alması algoritmanın doğru sınıflandırma işlemini yaptığını göstermesi açısından önemlidir.

$$F\text{-Measure} = 2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall})$$

$$\text{Talep sınıfı için F-Ölçütü değeri: } 2 * 0.790 * 0.882 / (0.790 + 0.882) = 0.833$$

Algoritmaların çalıştırılması neticesinde elde edilen ölçümlerde eşik değerinin altında veya üstünde yer alması sonucunda pozitif veya negatif tanı konulmaktadır. İşlem Karakteristik (Receiver Operating Characteristic) Eğrisi, farklı eşik değerleri için hesaplanan, dikey eksen üzerinde doğru pozitiflik (duyarlılık) ve yatay eksen üzerinde yanlış pozitiflik (1- özgüllük) oranlarının yer aldığı bir grafikdir. ROC eğrisinin altında kalan alan (Area Under Curve), tanı testlerinin üstünlüğü için bir karşılaştırma ölçeği olarak kullanılmaktadır. AUC ne kadar büyük ise, durumun tahmin edilmesinde söz konusu test, o kadar iyi bir test olmaktadır. AUC tahmininde genellikle binormal ve

nonparametrik yöntemler olmak üzere iki yöntem kullanılmaktadır (Tomak ve Bek, 2010: 58).

Algoritmaların çalıştırılması neticesinde veri setinde doğru sınıflandırılan veri, doğru sınıflandırılan örnekler (Correctly Classified Instances) olarak belirtilmektedir. Hata matrisindeki köşegen değerlerinin tümünün toplamına eşittir. Bahse konu algoritma sonucuna göre  $38.991 (1\_Talep) + 33751 (2\_sikayet) + 628 (3\_bilgi) = 73370$  adet veri doğru sınıflandırılmıştır. Doğru sınıflandırılan veri sayısının tüm veri sayısına oranı sınıflandırma algoritmasının doğruluğunu vermektedir. Bu örnekte  $73370/90911 = \%80.7053$  olarak görülmektedir.

Algoritmaların çalıştırılmasının ardından yanlış sınıflandırılan veri (Incorrectly Classified Instances) veri seti üzerinde doğru sınıflandırılmayan veri sayısıdır. Hata matrisinde doğru olarak sınıflandırılmayan veri sayısıdır. Hata matrisindeki köşegen değerlerinin dışındaki değerlerin tümünün toplamına eşit olarak hesaplanmaktadır. Bahse konu algoritma sonucuna göre  $4.771+430 + 7.380+376+3.010+1.574 = 17.541$  adet veri yanlış sınıflandırılmıştır. Yanlış sınıflandırılan veri sayısının tüm veri sayısına oranı sınıflandırma algoritmasının yanlışlık oranını vermektedir. Bu örnekte  $17.541/90.911 = \%19.2946$  olarak görülmektedir.

Kappa istatistiği (Viera ve Garret, 2005: 362); 1 kappa değerinin mükemmel uzlaşmayı, 0 değerinin ise şans uzlaşmasını ifade ettiğini belirtmiştir. Bahse konu algoritma için kappa istatistiği 0.6385'tir.

Çözüm masası verisini sınıflandırmak için en uygun sınıflandırma algoritmasını seçmek için öncelikle algoritmaların doğruluklarını karşılaştırmak gerekmektedir.

#### **4.6. Modelin Sınırlılıkları**

Veri madenciliği uygulamaları geliştirmek üzere birden fazla model ve her model için kullanılabilir farklı algoritmalar bulunmaktadır. Tüm algoritmaları, zaman ve maliyet kısıtları nedeniyle örnek veri seti üzerinde denemek oldukça güçtür. Bu nedenle bu çalışmada model olarak sınıflandırma modeli, algoritma olarak ise literatürde en çok tercih edilen K-NN, karar ağaçları, destek vektör makineleri, yapay sinir ağları ve Naive Bayes algoritmaları denenerek en yüksek doğruluk oranına erişilmeye çalışılacaktır. Bu algoritmaların bir kısmı bir kısmı beklenildiği gibi başarılı

sonular retirken, bir kısmı veriyi yeterince iyi tanımlanmış sınıflara ayıramamıştır. Sınıflandırma modeli veriyi tasniflemeyi amaçlamaktadır. Sınıflandırma sayesinde yeni nesnelerin, rneğın zelliklerine bakılarak veri sınıfı tahmin edilebilir. Bu uygulamada da sınıflandırma modeli kullanılarak, belirlenen algoritma ile nce zm masasına gelen isteklerden oluřan veri setiyle eğitilmesi ikinci olarak model sınıfı belirli olmayan verilere uygulanarak sınıfları tahmin edilmesi amaçlanmıştır.

alıřmada kullanılan veri setinin hacmi de veri madencilięi uygulamasının farklı bir kısıtıdır. Uygulanacak algoritmalar ile alınacak sonuların uzun srelerde alınmasına neden olacak seviyede byk hacimli olması sre kısıtını oluřurmakta ve alıřmanın maliyetini ykseltmektedir. alıřmada kullanılan veri seti zerindeki analizlerin uzun zaman alması, veri madencilięi analizlerini pratik ve uygulanabilir olmaktan uzaklařtırmaktadır. Uygulamada veri seti zerinde veri azaltma teknikleri, znitelik seme iřlemleri uygulanarak elde edilecek veri setinin hacim olarak daha kk olmasına alıřılmıştır. Bu sayede elde edilen veri setinin hacminin kk olmasına raęmen, orijinal veri kmesinin tamamına yakınının zelliklerini koruması saęlanmıştır.

Veri madencilięi uygulaması gerekleřtirilmek zere kullanılacak veri seti doęal (yani hata kontrolleri yapılmamıř ve grltl veri ieren) metin verisinin sayısal znitelik vektrlerine dnřtrlmesinde szlkteki kelimelerin metinde yazılıřlarına uygun bulunmaması, eř anlamlı kelimeler ve ok anlamlı kelimeler gibi problemler bulunmaktadır. Veri madencilięi alıřmasının bir alt dalı olarak metin madencilięi, bu kısıtlarıyla dięer veri madencilięi alıřmalarından ayırmakta ve daha kapsamlı bir veri hazırlama sreci gerektirmektedir.

Kamu kurumlarında gerekleřtirilmek istenen veri madencilięi alıřmalarında karřımıza ıkan dięer kısıt ise eęitim verisinin kurumlar tarafından paylařılmasında yařanan zorluklardır. Veri madencilięi alıřmalarında eęitim verisinin az olması, modeli eęitimek zere alıřtırılacak veri madencilięi algoritmalarının doęru sonular retmesini engellemektedir. Bu kapsamda kurumların elindeki veri seti zerinde uygulamıř oldukları gizlilik prensipleri nedeniyle de veriye ulařmakta zorluklar yařanmaktadır.

Verinin, veri madencilięi modelini alıřtıracak uygulamanın anlayabileceęi girdi formatına dnřtrlmesinde karřılařılan problemler modelin sınırlılıęını oluřtıran farklı bir etkidir. Farklı kaynaklardan, farklı formatta elde edilen verinin veri

madenciliği uygulamasını gerçekleştirmek üzere derlenmesi, anlamlı hale getirilmesi ve gerekli dönüşümlerin sağlanması gibi veri hazırlıklarının gerçekleştirilmesi gerekmektedir.

Kurumların elindeki veri setinin büyüklüğü ile doğru orantılı olarak ihtiyaç duyulan karmaşık veri madenciliği modellerinin eğitilmesinde donanım yetersizliği yaşanabilmektedir. Küçük hacimli veri seti üzerinde basit veri madenciliği algoritmaları çalıştırmak suretiyle sonuçlar elde etmek mümkünken, büyük veri setine sahip kurumların verisi üzerinde karmaşık modellerin eğitilmesinde problemler oluşabilmektedir. Örneğin, bahse konu “çözüm masası” veri seti 90 bin üzerinde veri içermektedir. Çözüm masası verisi modellerinin eğitilmesi için 8 GB RAM kapasitesi olan dizüstü bilgisayarın fiziksel kapasitesi yeterli olmamıştır. Dizüstü bilgisayar yerine, modellerin eğitilmesi için 64 GB kapasiteli orta ölçekli sunucu üzerinde ihtiyaç duyulan çalışmalar gerçekleştirilmiştir.

Veri seti büyüklüğünün etkilediği diğer bir kısıt ize zamandır. Veri setinin büyüklüğü ile doğru orantılı olarak eğitim süreci günler, hatta haftalar almaktadır. Örneğin, 90 bin veri seti için NaiveBayes algoritmasının eğitimi süreci 2 gün sürerken, SMO ve J48 algoritmalarının eğitimi süreci 2-3 hafta arasında sürmüştür.

Kurumların elindeki veriler üzerinde analiz ve tahminleri geliştirebilecek, veri madenciliği ile makine öğrenmesi model ve algoritmalarını işleyebilecek, anlamlandırabilecek veri bilimcilerine ihtiyaç duyulduğu bir dönemde, kamu kurumlarında nitelikli personel istihdamının sürekliliğini sınırlandıran faktörler fazla olduğundan bu tür modellerin geliştirilmesi ve kamu hizmetlerine uygulanması zorlaşmaktadır.

Kurumların elindeki veriden nasıl bir fayda sağlayacağını tam olarak bilmemesi, kullanıcı istek ve talep tanımlamalarındaki belirsizlikler, veri sözlüğünün olmaması ve süreç tanımlamalarının eksik olması bu çalışmada uygulanan veri madenciliği sürecinin modele dönüşmesini zorlaştırmaktadır.

Özel sektörde daha yoğun bir şekilde uygulanan veri madenciliği, makine öğrenmesi ile birleştirilerek belediye çözüm masası verilerinden anlamlı sonuçlar elde edilmesi, farklı kamu hizmetleri alanlarında veri madenciliğine dayanan modellerin

geliştirilebileceğini göstermektedir. Bu çalışmada önerilen modelin iş zekâsı uygulamaları ile ilişkilendirilerek, daha kapsamlı ürünlere dönüştürülmesi potansiyeli olduğu da tespit edilmiştir.

#### 4.7. Tartışma

Çözüm masası verisi üzerinde çalıştırılan sınıflandırma algoritmalarına ait doğruluk oranları Tablo 12’de verilmiştir. Çözüm masasına vatandaşların yapmış olduğu 90.911 adet başvuruya ait 9.493 adet özellik üzerinde sınıflandırma algoritmaları çalıştırılmıştır.

**Tablo 12**  
**Sınıflandırma Algoritmaları Doğruluk Karşılaştırması**

<b>Algoritma</b>	<b>Doğruluk (TFIDF)</b>	<b>Doğruluk (TF)</b>
NaiveBayes	% 64.6247	% 64.6181
<b>SMO</b>	<b>% 80.7053</b>	<b>% 80.6976</b>
J48	% 77.6001	% 77.6012

Algoritmanın doğruluğunu ölçmek üzere kullanılan diğer yöntem ROC (Receiver Operating Characteristic) olarak tanımlanan “Alıcı Çalışma Karakteristik Grafiği”dir. Bahse konu yöntem ikili (binary) ve parametrik olmayan sınıflandırma algoritmalarının değerlendirilmesinde kullanılmaktadır. Sınıflandırma algoritmasındaki ikili çıktı alanı olarak pozitif veya negatif olarak tahmin edilmektedir. Şekil 49 ve Şekil 50’de J48 Algoritması sonucunda da görüleceği üzere algoritmanın risk matrisi yanlışlıkla doğru pozitif oranları, yanlış pozitif oranları değerlerinden oluşmaktadır. ROC grafiği (0,1) noktası tüm pozitif negatif durumların doğru olarak tahmin edildiği mükemmel bir sınıflamanın gerçekleştiğini göstermektedir. Grafikte tüm noktalar negatif (0,0), tüm noktaların pozitif (1,1), tüm durumların hatalı (1,0) şekilde tahmin edildiğini göstermektedir (Aydın, 2007: 60-61). TFIDF ağırlıklandırma yöntemine göre doğruluk oranı en yüksek olan SMO ve J48 algoritmalarının doğru pozitif oranları, yanlış pozitif oranları, kesinlikleri, F-Ölçütleri, ROC alanları ve kappa istatistikleri Tablo 13’te görülmektedir. Tablodan görüleceği gibi algoritmaların TP oranı değerleri birbirine çok yakındır. FP oranı, kesinlik, F-Ölçütü, ROC, kappa istatistiği ve hata metrikleri bakımından SMO J48 algoritmasına göre daha yüksek performans göstermiştir. Bu nedenle çözüm masası verisi üzerinden vatandaş başvurusunun tahmini için en uygun algoritma, SMO olarak belirlenmiştir.

**Tablo 13**  
**SMO ve J48 Algoritmalarının Doğru Pozitif Oranları, Yanlış Pozitif Oranları, Kesinlikleri, F-Ölçütleri, ROC Alanları ve Kappa İstatistikleri**

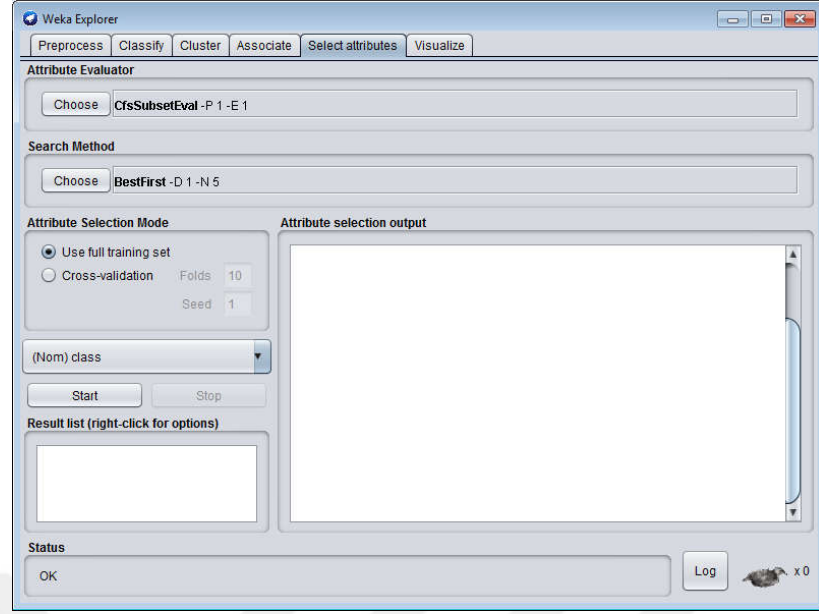
Algoritma	TP Oranı	FP Oranı	Kesinlik	F-Ölçütü	ROC Alanı	Kappa
SMO	0.807	0.167	0.793	0.794	0.830	0.6385
J48	0.776	0.180	0.765	0.770	0.813	0.587

Tablo 14'te görüleceği üzere çözüm masası verisi üzerinden vatandaşın başvurusunu tahmin etmek üzere doğruluğu en yüksek çıkan algoritmalar SMO ve J48 algoritmalarıdır. Algoritmaların doğruluk oranlarının eşit çıkması halinde Tablo 14'te yer alan SMO ve J48 algoritmaları hata metrikleri incelenerek hata metrik değerleri daha düşük olan algoritma seçilir.

**Tablo 14**  
**SMO ve J48 Algoritmaları Hata Metrikleri**

Algoritma	Ortalama Mutlak Hata	Ortalama Karesel Hata Karekökü	Bağıl Mutlak Hata	Bağıl Karesel Hata Karekökü
SMO	0.2751	0.3559	% 74.7675	% 82.9631
J48	0.1738	0.361	% 47.2422	% 84.1596

Çözüm masasına ait 90.911 adet başvuruya ait 9.493 adet özellik üzerinde sınıflandırma algoritmaları çalıştırılması halinde makine öğrenmesi işlemleri ve gerekli modelin oluşturulması süreci uzun sürmektedir. Kamu kurumlarında karar verme ve gerekli politikaların oluşturulması için veri madenciliği uygulamasına ihtiyaç duyulduğunda uzun süreler beklenilmesi elde edilen faydanın sağlanmasının önünde engel teşkil etmekte, veri madenciliğinin pratik ve uygulanabilir olmaktan çıkarmaktadır. Veri kümesi ne kadar büyük olursa veri analizi aynı ölçüde karmaşık olmaktadır. Veri madenciliği modelini oluşturmak üzere kullanılan çözüm masası eğitim setinde başvuru sayısının fazla olması ve elde edilen kök kelimelerin sayısının fazla olması daha iyi sonuçlar elde edilebileceği anlamına gelmemektedir. Bu sebeple niteliğin yanında nicel olarak verinin kalitesini artırmak üzere veri azaltma teknikleri uygulanmıştır. Veri azaltma tekniği uygulamak üzere öznitelik seçme işlemini uygulamak üzere WEKA yazılımında yer alan özellik seçimi Şekil 40'ta görüldüğü üzere Öznitelik Değerlendiricisi (Attribute Evaluator) ve Arama Yöntemi (Search Method) şeklinde iki kısma ayrılmaktadır.



**Şekil 40:** WEKA Uygulaması Öznitelik Seçimi Ekranı

Öznitelik değerlendiricisi, çözüm masası veri setindeki her öznitelik (sütun veya özellik) çıktı değişkeni bağlamında değerlendirilir (kelime kökü). Arama yöntemi, seçilen özelliklerin kısa bir listesine ulaşmak için veri setindeki farklı özellik kombinasyonlarını denemek ya da gezinmek için kullanılan tekniktir. Bazı öznitelik değerlendiricisi teknikleri, belirli arama yöntemlerinin kullanılmasını gerektirir. Veri seti için hangi öznitelik seçme yönteminin daha doğru sonuç üreteceğine dair net bir yanıt bulunmamakla birlikte verinin yapısına göre değişkenlik göstermektedir. Bu sebeple “CfsSubsetEval”, “CorrelationAttributeEval” ve “InfoGainAttributeEval” olmak üzere üç farklı öznitelik seçme yöntemi denenmiş olup arama yöntemi olarak “CfsSubsetEval” tercih edilmesine karar verilmiştir. Bu kapsamda çözüm masası veri setine uygulamak üzere öznitelik değerlendiricisi olarak “CfsSubsetEval”, arama yöntemi olarak ise “BestFirst” seçilmiştir.

Öznitelik seçme işlemi sayesinde çözüm masası veri setinde yer alan 90911 adet başvuruya ait 9.493 adet özellik içerisinde 62 adet özellik (kelime kökü) belirlenmiştir. Bulunan kelime kökü listesi ise Ek 2’de yer almaktadır. Bulunan kelime köklerine ait liste örneği Tablo 15’te verilmektedir.

**Tablo 15**  
**Öznelik Seçimi Sonucunda Bulunan Kök Kelime Örnekleri**

Sıra	Kelime
1	<b>talep</b>
2	saat
3	durak
4	hat
5	<b>teşekkür</b>
6	asfalt
7	ağaç
8	<b>şikayet</b>
9	sicil
10	müjde
11	sergi
12	dekoratif
13	tabiat
14	çörek
15	amonyak

Tablo 16’da öznelik seçimi sonucunda bulunan kök kelimeler üzerinden çözüm masası verisi üzerinde çalıştırılan sınıflandırma algoritmalarına ait doğruluk oranları verilmiştir. Çözüm masasına vatandaşların yapmış olduğu 90.911 adet başvuruya ait 62 adet özellik üzerinde sınıflandırma algoritmaları çalıştırılmıştır.

