



T.C.

**BURSA ULUDAĞ ÜNİVERSİTESİ**  
**SOSYAL BİLİMLER ENSTİTÜSÜ**  
**EKONOMETRİ ANABİLİM DALI**  
**İSTATİSTİK BİLİM DALI**

**MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE MÜŞTERİ**  
**SEGMENTASYONU VE HEPSİBURADA E-TİCARET**  
**PLATFORMU ÜZERİNE BİR UYGULAMA**

**(YÜKSEK LİSANS TEZİ)**

**OKTAY ERGUN**

**BURSA – 2023**





**T.C.**

**BURSA ULUDAĞ ÜNİVERSİTESİ  
SOSYAL BİLİMLER ENSTİTÜSÜ  
EKONOMETRİ ANABİLİM DALI  
İSTATİSTİK BİLİM DALI**

**MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE MÜŞTERİ  
SEGMENTASYONU VE HEPSİBURADA E-TİCARET  
PLATFORMU ÜZERİNE BİR UYGULAMA**

**(YÜKSEK LİSANS TEZİ)**

**OKTAY ERGUN**

**Danışman:**

**Prof. Dr. Erkan IŞIĞIÇOK**

**BURSA – 2023**

## ÖZET

Yazar adı soyadı	Oktay ERGUN
Üniversite	Bursa Uludağ Üniversitesi
Enstitü	Sosyal Bilimler Enstitüsü
Anabilim dalı	Ekonometri
Bilim dalı	İstatistik
Tezin niteliği	Yüksek Lisans
Mezuniyet tarihi	...../...../2023
Tez danışmanı	Prof. Dr. Erkan IŞIĞIÇOK

### MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE MÜŞTERİ SEGMENTASYONU VE HEPSİBURADA E-TİCARET PLATFORMU ÜZERİNE BİR UYGULAMA

E-ticaret, geleneksel ticaret yöntemlerinin teknolojik gelişmeler sayesinde günümüze uyarlanmış halidir. Geçtiğimiz yıllarda Covid-19 pandemisinin de etkisiyle e-ticaret hızla büyümüşür ve e-ticaretteki bu büyüme hala devam etmektedir. Bu hızlı büyüme, işletmelere müşterilerini anlama ve müşterilerine özel ürün ve hizmet önerileri sunma ihtiyacını birlikte getirmiştir. Bu ihtiyaç doğrultusunda, işletmeler müşteri segmentasyonu yöntemi ile müşterilerini belirli gruplara ayırabilmektedir. Belirlenen her bir grubun karakteristik yapıları tespit edilerek bu gruplara özel ürün ve hizmet önerileri sunulabilmektedir. Bu sayede uzun vadede müşteri aidiyetinin sağlanması ve işletmenin karlılığının artması sağlanabilmektedir. Bu tez çalışmasında, 2022 yılı içerisinde Hepsiburada e-ticaret platformundan seçilen yaklaşık 5.000 kullanıcının 9 ana kategoriye ait toplamda 72.863 adet ürün siparişinden oluşan alışveriş verileri alınmıştır. Bu veri kümesinin betimsel istatistikleri özetlenmiştir ve uygulama öncesinde veri ön işleme işlemleri uygulanmıştır. Sonrasında temel bileşenler analizi ve k-ortalamlar makine öğrenmesi algoritmaları ile müşteri segmentasyonu işlemi gerçekleştirilmiştir. Yapılan analiz sonucunda kullanıcılar 7 farklı kümeye bölünmüştür ve her bir kümenin alışveriş alışkanlıkları ayrıntılı bir şekilde incelenmiştir.

**Anahtar kelimeler:** Makine Öğrenmesi, Veri Analizi, Veri Madenciliği, E-Ticaret, Python, Müşteri Segmentasyonu, K-ortalamlar, PCA

## ABSTRACT

<b>Name &amp; surname</b>	<b>Oktay ERGUN</b>
<b>University</b>	<b>Bursa Uludağ University</b>
<b>Institute</b>	<b>Institute of Social Sciences</b>
<b>Field</b>	<b>Econometrics</b>
<b>Subfield</b>	<b>Statistics</b>
<b>Degree awarded</b>	<b>Master</b>
<b>Date of degree awarded</b>	<b>...../...../2023</b>
<b>Supervisor</b>	<b>Prof. Dr. Erkan İŞİĞİÇOK</b>

### **CUSTOMER SEGMENTATION WITH MACHINE LEARNING ALGORITHMS AND AN APPLICATION OF HEPSIBURADA E-COMMERCE PLATFORM**

E-commerce is the adaptation of traditional commerce methods to the present day, regarding to technological possibilities. In the past years, e-commerce has grown rapidly due to the Covid-19 pandemic, and this growth in e-commerce still continues. This rapid growth has brought the need for businesses to understand their customers and offer them specific product and service recommendations. In order to meet with this need, businesses can divide their customers into specific groups with the customer segmentation method. By determining the characteristics of each determined group, unique product and service suggestions can be offered to these groups. In this way, customer loyalty and business profitability can be increased in the long term. In this thesis, shopping data consists of 72,863 product orders belonging to 9 main categories of approximately 5,000 users selected from the Hepsiburada e-commerce platform in 2022 were obtained. The descriptive statistics of this dataset were summarized, and data preprocessing was applied before the application. Afterward, the customer segmentation process was performed with principal component analysis and k-means machine learning algorithms. As a result of the analysis, users were divided into 7 different clusters, and the shopping habits of each cluster were examined in detail.

