

T.C.
İSTANBUL BEYKENT ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI
ENDÜSTRİ MÜHENDİSLİĞİ BİLİM DALI

**VERİ ANALİZİ TEKNİKLERİYLE MÜŞTERİ DEĞER
SEGMENTASYONU OLUŞTURULMASI VE
PERAKENDE SEKTÖRÜNDE BİR UYGULAMA**

Yüksek Lisans Tezi

Tezi Hazırlayan
Dağhan BATAL

İstanbul, 2023

T.C.
İSTANBUL BEYKENT ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI
ENDÜSTRİ MÜHENDİSLİĞİ BİLİM DALI

**VERİ ANALİZİ TEKNİKLERİYLE MÜŞTERİ DEĞER
SEGMENTASYONU OLUŞTURULMASI VE
PERAKENDE SEKTÖRÜNDE BİR UYGULAMA**

Yüksek Lisans Tezi

Tezi Hazırlayan

Dağhan BATAL

2020006002

0009-0000-3557-6091

Dr. Öğr. Üyesi Selçuk ŞEN

İstanbul, 2023

YEMİN METNİ

Yüksek Lisans Tezi olarak sunduğum “**Veri Analizi Teknikleriyle Müşteri Değer Segmentasyonu Oluşturulması ve Perakende Sektöründe Bir Uygulama**” başlıklı bu çalışmanın, bilimsel ahlak ve geleneklere uygun şekilde tarafımdan yazıldığını, bu tezdeki bütün bilgileri akademik ve etik kurallar içinde elde ettiğimi, yararlandığım eserlerin tamamının kaynaklarda gösterildiğini ve çalışmanın içinde kullanıldıkları her yerde bunlara atıf yapıldığını, patent ve telif haklarını ihlal edici bir davranışımın olmadığını belirtir ve bunu onurumla doğrularım. 08.05.2023

Dağhan BATAL

T.C.
İSTANBUL BEYKENT ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ MÜDÜRLÜĞÜ
TEZLİ YÜKSEK LİSANS SINAV TUTANAĞI

08.05.2023

Enstitümüz *Endüstri Mühendisliği* Anabilim Dalı *Endüstri Mühendisliği* Programı Yüksek Lisans öğrencilerinden 2020006002 numaralı *Dağhan BATAL*'ın "Beykent Üniversitesi Eğitim-Öğretim Yönetmeliği"nin ilgili maddesine göre hazırlayarak Enstitümüze teslim ettiği "*Veri Analizi Teknikleriyle Müşteri Değer Segmentasyonu Oluşturulması : Perakende sektöründe bir uygulama*" konulu tezini, Yönetim Kurulumuzun 18/04/2023 tarih ve 2023/16 sayılı toplantısında seçilen ve On-line toplanan biz jüri üyeleri huzurunda, Beykent Üniversitesi Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 29. maddesinin 3. fıkrası gereğince 60 dakika süre ile Zoom programı aracılığıyla on-line olarak aday tarafından savunulmuş ve sonuçta adayın tezi hakkında "*OYBİRLİĞİ*" ile "*KABUL*" kararı verilmiştir.

İşbu tutanak, 2 nüsha olarak hazırlanmış ve Enstitü Müdürlüğü'ne sunulmak üzere tarafımızdan düzenlenmiştir.

DANIŞMAN
Dr. Öğr. Üyesi Se*** ŞE***
(İstanbul Beykent Üniversitesi)

ÜYE
Dr. Öğr. Üyesi Gü*** ÜN***
(İstanbul Beykent Üniversitesi)

ÜYE
Doç. Dr. Ceyda GÜ*** ŞE***
(Yıldız Teknik Üniversitesi)

Adı Soyadı : Dağhan BATAL
Danışmanı : Dr. Öğr. Üyesi Selçuk ŞEN
Türü ve Tarihi : Yüksek Lisans Tezi, 2023
Alanı : Endüstri Mühendisliği
Anahtar Kelimeler : Veri Madenciliği, Müşteri Segmentasyonu, Kümeleme
Analizi, Perakende Sektörü

ÖZ

VERİ ANALİZİ TEKNİKLERİYLE MÜŞTERİ DEĞER SEGMENTASYONU OLUŞTURULMASI VE BİR PERAKENDE SEKTÖRÜNDE UYGULAMA

Gelişen teknoloji ve artan lojistik faaliyetleri işletmelerin daha geniş müşteri ağına sahip olmasını sağlamaktadır. Bu değişim nedeniyle işletmeler artık müşterilerini tanımak zorunda kalmaktadır. İşletmeler müşterilerin davranış ve alışkanlıklarından yola çıkarak bir çıkarım sağlamak istemektedir. Bu çıkarımlar neticesinde müşterilerine stratejik tavırlar oluşturmakta ve yüksek oranda olumlu geri dönüşler beklemektedir.

Tez kapsamında öncelikle veri madenciliğine ilişkin literatür taraması gerçekleştirilmiştir. Literatür taraması neticesinde bir perakende sektörüne ait 1.000.000’u aşkın müşterinin davranış ve alışkanlıkları belirli kriterler dahilinde incelenmiştir. Veriler önce belirli bir ön aşamadan geçirilerek uygun hale getirilmiştir. Sonrasında küme sayısı belirlenerek K-Means yöntemi ile veriler kümelere ayrılmış benzer davranış ve alışkanlıklara sahip müşterilerin bir arada bulunması sağlanmıştır. Oluşan kümeler neticesinde her kümenin değer tanımlamaları yapılarak kümeler hakkında oluşturulabilecek stratejik tavırlar belirlenmiştir. Son bölümde sonuçlara ilişkin çıkarımlar değerlendirilmiştir.

Name and Surname : Dağhan BATAL
Advisor : Dr. Lecturer Selçuk ŞEN
Type and Date : Master Thesis, 2023
Field : Industrial Engineering
Keywords : Data Mining, Customer Segmentation, Cluster Analysis, Retail Industry

ABSTRACT

CREATING CUSTOMER VALUE SEGMENTATION USING DATA ANALYSIS TECHNIQUES AND ITS APPLICATION IN THE RETAIL INDUSTRY

Advancements in technology and logistics activities allow businesses to have a wider customer base. As a result of this change, businesses now need to know their customers better. Businesses aim to infer insights from customers' behavior and habits, and create strategic approaches for their customers, expecting high levels of positive feedback in return.

In this thesis, a literature review on data mining was initially conducted. As a result of the literature review, the behaviors and habits of over 1,000,000 customers in a retail sector were examined based on certain criteria. The data was processed through a preliminary stage and made suitable for analysis. Then, by determining the number of clusters, the data was divided into clusters using the K-Means method, grouping customers with similar behaviors and habits together. Strategic approaches that can be created for each cluster were identified by defining the value of each cluster. In the final section, conclusions were drawn based on the results.

İÇİNDEKİLER

Sayfa No.

ÖZ

ABSTRACT

TABLolar LİSTESİ	iv
ŞEKİLLER LİSTESİ	v
KISALTMALAR	vi
SÖZLÜK.....	vii
GİRİŞ.....	1

BİRİNCİ BÖLÜM

1. LİTERATÜR ARAŞTIRMASI	2
---------------------------------------	----------

İKİNCİ BÖLÜM

2. VERİ MADENCİLİĞİ.....	7
2.1 Veri Madenciliği Aşamaları.....	10
2.1.1 Problemin Tanımı	10
2.1.2 Veri önişleme	10
2.1.3 Veri Birleştirme	11
2.1.4 Veri Dönüştürme.....	12
2.1.5 Veri İndirgeme	12
2.1.6 Modelin Kurulması	12
2.1.7 Modelin Çalıştırılması	12
2.1.8 Değerlendirme	13
2.1.9 Modelin İzlenmesi	13
2.2 Veri Madenciliği Uygulama Alanları	13
2.2.1 Bankacılık	13
2.2.2 Sağlık	14
2.2.3 Eğitim.....	15
2.2.4 Pazarlama.....	16
2.2.5 E Ticaret.....	16

2.2.6 Sigortacılık.....	17
2.2.7 Telekomünikasyon.....	18
2.2.8 Müşteri İlişkileri Yönetimi	19
2.2.9 Risk Analizi	20
2.3 Veri Madenciliğinde Kullanılan Araçlar	21
2.3.1 Rapid Miner	21
2.3.2 Knime.....	22
2.3.3 Weka	23
2.3.4 Tanagra	23
2.3.5 Elki.....	23
2.3.6 Massive Online Analyss	24
2.3.7 Orange.....	24
2.3.8 KEEL	24
2.3.9 R.....	25
2.4 Veri Madenciliği Teknikleri	25
2.4.1 K Means Yöntemi.....	26
2.4.2 KNN En Yakın Komşu Yöntemi	26
2.4.3 Karar Ağaçları.....	27
2.4.4 Diskriminant Analizi.....	28
2.4.5 Birliktelik Kuralı.....	29
2.4.6 Genetik Algoritmalar	30
2.4.7 Yapay Sinir Ağları	31
2.4.8 Aykırılık Analizi	32
2.4.9 Aşamalı Kümeleme.....	33
2.4.10 Faktör Analizi	34

ÜÇÜNCÜ BÖLÜM

3. METODOLOJİ.....	35
3.1 Müşteri Segmentasyonu.....	35
3.2 Literatür ve Yöntem.....	37
3.3 K-Means.....	39
3.4 Silüet Katsayısı	42

DÖRDÜNCÜ BÖLÜM

4. UYGULAMA	45
4.1 Uygulamanın Amacı	45
4.2 Veri İçeriği ve Veri Önışleme Adımı:	46
4.3 En İyi K Değerinin Belirlenmesi:	48
4.4 Kümeleme Analizi:	50
4.5 Çıktıların Değerlendirilmesi	59
4.6 Kümeler İçin Yaklaşımların Belirlenmesi:	62
SONUÇ	64
KAYNAKÇA.....	66



TABLÖLAR LİSTESİ

	Sayfa No.
Tablo 1.1 Veri Madenciliğinde Kullanılan Modellerin Özet Tablosu	6
Tablo 4.2 Veriye İlişkin Örnek Tablo	46
Tablo 4.3 Yinelenen Verilere ilişkin Örnek Tablo.....	48
Tablo 4.4 Cluster 0 Verilerinden Örnek Kısım.....	51
Tablo 4.5 Cluster 1 Verilerinden Örnek Kısım.....	53
Tablo 4.6 Cluster 2 Verilerinden Örnek Kısım.....	54
Tablo 4.7 Cluster 3 Verilerinden Örnek Kısım.....	56
Tablo 4.8 Küme Kriterlerine Ait Ortalama ve Toplamlar.....	57
Tablo 4.9 Kümelerin Kriterlere Ait Alt ve Üst Değerleri	59
Tablo 4.10 Cluster Segmentasyon Değerlendirmesi	61
Tablo 4.11 Kriter Değerlerine Göre Segment Oluşturma	62

ŞEKİLLER LİSTESİ

	Sayfa No.
Şekil 4.1 Rapid Miner çıktısı.....	49
Şekil 4.2 Knime girdi görüntüsü.....	50
Şekil 4.3 Knime çalışma tasarımı görüntüsü	51



KISALTMALAR

BT	: Biliřim Teknolojileri
CRM	: Müşteri İliřkileri Yönetimi
DNA	: Deoksiribo Nükleik Asit
GAP	: Bořluk
İD	: Kullanıcı Numarası
RFM	: Güncellik, Sıklık, Parasallık
KOBİ	: Küçük ve Orta Büyüklükteki İşletmeler
Max	: En Yüksek Seviye
Min	: En Düşük Seviye
PSO	: Parçacık Sürü Optimizasyonu

SÖZLÜK

Algoritma: Belli bir problemi çözmek veya belirli bir amaca ulaşmak için tasarlanan yol.

Bulut Bilişim: Bilgisayarlar ve diğer cihazlar için, istendiği zaman kullanılabilen ve kullanıcılar arasında paylaşılan bilgisayar kaynakları sağlayan, internet tabanlı bilişim hizmetlerinin genel adıdır.

Büyük Veri: Daha fazla çeşitlilik içeren ve hacmi hızlıca artan verilerdir.

Değer: Bir şeyin gerekliliğini, önemini belirlemeye yarayan saymaca ölçü.

Derin Öğrenme: Bir veya daha fazla gizli katman içeren yapay sinir ağları ve benzeri makine öğrenme algoritmalarını kapsayan çalışma alanıdır.

E-Ticaret: Hizmet ve ürünlerin tüketiciler ile online kanallar vasıtasıyla bir araya getirildiği ticaret türüne verilen isimdir.

İstatistik: Doğru bir şekilde veri toplama ve verileri bilgiye dönüştürme bilimidir.

İş Zekâsı: Kuruluşların daha iyi kararlar almasına, bilgiye dayalı aksiyonlarda bulunmasına ve daha verimli iş süreçleri uygulamasına olanak sağlayan özellikler anlamına gelir.

K-Means: Denetimsiz bir bölümleme algoritmasıdır.

Kriter: Bir yargıya varmak veya değer vermek için başvurulmuş ilke.

Küme: İyi tanımlanmış nesneler topluluğuna denir.

Kümeleme Analizi: Bir nesne kümesini gruplama problemidir.

Literatür Taraması: Mevcut olan kaynaklar içerisinde belirlenmiş bir konu üzerinde yapılan detaylı araştırma sonucu o konuya ait verilerin sistemli biçimde toplanması sürecidir.

Makine Öğrenmesi: Tükettikleri verilere göre öğrenen ya da performansı iyileştiren sistemler oluşturmaya odaklanan bir yapay zekâ alt kümesidir.

Müşteri Segmentasyonu: Müşterilerin ortak özelliklerine göre gruplandırılması sürecidir.

Perakende: Bir şeyin gerekliliğini, önemini belirlemeye yarayan saymaca ölçü.

Strateji: Genel olarak bir şeyi elde etmek için izlenilen yol ya da amaca ulaşmak için geliştirilen bir plan olarak tanımlanabilir.

Telekomünikasyon: İki ya da daha fazla kişinin teknolojiyi kullanarak bilgi alışverişinde bulunmasına denir.

Veri: Bir araştırmada, bir tartışmada, bir akıl yürütmede sonuca ulaşabilmek için gereken ilk bilgi.

Veri Madenciliği: Karmaşık yazılım ve algoritmalar kullanarak verilerin daha fazla işlenmesi, depolanması ve analiz edilmesidir.

Yapay Zekâ: Bir bilgisayarın veya bilgisayar kontrollü robotun, genellikle akıllı varlıklarla ilişkili görevleri yerine getirme yeteneğidir.

GİRİŞ

Yaşamın varoluşundan beri insanlar ve diğer tüm canlılar sürekli olarak izler bırakmaktadır. Özellikle son yüzyılda teknolojinin oldukça hızlı gelişim göstermesi ve bu gelişmenin devam etmesiyle hayatımızın hemen hemen her alanında yer edinmesiyle de bırakılan bu izler bir hayli artmaktadır. İlerleyen süreçlerde de bu bilgi birikim artışının oldukça hızlı büyüyerek artması beklenmektedir. Biriken bu izlerden birtakım çıkarımlar sağlanması mümkün olmaktadır. Bu çıkarımlar sayesinde tarih, biyoloji, coğrafya, arkeoloji, istatistik gibi birçok alanda bilgiler edinilmesi sağlanılmaktadır. İlk canlı türleri, ilk canlı türlerinin özellikleri bilinen en eski insan topluluklarının yaşayış biçimleri, dünyanın coğrafik özellikleri ve dünyanın değişen iklim özellikleri dahi bu biriken izler sayesinde sağlanabilmektedir. İlkel insanlıktan günümüze kadar insan topluluklarının nasıl gelişim evreleri gösterdikleri, nesli tükenmiş canlıların kaç yüzyıl önce yaşadıkları hangi canlının hangi sınıflandırma içinde yer aldıklarıyla ilgili bilgiler de sağlanabilmektedir. Yaşanılan doğa olaylarını ve yaşanılan bu doğa olaylarının dünya üzerindeki etkilerinin saptanıp değerlendirilmesi de eklenebilmektedir. Günümüzde ise bilgi birikimi yalnızca geçmiş olaylar üzerinde değil mevcut zaman ve gelecek zaman hakkında da bilgi çıkarımlarında bulunmamızı sağlamaktadır. Günümüzde bilgilerin daha farklı şekilde özellikle teknolojik araç ve yöntemler kullanılarak öngörü denilen bir yaklaşım ile tahminler ve olası sonuçlar belirlenmeye çalışılmaktadır. Çünkü zaman içerisinde fark edilen olaylardan biride benzer eylemlerin benzer sonuçlar doğurmasıdır. Bu mantık ışığında ise elde edilen veriler toplanıp gruplanarak belli bir sentez sonrasında bir çıktı elde edilmekte ve bu çıktılar sayesinde bir öngörü sağlanmaya çalışılan eylemler bütünü oluşturulmuştur. Bu eylemler topluluğuna ise “Data Mining” dilimizde ise veri madenciliği uygulamaları denilmektedir.

Tezin ilk bölümde veri madenciliği kavramıyla alakalı olarak araştırılan literatür çalışması bulunmaktadır. İkinci bölüm veri madenciliğinin önemini, tarihçesini ve alt başlıklarını içermektedir. Üçüncü bölümde müşteri segmentasyonuna değinilmektedir. Dördüncü bölümde veri madenciliğinde kullanılan araçlar ve kullanılan teknikler ışığında bir perakende sektörüne ait müşteri alışkanlık verileriyle müşteri değer segmentasyonu oluşturan bir uygulama gerçekleştirilmektedir. Son bölümde ise tez çalışması ve uygulama sonuçlarına ilişkin bir sonuç değerlendirilmesinde bulunulmuştur.

BİRİNCİ BÖLÜM

1. LİTERATÜR ARAŞTIRMASI

Veri madenciliği ile ilgili yazılmış birçok makale ve tez incelenmiştir. İncelemelerin birçoğunda veri madenciliği tarihçesi hakkında bilgiler verilerek giriş yapıldığı görülmektedir. Veri madenciliği gelişimi bir süreç olarak anlatılmaktadır. Veri madenciliği anlamında geliştirilen yöntem ve araçlar yapılan araştırmaya ilişkin öncesinde anlatılmaktadır. Veri madenciliğinin uygulandığı alanlardan ve sektörlerden bahsedilmektedir. Özellikle makine öğrenmesi, yapay zekâ ve derin öğrenme konularına da değinilmektedir. Veri madenciliğine ilişkin çalışmaların bir aşamadan oluştuğu ve aşamaların neler olduğu hakkında bilgiler verilmektedir. Çalışmaların yapıldığı alanlardan bazıları finans, bankacılık, müşteri ilişkileri yönetimi, eğitim ve sağlık alanları olmaktadır. Çalışmalar çoğunlukla tanımlayıcı veya tahmin edici modellerden oluşmaktadır. İncelemelerde çoğunlukla kullanılan yöntemlerin kümeleme ve birliktelik analizi oldukları görülmektedir. Performans yönetimi açısından da veri madenciliği tekniklerinden yararlanıldığı görülmektedir. 2021 yılında yazılan bir makalede turizm alanında yapılan bir çalışmaya yer verilmiştir. Büyük veri derin öğrenme yapay zekâ kavramlarına da yer verilmiştir. (Küpeli ve Ünlüöner, 2021) Bankacılık alanında ise veri zarflama ile veri madenciliği teknikleri bir arada kullanılarak mevduat bankalarının performansları kıyaslanmıştır. 20 Banka etkinliklerine göre sıralanmıştır (Seyrek ve Ata, 2010) 2006 yılındaki makalede ise veri madenciliği uygulama alanları hakkında bilgiler verilmiş, veri ambarlarının özellikleri ve yapısı anlatılmıştır. (Baykal, 2006) Bir muhasebe ve finansman dergisinde yayınlanan çalışmada hile ve usulsüzlükler, veri madenciliği yöntemi ile belirlenmeye çalışılması anlatılmıştır. (Terzi, 2012) Veri madenciliğinin en çok kullanılan tekniklerinden olan K-Means yöntemi ile Sakarya ilindeki trafik kazalarının sebepleri ile kaza haritasının desenleri oluşturulmuştur. (Güner vd., 2018) Veri madenciliğinin en sık kullanım alanlarından biride eğitim alanıdır. Öğrenci başarılarının değerlendirme ve kıyaslanmasında eğitsel veri madenciliği kullanılmıştır. (Hotaman, 2020). Kriminal alanda ise makine öğrenmeleri ile birlikte kullanılarak suç analizinde veri madenciliği uygulamaları kullanılmıştır. (Orakçı vd., 2016) Daha önce KPSS sonuçların tahmin edilmesinde yapılan çalışmaya benzer bir çalışma

öğrencilerin mezuniyet notlarının tahminleri yapay sinir ağları ve karar ağaçları kullanılarak geçerlilik analizi ile modellenmiştir. (Şengür ve Tekin, 2013) Bir farklı makalede ise veri analizi süreçlerinde istatistik mi, yoksa veri madenciliği uygulamalarının mı daha etkin olduğu kıyaslanmıştır. (Emre ve Erol, 2017) Sadi Evren Şeker ve Merve Yıldız'ın hazırladığı makalede veri madenciliği uygulama alanlarında kullanılan yazılım ve araçlar anlatılmış ücretli-ücretsiz, kapladıkları alan, sonuç görselleştirme, kullanım kolaylıkları gibi özellikleri üzerinden değerlendirilmiş ve açıklanmıştır. (Şeker ve Yıldız, 2016) Adem Tekerek de 2011 yılında veri madenciliği araçlarını anlatan benzer bir makale yazmıştır. (Tekerek, 2011) Bir bilişim dergisinde ise tıp ve sağlık hizmetleri kullanım alanlarında veri madenciliğinin yerinden bahsedilmiş veri madenciliği aşamaları anlatılarak iş zekâsı kavramına da değinilmiştir. (Koyuncugil ve Özgülbaş, 2009) Ayrıca yapay sinir ağları kullanılarak 2012 yılındaki bir çalışmada hastanelerin hasta yoğunlukları belirlenmeye çalışılmıştır. (Irmak ve vd., 2012) Bir müşteri analizi çalışmasında RFM ve uyum analizi yöntemlerinden faydalanılarak analiz yapılmıştır Çalışma 2020 yılında yapılmıştır (Başkol, 2020) Özmen ve arkadaşlarının yaptığı bir çalışmada ise meta sezgisel bir yöntem olan parçacık sürü optimizasyonu kullanılarak bir telekomünikasyon firmasında müşteri bölümlenmesi çalışması yapılmıştır. (Özmen ve vd., 2018) 2021 yılında ise yine veri madenciliği tekniklerinden faydalanılarak ayrılan müşteri analizi yapılmıştır. Çalışma Düzce Üniversitesi Bilim ve Teknoloji dergisinde yayınlanmış Furkan Uyanık ve Mustafa Cem Kasapbaşı tarafından hazırlanmıştır (Uyanık ve Kasapbaşı, 2021) Yine 2021 yılında yayınlanan bir makalede veri madenciliğinin finans sektöründeki yerini ve hangi alanlarda fayda sağladığı anlatılmış ve veri madenciliği teknikleri üzerinde durulmuştur. (Doğan ve Kevser, 2021) Burcu Oralhan'ın yaptığı bir çalışmada telekomünikasyon sektörüne ait arızalara ilişkin analizler ve madenciliği teknikleri ile incelenmiştir. Çalışma 2019 yılında hazırlanmış 2020 de yayınlanmıştır. (Oralhan, 2020) Dener ve arkadaşlarının hazırladığı makalede veri madenciliğinde kullanılan yazılım ve araçlar anlatılmış ve WEKA aracı kullanılarak örnek bir çalışma yapılmıştır. (Dener ve vd., 2009) Özçakır ve Çamurcu tarafından hazırlanan makalede veri madenciliği tekniğinde kullanılan birliktelik kuralı yöntemi için bir veri madenciliği yazılımı tasarlanmış ve örnek bir uygulama yapılmıştır (Özçakır ve Çamurcu, 2007) Bulut bilişim kavramı son dönemin popüler

bir kavramı haline gelmiştir. Yıldırım ve Birant'ın 2018 yılında yayınlanan makalesinde veri madenciliği teknikleri ile olan ilişkisi anlatılmış yapılan literatür taraması sunulmuştur. (Yıldırım ve Birant, 2018) Eda Çoşlu 2013 yılında hazırladığı makalede ise veri madenciliği aşamaları üzerinde durarak süreçleri anlatmıştır. (Çoşlu, 2013) 2020 yılından yayınlanan Alan ve Karabatak'a ait makale de sınıflandırma ilişkisine etki eden faktörler değerlendirilmekte ve veri madenciliği teknikleri anlatılmaktadır. (Alan ve Karabatak, 2020) Ayşe Oğuzlar'ın hazırladığı makale de veri madenciliği süreçlerinin önemli aşamalarından olan veri ön işleme adımları anlatılmaktadır. (Oğuzlar, 2003) Doğan ve arkadaşlarının hazırladığı bir araştırma makalesi veri madenciliği tekniklerinden kümeleme analizi kullanılarak sigortacılık alanında bir uygulama gerçekleştirilmiştir. Çalışma Karaelmas Fen ve Mühendislik dergisinde yayınlanmıştır. (Doğan ve vd., 2017) Cem Gürler 2023 yılında yaptığı çalışma da K-Ortalamlar yöntemi kullanarak ülkeleri turizm potansiyellerine göre değerlendirerek kümelenmesi sağlayan bir araştırma makalesi bulunmaktadır. (Gürler, 2023) Endüstri Mühendisliği Dergisi'nde yayınlan bir çalışmada karar ağacı algoritmaları kullanılarak bilgisayar ve internet güvenliği üzerine bir uygulama gerçekleştirilmiştir. Çalışma 2014 yılına aittir. Günay Kılıç ve arkadaşlarının 2019 yılında hazırladığı 2020 yılında yayınlanan çalışmasında kümeleme analizi kullanılarak belediyelere ait çevre hizmetlerinin değerlendirilmesini içeren bir çalışmaları bulunmaktadır. (Kılıç ve vd., 2020) Esra Çelik ve arkadaşlarının 2021 yılında hazırladığı araştırma makalesinde bir duygu analizi için veri madenciliği sınıflandırma algoritmaları kullanılarak değerlendirilmesi yapılmıştır. Çalışma Avrupa Bilim ve Teknoloji Dergisi'nde yayınlanmıştır. (Çelik ve vd., 2021) Ersin Namlı ve Sümeyra Murat bir perakende sektöründeki, müşterilere yönelik strateji pazarlamaları oluşturmak veri madenciliği tekniklerinden olan kümeleme analizini kullanarak bir uygulama gerçekleştirmiştir. Çalışma 2019 yılında yayınlanmıştır. (Namlı ve Murat, 2019) Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi'nde yayınlanan bir araştırma makalesinde kümeleme algoritmalarından faydalanılarak perakende sektöründeki zincir mağazaların gruplanarak segmente edilmesi sağlanmıştır. Çalışma 2019 yılında yayınlanmış Emrah Bilgiç ve Özgür Çakır tarafından hazırlanmıştır. (Çakır ve Bilgiç, 2019) Yine aynı dergide yayınlanan bir makalede KNN algoritması kullanılarak benzer bir çalışma gerçekleştirilmiş bilimsel