**Tablo 16**  
**Öznelik Seçimi İşlemi Sonrası Sınıflandırma Algoritmaları Doğruluk Karşılaştırması**

Algoritma	Doğruluk (TFIDF)	Doğruluk (TF)
NaiveBayes	% 73.8997	% 73.892
SMO	% 73.4928	% 73.4917
J48	% 76.6101	% 76.6101
IBk	% 76.3252	% 76.3439
MultiLayer Perceptron	% 71,6217	% 73.3113

Çözüm masası verisi üzerinden başvuruların tahmin edilmesine yönelik çalıştırılan algorimalara ait çıktılar aşağıda olduğu gibidir. Ağırlıklandırma yöntemi olarak TFIDF ve TF olmak üzere iki farklı ağırlıklandırma yöntemi seçilerek başvuru seti üzerinden özellik (kelime kökü) elde edilmiştir. Çalışmada kullanılan algoritmalara ait bilgiler ve elde edilen sonuçların özet bilgileri aşağıda olduğu gibidir.

Naive Bayes, bayes teorimine ait istatistik sınıflandırma algoritmasıdır. Bayes sınıflandırıcılar, belirli sınıfa ait doküman içindeki özellikleri eğitim verisi yardımıyla birbirinden bağımsız olarak düşünerek kelimelerin ve sınıfları tahmin etmek için kullanılmaktadır. Bayes sınıflandırıcılara performans açısından karar ağacı modeli ve yapay sinir ağları algoritmaları ile karşılaştırılabilir. Bayes sınıflandırıcılara büyük verilere uygulandığında yüksek doğruluk ve hızda çalışır (Han ve diğerleri, 2012: 350). Naive Bayes Algoritması, kategorisi belirlenmiş bir sınıf için terim olasılıklarının hesaplanmasına göre terimlerin geçiş sayıları üzerinden çok terimli (multinomial) ve terimlerin mevcut olup olmadığı üzerinden çok değişkenli (multivariate) olarak ikiye ayrılır (Kesgin, 2007: 35).

NaiveBayes algoritması çözüm masası verisi üzerinden başvuruların tahmin edilmesine yönelik sınıflandırma işleminde TFIDF ağırlıklandırma yönteminde %64.6247'lik oranla 90.911 başvurunun 58751 adedi, TF ağırlıklandırma yönteminde %64.6181'lik oranla 90911 başvurunun 58.745 adedi doğru sınıflandırmıştır. Algoritma özet sonuçları; NaiveBayes Algoritması (TFIDF Ağırlıklandırma) Sonucu Şekil 41'de ve NaiveBayes Algoritması (TF Ağırlıklandırma) Sonucu Şekil 42'de görülmektedir.

```

Time taken to build model: 274.88 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      58751           64.6247 %
Kappa statistic                    0.4123
Mean absolute error                 0.2371
Root mean squared error            0.4684
Relative absolute error            64.4407 %
Root relative squared error        109.1918 %
Total Number of Instances          90911

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.622	0.184	0.762	0.622	0.685	0.447	0.764	0.738	1_Talep
	0.696	0.202	0.743	0.696	0.719	0.497	0.791	0.784	2_sikayet
	0.460	0.159	0.150	0.460	0.226	0.184	0.753	0.130	5_Bilgi
Weighted Avg.	0.646	0.191	0.716	0.646	0.674	0.455	0.776	0.724	

```

=== Confusion Matrix ===
      a    b    c  <-- classified as
27474  8651  8067 |  a = 1_Talep
 7090 28882  5535 |  b = 2_sikayet
 1499  1318  2395 |  c = 5_Bilgi

```

**Şekil 41:** NaiveBayes Algoritması (TFIDF Ağırlıklandırma) Sonucu

```

Time taken to build model: 257.22 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      58745          64.6181 %
Kappa statistic                    0.4122
Mean absolute error                 0.2372
Root mean squared error            0.4685
Relative absolute error            64.4691 %
Root relative squared error        109.2251 %
Total Number of Instances          90911

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.621	0.184	0.762	0.621	0.684	0.447	0.763	0.738	1_Talep
	0.696	0.202	0.743	0.696	0.719	0.497	0.791	0.782	2_sikayet
	0.460	0.159	0.150	0.460	0.226	0.184	0.753	0.130	5_Bilgi
Weighted Avg.	0.646	0.191	0.718	0.646	0.674	0.455	0.776	0.723	

```

=== Confusion Matrix ===

```

a	b	c	<-- classified as
27449	8672	8071	a = 1_Talep
7083	28897	5527	b = 2_sikayet
1491	1322	2399	c = 5_Bilgi

### Şekil 42: NaiveBayes Algoritması (TF Ağırlıklandırma) Sonucu

NaiveBayes algoritması çözüm masası verisi üzerinden başvuruların tahmin edilmesine yönelik öznitelik seçimi yapıldıktan sonra algoritmaların çalıştırılması neticesinde çözüm masası verisi üzerinden başvuruların tahmin edilmesine yönelik gerçekleştirilen sınıflandırma işleminde TFIDF ağırlıklandırma yönteminde % 73.8997'lik oranla 90.911 başvurunun 67.183 adedi, TF ağırlıklandırma yönteminde % 73.892'lik oranla 90.911 başvurunun 67.176 adedi doğru sınıflandırmıştır. Algoritma özet sonuçları; Öznitelik Seçimi Sonrası NaiveBayes Algoritması (TFIDF Ağırlıklandırma) Sonucu Şekil 43'te ve Öznitelik Seçimi Sonrası NaiveBayes Algoritması (TF Ağırlıklandırma) Sonucu Şekil 44'te görülmektedir.

```

Time taken to build model: 1.33 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      67183          73.8997 %
Kappa statistic                    0.5109
Mean absolute error                 0.2377
Root mean squared error            0.3635
Relative absolute error            64.5944 %
Root relative squared error        84.7353 %
Total Number of Instances          90911

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.875	0.359	0.697	0.875	0.776	0.528	0.824	0.792	1_Talep
	0.677	0.111	0.836	0.677	0.748	0.583	0.832	0.781	2_sikayet
	0.084	0.017	0.234	0.084	0.124	0.110	0.737	0.142	5_Bilgi
Weighted Avg.	0.739	0.226	0.734	0.739	0.726	0.529	0.823	0.750	

```

=== Confusion Matrix ===

```

a	b	c	<-- classified as
38661	4622	909	a = 1_Talep
12891	28082	534	b = 2_sikayet
3898	874	440	c = 5_Bilgi

### Şekil 43: Öznitelik Seçimi Sonrası NaiveBayes Algoritması (TFIDF Ağırlıklandırma) Sonucu

```

Time taken to build model: 0.86 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      67176          73.892 %
Kappa statistic                    0.5107
Mean absolute error                 0.2378
Root mean squared error             0.3635
Relative absolute error             64.6147 %
Root relative squared error         84.7464 %
Total Number of Instances          90911

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,875	0,360	0,697	0,875	0,776	0,528	0,824	0,792	1_Talep
	0,677	0,111	0,836	0,677	0,748	0,583	0,832	0,781	2_sikayet
	0,083	0,017	0,230	0,083	0,122	0,108	0,736	0,141	5_Bilgi
Weighted Avg.	0,739	0,227	0,734	0,739	0,726	0,529	0,823	0,750	

```

=== Confusion Matrix ===

```

	a	b	c	<-- classified as
38665	4622	905		a = 1_Talep
12892	28080	535		b = 2_sikayet
3907	874	431		c = 5_Bilgi

#### Şekil 44: Öznitelik Seçimi Sonrası NaiveBayes Algoritması (TF Ağırlıklandırma) Sonucu

Destek vektörü (SV) yöntemi, fonksiyon tahmini problemlerine genel bir yaklaşımdır. Model tanıma için (gösterge fonksiyonlarını tahmin etmek için), regresyon için (gerçek değerli fonksiyonları tahmin etmek için) ve doğrusal operatör denklemlerini çözmek için kullanılmaktadır. Destek vektör makinesi algoritması 1960'ların başında desen tanıma problemlerinde hipredüzlemler oluşturmak için kullanılmaya başlanmıştır (Vapnik, 1998: 25-26). Destek Vektör Makinesi (SVM), çok büyük karesel programlama (QP) optimizasyonunun çözümünde kullanılmaktadır. SMO algoritması (Deng ve diğerleri, 2013: 198) büyük parçaları olabildiğince küçük parçalara ayırmak suretiyle analitik olarak çözmektedir. Çok büyük eğitim veri setleri büyük matris hesaplamaları gerektirirken SMO bu verisetlerini doğrusal şekilde ölçeklendirir (Platt, 1998: 185). Etkili ve basit bir sınıflandırıcı algoritması olarak bu yöntemde bir düzlemde bulunan iki grup bir sınır çizgisi ile iki bölgeye ayrılmak istenmektedir. İşlemi gerçekleştirmek amacıyla her iki grubun üyelerine yakın ve paralel çizgiler çekilmekte, sınır çizgileri birbirine yaklaştırılarak ortak sınır çizgisi üretilerek eğitim işlemi tamamlanmaktadır (Bulut, 2016: 2).

SMO algoritması kullanılarak çözüm masası verisi üzerinden başvuruların tahmin edilmesine yönelik gerçekleştirilen sınıflandırma işleminde TFIDF ağırlıklandırma yönteminde %80.7053'lük oranla 90.911 başvurunun 73.370 adedi, TF ağırlıklandırma yönteminde yönteminde %80.6976'luk oranla 90.911 başvurunun 73.363 adedi doğru sınıflandırmıştır. Algoritma özet sonuçları; SMO Algoritması (TFIDF Ağırlıklandırma)

Sonucu Şekil 45'te ve SMO Algoritması (TF Ağırlıklandırma) Sonucu Şekil 46'da görülmektedir.

```
Time taken to build model: 108372.17 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      73370          80.7053 %
Kappa statistic                    0.6385
Mean absolute error                 0.2751
Root mean squared error             0.3559
Relative absolute error             74.7675 %
Root relative squared error         82.9631 %
Total Number of Instances          90911

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0.882   0.222   0.790     0.882   0.833     0.662   0.833   0.758   1_Talep
                0.813   0.128   0.842     0.813   0.827     0.687   0.847   0.775   2_sikayet
                0.120   0.009   0.438     0.120   0.189     0.207   0.660   0.143   5_Bilgi
Weighted Avg.   0.807   0.167   0.793     0.807   0.794     0.647   0.830   0.731

=== Confusion Matrix ===
      a    b    c  <-- classified as
38991  4771  430 |  a = 1_Talep
7380  33751  376 |  b = 2_sikayet
3010  1574  628 |  c = 5_Bilgi
```

Şekil 45: SMO Algoritması (TFIDF Ağırlıklandırma) Sonucu

```
Time taken to build model: 118891.84 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      73363          80.6976 %
Kappa statistic                    0.6383
Mean absolute error                 0.2751
Root mean squared error             0.3559
Relative absolute error             74.7728 %
Root relative squared error         82.9692 %
Total Number of Instances          90911

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0.882   0.222   0.790     0.882   0.833     0.662   0.833   0.758   1_Talep
                0.813   0.128   0.842     0.813   0.827     0.687   0.847   0.775   2_sikayet
                0.120   0.009   0.437     0.120   0.188     0.207   0.660   0.143   5_Bilgi
Weighted Avg.   0.807   0.167   0.793     0.807   0.794     0.647   0.830   0.730

=== Confusion Matrix ===
      a    b    c  <-- classified as
38987  4773  432 |  a = 1_Talep
7381  33750  376 |  b = 2_sikayet
3012  1574  626 |  c = 5_Bilgi
```

Şekil 46: SMO Algoritması (TF Ağırlıklandırma) Sonucu

SMO algoritması çözüm masası verisi üzerinden başvuruların tahmin edilmesine yönelik öznitelik seçimi yapıldıktan sonra algoritmaların çalıştırılması neticesinde çözüm masası verisi üzerinden başvuruların tahmin edilmesine yönelik gerçekleştirilen sınıflandırma işleminde TFIDF ağırlıklandırma yönteminde %73.4928'lik oranla 90911 başvurunun 66.813 adedi, TF ağırlıklandırma yönteminde yöntemde % 73.4917'lik oranla 90.911 başvurunun 66.812 adedi doğru sınıflandırmıştır. Algoritma özet sonuçları; Öznitelik Seçimi Sonrası SMO Algoritması (TFIDF Ağırlıklandırma) Sonucu Şekil 47'de ve Öznitelik Seçimi Sonrası SMO Algoritması (TF Ağırlıklandırma) Sonucu Şekil 48'de görülmektedir.

```

Time taken to build model: 70.25 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      66813      73.4928 %
Kappa statistic                    0.4943
Mean absolute error                 0.2928
Root mean squared error             0.3802
Relative absolute error             79.5582 %
Root relative squared error         88.6433 %
Total Number of Instances          90911

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.912	0.415	0.675	0.912	0.776	0.524	0.753	0.661	1_Talep
	0.638	0.095	0.849	0.638	0.729	0.570	0.772	0.708	2_sikayet
	0.003	0.000	0.483	0.003	0.005	0.033	0.527	0.064	5_Bilgi
Weighted Avg.	0.735	0.245	0.744	0.735	0.710	0.516	0.749	0.648	

```

=== Confusion Matrix ===

```

	a	b	c	<-- classified as
40314	3875	3		a = 1_Talep
15010	26485	12		b = 2_sikayet
4374	824	14		c = 5_Bilgi

**Şekil 47:** Öznitelik Seçimi Sonrası SMO Algoritması (TFIDF Ağırlıklandırma) Sonucu

```

Time taken to build model: 35.59 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      66812      73.4917 %
Kappa statistic                    0.4942
Mean absolute error                 0.2927
Root mean squared error             0.3802
Relative absolute error             79.5556 %
Root relative squared error         88.6398 %
Total Number of Instances          90911

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.912	0.415	0.675	0.912	0.776	0.524	0.753	0.661	1_Talep
	0.638	0.095	0.849	0.638	0.729	0.570	0.772	0.708	2_sikayet
	0.001	0.000	0.318	0.001	0.003	0.017	0.528	0.064	5_Bilgi
Weighted Avg.	0.735	0.245	0.734	0.735	0.710	0.516	0.749	0.648	

```

=== Confusion Matrix ===

```

	a	b	c	<-- classified as
40316	3873	3		a = 1_Talep
15006	26489	12		b = 2_sikayet
4380	825	7		c = 5_Bilgi

**Şekil 48:** Öznitelik Seçimi Sonrası SMO Algoritması (TF Ağırlıklandırma) Sonucu

Karar ağacı, birbirine benzerlik gösteren verinin ağaç yapısı şeklinde dallara bölen görsel istatistik bir süreçtir. Karar ağacı öğrenme ve sınıflandırma olmak üzere iki basamaklıdır. İlk basamağında bilinen bir eğitim verisi model oluşturmak amacıyla sınıflandırma algoritması tarafından çözümlenir. Sınıflama basamağında ise test verisi sınıflama kuralı ve karar ağacı modelinin doğruluğunu belirlemek amacıyla kullanılır. Doğruluk kabul edilebilir ise kurallar yeni verinin sınıflandırılmasında kullanılır. Modelleme ve veritabanları ile entegrasyonun kolay olması, güvenilirliğinin yüksek olması sebebiyle tercih edilmektedir.

Veri madenciliğinde karar ağacı modeli, verimli ve anlaşılabilir olağanüstü veri analizi sebebiyle oldukça popülerdir. CHAID, CART, C5.0 ve QUEST algoritmalarını aşağıdaki şekilde açıklamak mümkündür (He ve diğerleri, 2013: 124-125).

1. CHAID Algoritması, deęişken seçimi ve kümeleme için hedef verinin optimizasyonunu esas alan sınıflandırma ve sıralanmış seviyedeki verinin analizi için 1980 yılında Kass tarafından yayınlanmış uygun bir analiz metodudur.
2. CART; Leo Breiman, Jerome Friedman, Riehad Olshen ve Charles Ston tarafından 1984'te veri madencilięi sınıflandırma algoritması olarak yayınlanmıştır. Hedef nitelięin deęeri istendięinde bu işlem regresyon ağacı olarak tanımlanır, deęer ayırık olduęunda ise bu işlem sınıflandırma ağacı olarak tanımlanır.  
İkili özyinelemeli segmentasyon teknolojisi, mevcut örnek setini karar ağacında olmayan iki alt örnek setine böler. Bu nedenle karar ağacı, CART algoritması tarafından üretilmiş, basit yapılandırılmış ikili bir ağaçtır.
3. QUEST Algoritması, yeni ikili ağaç algoritmasının geliştirilmesiyle 1997'de Loh ve Shih tarafından çıkarılmış bir algoritmadır. Bu algoritma, deęişkenlerin ve çapraz noktaların seçimini ayırır. CHAID'in bazı dezavantajlarını gidermekle birlikte herhangi bir deęişken seçimi için uygundur.
4. C5.0 algoritması; karar ağacı modellerindeki sınıflandırma, karar ağacı veya kural seti kurgusu ve tekrarlama yolu ile doęruluęunu artırır. 1992 yılında Quinlan tarafından önerilmiştir. Başlangıçta, eğitim setindeki her örneęe aynı başlangıç aęırlığı verilir. Ardından algoritma, karar ağacında yanlış aęırlık verilmiş olan eğitim setinin aęırlığını arttıracaktır. Bu şekilde yanlış şekilde eğitim setinde yanlış sınıflandırılmış veri seti artacaktır. Bu süreç belirlenen eęięe gelinene kadar devam edecektir. C5.0 algoritması sadece sınıflandırma için kullanılır. C.5.0 algoritması, C4.5 karar ağacı algoritmasının ticari versiyonudur.
5. J48 algoritması; sınıflandırma işlemleri gerçekleştirmek üzere kullanılan C4.5 karar ağacı algoritmasının Weka'da kullanılan versiyonudur. Karar ağacı yaklaşımı sınıflandırma problemlerinde tercih edilen en kullanışlı yöntemdir (Patil ve Sherekar, 2013: 256). J48 algoritması, enformasyona dayalı veriden ilgili özellikleri seçmek amacıyla veriyi otomatik işleyen Naive Bayes, ID3, Lojistik Regresyon gibi algoritmalarına göre sınıflandırması en yüksek karar ağacı algoritmasıdır. J48 algoritması, enformasyon kazancının en iyi olduęu noktadan örnekleri bölerek IF-THEN kurallarına dayalı bir karar ağacı ve üyelik

fonksiyon kümeleri çıktısı verir. Örnek verisetini ağaç yapısı şeklinde bölerek ağacın en iyi kök değişkeninin seçilmesi ile başlayarak yukarıdan aşağıya inşa edilmesi şeklinde çalışmaktadır.

WEKA uygulaması üzerinden J48 algoritması kullanılarak çözüm masası verisi üzerinden başvuruların tahmin edilmesine yönelik gerçekleştirilen sınıflandırma işleminde TFIDF ağırlıklandırma yönteminde %76.6001'lik oranla 90.911 başvurunun 70.547 adedi, TF ağırlıklandırma yönteminde %77.6012'lik oranla 90.911 başvurunun 70548 adedi doğru sınıflandırmıştır. Algoritma özet sonuçları; J48 Algoritması (TFIDF Ağırlıklandırma) Sonucu Şekil 49'da ve J48 Algoritması (TF Ağırlıklandırma) Sonucu Şekil 50'de görülmektedir.

```

Time taken to build model: 46296.82 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      70547          77.6001 %
Kappa statistic                    0.587
Mean absolute error                 0.1738
Root mean squared error             0.361
Relative absolute error             47.2422 %
Root relative squared error        84.1596 %
Total Number of Instances          90911

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.826	0.216	0.784	0.826	0.804	0.610	0.819	0.750	1_Talep
	0.798	0.162	0.805	0.798	0.801	0.636	0.830	0.756	2_sikayet
	0.180	0.026	0.292	0.180	0.223	0.193	0.629	0.154	5_Bilgi
Weighted Avg.	0.776	0.180	0.765	0.776	0.770	0.598	0.813	0.719	

```

=== Confusion Matrix ===

```

a	b	c	<-- classified as
36492	6366	1334	a = 1_Talep
7455	33118	934	b = 2_sikayet
2623	1652	937	c = 5_Bilgi

Şekil 49: J48 Algoritması (TFIDF Ağırlıklandırma) Sonucu

```

Time taken to build model: 39506.06 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      70548          77.6012 %
Kappa statistic                    0.587
Mean absolute error                 0.1738
Root mean squared error             0.361
Relative absolute error             47.2428 %
Root relative squared error        84.1599 %
Total Number of Instances          90911

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.826	0.216	0.784	0.826	0.804	0.610	0.819	0.751	1_Talep
	0.798	0.162	0.805	0.798	0.802	0.636	0.831	0.756	2_sikayet
	0.180	0.027	0.292	0.180	0.222	0.193	0.629	0.153	5_Bilgi
Weighted Avg.	0.776	0.180	0.765	0.776	0.770	0.598	0.813	0.719	

```

=== Confusion Matrix ===

```

a	b	c	<-- classified as
36490	6365	1337	a = 1_Talep
7449	33121	937	b = 2_sikayet
2624	1651	937	c = 5_Bilgi

Şekil 50: J48 Algoritması (TF Ağırlıklandırma) Sonucu

J48 algoritması çözüm masası verisi üzerinden başvuruların tahmin edilmesine yönelik öznitelik seçimi yapıldıktan sonra algoritmaların çalıştırılması neticesinde çözüm masası verisi üzerinden başvuruların tahmin edilmesine yönelik gerçekleştirilen sınıflandırma işleminde TFIDF ve TF ağırlıklandırma yönteminde % 76.6101'lik oranla 90.911 başvurunun 69.647 adedi doğru sınıflandırmıştır. Algoritma özet sonuçları; Öznitelik Seçimi Sonrası J48 Algoritması (TFIDF Ağırlıklandırma) Sonucu Şekil 51'de ve Öznitelik Seçimi Sonrası J48 Algoritması (TF Ağırlıklandırma) Sonucu Şekil 52'de görülmektedir.

```

Time taken to build model: 46.85 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      69647          76.6101 %
Kappa statistic                    0.5551
Mean absolute error                 0.2247
Root mean squared error             0.3358
Relative absolute error              61.0758 %
Root relative squared error         78.2979 %
Total Number of Instances          90911

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.914	0.350	0.712	0.914	0.800	0.581	0.855	0.809	1_Talep
	0.703	0.097	0.859	0.703	0.773	0.624	0.860	0.858	2_sikayet
	0.015	0.001	0.437	0.015	0.028	0.071	0.754	0.146	5_Bilgi
Weighted Avg.	0.766	0.215	0.763	0.766	0.744	0.572	0.852	0.794	

```

=== Confusion Matrix ===
  a   b   c  <-- classified as
40376 3746  70 |  a = 1_Talep
12284 29195  28 |  b = 2_sikayet
 4087  1049  76 |  c = 5_Bilgi

```

Şekil 51: Öznitelik Seçimi Sonrası J48 Algoritması (TFIDF Ağırlıklandırma) Sonucu

```

Time taken to build model: 26.73 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      69647          76.6101 %
Kappa statistic                    0.5551
Mean absolute error                 0.2247
Root mean squared error             0.3358
Relative absolute error              61.0758 %
Root relative squared error         78.2979 %
Total Number of Instances          90911

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.914	0.350	0.712	0.914	0.800	0.581	0.855	0.809	1_Talep
	0.703	0.097	0.859	0.703	0.773	0.624	0.860	0.858	2_sikayet
	0.015	0.001	0.437	0.015	0.028	0.071	0.754	0.146	5_Bilgi
Weighted Avg.	0.766	0.215	0.763	0.766	0.744	0.572	0.852	0.794	

```

=== Confusion Matrix ===
  a   b   c  <-- classified as
40376 3746  70 |  a = 1_Talep
12284 29195  28 |  b = 2_sikayet
 4087  1049  76 |  c = 5_Bilgi

```

Şekil 52: Öznitelik Seçimi Sonrası J48 Algoritması (TF Ağırlıklandırma) Sonucu

Naive bayes algoritması sınıflandırma işleminde eğitim verisini incelerken, K-en yakın komşuluk algoritması (K-NN) tümevarım yöntemini kullanmaktadır. Bu algoritma ile sınıflandırma uygulandığında tüm metinler algoritmanın belleğinde tutularak yeni metin kendisine en yakın k adet komşu metni seçer. İlgili seçme işlemi öklid uzaklığı, kosinüs

benzerliđi ölçüsü ve daha karmaşık bir sınıflandırma yöntemi olan uzaklıkları ağırlıklandırma yöntemi kullanılarak gerçekleştirilir (Kesgin, 2007: 36).

K-en yakın komşuluk algoritması örnek tabanlı bir yaklaşım olması, sadeliđi ve doğruluk yüzdesi sebebiyle metin sınıflandırma işlemlerinde sıklıkla kullanılmaktadır. Bahse konu algoritmada bir belgeyi sınıflandırma işlemi için, eğitim veri seti içinde belgenin komşuları e-en yakın komşuların sınıf etiketini kullanmaktadır. İki örnek belge arasındaki mesafe Öklid ölçü mesafesi ile ölçülebilmektedir. Sınıflandırmanın K-NN aracılığıyla yapıldığı aşamada, sınıflandırmayı etkileyen en önemli parametre K-NN parametresidir (Uguz, 2011: 1027). Metnin kendisine en yakın komşu metinleri seçebilmesi amacıyla yapılacak inceleme işlemi için algoritma tarafından yeniden hesaplama gerektiđi için işlem hızı diğer algoritmalara göre düşüktür. Kategorilerdeki metin sayılarının birbirinden farklı olmasına rağmen birbirine benzemeyen metinleri de aynı sınıfa dahil etmesi K-NN algoritmasının dezavantajı olarak görülmektedir. IBk algoritması; sınıflandırma işlemleri gerçekleştirmek üzere kullanılan K-NN algoritmasının Weka'da kullanılan versiyonudur. IBK algoritması kullanılarak tüm veri seti üzerinden makine öğrenmesi süreci algoritmanın çalışma sürecinin çok uzun olması ve veri setinin büyüklüğünde yaşanan kısıt nedeniyle çalıştırılmamıştır.

IBk algoritması çözüm masası verisi üzerinden başvuruların tahmin edilmesine yönelik öznitelik seçimi yapıldıktan sonra algoritmaların çalıştırılması neticesinde çözüm masası verisi üzerinden başvuruların tahmin edilmesine yönelik gerçekleştirilen sınıflandırma işleminde TFIDF ağırlıklandırma yönteminde %76.3252'lik oranla 90.911 başvurunun 69.388 adedi, TF ağırlıklandırma yönteminde %76.3439'luk oranla 90.911 başvurunun 69.405 adedi doğru sınıflandırmıştır. Algoritma özet sonuçları; Öznitelik Seçimi Sonrası IBk Algoritması (TFIDF Ağırlıklandırma) Sonucu Şekil 53'te ve Öznitelik Seçimi Sonrası IBk Algoritması (TF Ağırlıklandırma) Sonucu Şekil 54'te görülmektedir.

```

Time taken to build model: 0.05 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      69388          76.3252 %
Kappa statistic                    0.5507
Mean absolute error                 0.2189
Root mean squared error             0.3377
Relative absolute error             59.4923 %
Root relative squared error         78.7216 %
Total Number of Instances          90911

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.913	0.355	0.709	0.913	0.798	0.577	0.855	0.818	1_Talep
	0.696	0.095	0.861	0.696	0.769	0.620	0.861	0.847	2_sikayet
	0.032	0.003	0.372	0.032	0.060	0.096	0.758	0.160	5_Bilgi
Weighted Avg.	0.763	0.216	0.759	0.763	0.743	0.569	0.852	0.793	

```

=== Confusion Matrix ===

```

	a	b	c	<-- classified as
40346	3676	170		a = 1_Talep
12519	28873	115		b = 2_sikayet
4044	999	169		c = 5_Bilgi

**Şekil 53:** Öznitelik Seçimi Sonrası IBk Algoritması (TFIDF Ağırlıklandırma) Sonucu

```

Time taken to build model: 0.05 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      69405          76.3439 %
Kappa statistic                    0.551
Mean absolute error                 0.2188
Root mean squared error             0.3374
Relative absolute error             59.4593 %
Root relative squared error         78.6558 %
Total Number of Instances          90911

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.913	0.355	0.709	0.913	0.798	0.577	0.855	0.819	1_Talep
	0.696	0.094	0.861	0.696	0.769	0.621	0.862	0.848	2_sikayet
	0.033	0.003	0.383	0.033	0.060	0.098	0.758	0.160	5_Bilgi
Weighted Avg.	0.763	0.216	0.760	0.763	0.743	0.570	0.852	0.794	

```

=== Confusion Matrix ===

```

	a	b	c	<-- classified as
40363	3665	164		a = 1_Talep
12525	28872	110		b = 2_sikayet
4044	998	170		c = 5_Bilgi

**Şekil 54:** Öznitelik Seçimi Sonrası IBk Algoritması (TF Ağırlıklandırma) Sonucu

Dijitalleşmenin neticesinde bilgisayarların ortaya çıktığı ve moderne teorilerin gelişimi ile yapay sinir ağlarına ait işlemler 1940'lı yılların sonunda birlikte başlamıştır. Bilgisayarlar o dönemden bu yana bireysel nöronların modellenmesi işlemlerinde kullanılmaya devam etmektedir. Bugün gerçek dünyada bilgisayarların karmaşık problemleri ve desenleri tanıma problemlerine çözüm üretmesi beklenmemektedir. Normal bilgisayar işlemleri ile bu tarz karmaşık problemlere çözüm üretmek mümkün değildir. Bu noktada, insan beynindeki fizyolojik yapı ödünc alınmaktadır. Bahse konu işlemleri gerçekleştirmek üzere tercih edilen algoritmalara yapay sinir ağları (artificial neural systems, ANS) ya da sinir ağları (neural networks) denilmektedir (Freeman ve Skapura, 1991: 3).

Yapay sinir ağı çok basit ve çok sayıda birbirine bağlı nöronlar olarak da adlandırılan birbirine bağlı işlemciler olarak adlandırılabilir beyindeki biyolojik nöronlardan oluşmaktadır. Nöronlar birbirine bağlantılar ile bağlı durumdadır. Her bir nöron bağlantılarından birden fazla sinyal almasına rağmen tek bir çıkış sinyali üretmektedir. Bu şekilde bir nöronun diğerine sinyaller iletilmesiyle bir nöron dizisi oluşturulmaktadır. Gelen sinyal diğer nöronların gelen bağlantısı olacak şekilde sona ermektedir (Negnevitsky, 2005: 167).

Yapay sinir ağı algoritması en uygun mimarinin temel işlevinin öğrenmek için yeterince büyük ve iyi, genelleme için yeterince küçük bir ağı olmasına dayanmaktadır. Daha küçük bir ağı optimal mimari ile problemi iyi öğrenemezken diğer yandan geniş bir ağı ise zayıf genelleme yapmak suretiyle eğitim verisini aşacaktır. Yapay sinir ağı küçük seçilip öğrenme sürecinde büyüyor ise büyüyen/yapıcı bir yaklaşım, büyük seçilip öğrenme süresince küçülüyor ise budama/yıkıcı bir yaklaşım kullanılmaktadır. İkisini de kullanan hibrit algoritmalarda bulunmakla birlikte genellikle yapıcı yaklaşım yıkıcı yaklaşıma göre daha çok tercih edilmektedir. (Aran ve diğerleri, 2009: 160). Multilayer Perceptron algoritması; sınıflandırma işlemleri gerçekleştirmek üzere kullanılan yapay sinir ağı algoritmasının Weka’da kullanılan versiyonudur. Multilayer Perceptron algoritması kullanılarak tüm veri seti üzerinden makine öğrenmesi süreci algoritmanın çalışma sürecinin çok uzun olması ve veri setinin büyüklüğünde yaşanan kısıt nedeniyle çalıştırılmamıştır.

MultiLayer algoritması çözüm masası verisi üzerinden başvuruların tahmin edilmesine yönelik öznelik seçimi yapıldıktan sonra algoritmaların çalıştırılması neticesinde çözüm masası verisi üzerinden başvuruların tahmin edilmesine yönelik gerçekleştirilen sınıflandırma işleminde TFIDF ağırlıklandırma yönteminde %71.6217’lik oranla 90911 başvurunun 65.112’adedi, TF ağırlıklandırma yönteminde %73.3113’lük oranla 90.911 başvurunun 66.648’adedi doğru sınıflandırmıştır. Algoritma özet sonuçları; Öznelik Seçimi Sonrası MultiLayer Algoritması (TFIDF Ağırlıklandırma) Sonucu Şekil 55’te ve Öznelik Seçimi Sonrası MultiLayer Algoritması (TF Ağırlıklandırma) Sonucu Şekil 56’da görülmektedir.

```

Time taken to build model: 1455.47 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      65112          71.6217 %
Kappa statistic                    0.464
Mean absolute error                0.2248
Root mean squared error            0.3547
Relative absolute error            61.0807 %
Root relative squared error        82.7028 %
Total Number of Instances          90911

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.736	0.245	0.740	0.736	0.738	0.491	0.842	0.811	1_Talep
	0.784	0.290	0.695	0.784	0.737	0.493	0.845	0.833	2_sikayet
	0.006	0.001	0.372	0.006	0.011	0.040	0.704	0.119	5_Bilgi
Weighted Avg.	0.716	0.251	0.698	0.716	0.696	0.466	0.835	0.781	

```

=== Confusion Matrix ===
  a    b    c  <-- classified as
32542 11620  30 |  a = 1_Talep
 8947 32541  19 |  b = 2_sikayet
 2499  2684  29 |  c = 5_Bilgi

```

**Şekil 55: Öznitelik Seçimi Sonrası MultiLayer Algoritması (TFIDF Ağırlıklandırma) Sonucu**

```

Time taken to build model: 1384.35 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      66648          73.3113 %
Kappa statistic                    0.4944
Mean absolute error                0.2169
Root mean squared error            0.3514
Relative absolute error            58.9381 %
Root relative squared error        81.917 %
Total Number of Instances          90911

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,805	0,290	0,724	0,805	0,762	0,516	0,843	0,809	1_Talep
	0,748	0,216	0,744	0,748	0,746	0,532	0,847	0,833	2_sikayet
	0,006	0,001	0,369	0,006	0,012	0,041	0,703	0,122	5_Bilgi
Weighted Avg.	0,733	0,240	0,713	0,733	0,712	0,496	0,837	0,780	

```

=== Confusion Matrix ===
  a    b    c  <-- classified as
35566  8589  37 |  a = 1_Talep
10440 31051  16 |  b = 2_sikayet
 3098  2083  31 |  c = 5_Bilgi

```

**Şekil 56: Öznitelik Seçimi Sonrası MultiLayer Algoritması (TF Ağırlıklandırma) Sonucu**

Makine öğrenmesi algoritmalarının uygulanacağı verinin kalitesi ve hedeflenen amaca göre uygulanması önem arz etmektedir. Çözüm masası verisine en uygun makine öğrenmesi algoritmasını belirlemek adına farklı algoritmalar denenmiş ve sonuçları listelenmiştir.

## SONUÇ VE ÖNERİLER

Kamu kurumları sağlık, çevre, kent ve enerji gibi çeşitli hizmet alanlarında vatandaşa kaliteli hizmet sunmak, karşılaşılan problemlere çözüm üretmek, karar vericilere geleceğe dönük tahminlere dayalı gerçekçi politikalar üretmede yardımcı olmak üzere bilgi ve iletişim teknolojilerini kullanmaktadır. Yönetsel, toplumsal ve ekonomik anlamda yaşanan dönüşüm ve özellikle 1980'ler sonrası teknolojik anlamda yaşanan ilerlemeler neticesinde disiplinler arası bir alan olarak veri madenciliği, son zamanlarda ise büyük veri uygulamaları özel sektör ve kamu sektörünün ilgisini çekmektedir. Devletlere, kurumlara ve özel sektöre yönetsel süreçlerde rekabet üstünlüğü sağlayan bilginin önemi bulunduğumuz çağda artmıştır. Geleneksel donanım ve yazılım çözümleriyle saklanması, işlenmesi, paylaşılması ve analiz edilmesi kurumlara yüksek maliyetlere mal olan bilgi günümüzde gelişen bilişim teknolojilerine yapılan yatırımlar sayesinde karar alma, politikalar üretme, hizmet sunma, kar elde etme amacıyla özel sektör ve kamu kurumları tarafından kullanılmaya başlanmıştır. Örgütler veya bireysel kullanıcıların özel ya da kamusal işlemler için kurum içi ve kurumlar arası gerçekleştirdikleri işlemler neticesinde oluşan büyük verinin saklanması, işlenmesi, anlamlı hale getirilmesi, analizi ve geleceğe dönük tahminlerde kullanılabilmesi amacıyla yeni teknolojilere ihtiyaç duyulması neticesinde büyük veri ve veri madenciliği uygulamaları ortaya çıkmıştır.

Büyük veri usulsüzlüklerin belirlenmesi ve yasal uyumsuzlukların tespiti maksadıyla kamu denetiminde, sosyal davranış ve ilişkileri düzenlemek üzere kamu düzenini sağlamada, altyapı, yollar, kent yaşamı gibi belirli hizmetleri sağlamak üzere kamu hizmeti sunumunda kamu kurumlarına yeni imkânlar sağlamıştır. Kamu kaynaklarının etkin, verimli şekilde kullanılması, uluslararası ortamda stratejik üstünlük sağlamak üzere eldeki büyük verinin karar verme süreçlerinde ve sunulan hizmetlerde hızlı biçimde değerlendirilmesi kamu kurumları açısından önemlidir. Büyük veri sağlamış olduğu imkânların yanında gizlilik, veri analizi, görselleştirme ve nitelikli personel konusunda zorluk ve riskleri de beraberinde getirmiştir.