makalelerin sınıflandırılması yapılmıştır. Çalışma Deniz Kılınç ve arkadaşları tarafından yapılmış 2016 yılında yayınlanmıştır. (Kılınç ve vd., 2016) 2007 yılında yayınlanan bir makalede kümeleme algoritmaları bir arada kullanılarak karşılaştırmalı bir şekilde performansları incelenerek tespitlerde bulunulmuştur. (Işık ve Çamurcu, 2007) Muhammet Atalay ve Enes Çelik'in büyük veri analizlerinde kullanılan yapay zekâ ve makine öğrenmesi algoritmalarına değinilerek anlatımlarda bulunulmuştur. Makale 2017 yılında yayınlanmıştır. (Atalay ve Çelik, 2017) Selay Giray'ın hazırladığı 2016 yılında yayınlanan çalışmasında hükümlülere ait veriler iki aşamalı kümeleme analizi tekniği kullanılarak değerlendirilmiştir. (Giray, 2016) Özgür Akpınar'ın 2018 yılında yayınlanan çalışmasında veri madenciliğinin sigorta alanındaki önemi anlatılmaktadır. Veri madenciliği kavramlarına değinilerek işletmelere sağladığı faydalar üzerinde durulmuştur. (Akpınar, 2018) Bir derginin özel sayısında yayınlanan araştırma makalesinde Neslihan Köse ve Filiz Ersöz demir çelik endüstrisinde gerçekleşen kazalar üzerine karar ağacı algoritmalarını kullanarak bir çalışma gerçekleştirmiştir. 2020 yılında yayınlanmıştır. (Köse ve Ersöz, 2020) 2014 yılına ait bir makalede veri madenciliği yazılımları anlatılarak özellikleri üzerinden birbirleriyle kıyaslanmıştır. Çalışma Mümine Kaya ve Selma Ayşe Özel'e aittir. (Kaya ve Özel, 2014) Teknolojinin ve lojistik faaliyetlerin gelişmesi ile daha da artan e-ticaret faaliyetleri üzerine 2016 yılında yayınlanan makalede bir değerlendirme gerçekleştirilmiştir. E-ticaret verileri ile müşteri profili arasındaki bağın değerlendirilmesi yapılmıştır. Çalışma da K-Means kümeleme analizi kullanılmıştır. (Özseven ve Ersöz, 2016) Veri bilim dergisinde yayınlanan Serpil Sevimli Deniz'in çalışmasında kural tabanlı sınıflandırma algoritmaları üzerinde durularak anlatılmıştır. Bu algoritmalarla ilişkin değerlerin karşılaştırılması yapılmıştır. Çalışma 2021 yılında yayınlanmıştır. (Sevimli Deniz, 2021) Yine Serpil Sevimli Deniz'e ait veri madenciliğine ilişkin bir çalışma da yapay sinir ağlarından faydalanılarak Türkiye'deki turizm gelirleri aylık bazlı olarak tahmin edilmeye çalışılmıştır. Makale 2019 yılında yayınlanmıştır. (Sevimli Deniz, 2021) Serkan Savaş ve arkadaşları 2012 yılında yayınlanan bir çalışmasında veri madenciliğine ilişkin Türkiye'deki örnek uygulamalar üzerinde durarak anlatmıştır. (Savaş ve vd., 2012) Literatürde veri madenciliği tekniklerini ayırmada en sık kullanılan ayırma şekli tanımlayıcı modeller ve tahmin edici modeller şeklinde olmaktadır.

Tablo 1.1 Veri Madenciliğinde Kullanılan Modellerin Özet Tablosu

Yazarlar	Yayınlanma Yılı	Kullanılan Model
Cengiz ve Öztürk	2012	Tanımlayıcı
Aydın	2013	Tanımlayıcı
Çalış vd.	2014	Tahmin edici
Onan	2014	Tahmin edici
Çalış ve Baynal	2016	Tanımlayıcı
Kılınç vd.	2016	Tahmin edici
Giray	2016	Tanımlayıcı
Doğan vd.	2018	Tanımlayıcı
Güner vd.	2018	Tanımlayıcı
Namlı ve Murat	2019	Tanımlayıcı
Alan ve Yeşilyurt	2019	Tanımlayıcı
Yıldız vd.	2019	Tanımlayıcı
Alan ve Karabatak	2020	Tahmin edici
Kılıç vd.	2020	Tanımlayıcı
Sabah ve Bayraktar	2020	Tanımlayıcı
Köse ve Ersöz	2020	Tahmin edici
Çoşkun vd.	2021	Tahmin edici
Tuncer ve Karaboğa	2021	Tanımlayıcı
Doğruel ve Fırat	2021	Tahmin edici
Gürler	2022	Tanımlayıcı
Ödev ve Peker	2022	Tanımlayıcı
Dilki ve Başar	2022	Tahmin edici
Gürler	2023	Tanımlayıcı

İKİNCİ BÖLÜM

2. VERİ MADENCİLİĞİ

Daha önceden bilinmeyen bir durumun ve bir bilgi yığınının dinamik bir süreçten geçirilmesi ile elde edilen bir sonuç olarak tanımlanabilir. (Baykal, 2006) Daha geniş ve benzer anlamıyla ise eldeki çok büyük verilerden faydalı olan bilgiyi çıkarabilme veya ulaşılabilme kabiliyeti yani madenleme sürecidir. Veri madenciliği istenilen veya beklenen sonuçları öngörebilme kabiliyetidir. Daha öncede belirtildiği gibi her eylem bir iz bırakmaktadır. Bu bırakılan izler birikerek bir bilgi diğer dildeki anlamıyla datayı oluşturmaktadır. İnsanlar bu geçmişteki bilgileri kullanarak gelecekte oluşabilecek sonuçları önceden bilebilmek öngörebilmek ister. Bir sonraki olabilecek doğal afetleri geçmiş verilerden analiz etmek olası bir doğal afette de etkileri en aza indirmeye yarayabilmektedir. Bir firma ürün veya hizmeti müşterisine sunarken veri madenciliği uygulamalarını kullanarak doğru hedeflerle müşterisini yönlendirebilmektedir. Aynı zamanda bir ürün veya hizmeti sunarken ikinci bir ürün ya da hizmeti birlikte sunabilmek için bu uygulamalardan faydalanabilmektedir. Müşteri memnuniyet anketleri analiz edilerek, herhangi bir işletme için farklı politikalar izlenmesi sağlanabilir. Kayıp müşteri nedenleri incelenerek ilerleyen süreçlerde daha az sayıda müşteri kaybı yaşanması sağlanabilir. Bu sebeplerden dolayı veri madenciliği hemen hemen her alanda kullanılmakta ve kullanılabilir. Bankacılık, finans, üretim, sağlık, teknoloji, havacılık, pazarlama, müşteri ilişkileri yönetimi veri madenciliği uygulamalarının kullanıldığı başlıca alanlardandır. Özellikle son 20 yılda teknolojinin gelişmesi ile e-ticaret alanında da veri madenciliği uygulamalarından faydalanılmaktadır. Hatta yakın dönem içinde e-ticaret veri madenciliği uygulamalarının neredeyse merkezi haline gelmektedir. Bir sonraki kısımda bu alanlarla ilgili detaylı bilgi verilmektedir. Veri madenciliği neden önemlidir sorusu sorulduğunda ise birkaç durum ön plana çıkmaktadır. Teknolojinin gelişmesi ile geçmişe oranla çok daha fazla iz bırakılmaktadır. Bu durum hem avantaj hem de dezavantaj sağlamaktadır. Her alanda bilgiye ulaşabilmek hem de bunu hızlı şekilde sağlayabilmek bir avantaj sağlarken, bu durum çok sayıda gereksiz ve yanlış bilgiyi de karşımıza çıkarabilmektedir. Bu bilgilerin işlenmesiyle de çok yanlış öngörü sonuçları elde edilebilir ki bu istenilen bir durum değildir. Bu açıdan veri

madenciliği uygulamaları önemli bir süreçtir. Tarihteki olaylara bakıldığında icat ve keşiflerin birçoğu deneme-yanılma denilen olaylardan elde edilmiştir. Ancak bazı durumlar içinde günümüzde böyle bir duruma yeltenmek zaman, maliyet, ürün ve işçilik gibi kısıtlardan dolayı mümkün olamamaktadır. Hatta imkân sınırları içinde dahi olsa içinde bulunduğumuz yüzyıl içerisinde kabul edilebilir bir yöntem değildir. Bu yöntemleri denemek ise bazen geri dönülemez müşteri ve ürün kayıplarına neden olabilmektedir. Artık elde edilen verilerin fiziki olarak depolanması her zaman olmasa da yeterli olamamaktadır. Verilerin büyük bir kısmı elektronik ortamlarda saklanmaktadır. Bu verilerin güvenli bir şekilde saklanabilmesi önemlidir. Aynı zamanda tutarsız bilgilerden arındırılmış olması önemlidir. Tutulan bu verilerin birlikteliği, birliktelikteki uyumluluğu, tutarlı bir şekilde filtrelenebilir olması artık verilerden arındırılmış olması hem genel anlamda hem de veri madenciliği anlamında çok önemlidir. Özetle veriler sınırsız sayılabilecek kadar fazlayken, bir işletmenin sahip olabildiği zaman, sermaye, ürün işçilik, enerji vb. sahip olunabilen yetkinlikler oldukça kısıtlıdır. Bu iki ikilemde dengeli bir sonuç alabilme durumu için veri madenciliği uygulamaları önemli bir yer eder. Veri madenciliği temelleri 1950'ye dayanmaktadır. Ancak uygulamaların temelini bilgiye dayandığı varsayıldığında bilinen en eski tarih dahi veri madenciliği uygulamalarına katkı sağlamaktadır. Çünkü buradaki amaç o ana kadar oluşturulmuş dizinlerden bir sonraki süreç ve oluşumlardan tahmin ya da tahmine yakın sonuçlar oluşturabilme çabasıdır. Farklı bir bakış açısıyla ise veri madenciliğinin temeli ilk bilgisayarın üretilmesiyle başlamıştır. Çünkü ilk bilgisayarların temel amaçlarından biri veriyi bir arada tutabilmektir. Devam eden süreçte karşılaşılan problemlere çözüm arayışları ise veri madenciliğinin gelişimlerine de katkı sağlamıştır. 1960'lı yıllara gelindiğinde ise veri depolama ve veri tabanı gibi kavramlar literatürde yer edinmeye başlamıştır. Bu yıllarda regresyon analizi sinir ağları gibi benzer algoritmalar üzerinde çalışılmıştır ve ilk basit öğrenmeli bilgisayarlar geliştirilmiştir. İlerleyen 10 yıl içerisinde ise artık çok önemli bir kavram haline gelen ilişkisel veri kavramı kullanılmaya başlamış ve basit öğrenmeli makineler üzerinde geliştirilme sağlanmıştır. 1980'li yıllara gelindiğinde ise artık uygulama alanlarının görülmesi başlanmış ve bu süreçlerden fayda sağlanılmaya başlanmıştır. Yayılarak gelişmekte olan bu alan başta mühendislik alanlarında kullanılmaya başlanmış. Büyük firmalar hem ürün ve müşteriler hem de rakip firmaları ile ilgili

bilgileri toplamakta ve bunlar arasında ilişkisel bağ kurma aşamalarına girmiştir. Büyük veriler üretilmeye devam edilmekte olup artık veri sorgulama dili ve vb. uygulamaları kullanılarak erişim sağlanmaktadır. 1990'lı yıllarda ise biz bu verilerden nasıl fayda sağlarız diye sorular sorulmaya ve cevap aranmaya başlanmıştır. Aynı zamanda geçmiş yıllara göre çok fazla oranla bilgi yığını birikmeye başlanmıştır. Hala geçerliliği koruyan birçok yazılım ve uygulamanın temeli de bu yıllarda atılmış ve geliştirme süreci devam etmiştir. 1992 yılı ise ilk yazılımın oluşturulduğu tarih olmuştur. Ayrıca bu tarihler içinde 89 Workshop makalesi de veri madenciliği ile ilgili temel kavram ve tanımları ortaya koyan aynı zamanda süreçleri hızlandıran önemli bir makale olmuştur. 2000'li yıllarda artık bu gelişim sürecinden gözle görülür bir şekilde fayda sağlanmaya başlanılmış fayda sağlanmaya devam ettikçe ise kullanım alanları hemen hemen her alanda etkisini göstermeye başlanmıştır. Bu devamlılık hali veri madenciliği teknikleri ve yazılımları geliştirilerek sürmüştür. 2000'li yıllarda ayrıca internetin hayatımızda aktif yer edinmesiyle birlikte geçmişe kıyasla olağanüstü bir fark oluşarak bilgi birikimi artmaya başlamış çok fazla sayıda yazılım ve teknik de literatüre dahil olmuş Halihazırda mevcut olan teknik ve yazılımlar ise geliştirilmiştir. Bu sürecin devamı olan 2010'lu yıllarda ise WhatsApp, Twitter, Facebook, Tumblr, Instagram gibi sosyal araçların kullanımının artmasıyla büyük veri yığını birikmeye devam etmiştir. Ayrıca e-alışveriş siteleri olan Amazon, Aliexpress, Etsy kullanımı özellikle bu yıllarda fazlasıyla yaygınlaşmış olup veri madenciliği teknik ve süreçlerinden daha fazla yararlanılmaya ihtiyaç duyulmasına neden olmuştur. Ayrıca ülkemizde ise Hepsiburada, N11 Trendyol gibi alışveriş siteleri de bu yıllarda aktif olarak kullanılmaya başlanmıştır. Kişilerin alışveriş alışkanlıkları ürün ve hizmet aramaları gibi veriler analiz edilerek onlara doğru ürün ve ilgi çeken kampanyalar sunulmaya çalışılıp alışveriş alışkanlıklarının devamlılığı sağlanması veri madenciliğinin temel amaçlarından birisi olmuştur. 2020 yılına geldiğimiz de ise telefonlara ve bilgisayarlara ek olarak nesnelerin interneti olarak adlandırılan teknolojiler geliştirilmiştir. Yazıcılar televizyonlar saatler hatta beyaz eşyalarda internete bağlantı özellikleri geliştirilmiştir. Günümüzde özellikle yeni kavramlar haline gelen derin öğrenme yapay zekâ gibi kavramlar veri madenciliğinin de temel prensiplerini kapsamaktadır. Geline son noktada ise belli bir kapasite ve teknolojiye sahip tüm işletmeler, sağlık kurumları, eğitim kurumları KOBİ'ler veri

madenciliğinden faydalanmaktadır. Tüm alanlarda artık bir ihtiyaç haline gelmiş mevcut pazarlama stratejilerini geliştirmek isteyen firmalar için kaçınılmaz bir ihtiyaç haline gelmiştir. Çünkü daha öncede belirtilmiş olup artık günümüzde veri birikimi çok faktörlü ve sonsuz denilebilecek kadar büyüktür. Ancak geleneksel yöntemler ile bunların hepsinin tek tek işlenip anlamlı bir sonuç elde edilebilmesi için de yeterli zaman, maliyet ve işçilik bulunamaya bilmektedir. Bu nedenle veri madenciliğinin geliştirilmiş teknik ve yazılımı özellikle büyük veri denilen verilerde hızlı ve anlamlı sonuçlar elde etmemizi sağlamaktadır. Veri madenciliği uygulamalarının gelecekte nasıl bir ilerleme kat edeceği merakla beklenmektedir. Ayrıca ne kadar olumlu ne kadar olumsuz sonuçlar karşımıza çıkaracağı tartışma konuları olmaktadır.

2.1 Veri Madenciliği Aşamaları

Her çıktının elde edilme süreci olduğu gibi veri madenciliği uygulamalarının belli bir prosesi bulunmaktadır. Herhangi bir soruya cevap veya bir öngörü aradığımız bu süreçlerde, girdi aşamasında ne kadar titiz ve doğru davranılırsa çıktıda da doğru ve daha kolay sonuç alınması sağlanır. Öncelikle problemin tanımını ve çalışmadan istenilen hedef belirlenmelidir. Daha sonra veri ön işleme aşamaları yapılmalı sonrasında ise ortaya çıkan sonuç ve örüntülerden anlamlı bilgiler değerlendirilerek bir sonuç elde edilmelidir. Bu bölümde bu aşamalardan detaylı olarak bilgiler verilecektir.

2.1.1 Problemin Tanımı

Her bilimsel araştırmada olduğu gibi öncelikle yapılacak çalışma ile ilgili problemin tanımlanması gerekir. Problemin tanımı yapıldıktan sonra çalışmayla ilgili istenilen amaç net olarak ifade edilmelidir. Bilimsel çalışmalarda elde edilen girdi ve hedeflenen çıktı beklentisi açıkça belirtilmelidir. Problem çözüm aşamasına geçilmeden önce verinin toparlanarak işlenmesi ve model olarak hazır hale getirilmesi gereklidir. Bu aşamalara veri madenciliğinde veri ön işleme adımları denilmektedir.

2.1.2 Veri ön işleme

Bu adım verilerin toplandığı eksik verilerin tamamlandığı gürültülü ve anormal verilerin çıkarıldığı adımdır. Sonrasında elde edilen veriler toplanarak aynı anlamlara gelen veriler ortak bir dönüşüme çevrilir. Sonunda ise çalışmaya uygun hale getirilmesi için gerekli bilgi indirgenir. Örneğin mevcut veri kullanacağımız algoritma

ile uyumlu olmayabilir. Çünkü kimi algoritmalar sayısal değerler kullanırken kimi algoritmalar ise kategorik değer kullanabilmektedir. Veri önışleme aşamasında uygun hale getirmek gerekir. (Köseoğlu ve Demirci, 2017)

Genel olarak veri önışleme řu adımlardan oluşur. Veri toplama, veri temizleme, veri birleřtirme, veri dönüřtürme ve veri seçme (indirgeme) adımlarıdır. (Tüzüntürk, 2010)

Veri toplama aşamasında çalışmaya ilişkin veriler toplanmaktadır. Veriler birçok farklı ortamda bulunabilmektedir. (Tekerek, 2011) Bu verilerin çalışma yapılacak ortak bir sunucu üzerinde bir araya getirilmekte ve toplanmaktadır. Sağlanan bu bütün sonrasında veri temizleme aşamasına geçilmektedir.

Veri Temizleme aşaması aşamalardaki en önemli süreçtir. Gereksiz verilerin arındırılması eksik verilerin ise doldurulması gerekli olabilmektedir. Bunun için yapılacak adımlardan birkaçı řunlardır; diğeri verilerin ortalaması hesaplanarak bir değeri atanabilir veya sabit bir değeri yazılabilir. (Çoşlu, 2013) Regresyon ve karar analizi yöntemleri kullanılarak da eksik değeri atama yapılabilir veya eksik değeri içeren veriler silinerek dışarı atılabilir. (Oğuzlar, 2003) Veri içerisinde gürültü ve anormal değeri bulunabilmektedir. Örneğinin varlık içeren bir değeri negatif değeri alması anormal bir durumdur. -10 cm uzunluk -10 elma gibi değeri uygun değeri değildir veriden atılması gereklidir. Ya da bir insan kilosu için 900 kg değeri girilmesi gürültü bir veridir. Özetle bu aşama eksik verinin tamamlandığı gürültü verilerin temizlendiğı aşamadır. Bu aşamada yapılacak hatalar çıktıdan yanlış sonuçların alınması ve tekrar tekrar geri aşamalara dönülmesi nedeniyle uzun süreç ve maliyetler çıkarabilmektedir. Bu aşamada titiz bir süreç geçirilmesi sonuç ile ilgili kaliteli sonuçlar elde edilmesini sağlamaktadır.

2.1.3 Veri Birleřtirme

Veri toplama aşamasında birçok farklı kaynaktan veri bir araya getirilebilmektedir. Ancak bu verilerin bir arada iken standart bir kalıp haline getirilmesi gereklidir. Herhangi bir kaynaktan alınan veri birinde “tüketici numarası” şeklindeyken diğeri kaynaktaki “tüketici id” olarak belirtilmiş olabilir. Bu duruma řema birleřtirme hatası adı verilir. (Oğuzlar, 2003)

2.1.4 Veri Dönüştürme

Önceki adımlarda toplanıp temizlenen ve uygun şema haline getirilen veriler anlamları değiştirilmeden uygun bir hale getirilmelidir. Buradaki amaç veriyi daha başarılı kullanılabilecek hale getirmektir. (Yıldız ve Şeker, 2016) Uygun modele göre içerik değiştirilirken şekil korunmaktadır. Çünkü kullanılacak model ve algoritma önemli bir rol oynamaktadır. Veri dönüştürme işlemlerinde genellikle normalleştirme yöntemi uygulanır. En sık kullanılan normalleştirme yöntemleri ise min-max, Z skor ve ondalık ölçeklemedir. (Oğuzlar, 2003)

2.1.5 Veri İndirgeme

Veri madenciliğinde elde edilecek çıktıya ilişkin amaç en sağlıklı sonuçları almaktır. Sağlıklı bir çıktı için girdi verilerinin aykırı verilerden arındırılması gereklidir. Analizi yapılacak veri setinden de amaca uygun verilerin seçilmesi gereklidir. (Sabah ve Bayraktar, 2020) Veri indirgeme aşaması veri ön işleme aşamasının en önemli adımı olmakta olup zaman alıcı bir süreçtir. Veri indirgeme adımının sağlıklı işlemesi çıktıdan alınan olumlu bir sonuca büyük katkı sağlamaktadır.

2.1.6 Modelin Kurulması

Ön işleme adımı hazır haline getirilen veri için çalıştırılmak üzere modelin kurulması gereklidir. Model kurulum süreci probleme ilişkin olarak farklı olabilmektedir. Modelin kurulması aşamasına gelindiğinde uygun model seçilerek kurulabilmesi için projenin amaçları iyi bir şekilde kavranıp açıkça belirtilerek ortaya konulmalıdır. Modelin çalıştırılabilmesi için çok sayıda algoritma mevcut olabilmektedir. (Tekerek, 2011) Çalışmaya ilişkin uygun modelin bulunabilmesi için uzun bir süreye ve çok fazla sayıda tekrara ihtiyaç duyulabilmektedir. Tekrar süreci uygun model bulununcaya kadar devam etmektedir.

2.1.7 Modelin Çalıştırılması

Hazır hale getirilen model kullanılarak çalıştırılır ve bir çıktı elde edilir. Bu model çıktısı doğrudan uygulamanın kendisi olabildiği gibi bir başka uygulamanın alt parçası da olabilmektedir. (Tüzüntürk, 2010) Genellikle çalışmaya ilişkin tahmin edici veya tanımlayıcı bir sonuç elde edilmektedir. Çalıştırılan modelden çıkarılan sonuç

incelenir ve değerlendirilir. Çıkarılan sonuçtan çalışma maliyetine karşılık daha yüksek bir fayda sağlanması beklenmektedir.

2.1.8 Değerlendirme

Veri madenciliği aşamalarının tamamlandığı süreçtir. Artık elde edilen çıktının değerlendirilmesi gereklidir. Kullanılan yöntemlerin ve araçların zaman ve maliyet gibi faktörlerle değerlendirilmesinin yapılması gereklidir. Mevcut yöntem ve araçların hangi şartlarda farklı çalıştığının gözlenebilir olması gereklidir. Mevcut çalışmanın farklı yöntemlere ve araçlara karşılık avantaj ve dezavantajlarının fark edilip değerlendirilmesi gereklidir.

2.1.9 Modelin İzlenmesi

Mevcut veriye ilişkin bir çalışma yapılmakta çıktı elde edilmekte ve değerlendirilmektedir. Ancak veri sürekli olarak artan bir değerdir. (Tüzüntürk, 2010) Modeldeki verilen giderek artmakta veya mevcut veriler elenmek zorunda kalabilmektedir. Değişen veriler nedeniyle bazı değerler aykırı veri kabul edilebilmektedir. Bu nedenler doğrultusunda sürekli olarak model takip edilmekte ve güncellemeler doğrultusunda yeni çalışmalar yapılabilmektedir.

2.2 Veri Madenciliği Uygulama Alanları

Veri madenciliği işletmenin yapısına göre birçok alanda kullanılmaktadır. Yakın bir gelecekte ise hemen hemen her alanda kullanılması beklenmektedir. Başlıca kullanım alanları pazarlama bankacılık ve sigortacılıktır. (Baykal, 2006) Diğer kullanım alanlarına ek olarak ise sağlık, perakende, giyim, eğitim, telekomünikasyon, müşteri ilişkileri yönetimi gibi alanlar da sayılabilmektedir. Teknolojinin gelişmesi ve veri yığın birikiminin büyük ölçüde artması ile birlikte e-ticaret alanı da veri madenciliğinin yoğun olarak kullanıldığı alan haline gelmektedir.