Teknolojideki gelişmeler neticesinde otonom sistemler, bulut bilişim, dronlar, yapay zeka, nesnelerin interneti, giyilebilir teknoloji, artırılmış ve sanal gerçeklik uygulamalarının yanı sıra Endüstri 4.0 yaklaşımının da etkisiyle kamu kurumlarının iş

yapma ve hizmet verme şekli dönüşüme uğramış ve uğramaya devam etmektedir. Akıllı kent yaklaşımıyla birlikte nesnelerin interneti teknolojileriyle sensörler üzerinden üretilen veri, sosyal medyanın artan kullanımıyla artan veri, akıllı saat gibi giyilebilir teknolojiler yardımıyla kişilere ait sağlık verisi, lokasyon verisi, dronlar ile toplanan milli savunmaya yönelik sınır güvenliğiyle ilgili veri, kentlerde trafik güvenliğine yönelik toplanan veri ve son dönemde yapay zeka verisi gibi diğer akıllı teknolojilerin kullanımıyla üretilen büyük veri üzerinde gerekli veri madenciliği aşamaları takip edilerek anlamlı bilgi elde edilebilmektedir.

Vatandaş; internet ve bulut bilişim yaklaşımıyla sunulan hizmetlere her noktadan ulaşabilir hale gelmiştir. Aynı zamanda kamu sektörü hizmetlerini özel sektörden almak suretiyle kamu-özel işbirliğine gidilmiş, hizmetlerin etkin, verimli ve kamu kaynaklarından tasarruf edilerek sunulması sağlanmıştır. Bir taraftan bilişim hizmetlerinin idamesi noktasında donanım, işletme giderleri ve uzman personel noktasında maliyet tasarrufu sağlanırken diğer taraftan profesyonel kurumlar tarafından işletilmesi sayesinde vatandaşa ve ilgili kurumlara kaliteli şekilde sunulması sağlanmıştır. Uygulamaların yavaş çalışması, güvenlik sorunları, hizmet sağlayıcıların güvenilirliği, bilgi güvenliği, erişim denetimi, siber saldırılar ve adli incelemelerde delillerin elde edilmesindeki hukuki zeminin sağlanamamış olması hizmetlerin bulut bilişim altyapısı ile sunulmasına doğru yaşanan dönüşüm esnasında kurumların yaşadığı olumsuzlar olarak görülmektedir.

Büyük veri uygulamaları öncelikli olarak özel sektör tarafından uygulanmaya başlanmış olmakla birlikte öze sektörün elde ettiği faydaların kamu sektörü tarafından görülmesiyle kamu tarafından da büyük veri projeleri geliştirilmeye başlanmıştır. Endüstri 4.0 devriminin kamuya yansması ile birlikte büyük veriye sahip olan kurumlar arasında kamu kurumlarının ilk sıralarda yer alacağı düşünüldüğünde büyük veri projelerinden elde edilecek fayda ve kamu tasarrufunun da büyük olacağı değerlendirilmektedir. Bu noktadan hareketle veri üzerinde gerçekleştirilecek veri madenciliği, makine öğrenmesi ve yapay zeka uygulamaları kamu yöneticilerine kuruma yön verme adına alacakları geleceğe dönük kararlarda olumlu etki yapacaktır.

Kurumların sahip olduğu muhasebe, tıbbi tetkik, ulaşım, ekonomik faaliyet ve mekânsal veri gibi farklı kaynaklardan elde edilen büyük veri yığınları ham biçimde bulunduğu

için karar vericilere kurumun geleceğe dönük kararlarını vermek üzere yardımcı olamamaktadır. Stratejik kararların alınması, kamu kaynağının etkin kullanılması ve geleceğe dönük tahminlerde kullanılmak üzere ham verinin nitelikli bilgiye dönüştürülmesi ve analiz edilmesi amacıyla kullanılan yöntem ve tekniklerin yetersiz kalmaya başlaması neticesinde istatistik, makine öğrenmesi, veritabanı teknolojileri ve yapay zekâ disiplinlerini bir araya getiren veri madenciliği günümüzde kurumların ilgisini çekmektedir. Verinin toplanması, sınıflandırılması, analiz edilmesi ve yorumlanması istatistik bilimini, bilgisayar faaliyetlerinin canlılara benzer şekilde yerine getirilmesi yapay zekâyı, verinin depolanması veritabanı ve veri ambarını, verinin türüne göre öğrenilmesi makine öğrenmesini gerektirdiği için veri madenciliği disiplinler arası bir alan haline gelmiştir. Metin içerikli veri üzerinde yapılan uygulama metin madenciliği, yapılandırılmamış link, istatistik şeklinde web içeriği üzerinde yapılan uygulama web madenciliği olarak tanımlanmaktadır.

Veri madenciliği işin tanımlanması, verinin tanımlanması, verinin hazırlanması, modelleme, değerlendirme ve uygulama olmak üzere 6 aşamadan meydana gelmektedir. Kurumlar tarafından ortak bir süreç modelinin kabul edilmesi, tüm paydaşlar tarafından veri madenciliği sorunlarının anlaşılması açısından önemlidir. Bu sayede kurumlar projenin nasıl ilerleyeceği ve sonunda ne beklemeleri gerektiği konusunda bir fikir sahibi olabilmektedir. Ortak süreç modelinin kullanılması kurum içindeki bilgi ve deneyimin paylaşılması açısından da önemlidir.

Türkiye özelinde yürütülen kamu hizmetleri vatandaşın değişen talepleri doğrultusunda büyümeye devam etmektedir. Kamu kurumları geçmişten günümüze bir yığın ham veriye sahip olmasına karşın sahip olduğu verinin kurum için ne ifade ettiğinin farkında değildir. Son dönemde büyük veri projeleriyle elde edilen kazanımlar kamu yöneticilerinin de dikkatini çekmeyi başarmıştır. Ekonomik anlamda iş ve stratejik kararların alınması, kamu yatırımlarının ülkenin gelir düzeyi ile doğru orantılı olarak gerçekleştirilmesi gibi kamu yöneticilerinin alacağı kararlarda büyük veri ve veri madenciliği uygulamaları ile karar desteği sağlanmaktadır. Veri madenciliği uygulamaları ve büyük veri projeleri özel sektöre rekabet avantajı sağlarken sağlık, eğitim, güvenlik, savunma, enerji ve ulaşımaya yönelik kamu hizmetlerinin etkin verimli

ve kaliteli biçimde vatandaşa ve diğer paydaşlara sunulmasında, kamu yöneticilerinin daha etkin karar almasında önemli rol oynamaktadır.

Kamuya ait stratejik planları, raporlar, hükümet programları, kalkınma planları ve eylem planlarının incelenmesi neticesinde kamu kurumlarında özellikle son dönemde büyük veriye yönelik farkındalık oluşmuş durumdadır. Merkezi idare kurumlarının hazırlamış olduğu eylem planları ve kurumların stratejik planlar doğrultusunda belirlemiş olduğu stratejik hedefleri gerçekleştirmek üzere geliştirmiş olduğu projeler kamunun yenilikçi teknolojiler olarak büyük veri ve veri madenciliği uygulamalarını kullanma yönündeki isteğini ortaya koyması açısından önemlidir. Aynı şekilde büyük veri ve veri madenciliğine atıfta bulunan yukarıda açıklanan yasaların yürürlüğe girmesi de kullanılan teknolojilerin yasal zemininin oluşturulması açısından olumlu adımlardır. Bu noktada dikkat edilmesi gereken husus, yayımlanan resmi politika belgeleri, kabul edilen kanunlar ve idari düzenlemeler doğrultusunda kamu kurumları tarafından gerekli uygulamaların, üst yönetimin desteği ve teknik/uzman personelin katkılarıyla hayata geçirilmesidir.

Türkiye’de yeni geçilen başkanlık sisteminin etkilerinin gözlemlenmesi için zamana ihtiyaç olmakla birlikte, Cumhurbaşkanlığı teşkilatında kurulan “Bilim, Teknoloji ve Yenilik Politikaları Kurulu”, büyük veri teknolojileri ve veri madenciliği gibi yenilikçi teknolojilerin kamu kurumları tarafından hangi hizmet alanlarında yararlanabileceğine ilişkin ciddi öneriler getirebilir. Ayrıca “Dijital Dönüşüm Ofisi”nin görevi içinde büyük veri analizi yapmak ve öncelikli proje alanlarında yapay zekâ uygulamalarına öncülük etmek yer almaktadır. Bu kapsamda önümüzdeki dönemde büyük veri analizi, veri madenciliği uygulamaları ve makine öğrenmesi metodlarının kamu tarafından daha kapsamlı ve yoğun bir şekilde kullanması beklenebilir.

Yerel yönetimler açısından büyük veri ve veri madenciliği uygulamaları değerlendirildiğinde, özellikle büyük kentlerde büyükşehir belediyelerinin bu alanlarda yatırımlar yaptığı görülmektedir. Ancak büyük veri teknolojilerine temel oluşturan e-belediyeçilik ve kurumlar arasında açık veri paylaşım ve protokollerinin istenilen düzeyde olmaması veri tekrarlarının oluşmasına neden olabilmektedir. Ayrıca kamu politikası üretmek üzere büyükşehir belediyelerinin sahip olduğu verilerden anlamlı bilgiler elde edecek yazılım ve kaliteli personel ihtiyacı bulunmaktadır. Özellikle bu

tezin konusunu oluşturan büyükşehir belediyelerine ait stratejik planlar incelendiğinde yeni kurulan büyükşehir belediyelerinde yönetim bilişim sistemlerinin kurulması ya da yaygınlaştırılmasının ortak stratejik hedef olarak belirlendiği görülmüştür. Bununla birlikte, uygulamada veri madenciliğinin büyükşehir belediyeleri ölçeğinde yoğun ve sistematik bir şekilde kullanılmadığı anlaşılmaktadır.

Çalışmanın son bölümünü oluşturan veri madenciliği uygulama sürecinde kullanılan araç ve yöntemlerin kurumların problemlerine kendiliğinden çözüm bulması beklenmemeli, her aşamasında insan tecrübesi dâhil edilerek uygulama aşamaları planlı şekilde yapılandırılmalıdır. Bu noktadan hareketle verinin ön hazırlık aşaması büyük önem taşımaktadır. Çözüm masası verisi üzerinde gerçekleştirilen veri madenciliği uygulaması vatandaş tarafından serbest metin şeklinde girilen tüm harflerin küçük harfe çevrilmesi, noktalama işaretlerinin kaldırılması, anlamsız, yanlış kelimelerin çıkarılması, gürültülü veri, hatalı veri toplama araçlarının kullanılması, veri giriş problemleri, teknolojik sınırlamalar, veri yığını içinde, anlam ifade etmeyen boş verinin örnek veri seti hazırlığında temizlenmesi ve köklerin bulunması vb. işlemler ön hazırlık aşamasının ne kadar önemli olduğunu göstermiştir. Diğer bir husus makale, köşe yazısı ya da yapılandırılmış metin gibi küçük ve tutarlı veri kümelerine hızlı ve doğru biçimde veri madenciliği teknikleri uygulanabilirken çözüm masası örneğinde olduğu üzere yapılandırılmamış ve 90 bin üzerinde büyük veri kümelerine veri madenciliği teknikleri uygulandığında yüksek doğrulukta sonuçlar elde edilmesi güçleşmiştir. Başvuru çeşidine göre 100'er ve 500'er örnek veri üzerinden yapılan veri madenciliği uygulaması ile 90 bin üzerinde tüm başvuru verisi üzerinde gerçekleştirilen veri madenciliği uygulaması doğruluk açısından farklılık göstermiştir. Örnek veri miktarının artması algoritmanın modeline yönelik olumlu yönde katkı sağlarken veri madenciliği sürecinin özellikle ön işleme ve makine öğrenmesi sürecinin uzamasına sebep olmaktadır.

Veri madenciliği uygulamaları 7x24 esasına göre hizmet veren üzerinde dinamik veri barındıran operasyonel veritabanları yerine veri ambarları ya da veri madenciliği uygulamasını gerçekleştirmek üzere hazırlanmış doyalar üzerinde gerçekleştirilmelidir. Çözüm masası verisi operasyonel veritabanından farklı bir veritabanına taşınarak üzerinde gerekli ön hazırlık işlemleri uygulanmıştır. Veritabanı boyutları artmaya

devam ederken veri madenciliği örnek veri kümeleri üzerinde çalıştırılan algoritmalar ile test edilerek oluşturulmaktadır. Büyük veri üzerinde veri madenciliği uygulamasının çalıştırılması örüntülerin gerçekliğinin tespiti açısından önemli olmakla birlikte örneklem sayısının fazla olması sürecin uzun olmasına sebebiyet vermektedir. Bu nedenle veri kümesinde, çözüme yönelik olmayan gereksiz, nitelsiz artık verinin veri madenciliği teknikleri uygulanmadan önce temizlenmesi gerekmektedir.

Bu çalışmada, makine öğrenmesi algoritmalarının metinlere uygulanabilmesi için metinler sayısallaştırma işlemine tabii tutularak literatürde yaygın olarak kullanılan vektör uzay modeline çevrilmiştir. Yapısal forma dönüştürme esnasında indeksleme aşamasında metinlere frekansa göre ağırlıklandırma (tf) ve terim frekansı x ters doküman frekansı ağırlıklandırma (tfidf) indeksleme yöntemleri uygulanmıştır. Ön hazırlık aşamasından geçirilen veri seti üzerinde veri madenciliği için makine öğrenmesi algoritmalarını çalıştırmak üzere WEKA uygulamasına özel ARFF formatında dosya hazırlanmıştır. Bu sayede sonuçların güvenilirliği artırılmıştır.

Çözüm masası örneğinde vatandaş tarafından yapılan başvuruların sınıflandırılmasına yönelik makine öğrenmesi algoritmaları çalıştırılmıştır. Bu sayede sınıflandırma işlemini en yüksek doğruluk oranı ile gerçekleştirilen algoritma tespit edilmiş olup SMO algoritması J48 algoritmasına göre daha yüksek performans göstermiştir. Bu nedenle çözüm masası verisi üzerinden vatandaş başvurusunun tahmini için en uygun algoritma, SMO algoritması olarak belirlenmiştir.

Büyük veri ve veri madenciliği uygulamaları, veriden öğrenen ve kendini geliştiren bir yapay zekâ ortamı yaratmaktadır. Örnek çalışmamızda çözüm masası uygulamasını kullanan sorumlu personelin bilgi ve tecrübesi dahilinde gelen başvuruları ilgili bölümlere ve başkanlıklara ataması ile süreç ilerlemektedir. Dolayısıyla çözüm masasına yeni bir personelin atanması halinde yönlendirmeleri istenilen doğrulukta ve isabetli olmayacak, sürekli değişken bir ortam olacaktır. Oysa makine öğrenmesi sayesinde gelen veri ile kendini sürekli geliştiren bir çözüm masası uygulaması sayesinde süreç ve iş yapış yöntemi sürekli gelişecek, daha kaliteli hizmet verilecektir. Diğer taraftan birim, yönlendirmelerinde çoğu zaman insanların kararsız kaldıkları durumlar söz konusudur ve her insan farklı karar verebilir. Oysa makine öğrenmesi sayesinde geçmişte verilen doğru ve yanlış kararlardan öğrenen makine kendine göre

daha yeni bir algoritma (generative) geliştirebilecektir. Böylece geçmişte hiç karşılaşılmayan bir duruma özgü kararlar da alabilecektir.

Son olarak yönetim bilişim sistemlerinin entegrasyonu sayesinde veri madenciliği uygulamaları üzerinde biriken bilgi web servisleri üzerinden paylaşılabilir. Vatandaşa ve karar vericilere sunulabilecek ve diğer uygulamalar da bu hizmetten faydalanacaktır. Örneğin iş zekâsı uygulamaları ile çözüm masasının ürettiği dönemsel raporlar belirli formatta ilgili başkanlıklara gönderilebilecek, istatistiki veri, karar destek uygulamalarında karar vericilerinin önünde sunulabilecektir. Mesela çözüm masasına ulaşan çevre düzenlemesine yönelik şikayetlerin A ilçesinde yoğunlaşması geleceğe yönelik yatırımların bu alana yönlendirilmesini sağlayacaktır.

Bu çalışmada, yerel yönetimlerde özelinde veri madenciliğine yönelik yapılan araştırma ile hem bu alanda ilgili literatürdeki eksiklik giderilmeye çalışılmış hem de bir uygulama modelinin geliştirilmesi için ilk adımlar atılmıştır. Bu çalışmada veri madenciliğine odaklanılmıştır, ancak büyük veri analizleri ve Endüstri 4.0'ın kamu kurumlarını ve işleyişini ne şekilde değiştireceği, iş zekâsı uygulamalarının kamu yöneticilerine yapacağı katkının da araştırılmasına ihtiyaç bulunmaktadır. Söz konusu yeni teknolojilerin gelecekte kamuda kullanılmakta olan yönetim bilişim sistemleri ve e-devlet uygulamalarıyla bütünlüğün sağlanması ve etkisinin araştırılarak, birbirinden farklı çalışan bu sistemleri bütünleştirecek uygulama ve modellerin geliştirilmesi gerekmektedir.

Günümüzde büyük kapasiteli veri depolama ve hızlı donanım maliyetlerinin düşmesi, gelişen veri madenciliği uygulamaları ve makine öğrenmesinde kullanılan algoritmaları sayesinde veri madenciliği süreci daha kolay hale gelmiştir. Fakat veri madenciliğinin büyük veri üzerinde uygulanması kamu kurumlarında uzman personeli bulundurma güçlüğü ve karar vericilerin veri madenciliği kabiliyetleri konusunda bilgilendirilmesi ve eğitilmesi ihtiyacı nedeniyle güç olabilmektedir. Her kurumda veri madenciliği konusunda uzman çalıştırılmayacağı ve her kamu yöneticisine eğitim verilemeyeceği göz önünde bulundurulduğunda veri madenciliği süreçlerine ait akışın otomatik hale getirilerek modellenmesi bulut bilişim hizmetlerine dönüştürmek suretiyle ilgili kamu kurumları tarafından kullanılabilir hizmet haline dönüştürülmesi bu güçlüğü ortadan

kaldıracağı gibi farklı kurumlara ait veri kümeleri içinde kullanılması mümkün hale getirecektir.

Büyük veri ve veri madenciliği uygulamalarına yönelik gerçekleştirilen örnek uygulama özelinde büyükşehir belediyelerinde daire başkanlıkları düzeyinde çalışma gerçekleştirilerek bu birimlerin geleceğe dönük karar almasında ve hizmet kalitesini iyileştirmesinde faydalanması sağlanabilir. Örneğin, Ulaşım Daire Başkanlığının temel problemi olarak otobüslerin optimum hizmet verecek şekilde kullanılması ele alındığında bu kapsamda mevcut otobüslerin hangi güzergahlarda, hangi saatlerde çalışacağı, bu kapsamda ne kadar şoföre ihtiyaç duyulacağı, çalışma saatlerine uygun olarak gelecek dönemde otobüslerin bakım ihtiyacı ve harcamaların nasıl yapılandırılacağına dair geleceğe dönük tahmin ihtiyaç duyulan verinin sağlanması halinde gerçekleştirilebilir. Diğer geleceğe dönük gerçekleştirilebilecek çalışma vatandaşların talepleri doğrultusunda gerçekleştirilen yatırımlar ve yerel seçim sonuçlarının karşılaştırılmasıdır. Yapılan yatırımların yerel seçimlerde geri dönüşü tahmin edilebilir. Bu sayede yatırım yapılan bölgelerde vatandaş memnuniyetinin seçim sonuçlarına ne ölçüde yansıdığı tahmin edilebilir.