2.2.1 Bankacılık

Veri madenciliğinin başladığı esas durumlardan biride müşteri ilişkileridir. Müşteri ilişkilerinin en yoğun olduğu alanların başında da bankacılık gelmektedir. Çünkü bankacılığın temelinde doğru müşteriye doğru ürünü doğru zamanda sunmak gereklidir. (Tüzüntürk, 2010) Bankacılıkta yapılan her alışveriş her telefon görüşmesi, yapılan nakit transfer işlemleri işlemi gerçekleştiren kişi bilgileri, kısacası her işlem

kayıt altına alınmaktadır. Bankacılıkta veri madenciliğinde yapılan uygulamaların başında kredi risk yönetimi, davranışsal skorlama analizi, kredi skorlama analizi portföy yönetimi müşterilerin kredi-urun kullanma sıklığının belirlenmesi gibi yöntemler bankacılığın temelinde araştırılan yöntemlerdir. Fayda sağlanan durumların başında ise şunlar gelir imza tanıma, kredi uygulamaları, sahtekarlıkların tespit edilmesi, kredi kartı kullanımı ve sıklığının öğrenilmesi müşteri davranışları belirlenerek olağan dışı hareketlerin fark edilip güvenliğin sağlanması, müşterinin aldığı krediyi geri ödeyebilme durumlarının belirlenebilmesi gibi ihtiyaç durumlarında veri madenciliği fayda sağlamaktadır. (Çoşlu, 2013) Ayrıca risk analizleri ve usulsüzlük tespiti de veri madenciliği uygulamaları sayesinde sağlanmaktadır. Yeni müşteri kazanma mevcut müşterinin devamlılığını sağlama, kredi kartı dolandırıcılıklarının önüne geçilmesi açısından da önemlidir (Köseoğlu ve Demirci, 2017) Bu açıdan müşteri bilgileri ve davranışları bankacılıkta her işlem için kayıt altına alınmaktadır.

2.2.2 Sağlık

Veri madenciliği her alanda etkin olduğu gibi sağlık alanında da oldukça etkili olmaktadır. Sağlık verileri yapısal olmayan metinlerdir ve değerli bilgileri içermektedir. Bu metin yapısındaki bilgilerin birbirleriyle olan ilişkileri araştırılarak bilgi keşfi yapılması gereklidir. (Yıldırım ve vd., 2008) Bu çalışmalara ve bu çalışmaların sağladığı faydalara birkaç örnek şu şekilde verilebilmektedir. Kişilerin genetik alışkanlıkları ve sağlık durumları incelenerek kişilerin prostat kanserine yakalanma olasılıkları hesaplanabilir. Performans değerlendirme yöntemleri kullanılarak sağlık alanındaki işletmenin başarısı artırılabilen, finansal kriz olasılıkları değerlendirilip kötü gidişat öngörülerek kriz öncesi önlemler alınabilmektedir. (Koyuncugil ve Özgülbaş, 2009) Salgın hastalık dönemlerinde hastalık yayılma hızı hastalığa yakalanma olasılığı ölüm ve iyileşme oranlarının tahmin edilmesinde de veri analizleri yapılabilmektedir. Benzer hastalıklara sahip kişilerin ortak özellikleri üzerinde inceleme yapılarak o hastalıkla ilgili risk taşıyan kişilerin hastalığa sahip olma olasılıkları üzerinde veri madenciliği çalışmaları yapılmaktadır. Tedavi süreçlerinde uygulanacak yöntemler sonucundan ne kadar başarılı olunacağını tahmin edilmesinde de veri madenciliği uygulamalarından yararlanılmaktadır. Kişinin demografik bilgileriyle de bir ilişkilendirme kurularak

yakalanabileceği hastalıklar ile ilgili bir ilişki çıkartılabilmekte ve olası hastalıklar üzerinde önlemler alınabilmektedir. (Yıldırım ve vd., 2008) Hastanenin maliyet tahminlerinde büyüme politikalarında atılması gereken adımların öngörüsü de bu çalışmalar neticesinde elde edilebilmektedir. Hasta memnuniyet anketleri analiz edilerek misyon ve vizyon politikaları üzerinde çalışmalar yapılabilmektedir. Özetle veri madenciliği her alanda etkili olduğu gibi sağlık alanında da birçok konuda etkili ve fayda sağlayan bir durum oluşturmaktadır.

2.2.3 Eğitim

21.yüzyıl ile birlikte artan veri madenciliği uygulamaları eğitim ve mühendislik alanında da etkili olmaktadır. Örneğin mühendislik alanında yeni bir ürün üretildiğinde veya mevcut ürünle ilgili geliştirme çalışmaları yapılacağı zaman veri madenciliği uygulamalarından yararlanılmaktadır. Ürünün üretilmesinden sonrasında karşılaşılabilecek olası sorunlar saptanmak ve azaltılmak istenmektedir. (Savaş ve vd., 2012) O ürüne ilişkin talep arz dengeleri rekabet piyasasındaki edineceği yerler hangi fiyat bazı ve hangi segmentlerde daha yüksek kar sağlanacağı bu bilgi keşfi çalışmaları sonrasında bilinebilmekte veya yaklaşık bir saptanma olarak öngörülebilmektedir. Eğitim alanında başarıyı etkileyen temel faktörlerin belirlenip bu faktörlerinde başarı üzerinde ne denli etkili oldukları belirlenmeye çalışılmaktadır. Sonrasında ise bu faktörler üzerinde çalışmalar yapılarak olumsuz faktörler azaltılıp başarı oranlarının artırılması sağlanmaktadır. Ancak bunun yanında başarının da ne olduğunun tanımlanabilmesi gerekli olmaktadır. Mevcut başarı düzeyinin ne olduğu nasıl tanımlanabilir olduğunun açıklanması ve kabul edilebilir sınır aralıklarının belirlenmesi gereklidir. Belirlenen bu sınır aralıklarının da nitel ve nicel olarak ifade edilip kabul edilen sınırlara uygun olan durumların başarılı kabul edilmesinde bu bilgi keşfi analizlerinden faydalanılmaktadır. Daha net açıklanırsa başarı konusunda etkili 3 faktörün beslenme, spor ve yaşam alanı olduğunu varsayabilmekteyiz. Bu faktörlerin ne kadar etkili olduğu hangi şartlar altında daha başarılı olunduğu belirlenebilmekte ve bu durumlar için şartlar iyileştirilebilmektedir. Eğitim alanında bir konuyu öğretebilme konusunda birçok farklı metot ve teknik bulunmaktadır. Bu metotların ne kadar faydalı olabileceği metotların birbirlerine kıyasla performans açısından ne kadar başarılı olabilecekleri hangi durumlar arasında hangi metodun daha etkin olabileceğinin belirlenebilmesi içinde veri madenciliği uygulamaları kullanılmaktadır.

Bir sonraki aşamada ise başarılı olarak kabul edilen sınırlar içerisinde kalan kişileri sahip oldukları ortak özellikler üzerinde çalışmalar yapılabilir. Bu ortak özelliklerin hangisinin başarı konusunda gerekli oldukları tespit edilebilir. Bir sonraki aşamada bu ise bu özelliklere sahip kişilerden kimlerin başarılı olabileceği öngörüsü için çalışmalar yapılabilir.

2.2.4 Pazarlama

Pazarlama Bir ürünü doğru bir kitleye doğru bir zamanda uygun bir bütçe ile sunmayı hedeflemek yani müşterinin karşılaştığı tanıtım karşısında yüksek bir oranla geri dönüş alabilmeyi hedeflemektir. Bu açıdan müşterilerin demografik kimlikleri ile sunulacak ürün ya da hizmet ilişkilendirilmelidir. Örneğin kişinin cinsiyeti yaşı veya yaşadığı bölgenin iklimleri baz alındığında sunulması gereken veya sunulmaması gereken temel ürün ve hizmetler olabilir. Açıklanan ayırım sonrasında yapılan daraltma pazarlama oranını iyileştirecektir. Pazarlamanın en önemli hedeflerinden biri ise talep arz dengesi oluşturabilmektir. Bir ürün talepten fazla üretildiğinden depolama ve elde kalma maliyeti oluşturur. Satılan ürünün kar getirisine bakılmaksızın talepten fazla ürün sunulması zarara neden olabilmektedir. Diğer bir durumda ise talepten daha az ürün üretilmesi sağlanacak fayda konusunda geride kalmaya neden olabilmektedir. Bu noktada veri madenciliği uygulamaları bir öngörü oluşturmaktadır.

2.2.5 E Ticaret

Veri madenciliği uygulamalarının en etkin olduğu alanlardan biride e-ticaret alanıdır. E ticaret müşterinin alışverişini hizmet sağlayıcı ile temas halinde bulunmadan teknolojik araç ve uygulamalarla sağladığı bağlantıdır. Kişiler artık bir ürün için herhangi bir mesafe almadan istedikleri ürün ve hizmeti e-ticaret faaliyetleri ile elde edebilmektedir. E ticarete her alıcı bir kimlikten ve yaptığı her işlem bir veri bir eylem oluşturmaktadır. Bu işlemler büyük sunucularda depolanmakta ve saklanmaktadır. Bu depolanan veriler gerekli çalışmalar yapıp anlamlandırıldığı zaman e ticaret faaliyetleri ile ilgili bilgiler ve pazarlama stratejileri hakkında bilgiler verir. Her kullanıcının kendine ait demografik coğrafik bilgileri bulunmaktadır. Ayrıca her kullanıcının kendine ait ilgi alanları ve bu alanlara ilişkin bütçe gibi kısıtları da mevcut olabilir. Veri madenciliği uygulamaları ile bu girdi ve çıktıları bir arada işleyerek faydalı sonuçlar elde edilebilir. E ticaret firmaları doğru pazarlama, doğru

strateji ve uygun hizmet önerileri ile e-ticaret faaliyetlerini geliştirirken yüksek oranda olumlu geri bildirim alarak daha büyük bir işletme haline gelebilirler. İlişki analizleri, çoklu uyum analizi ve K-Means metodu e-ticaret alanında en çok kullanılan araç ve tekniklerden birkaçıdır. (Özseven ve Ersöz, 2016)

2.2.6 Sigortacılık

Sigortacılık bir müşterinin ürün veya hizmetiyle ilgili karşılaşılabileceği risk faktörünü üstlenmektir. Pazarlama stratejisinde olduğu gibi doğru ürün doğru müşteriye doğru zamanda ürün risk müşteri çerçevesinde yüksek olumlu geri bildirim alınacak şekilde sunulmalıdır. Veri madenciliği uygulamaları bu süreç bütününe yardımcı olmaktadır. Veri madenciliği uygulamaları sigortacılık alanında artık kullanılmaması neredeyse imkânsız olmaktadır. Geçmişte veri madenciliği uygulamalarından faydalanmayan firmalar yığılan veri kontrolünü sağlayamamakta verilerini işleyememekte ve elindeki verilerden anlamlı çıkarımlar yapamamaktadır. Bu süreçleri etkin kullanmayan firmalar gelişen teknolojilere de kolay entegre olamamaktadır. (Akpınar, 2018) Sigorta firmaları her alanda olduğu gibi müşteri kitlesini büyütmek isterler. Mevcut müşterilerine daha fazla ürün sunabilmek. Mevcut müşterilerini kaybetmemek yani müşterilerinin sürdürülebilirliğini sağlamak isterler. Firmalar riske ilişkin ürün ya da hizmetlerine bir fiyat belirlemelidir. Ancak her müşteri aynı riske aynı bütçeye veya aynı ürün ya da hizmete sahip değildir. Bu açıdan firmalar müşterilerini çok iyi tanımaları gereklidir. Hangi müşterinin ne kadar yüksek risk potansiyeline sahip olduğunun bilinmesi gereklidir. Risk faktörüne bağlı olarak müşterilerine ürün ya da hizmet bedeli çıkarabilmesi sigorta firması açısından oldukça önemlidir. Örneğin hayat sigortası sunan bir sigorta firması ortalama yaşam süreleri hakkında kişinin genetik veya kronik rahatsızlıkları hakkında fikir sahibi olması ve bu verileri işleyerek uygun politikalar ve sınırlar çizmesi gereklidir. Bu da firmanın doğru ürün doğru müşteriye doğru zamanda uygun bütçe ile sunulması açısından iyileştirmeler sağlamaktadır. Bu yapılan çalışmalardan birkaçı müşteri bölümlenmesi, birliktelik analizi risk oranının yüksek olduğu müşteri örüntülerinin belirlenmesi örnek olarak gösterilen çalışmalardır. Bu çalışmalar neticesinden müşteriye kişiselleştirilmiş ürün veya hizmet sunumu da yapılabilmektedir. Ayrıca sunulan ürün ya da hizmete ilişkin sigortacılık hizmetleri şahıs veya firmalar tarafından art niyetli fayda sağlayıcı bir tavır neticesiyle suistimal edilebilmekte ve edilmektedir. Veri madenciliği

uygulamaları ile sigorta dolandırıcılık tespitleri yapılmaktadır. Veri madenciliği uygulamaları neticesinde yapılan analizle ile kötü niyetli eylemler belirlenebilmekte ve sigorta firmaları bunlara ilişkin haklarını koruyabilmektedir. Ancak veri politikası nedeniyle müşteriye ilişkin verilerin paylaşılabilmesi ve işlenebilmesi açısından bazı kısıtlar mevcut olabilmektedir. Özetle veri madenciliği uygulamaları sigorta firmalarının yeni müşteri elde edip mevcut müşterilerin daha fazla hizmet alıp sürdürebilirliğini sağlaması açısından oldukça önemli ve uygulaması zorunlu haline gelen çalışmalardır.

2.2.7 Telekomünikasyon

Bankacılık, sigortacılık, pazarlama, e-ticaret alanlarında olduğu gibi telekomünikasyon alanı da rekabet piyasasının yoğun olarak yaşandığı bir alandır. Ancak diğer alanlardan farklı olarak telekomünikasyon alanı müşteri geçiş ve kayıplarının daha hızlı yaşandığı alanlardan biridir. Telekomünikasyon rakip firmalardan müşteri kazanımlarının sağlandığı mevcut müşterilerin devamlılığının sağlanması amacıyla yoğun olarak faaliyet gösterirler. Telekomünikasyon alanında faaliyet gösteren firmaların temel amacı yeni müşteriler kazanmaktır. Ancak bu yeterli bir hedef olmamaktadır. Ürün ve hizmet sonrası müşteri memnuniyetin sağlanması mevcut müşterilerin devamlılığını sağlamak açısından oldukça önemlidir. Ayrıca rakip firmaların mevcut kullanıcıları da bünyesine katmak istemektedir. Her kullanıcının kendine ait demografik, coğrafik birtakım bilgileri mevcuttur. Bu bilgilere ek her kullanıcının yine kullanım geçmişi ve kullanım alışkanlıkları bulunmaktadır. Depolanan kullanım alışkanlıklarını içeren bu veriler sayesinde analizler yapılarak kullanıcılara ilişkin hizmetler sunulabilmektedir. Telekomünikasyon alanında mevcut müşterilere ilişkin daha yoğun çalışmaların olması yeni müşteri kazanımlarının mevcut müşteri devamlılığı sağlamaktan daha maliyetli olmasıdır. (Uyanık ve Kasapbaşı, 2021) Ayrıca firmalar müşterilerini rakip firmalara karşı kaybetme riskinden kaynaklı çalışmalar yapmak zorundadır. Telekomünikasyon firmaları veri analizleri yaparak müşterilerine uygun hizmetler sağlayıp kayıpların önüne geçmek kullanıcı devamlılığını amaçlamaktadır. Bu noktalarda veri madenciliği uygulamaları firmalara bir öngörü sunmaktadır. Telekomünikasyon alanında kullanıcıya bir ürün ya da hizmeti sunmak yeterli değildir. Ürün kullanımı ya da hizmet süresi boyunca karşılaşılan problemlerde destek sağlamak karşılaşılabilecek problem sayısını da en aza

indirmek gereklidir. Karşılaşılan arıza ya da aksaklıklar müşterilerin güven kaybı yaşamasına neden olur. Süreç devamlılığı da müşteri kayıplara haliyle gelir kaybı yaşanmasına da neden olur. Ürün ve hizmet kullanım sırasında yaşanan aksaklık ve arızalara ilişkin veri madenciliği çalışmaları bulunmaktadır.

2.2.8 Müşteri İlişkileri Yönetimi

Bir ticari faaliyeti oluşturan unsurlar işçi, işveren ve müşterilerdir. Bir işletme için müşteri hakkındaki sahip olunan bilgiyi temel bakış açılarıyla anlamlandırmaya çalışmak hem oldukça zor hem de oldukça uzun süren bir süreçtir. Özellikle teknolojinin gelişmesi ve e-ticaret faaliyetlerinin oldukça artması müşteri bilgi ve alışkanlıklarıyla ilgili yoğun bilgi birikmesine neden olmaktadır. Firmaların bu bilgi yığını işleme geleneksel yöntemlerle mümkün olmadığından artık müşteri ilişkileri yönetiminde de veri madenciliği uygulamalarından faydalanılmaktadır. Müşteri ilişkileri yönetimi İngilizce kısaltması ile CRM müşteriye ilişkin bilgilerin depolanmasını bilgileri toplanan müşterilerin hareket ve alışkanlıkları ile ilgili de bilgilerin depolanarak saklanması sağlar. Ancak bu depolanan bilgiler oldukça yer kaplayabilmektedir. Bu bilgilerin veri madenciliği teknikleri ile gelişmiş analizler yapılarak bir bilgi çıkarımının sağlanması gereklidir. Bu eylemlerin sağladığı birtakım avantajlar bulunmaktadır. Müşterilerin daha iyi tanınmasını sağlamaktadır. Benzer müşterilerin bir arada gruplanmasını sağlamaktadır. Müşteri bölümlenmesi yapıp aynı müşteri grubundan yer alan kişilere benzer pazarlama stratejileri uygulanabilmektedir. CRM ve veri madenciliği uygulamaları mevcut müşterilere ilişkin daha geniş bir perspektiften bakma şansı kazandırarak daha derin bir inceleme olanağı da tanımaktadır. Bu durumda müşterilere karşı daha iyi bir hizmet sunmayı sağlarken daha iyi bir hizmet de devamlılığı sağlanabilen müşteriler kazanmayı sağlamaktadır. Müşteri ve işletme arasında bir hizmet veya ürün ile ilgili bir iş akışı bulunmaktadır. Bu akış kimi işletmeler için oldukça maliyetli olabilmektedir. İşletme olarak bu akışları iyileştirmek maliyetlerin azalmasını sağlayabilmektedir. Ayrıca bu durumda daha fazla müşteriye hizmet sunabilme imkanını da beraberinde getirmektedir. CRM ve veri madenciliği yalnızca müşteri ve işletme arasındaki bağları iyileştirmek için kullanılmamaktadır. Müşteri ilişkileri yönetimi işletme içi organizasyon ve bilgi akış süreçlerini de iyileştirmektedir. CRM organizasyon içi iletişim süreçlerini hızlandırarak daha az maliyetli iş akışı daha verimli çalışan ve müşteri açısından daha

olumlu geri bildirimlerin alınmasını sağlamaktadır. Temel olarak hedef ve yapı farkı olarak birkaç sayıda CRM sistem türleri bulunmaktadır. Organizasyonel CRM türü bir iş akışının süreçlerini daha iyi hale getirmek için kullanılmaktadır. Pazarlama verilerini bir arada tutup analizini yapar. Müşterilere otomatik mesaj göndermek ve müşteri alışkanlık kayıtlarını tutmaktadır. İş birlikçi CRM ise ekipler arasındaki bilgi akışını kolaylaştırarak iletişimi sağlamaktadır. Gelişmiş analizler için verilerin toplanması ve derlenmesini sağlamaktadır. Müşteriye ait birçok bilgi, kriter ve davranış bulunmaktadır. Bunların bir arada tutularak düzenlenip işlenmesi gereklidir. Çünkü müşteri tüketim eğilimleri belirlenerek müşteriye ürün sunmak geri dönüşleri daha etkin hale getirir. Bu süreç içinse çoğunlukla analitik CRM kullanılmaktadır. (Wix Blog, 2021)

2.2.9 Risk Analizi

Riskin tanımı yapıldığında risk zarar veren hasar oluşturan kayıplara neden olan belirsizlik içeren durumlardır. Her oluşum veya her girişim mutlaka beraberinde bir belirsizlik ve riski getirmektedir. Risk kesin anlamıyla engellenemez ve her zaman bir olasılık dahilinde mevcut bir durumdur. Ancak oluşabilecek risklerin tanımlamalarını yapmak mümkündür. Çalışılan alana veya yapılan girişimi ilişkin olabilecek risklerin belirlenmesi ve tanımlanması zorunluluk içeren bir durumdur. Sonrasında belirlenen risklerin hangi durumlarda ne kadar olasılıkla karşılaşılabileceği de bilinmek istenmektedir. Bu olasılıklar sonrasında risk öncesi alınması gereken önlemler belirlenmektedir. Riskin gerçekleşme durumunda ise yapılması gereken stratejiler karşılaşılan kayıplar neticesinde öngörülen zararın azaltılması için de çalışmalar yapılmalıdır. Veri madenciliği uygulamaları diğer alanlarda olduğu gibi risk analizi çalışmalarında da etkin bir biçimde kullanılmaktadır. Geçmiş veriler kullanılarak karşılaşılabilecek riskler belirlenir o risklere ilişkin olasılıklar öngörülür. Öncesi için alınan önlemlerin ve risk gerçekleşme durumunda yapılacaklar veri madenciliği ile saptanmakta ve riske karşı zararlar minimize edilmektedir. Risk analizine ilişkin veri madenciliği ile yapılan çalışmalara ilişkin birkaç örnek şu şekildedir; Levent Sabah ve Hüseyin Bayraktarın yazdığı Gazi Mühendislik Bilimleri Dergisi'nde yayınlanan “Veri Madenciliği Birliktelik Kuralları ile Binaların Risk Durumlarının Analizi:” isimli makalede depreme karşı riskli binaların birliktelik kuralları ile analizi yapılmıştır. (Sabah ve Bayraktar, 2020)

2.3 Veri Madenciliğinde Kullanılan Araçlar

Her eylem beraberinde bir iz bırakmaktadır. Bu izler birikerek zamanla bir veri yığını oluşturmaktadır. Ancak teknolojinin gelişmesiyle bu veriler hayli bir büyük yer kaplamaktadır. Verilerin yazılması, sınıflandırılması, analiz edilmesi, sonuçlandırılıp değerlendirilmesi basit el yöntemleri ile çok zor, hatta mümkün olmamaktadır. Özellikle “BİG DATA” olarak adlandırılan büyük verilerin işlenmesinde mutlak olarak teknolojiden bilgisayar ortamından destek alınması kaçınılmaz bir durum haline gelmektedir. 21. yüzyılda ise teknolojik gelişmeler ile veri madenciliğinde kullanılan birçok program geliştirilmiştir. Donanımsal olarak ise işlemcilerin güçlenmesi ile büyük verilerin çözümü daha kolay hale gelmektedir. Ancak işlemcilerin güçlenmesi ile doğru orantılı olarak da işlenmesi gereken veriler daha da büyük bir hal alabilmektedir. Oluşturulan ve geliştirilmekte olan bu program ve yazılımlar ücretli ya da ücretsiz olmaları gibi, kapladıkları yer, kullanım kolaylığı, program arayüzü, ücretli olanların fiyatları gibi özellikleri ile değerlendirilmektedir. Uygulanacak tekniklere göre de kullanılan programlar değişiklik gösterebilmektedir.

2.3.1 Rapid Miner

Rapid Miner YALE Üniversitesinde bilim adamları tarafından geliştirilen bir araçtır. Dili javadır ve sayısız uzantılı dosyayı desteklemektedir. Bu özelliği ile diğer veri madenciliği araçlarından ayrılmaktadır. Rapid Miner veri madenciliğinin tüm aşamalarını desteklemektedir. Makine öğrenme algoritmalarını, karar ağaçlarını kümeleme algoritmalarını ve 400’den fazla algoritmayı desteklemektedir. Ayrıca büyük miktardaki verileri işleyebilmektedir. Oracle Microsoft SQL Server Postrage SQL ve My SQL veri tabanlarından R dosya aktarılabilir. (Tekerek, 2011) Rapid Miner kullanıcılara hem kısıtlı olan ücretsiz hem de profesyonel olan ücretli versiyonunu sunmaktadır. Dortmund teknoloji üniversitesinde uygulama ile aynı isimdeki firma ile geliştirilmiş bir yazılımdır. Uygulamayı diğerlerinden ayıran en önemli yönü çıktıları görselleştirebilme açısından zengin olmasıdır. İş ve Ticari amaçlarının dışında eğitim ve araştırma amacı ile de tercih edilmektedir. Veri madenciliğini aşamalarından olan veri hazırlama sonuçları görselleştirme doğrulama ve optimizasyon süreçlerini destekler görselleştirir. Ayrıca bilgisayara kurulum süresinin çok hızlı gerçekleşmesi ve boyut olarak az (150mb civarı) yer kaplaması

diğer sağladığı avantajlardandır. (Yıldız ve Şeker, 2016) Rapid Miner’ın en önemli özelliklerinden biri ise Marketplace geliştiricilerinin veri analiz algoritmalarını oluşturur ve bunları topluluğa yayınlanması için bir platform sağlamaktadır. (Cerebro, 2018)

2.3.2 Knime

Knime Almanya’da Konztanz Üniversitesinde biyoenformatik ve bilgi madenciliği bölümü tarafından geliştirilen birçok yöntemin entegre edildiği bir veri madenciliği aracıdır. (Tekerek, 2011) Ücretsiz açık kaynak kodlu bir yazılımdır. Görselleştirme açısından oldukça zengin olup makine öğrenmeleri içinde uygun olan bir araçtır. Ortalama bir kurulum hızı ve ortalama bir alan kaplamaktadır. (Şeker ve Yıldız, 2016) Genellikle veri madenciliği adımlarını 3 aşamadan oluşturmaktadır. Bunlar ekstraksiyon, dönüşüm ve yüklemedir. Verileri türüne göre sınıflar sonrasında anlamlandırır. İkinci olarak ise istenilen veya kullanılan teknik ile verileri bir düğüm haline getirir. İlgili teknik ile belirtilmesi gereken veriler varsa ayrıca belirtilir. Düğüm haline getirildikten sonra çıktı alınabilmesi verinin büyüklüğü kullanılan teknik ve çalışmanın yapıldığı işlemci türüne göre değişiklik göstermektedir. Ayrıca Veri madenciliği için çeşitli bileşenleri entegre ederek iş zekâsı ve finansal veri analizinin önünü açmaktadır. Yanlış bir çıktı elde etmemek için başlangıçta girilen verinin yazılımın doğru entegrasyonunun sağlayıp sağlamadığını kontrol etmek daha doğru çıktı elde etmemizi sağlamaktadır. Bu iki veri ve tekniğin bir arada olması düğümler sayesinde olmaktadır. Bu düğümlerin birleşebilmesi için de bir grafik kullanıcı arabirimi sağlamaktadır. Son olarak ise çıktının gösteriminin yapılması gereklidir. Yoğun olarak sonuçları renkler, grafikler, koordinat düzlemleri, tablolar halinde problemin teknik ve kullanıcı tercihlerine göre karşımıza çıkarmaktadır. Programın kolay kullanım arayüzü, ücretsiz olması, hızlı sonuçlar üretebilmesi, basit donanım gereksinimlerinde bile çalışabilir olması ve çok amaçlı kullanılabilir oluşuyla, zengin teknik barındırması nedeniyle oldukça bilinen ve tercih edilen bir yazılımdır. Java kodludur ve çekirdek sürümünde birçok veri entegrasyon modülü bulunmaktadır. (Cerebro, 2018)

2.3.3 Weka

Weka ilk amacı tarımsal faaliyetleri analiz edip yönetmek amacıyla 1993 yılında Waikato Üniversitesi tarafından oluşturulmuştur. 1997 yılında Java tabanlı sürüme geçildikten sonra ise birçok alanda kullanılmaya başlanmıştır. (Şeker ve Yıldız, 2016) Weka açık kaynak koda sahip olup java dilinde yazılmıştır. Weka'nın en önemli özelliklerinden birisi makine öğrenmesi sağlayabilmekte ve ardından bunları geliştirebilecek bir yapıya sahip olmasıdır. Weka da mevcut algoritmalar doğrudan bir veri setine uygulanabilmekte ya da java kodu içerisinde işlenebilmektedir. Weka veri işleme, sınıflandırma, regresyon kümeleme, ilişki kuralları ve görüntüleme araçlarını içermektedir. (Tekerek, 2011) Weka diğer veri madenciliği araçlarına göre daha az sayıda algoritma içerse de kullanım arayüzü kolaylığı nedeniyle daha ön plana çıkmaktadır. Ayrıca kapladığı alan ve kurulum süresi açısından oldukça alternatif bir veri madenciliği aracıdır.

2.3.4 Tanagra

Tanagra Fransa'da Lumiere Üniversitesinde akademik ve araştırma amacıyla Ricco Rakotomala tarafından geliştirilmiştir. Tanagra birçok veri madenciliği görevini diyagram içerisinde görsel olarak tasarlayarak gösterir. Görsel arayüzü diğer veri madenciliği araçlarına oranla daha basittir. İstatistiksel yaklaşımları, çok değişkenli analiz yöntemlerini ve makine öğrenme tekniklerini içermektedir. (Cerebro, 2018)

Tanagra ücretsiz açık kaynak kodlu herkesin erişim sağladığı bir veri madenciliği aracıdır. Ayrıca Tanagra kişilerin kendi veri madenciliği uygulamalarını ekleyip geliştirebileceği bir araçtır. Tanagra SIPINADAN sonra geliştirilen daha güçlü bir araçtır. Fakat diğer yazılımlara kıyasla geniş bir sete veri kaynağına sahip değildir. (Tekerek, 2011)

2.3.5 Elki

Elki Almanya'nın Münih şehrinde bulunan Ludwig Maximilian Üniversitesi'nde Profesör Hans ve Peter Kriegel'in geliştirmiş olduğu java ile yazılmış bir araçtır. Elki açık kaynak kodlu bir araçtır. Elki'nin kurulumu çok hızlı gerçekleşmekte ve diğer araçlara karşın çok az yer kaplamaktadır. Elki diğer araçlara

kıyasla veri madenciliği algoritmalarının geliştirilmesi ve bunların değerlendirilebilmesini sağlamaktadır. Elki veri tabanı diziniyle veri madenciliği algoritmalarının etkileşimlerini sağlamaktadır. (Şeker ve Yıldız, 2016)

2.3.6 Massive Online Analyss

Massive Online Analyss (MOA) Weka ile aynı üniversitede ve aynı ekip tarafından geliştirilmiş bir yazılımdır. Açık kaynak kodlu bir yazılım olup birçok yazılımda olduğu gibi dili javadır. MOA büyük veri analizleri ve zamana bağlı biriken verileri işlemek için geliştirilmiştir. Büyük verileri işlemesine karşın yazılım çok az yer kaplar ve kurulumu oldukça hızlı gerçekleşmektedir. En sık kullanılan veri jeneratörlerini içermektedir. (Şeker ve Yıldız, 2016)

2.3.7 Orange

Orange 1996 yılında Ljubljana üniversitesinde yapay zekâ araştırmaları yapan bir ekip tarafından geliştirilmeye başlanmış bir yazılımdır. Açık kaynak kodlu yazılım olan Orange veri görselleştirme, makine öğrenmesi ve veri madenciliği araç setlerini sunan bir yazılımdır. Orange 1996’da geliştirilmeye başlanmış ve yıllar içerisinde sürekli bir gelişim göstermiştir. En büyük gelişimini 2008 yılında Mac OS DMG ve Fink tabanlı kurulum paketlerini gerçekleştirerek yapmıştır. (Tekerek, 2011) Birçok yazılımda olduğu gibi açık kaynak kodlu bir yazılım olup dili ise Python’dur. Yazılım dilinin Python olması nedeniyle script yazılabilmektedir. Orange kurulumu hızlı gerçekleştirilen ve çok az yer kaplayan bir yazılımdır. Çıktıları görselleştirme imkânı sunmaktadır. Biyoenformatik ve metin madenciliği için eklentileri bulunmaktadır. Ayrıca makine öğrenmesini de içermektedir. Hızlı ve geniş bir nitel veri analizi sunan araç veri hazırlama, keşifsel veri analizi modelleme gibi imkanları da sağlamaktadır. (Tekerek, 2011)

2.3.8 KEEL

Java dilinde kodlanan KEEL (CSIC) desteği ile İspanya’da bulunan GRANADA üniversitesinde geliştirilmiştir. (Şeker ve Yıldız, 2016) Çok hızlı bir şekilde kurulumu sağlanmakta ve çok az yer kaplamaktadır. Kümeleme gibi klasik veri madenciliği uygulamaları açısından zengin değildir. Ayrıca veri görselleştirme açısından diğer araçlar kadar geniş bir görselleştirmeye sahip değildir. Fuzzy

sınıflandırıcılar, yapay zekâ tabanlı sınıflandırma ve kural tabanlı kümeleme algoritmalarının birçoğunu içermektedir. (Kaya ve Özel, 2014)

2.3.9 R

R Yeni Zelanda Auckland Üniversitesinde geliştirilmiştir. R & R olarak da bilinmektedir. Windows Mac OS Linux ve Unix işletim sistemleri içinde uygundur. R grafikler istatistiksel hesaplamalar ve veri analizler için geliştirilmekte olan bir araçtır. Kullanım için uzman desteği gerekli olabilmektedir. (Tekerek, 2011) Diğer veri madenciliği uygulamalarının aksine çok fazla tercih edilmemektedir. (Kaya ve Özel, 2014)

2.4 Veri Madenciliği Teknikleri

Veri madenciliği tanımı literatüre girdiğinden biri veri analizlerine ilişkin birçok yöntem ve algoritma bulunmakta ve bulunan bu yöntem ve algoritmalar zaman içerisinde geliştirilmektedir. Kimi yöntem ve algoritmalar zaman içerisinde geçerliliğini yitirmekte olsa da bazı algoritmalar bilim ve teknolojinin geldiği son noktada da popülerliğini korumaktadır. Veri madenciliği yöntemleri literatürde birçok farklı model ve şekilde ana ve alt başlıklara ayrılmaktadır. Ancak geçerliliğini en çok koruyan sınıflandırma şekli veri madenciliği tekniklerini tahmin edici ve tanımlayıcı modeller olarak ayıran sınıflandırma şeklidir. Geçerliliğini koruyan şekle göre veri madenciliği teknikleri tahmin edici ve tanımlayıcı modeller olarak ana başlıklara ayrılmaktadır. Tahmin edici modeller daha önce sonuçları bilinen verilerden yola çıkarak modeller oluşturmak ve bilinmeyen veri setleri hakkında değerlere ilişkin tahminlerin yapılması olarak tanımlanmaktadır. Tanımlayıcı modeller ise bir durumla ilgili karara ilişkin çıkarım sağlamaya yönelik modellerdir. Tahmin edici ve tanımlayıcı modeller kendi içerisinde alt başlıklara ayrılmaktadır.

Tahmin edici modeller sınıflandırma ve tahmin, regresyon modelleri ve zaman serisi analizi algoritmalarından oluşmaktadır. Bu tahmin edici modellere ilişkin en çok kullanılan yöntemler karar ağaçları, KNN yöntemi ve Navie Bayes sayılabilmektedir. Tanımlayıcı modeller ise kümeleme yöntemi, birliktelik kuralları gibi alt başlıklara ayrılmaktadır. Bu bölümde en sık kullanılmakta olan veri madenciliği teknikleri anlatılmaktadır. (Alaybeg, 2019)

2.4.1 K Means Yöntemi

K Means yöntemi bilinen en eski veri madenciliği tekniklerinden birisidir. Hızlı çalışması ve çoğu veri madenciliği araçlarında kullanılabilir olması büyük avantajlarındandır. Büyük verilerin çözülmesinde de büyük avantaj sağlamaktadır. K Means yönteminin temel amacı ise bir veri yığınının birbirine yakın olan verileri kümelemek gruplamaktır. K Means sayısız veri için k adet merkez belirlenerek kalan verilerin k merkezlere yakın olan verilerle kümelendiğini sağlamaktadır. (Yıldız ve vd., 2019) K Means belge sınıflandırma müşteri segmentasyonu gibi alanlarda yoğun olarak kullanılmaktadır. Dolandırıcılık tespiti ve suç yerlerinin analizinde K Means kullanılmaktadır. Oyuncu analizi değerlendirmesi gibi durumlarda da K Means yönteminden faydalanılmaktadır. Çağrı kaydı detaylarının analizi ve BT uyarılarının otomatik olarak kümelendiği K Means yönteminin kullanıldığı diğer alanlardır. (Ay, 2019)

K Means yöntemi ise şu şekilde çalışmaktadır. Önce verilerin kaç gruba bölünmek istendiği belirlenmektedir. Bu sayı k ile ifade edilmektedir. Sonra belirlenen sayı kadar rastgele olarak grup merkezleri seçilir. Öklid veya diğer uzaklık ölçüm teknikleri kullanılarak verilerin grup merkezlerine olan uzaklığı ölçülür. Her bir veri kendisine en yakın grup merkezine atanmaktadır. Bu adımlar bütün veriler bir merkeze atanana kadar ve aynı küme gruplarında artık değişiklik göstermeyene kadar tekrar etmektedir. (Yıldız ve vd., 2019) Bu işlemlerdeki temel amaç küme içi benzerliklerin yüksek seviyede olmasını kümeler arasında ise benzerliğin ise düşük seviyede olmasını sağlamaktır. (Sarıman, 2011) Başlıca kümeleme yöntemleri ise bölümleme, hiyerarşik kümeleme ve yoğunluk bazlı yöntemlerdir. (Ay, 2019)

2.4.2 KNN En Yakın Komşu Yöntemi

Hem regresyon hem sınıflama algoritması olan KNN yöntemi literatüre ilk kez 1950 yılında girmiştir. Girdiği dönemde yöntemin oldukça zaman alıcı olması ve veri madenciliği kavramının yeteri kadar bilinmemesi nedenleriyle tutan bir yöntem olmamıştır. Henüz atanmamış sınıfı bilinmeyen veriler için diğer bilinen veriler ile uzaklık ölçümü yapılarak diğer verilere en optimal sınıflar belirlenmektedir. KNN de temel amaç benzer veri ya da grupların aynı sınıfta yer alabilmesini sağlamaktır. KNN de sınıfı bilinmeyen veriler diğer veriler ile kıyaslanarak uzaklık ölçümleri

yapılmaktadır. Uzaklık ölçümlerinde çoğunlukla Manhattan, Öklid, Minkowski ve Chebyshev kullanılmaktadır. Atanmayan verilere ise en optimal sınıflar aranmaktadır. Her sınıflayıcı kendi içerisinde artık bir sınıflayıcı kullanmaktadır. Gelen her örnek için en yakın komşu kümesi tekrardan taranır. Ancak her bir veride tek tek tarama yapıldığından süreç uzun sürmektedir. Temel amaç ise optimal k sınıf değerini bulmaktır. Ancak K değeri gerekenden büyük olursa çok benzer olmayan verilerin bir arada olmasını yani doğruluk değerinin düşmesini sağlayabilmektedir. Küçük değerlerde ise bir arada olması gereken değerlerin sınıf dışında kalmasına neden olabilmektedir. Her iki durumda olumsuz sonuç alınmasına neden olabilmektedir. (Dilki ve Deniz Başar, 2020) KNN örüntü tanıma, veri madenciliği, yapay zekâ ve istatistik alanlarında yoğun olarak kullanılmaktadır. Ayrıca bilişsel psikoloji, tıp, biyoinformatik gibi sağlık alanlarında da yoğun olarak kullanılmaktadır. KNN yönteminin avantajları arasından analitik bir yöntem olması uygulama açısından kolay ve uygulama süreçlerinin izlenebilir olması avantajlarındandır. Öğrenme açısından kolay ve gürültülü verilere karşı direnç göstermektedir. Yüksek belleğe ihtiyaç duyması işlem süresinin uzun olması ve maliyet açısından yüksek olması da dezavantajlarındandır. (Taşcı ve Oran, 2017)

2.4.3 Karar Ağaçları

Karar ağaçları bir olay üzerine yapılması gereken tercihin risklerini, kazançlarını ve hedeflerini tanımlamak ve bu tanımlamalar üzerinde doğru stratejiler geliştirmeyi hedeflemektedir. Bir yatırım, oyun veya herhangi olay belirli bir miktar girdi bedeli içermektedir. Bu girdi bedeline karşılık olarak yapılan yatırım ya da oyunun getiri beklentisi bulunmaktadır. Ancak bir oyun kazanç getirisinin yanında kayıp riskini de içermektedir. Ayrıca bu olaylara ilişkin birden fazla strateji bulunabilmektedir. Her stratejide kendi içerisinde farklı kazanç ve kayıp oranlarına sahip olmaktadır. Buda durumlara ilişkin olayların iç içe dallanarak uzamasına neden olur şekilsel olarak ifade edilmek istenen bu durum ağaca benzemektedir. Sınıflandırma ve regresyon problemlerinde tercih edilen karar ağaçları karmaşık veri setlerinde de kullanılabilmektedir. Karar ağaçları 3 bölümden oluşmaktadır. Bunlar kök, düğümler ve yapraktır. Düğümlerde kendi içinde ayrılır. Kök başlangıçtaki karar ağaçlarının hücrelerine denmektedir. Hemen altında ise düğümler bulunmaktadır. Düğüm sayısı problemin karmaşıklığı ile doğru orantıdadır. En altta da sonucu veren

alıřılmıştır. (Bardi ve Can, 2021) Yakut ve Elmasın yayınlanan bir alıřmasında da diskriminant analizi kullanılarak bařarısızlıklar zerine bir alıřma yapılmıştır. (Yakut ve Elmas, 2008) 2018 yılında yayınlanan İslamođlu ve ankaya’ ya ait makalede ise bu sefer finansal bařarısızlık modelleri kullanılarak finansal bařarının tahmini amalanmıştır.

2.4.5 Birliktelik Kuralı

Veri madenciliđi denildiđinde akla gelen ilk tekniklerden olan birliktelik analizinin temelleri 1993 yılında Agrawal tarafından atılmıştır. Birliktelik analizinin amacı bir destek ya da gven eřiđi etrafında sık tekrarlanan iřlem ve veriler zerinde tespit yapılarak onların analiz edilmesini sađlamaktır. (Sabah ve Bayraktar 2020) Birliktelik analizin genel tanımında ise birlikte alınan rnlerin birbiriyle olan iliřkisini incelemek vardır. Birliktelik analizinde hangi rnlerin yanında hangi rnlerin daha sık tercih edildiđinin tespiti yapılarak rn dizilim planı ve rnlere iliřkin promosyon stratejileri geliřtirilmektedir. Aynı sreler ile gerekleřen olayların ve nemli iliřkilerin birbirleriyle olan bađlantısı da incelenip ıkarım sađlanmaktadır. Web tıklanma analizleri iin de birliktelik analizi kullanılmaktadır. Ayrıca bir olaya iliřkin risk durumlarının analizinde de kullanılabilmektedir. Birliktelik analizi pazarlama, finans, sađlık sigorta gibi alanlarda sıka kullanılmaktadır. zetle bu alıřmalar sayesinde gemiřteki verilerin analizi yapılarak geleceđe ynelik nlenebilir veya artırılabilir yeni alıřmalar yapılmaktadır. Firmalar birliktelik analizini kullanarak satın alınan rnlerin birliktelik iliřkisini tanımlayabilir tketicisi satın alma alışkanlıklarını tanımlayarak byk karlar elde edebilmektedir. Birliktelik analizi iin kullanılan algoritmalarından bazıları ise FP-Growth, GRI, Carma Sequence ve en sık kullanılan apriori algoritmalarıdır. (Uslu, 2018)

Trkede nsel anlamına gelen “priori” kelimesinden oluřturulan Apriori algoritması 14 yy. da Albert le Grand de Saxe tarafından temelleri oluřturulmuřtur. Literatrde birliktelik analizi ile yapılan alıřmalardan ise Levent Sabah ve Hseyin Bayraktar’ın 2020 yılında yayınlandıđı alıřması “Veri Madenciliđi Birliktelik Kuralları ile Binaların Risk Durumlarının Analizi” rnek gsterilebilir. (Sabah ve Bayraktar 2020) 2019 yılında yapılan bir alıřma ise Alan ve Yeřilyurt’ a ait

“Birliktelik Kuralları Madenciliği ile Yatan Hasta Profiline Çıkarılması” diğerk bir örnek çalışmadır. (Alan ve Yeşilyurt 2019)

2.4.6 Genetik Algoritmalar

Veri madenciliği analizlerine ilişkin kullanılan yöntemlerden bir diğeri ise genetik algoritmalarıdır. Genetik Algoritmalar ilk defa John Holland ve arkadaşlarının canlılardaki genetik yaşam sürecini bilgisayar ortamında canlandırmayı düşünmesiyle başlamıştır. Genetik algoritmalar doğadaki evrimsel süreçleri mühendislik alanındaki problemlere uyarlama yaparak daha optimal bir çözüm elde etmeyi hedefleyen bir algoritmadır. Genetik algoritmalar birçok alanda karmaşık problemleri çözmek için kullanılmaktadır. Finans alanında veri madenciliğinden iflas tahminlerine, gelir kaybı nedenlerinden yatırım araçlarının performansını belirlemeye kadar bu yöntemden yararlanılmaktadır. Pazarlama alanında mevcut verileri çok iyi analiz etmek ve uygun stratejileri belirlemek pazarlamadaki etkinliği göstermektedir. Bu süreçlerde yine bu yöntemden yararlanılmaktadır. Üretim sektöründe birçok temel problemin çözümünde de genetik algoritmalar kullanılmaktadır. Montaj hattı dengeleme, tesis yerleşimi, çizelgeleme, hücresele üretim, atama, taşıma, gezgin satıcı problemleri gibi problemlerin çözümünde genetik algoritma kullanılmaktadır. Diğerk bir kullanım alanı da bilgi sistemleridir. Bilgi sistemleri üretim süreçlerinde büyük bir role sahip olsalar da arızalanmaya ve etkisiz kalma durumuna açıktırlar. Sistemi oluşturan parçanın güvenliği, verimliliği, arızalanma olasılık ve sürelerinin belirlenmesinde kullanılmaktadır. (Emel ve Taşkın 2002). Genetik algoritma parametreleri kodlama, eşler ve bireyler oluşturma, mutasyon uğratma ve bu adımların tekrarlanması gibi adımlardan oluşmaktadır. Parametreleri Kodlama aşamasında verilerin bir kod haline getirilmesi gereklidir. Bu aşamalarda kodlama için bitlerden alfabeledeki harflerden, sayılardan veya problemin gerçek reel değerlerinden faydalanılabilmektedir. (Çolak, 2010) Parametreler kodlandıktan sonra çözüme ilişkin rastgele bireyler yani eşler üretilerek amaç fonksiyonları hesaplanmaktadır. Üretilen bireylerden daha kötü sonuç verenler elenmektedir. (Özdağlar ve Kahraman, 2004) Elenen bireylerden sonra sezgisel bir yaklaşım kullanılarak DNA’lar mutasyona uğratılıp daha ideal bir sonuç bulunmaya çalışılır. Oluşan çözüm alternatifleri yani DNA’da değişikliklerle amaç fonksiyonlarında iyileştirme sağlanması amaçlanır. Bu tüm aşamalar belirlenen veya istenen idealliklere kadar tekrar tekrar uygulanabilir. İstenilen değerleri vermeyen

DNA'lar elenir ve kalanlara mutasyon uygulanır. (Çolak, 2010) Bitirme adımına gelindiğinde sonlanmayı belirleyen birkaç durum bulunmaktadır. Bunlardan ilki ideal çözümün bulunmasıdır istenilen düşüklük veya yükseklikte değeri veren DNA'ya erişilmesi durumudur. Benzer sonuçlarının sürekli tekrar etmesi yani adım ilerletildikçe istenilen hassasiyette fark oluşturan sonuçlar elde edilememesi tekrar adımını bitirebilmektedir. Belirli bir aşamadan sonra harcanan maliyetin istenilen optimalliğin üzerine çıkmasının öngörülmesi ve başlangıçta belirlenen hedef tekrara ulaşılması bitirme adımını oluşturan nedenlerdendir. Genetik algoritma ile ilişkin literatürden Arif Gülten ve Şengül Doğan'a ait "Genetik Algoritmalar Yönteminin Biyomedikal verileri üzerindeki uygulamaları" isimli çalışması bulunmaktadır. (Gülten ve Doğan, 2008) Murat Kahraman ve Davut Özdağlar su dağıtım ağının oluşturulmasında yani tasarımında genetik algoritma ile uygulama yapmış ve olumlu kabul edilen bir sonuç elde etmişlerdir. (Kahraman ve Özdağlar, 2004) Bahadır ve arkadaşları ise alternatiflerden tesis yer seçimini bu teknikle belirlemiştir (Gülsün vd. 2009) Yusuf Şahin ve Kenan Karagül yöntemin farklı bir çeşidi olan melez akışkan genetik algoritma ile gezgin satıcı probleminin çözümünü gerçekleştirmiştir. (Şahin ve Karagül, 2019) Yine öğretim görevlisi Selçuk Çolak 2010 yılında benzer bir gezgin satıcı problemini çözmüştür. (Çolak, 2010) Günümüzde popüler olarak kabul edebileceğimiz bir çalışma olarak insansız hava araçlarının hedeflere yönlendirilmesinde genetik algoritma yöntemi kullanılmıştır. Bu çalışma Yüzbaşı Baha Pakkan ve Murat Ermiş tarafından hazırlanmıştır. (Pakkan ve Ermiş, 2010) Farklı bir çalışmada ise yöntem dalgacık paket dönüşümü ile birlikte kullanılarak zayıf olan radar sinyallerinin gürültüden arındırılması sağlanmıştır. Çalışma Mehmet Üstündağ ve arkadaşları tarafından yapılmıştır. (Üstündağ vd. 2014)

2.4.7 Yapay Sinir Ağları

Gözetimli öğrenme algoritmalarından olan yapay sinir ağları insan beynini örnek olarak öğrenme sürecinin modellenmesini esas almaktadır. Yapay sinir ağları otomotiv havacılık ve üretim alanlarında kullanılmaktadır. Sesli asistan, duygu analizi gibi doğal dil işleme alanında da yapay sinir ağları modelinden faydalanılır. Tıp ve biyoloji alanlarında oldukça fayda sağlayan yapay sinir ağları ilaç keşfi tümör bulma gibi konularda da fayda sağlamaktadır. Finans ve bankacılık da bir diğer kullanım alanlarındandır. Ayrıca yapay zekâ uygulamalarında da yapay sinir ağları

kullanılmaktadır. İnsan beyninde temel işlevi bilgi aktarmak olan milyarlarca nöron bulunmaktadır. Her bir sinir hücresi de 50.000 ile 250000 arası nöron ile bağlantı kurmaktadır. Yapay sinir ağları üretebilme, oluşturabilme ve keşfedebilme yeteneğine sahiptir. Sinir ağının bu yapısı taklit edilerek yapay bir tasarım oluşturulur. Yapay sinir ağlarından bir giriş ve çıkış oluşturularak belirli kurallar ışığında tasarım uygulanır. Yapay sinir hücresi 5 bölümden oluşmaktadır. Bunlar girdiler, ağırlıklar, toplama fonksiyonu, aktivasyon fonksiyonu ve çıktılardır. (Yıldırım, 2020) Yapay sinir ağları 3 katmandan oluşmaktadır. Bunlar giriş katmanı, gizli katman ve çıktı katmanıdır. Giriş katmanı dışardan bilgileri alıp gizli katmana aktarmaktadır. Gizli katman giriş katmandan gelen bilgileri alıp işleyerek çıktı katmanına göndermektedir. Çıktı katmanı ise gelen bilgileri işleyerek çıkış verisini oluşturmaktadır. Bu bilgelere ek olarak yapay sinir ağları geleneksel programlama ve yapay zekâ yöntemlerinden farklı bilgi işlem yöntemlerine sahip olması açısından makine öğrenme süreçlerinde de kullanılabilmektedir.