## KAYNAKÇA

### Kitaplar

- Aggarwal, C. C. ve P. S. Yu. (1999). Data Mining Techniques for Associations, Clustering and Classification. Ning Zhong ve Lizhu Zhou (Ed.). *Methodologies for Knowledge Discovery and Data Mining (Third Pacific-Asia Conference, PAKDD-99 Beijing, China, April 26-28, 1999 Proceedings)* içinde. Berlin: Springer, 1999,13-24.
- Aktan, C. C. ve İ. Y. Vural. (2005). Bilgi Çağında Bilginin Yönetimi. C.C. Aktan ve İ.Y. Vural (Ed.). *Bilgi Çağı Bilgi Yönetimi ve Bilgi Sistemleri* içinde. Konya: Çizgi Kitabevi, 2005, 1-30.
- Al, H. (2002). *Bilgi Toplumu ve Kamu Yönetiminde Paradigma Değişimleri*. Ankara: Bilimadamları Yayınları.
- Bai, J. ve J. Sarkis. (2016). Green Government Procurement: Decision-Making with Rough Set, TOPSIS, and VIKOR Methodologies. C.G. Reddick (Ed.). *Public Administration and Information Technology* içinde. New York: Springer, 2016, 93-121.
- Barutçugil, İ. (2002). *Bilgi Yönetimi*. İstanbul: Anı Yayıncılık.
- Bensghir, T.K. (1996). *Bilgi Teknolojileri ve Örgütsel Değişim*. Ankara: TODAİE Yayınları.
- Buss, T.ve F.S.Redburn. (2006). Information Technology and Governance. T.Buss (Ed.). *Modernizing Democracy: Innovations in Citizen Participation* içinde. ABD: M.E. Sharpe Inc., 2006, 114-135.
- Castells, M. (2013). *Enformasyon Çağı: Ekonomi, Toplum ve Kültür Birinci Cilt: Ağ Toplumunun Yükselişi*. İstanbul: İstanbul Bilgi Üniversitesi Yayınları.
- Castells, M. (2008). *Enformasyon Çağı: Ekonomi, Toplum ve Kültür İkinci Cilt: Ağ Kimliğinin Gücü*. İstanbul: İstanbul Bilgi Üniversitesi Yayınları.
- Chaouchi, H. (2010). Introduction to the Internet of Things. H.Chaouchi (Ed.). *The Internet of Things Connecting Objects to the Web* içinde. Great Britain: Wiley, 2010, 1-32.
- Davenport, T. H. ve L. Prusak. (2001). *İş Dünyasında Bilgi Yönetimi: Kuruluşlar Elleriindeki Bilgiyi Nasıl Yönetirler*. G. Güney (çev.). İstanbul: Rota Yayın Yapım (orijinal baskı tarihi 1998).
- Deng, N., Y. Tian ve C. Zhan. (2013). *Support Vector Machines Optimization Based Theory, Algorithms, and Extensions*. Boca Raton: CRC Press.
- Feldman, R. ve J. Sanger. (2007). *The Text Mining Handbook: Advanced Approaches in*

*Analyzing Unstructured Data*. Cambridge University Press, New York.

- Freeman, J.A. ve D.M. Skapura. (1991). *Neural Networks: Algorithms, Applications, And Programming Techniques*. Addison-Wesley Publishing, *Massachusetts*.
- Foster, I., R. Ghani, R.S. Jarmin, F. Kreuter ve J. Lane. (2017). *Big Data and Social Science A Practical Guide to Methods and Tools*. Boca Raton: CRC Press Taylor & Francis Group.
- Geddes, S. ve K. Lai. (2017). Analyzing Predictors Of Severe Traffic Accidents. G. Richards (Ed.). *Big Data and Analytics Applications in Government Current Practices and Future Opportunities* içinde. New York: Auerbach Publications, 2017, 190-205.
- Gilchrist, A. (2016). *Industry 4.0: The Industrial Internet of Things*. Thailand: APress.
- Goh, S.C., C. Elliott ve G. Richards. (2017). Setting the Context for Analytics: Performance Management in Canadian Public Organizations: Findings of a Multi-Case Study. G. Richards (Ed.). *Big Data and Analytics Applications in Government Current Practices and Future Opportunities* içinde. New York: Auerbach Publications, 2017, 29-55.
- Gürsoy, U.T.Ş. (2012). *Uygulamalı Veri Madenciliği Sektörel Analizler*. Ankara: Pegem.
- Han, J., M. Kamber ve J. Pei. (2012). *Data Mining Concepts and Techniques*. London Elsevier.
- Kale, V. (2018). *Creating Smart Enterprises Leveraging Cloud, Big Data, Web, Social Media, Mobile and IoT Technologies*. New York: CRC Press Taylor & Francis Group.
- Kendall, K. E. ve J.E. Kendall. (2010). *Systems Analysis and Design*. 8.Baskı. New Jersey: Prentice Hall.
- Korkontzelos, I. (2014). Mining Big Textual Data. S. Kudyba ve T. H. Davenport (Ed.). *Big Data, Mining, and Analytics Components of Strategic Decision Making* içinde. Boca Raton: CRC Press Taylor & Francis Group, 2014, 231-257.
- Köseçik, M. ve N. Karkın. (2004). Elektronik Devlet: Amaçlar, Sorunlar ve Uygulamalar. A. Yılmaz ve M. Ökmen (Ed.). *Kamu Yönetimi* içinde. Ankara: Gazi Kitabevi, 2004, 97-112.
- Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. New Jersey: John and Wiley Sons Incorporated.
- Larose, D. T. (2006). *Data Mining Methods and Models*. New Jersey: John Wiley & Sons Inc. Publication.
- Laudon, K.C. ve J.P. Laudon. (1998). *Management Information Systems : New*

- Approaches To Organization And Technology*. Upper Saddle River: Prentice Hall.
- Lehr, T. (2018). Smart Cities: Vision on-the-Ground. S. McClellan ve J. A. Jimenez ve G. Koutitas (Ed.). *Smart Cities Applications, Technologies, Standards, and Driving Factors* içinde. Switzerland: Springer, 2018, 3-17.
- Liu, B. (2007). *Web Data Mining Exploring Hyperlinks, Contents, and Usage Data*. New York: Springer.
- Marks, E.A. ve B. Lozano. (2010). *Executive's Guide to Cloud Computing*. New Jersey: John Wiley & Sons, Inc.
- Miller, D., E. Costa, N. Haynes, T. McDonald, R. Nicolescu, J. Sinanan, J. Spyer, S. Venkatraman ve X. Wang (Ed.). (2016). *How the World Changed Social Media*. London: UCL Press.
- Milner, E.M. (2000). *Managing Information and Knowledge in the Public Sector*. London ve New York: Routledge Taylor&Francis Group.
- Munné, R. (2016). Big Data in the Public Sector, J. M.Cavanillas, E. Curry ve W. Wahlster (Ed.). *New Horizons for a Data-Driven Economy A Roadmap for Usage and Exploitation of Big Data in Europe* içinde. Cham: Springer, 2016, 195-208.
- Negnevitsky, M. (2005). *Artificial Intelligence A Guide to Intelligent Systems*. Harlow-England: Addison-Wesley Publishing.
- Nonaka, I. ve H.Takeuchi. (1995). *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. New York: Oxford Universty Pres.
- Osborne, D. ve T. Gaebler. (1992). *Reinventing Government: How The Entrepreneurial Spirit Is Transforming Government*. Reading Mass: Adison Wesley Public Comp.
- Platt, J. C. (1998). Fast Training of Support Vector Machines Using Sequential Minimal Optimization. B. Schölkopf, C. Burges, ve A. Smola (Ed.). *Advances in Kernel Methods - Support Vector Learning* içinde. Cambridge: MIT Press, 1998,185-208.
- Pries, K. H ve R. Dunnigan. (2015). *Big Data Analytics A Practical Guide For Managers*. Boca Raton: CRC Press.
- Rittinghouse, J., W. ve J. F. Ransome. (2009). *Cloud Computing: Implementation, Management, and Security*. Boca Raton: CRC Press.
- Rocheleau, B. (2006). *Public Managment Information Systems*, Hershey: Idea Group Publishing.

- Rodriguez, E. (2017). Big Data Analytics and Public Bus Transportation Systems in China, G. Richards (Ed.). *Big Data and Analytics Applications in Government Current Practices and Future Opportunities* içinde. New York: Auerbach Publications, 2017,93-113.
- Rokach, L. ve O. Maimon. (2015). *Data Mining With Decision Trees Theory And Applications*. Singapore: World Scientific Publishing.
- Salkin, C., M. Oner, A. Ustundag ve E. Cevikcan. (2018). A Conceptual Framework for Industry 4.0. A. Ustundag ve E. Cevikcan (Ed.). *Industry 4.0: Managing The Digital Transformation* içinde. Switzerland: Springer, 2018, 3-22.
- Schedler, K. ve M.C. Scharf. (2002). Exploring the Interrelations Between Electronic Government and the new Public Management. B. Schmid, K. Stanoevska-Slabeva ve V. Tschammer(Ed.). *Towards the E-Society. IFIP International Federation for Information Processing* içinde. Boston: Springer, 2002, 775-788.
- Sharda, R., D. Delen ve E. Turban. (2014). *Business Intelligence A Managerial Perspective on Analytics*. Harlow: Pearson.
- Silahtaroglu, G. (2013). *Veri Madenciliği Kavram ve Algoritmaları*. İstanbul: Papatya Yayıncılık.
- Sobacı, M. Z. ve Ö. Köseoğlu. (2015). Yeni Kamu Yönetişimi: Birlikte Üretmenin Ve İşbirliğinin Teorik Çerçevesi. Ö. Köseoğlu ve M.Z. Sobacı (Ed.). *Kamu Yönetiminde Paradigma Arayışları: Yeni Kamu İşletmeciliği ve Ötesi*. Bursa: Dora, 2015, 231-248.
- Sobacı, M.Z. (2012). E-Devlet: Kuramsal Bir Bakış. M.Z. Sobacı ve M. Yıldız (Ed.). *E-Devlet Kamu Yönetimi ve Teknoloji İlişkisinde Güncel Gelişmeler* içinde. Ankara: Nobel Akademik Yayıncılık, 2012, 3-37.
- Sobacı, M. Z., Ö. Köseoğlu ve N. Karkın. (2015). *Belediyelerde Sosyal Medya: Değişim İçin Yeni Fırsatlar*. İstanbul: Marmara Belediyeler Birliği Kültür Yayınları.
- Şentürk, A. (2006). *Veri Madenciliği Kavram ve Teknikler*. Bursa: Ekin Yayınevi.
- Taylor, F. (2012). *Bilimsel Yönetimin İlkeleri*. H.B.Akın (çev.). Ankara: Adres Yayınları (orijinal baskı tarihi 1911).
- Trajkovik, V. (2013). *Yerel Yönetimler için Bilgi ve İletişim Teknolojileri BİT: Standartlar, İlkeler ve En İyi Uygulamalar*. İstanbul: T.C. Marmara Belediyeler Birliği Yayını.
- Vapnik, V. (1998). Three Remarks On The Support Vector Method Of Function Estimation. B. Schölkopf, C. Burges, ve A. Smola (Ed.). *Advances in Kernel Methods - Support Vector Learning* içinde. Cambridge: MIT Press,25-43.
- Varma, N. ve R.K. Tyagi. (2017). Government of India Prepares for Big Data Analytics

- Using Aadhaar Card Unique Identification System. G. Richards (Ed.). *Big Data and Analytics Applications in Government Current Practices and Future Opportunities* içinde. New York: Auerbach Publications, 2017, 115-130.
- Velte, A.T., T.J. Velte ve R. Elsenpeter. (2010). *Cloud Computing: A Practical Approach*. New York: McGraw-Hill Companies.
- Vinh, N. X. ve M. E. Houle. (2010). A Set Correlation Model for Partitional Clustering. R. Mohammed, J. Zaki, J. X. Yu, B. Ravindran ve V. Pudi (Ed.). *Advances in Knowledge Discovery (14th Pacific-Asia Conference, PAKDD 2010 Hyderabad, India, June 21-24, 2010 Proceedings Part I)* içinde. Berlin: Springer, 1999, 4-15.
- White, J.D. (2007). *Managing Information in the Public Sector*. Londra: M.E. Sharpe.
- Yavuz, N. (2015). Dijital Çağ Yönetişimi: Kamu Yönetiminde İhtiyaç Temelli Bütünleşme Çabası. Ö. Köseoğlu ve M.Z. Sobacı (Ed.). *Kamu Yönetiminde Paradigma Arayışları: Yeni Kamu İşletmeciliği ve Ötesi*, Bursa: Dora, 2015, 273-296.
- Zhang, J., L. F. Luna-Reyes ve T. A. Pardo. (2016). Information, Policy, and Sustainability: The Role of Information Technology in the Age of Big Data and Open Government. C.G. Reddick (Ed.). *Public Administration and Information Technology* içinde. New York: Springer, 2016, 1-21.
- Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort an Introduction to Human Ecology*, Reading: Addison-Wesley.

## Sürelî Yayınlar

- Ahlgren, B., M. Hidell ve C.H. Ngai. (2016). Internet of Things for Smart Cities: Interoperability and Open Data. *IEEE Computer Society*. 20.6, 52-56.
- Ak, A. (2014). Türk Kamu Yönetiminde Açık Yönetim İlkesi ve Açık Yönetim Ortaklığı. *İdarecinin Sesi*. 163, 49-53.
- Akdamar, E. (2017a). Akıllı Kent İdealine Ulaşmada Büyük Verinin Rolü. *Kent Kültürü ve Yönetimi*. 10.2, 200-215.
- Akdamar, E. (2017b). Akıllı Kent İdealine Ulaşmada Açık Verinin Rolü. *Social Sciences Research Journal*. 6.1, 45-52.
- Akın, A. ve M.D. Akın. (2007). Zemberek. An Open Source NLP Framework For Turkic Languages. *Structure*. 10, 1-5.
- Akpınar, H. (2000). Veritabanlarında Bilgi Keşfi ve Veri Madenciliği. *İ.Ü. İşletme Fakültesi Dergisi*. 29.1, 1-22.
- Alaeddinoğlu, M.F., T. Aydın ve D. Dal. (2012). Birliktelik Kuralları İle Mekânsal-Zamansal Veri Madenciliği. *EÜFBED - Fen Bilimleri Enstitüsü Dergisi*. 5.2, 191-212.
- Alçın, S. (2016). Üretim İçin Yeni Bir İzlek: Sanayi 4.0. *Journal of Life Economics*. 8, 19-30.
- Alkan, T., (2015). Akıllı Kentler ya da 21. Yüzyıl Şehirleri. *Bilişim Dergisi*. 182, 71-77.
- Altun, T., F. Şahin ve N.Öztaş. (2017). Kamu Politikalarının Belirlenmesi Ve Uygulanmasında Büyük Veri. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*. 22.Kayfor15 Özel Sayısı, 2021-2044.
- Amasyalı, M.F., M. Çetin ve C. Akbulut. (2013). Metinlerin Anlamsal Uzaydaki Temsil Yöntemlerinin Sınıflandırma Performansına Etkileri. *Yıldız Teknik Üniversitesi Sigma Dergisi*. 5.1, 8-14.
- Amasyalı, M.F., S. Balcı, E. Varlı ve N. Mete. (2012). Türkçe Metinlerin Sınıflandırılmasında Metin Temsil Yöntemlerinin Performans Karşılaştırılması. *EMO Bilimsel Dergi*. 2.4, 95-104.
- Aran, O., T. Yıldız ve E. Alpaydın. (2009). An Incremental Framework Based on Cross-Validation For Estimating The Architecture Of A Multilayer Perceptron. *Int. J. Pattern Recognit Artif Intell*. 23.2, 159-190.
- Arkun, M.E. (2003). Türkiye İçin Bir Enformasyon Politikasının Ana Öğeleri Neler Olmalı?. *Bilgi Dünyası*. 4.2, 175-191.
- Arslan, N.T. (2010). Klasik - Neo Klasik Dönüşüm Süreci: Yeni Kamu Yönetimi. *C.Ü. İktisadi ve İdari Bilimler Dergisi*. 11.2, 21-38.

- Atalay, M. ve E. Çelik. (2017). Büyük Veri Analizinde Yapay Zekâ Ve Makine Öğrenmesi Uygulamaları. *Mehmet Akif Ersoy Üniversitesi Sosyal BilimlerEnstitüsü Dergisi*. 9.22, 155-172.
- Ayvaz, Z. R., F.G. Kırbaşlar ve Z. Ö. Güneş. (2010). Fen Bilgisi Öğretmen Adaylarının Kimya Öğretiminde BDE Materyali Kullanımına İlişkin Düşünceleri. *Hasan Ali Yücel Eğitim Fakültesi Dergisi*. 14.2, 1-18.
- Aytaç, M.B. ve H.Ş. Bilge (2013). Tele Pazarlama Verilerinin Birliktelik Kurallarıyla Ve CRISP-DM Yöntemiyle Analiz Edilmesi. *Aksaray Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*. 5.2, 25-40.
- Baesens, B., C. Mues, D. Martens ve J. Vanthienen. (2009). 50 Years Of Data Mining And Or: Upcoming Trends And Challenges, *Journal Of The Operational Research Society*. 60, 16-23.
- Baltacı, C. (2004). Yeni Sağ Üzerine Bir Eleştiri. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*. 9.2, 359-373.
- Batty, M. (2013). Big Data Smart Cities and City Planning. *Dialogues in Human Geography*. 3.3, 274-279.
- Baykal, A. (2006). Veri Madenciliği Uygulama Alanları. *D.Ü.Ziya Gökalp Eğitim Fakültesi Dergisi*. 7, 95-107.
- Bettencourt, L. M. (2014). The Uses Of Big Data In Cities. *Big Data*, 2.1, 12-22.
- Bhatt, G. D. (2001). Knowledge Management In Organizations. *Journal of Knowledge Management*. 5.1, 68-75.
- Boyd, D., K. Crawford ve B. Savcı. (2015). Büyük Veri Üzerine Eleştirel Sorular: Kültürel, Teknolojik ve Bilimsel Bir Olgu Hakkında Eleştirel Sorgulamalar. *Folklor/Edebiyat*. 21.83, 199-215.
- Bozkurt, A. (2014). Türkiye’de Açık Verinin Farkında, Bir Şeyler Yapacak Ama Nasıl?. *Bilişim Dergisi*. 169, 87-91.
- Branke, J., S. Farid, ve N. Shah (2016). Industry 4.0 - A Vision Also For Personalized Medicine Supply Chains?. *Cell and Gene Therapy Insights*. 2.2, 263-270.
- Brian, W. H. (2010). Reconsidering Evidence-Based Policy: Key Issues and Challenges. *Policy and Society*. 29, 81-82.
- Brown, J., C. Cooper ve M. Pidd. (2006). A Taxing Problem: The Complementary Use Of Hard And Soft Or In The Public Sector. *European Journal of Operational Research*. 172, 666-679.
- Bulut, E. ve T.Akçacı (2017). Endüstri 4.0 Ve İnovasyon Göstergeleri Kapsamında Türkiye Analizi. *ASSAM Uluslararası Hakemli Dergi*. 7, 50-72.