2.4.8 Aykırılık Analizi

Aykırı veri ya da diğer anlamıyla uç değer diğer gözlemlerden kayda değer derecede uzak olan değer demektir. (Dayanaklı, 2021) Aykırılık analizi de verinin diğer kalan verilerden önemli derecede farklı veya sapmaya sahip olması durumunda belirlenip elenmesi süreci olarak tanımlanmaktadır. (Doğan ve Kevser, 2021) Veri analizlerinde temel amaç girdiden anlamlı çıktılar oluşturabilmektir. Önışleme adımı da veri analizinin hem önemli hem de zaman alan süreçlerindendir. Sağlıklı bir çıktı için önışleme adımlarının doğru bir şekilde gerçekleştirilmesi gereklidir. Aykırılık analizi ise doğru önışleme adımları için gerekli bir aşamadır. (Taş, 2019) Aykırı verinin ayıklanmaması sonuçta alacağımız çıkarımlar için sapmalara neden olabilmektedir.

Aykırı değer türleri nokta aykırı değerler, bağlamsal aykırı değerler ve toplu aykırı değerler olarak ayrılmaktadır. Ayrıca aykırı değer türleri tek değişkenli aykırı türleri ve çok değişkenli aykırı türleri olarak da sınıflandırılabilir. Aykırılık analizinde tek değişkenli aykırılık analizleri ve çok değişkenli aykırılık analizleri için ayrı yöntemler bulunmaktadır. Tek değişkenli aykırılık analizine ilişkin grafiksel yöntemler olarak kutu grafiği ve histogram örnek verilebilmektedir. Tek değişkenli

aykırılık analizinde istatistiksel yöntemlere çeyrekler açıklığı, ortanca mutlak sapma, Z Skoru Hampel Filtresi verilen örneklerden olmaktadır. Çok değişkenli analizlerde ise izolasyon ormanı en sık kullanılmakta olan yöntemlerden birisidir. Aykırılık analizi aşamalar şeklinde oluşabilmektedir. Aykırılık analizi aşamalarında ilk önce aykırı verilerin analiz sonucu tespit edilerek silinmesinden başlanmaktadır. Gerekli olması durumunda aykırı değer yerine değer atanması ile devam edilmektedir. Değerlerin gözlenen sınırlar içerisinde olmasına dikkat edilmektedir. Son aşamada ise gözlemler analize ilişkin uygun formatlarda dönüştürülmesi ile tamamlanmaktadır. (Dayanaklı, 2021) Aykırılık analizinden sağlanan çıkarım yalnızca aykırı verinin belirlenip ayklanması değildir. Aykırılık analizinde aykırı veri aranan değer hedefin kendisi olabilmektedir. Diğer verilerden anlamlı derecede farklı olan veriler sürece ilişkin aranan değerlerde olabilmektedir. Bu doğrultuda saldırı tespit yöntemlerinde, kredi kartı dolandırıcılıklarını önlemede, suçla mücadelede aykırılık analizinden faydalanılmaktadır. Ayrıca tıbbi tanı ve bilimsel çalışmalarda da aykırılık analizi uygulamalarından faydalanılmaktadır. Aykırılık analizinde verinin türü, büyüklüğü, modelden sağlanmak istenen çıkarımlar model seçimini etkileyen faktörlerdendir. (Şirin,2018)

2.4.9 Aşamalı Kümeleme

Veri analizi tekniklerinden olan aşamalı kümeleme verilerin birbirine olan benzerlikleri dikkate alınarak belirli düzeyde bir araya getirilmelerini hedeflemektedir. Aşamalı kümeleme yöntemleri birbirine yakın olan değerleri değişik aşamalar ile bir araya getirerek kümeler oluşturmaktadır. Aşamalı kümelemede yöntemler ardışık bir biçimde gerçekleştirilmektedir. Bu ardışık yöntemler hangi verinin hangi kümeye ait olduğunu ve hangi düzeyde kümenin özelliklerine yakın olduğunu belirtmektedir. Aşamalı kümeleme analizinde başlangıçta belirlenen hedef durumuna ve belirlenen kriterlere ilişkin aşamalı yöntemler birleştirici aşamalı kümeleme yöntemleri ve ayırıcı aşamalı kümeleme yöntemleri olarak ikiye ayrılmaktadır. Birleştirici aşamalı kümeleme analizleri daha yaygın olarak tercih edilmektedir. (Kayaalp vd., 2000) Aşamalı kümeleme analizi çözüme ilişkin aşamalarda hiyerarşik olmayan yöntemler ile başlamakta ve hiyerarşik olan yöntemler ile devam etmektedir. Hiyerarşik olmayan yöntemlere k ortalamalar yöntemi örnek verilirken hiyerarşik olan yöntemlere en küçük varyans tekniği örnek verilmektedir. Ayrıca aşamalı kümeleme analizi küme

sayısında ön bilginin olmadığı durumlarda kullanılabilir. İşleyişinde kategorik ve sürekli değişkenlere yer verilebilen bir teknik olmaktadır. (Giray, 2016) Literatürde Ceylan ve arkadaşlarının yaptığı “İki Aşamalı Kümeleme Analizi ile Bireysel Emeklilik Sektöründe Müşteri Profiline Değerlendirilmesi” isimli çalışma bu tekniğe ilişkin örnek bir çalışmadır. (Ceylan ve vd., 2017)

2.4.10 Faktör Analizi

Faktör analizi çok fazla sayıda veri ya da değişkenin bir arada olduğu durumlarda birbirleriyle ilişkisi olan değerlerin bir arada toplanılması ve faktör oluşturulmasıdır. İlişkili veriler bir kategoride toplanarak değişken sayısından daha az sayıda faktöre ulaşılması hedeflenmektedir. (Büyüköztürk, 2002) Faktör analizinin birden fazla temel amacı bulunmaktadır. Oluşan değişkenleri yorumlayarak veri setindeki deney birimlerin değerlendirmesini sağlayabilmek istenen hedeflerdendir. Korelasyonu olmayan veri setinin oluşturulması ve indirgenen değişkenlerin tanımlanmak istenmesi de faktör analizinin amaçlarındandır. Elde edilen değişkenleri diğer analizlerde kullanılması da diğer amaçlarındandır. (Stat, t.y)

Faktör analizi verileri korelasyonlu olarak bir kategoride toplayıp gruplandırarak indirgemektedir. Faktör analizi örnek olarak açıklanırsa veriye ilişkin 10 değer olduğu varsayıldığında bu verilerin alt başlıkta toplanarak indirgenmek istenmesidir. Belirtilen 10 veriden hangilerinin birbiriyle benzer olup o kategori altından toplanması gerektiği faktör analizi ile sağlanmaktadır. Faktör analizi aşamalardan oluşmaktadır. Faktör analizi varsayımların test edilmesi ile başlamaktadır. Varsayımların test edilmesi sürecinden sonra korelasyon matrisi ortaya çıkarılmaktadır. Hedefe yönelik faktör analizinin metodu belirlenmektedir. Faktör analizi için gerekli metot belirlendikten sonra faktör sayısı belirlenerek devam edilmektedir. Sonraki aşamada mevcut faktörler döndürülerek devam edilmektedir. Son aşamada ise elde edilen çıktı yorumlanmaktadır. Faktör analizi aşamasında dikkat edilmesi gereken unsurlar bulunmaktadır. Veri setinde aykırı değer bulunmaması gereklidir. Veri setinin normal dağılması ve ölçülebilir olması sağlıklı çıktılar için önemli varsayımlardandır. (Yılmaz, 2018)

ÜÇÜNCÜ BÖLÜM

3. METODOLOJİ

Bu bölümde kullanılan yöntemler hakkında bilgiler verilmektedir. Müşteri segmentasyonu üzerinde durularak kullanılan yöntemler ile ilgili literatürdeki örneklerle yer verilmektedir

3.1 Müşteri Segmentasyonu

Herhangi bir teorik ders ya da uygulama gerektiren bir süreci sınıf ortamında herkes içerisinde öğretilmesi ile kişiye özel birebir süreçler ile anlatılması arasında bilinen ve göz ardı edilemeyen farklar mevcuttur. Sınıf ortamında yetkinlikleri, konuya ilişkin mevcut seviyeleri, anlama ve kavrama kapasiteleri, ilgi alanlarının değişkenliği gibi farklılık gösteren durumlarda birçok insan bir arada bulunabilmektedir. Ayrıca öğretmen ve öğrenci ilişkisinde konuya ilişkin birebir bildirim ve geri dönüşlerin daha zayıf olduklarını görebilmekte mümkün olabilmektedir. Birebir etkileşimin daha maliyetli olduğu bilinmektedir. Ancak sınıf ortamının yerine birebir gerçekleşen eğitimin daha başarılı ve imkanlar dahilinde daha çok tercih edilen bir durum olduğunu da görebilmekteyiz. Ancak bu iki duruma ilişkin daha optimal bir yaklaşım gerçekleştirilen durumlar mevcut olmaktadır. Dil kurslarına bakıldığında öğrencilerin öncelikle mevcut seviyeleri belirlenmekte öğrenciler mevcut seviyelerine göre bir arada eğitim görmekte seviyelerine uygun nitelikte dersler verilmektedir. Aynı durum sınava hazırlık kurslarında da görülmektedir. Bu süreç eğitimin kalitesini arttırmakta yetkinlik ve seviye kaynaklı öğrenim süreç zorluklarını da daha kolaylaştırmaktadır. Tüm bu süreçleri işletme ve müşteri ilişkileri bakımından benzetebilmek aynı kalıplar üzerinde değerlendirebilmek mümkündür. Bir işletme müşterisine sunduğu hizmetten yüksek oranda geri dönüş ve kar sağlamak istemektedir. Bunun için işletme müşterisini tanımak zorundadır. Her müşterinin sahip olduğu kişiye özgü bilgileri mevcuttur. Kişisel bilgiler dışında ilgi alanları, sahip olduğu imkanlar, hizmete ilişkin tercihleri ve alışkanlıkları da mevcut olmaktadır. Bu bilgiler bize müşteriye tanıtan bilgilerdir. Bu bilgiler sayesinde müşteriye bir yaklaşım sergilenmekte ve bu yaklaşım neticesinde geri dönüşler alabilmekteyiz. Sahip olunan bilgilere karşılık nasıl tavır ve stratejik yaklaşımlar gösterilmesi de araştırılması gereken ayrı bir durum olmaktadır. Günümüzde teknolojinin ve lojistik faaliyetlerin

gelişmesi ile bazı işletmeler yüksek sayıda müşteriye sahip olabilmektedir. Müşteri işletme ilişkisinde en önemli durum bu ilişkinin devamlılığını sağlayacak yaklaşımlarda bulunmaktır. Bu açıdan işletme gerekli kriterler ölçüsünde müşterisini tanımak ve onlara karşı stratejik tavırlar geliştirmektedir. Ancak günümüzde işletmeler tüm müşterilerine benzer yaklaşımlar sergilemenin yeterli olmadığını bilirken her müşterisi için ayrı ayrı yaklaşım sergilemenin mümkün olmayacağını görebilmektedir. Aynı eğitim sürecinde olduğu gibi müşterilere ilişkin benzer bir yaklaşımın olması gerektiğini fark etmektedirler. İstenen ya da belirlenen kriterlere ilişkin benzer eylemlere sahip olan müşterileri bir arada görmek istemektedirler. Benzer davranış sergileyen müşterilere ayrı yaklaşım sergilemenin daha başarılı olacağını görebilmek mümkün olmaktadır. Bu süreçler bağlamında müşteri segmentasyonu tanımı oluşturulmakta ve oluşturulan bu tanıma ilişkin çözüm teknikleri ile çalışmalar yapılmaktadır. Müşteri segmentasyonun genel tanımı için herhangi bir pazarın özellik ve beklentileri dahilinde farklı gruplara bölünmesi olarak ifadesi kullanılabilmektedir. (Euromsg Express,2022) Özetle müşteri segmentasyonu müşterileri tanımamızı sağlarken mevcut bilgileri ile ilişkilendirerek davranış ve alışkanlıklarını görmemizi sağlamaktadır. Bu süreçte bize müşterilerimize nasıl bir yaklaşım içerisinde olmamız gerektiği ile ilgili çıkarımlar sunmaktadır. Bu çıkarımlar sayesinde müşterilere sunulan hizmet ya da ürünler ile ilgili olarak daha yüksek oranda geri dönüş alırken, geleceğe yönelikte gerçekleştirmemiz gereken faaliyet ve değişiklikler hakkında fikir sahibi olmamızı sağlamaktadır.

Hedef değişkenliklerine göre sık kullanılan 5 segmentasyon çeşidi bulunmaktadır. Bunlar demografik segmentasyon, psikolojik segmentasyon, davranış segmentasyonu, coğrafik segmentasyon ve değer segmentasyonudur. Demografik segmentasyon hemen hemen her sektörde kullanılan müşterileri tanımamızı sağlayan bir segmentasyon çeşididir. Demografik segmentasyon bireysel veya kurumsal olmasına göre farklı kriterler üzerinden değerlendirilmektedir. Bireysel demografik segmentasyonda yaş, cinsiyet ve meslek durumu gibi kriterler mevcutken kurumsal demografik segmentasyonda şirket tipi ve çalışan sayısı gibi kriterler etkin olmaktadır. Psikolojik segmentasyon da müşterilerin yaşam tarzı, karakter tipi, öncelikler, değerler gibi kriterlerine göre değerlendirmekte ve çıkarımlar sağlamakta olan segmentasyon çeşididir. Müşterilerin alışveriş sıklığına ve bütçe aralıklarına göre yaklaşımda

bulunmamızı sağlamaktadır. Davranış segmentasyonu müşteri alışkanlıklarını gözlemleyerek davranışları özellikleri açısından bölümlememizi sağlayan segmentasyondur. Davranış segmentasyonun da kullanım davranışı, iletişim davranışı ve marka güveni kullanılan kriterlerdendir. Çapraz satış faaliyetlerinin sunulabilmesi için önemli segmentasyon çeşitlerindendir. Coğrafik segmentasyon uluslararası alanda faaliyet gösteren şirketler için önemli bir segmentasyon çeşididir. Coğrafik segmentasyon çeşidin de şirketler müşteri yoğunluklarını ve dağılımlarını gözlemleyebilmektedir. İklim, şehir, bölge coğrafik segmentasyon kriterlerindendir. Şube ya da depo gibi yatırımların lokasyonunun belirlenmesinde fayda sağlamaktadır. Bir ürün ya da hizmetin farklı bölgelerdeki iklim nedeniyle farklı şekillerde sunulması gerekli olabilmektedir. Bu farklılığı anlayabilmemizde coğrafik segmentasyon büyük oranda etki göstermektedir. Değer segmentasyonunu ise diğer 4 segmentasyonun bir arada bulunabildiği karışım bir segmentasyon olarak değerlendirebilmekteyiz. Finansal açıdan girdi ve çıktılarında eklenmediği segmentasyon çeşididir. Belirli aşama ve süreçlerden geçerek segmente oluşturmaktadır. (Şener, 2020)

3.2 Literatür ve Yöntem

Problemin tanımına ilişkin bakıldığında perakende sektöründeki bir firma elindeki müşteri alışkanlıklarına karşılık bir stratejik tavır sergilemek istemektedir. Yüksek ölçekte müşteri sayısına sahip firma her müşteriye aynı yaklaşımının sergilenmesini yeterli görmezken her müşterisi için ayrı ayrı bir tutum sergilemenin de mümkün olamadığını bilmektedir. Alışveriş sayısı, alışveriş sayısına ilişkin ziyaret sıklığı ve müşteri tutarı gibi kriterler üzerinden benzer tutum ve alışkanlıklar içerisindeki müşterilerini bir arada tutmak istemektedirler. Bir arada benzer alışkanlıklara sahip müşterilerin nasıl bir ortak alışkanlık sergilediklerini görmek tanımlamak ve bu tanımlamalar üzerinden pazarlama stratejileri uygulamak ve bu stratejilerden yüksek oranda olumlu geri bildirimler almak istemektedirler. Mevcut veri ve istenen hedefler doğrultusunda literatür araştırması yapılmakta ve benzer problemlerde ve istenen hedeflere ilişkin olumlu sonuçları veren yöntem ve araçlar incelenerek bir çalışma süreci belirlenmektedir. Müşteri ilişkileri yönetimine ilişkin yapılan çalışmalar incelendiğinde bazı sonuçlar elde edilmektedir.

Mihrimah Özmen ve arkadaşlarının yaptığı telekomünikasyon alanındaki müşteri bazlı çalışması incelenmiştir. Çalışmanın hedeflerine bakıldığında mevcut problemimize ilişkin benzer hedefler bulunmaktadır. Müşterilerin birbirlerinden farklı tercih ve alışkanlıklarının olduklarını görmektedirler. Bu bağlamda benzer tutum ve alışkanlık sergileyen müşterilerinin bir arada tutarak tavır sergilemek istemektedirler. Bu doğrultuda müşterilerini bölümlenmek istemektedirler. Çalışma olarak müşterileri için bir kümeleme çalışması gerçekleştirmiş müşterilerini bölümlenmişlerdir. (Özmen ve vd., 2018) İlhami Tuncer ve Kazım Karaboğa 2021 yılında perakende sektöründeki müşterilerine ilişkin bir çalışma gerçekleştirmişlerdir. Mevcut yoğun müşteri verilerinden müşterilerini kar oranlarına göre gruplamak yüksekten düşüğe doğru kar sağlayan müşterilerini bir arada tutmak gruplamak istemektedir. Çünkü artan müşteri sayısı farklı tercih ve alışkanlıkların oluşmasına neden olmakta firmalarda müşteri tercih ve alışkanlıkları benzer olan müşterilerini bir arada görmek ve bu müşterileri tanımlayarak gruplara farklı stratejik yaklaşımlar geliştirmek istemektedirler. Çalışmada RFM tekniklerini kullanarak bir bölümlenme gerçekleştirilmiş ardından K-Ortalamalar K-Means yöntemi kullanılarak devam edilmiştir. K-Means yönteminin geçmişi, kullanım alanları aşamaları gibi süreçler anlatılarak yöntemin temel amacı üzerinde durulmuştur. (Tuncer ve Karaboğa, 2021) Bankacılık alanı müşteri ilişkileri yönetiminin en yoğun olduğu alanlarından başında gelmektedir. Bankalar müşterilerini iyi tanımlamak tanımladığı müşterilerini de doğru bir şekilde yönlendirerek hedef sunmak istemektedirler. Bu bağlamda 2016 yılında yayınlanan Aslı Çalış ve Kasım Baynal'a ait çalışma içerisinde satış stratejilerinin belirlenmesine yönelik bir çalışma yapılmıştır. Bankacılık alanında yapılan veri madenciliği çalışmaları anlatılarak uygulama gerçekleştirilmiştir. Gerçekleştirilen uygulama da kümeleme analizleri anlatılarak üzerinde durulmuştur. Çalışma içerisinde kümeleme analizlerimden K-Means yöntemi üzerinde durularak anlatılmış ve çalışmada K-Means tekniği kullanılmıştır. (Çalış ve Baynal, 2016) Yine müşteri ilişkilerine dayalı olan Gamze Ödev ve Serhat Peker'in 2022 yılında yayınlanan bir çalışması bulunmaktadır. Farklı firmaların müşterilerine ait şikayetleri bulunmaktadır. Firmaların şikayetler üzerinden değerlendirilmesi performanslarının kıyaslanması istenmektedir. Yine bu çalışmada veri madenciliği kümeleme tekniklerinden olan K-Means yönteminden faydalanılmış ve bir sonuç elde edilmiştir. Firmalar aldığı şikâyet

sayısı, şikayetlere dönüş sayıları ve çözüm üretme kriterlerine göre değerlendirilmeye alınmıştır. (Ödev ve Peker, 2022) Muhammed Bektaş'a ait bir internet sitesindeki çalışmada müşteri segmentasyon çalışması yapılmıştır. Müşteri segmentasyon çalışmasında kümeleme yöntemlerinden faydalanılmıştır. Kümeleme yöntemlerinden 6 farklı yöntem ile çalışma yapılmış çalışmalardan elde edilen değerler ile kıyaslama yapılmıştır. Bu çalışmada kullanılan K-Means, MiniBatch K-Means, Gaussian Mixture Model, MeanShift, Hierarchical Clustering ve DBSCAN kümeleme yöntemleri kullanılmıştır. Davies Bouldin ve Silhouette Skorlarına göre en iyi sonucu veren yöntem K-Means yöntemi olarak belirlenmiştir. (Bektaş, 2020) Farklı bir internet sitesindeki Emre Sırma'ya ait yazıda bir e-ticaret firmasının müşterilerini segmente etmesi üzerinde durulmuştur. Firmanın müşterilerini segmente ederek pazarlama stratejileri oluşturmak istemektedir. Çalışma da veri ve veriye ilişkin kriterler anlatılarak veri ön işleme aşamasından geçirilmiş ve kümeleme yöntemi olarak K-Means yöntemi kullanılmıştır. (Sırma, t.y)

3.3 K-Means

Denetimsiz algoritma tekniklerinden olan K-Means ilk defa Mac Queen tarafından geliştirilmiştir. Denetimsiz öğrenme modeli küme içerisindeki sınıflandırılmamış ya da etiketlenmemiş verileri tanımlamak için kullanılan yöntemlerdir. (İnnova, 2021)

Denetimsiz öğrenme yöntemleri denetleme gerektirmeyen modellerdir. Makine öğrenmesi olan denetimsiz öğrenme yöntemleri kendi başına çalışabilmektedir. Denetimsiz öğrenme yöntemleri veri içerisinde bilinmeyen kalıpları bulabilmektedir. Ayrıca denetimsiz öğrenme modelleri kategorizasyon için fayda sağlayabilecek özellikleri görmemizi de sağlamaktadır. Denetimsiz öğrenme algoritmaları verileri gerçek zamanlı olarak işleyebilmektedir. Tüm girdi verilerinin analiz edilerek etiketlenmesini sağlamaktadır. Etiketlenmeyen verileri bilgisayarda elde etmek elle müdahale gerektiren etiketli verilerden daha kolay olmaktadır. (Ay,2019)

Denetimsiz öğrenme çıktılarının doğruluğu her zaman kesin değildir. Bu durum denetimsiz öğrenme çıktılarının dezavantaj oluşturan durumlarından biridir. Kümeleme yöntemleri genelde iki başlık altında toplanmaktadır. Bunlar hiyerarşik

kümeleme ve hiyerarşik olmayan kümeleme yöntemleridir. Hiyerarşik kümeleme yöntemleri küme sayısının bilinmediği görsel olarak karar verilen kümeleme yöntemleridir. Hiyerarşik kümelerde veriler birleştirici veya ayırıcı bir biçimde oluşturulmaktadır. Hiyerarşik olmayan kümeleme yöntemlerinde küme sayısı ile ilgili bilgi veya küme sayısı üzerinde çalışmaların yapılabildiği kümeleme yöntemleridir. En sık kullanılan hiyerarşik olmayan kümeleme yöntemi K-Means olmaktadır. (Erdem, 2021)

K-Means yöntemi belirli sayıdaki veriden oluşan bir veri kümesini K sayısınca kümeye bölümlenmek olarak tanımlanabilmektedir. K-Means 'in temel amacı işlem sonrasında elde edilen kümeler sonucunda küme içindeki verilerin benzerliklerinin en yüksek seviyede kümeler arası benzerliklerin ise en az seviyede olmasıdır. K-Means yöntemi en sık kullanılan kümeleme algoritmalarındandır. Uygulama süreçleri oldukça kolay olmakta büyük verileri ise hızlı ve etkin bir şekilde işleyebilmektedir. K-Means iteratif tekrarlayıcı bir sürece sahip olmaktadır. Bu durum nedeniyle her aşamada her verinin ait olduğu kümeye olan uzaklıklarını küçültmektedir. Özetle K-Means algoritması karesel hatayı en küçük yapacak olan K adet kümeyi tespit etmeye çalışmaktadır. (Vikipedi, t.y)

K-Means aşamalarına gelindiğinde ise K-Means tekrarlayıcı aşamalardan oluşmaktadır. İlk önce küme sayısını ifade eden K sayısı belirlenmektedir. K sayısı rastgele belirlendiği gibi analitik çalışmalar sonucunda da elde edilebilmektedir. K sayısı yani küme sayısı belirlendikten sonra aynı sayı kadar veri içerisinde rastgele grup merkezleri belirlenmektedir. Kalan tüm veriler nitel veya nicel olmalarına göre uzaklık ölçüm yöntemleri kullanılarak küme merkezlerine olan uzaklıkları ölçülmektedir.

Her veri kendisine en yakın olan küme merkezine atanmaktadır. Bu atamanın sonunda tekrar küme merkezleri belirlenerek uzaklık ölçümleri gerçekleştirilir ve yeni atamalar yapılır. Bu iteratif tekrarlayıcı süreçler küme verileri aynı kalana kadar tekrar edilmektedir. Sonuçta istenen K sayısında küme elde edilmektedir. (Aydın ve Başkır, 2012) İteratif süreçlerde sonlanmayı belirlemek için optimizasyon kriteri oluşturmak kullanılan yöntemlerden birisidir. (Mesci, 2020)

$$J(D, C) = \sum_{x_i \in D} \min_{c_i \in C} \|c_i - x_i\|^2 \quad (1)$$

K-Means kümeleme yöntemi müşteri ilişkilerinin yoğun bulunduğu bankacılık, sigortacılık, pazarlama, perakende, e-ticaret gibi alanlarda kullanılmaktadır. K-Means yöntemi adli süreçlerde de kullanılmaktadır. Suç yerleri ile işlenen suçlar arasındaki ilişki K-Means yöntemi ile incelenebilmekte ve fikir edinebilmesi sağlanabilmektedir. Adli vaka kaydı bulunan kişiler içerisinde kümeleme yapılarak suç ve kişiler arasındaki ilişki görülebilmektedir. Dolandırıcılık tespiti K-Means yöntemleri ile incelenerek anlaşılabilir. Sigorta ile ilgili dolandırıcılıklar 2.el ürün satımına ilişkin analizlerin tespiti verilen ifadelerdeki gerçeklik uyumları K-Means yöntemi ile araştırılabilir. K-Means performans değerlendirmelerinde oyuncu analizlerinde kullanılabilir. İstatistiklerin görülmesi ve analiz edilmesinde K-Means 'den faydalanılır. Telekom şirketleri tarafından müşteri ile sağlanan iletişim bilgileri saklanmaktadır. Demografik bilgiler ile ilişkilendirilen iletişim bilgileri müşteri ihtiyaçları hakkında bilgi verebilmektedir. Denetimsiz öğrenme yöntemlerinde verilerin gerçek zamanlı işlenmesi de bu süreçler açısından büyük avantaj sağlamaktadır. K-Means yönteminin en etkin olduğu alanlardan bir diğeri ise müşteri segmentasyonu alanıdır. (Ay, 2019)

Firmalar müşterilerinin davranış ve alışkanlıklarına ilişkin tanımlamaları yaparak onlara bir stratejik tavır sergilemek istemektedirler. Ancak yüksek sayıda müşteri portföyüne sahip alanlarda ve firmalarda bu geleneksel yöntemlerle mümkün olamamaktadır. Bu tarz işletmelerin genellikle istedikleri benzer davranış ve alışkanlıklara sahip müşterileri bir arada görebilmektir. Farklı davranış ve alışkanlıklara sahip müşterileri de birbirinden bağımsız incelemek istemektedirler. Bu açıdan K-Means yöntemin genel tanımına bakıldığında N adet verinin K sayıda kümelere ayrılması olduğu görülmektedir. K-Means deki temel amacında her küme içerisindeki veri benzerliklerin en yüksek seviyedeysen küme dışı veriler ile de en az seviyede benzerlik olması tanımı bulunmaktadır. Müşteri segmentasyonu temel prensibi ve K-Means yöntemi amaçları incelendiğinde yüksek oranda tanımlama uyumlamasını görmek mümkün olmaktadır. K-Means da verilerin küme merkezlerine olan uzaklıklarını ölçmede yoğunlukla Manhattan ve Öklid uzaklık ölçümleri kullanılmaktadır. Öklid uzaklık ölçümü gösterilmektedir. (Dicle ve Öztürk, 2012)

$$d_{ij} = \left[\sum_{k=1}^p (x_{ik} - x_{jk})^2 \right]^{\frac{1}{2}} \quad (2)$$

K-Means sonucunun başarısı da genellikle karesel hata değeri formülü ile değerlendirilir. İngilizce “Sum Of Squared Error” anlamına gelmektedir. En düşük karesel hata değeri en yüksek başarı değerini vermektedir. En düşük karesel hata değeri bir küme içerisindeki tüm verilerin küme merkezine olan uzaklığın karelerinin toplamı ile bulunmaktadır. Karesel hata değeri ile ilgili formül gösterilmektedir. (Işık ve Çamurcu)

$$SSE = \sum_{i=1}^k \sum_{x \in c_i} dist^2(m_i, x) \quad (3)$$

K-Means 'i diğer kümeleme yöntemlerinden ayıran en büyük özelliği birçok veri madenciliğine ait araç ve yazılım üzerinde kullanılabilmekte olmasıdır. Bu araç ve yazılımların bir kısmı da açık kaynak kodlu ve ücretsiz olmaktadır. Ayrıca çoğu kümeleme yöntemine göre daha büyük verilerde çalıştırılabilir ve çok daha hızlı sonuçlar verebilmektedir. Bu durumda teknolojinin gelişmesi ile artan veri yığını çağında da büyük avantaj sağlamaktadır. K-Means 'in temel prensibi olan küme içi verilerin yüksek seviyede benzerlikleri bulunurken küme dışı benzerliklerinde en az seviyede olmasıdır tanımı birçok işletmenin aradığı probleme ilişkin aranan temel cevapla örtüşmektedir.

3.4 Silüet Katsayısı

K-Means yöntemindeki K ifadesi istenen küme sayısını belirtmektedir. K-Means yönteminin ilk aşaması da K sayısını belirlemektir. K ifadesi kimi zaman rastgele belirlenirken kimi zaman da sağlanmak istenen faydayla ilişkili olabilmektedir. K sayısını belirlemek kolay gibi görülse de bazı durumlar için K sayısını yani küme sayısını belirleyebilmek K-Means yöntemi için büyük problemler oluşturan aşamalardan oluşabilmektedir. Bazı durumlarda K sayısını belirleyebilmek için yapılan çalışmalar K-Means yönteminin diğer süreçleri kadar zaman alıcı ve maliyetli olabilmektedir. Literatürde K-Means yöntemine ilişkin K sayısının belirlenmesi için yapılan çalışmalar ve yöntemler bulunmaktadır. K sayısı yani küme sayısına ilişkin en sık kullanılan yöntemler Elbow metodu ve Silhouette katsayısıdır. Diğer kullanılan yöntemlere ise başparmak kuralı, bilgi ölçütü yaklaşımı ve çapraz

doğrulama örnek olmaktadır. (Kodinariya ve Makwana, 2013:91, akt: Kılıç vd., 2020) Ayrıca GAP İstatistiği yöntemi de K sayısını belirlemede kullanılan yöntemlerden biri olmaktadır. (Gürler, 2022) Elbow metodu her K değeri için noktaların küme merkezlerine olan uzaklarının karesi toplamını hesaplamaktadır. Bu değere Within Clusters Sum of Square (WCSS) yani kümeler içi kareler toplamı denilmektedir. Her K değeri için hesaplanan bu değerler bir grafik haline getirilerek izlenir. Değerleri düşerken keskin bir kırılım gerçekleştiği değer en iyi K değeri olarak kabul edilmektedir. Bu problemin çözümüne ilişkin sezgisel bir yaklaşım şeklidir. (Şirin, 2017) Eğer elimizde mevcut veriler için temel doğruluk etiketleri var ise dışsal yöntemlerden faydalanabiliriz. Eğer bu doğruluk etiketleri elimizde mevcut değil ise içsel yöntemlerden faydalanmamız optimallik açısından daha doğru olabilmektedir. Silüet katsayısı yöntemi K değerini belirlemede kullanılan içsel yöntemlerden olmaktadır. Silüet katsayısı ilk defa 1987 yılında Rousseeuw tarafından ortaya konularak geliştirilmiştir. (Cengiz ve Öztürk, 2012) Silüet katsayısı için öncelikle bir veri noktası ile ait olduğu kümedeki diğer veri noktaları arasındaki uzaklığının ortalaması alınmaktadır. İkinci olarak o veriye ait en yakın kümedeki tüm verilere olan uzaklıklarının ortalaması alınmaktadır. Verinin kendi içerisindeki diğer verilere olan ortalama uzaklığını ve en yakın küme içerisindeki verilere olan ortalama uzaklığını bularak çözüm sağlanmaktadır. Bu bağlamda silüet katsayısı formülü şu şekilde olmaktadır. (Aydın ve Başkır, 2013)

$$sil(x_i) = \frac{b(x_i) - a(x_i)}{\max(a(x_i), b(x_i))} \quad (4)$$

Silüet katsayısı yönteminde her K değeri için kümelerin farklılığına bakılmakta Her bir verinin yer aldığı kümeye olan uygunluğu hesaplanmaktadır. Kullanılan formül neticesinde ise -1 ve 1 aralığında sonuçlar elde edilebilmektedir. -1 değerine yakın olan değerler K değeri için uygunsuzluğu ifade etmektedir. 0 değeri istenen K değerine ilişkin tam ortada bir ifade sunmaktadır. 1'e yakın değerler özellikle 0.5 üstü değerler bize K değeri için daha optimal bir tercih olduğunu sunmaktadır. Belirtilen K değerleri içerisinde 1'e en yakın değer K değeri için en uygun değer kabul edilebilmektedir. (Güven, 2019)

Silüet katsayısının diğer en önemli özelliği küme içerisindeki yakınlığın yani başarısının değersel olarak görebilmemizi sağlamaktadır. Silüet katsayısı K değerine

ilişkin bir değer bulmamızı sağlarken mevcut kümenin başarısını aynı zamanda validasyonunu da göstermektedir. Bu durum Silüet katsayısının K değerini bulmaya ilişkin diğer yöntemlerden ayıran en önemli özelliğidir.



DÖRDÜNCÜ BÖLÜM

4. UYGULAMA

Bu bölümde veri madenciliği yöntem ve araçları ışığında bir uygulama gerçekleştirildi. Uygulamaya ilişkin veri bir perakende sektörüne ait bir yıllık müşteri datasını içermektedir. Uygulama da Silüet Katsayısı ve K-Means yöntemleri kullanılarak bir yıllık data içerisindeki müşteri dataları kümelere ayrılarak değerlendirildi ve müşteriler segmente edilerek stratejiler oluşturuldu. Uygulama aşamalarında veri madenciliği araçlarından olan Rapid Miner ve Knime kullanıldı. Verinin tablolar halinde incelenebilmesi için excelden faydalanıldı.

4.1 Uygulamanın Amacı

Gelişen teknoloji, büyüyen lojistik faaliyetleri ile e-ticareti aktif olarak kullanmaya başlayan perakende sektörleri son dönemlerde hızlı bir büyüme gerçekleştirmiştir. Büyüme neticesinde müşteri portföyleri artmıştır. Her alanda olduğu gibi perakende alanında da müşterilere yönelik temel amaç daha fazla müşteri kazanmak, mevcut müşterileri daha aktif hale getirmek ve eldeki müşterileri de kaybetmemektir. Her müşterinin kendine özgü bilgileri bulunmaktadır. Kendine özgü bilgileri bulunan her müşterinin de alışverişine ilişkin davranış ve alışkanlıkları bulunmaktadır. Bilgiler ve davranışlar doğrultusunda müşterilere bir yaklaşım gerçekleştirilmesi gereklidir. Ancak müşteri sayısı ve müşteriye ilişkin verilerin yoğun bir şekilde artması ile gerçekleştirilmesi gereken tutum ve stratejilerin bireysel olarak değerlendirilmesi neredeyse imkânsız hale gelmiştir. Elimizde 1.000.000’u aşkın müşteri olduğu varsayıldığında tüm müşterilere aynı tutum ve stratejik yaklaşım içinde bulunmak fayda sağlamayacağı gibi 1.000.000 müşteri için ayrı ayrı yaklaşım sergileyebilmekte mümkün olamamaktadır. Bu açıdan süreçler değerlendirildiğinde yapılması gereken analitik çalışmalar oluşturulmaktadır. Her müşteriye ayrı yaklaşım sergilemek mümkün değildi ancak benzer bilgi, davranış ve alışkanlıklara sahip müşterileri bir araya getirip bütün içerisinde değerlendirmek hem mümkün hem de daha avantajlı bir süreç haline gelmemektedir. Bu durumu bir kütüphane de kitapların kategorik olarak yerleştirilmesi bir arada bulunması gibi düşünülebilmektedir. Bu amaçlar doğrultusunda veri madenciliği kümeleme tekniklerinden Silüet Katsayısı ve K-Means kullanılarak bir çalışma gerçekleştirilmiştir.

4.2 Veri İerięi ve Veri niřleme Adımı:

Uygulamaya iliřkin veri bir perakende sektrndeki bir yıllık mřteri alışkanlıklarından oluřmaktadır. Uygulama verisi 1048575 mřteri bilgisi ve alışkanlıklarını iermektedir. Mřteri bilgileri veri ierisinde bulunmamakta olup sadece id řeklinde gzkmektedir. Mřterilere iliřkin bilgileri ieren veriler ise mřterilerin alışveriş sayısını belirten frekans toplamı, mřterilerin en son ka gn nce alışveriş yaptıęı gn sayısını belirten son alışveriş tarihi, mřterilerin toplam harcama miktarını belirten mřteri alışveriş tutarı ve mřteri alışveriş tutarı ile frekans toplamının blnmesi ile elde edilen frekans bařına tutardır.

Tablo 4.2 Veriye İliřkin rnek Tablo

CRMID	FREAKANS TOPLAM	SON AV TARİH	FREKANS BASINA TUTAR	MUSTER AV TUTAR
899EBD73- F0F6-E811-	1	190	3	3
BFC331D2- 0572-E711-	1	108	3	3
BD05CDAA- 13EB-EB11-	1	235	3	3
F06BEDF7- 34D0-E711-	1	201	4	4
B6609877- B906-E911-	1	48	6	6
2E20C550- 0D97-EC11-	1	16	20	20
EB2CBAE5- 768D-EB11-	2	353	38	76

CRMID	FREAKANS TOPLAM	SON AV TARİH	FREKANS BASINA TUTAR	MUSTER AV TUTAR
66F935D0- 8B5D-EC11-	1	89	40	40
FE7EED2B- 8CDB-E811-	1	326	50	50
F2053787- 3D6B-EC11-	1	70	54	54
E9EB6477- 0EAB-E911-	2	285	56	112
A15A382C- 5E3A-E811-	1	291	60	60
B4FC9297- DBC6-E711-	1	206	60	60
7FAC6F65- D1C2-E611-	1	173	60	60
C5F27C6A- 5CD3-EB11-	1	265	60	60

Veri önişleme adımına gelindiğinde ise her kümeleme analizinde olduğu gibi bu çalışmadaki verilerde de elenmesi gereken veriler bulunabilmektedir. İlk bakışta yinelen verilerin olduğu görülmektedir. Veri içerisinde 21536 müşteri verisinin yinelandığı aynı müşterilere ait aynı verilerin bulunduğu fark edilerek bu veriler silinmiştir. Sonraki aşamada her kritere ilişkin minimum ve maksimum değerlerin ölçüm içerisinde yer alması ve değerlendirilmesinin istenmesi nedeniyle hiçbir veri aykırı kabul edilememektedir. Kalan 1027039 müşteri verisi ile çalışmaya devam edilmektedir. Çalışmanın devam eden aşamasında Silüet Katsayısı ile K sayısı belirlenerek K-Means kümeleme tekniği ile 1027039 veri kümelerine ayrılmaktadır.

Tablo 4.3 Yinelenen Verilere ilişkin Örnek Tablo

CRMID	FREAKANS TOPLAM	SON AV TARİH	FREKANS BASINA TUTAR	MUSTER AV TUTAR
8D3CE7FF- 9391-EC11-	1	23	37063	37063
8D3CE7FF- 9391-EC11-	1	23	37063	37063
8D3CE7FF- 9391-EC11-	1	23	37063	37063
5254D03D- 5643-EB11-	106	2	894	94854
5254D03D- 5643-EB11-	106	2	894	94854
5254D03D- 5643-EB11-	106	2	894	94854
5254D03D- 5643-EB11-	106	2	894	94854
5254D03D- 5643-EB11-	106	2	894	94854
5254D03D- 5643-EB11-	106	2	894	94854

4.3 En İyi K Değerinin Belirlenmesi:

K-Means kümeleme tekniğinin aşamaları şu şekilde oluşmaktadır. Öncelikle küme sayısını belirten bir K değeri belirtilmektedir. K sayısı kadar da rastgele küme merkezleri atanmaktadır. Geriye kalan her noktanın bu rastgele belirlenen merkezlere olan uzaklığı hesaplanmaktadır. Her nokta minimum uzaklığa sahip olduğu küme merkezine atanmaktadır. Bu aşamadan sonra küme merkezleri tekrar belirlenerek

uzaklıklar tekrardan hesaplanır ve sonuç aynı kalana kadar bu adımlar tekrar etmektedir. Bu aşamalarda ise K sayısını belirlemek bir problem oluşturmaktadır.

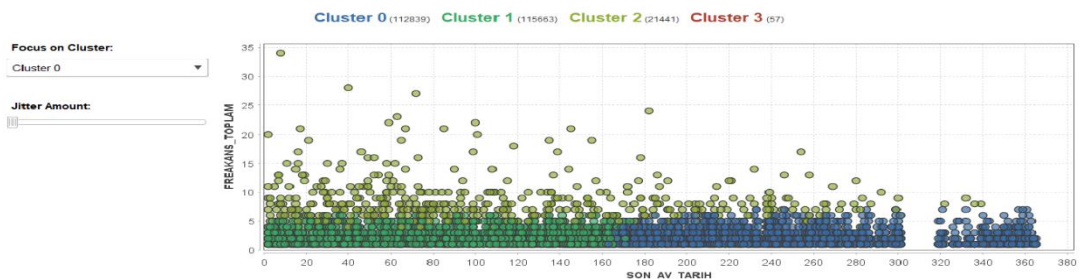
K probleminin belirlenmesine ilişkin literatürde yöntemler bulunmaktadır. Bilgi ölçütü, Elbow Metodu, Başparmak Kuralı Silüet Katsayısı bu yöntemlere örnektir. (Kodinariya ve Makwana, 2013:91, akt: Kılıç vd., 2020) Bu uygulamada ise veri madenciliği araçlarından Rapid Miner kullanılarak X-Means aracından faydalanılmıştır. Rapid Miner aracı X-Means aracında belirtilen aralıklarda sırasıyla her K sayısı için analiz yaparak her analize ilişkin silüet katsayısını hesaplamaktadır. Silüet Katsayısı -1 ile 1 arasında değerlerden oluşmaktadır. (Güven, 2019)

Silüet katsayısı için öncelikle bir veri noktası ile ait olduğu kümedeki diğer veri noktaları arasındaki uzaklığının ortalaması alınmaktadır. İkinci olarak o veriye ait en yakın kümedeki tüm verilere olan uzaklıklarının ortalaması alınmaktadır. Verinin kendi içerisindeki diğer verilere olan ortalama uzaklığını ve en yakın küme içerisindeki verilere olan ortalama uzaklığını bularak çözüm sağlanmaktadır. (Aydın ve Başkır, 2013)

$$sil(x_i) = \frac{b(x_i) - a(x_i)}{\max(a(x_i), b(x_i))} \quad (5)$$

Rapid Miner 'da X-Means aracı belirtilen aralıkta her K değeri için silüet katsayısı hesaplamaktadır. Bu etken hem K değerini hem de K değerine ilişkin oluşturulan kümelerin validasyonunu doğrulamasını sağlamaktadır. Ortaya çıkan sonuca ilişkin kümeleme modelinin de doğrulanması sağlanmaktadır.

x-Means - Scatter Plot

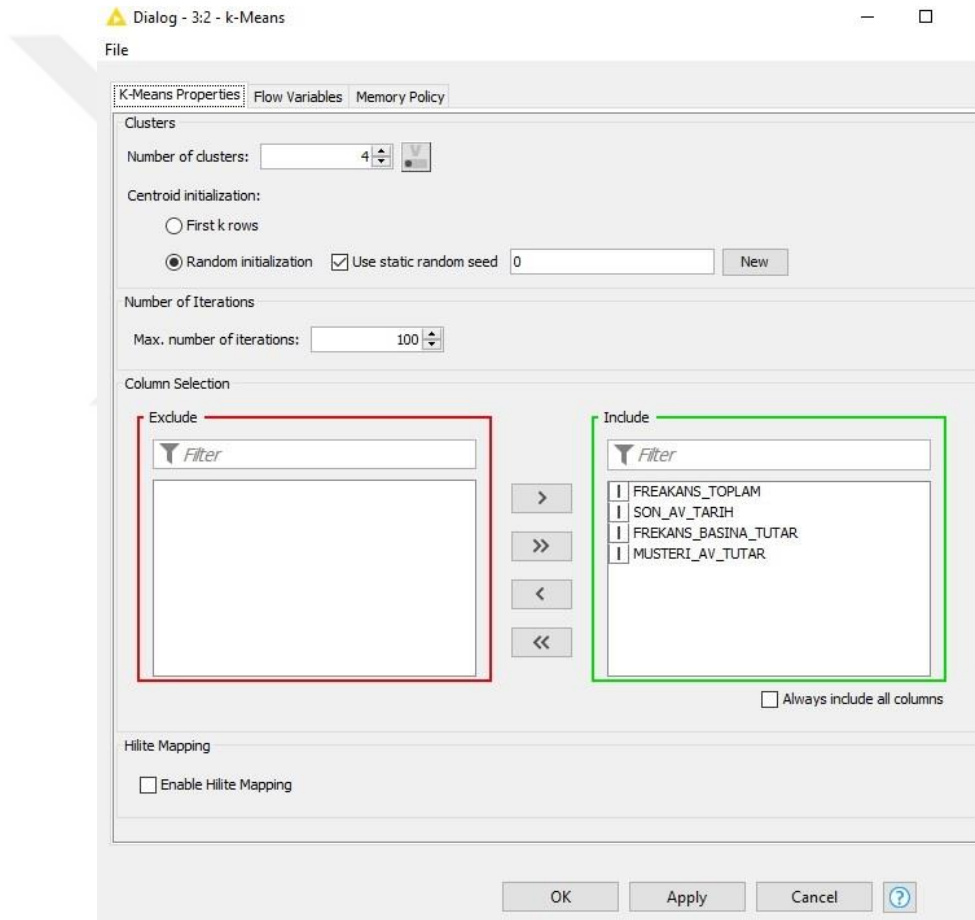


Şekil 4.1 Rapid Miner Çıktısı

Rapid Miner aracında X-Means çalıştırılarak K değeri için 2-12 aralığında test edilmesi istenmiştir. Rapid Miner verinin orta büyüklükte olması nedeniyle uzun bir süre sonucunda sonuç çıktısı vermiştir. Çalışma sonucunda çıktı olarak en iyi K değeri 4 olarak gösterilmektedir.

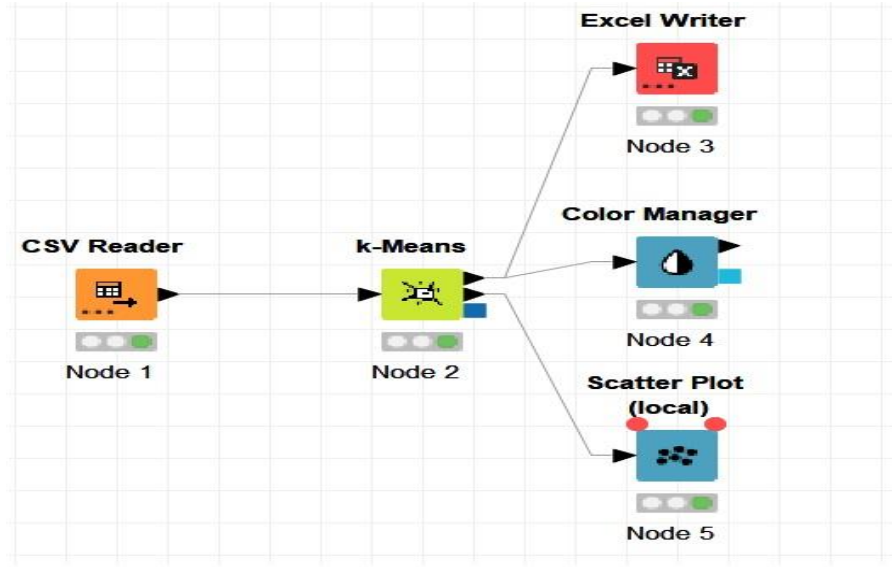
4.4 Kümeleme Analizi:

K-Means aşamalarından olan K sayısının belirlenmesi aşamasında K değerinin 4 bulunmasından sonra KNİME aracı kullanılarak K-Means yöntemi ile kümeleme gerçekleştirilmiştir. Mevcut 1027039 veri K değeri 4 belirlenerek merkezleri rastgele atanması istenmiş ve iterasyon sayısı 100 olarak belirlenmiştir.



Şekil 4.2 Knime Girdi Görüntüsü

Knime'da problemin tasarımı birbirine bağlanan nodelar sayesinde oluşturulmaktadır. Veri CSV Reader ile okunmaktadır. K-Means node ile istenen değerler girilerek kümeleme oluşturulmuştur. Çıktı değerlerinin incelenebilmesi için Color Manager, Scatter Pilot (local) ve Excel Writer kullanılmıştır.



Şekil 4.3 Knime Çalışma Tasarımı Görünütüsü

Kümeleme analizinden sonra bir çıktı elde edilmiştir. İterasyon sayısı 150 yapılarak aynı girdi tekrar edilmiştir. Çıktıya ilişkin sonuçlar karşılaştırıldığında sonuçların değişmediği gözlemlenmiş ve iterasyon sayısı için 100 tekrarın yeterli bir sayı olduğu doğrulanmıştır. Çalışma sonucunda 4 küme elde edilmiştir.

Cluster 0 4899 veriden oluşan bir kümedir. Yüksek alışveriş tutarına sahip müşteriler bu kümede yer almaktadır. Aynı zamanda yüksek alışveriş sayısına ve tek seferde yüksek harcamaya sahip müşterilerin olduğu kümelemedir.

Tablo 4.4 Cluster 0 Verilerinden Örnek Kısım

CRMID	FREAKANS TOPLAM	SON AV TARİH	FREKANS BASINA TUTAR	MUSTERI AV TUTAR
57187952- 2A2C-EC11-	12	39	351	4215
6D903C8D- EA66-EA11-	14	16	298	4175
8424EA0E- 2A8B-E811-	22	46	178	3917

CRMID	FREKANS TOPLAM	SON AV TARİH	FREKANS BASINA TUTAR	MUSTERİ AV TUTAR
077B550F- 1A52-E811-	38	49	172	6558
61A0C9E5- 8EC1-E611-	12	71	341	4103
54E24AEB- B51C-EA11-	16	5	381	6102
AA321525- 64DB-E711-	16	78	254	4068
D9E40D4F- 74DE-E711-	12	22	330	3971
15AE99C4- EDAE-E911-	15	116	260	3907
4F69C2C7- B50E-EA11-	21	37	277	5832
AFEB6CA8- 64FB-E711-	28	94	162	4536
18662AAD- DA0F-E811-	16	16	305	4892
0F291012- F801-E811-	12	162	347	4166
DA9BBD6D- 6E8F-E811-	20	2	260	5217
AD20341E- 062A-E811-	16	65	247	3961

Cluster 1 66268 veriden oluşmaktadır. Cluster 0 değerlerinden daha düşük frekans ve alışveriş tutarlarına sahip olan kümedir. Toplam müşteri sayısının yaklaşık %6,5 gibi bir değerini oluşturmaktadır. Küme içerisinde Cluster 0 müşterilerinden daha yüksek alışveriş tutarına sahip müşterilerde bulunmaktadır.