- Castells, M. (2000). Materials For An Exploratory Theory Of The Network Society. *British Journal of Sociology*. 51.1, 5-24.
- Chatfield, A.T. ve O. Alhujran. (2009). A Cross-Country Comperative Anlaysia of e-Government Service Delivery among Arab Countries.. *Information Teechnology for Development*. 15.3, 151-170.
- Chen, Y., K. Tang, R. Shen ve Y. Hu. (2005). Market Basket Analysis in a Multiple Store Environment. *Decision Support Systems*. 40, 339-354.
- Codd, E. F. (1970). A Relational Model of Data for Large Shared Data Banks. *Communications of the ACM*. 13.6, 377-387.
- Cook, M.E., M.F. LaVigne, C. M. Pagano, S.S. Dawes ve T. A. Pardo. (2002). Making a Case for Local E-Government. *Center for Technology in Government*. 1-16.
- Çalış, A., S Kayapınar ve T. Çetinyokuş, (2014). Veri Madenciliğinde Karar Ağacı Algoritmaları İle Bilgisayar Ve İnternet Güvenliği Üzerine Bir Uygulama. *Endüstri Mühendisliği Dergisi*. 25.3-4, 2-19.
- Çekin, M. (2016). 6698 Sayılı Kişisel Verilerin Korunması Hakkında Kanun'un Big Data (Büyük Veri) Ve İrade Serbestisi Açısından Değerlendirilmesi. *İstanbul Üniversitesi Hukuk Fakültesi Mecmuası*. 74.2, 629-644.
- Çelik, S. ve E. Akdamar. (2018). Büyük Veri ve Veri Görselleştirme. *Akademik Bakış Dergisi*. 65, 253-264.
- Çeliktaş, M. S., G. Sonlu, S. Özgel ve Y. Atalay (2015). Endüstriyel Devrimin Son Sürümünde Mühendisliğin Yol Haritası. *Endüstri ve Mühendislik Dergisi*. 54.662, 24-34.
- Çımrın, F. K. (2009). Küreselleşme, Neo-Liberalizm Ve Refah Devleti İlişkisi Üzerine, *Muğla Üniversitesi Sosyal Bilimler Enstitüsü Dergisi (İLKE)*. Güz.23, 195-204.
- Çımrın, F. K. (2011). Manuel Castells'i Yeniden Okumak: Küresel Ağ Hareketleri Yaklaşımının Eleştirel Bir Değerlendirmesi. *Sosyal Bilimler Dergisi*. 4.2, 65-77.
- Demirel, D. (2010). Yönetişimde Yeni Boyut:E-Yönetişim. *Türk İdare Dergisi*. 466, 64-94.
- Demirtaş, M. (2012). Yerel Yönetimlerde Bir Halkla İlişkiler Çalışması Olarak Sosyal Paylaşım Ağlarının Kullanımı: Üsküdar Belediyesi Örneği. *e-Journal of New World Sciences Academy*. 7.4, 91-314.
- Dengiz, O. (2017). Endüstri 4.0: Üretimde Kavram ve Algı Devrimi. *Makina Tasarım Ve İmalat Dergisi*. 15.1, 38-45.
- Denhardt R.B. ve J.V. Denhardt (2000). The New Public Service: Serving Rather Than Steering. *Public Administration Review*. 60.6, 549-559.

- Dođan, K.C. ve F. Ustakara. (2013). Kamuda Bir Yapılanma Dönüşümü Olarak E-Devlet ve E-Yönetişim İlişkisi Üzerine., *Global Journal of Economics and Business Studies*. 2.3, 1-11.
- Dolgun, M.Ö., T.G. Özdemir ve D. Ođuz. (2009). Veri Madenciliđinin Yapısal Olmayan Analizi: Metin ve Web Madenciliđi. *İstatistikçiler Dergisi*. 2, 48-58.
- Dunleavy, P., H. Margetts, S. Bastow ve J. Tinkler. (2006). New Public Management Is Dead-Long Live Digital-Era Governance. *Journal Of Public Administration Research And Theory*. 16.3, 467-494.
- Durna, U. ve M. Özel. (2008). Bilgi Çađında Bir Yönetişel Dönüşüm Yaklaşımı: E-(Yerel) Yönetim. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*. 5.10, 1-32.
- Emel, G. G. ve Ç. Taşkın. (2005). Veri Madenciliđinde Karar Ağaçları Ve Bir Satış Analizi Uygulaması. *Eskişehir Osmangazi Üniversitesi sosyal Bilimler Dergisi*. 6.2, 221-239.
- Eren, V. (2003). Kamu Yönetiminde Rekabet, Rekabetin Kurumsallaştırılması ve Rekabet Mekanizmaları. *Ankara Üniversitesi SBF Dergisi*. 58.4, 83-105.
- Erkek, S. (2016). Kamu Kurumlarında Sosyal Medya Kullanımı: Sağlık Bakanlığı Örneđi. *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*. 35, 141-150.
- Eyüpođlu, C., M.A. Aydın, A. Sertbaş, A.H. Zaim ve O. Öneş. (2017). Büyük Veride Kişi Mahremiyetinin Korunması. *Bilişim Teknolojileri Dergisi*. 10.2, 177-184.
- Ertuđrul, İ., A. Organ ve A. Şavlı. (2013). Veri Madenciliđi Uygulamasına İlişkin PAÜ Hastanesinde Hasta Profilinin Belirlenmesi. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*. 19.2, 97-103.
- Fan, W., L. Wallace, S. Rich ve Z. Zhang. (2006). Tapping Into The Power Of Text Mining. *Communications of ACM*. 49.9, 76-82.
- Fayyad, U.M. (1996). Data Mining and Knowledge Discovery: Making Sense Out of Data, *IEEE Intelligent Systems*. 11.5, 20-25.
- Fayyad, U. M., G. Piatetsky-Shapiro ve P. Smyth. (1996). From Data Mining to Knowledge Discovery in Databases, *Artificial Intelligence Magazine*.17.3, 37-54.
- Fırat, O.Z. ve S.Ü. Fırat. (2017). Endüstri 4.0 Yolculuđunda Trendler ve Robotlar. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*. 46.2, 211-223.
- Giones, F. ve A. Brem. (2017). From Toys To Tools: The Co-Evolution Of Technological And Entrepreneurial Developments İn The Drone Industry, *Business Horizons*. 60, 875-884.

- Goh, Y.M. ve C.U. Ubeynarayana. (2017). Construction Accident Narrative Classification: An Evaluation Of Text Mining Techniques. *Accident Analysis and Prevention*. 108, 122-130.
- Göker, G. ve A. Doğan (2011). Ağ Toplumunda Örgütlenme: Facebook'ta Çevrimiçi Tekel Eylemi. *Balikesir University Journal of Social Sciences Institute*. 14.25, 175-203.
- Göküş, M. (2010). Küreselleşme Sürecinin Kamu Hizmetine Yansımaları. *Journal of Life Economics*. 10.20, 193-218.
- Guo, X. ve J. Lu. (2007). Intelligent e-Government Services with Personalized Recommendation Techniques. *International Journal of Intelligent Systems*. 22, 401-417.
- Hamet, P. ve J. Tremblay. (2017). Artificial Intelligence In Medicine. *Metabolism*. 69, 36-40.
- Hanumanthappa, M., B.R. Prakash ve M. Kumar. (2012). Applications of Data Mining in e-Governance: A Case Study of Bhoomi Project. *ICDEM 2010. LNCS 6411*. 208-218.
- Harrison, T. M., S. Guerrero, G. B. Burke, M. Cook, A. Cresswell, N. Helbig, J. Hrdinova ve T. Pardo. (2012). Applications of Data Mining in e-Governance: A Case Study of Bhoomi Project. *ICDEM 2010. LNCS 6411*. 208-218.
- He, Y., J. Liu ve L. Lin (2013). A Research On Application Of Data Mining in Cyber-Pursuit, *International Journal of Management Science and Engineering Management*, 6.2. 123-129.
- Henden, H.B. ve R. Henden. (2005). Yerel Yönetimlerin Hizmet Sunumlarındaki Değişim Ve E-Belediyecilik. *Elektronik Sosyal Bilimler Dergisi*. 4.14, 468-66.
- Henkoğlu, T. ve Ö. Külcü. (2013). Bilgi Erişim Platformu Olarak Bulut Bilişim: Riskler Ve Hukuksal Koşullar Üzerine Bir İnceleme. *Bilgi Dünyası*. 14.1, 62-86.
- Ho, A.T. (2002). Reinventing Local Governments And The E-government Initiative. *Public Administration Review*. 62.4, 434-444.
- Hood, C. (1991). A Public Management For All Seasons?. *Public Administration*. 69.1, 3-19.
- Irmak, S., C.D. Köksal ve Ö. Asilkan. (2012). Hastanelerin Gelecekteki Hasta Yoğunluklarının Veri Madenciliği Yöntemleri İle Tahmin Edilmesi. *Uluslararası Alanya İşletme Fakültesi Dergisi*. 4.1, 101-114.
- Ivanov, D., A. Dolgui, B. Sokolov, F. Werner ve M. Ivanova. (2016). A Dynamic Model And An Algorithm For Short-Term Supply Chain Scheduling In The Smart Factory Industry 4.0. *International Journal of Production Research*. 54.2, 386-402.

- Kalay, M. (2018). Database System Suggestions For The Internet Of Things (IoT) Systems. *Mugla Journal of Science and Technology*. 4.1, 46-52.
- Kalfa, C. ve F. Ataay. (2008). Yönetişim: Devlet-Toplum İlişkilerinde Yeni Bir Aşama. *Ç.Ü. Sosyal Bilimler Enstitüsü Dergisi*. 17.3, 229-240.
- Karacan, H. ve M. Yeşilbudak. (2010). Kullanıcı Merkezli İnteraktif Veri Madenciliği: Bir Literatür Taraması. *Bilişim Teknolojileri Dergisi*. 3.1, 17-22.
- Karagöz, K. (2013). Yeni Medya Çağında Dönüşen Toplumsal Hareketler Ve Dijital Aktivizm Hareketleri. *İletişim ve Diplomasi*, 1.1, 131-156.
- Karcı, Ş.M. (2008). Yeni Kamu İşletmeciliği Yaklaşımının Temel Değerleri Üzerine Bir İnceleme. *Akdeniz İ.İ.B.F. Dergisi*.16, 40-64.
- Kaşıkçı, T. ve H. Gökçen. (2014). Metin Madenciliği ile E-Ticaret Sitelerinin Belirlenmesi. *DBilişim Teknolojileri Dergisi*. 7.1, 25-32.
- Kaya, H. ve K. Köymen. (2008). Veri Madenciliği Kavramı ve Uygulama Alanları. *Doğu Anadolu Bölgesi Araştırmaları*. 159-164.
- Kayabaşı, Y. (2005). Sanal Gerçeklik Ve Eğitim Amaçlı Kullanılması. *The Turkish Online Journal of Educational Technology*. 4.3, 151-158.
- Kaymas, S. (2007). İnternet ve Ulusal Kamu Politikaları: İnternet Yönetiminde Türkiye için Alternatif Öneriler. *İletişim Araştırmaları Dergisi*. 97-135.
- Kaymas, S. (2012). İletişim Teknolojileri Politikaları Ve Türkiye: Küresel Medya Yönetimi Çağında Ulus Devlet. *Akademik Bakış Dergisi*. 28, 1-32.
- Kaplan, A. M. ve M. Haenlein. (2010). Users of the World, Unite! The Challenges and Opportunities of Social Media. *Business Horizons*. 53, 59-68.
- Keyvanpour, M.R., M. Javşdeh ve M.R. Ebrahimi. (2011). Detecting And Investigating Crime by Means of Data Mining: A General Crime Matching Framework. *Procedia Computer Science*. 3, 872-880.
- Kleineidam, G., M. Krasser ve M. Reichböck. (2016). The Cellular Approach: Smart Energy Region Wunsiedel. Testbed For Smart Grid, Smart Metering And Smart Home Solutions. *Electrical Engineering*. 98.4, 335-340.
- Koç, M. ve M. Karabatak. (2012). Sosyal Ağların Öğrenciler Üzerindeki Etkisinin Veri Madenciliği Kullanılarak İncelenmesi. *Nwsa-Education Sciences*. 7.1, 155-164.
- Korde, V. (2012). Text Classification and Classifiers: A Survey, *International Journal of Artificial Intelligence & Applications (IJAlA)*. 3.2, 85-99.
- Köktaş, Ö.F. ve Ö. Köseoğlu. (2015). Kanıta Dayalı Kamu Politikası Yapımı: Sosyal Bilim Araştırması ve Kamu Politikaları İlişkisini Yeniden Dizayn Etmek için Bir Fırsat Mı?. *Yasama Dergisi*. 29, 32-57.

- Köktürk, F., H. Ankaralı ve V. Sümbüloğlu. (2009). Veri Madenciliği Yöntemlerine Genel Bakış. *Türkiye Klinikleri J Biostat.* 1.1, 20-25.
- Köseoğlu, Ö. ve A. Tuncer. (2014). Kamu Yönetiminde Yeni Bir Yaklaşım Olarak Kamu Değeri: Kavramsal ve Kuramsal Açından Bir Değerlendirme. *Yönetim Bilimleri Dergisi*, 12.24, 145-170.
- Köseoğlu, Ö. ve Y. Demirci. (2017). Türkiye’de Büyük Veri Ve Veri Madenciliğine İlişkin Politika Ve Stratejiler: Ulusal Politika Belgelerinin İçerik Analizi. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*. 22. Kayfor15 Özel Sayısı, 2223-2239.
- Ku, C.H. ve G. Leroy. (2014). A Decision Support System: Automated Crime Report Analysis and Classification for E-government. *Government Information Quarterly*. 534-544.
- Li, W. (1992). Random Texts Exhibit Zipf's-Law-Like Word Frequency Distribution. *IEEE Transactions on Information*. 38.6, 1842-1845.
- Longo, J. (2011). #Opendata: Digital-Era Governance Thoroughbred or New Public Management Trojan Horse?. *Public Policy&Governance Review*. 2.2, 38-51.
- Maciejewski, M. (2017). To Do More, Better, Faster And More Cheaply: Using Big Data in Public Administration. *International Review of Administrative Sciences*, 83.1, 120-135.
- Macit, İ. (2017). Kurumsal Kaynak Planlamasının Endüstri 4.0 Kazanımları: Bir Yapısal Çatı Modeli Önerisi. *Yönetim Bilişim Sistemleri Dergisi*, 2.1, 50-60.
- Mell, P. ve T. Grance. (2011). The NIST Definition Of Cloud Computing. *U.S. Department of Commerce*. 800.145, 1-3.
- Moniruzzaman, A.B.M. ve S. A. Hossain. (2013). NoSQL Database: New Era of Databases for Big Data Analytics - Classification, Characteristics and Comparison. *International Journal of Database Theory and Application*. 6.4, 1-13.
- Oğuzlar, A. (2003). Veri Ön İşleme. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*. 21, 67-76.
- Ozyirmidokuz, E.K., K. Uyar ve M.H. Ozyirmidokuz. (2015). A Data Mining Approach to A Firm’s Marketing Channel. *Procedia Economics and Finance*. 27, 77-84.
- Özdemir, A., F. Y. Aslay ve H. Çam. (2010). Veri Tabanında Bilgi Keşfi Süreci: Gümüşhane Devlet Hastanesi Uygulaması. *Sosyal Ekonomik Araştırmalar Dergisi*. 10.20, 347-366.
- Özekes, S. (2003). Veri Madenciliği Modelleri Ve Uygulama Alanları. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*. 2.3, 65-82.

- Özer, M.A. (2012). Kamu Yönetimi Disiplininin Geleceği Üzerine Tartışmalar. *International Journal of Economic and Administrative Studies*. 5.9, 1-20.
- Özsoylu, A.F. (2017). Endüstri 4.0. *Çukurova Üniversitesi İİBF Dergisi*. 21.1, 41-64.
- Palmarini, R., J. A. Erkoyuncu, R. Roy ve H. Torabmostaedi. (2018). A Systematic Review Of Augmented Reality Applications In Maintenance. *Robotics and Computer-Integrated Manufacturing*. 49, 215-228.
- Paşayığıt, G. (2012). Biçimbilimsel Çözümleme. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*. 5.2, 1-23.
- Patil, T., ve S.S. Sherekar. (2013). Performance Analysis Of Naive Bayes And J48 Classification Algorithm For Data Classification. *Journal Of Computer Science and Applications*. 6.2, 256-261.
- Prakash, M., ve G. Singaravel. (2015). An Approach For Prevention Of Privacy Breach and Information Leakage In Sensitive Data Mining. *Computers and Electrical Engineering*. 45, 134-140.
- Raja, V. ve P. Calvo. (2017). Augmented Reality: An Ecological Blend. *Cognitive Systems Research*. 42, 58-72.
- Ralha, C.G. ve C.V.S. Silva. (2012). A Multi-Agent Data Mining System For Cartel Detection In Brazilian Government Procurement. *Expert Systems with Applications*. 39, 11642-11656.
- Ramkumar, G.D. ve A. Swami (1998). Clustering Data Without Distance Functions. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*. 21.1, 1-5.
- Rhodes, R. (1996). The New Governance: Governing Without Government. *Political Studies*. 44, 652-657.
- Rygielski, C., J.C. Wang ve D.C. Yen. (2002). Data Mining Techniques For Customer Relationship Management. *Technology in Society*. 24, 483-502.
- Salton, G., A. Wong ve C.S. Yang. (1975). A Vector Space Model for Automatic Indexing. *Communications of the ACM*. 18.11, 613-620.
- Savaş, S., N. Topaloğlu ve M. Yılmaz. (2012). Veri Madenciliği Ve Türkiye'deki Uygulama Örnekleri. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*. 11.21, 1-23.
- Sener, S. ve B. Eevli. (2017). Endüstri 4.0'da Yeni İş Kolları ve Yüksek Öğrenim. *Mühendis Beyinler Dergisi*. 1.2, 25-37.
- Senyucel. Z. (2007). Assessing the Impact of e-Government on Providers and Users of the IS Function: A Structuration Perspective. *Transforming Government: People Process and Policy*. 1.2, 131-144.

- Sevindik, T., K. Kayışlı ve O. Ünlükahraman. (2012). Web Tabanlı Eğitimde Veri Madenciliği. *Turkish Journal of Computer and Mathematics Education*. 3.3, 183-193.
- Shearer, C. (2000). The CRISP-DM Model: The New Blueprint for Data Mining. *Journal of Data Warehousing*. 5.4, 13-22.
- Sun, S., C. Luo ve J. Chen. (2017). A Review Of Natural Language Processing Techniques For Opinion Mining Systems. *Information Fusion*. 36, 10-25.
- Tan, A.H. ve P.S. Yu. (2003). Guest Editorial: Text and Web Mining. *Applied Intelligence*. 18, 239-241.
- Thompson, J. (2008). Don't Be Afraid to Explore Web 2.0. *Phi Delta Kappa International*. 89.10, 711-718.
- Tomak, L., ve Y. Bek. (2010). İşlem Karakteristik Eğrisi Analizi Ve Eğri Altında Kalan Alanların Karşılaştırılması. *Journal of Experimental and Clinical Medicine*. 27.2, 58-65.
- Tuncer, A. ve S. Usta. (2013). İki Kriz Arasında Yönetim: Yeni Kamu İşlemciliği. *Journal of Institute of Social Sciences*. C:30, 181-195.
- Tüzüntürk, S. (2010). Veri Madenciliği ve İstatistik. *Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*., 65-90.
- Usama, F. ve S. Paul. (1997). Data Mininig and KDD: Promise and Challenges. *Future Generation Computer Systems*. 13, 99-115.
- Uguz, H. (2011). A Two-Stage Feature Selection Method For Text Categorization By Using Information Gain, Principal Component Analysis And Genetic Algorithm. *Knowledge-Based Systems*. 24, 1024-1032.
- Wu, X., V. Kumar, J.R Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A.Ng, B. Liu P. S. Yu, Z.Zhou, M. Steinbach, D. J. Hand ve D. Steinberg. (2008). Top 10 Algorithms In Data Mining. *Knowledge and Information Systems*. 14.1, 1-37.
- Vanolo, A. (2014). Smartmentality: The Smart City as Disciplinary Strategy. *Urban Studies*. 51.5, 883-898.
- Viera, A.J., ve J.M. Garret. (2005). Understanding Interobserver Agreement: The Kappa Statistic. *Family Medicine*. 37.5, 360-363.
- Vural, Z., B.Akıncı ve M. Bat. (2010). Yeni Bir İletişim Ortamı Olarak Sosyal Medya: Ege Üniversitesi İletişim Fakültesine Yönelik Bir Araştırma. *Journal of Yasar University*, 20.5, 3348-3382.
- Yağmurlu, A. (2011). Kamu Yönetiminde Halkla İlişkiler Ve Sosyal Medya. *Selçuk Üniversitesi İletişim Fakültesi Akademik Dergisi*. 7.1, 5-15.

- Yanmaz, E., S. Yahyanejad, B. Rinner, H. Hellwagner ve C. Bettstetter. (2018). Drone Networks: Communications, Coordination, And Sensing, *Ad Hoc Networks*. 68, 1-15.
- Yaşamış, F.D. (1991). Hızlı Kentleşmenin Sonuçları ve Belediyelerin Kurumsal, Örgütsel ve Yönetimsel Olanakları. *Amme İdaresi Dergisi*. 24.3, 163-188.
- Zanella, A., N. Bui, A. Castellani, L. Vangelista, ve M. Zorzi. (2014). Internet of Things for Smart Cities. *IEEE Internet Of Things Journal*. 1.1, 22-32.
- Zuiderwijk , A., M. Janssen ve C.Davis. (2014). Innovation With Open Data: Essential Elements Of Open Data Ecosystems. *Information Polity*. 19, 17-33.



## Diğer Yayınlar

- Adalet ve Kalkınma Partisi, AKP. (2018). 2018 Seçim Beyannamesi. Ankara.
- Agrawal, R., T. Imielinski ve A. Swami. (Mayıs 1993). Mining Association Rules Between Sets of Items in Large Databases. *Proceedings of the 1993 ACM SIGMOD Conference*, Washington, 1-10.
- Aksu, H. (2014). Efficient Analysis Of Large-Scale Social Networks Using Big-Data Platforms, *Doktora Tezi*. Ankara: Bilkent Üniversitesi FBE.
- Aktaş, M.S. ve O. Kalıpsız. (23 Ekim 2015). Veri Madenciliğinde Özellik Seçim Tekniklerinin Bankacılık Verisine Uygulanması Üzerine Araştırma ve Karşılaştırmalı Uygulama. *2015 Turkish National Software Engineering Symposium*, Yaşar Üniversitesi: İzmir, 72-83.
- Alguliyev, R. ve Y. Imamverdiyev. (15-17 Ekim 2014). Big Data: Big Promises for Information Security. *2014 IEEE 8th International Conference on Application of Information and Communication Technologies (AICT)*, Astana, 1-4.
- Amasyalı, M.F. ve T. Yıldırım. (30-31 Nisan 2004). Otomatik Haber Metinleri Sınıflandırma. *Signal Processing and Communications Applications Conference*. Kuşadası, 224-226.
- Amasyalı, M.F., F. Davletov, A. I. Torayew ve Ü. Çiftçi. (22-24 Nisan 2010). Text2arff: Automatic Feature Extraction Software For Turkish Texts. *2010 IEEE 18th Signal Processing and Communications Applications Conference*. Diyarbakır, 629-632.
- Aydın, S. (2007). Veri Madenciliği Ve Anadolu Üniversitesi Uzaktan Eğitim Sisteminde Bir Uygulama, *Doktora Tezi*. Eskişehir: Anadolu Üniversitesi SBE.
- Azevedo, A. ve M.F. Santos. (2008). KDD, SEMMA and CRISP-DM: A Parallel Overview, *IADS-DM*, 1-6.
- Aziza, B. (2018). Facebook Privacy Scandal Hearings: What You Missed. *Forbes*. <https://www.forbes.com/sites/ciocentral/2018/04/16/facebook-privacy-scandal-hearings-what-you-missed/#4a18d9287ab9> (16 Nisan 2018)
- Başbakanlık. (2002). E-Türkiye Girişimi Eylem Planı. Ankara
- Başbakanlık. (2008). Kamu Kurum ve Kuruluşlarının Büyük Ölçekli Bilgi İşlem Birimlerinde Sözleşmeli Bilişim Personeli İstihdamına İlişkin Esas ve Usuller Hakkında Yönetmelik, <http://www.resmigazete.gov.tr/eskiler/2008/12/20081231-14.htm>.(29 Nisan 2018) Türkiye.
- Başbakanlık. (2016). 65. Hükümet Programı, <http://reformlar.gov.tr/> (25 Mayıs 2016) Türkiye.

- Başbakanlık. (2016). Saydamlığın Artırılması ve Yolsuzlukla Mücadelenin Güçlendirilmesi Eylem Planı, [www.resmigazete.gov.tr/eskiler/2016/04/20160430-13.pdf](http://www.resmigazete.gov.tr/eskiler/2016/04/20160430-13.pdf) (16 Temmuz 2018) Türkiye.
- Belediye Kanunu. (2003). Kanun Numarası: 5393, Kabul Tarihi: 03/07/2005.
- Biricik, G. (2011). Metin Sınıflama İçin Yeni Bir Özellik Çıkarım Yöntemi. *Doktora Tezi*. İstanbul: Yıldız Teknik Üniversitesi FBE.
- Birleşmiş Milletler. (2017). *BM E-Devlet Endeksleri*, <https://publicadministration.un.org/egovkb/en-us/Data/Country-Information/id/176-Turkey> (01 Eylül 2017).
- Birleşmiş Milletler. (2017). *Frequently Asked Questions, What is E-Government*. <https://publicadministration.un.org/egovkb/en-us/About/UNeGovDD-Framework#whatis> (21 Ağustos 2017).
- Boran, L. (2012). Veri Madenciliğinin Türk İşletmelerin Finansal Tablolarına Uygulanması Ve Uygulama Örneği. *Doktora Tezi*. İstanbul: Marmara Üniversitesi SBE.
- Brohman, M.K. (04-07 Ocak 2006). Knowledge Creation Opportunities in the Data Mining Process. *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*. Hawaii, 1-10.
- Bulut, F. (26-27 Nisan 2016). Otomatik Haber Metinleri Sınıflandırma. *2016 Electric Electronics, Computer Science, Biomedical Engineerings Meeting (EBBT)*. İstanbul, 1-4.
- Bükey, F.Ö. (2014). Data Mining Applications In Customer Relationship Management and A Comparative Study In The Banking Sector. *Doktora Tezi*, İstanbul: Marmara Üniversitesi FBE.
- Büyükşehir Belediyesi Kanunu. (5216). Kanun Numarası: 5393, Kabul Tarihi: 10/07/2004.
- Can, E. ve F. Amasyalı. (2016). Text2arff: A Text Representation Library, *24th Signal Processing and Communication Application Conference (SIU)*. Zonguldak, 197-200.
- Can, M. B., E. Çamur., M. Kuru, Ö. Özkan ve Z. Rzaeva. (2012). Veri Kümelerinden Bilgi Keşfi: Veri Madenciliği. *Başkent Üniversitesi Tıp Fakültesi XIV. Öğrenci Sempozyumu*. Ankara, 1-14.
- Canpolat, Ö. (2001). E-Ticaret ve Türkiye'deki Gelişmeler. T.C. Sanayi Ve Ticaret Bakanlığı Hukuk Müşavirliği. Ankara.
- Cavnar, W. B. ve J. M. Trenkle (11-13 Nisan 1994). N-Gram-Based Text Categorization. *In Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval*. Las Vegas, 161-175.

- Chapman, P., J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer ve R. Wirth. (2000). *CRISP-DM 1.0 Step-by-step Data Mining Guide*. SPSS Inc.
- Cline, G. (2017). Industry 4.0 and Industrial IoT in Manufacturing: A Sneak Peek. <https://www.aberdeen.com/opspro-essentials/industry-4-0-industrial-iot-manufacturing-sneak-peek/> (03 Mart 2018).
- Cormode, G. ve B. Krishnamurthy. (2008). Key Difference Between Web 1.0 and Web 2.0. *First Monday*. 13.6. <http://www.ojphi.org/ojs/index.php/fm/article/view/2125/1972> (13 Mart 2018).
- Cumhuriyet Halk Partisi. (2018). 2018 Seçim Bildirgesi. Ankara.
- Çakır, Ö. (2008). Veri Madenciliğinde Sınıflandırma Yöntemlerinin Karşılaştırılması Bankacılık Müşteri Veri Tabanı Üzerinde Bir Uygulama. *Doktora Tezi*. İstanbul: Marmara Üniversitesi SBE.
- Çankırı, S., E. Kartal, K. Yıldırım ve S. Gülseçen. (01-02 Ekim 2009). Organizasyonlarda Bilgi Yönetimi Sürecinde Veri Madenciliği Yaklaşımı. *Bilgi Çağında Varoluş: "Fırsatlar ve Tehditler" Sempozyumu*. İstanbul: Yeditepe Üniversitesi, 148-168.
- Çelik, S. (2018). Büyük Veri ve İstatisteki Uygulamaları, *Doktora Tezi*. Bursa: Uludağ Üniversitesi SBE.
- Çetin, M. ve M.F. Amasyalı. (24-26 Nisan 2013). Eğiticili ve Geleneksel Terim Ağırlıklandırma Yöntemleriyle Duygu Analizi. *21st Signal Processing and Communications Applications Conference (SIU)*. Haspolat, 1-4.
- Dalgaldere, S. (2016). Epistemolojik Açından Büyük Veri Ve Gelecek Tahmin Sistemleri, *Doktora Tezi*. İstanbul: Marmara Üniversitesi SBE.
- Deloitte. (2016). *Akıllı Şehir Yol Haritası*, Türkiye.
- Dener, M., M. Dörtlerler ve A. Orman. (11-13 Şubat 2009). Açık Kaynak Kodlu Veri Madenciliği Programları: Weka'da Örnek Uygulama. *Akademik Bilişim '09 - XI. Akademik Bilişim Konferansı Bildirileri*. Şanlıurfa: Harran Üniversitesi, 787-796.
- Demir, M. (2017). Yenilikçi Akıllı Şehir Uygulamalarında Öncüyüz. *Fortune Dergisi*. <http://www.fortuneturkey.com/yol-acin-akilli-sehirler-geliyor-45878> (29 Haziran 2018)
- Dijital Türkiye Platformu. (2014). Avrupa Dijital Gündemi ve Türkiye: Avrupa Dijital Gündemine Uyum Projesi Ön Raporu. İstanbul.
- Doğan, M. E. (02-04 Şubat 2011). Enformasyonel İş Gücü İçin Yeni Bir Öğrenme Yaklaşımı: Bağlantıcılık. *Akademik Bilişim*. Malatya: İnönü Üniversitesi, 171-177.

- Doğan, O., H. Büyükkacı, A. Darılmaz, E. Kara ve K. Çağiltay. (20-21 Aralık 2017). FeTeMM Eğitiminde Giyilebilir Meteoroloji İstasyonu – GiyMİ. *TBD 34. Ulusal Bilişim Kurultayı*. Ankara, 25-27.
- Dolgun M.Ö. (2014). Veri Madenciliği Sınıflama Yöntemlerinin Başarılarının; Bağımlı Değişken Prevelansı, Örneklem Büyüklüğü ve Bağımsız Değişkenler Arası İlişki Yapısına Göre Karşılaştırılması. *Doktora Tezi*. Ankara: Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü.
- Dunleavy, P. ve H. Margetts. (2-5 Eylül 2010). The Second Wave Of Digital Era Governance. *American Political Science Association Conference*. Washington, 1-32.
- E-Dönüşüm Türkiye Projesi*. (2017), <http://www.bilgitoplumu.gov.tr/bilgi-toplumu/e-donusum-projesi/> (01 Eylül 2017).
- EMC. (2014). *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things*, <https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm> (07 Mayıs 2014)
- Facebook Research. (2018). *Advancing The Field Of Machine Intelligence*, <https://research.fb.com/category/facebook-ai-research/> (20 Ağustos 2018)
- Geman, O., I. Chiuchisan, ve M. Covasa. (20-21 Ekim 2016). Data Mining and Knowledge Discovery Tools for Human Microbiome Big Data. *2016 6th International Conference on Computers Communications and Control (ICCC)EEE 18th Signal Processing and Communications Applications Conference*. Mashhad: Ferdowsi University, 91-96.
- Gürbüz, F. (2009). Havacılık Sektöründe Veri Madenciliği İle Farklı Sınıflandırma Tekniklerinin Karşılaştırmalı Olarak Uygulanması. *Doktora Tezi*. Kayseri: Erciyes Üniversitesi FBE.
- Halkbank. (2013). *Faaliyet Raporu*, Türkiye.
- İçişleri Bakanlığı. (2015). 2015-2019 Türkiye Cumhuriyeti Ulaştırma Bakanlığı Stratejik Planı. Ankara.
- İnce, N.M. (2001). Kamu Hizmetlerinin Sunulmasında Yeni İmkanlar. *Elektronik Devlet*. Ankara: Devlet Planlama Teşkilatı.
- Kalkınma Bakanlığı. (2010). *2011 Yılı Programı*. Ankara.
- Kalkınma Bakanlığı. (2013). *Kamu Yatırım Programı*. Ankara.
- Kalkınma Bakanlığı. (2013). Onuncu Kalkınma Planı (2014-2018). Ankara.
- Kalkınma Bakanlığı. (2006). 2006-2010 Bilgi Toplumu Stratejisi ve Eylem Planı. Ankara.

- Kalkınma Bakanlığı. (2014). 2015-2018 Bilgi Toplumu Stratejisi ve Eylem Planı. Ankara.
- TÜBİTAK. (2018). Kamu İnternet Siteleri Rehberi. Ankara.
- Kalkınma Bakanlığı. (2018). Kamu İdareleri İçin Stratejik Planlama Kılavuzu (3.Sürüm). Ankara.
- Kamu Mali Yönetimi ve Kontrol Kanunu. (2003). Kanun Numarası: 5018, Kabul Tarihi: 24/12/2013.
- Karimi, O. ve A.Korkmaz. (9-11 Aralık 2013). Kişisel Verilerin Korunması, *XVIII. Türkiye'de İnternet Konferansı*. İstanbul, 193-199.
- Kaya, M. ve S.A. Özel. (05-07 Şubat 2014). Açık Kaynak Kodlu Veri Madenciliği Yazılımlarının Karşılaştırılması. *Akademik Bilişim*. İçel: Mersin Üniversitesi, 1-8.
- Kesgin, F. (2007). Türkçe Metinler İçin Konu Belirleme Sistemi. *Yüksek Lisans Tezi*. İstanbul Teknik Üniversitesi FBE.
- Khan, Z., A. Anjum ve S. L Kiani. (9-12 Aralık 2013). Cloud Based Big Data Analytics For Smart Future Cities. *2013 IEEE/ACM 6th International Conference on Utility and Cloud Computing*. Dresden, 381- 386.
- Kleissner, C. (January 1998). Data Mining For Enterprise. *31st Annual Hawaii International Conference on System Screens*. Hawai, 295–304.
- Kuru, E. (2017). İstanbul Küresel Akıllı Şehir Olma Hedefinde. *Fortune Dergisi*. <http://www.fortuneturkey.com/yol-acin-akilli-sehirler-geliyor-45878> (29 Haziran 2018)
- Kuzey, C. (2012). Veri Madenciliğinde Destek Vektör Makinaları ve Karar Ağaçları Yöntemlerini Kullanarak Bilgi Çalışanlarının Kurum Performansı Üzerine Etkisinin Ölçülmesi ve Bir Uygulama. *Doktora Tezi*, İstanbul: İstanbul Üniversitesi SBE.
- Küçüksille, E. (2009). Veri Madenciliği Süreci Kullanılarak Portföy Performansının Değerlendirilmesi ve İMKB Hisse Senetleri Piyasasında Bir Uygulama. *Doktora Tezi*, Isparta: Süleyman Demirel Üniversitesi FBE.
- Kesayak, B. (2017). *Akıllı Şehir İçin Veri Madenciliği Kullanılacak*. <https://www.techinside.com/akilli-sehir-icin-veri-madenciligi-kullanilacak/> (08 Haziran 2018).
- Kişisel Verilerin Korunması Kanunu. (2016). Kanun Numarası: 6698, Kabul Tarihi: 24/03/2016.
- Koyuncu, E. (2016). Kalkınma İçin Büyük Veri. *Türkiye Ekonomi Politikaları Araştırma Vakfı*. Ankara, 1-7.

- Koyuncugil, A.S. (2007). Ülkelerin Gelir Düzeyleri İle Altyapı Yatırımları Arasındaki İlişkinin Veri Madenciliği İle Araştırılması. *Sermaye Piyasası Kurulu Araştırma Raporu*. Ankara,1-27.
- Kutup, N. (30 Kasım - 2 Aralık 2011). Nesnelerin İnterneti; 4H Her Yerden, Herkesle, Her zaman, Her Nesne İle Bağlantı. *16. Türkiye'de İnternet Konferansı*. İzmir, 151-156.
- Liu, Q., Y. Wei, S. Leng ve Y. Chen. (27-30 Ekim 2017). Task Scheduling in Fog Enabled Internet of Things for Smart Cities. *17th IEEE International Conference on Communication Technology*. Chengdu, 975-980.
- Liu, M. ve J. Yang. (18-19 Ağustos 2012). An Improvement Of TFIDF Weighting in Text Categorization. *International Conference on Computer Technology and Science (ICCTS 2012)*. New Delhi, 44-47.
- Microsoft. (2017). *Microsoft's AI Vision, Rooted In Research, Conversations*, <https://news.microsoft.com/features/microsofts-ai-vision-rooted-in-research-conversations/> (20 Ağustos 2018)
- Moskwa, W. (9 Mayıs 2016). World Drone Market Seen Nearing \$127 Billion in 2020. PwC Says. *Bloomberg*. <https://www.moneyweb.co.za/news/tech/world-drone-market-seen-nearing-127bn-2020-pwc-says/> (08 Mart 2018).
- Nadali, A., E. N. Kakhky ve H.E. Nosratabadi. (8-10 Nisan 2011). Evaluating The Success Level Of Data Mining Projects Based On Crisp-Dmmethodology By A Fuzzy Expert System. *3rd International Conference on Electronics Computer Technology*. Kanyakumari, 161-165.
- On Dört İlde Büyükşehir Belediyesi Ve Yirmi Yedi İlçe Kurulması İle Bazı Kanun Ve Kanun Hükmünde Kararnamelerde Değişiklik Yapılmasına Dair Kanun (2012). Kanun Numarası: 6360, Kabul Tarihi: 12/11/2012.
- Open Government Partnership. (2011). *Turkey Action Plan 2012-2013*, [www.opengovpartnership.org/documents/turkey-action-plan-2012-2013](http://www.opengovpartnership.org/documents/turkey-action-plan-2012-2013) (16 Temmuz 2018).
- Open Government Partnership. (2014). *Official Communication from the Open Government Partnership Support Unit*, <https://www.opengovpartnership.org/sites/default/files/attachments/OGP%20Letter%20-%20Turkey.pdf> (16 Temmuz 2018).
- Open Government Partnership. (2016a). *Resolution of the OGP Steering Committee Regarding the Status of the Government of Turkey in OGP*, <http://www.opengovpartnership.org/sites/default/files/Revised%20Turkey%20Resolution%20for%20SC%20May%204.pdf> (16 Temmuz 2018).
- Open Government Partnership. (2016b). *Press Release: Turkey made inactive in the Open Government Partnership*, [www.opengovpartnership.org/about/news-and-events/turkey-made-inactive-open-government-partnership](http://www.opengovpartnership.org/about/news-and-events/turkey-made-inactive-open-government-partnership) (16 Temmuz 2018).