Tablo 4.5 Cluster 1 Verilerinden Örnek Kısım

CRMID	FREKANS TOPLAM	SON AV TARİH	FREKANS BASINA TUTAR	MUSTERİ AV TUTAR
6454E949- D550-EA11-	1	308	1798	1798
30F75471- E784-EC11-	2	33	925	1850
3536184C- B5C2-E611-	9	86	183	1653
B8D4F892- 6442-E911-	12	36	144	1736
CA897FE8- DDA3-E911-	13	117	126	1643
9610288A- C556-EC11-	2	94	1007	2014
1D12D502- 3D20-EC11-	3	114	712	2138
B5AFF05C- 048C-E811-	6	8	264	1585
804013D1- 453D-E711-	9	292	234	2107
99E41DA0- DCA9-E811-	11	119	205	2264

CRMID	FREAKANS TOPLAM	SON AV TARİH	FREKANS BASINA TUTAR	MUSTERİ AV TUTAR
874FDEA0- EC53-EB11-	8	81	256	2054
76AE4969- 946D-E911-	8	15	197	1576
ECFE1D91- 8554-E911-	7	56	219	1539

Cluster 2 291098 veriden oluşan kümedir. Cluster 3 veri sayısından sonra en çok müşteri sayısına sahip kümedir. İlk 2 kümeye göre düşük frekans ve alışveriş tutarına sahip olsa da müşteri sayısı nedeniyle önemli yer kaplamaktadır.

Tablo 4.6 Cluster 2 Verilerinden Örnek Kısım

CRMID	FREAKANS TOPLAM	SON AV TARİH	FREKANS BASINA TUTAR	MUSTERİ AV TUTAR
A44EB802- 62BB-E811-	3	196	208	624
24D78658- F55C-EC11-	2	89	480	960
7AFA49A1- 45D4-EB11-	4	179	160	641
7792FC32- 8F70-EC11-	1	65	641	641
5D01A967- 75A8-E911-	3	283	208	625

CRMID	FREKANS TOPLAM	SON AV TARİH	FREKANS BASINA TUTAR	MUSTERİ AV TUTAR
9CB0E3A2- C119-EA11-	2	27	321	642
89299FDF- FA9A-E711-	4	71	192	770
432722AC- ABD6-E711-	4	245	208	834
D459DCB9- 3135-EA11-	2	169	305	610
021DA733- C3D5-EB11-	4	75	192	771
15472DDE- 72BA-E711-	4	143	184	739
8036D36F- F4D6-E911-	5	326	161	805

Cluster 3 Verinin büyük kısmını 664774 tanesini oluşturmaktadır. Frekans değeri ve müşteri alışveriş değerleri oldukça düşüktür. Diğer kümelere göre alt değerler almaktadır. Ancak veri sayısının yani müşteri sayısının büyük kısmını kaplayan kümedir.

Tablo 4.7 Cluster 3 Verilerinden Örnek Kısım

CRMID	FREAKANS TOPLAM	SON AV TARİH	FREKANS BASINA TUTAR	MUSTERİ AV TUTAR
899EBD73- F0F6-E811- 80DD-	1	190	3	3
BFC331D2- 0572-E711- 80D2-	1	108	3	3
BD05CDAA- 13EB-EB11- 810C-	1	235	3	3
F06BEDF7- 34D0-E711- 80D5-	1	201	4	4
B6609877- B906-E911- 80DF-	1	48	6	6
2E20C550- 0D97-EC11- 811B-	1	16	20	20
EB2CBAE5- 768D-EB11- 8101-	2	353	38	76
66F935D0- 8B5D-EC11- 810C-	1	89	40	40
FE7EED2B- 8CDB-E811- 80DD-	1	326	50	50
F2053787- 3D6B-EC11- 810F-	1	70	54	54
E9EB6477- 0EAB-E911- 80E7-	2	285	56	112
A15A382C- 5E3A-E811- 80D7-	1	291	60	60

CRMID	FREAKANS TOPLAM	SON AV TARİH	FREKANS BASINA TUTAR	MUSTERİ AV TUTAR
B4FC9297- DBC6-E711- 80D5-	1	206	60	60
7FAC6F65- D1C2-E611- 80CB-	1	173	60	60

Her kümeye ilişkin kriterlerin ortalamaları ve toplam değerleri bulunmuş ve bir tablo haline getirilmiştir. Tablodaki veriler müşterilerin segmente edilip bir değerlendirme yapılabilmesi için çıkarım sağlayan önemli bir tablodur.

Tablo 4.8 Küme Kriterlerine Ait Ortalama ve Toplamlar

	KÜME SAYISI	FREAKANS TOPLAM ORT	SON AV TARİH ORT	FREKANS BASINA TUTAR ORT	MUSTERİ AV TUTAR ORT
0	C 4899	15,71484	86,166565	629,45234	5632,1343
1	C 66268	6,920656	112,2656	462,72625	2137,1584
2	C 291098	3,324368	143,60653	391,10696	903,82214
3	C 664774	1,712493	173,48715	182,64666	276,70513
	1027039				

	KÜME SAYISI	FREKANS TOPLAM TOP	SON AV TARİH TOP	FREKANS BASINA TUTAR TOP	MUSTERİ AV TUTAR TOP
0	C 4899	76987	422130	3083687	27591826
1	C 66268	458618	7439617	30663943	141625214
2	C 291098	967717	41803573	113850455	263100818
3	C 664774	1138421	11532974 5	121418749	183946378
	1027039	2641743	16499506 5	269016834	616264236

Kümeleme sonuçlarına ilişkin çıktı sağlayabilme konusundaki diğer önemli bilgide kümelere ait verilerdeki kriterlerin üst ve alt değerleridir. Bu tablo hangi verilerin hangi diğer kümelere daha yakın olduğunun görülebilmesi ve kriter değeri baz alınarak müşterilerin değerlendirilmesi açısından önemli olmaktadır.

Tablo 4.9 Kümelerin Kriterlere Ait Alt ve Üst Değerleri

	FREKANS TOPLAM	SON AV TARİH	FREKANS BASINA TUTAR	MUSTERİ AV TUTAR
C 0 Üst Değer	316	365	37063	94967
C 0 Alt Değer	1	2	56	3752
C 1 Üst Değer	60	365	3726	3898
C 1 Alt Değer	1	2	40	1459
C 2 Üst Değer	48	365	1462	1539
C 2 Alt Değer	1	2	22	510
C 3 Üst Değer	28	365	521	672
C 3 Alt Değer	1	2	0	0

4.5 Çıktıların Değerlendirilmesi

Kümeleme analizi sonrasında çıktılar değerlendirilerek bir sonuç elde edilmiştir. Kriterlere ilişkin son alışveriş tarihi için minimum değerler fayda sağlarken diğer kriterler için maksimum değerler daha faydalı olarak kabul edilmektedir.

Cluster 0 incelendiğinde en yüksek alışveriş ortalamasına sahip olan küme olarak görülmektedir. Cluster 0 da frekans toplam ortalamasına bakıldığında 15,71 olduğu görülmektedir. Bir yıldaki gün sayısına bakılarak oranlandığında müşteriler ortalama olarak 23 günde bir alışveriş yapmakta ve her alışveriş sıklığında ise yüksek miktarlarda harcama yapmaktadır. Son alışveriş tarihi de diğer kümelere göre daha düşük bir ortalamayla değer kazanmaktadır. Cluster 0 için üst değerler diğer kümelere göre oldukça yüksektir. Hatta Cluster 0 da bulunan müşteri alışveriş tutarındaki alt değer ona en yakın kümedeki üst değere neredeyse çok yakındır. Bu açıdan değerlendirildiğinde Cluster 0 yüksek segment olarak kabul edilebilmektedir.

Cluster 1 incelenirse Cluster 0 küme verilerinin çalışmaya dahil edilmediği varsayıldığında Cluster 1 için diğer kümelemelere kıyasla Cluster 0 için söylenen her durumu söylemek mümkündür. Cluster 1 toplam verinin yani müşteri sayısının yaklaşık %6,5'ini oluşturmaktadır. Ancak diğer 2 kümenin müşteri alışveriş toplam değerine göre bu açıdan değerlendirildiğinde çok iyi bir değere sahiptir. Cluster 2 de toplam müşteri alışveriş değeri 291098 müşteri ile 263100818 olurken Cluster 3 de toplam müşteri alışveriş değeri 664774 müşteri ile 183946378 olmaktadır. Cluster 1 de bulunan 66298 müşterinin ise toplam harcaması 141625214 olmaktadır. Cluster 0'ı en yüksek segment kabul ettiğimizde Cluster 1 içinde orta segment dememiz mümkündür.

Cluster 2 için geriye kalan Cluster 3 kümesi ile kıyaslandığında Cluster 3'ün yarısından daha az müşteriye sahip olmasına rağmen 4 kriter de ortalama ve beklenen değerler gözetildiğinde daha iyi durumda olduğu görülmektedir. Özellikle müşteri alışveriş tutarı kriterine bakıldığında bu durum daha belirgin olmaktadır. Tüm değerlendirmeler açısından Cluster 2 için alt segment dememiz mümkündür.

Cluster 3 segment altı kalmıştır ancak bunu söylemek doğru olmayabilir. Çünkü Cluster 3 toplam veri yani müşteri sayısının yaklaşık %65'ini oluşturmaktadır. Kriterler açısından diğer kümelerden bağımsız incelendiğinde göz ardı edilemeyecek kriter değerlerine sahip olmaktadır. Toplam değerleri açısından bakıldığında da bütün içerisinde önemli bir pay edinmektedir. Üst ve alt değerler açısından bakıldığında birçok müşterinin de Cluster 2 'deki bazı müşteri değerlerden daha yüksek değerlerde müşterileri olduğu da görülebilmektedir. Bu açılardan değerlendirildiğinde kriterlere ilişkin değerler göz önüne alınarak çoğu müşteri segment içine alınabilmektedir. Diğer bir durum ise müşterilerin yaklaşık %65'ini kaplayan Cluster 3'ün sadece kendisini bütün bir veri kabul edilerek Cluster 3 için anlatılan süreçlerle aynı bir alt çalışma yapılması daha faydalı olabileceğini de söylemek mümkün olabilir.

Tablo 4.10 Cluster Segmentasyon Değerlendirmesi

Cluster 0	Üst
Cluster 1	Orta
Cluster 2	Alt
Cluster 3	Segment altı ve diğer

Ancak buraya kadar yapılan değerlendirmeler yeterli olmamaktadır. Örneğin Cluster 0 da müşteri alışveriş tutarının alt değerinden daha yüksek değerler Cluster 1 de bulunabilmektedir. Bu durumun diğer kriter için de geçerli olduğu görülebilmektedir. Ayrıca her kümenin birbiriyle kıyaslanması durumunda da bu görülebilmektedir. Bu süreçte aşamalı olarak ilerleyerek incelemek mümkündür. Başlangıçta Cluster 0'a ait olan alt değerleri inceleyelim. Frekans toplamı ve son alışveriş tarihi Cluster 0 da olduğu gibi tüm kümelerde aynı değere sahiptir. Cluster 0 da frekans başına alt tutar 56 müşteri alışveriş tutarı ise 3752'dir. Cluster 1 de bu değerleri aşan 601 veri bulunmaktadır. Cluster 0'a ait 4899 müşteri bulunmaktaydı. Cluster 0 da belirtilen alt değerleri geçen 601 veri oran anlamında az sayılamayacaktır. Bu açıdan 56 frekans başına tutarı ve müşteri alışveriş tutarı olan 3752'i aşan verileri de üst segment kabul edebiliriz. İkinci aşamada Cluster 1' e ait alt değerler frekans başına tutar için 40 müşteri alışveriş tutarı için 1459 olarak görülmektedir. Cluster 1 de ise 66268 veri bulunmaktaydı. Cluster 2 de belirtilen alt sınırları aşan yaklaşık Cluster 1'in 6938 müşterisi bulunmaktadır. Yine aynı açıdan frekans başına 40 ve müşteri alışveriş tutarı 1459'u aşan müşterileri bir segment yukarıda da değerlendirebilmekteyiz. Son aşamaya gelindiğinde ise Cluster 2'deki alt değerler frekans başına tutar için 22 müşteri alışveriş tutarı için 510 olmaktadır. Cluster 3 de bu değerleri aşan 55566 müşteri bulunmaktadır. Özellikle segment altı kabul etmek istemediğimiz bu grupta 55566 veriyi bir üst segment üzerinden değerlendirmek iyi bir sonuç vermektedir. Müşterileri segmente anlamında kriter değerine ilişkin değerlendirmemiz sonucunda kabul edilen alt sınırlar tablo şeklinde gösterilmektedir.

Tablo 4.11 Kriter Değerlerine Göre Segment Oluşturma

	FREKANS BASINA TUTAR	MUSTERI AV TUTAR
Üst Segment	56	3752
Orta Segment	40	1459
Alt Segment	22	510

4.6 Kümeler İçin Yaklaşımların Belirlenmesi:

Cluster 0 yılda ortalama 15 defa alışveriş yapmaktadır. Bu ortalama bakıldığında Cluster 0 'a ait müşteriler 23 günde bir alışveriş yapmaktadır. Her alışveriş de ortalama 630 TL harcamaktadır. Müşterilere ait toplam alışveriş tutarı ortalaması da 5632 TL olmaktadır. Yüksek ve sık harcama yapan Cluster 0 müşteri kümesi için 23 günden daha kısa sürelerde bilgilendirme ve kampanya mesajları sunulmalıdır. Cluster 1 ise yaklaşık 53 günde bir alışveriş yapmaktadır. Bir üst kümeye ait olan 23 günde bir bilgilendirme ve kampanya mesajları Cluster 1 'deki müşterilere sunulmalıdır. Devamında Cluster 2'ye bakıldığında 109 günde bir alışveriş yapmaktadır. Bir üst segmenti olan Cluster 1 müşterilerinde ise bu süre 53 günde bir idi. Cluster 2 için 53 günde bir bilgilendirme ve kampanya mesajları sunulmalıdır. Cluster 3 için ise 213 yılda 213 günde bir ortalamasına sahiptir. Yine aynı şekilde bir üst segmentine ait olan 109 gün de bir bilgilendirme ve kampanya mesajları sunulmalıdır. Bu oranlarda müşterilere yaklaşım sağlamak gerçekçi olmakta ve geri dönüş anlamında pozitif etki sağlayacaktır. Bu süreçlerden alınacak geri dönüşlerde tüm kümelerdeki frekans toplam ve son alışveriş tarihi kriterlerini iyileştirecektir.

Kümeler belirli kriterler üzerinden oluşturulmuştur. Çıkan değerler üzerinden ise kümeler yüksekten segment dışına doğru sıralanmıştır. İnceleme sonucunda her segmentin alt değerlerine sahip müşterileri bir alt segmentte değer açısından geçen müşterileri olduğu görülmektedir. Örneğin Cluster 0 ait en alt değer 3752 iken bu değerleri geçen müşteriler Cluster 1 de bulunmaktadır. Cluster 1 3752 TL'yi geçen fiyatlarda alışveriş yaparken bu fiyat bantlarındaki ürünleri indirimli olarak sunmak hem onların alış gücüne hitap ederken hem de indirimli ürünler olması daha da

alabilme isteęi uyandıracaktır. Aynı süreçleri Cluster 2 ve Cluster 3 içinde uygulamak aynı oranda müşterilerde alabilme isteęi uyandıracaktır. Bu süreçte Müşteri alışveriş toplam ve en son alışveriş tarihi açısından pozitif etki oluşturacaktır. Cluster 0 için ise ortalama müşteri toplam alışveriş olan 5632 TL civarındaki ürünler için indirim fırsatı sunmak yeterli olacaktır. Özetle 4 segment elde edilmiştir. Her segmentin bulunduğu segmente ait kriter değerlerini yani durumlarını koruması istenmektedir. Farklı bir kavramla müşterilerini kaybetmemek devamlılıklarını sağlamak istenmektedir. Bu stratejilerin bu amaçlar doğrultusunda fayda sağlaması beklenmektedir. Diğer mevcut amaç ise her segmente ait müşterilerin kendisini bir üst segment kriter değerlerine ulaşabilmesi sağlamaktır. Bu bağlamda değer bazlı değerlendirme sonuçlarında her segmentin üst segmentindeki alt değerlerini geçen değerleri bulunduğunu gözlemlemiştik. Bu aralıklarda müşterilere indirimli ürünler sunmak hem alıcıların alış gücüne hitap ederken hem de alabilme isteęini uyandırmasını ifade etmektedir. Bu süreçler müşterilerin daha üst segmentlere çıkabilmesini sağlayabilmede gerçekçi bir yaklaşım olacaktır.

SONUÇ

Bilgi sonu olmayan ve sürekli olarak artan doğaya ve insana değer katan bir durumdur. Bilgi geçmiş zamanlara bakıldığında sadece ulaşılması gereken bir değerdi. Ancak günümüzde her ne kadar bilgiye ulaşmak kolay hale gelse de günümüzde bilgi kirliliği gibi bir gerçek bulunmaktadır. Yanlış bilgi yani dezenformasyondan korunmak konuya ilişkin gerçek bilgilere ulaşmak gereklidir. Doğru bir bilgi için iyi bir araştırma, geniş bir kaynak yelpazesi ve düşüncesele anlamda süzebilme kabiliyeti gereklidir. Ancak artık bu süreçlerde bilgiden kazanım sağlamak için yeterli olmamaktadır. Mevcut bilgiden bir anlam ya da ifade ettiği durumdan bir çıkarım sağlamak gereklidir. Başlangıçta bu süreç kolay gibi görülse de günümüzde geleneksel yöntemler ile oldukça zor hatta imkânsız hale gelmiştir. Belirtildiği gibi bilgi sürekli olarak artmaktadır. Bilgi birikim artışı sürekli olarak kendini güncellerken bilgiyi işleyebilme gücünde zamanla zorlaşmaktadır. Son 50 yıl içerisindeki bilgi birikim hızının çok hızlı bir şekilde artması özellikle neden olan durumlardandır. Ancak bu bilgi birikim hızına karşılık teknolojik olanak ve bilimsel metodolojilerin geliştirilmesi istenen durumlara ilişkin olumlu olanak sağlamaktadır. Bu bağlamda 100'e yakın kaynak incelenmiş bir literatür çalışması gerçekleştirilmiştir. Edinilen bilgiler ışığında 1.000.000'u aşkın müşterisi olan bir perakende sektöründeki firmanın müşterilerine ait bilgi ve davranışlarından elde etmek istediği bilgi çıkarımı üzerinde çalışılmıştır. Firma müşterilerine ilişkin bir değer tanımı yapmak istemektedir. Bütün müşterilerin aynı değerde tutularak segmente edilmesi doğru olmazken yüksek sayıdaki müşterilerine tek tek bir değer oluşturmak da neredeyse imkânsız bir hal almaktadır. Elimizdeki müşteriler alışveriş davranışlarına göre tanımlanmış ve K-Means yöntemi ile kümelenecek bir sonuç elde edilmiştir. Süreçler öncesinde yinelenen artık veriler silinmiştir. K sayısının belirlenmesi ve modelin doğrulanmasında Rapid Miner aracı ve X-Means yönteminden faydalanılmıştır. Sonuçta K sayısı 4 bulunmuş ve müşteriler 4 grupta segmente edilmiştir. Uygulama kısmının sonuç bölümünde belirtilen değerlere ulaşılmıştır. Perakende sektöründeki işletme 1.000.000'u aşkın müşteri için ayrı ayrı değer ve stratejik tavır belirlemek yerine benzer müşteri davranış ve alışkanlıklarına sahip kişilerin bir arada bulunduğu 4 kümeye değer ve stratejik tavır sergilemesi hem daha doğru ve gerçekçi hem de daha kolay bir süreç olmaktadır. Bu durum göz ardı edilemez bir fayda sağlamaktadır. Ayrıca yapılan çalışmalar

sonucunda bireysel olarak kriterlere ilişkin hangi değerlere sahip müşterilerin hangi segmentte değerlendirilmesinin de gerektiği söylenebilmektedir. Yapılan çalışmaların süre maliyet ilişkisi açısından kazanç sağladığını söylemek mümkündür. Diğer perakende sektöründe yer alan ve müşteri ilişkileri yönetimi bulunan her firma için bu çalışmaların fayda sağlayabileceğini söylemek mümkün olmaktadır. Müşteriler segmente edilerek edilen segmentler sonucunda bir değer tanımlaması yapılabilir ve stratejik tavırlar oluşturulabilir. Bu durumlarda yapılan eylemlerin daha yüksek oranda geri dönüş alınmasını sağlarken süre maliyet açısından da verim alınmasını sağlamaktadır. Ancak yapılan her çalışmada harcanan maliyetin sağlanan faydayı geçmemesi gereklidir. Az sayıda müşteriye sahip firmaların bu çalışmayı yapması sağlanan fayda konusunda harcanan maliyetin altında kalabilmektedir. Ayrıca mevcut veri ve müşteri alışkanlıkları neredeyse her saniye kendini güncelleyebilmektedir. Haliyle yapılan çalışmaların da tekrar edilmesi gerekli olmaktadır. Bu açılarından bakıldığında sürekli olarak çalışmanın tekrar etmesi daha maliyetli ve dezavantajlı durumlardandır. İlgili yöntemle hem literatüre bakıldığında hem de kendi yaptığımız çalışmada ise K sayısının belirlenmesi ve kümeler arasında kıyaslamaların yapılması oldukça zor olmaktadır. K sayısının belirlenebilmesi ile çıkan kümelerin birbiri ile kıyaslamalarının yapılarak değerlendirilmesi kümeleme çalışmasından neredeyse daha uzun süren bir süreç olabilmektedir. Bu açıdan değerlendirildiğinde çalışmanın zorlu yönleri bu şekilde ifade edilebilmektedir. Çalışma kendi üzerinde alt çalışmalar gereksinimi de oluşturmaktadır. Müşteriler değer anlamında segmente edildikten sonra ürün tercih yelpazeleri üzerinden bir çalışma daha yapılması stratejik yaklaşımların daha geniş bir pencereden görülebilmesini sağlayacaktır. İlerleyen süreçlerde ise bilgi birikimine ilişkin miktarın artacağını ve bu bilgiden sağlanmak istenen çıkarımında önem bakımından artacağını söylemek mümkündür. Literatürde bu süreçlerle ilgili daha iyi kazanım sağlayan yöntem ve araçların bulunması beklenmektedir.

KAYNAKÇA

İnternet Yayınları

- Abdullah, A. L. A. N., & KARABATAK, M. (2020). Veri Seti-Sınıflandırma İlişkisinde Performansa Etki Eden Faktörlerin Değerlendirilmesi. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 32(2), 531-540.
- Akgöbek, Ö., & Çakır, F. (2009). Veri madenciliğinde bir uzman sistem tasarımı. *Akademik Bilişim*, 9, 801-806.
- AKGÜN, K., & ÖZEK, M. B. (2020). Eğitsel veri madenciliği yöntemi ile ilgili yapılmış çalışmaların incelenmesi: İçerik analizi. *Uluslararası Eğitim Bilim ve Teknoloji Dergisi*, 6(3), 197-213.
- AKPINAR, Ö. (2018). SİGORTA SEKTÖRÜNDE VERİ MADENCİLİĞİ VE KULLANIM ALANLARI. *Dumlupınar University Journal of Social Science/Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, (57).
- ALAN, M. A., & YEŞİLYURT, C. Birliktelik Kuralları Madenciliği İle Yatan Hasta Profiline Çıkarılması. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 23(4), 1917-1926.
- Alan, M. A. (2012). Veri madenciliği ve lisansüstü öğrenci verileri üzerine bir uygulama. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, (33).
- Aslanyürek, M., & Mesut, A. (2021). Kümeleme Performansını Ölçmek için Yeni Bir Yöntem ve Metin Kümeleme için Değerlendirmesi. *Avrupa Bilim ve Teknoloji Dergisi*, (27), 53-65.
- Aydın, D., & BAŞKIR, M. B. (2013). Bankaların 2012 yılı sermaye yeterlilik rasyolarına göre kümeleme analizi ve çok boyutlu ölçekleme sonucu sınıflandırılma yapıları. *Bankacılık ve Sigortacılık Araştırmaları Dergisi*, 1(5), 29-47.

- AYIK, Y. Z., Özdemir, A., & Yavuz, U. (2007). LİSE TÜRÜ VE LİSE MEZUNİYET BAŞARISININ, KAZANILAN FAKÜLTE İLE İLİŞKİSİNİN VERİ MADENCİLİĞİ TEKNİĞİ İLE ANALİZİ. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 10(2), 441-454.
- Aytaç, M. B., & Bilge, H. Ş. (2013). TELE PAZARLAMA VERİLERİNİN BİRLİKTELİK KURALLARIYLA VE CRISP-DM YÖNTEMİYLE ANALİZ EDİLMESİ. *Aksaray Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 5(2), 25-40.
- Aytuğ, O. N. A. N. (2015). Şirket iflaslarının tahminlenmesinde karar ağacı algoritmalarının karşılaştırmalı başarımların analizi. *Bilişim Teknolojileri Dergisi*, 8(1).
- BARDİ, Ş., & CAN, A. V. (2021). Diskriminant analizi ve C5. 0 algoritması ile finansal başarısızlığın tahmini: BİST Kobi Sanayi Endeksi'ndeki işletmeler örneği. *Ömer Halisdemir Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 14(3), 1071-1090.
- BAŞKOL, M. (2020). RFM VE UYUM ANALİZİ KULLANILARAK MÜŞTERİ SEGMENTASYONUNUN BELİRLENMESİ. *Business & Management Studies: An International Journal*, 8(4), 902-928.
- Baykal, A. (2006). Veri madenciliği uygulama alanları. *Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi*, (7), 95-107.
- Baykasoğlu, A. (2005). Veri madenciliği ve çimento sektöründe bir uygulama. *Akademik Bilişim*, 5.
- BİLGİÇ, E., & ÇAKIR, Ö. (2019). SOSYOEKONOMİK YAKLAŞIMLA ZİNCİR PERAKENDE MAĞAZALARININ SEGMENTASYONU. *Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 41(2), 338-363.
- Büyüköztürk, Ş. (2002). Faktör analizi: Temel kavramlar ve ölçek geliştirmede kullanımı. *Kuram ve uygulamada eğitim yönetimi*, 32(32), 470-483.