- Open Government Partnership. (2017). *Official Communication from the Open Government Partnership Support Unit - Letter informing ending of Turkey's participation in OGP*, <https://www.opengovpartnership.org/sites/default/files/Turkey-OGP-Letter-Government-Turkey-Sept2017.pdf> (16 Temmuz 2018).
- Open Knowledge International. (2018). *Open Data Handbook*, London.
- Open Knowledge International. (2018). *Open Data Handbook*, <http://opendatahandbook.org/guide/en/what-is-open-data/> (06 Haziran 2018)
- Pehlivan, E. (2017). Katılımcı, Sürdürülebilir Bir Akıllı Şehir Hedefliyoruz. *Fortune Dergisi*. <http://www.fortuneturkey.com/yol-acin-akilli-sehirler-geliyor-45878> (29 Haziran 2018).
- Peng, L. ve V. Gulshan. (2016). Deep Learning for Detection of Diabetic Eye Disease. *Google AI Blog*. <https://ai.googleblog.com/2016/11/deep-learning-for-detection-of-diabetic.html> (29 Kasım 2016).
- Pichai, S. (2018). AI At Google: Our Principles. <https://www.blog.google/technology/ai/ai-principles/> (7 Haziran 2018).
- Razboyanlı, C. (2016). Big Data (Büyük Veri) Ve Geleneksel Veri Saklama Ve İşleme Yöntemlerine Etkisi Üzerine Bir Araştırma, *Doktora Tezi*. İstanbul: Okan Üniversitesi FBE.
- SAS. (2018). *Data Mining Using SAS(R) Enterprise Miner(TM): A Case Study Approach, Third Edition*. <http://support.sas.com/documentation/cdl/en/emcs/66392/HTML/default/viewer.htm#n0pejm83csbja4n1xueveo2uoujy.htm/> (28 Mayıs 2018)
- Sayad, S. (2010). An Introduction to Data Science. [http://www.saedsayad.com/data\\_mining.htm](http://www.saedsayad.com/data_mining.htm) (22 Aralık 2016).
- Scherer, M. (2012). How Obama's Data Crunchers Helped Him Win. <https://edition.cnn.com/2012/11/07/tech/web/obama-campaign-tech-team/index.html> (20 Ağustos 2018).
- Sevinçhan, Y. (2016). Kişisel Verilerin Korunması Yasal Düzenlemesi. *Genç Hukukçular Okuma Grubu Sunum Raporu*. 1-5.
- Sönmez, S. (29-31 Mayıs 2002). Küreselleşme Ortamında Kamu İşletmeciliği İçin Bir Strateji Oluşturulabilir Mi?. *Türkiye 13. Kömür Kongresi Bildiriler Kitabı*. Zonguldak, 395-424.
- SPSS. (2015). *Kamu Ve Savunma Sektöründe SPSS Başarının Anahtarı Karar mı? Kader mi? Bilgiye Dayalı Yönetimde Son Trendler*, [https://sbu.saglik.gov.tr/Ekutuphane/kitaplar/biyoistatistik%20\(6\).pdf](https://sbu.saglik.gov.tr/Ekutuphane/kitaplar/biyoistatistik%20(6).pdf) (05.03.2018)
- Strategic Policy Making Team Cabinet Office. (1999). *Professional Policy Making For The Twenty First Century Report*. <https://ntouk.files.wordpress.com>

- /2015/06/professional-policy-making-for-the-21st-century-1999.pdf (19 Eylül 2017).
- Şeker, S.E. (2 Haziran 2015). Sosyal Ağlarda Veri Madenciliği. *YBS Ansiklopedi*. [www.ybsansiklopedi.com](http://www.ybsansiklopedi.com) (02 Mayıs 2016).
- Şener, M. ve A. Paşayığıt. (2006). E-Devlette Kalite, Güvenlik ve Kişisel Gizlilik. 23. *İTÜ EMÖS Proje Yarışması*. 1-8
- Tilley, A. (2017). Why Apple Joined Rivals Amazon, Google, Microsoft In AI Partnership. *Forbes*. <https://forbes.com/sites/aarontilley/2017/01/27/why-apple-joined-rivals-amazon-google-microsoft-in-ai-partnership/#6ef367cc5832> (8 Mart 2018).
- Tekerek, A. (02-04 Şubat 2011). Veri Madenciliği Süreçleri Ve Açık Kaynak Kodlu Veri Madenciliği Araçları. *Akademik Bilişim*. Malatya: İnönü Üniversitesi, 2-4.
- TUSEV. (2015). Açık Yönetim Ortaklığı Ve Türkiye Süreci Vaka Analizi Sivil Toplum İzleme Raporu 2015-2016. İstanbul.
- Tüm hükümet programları; <https://www.akparti.org.tr/site/dosya/66517>, 01.09.2017.
- Türkiye Bilişim Derneği. (2016). *Büyük Veri Uygulamaları Çalışma Grubu 4*, [www.kamu-bib.org.tr/CG4-Buyuk-Veri-Uygulamalari-2016.pdf](http://www.kamu-bib.org.tr/CG4-Buyuk-Veri-Uygulamalari-2016.pdf) (13 Temmuz 2018).
- Ulaştırma Denizcilik ve Haberleşme Bakanlığı. (2016). *2016-2019 Ulusal e-Devlet Stratejisi ve Eylem Planı*, Türkiye.
- Ulaştırma Denizcilik ve Haberleşme Bakanlığı (1999). Türkiye Ulusal Enformasyon Altyapısı Anaplanı (TUENA) Sonuç Raporu. Ankara.
- Ulaştırma Denizcilik ve Haberleşme Bakanlığı. (2013). 2013-2017 Türkiye Cumhuriyeti Ulaştırma Bakanlığı Stratejik Planı. Ankara.
- Ulaştırma Denizcilik ve Haberleşme Bakanlığı. (2016). 2016-2019 Ulusal e-Devlet Stratejisi ve Eylem Planı. Ankara.
- Ulaştırma Denizcilik ve Haberleşme Bakanlığı. (2017). 2017-2021 Türkiye Cumhuriyeti Ulaştırma Bakanlığı Stratejik Planı. Ankara.
- Ürey, H., ve . Akşit. *Sanal Gerçeklik ve Artırılmış Gerçeklik Nedir?*, <https://sarkac.org/2018/03/sanal-gerceklik-artirilmis-gerceklik-nedir/> (10 Eylül 2018).
- Veri Koruma Direktifi. (1995). Kanun Numarası: 95/46/AT.
- Victor, D. (2016). *Microsoft Created a Twitter Bot to Learn From Users. It Quickly Became a Racist Jerk..* <https://www.nytimes.com/2016/03/25/technology/>

[microsoft-created-a-twitter-bot-to-learn-from-users-it-quickly-became-a-racist-jerk.html](https://www.microsoft.com/en-us/news/story/microsoft-created-a-twitter-bot-to-learn-from-users-it-quickly-became-a-racist-jerk.html) (24 Mart 2016)

Yağcı, O. (2018) *Artırılmış Gerçeklik ile Sanal Gerçeklik Arasındaki Farklar*, <http://www.elektrikport.com/teknik-kutuphane/artirilmis-gerceklik-ile-sanal-gerceklik-arasindaki-farklar/18557#ad-image-0> (10 Eylül 2018).

Wikipedia. (2018). *Cloud Computing*, [https://en.wikipedia.org/wiki/Cloud\\_computing](https://en.wikipedia.org/wiki/Cloud_computing) (03 Mart 2018).

Wirth, R. ve J. Hipp. (27 Ağustos 1998). CRISP-DM: Towards A Standard Process Model For Data Mining. *Proceedings Of The 4th International Conference On The Practical Applications Of Knowledge Discovery And Data Mining*. New York, 29-39.

Weiss, C. H. (Haziran 2001). What Kind of Evidence in Evidence-Based Policy?. *Third International, Inter-Disciplinary Evidence-Based Policies and Indicator Systems Conference*. Durham, 284-291.

Yılmaz, S. (2009). Türkçe İçin İyileştirilmiş Biçimbirimsel Çözümleyici. *Yüksek Lisans Tezi*, İstanbul: İstanbul Teknik Üniversitesi FBE.

Yılmaz, Ş. (22 Eylül 2005). Bilgi Yönetimine Giden Yol: E-Türkiye Sürecinde Elektronik Belge Yönetiminin Önemi ve Mevcut Durum Çalışmaları, *ÜNAK'05: Bilgi Hizmetlerinin Organizasyonu ve Pazarlanması*. İstanbul, 313-315.

Yıldız, H.K., M. Gençtav, N. Usta, B. Diri ve M.F. Amasyalı. (11-13 Haziran 2007). Metin Sınıflandırmada Yeni Özellik Çıkarımı. *IEEE 15th Signal Processing and Communication Applications Conference*, Eskişehir, 1-4.

Zaitseva, E, M. Kvassay, V. Levashenko ve J. Kostolny (13-16 Eylül 2015). Introduction to Knowledge Discovery in Medical Databases and Use of Reliability Analysis in Data Mining, *Proceedings of the Federated Conference on Computer Science and Information Systems*, Lodz, 311-320.

## EKLER

### **Kamuda Veri Madenciliği**

Veri Madenciliği, büyük veri yığınları içerisinde gelecekle ilgili tahminde bulunabilmemizi sağlayabilecek bağlantıların bilgisayar programı kullanılarak aranmasıdır.

Çalışmanın konusu, kamu yönetiminde son yıllarda giderek stratejik bir önem kazanan ve sağlıktan eğitime, çevre sorunlarından trafığa kadar birçok kamusal hizmete yönelik politikalar üretilmesine imkân sağlayan veri madenciliğidir.

Çalışmanın amacı; Veri madenciliğinin çözüm masası verilerine uygulanması suretiyle kamu yönetimi kurumları özelinde yerel yönetimlerin politika ve çözüm üretimine katkı sağlayabileceğine yönelik bir model önerisi sunmaktır.

Bu anketten elde edilecek sonuçlar tamamen bilimsel amaçlı kullanılacak olup sorulara olabildiğince ayrıntılı bir şekilde cevap verilmesi akademik anlamda büyük katkı sağlayacaktır.

Değerli katılımınız için teşekkür ederiz.

### **SORULAR:**

1. Büyükşehir Belediyesinin adı: **ESKİŞEHİR BÜYÜKŞEHİR BELEDİYESİ**

2. Çözüm Masasına aylık ve yıllık olarak gelen şikâyet/istek/talep miktarı nedir?

Ocak : 492 , Şubat : 372 , Mart : 402 , Nisan : 393 , Mayıs : 642 , Haziran : 414 , Temmuz : 533 , Ağustos : 376 , Eylül : 462 , Ekim : 443 , Kasım : 471 , Aralık : 430  
2017 Yılına Ait şikâyet/istek/talep miktarı ; 5630

3. Çözüm Masasına gelen şikâyet/istek/taleplerin çözülmesi süreci nasıl işlemektedir?

Çözüm Masasına gelen şikâyet/istek/talepler konularına göre ilgili birimlere dağılmaktadır. İlgili birim kamuya göre 2 Hafta (14 Gün) içerisinde cevaplamak zorundadır.

4. Çözüm Masasına gelen şikâyet/istek/taleplerin çözülmesi sürecinde mobil uygulamalardan yararlanılmakta mıdır?

Evet yararlanılmaktadır. Mobil Belediyecilik uygulaması aktif olarak kullanılan bir mobil uygulamadır.

### **Ek 1.1: Kamuda Veri Madenciliği Anketi (1.sayfa)**

5. Çözüm Masasına gelen şikâyet/istek/taleplerin çözülmesi sürecinde otomasyon programları, yazılımlar ve bilişim sistemleri kullanılmakta mıdır?

Info mail adresi üzerinden gelen şikâyet/istek/talepler Kent Bilgi Sisteminde toplanmaktadır.

6. Çözüm Masasına gelen şikâyet/istek/taleplerin çözülmesinde hangi program, yazılım ve bilişim sistemi kullanılmaktadır?

Kent Bilgi Sistemi

7. Çözüm Masasına gelen şikâyet/istek/taleplerin çözülmesi sürecinde kullanılan otomasyon programını belediyeiniz mi geliştirmiştir?

Hayır. Kullanılan otomasyon özel bir firmaya aittir.

8. Çözüm Masasına gelen şikâyet/istek/talepler ve buna karşılık gerçekleştirilen eylemler düzenli kayıt altına alınmakta mıdır?

Evet gerek dilekçe olarak gerekse mobil uygulamaların depolama alanlarında kayıt altına alınmaktadır.

9. Çözüm Masasına gelen şikâyet/istek/taleplerin ilgili belediye birimlerine iletilmesi, çözüm sürecinin takip edilmesi ve ilgili vatandaşa geri bildirim/bilgilendirme yapılmasında veri madenciliği ya da metin madenciliği yöntemlerinden yararlanıyor musunuz?

Hayır veri madenciliği ya da metin madenciliği yöntemlerinden yararlanılmamaktadır.

10. Çözüm masası faaliyetleri (gelen başvuruların sayısı, türü, çözülen başvuru oranı vs.) hangi bilişim programlarıyla (Word veya excel tablosu, otomasyon üzerinden otomatik olarak vb.) depolanmakta ve raporlanmaktadır?

Kent Bilgi Sistemi otomasyonu üzerinden otomatik olarak ayrıca Excel dosyası olarak da sayısal veriler depolanıp raporlanmaktadır.

11. Kurumunuzda veri madenciliği uygulaması olarak değerlendirdiğiniz diğer çözümler bulunmakta mıdır? Bulunmakta ise kurumunuzun hangi bölümünde ne amaçla (politika üretimi/çözümü vb.) kullanılmaktadır?

Kurumumuzda veri madenciliği uygulaması olarak değerlendirdiğimiz başka çözümler bulunmamaktadır.

## Ek 1.2: Kamuda Veri Madenciliği Anketi (2. sayfa)

Sıra No	Kök Kelime	Sıra No	Kök Kelime	Sıra No	Kök Kelime
1	talep	22	paket	43	küheylan
2	saat	23	optik	44	çörek
3	durak	24	sicil	45	amonyak
4	hat	25	puan	46	sayla
5	teşekkür	26	şifre	47	miting
6	çiy	27	kutlu	48	kaşar
7	kon	28	ihata	49	sarkıntı
8	asfalt	29	müjde	50	ispirto
9	ağaç	30	tun	51	çekiliş
10	im	31	sergi	52	gökdelen
11	proje	32	dekoratif	53	flüt
12	yazar	33	kırkayak	54	vesika
13	kreş	34	pirelen	55	akustik
14	barınak	35	ekşi	56	türküle
15	kay	36	kovan	57	kılçık
16	hangi	37	plak	58	pasaport
17	ışıklan	38	celep	59	fizyoterapist
18	plaka	39	kıraç	60	klasör
19	cam	40	defterdar	61	asabiyet
20	şikayet	41	tabiat	62	barışsever
21	nergis	42	alev		

**Ek 2: Öznitelik Seçimi Sonucunda Bulunan Kök Kelime Listesi**

## ÖZGEÇMİŞ

Yılmaz DEMİRCİ, 01.01.1983 tarihinde Kütahya’da doğdu. 1997-2000 yılları arasında Deniz Astsubay Hazırlama Okulunda gördüğü lise eğitiminin ardından, 2001 yılında Muhabere ve Elektronik Bilgi Sistemleri Okulu K.lığı’ndan mezun oldu. 2001 - 2005 yılları arasında Aksaz Deniz Üs K.lığı’nda, 2005 - 2010 yılları arasında Deniz Eğitim ve Öğretim K.lığı’nda Yazılım Uzmanı olarak Visual Studio .NET ve Microsoft SQL Server veritabanı kullanılan yazılım projelerinde görev yaptı. 2010-2016 yılları arasında Donanma K.lığına bağlı Envanter Kontrol Merkezi K. lığı’nda Veritabanı Yöneticisi olarak görev yaptı. 2016-2018 yılları arasında Deniz Kuvvetleri K.lığı’nda Veritabanı Yönetimi Ekibi Takım Lideri olarak çalışmıştır. Görev yaptığı yazılım birimlerinde C#.NET ve JAVA dillerinde geliştirilen projelerde görev almıştır. 2010-2018 yılları arasında Deniz Kuvvetleri Komutanlığı’nda kullanılmakta olan kurumsal kaynak planlama uygulamasına ait Oracle ERP (E-Business Suite), Oracle veritabanı ve Oracle Weblogic Server üzerinde çalışmalarına devam ederek ilgili uygulamaların yöneticiliğini sürdürmüştür. Deniz Kuvvetleri Komutanlığı’ndaki görevi süresince felaket kurtarım sistemleri donanım ve yazılım, veritabanı, uygulama ve kurumsal kaynak planlama yazılımı başta olmak üzere çeşitli sürüm yükseltme ve yenileme projelerinde görev yapmıştır. 2018 Ekim ayından itibaren NATO İletişim ve Bilgi Ajansı (NCI Agency) Mons/Belçika’da veritabanı yöneticisi olarak çalışmaya başlayacaktır. 2006 yılında Anadolu Üniversitesi İktisat Fakültesi Kamu Yönetimi bölümünden mezun oldu. 2010-2012 yılında Sakarya Üniversitesi Sosyal Bilimler Enstitüsü, Kamu Yönetimi Yüksek Lisans programını “Yerel Yönetimlerde Bilgi Yönetimi Algısı: Kocaeli Büyükşehir Belediyesi Örneği” konulu tezle tamamlamıştır. 2012 yılında arasında Sakarya Üniversitesi Sosyal Bilimler Enstitüsü, Kamu Yönetimi Doktora programına başlamıştır. Yönetim bilişim sistemleri ve veritabanı yönetim sistemlerine yönelik edindiği mesleki bilgi birikimi yanında, akademik olarak edindiği kamu yönetimi başta olmak üzere disiplinler arası konularda akademik çalışmalarına devam etmektedir. İngilizce bilmektedir. Ebru DEMİRCİ ile evli olup, Mert ve Yiğit adında iki çocuk babasıdır.