- Cengiz, D., & Öztürk, F. (2012). Türkiye’de illerin eğitim düzeylerine göre kümeleme analizi ile incelenmesi. *Trakya Üniversitesi Sosyal Bilimler Dergisi*, 14(1), 69-84.
- Ceylan, Z., Gürsev, S., & Bulkan, S. (2017). İki aşamalı kümeleme analizi ile bireysel emeklilik sektöründe müşteri profilinin değerlendirilmesi. *Bilişim Teknolojileri Dergisi*, 10(4), 475-485.
- COŞKUN, E., GÜNDOĞAN, E., Mehmet, K. A. Y. A., & ALHAJJ, R. Veri madenciliği yöntemleri kullanarak yoğun bakım ünitesindeki hastaların sınıflandırması. *Computer Science*, (Special), 319-328.
- Coşlu, E. (2013). Veri madenciliği. *Akademik bilişim*, 23-25.
- Çalış, A., & Baynal, K. (2016). Kümeleme analizi ile bankacılık sektöründe satış stratejilerinin belirlenmesi, Beykent Üniversitesi. *Beykent Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, 9(1), 13-41.
- ÇALIŞ, A., KAYAPINAR, S., & ÇETİNYOKUŞ, T. (2014). VERİ MADENCİLİĞİNDE KARAR AĞACI ALGORİTMALARI İLE BİLGİSAYAR VE İNTERNET GÜVENLİĞİ ÜZERİNE BİR UYGULAMA. *Endüstri Mühendisliği*, 25(3), 2-19.
- Çamoglu, K., Akbayir, D., Yücalar, F., & Bayrakli, S. (2010). İNSANSIZ HAVA ARAÇLARININ GENETİK ALGORİTMA YÖNTEMİYLE ÇOKLU HEDEFLERE PLANLANMASI. *Journal of Aeronautics & Space Technologies/Havacılık ve Uzay Teknolojileri Dergisi*, 4(3).
- Çelik, E., Deniz, D. A. L., & AYDİN, T. (2021). Duygu Analizi İçin Veri Madenciliği Sınıflandırma Algoritmalarının Karşılaştırılması. *Avrupa Bilim ve Teknoloji Dergisi*, (27), 880-889.
- Çolak, S. (2010). Genetik algoritmalar yardımı ile gezgin satıcı probleminin çözümü üzerine bir uygulama. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 19(3), 423-438.

- Dener, M., Dörterler, M., & Orman, A. (2009). Açık kaynak kodlu veri madenciliği programları: WEKA’da örnek uygulama. *Akademik Bilişim*, 9, 11-13.
- DENİZ, S. S. Kural Tabanlı Sınıflandırma Algoritmalarının Karşılaştırılması. *Veri Bilimi*, 4(3), 72-80.
- Dilki, G., & Başar, Ö. D. (2020). İŞLETMELERİN İFLAS TAHMİNİNDE K-EN YAKIN KOMŞU ALGORİTMASI ÜZERİNDEN UZAKLIK ÖLÇÜTLERİNİN KARŞILAŞTIRILMASI. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 19(38), 224-233.
- Doğan, B., Buldu, A., Demir, Ö., & Bahar, E. R. O. L. (2018). Sigortacılık sektöründe müşteri ilişki yönetimi için kümeleme analizi. *Karaelmas Fen ve Mühendislik Dergisi*, 8(1), 11-18.
- DOĞAN, M., & KEVSER, M. VERİ MADENCİLİĞİ YÖNTEMİNİN FİNANS SEKTÖRÜNE UYGULANMASINA YÖNELİK KAVRAMSAL BGR DEĞERLENDİRME. *Türk Kamu Yönetimi Dergisi*, 2(1), 62-69.
- Doğruel, M., & FIRAT, S. Ü. (2021). Veri Madenciliği Karar Ağaçları Kullanarak Ülkelerin İnovasyon Değerlerinin Tahmini ve Doğrusal Regresyon Modeli ile Karşılaştırmalı Bir Uygulama. *İstanbul Business Research*, 50(2), 465-493.
- Emel, G. G., & Taşkın, Ç. (2002). Genetik algoritmalar ve uygulama alanları. *Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 21(1), 129-152.
- Emre, İ. E., & Erol, Ç. S. (2017). Veri Analizinde İstatistik mi Veri Madenciliği mi?. *Bilişim Teknolojileri Dergisi*, 10(2), 161-167.
- Gamze, Ö. D. E. V., & Peker, S. Müşteri şikâyet yönetiminde firmaların performanslarının değerlendirilmesi: Kümeleme analizi incelemesi. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 13(3), 447-456.
- Giray, S. (2016). İki aşamalı kümeleme analizi ile hükümlü verilerinin incelenmesi. *İstanbul Üniversitesi İktisat Fakültesi Ekonometri ve İstatistik Dergisi*, 25, 1-31.

- Gülsün, B., Tuzkaya, G., & Duman, C. (2009). GENETİK ALGORİTMALAR İLE TESİS YERLEŞİMİ TASARIMI VE BİR UYGULAMA. *Doğuş Üniversitesi Dergisi*, 10(1), 73-87.
- Güner, S., CODAL, K. S., GEÇER, H. S., & COŞKUN, E. (2018). TRAFİK KAZA DESENLERİNİN TANIMLANMASINDA K-MEANS KÜMELEME ALGORİTMASININ KULLANILMASI: SAKARYA İLİ UYGULAMASI. *İşletme Bilimi Dergisi*, 6(3), 89-105.
- GÜRLER, C. R Programlama Dili ile Kümeleme Analizi. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 22(Özel Sayı 2), 341-366.
- GÜRLER, C. (2023). Ülkelerin Turizm Potansiyellerine Göre K-Ortalamalar Yöntemi Kullanılarak Kümelenmesi. *MANAS Sosyal Araştırmalar Dergisi*, 12(1), 353-363.
- Hotaman, D. (2020). ÖĞRENCİ BAŞARISININ DEĞERLENDİRİLMESİNDE EĞİTSEL VERİ MADENCİLİĞİNİN KULLANIMI. *Ulakbilge Sosyal Bilimler Dergisi*, 8(48), 577-587.
- Irmak, S., Köksal, C. D., & Asilkan, Ö. (2012). Hastanelerin gelecekteki hasta yoğunluklarının veri madenciliği yöntemleri ile tahmin edilmesi. *Uluslararası Alanya İşletme Fakültesi Dergisi*, 4(1), 101-114.
- Meltem, I. Ş. I. K., & ÇAMURCU, A. Y. (2007). K-Means, K-Medoids ve Bulanık C-Means Algoritmalarının Uygulamalı Olarak Performanslarının Tespiti. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 6(11), 31-45.
- Kaya, M., & Özel, S. A. (2014). Açık kaynak kodlu veri madenciliği yazılımlarının karşılaştırılması. *Akademik Bilişim*, 1-8.
- KILIÇ, G., Budak, İ., & Organ, A. (2020). K-ortalamalar Kümeleme Analizi İle Belediyelerin Çevre Hizmetlerinin Değerlendirilmesi. *Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 15(1), 209-230.

- KILINÇ, D., BORANDAĞ, E., YÜCALAR, F., TUNALI, V., ŞİMŞEK, M., & ÖZÇİFT, A. (2016). KNN algoritması ve r dili ile metin madenciliği kullanılarak bilimsel makale tasnifi. *Marmara Fen Bilimleri Dergisi*, 28(3), 89-94.
- Koyuncugil, A., & Özgülbaş, N. (2009). Veri madenciliği: Tıp ve sağlık hizmetlerinde kullanımı ve uygulamaları. *Bilişim Teknolojileri Dergisi*, 2(2).
- Köseoğlu, Ö., & Demirci, Y. (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(Kayfor 15 Özel Sayısı), 2223-2239.
- KÜPELİ, T. Ş., & ÜNLÜÖNEN, K. (2021). Veri Madenciliği ve Turizmde Veri Madenciliği Çalışmaları**(Data Mining and Data. *Journal of Tourism and Gastronomy Studies*, 9(1), 275-298.
- Meltem, I. Ş. I. K., & ÇAMURCU, A. Y. (2007). K-Means, K-Medoids ve Bulanık C-Means Algoritmalarının Uygulamalı Olarak Performanslarının Tespiti. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 6(11), 31-45.
- Meltem, I. Ş. I. K., & Çamurcu, A. Y. (2011). K-MEANS VE AŞIRI KÜRESEL C-MEANS ALGORİTMALARI İLE BELGE MADENCİLİĞİ. *Marmara Fen Bilimleri Dergisi*, 22(1), 1-18.
- Atalay, M., & Çelik, E. (2017). Büyük veri analizinde yapay zekâ ve makine öğrenmesi uygulamaları-artificial intelligence and machine learning applications in big data analysis. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 9(22), 155-172.
- NAMLI, E., & Murat, S. (2019). Müşteri odaklı pazarlama stratejileri için veri madenciliği teknikleri kapsamında perakende sektöründe kümeleme analizi uygulaması. *Gümüşhane Üniversitesi Sosyal Bilimler Dergisi*, 10, 317-327.

- Neslihan, K. Ö. S. E., & ERSÖZ, F. Veri Madenciliğinde Karar Ağacı Algoritmaları İle Demir Çelik Endüstrisinde İş Kazaları Üzerine Bir Uygulama. *Avrupa Bilim ve Teknoloji Dergisi*, 397-407.
- Oğuzlar, A. (2003). Veri ön işleme. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, (21).
- Orakcı, M., Cıylan, B., Kök, İ., & Sevri, M. (2019). Suç Analizinde Veri Madenciliği Teknikleri Ve Makine Öğrenmesi Algoritmalarının Kullanılması. *Yayınlanmamış Yüksek Lisans Tezi*.
- Oralhan, B. (2020). VERİ MADENCİLİĞİ YAKLAŞIMI İLE TELEKOMÜNİKASYON SEKTÖRÜNDE ARIZA GİDERME ANALİZİ. *Business & Management Studies: An International Journal*, 8(1), 1026-1043.
- ÖZÇAKIR, F. C., & ÇAMURCU, A. Y. (2007). Birlikte kuralı yöntemi için bir veri madenciliği yazılımı tasarımı ve uygulaması. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 6(12), 21-37.
- Özekes, S. (2003). Veri madenciliği modelleri ve uygulama alanları.
- ÖZMEN, M., Delice, Y., & AYDOĞAN, E. K. (2018). Telekomünikasyon sektöründe PSO ile müşteri bölümlenmesi. *Bilişim Teknolojileri Dergisi*, 11(2), 163-173.
- Özseven, T., & Ersöz, T. (2016). E-TİCARET VERİLERİNİN MÜŞTERİ PROFİLİ AÇISINDAN DEĞERLENDİRİLMESİ. *Uluslararası Yönetim İktisat ve İşletme Dergisi*, 12(28), 85-98.
- SABAH, L., & BAYRAKTAR, H. (2020). Veri Madenciliği Birlikte Kuralları ile Binaların Risk Durumlarının Analizi: Kaynaşlı, Düzce Örneği. *Gazi Mühendislik Bilimleri Dergisi*, 6(1), 70-78.
- Sariman, G. (2011). Veri madenciliğinde kümeleme teknikleri üzerine bir çalışma: k-means ve k-medoids kümeleme algoritmalarının karşılaştırılması. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 15(3), 192-202.

- Savaş, S., Topaloğlu, N., & YILMAZ, M. (2012). Veri madenciliği ve Türkiye’deki uygulama örnekleri. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 11(21), 1-23.
- DENİZ, S. S. VERİ MADENCİLİĞİ ARAÇLARI KULLANILARAK TÜRKİYE’NİN TURİZM GELİRLERİNİN AYLARA GÖRE YAPAY SİNİR AĞLARI İLE TAHMİNLENMESİ. *Yüzüncü Yıl Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 241-255.
- Seyrek, İ. H., & Ata, H. A. (2010). Veri zarflama analizi ve veri madenciliği ile mevduat bankalarında etkinlik ölçümü. *BDDK Bankacılık ve Finansal Piyasalar Dergisi*, 4(2), 67-84.
- SİSTEMLERİNİN, S. D., KAHRAMAN, A. M., & ÖZDAĞLAR, D. (2004). SU DAĞITIM SİSTEMLERİNİN GENETİK ALGORİTMA İLE OPTİMİZASYONU. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, 6(3), 1-18.
- Şahin, Y., & Karagül, K. (2019). Gezgin satıcı probleminin melez akışkan genetik algoritma (MAGA) kullanarak çözümü. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 25(1), 106-114.
- Şengür, D., & TEKİN, A. (2013). Öğrencilerin mezuniyet notlarının veri madenciliği metotları ile tahmini. *Bilişim Teknolojileri Dergisi*, 6(3), 7-16.
- Taşcı, E., & Onan, A. (2016). K-en yakın komşu algoritması parametrelerinin sınıflandırma performansı üzerine etkisinin incelenmesi. *Akademik Bilişim*, 1(1), 4-18.
- Tekerek, A. (2011). Veri madenciliği süreçleri ve açık kaynak kodlu veri madenciliği araçları. *Akademik Bilişim*, 11, 2-4.
- Terzi, S. (2012). Hile ve usulsüzlüklerin tespitinde veri madenciliğinin kullanımı. *Muhasebe ve Finansman Dergisi*, (54), 51-64.

- Tuncer, İ., & Karaboğa, K. (2021). RFM Metriklerini Kullanarak Kümeleme Yöntemi ile Müsteri Bölümlendirme: Perakende Sektöründe Bir Uygulama. *Third Sector Social Economic Review*, 56(1), 411-425.
- Tüzüntürk, S. (2010). Veri madenciliği ve istatistik. *Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 29(1), 65-90.
- UYANIK, F., & KASAPBAŞI, M. C. (2021). Telekomünikasyon sektörü için veri madenciliği ve makine öğrenmesi teknikleri ile ayrılan müşteri analizi. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 9(3), 172-191.
- Üstündağ, M., Engin, A. V. C. I., Gökbulut, M., & Fikret, A. T. A. (2014). Dalgacık paket dönüşümü ve genetik algoritma kullanarak zayıf radar sinyallerinin gürültüden arındırılması. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 29(2).
- Yakut, E., & Elmas, B. (2013). İŞLETMELERİN FİNANSALBAŞARISIZLIĞININ VERİ MADENCİLİĞİ VE DİSKRİMİNANT ANALİZİ MODELLERİ İLE TAHMİN EDİLMESİ. *Afyon Kocatepe Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 15(1), 261-280.
- YAZGAN, T. K. A. G. E., & ŞAHİNLER, S. AŞAMALI KÜMELEME ANALİZİ (HIERARCHICAL CLUSTER ANALYSIS) YÖNTEMLERİNİN KARŞILAŞTIRILMALI OLARAK İNCELENMESİ.
- YILDIRIM, P., & Birant, D. (2018). Bulut bilişimde veri madenciliği tekniklerinin uygulanması: Bir literatür taraması. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 24(2), 336-343.
- Yildirim, P., Uludağ, M., & Görür, A. (2008). Hastane Bilgi Sistemlerinde Veri Madenciliği. *Çanakkale On Sekiz Mart Üniversitesi Akademik Bilişim*.
- Yıldız, M., & Şeker, Ş. E. (2016). Veri Madenciliği Araçları (Data Mining Tools). *YBS Ansiklopedi*, 3(4), 10-19.

YILDIZ, M. Ş., Kekezoğlu, B., & Evren, İ. Ş. E. N. (2019). K-ORTALAMALAR KÜMELEME YÖNTEMİ İLE GÜÇ TRANSFORMATÖRLERİNİN BAKIM STRATEJİLERİNİN BELİRLENMESİ. *Mühendislik Bilimleri ve Tasarım Dergisi*, 7(3), 505-513.

Zontul, M., & YANGIN, A. (2017). Yapay Sinir Ağı Teknikleri Kullanarak Eğitim Yayıncılığı Sektöründe Veri Madenciliği= Data Mining On Education Publishing Sector By Artificial Neural Network Tecniques. *AURUM Journal of Engineering Systems and Architecture*, 1(2), 1-15.

İnternet/Kurum:

Akca, M.F. (2020, Eylül 7). *Karar ağaçları (makine öğrenmesi serisi-3)*. Medium.

<https://medium.com/deep-learning-turkiye/karar-a%C4%9Fa%C3%A7lar%C4%B1-makine-%C3%B6%C4%9Frenmesi-serisi-3-a03f3ff00ba5>

Alaybeg, F. (2019, Haziran 24). *Veri Madenciliği Giriş, Yöntemleri ve Metodolojileri*. Medium.

<https://furkanalaybeg.medium.com/veri-madencili%C4%9Fi-ve-y%C3%B6ntemleri-d0e2fd238e44>

Ay, Ş. (2019, Ekim 28). *K-means algoritması*. Medium.

<https://medium.com/deep-learning-turkiye/k-means-algoritmas%C4%B1-b460620dd02a>

Bektaş, M. (2020, Mart 17). *Kümeleme Algoritmaları ile Müşteri Segmentasyonu*. Medium.

<https://medium.com/@mbektas/k%C3%BCmeleme-algoritmalar%C4%B1-ile-m%C3%BC%C5%9Fteri-segmentasyonu-498d63c56204>

Bilgi Uzmanı. tarih yok. *Veri Madenciliği Tarihçesi*. Bilgi Bilimi.

<https://bilgibilimi.net/veri-madenciligi-tarihcesi/>

Cerebro. (2018, Mart 28). *Popüler veri madenciliği araçları*. Medium.

<https://medium.com/kodcular/pop%C3%BCler-veri-madencili%C4%9Fi-ara%C3%A7lar%C4%B1-d8f1b85b1f20>

Dayanaklı, A.S. (2021, Şubat 11). *Aykırı değer (outlier) analizi nedir? Uç değerler nasıl tespit edilir*. Ravenfo.

<https://ravenfo.com/2021/02/11/aykiri-deger-analizi/>

Erdem, K. (2021, Mart 20). *Gözetimsiz Öğrenme ve Kümeleme Analizi (Unsupervised Learning and Cluster Analysis)*. Medium.

<https://medium.com/machine-learning-t%C3%BCrkiye/g%C3%B6zetimsiz-%C3%B6%C4%9Frenme-ve-k%C3%BCmeleme-analizi-unsupervised-learning-and-cluster-analysis-bcc969a69cd2#:~:text=K%C3%BCmeleme%20Y%C3%B6ntemleri%20Nele,dir%3F,k%C3%BCmeleme%20ve%20hiyerar%C5%9Fik%20olmayan%20k%C3%BCmeledir.&text=Temel%20amac%C4%B1m%C4%B1z%20g%C3%B6zlemleri%20birbirine%20olan%20benzerliklerine%20g%C3%B6re%20alt%20k%C3%BCmelere%20ay%C4%B1rmakt%C4%B1r.>

Euromsg express. (2022, Mayıs 11). *Müşteri veri segmentasyonu nedir?*. Euromsg Express.

<https://www.euromsgexpress.com/musteri-veri-segmentasyonu-nedir/>

Güven, A. (2019, Aralık 21). *K-means kümeleme & elbow yöntemi*. Medium.

<https://medium.com/@anilguven1055/k-means-k%C3%BCmeleme-y%C3%B6ntemi-ed1c537046c6>

İnnova. (2021, Ağustos 2). *Gözetimsiz makine öğrenimi: Makineler kendi kendilerini nasıl eğitirler?*. İnnova.

[https://www.innova.com.tr/tr/blog/buyuk-veri-blog/gozetimsiz-makine-ogrenimi-makineler-kendi-kendilerini-nasil-egitirler#:~:text=G%C3%B6zetimsiz%20%C3%B6%C4%9Frenme%20\(dene timsiz%20%C3%B6%C4%9Frenme%2D%20unsupervised,%C3%B6%C4%9Frenme%20ve%20bulma%20yetene%C4%9Fi%20verir.](https://www.innova.com.tr/tr/blog/buyuk-veri-blog/gozetimsiz-makine-ogrenimi-makineler-kendi-kendilerini-nasil-egitirler#:~:text=G%C3%B6zetimsiz%20%C3%B6%C4%9Frenme%20(dene timsiz%20%C3%B6%C4%9Frenme%2D%20unsupervised,%C3%B6%C4%9Frenme%20ve%20bulma%20yetene%C4%9Fi%20verir.)

Mesci, Y. (2020, Mart 23). *Kümeleme teknikleri ve python'da k-means*. Medium

<https://medium.com/deep-learning-turkiye/k%C3%BCmeleme-teknikleri-ve-pythonda-k-means-62d93c624e5c>

Savaş, S. (2020, Mayıs 18). *Veri madenciliğinin tarihi ve kullanıldığı alanlar*. Medium.

<https://medium.com/veri-madencili%C4%9Fi/veri-madencili%C4%9Finin-tarihi-ve-kullan%C4%B1d%C4%B1%C4%9F%C4%B1-alanlar-19a1896c1505>

Sırma, E. tarih yok. *K-means algoritması ile müşteri segmentasyonu*. Kaggle.

<https://www.kaggle.com/code/emresirma/k-means-algoritmas-ile-m-teri-segmentasyonu>

Stat. tarih yok. *Faktör Analizi*. Stat. Erişim tarihi: 2023, 15 Mart.

<http://www.stat.gen.tr/index.php?istek=sinif&dersid=ist01&konuid=fak&max=1>

Şener, Y. (2020, Temmuz 31). *Segmentasyon Kullanımı ve 5 Adet Segmentasyon Çeşidi*. Medium.

<https://yigitsener.medium.com/segmentasyon-kullan%C4%B1m%C4%B1-ve-5-adet-segmentasyon-%C3%A7e%C5%9Fidi-e7a5ad1976bd>

Şirin, E. (2017, Ağustos 4). *K-ortalamlar kümeleme (k-means clustering) tekniğinde küme sayısını belirlemek*. Veri bilim okulu.

<https://www.veribilimiokulu.com/kumeleme-notlari-3-k-ortalamalar-kume-sayisini-belirleme/#:~:text=Bunu%20sa%C4%9Flayacak%20bir%20metrik%20var,%C5%9F%C3%B6yle%3A%20K%C3%BCmeler%20i%C3%A7i%20kareler%20toplam%C4%B1.>

Şirin, E. (2018, Ocak 7). *Anormallik analizinde kullanılan yöntemler (outlier analysis methods)*. Veri bilim okulu.

<https://www.veribilimiokulu.com/anormallik-analizinde-kullanilan-yontemler-outlier-analysis-methods/>

Taş, B. (2019, Ekim 20). *Python ile aykırı değer (outlier) analizi*. Medium.

<https://bernatas.medium.com/python-ile-ayk%C4%B1r%C4%B1-de%C4%9Fer-outlier-analizi-2b693bbc886d>

Uslu, M. (2018, Eylül 4). *Birliktelik kuralları analizi (association rules analysis)*.

Veri bilim okulu.

<https://www.veribilimiokulu.com/associationrulesanalysis/>

Wix Blog. (2021, Ağustos 17). *Crm nedir? Yeni başlayanlar için 'müşteri ilişkileri yönetimi*. Wix

<https://tr.wix.com/blog/makale/crm-nedir>

Vikipedi. tarih yok. *Karar ağacı*. Vikipedi. Erişim Tarihi: 2023, Şubat 25.

https://tr.wikipedia.org/wiki/Karar_a%C4%9Fac%C4%B1#:~:text=Karar%20a%C4%9Fac%C4%B1%2C%20bir%20hedefe%20ula%C5%9Fma,%C3%B6%C4%9Frenmesinde%20kullan%C4%B1lan%20yayg%C4%B1n%20bir%20ara%C3%A7t%C4%B1r.

Vikipedi. tarih yok. *K-means kümeleme*. Vikipedi. Erişim Tarihi: 2023, Mart 26

[https://tr.wikipedia.org/wiki/K-](https://tr.wikipedia.org/wiki/K-means_k%C3%BCmeleme#:~:text=%E2%80%9CK%E2%80%9D%20algoritmaya%20ba%C5%9Flamadan%20%C3%B6nce%20ihtiya%C3%A7,adet%20k%C3%BCmeyi%20tespit%20etmeye%20%C3%A7al%C4%B1%C5%9Fmaktad%C4%B1r.)

[means k%C3%BCmeleme#:~:text=%E2%80%9CK%E2%80%9D%20algoritmaya%20ba%C5%9Flamadan%20%C3%B6nce%20ihtiya%C3%A7,adet%20k%C3%BCmeyi%20tespit%20etmeye%20%C3%A7al%C4%B1%C5%9Fmaktad%C4%B1r.](https://tr.wikipedia.org/wiki/K-means_k%C3%BCmeleme#:~:text=%E2%80%9CK%E2%80%9D%20algoritmaya%20ba%C5%9Flamadan%20%C3%B6nce%20ihtiya%C3%A7,adet%20k%C3%BCmeyi%20tespit%20etmeye%20%C3%A7al%C4%B1%C5%9Fmaktad%C4%B1r.)

Yıldırım, E. (2020, Mayıs 2). *Yapay sinir ağı(artificial neural network) nedir?*. Veri bilim okulu.

<https://www.veribilimiokulu.com/yapay-sinir-agiartificial-neural-network-nedir/>

Yılmaz, E. (2018, Temmuz 19). *Faktör analizi nedir? Nasıl uygulanır*. Veri bilim okulu.

<https://www.veribilimiokulu.com/faktor-analizi-nedir-nasil-uygulanir/>