**Keywords:** Machine Learning, Data Analysis, Data Mining, E-commerce, Python, Customer Segmentation, K-means, PCA

## ÖNSÖZ

Öncelikle, üzerimde çok büyük emeđi olan ve tez sürecimin tamamında bana yol göstererek son derece yardımcı olan deđerli tez danışmanım Prof. Dr. Erkan IŞIĞIÇOK'a çok teşekkür ederim.

Kariyerime büyük katkısı olan deđerli akıl hocam Sarp Gürol'a ve bu tez çalışmasının gerçekleştirilebilmesi için ihtiyacım olan veriyi sağlayan Hepsiburada'ya teşekkür ederim.

Son olarak bu süreçte bana her zaman inanan ve hayatım boyunca sonsuz desteklerini hiç eksik hissetmediđim çok deđerli aileme ve eşime çok teşekkür ederim.

Bursa, 2023

Oktay ERGUN

## İÇİNDEKİLER

ÖZET.....	i
ABSTRACT.....	ii
ÖNSÖZ.....	iii
TABLolar .....	viii
ŞEKİLLER .....	ix
KISALTMALAR .....	x
GİRİŞ .....	1

## BİRİNCİ BÖLÜM

### E-TİCARET VE MÜŞTERİ İLİŞKİLERİ YÖNETİMİ (CRM)

1.1. E-TİCARETİN TANIMI.....	3
1.2. E-TİCARETİN TARİHİ .....	4
1.3. E-TİCARETİN GELİŞİMİNİ ETKİLEYEN FAKTÖRLER .....	5
1.4. E-TİCARETİN AVANTAJLARI.....	6
1.5. E-TİCARETİN TARAFLARI .....	7
1.5.1. Satıcılar.....	8
1.5.2. Müşteriler .....	8
1.5.3. Teknoloji Hizmeti Sağlayıcıları .....	8
1.5.4. Nakliye ve Kargo Firmaları.....	9
1.5.5. Elektronik Ödeme Hizmeti Sağlayıcıları .....	9
1.6. E-TİCARET TÜRLERİ .....	9
1.6.1. İşletmeden İşletmeye.....	10
1.6.2. İşletmeden Tüketicie .....	10
1.6.3. Tüketiciden Tüketicie.....	10
1.6.4. Tüketiciden İşletmeye .....	11

1.6.5. İşletmeden Kamu Kurumuna .....	11
1.6.6. İşletmeden Çalışanlarına .....	11
1.6.7. Kamu Kurumundan İşletmeye .....	11
1.6.8. Kamu Kurumundan Tüketiciye .....	11
1.6.9. Kamu Kurumundan Çalışanlarına .....	12
1.6.10. Kamu Kurumundan Kamu Kurumuna .....	12
1.6.11. Çiftçiden Tüketiciye .....	12
1.6.12. İşletmeden İşletmeye ve Tüketiciye .....	12
1.6.13. Direkt Tüketiciye .....	13
1.6.14. Makineden Makineye .....	13
<b>1.7. E-TİCARETTE KULLANILAN ARAÇLAR .....</b>	<b>14</b>
1.7.1. E-ticaret Platformları .....	14
1.7.2. Elektronik Ödeme Sistemleri .....	15
1.7.3. Veri Yönetimi ve İstatistiksel Veri Analizi Araçları .....	15
1.7.4. Pazarlama ve İletişim Araçları .....	16
<b>1.8. E-TİCARET VE MÜŞTERİ İLİŞKİLERİ YÖNETİMİ (CRM) .....</b>	<b>17</b>
1.8.1. Müşteri İlişkileri Yönetiminin (CRM) Tanımı .....	17
1.8.2. Müşteri İlişkileri Yönetiminin Tarihsel Gelişimi .....	18
1.8.3. Müşteri İlişkileri Yönetiminin Avantajları .....	20
1.8.4. Müşteri İlişkileri Yönetiminin İlkeleri .....	21
1.8.5. Müşteri İlişkileri Yönetiminin Kurulması ve Uygulanması .....	22
1.8.6. Müşteri İlişkileri Yönetiminin Süreçleri .....	23
1.8.6.1. Veri Yönetimi Aşaması (Bilgi Toplama) .....	23
1.8.6.2. Satış Yönetimi Aşaması (Müşteri Edinme) .....	23
1.8.6.3. Müşteri Hizmetleri Yönetimi (Müşterinin Sadakatini Kazanma) .....	24
1.8.6.4. Müşteri Değerini Arttırma .....	24
1.8.7. Müşteri İlişkileri Yönetiminde Kullanılan Teknolojiler .....	24
<b>1.9. E-TİCARETİN EKONOMİYE ETKİLERİ .....</b>	<b>27</b>
<b>1.10. E-TİCARETİN GELECEĞİ .....</b>	<b>28</b>



## İKİNCİ BÖLÜM

### MAKİNE ÖĞRENMESİ

<b>2.1. MAKİNE ÖĞRENMESİ KATEGORİLERİ .....</b>	<b>31</b>
2.1.1. Denetimli Öğrenme .....	32
2.1.1.1. Regresyon .....	33
2.1.1.2. Sınıflandırma .....	34
2.1.2. Denetimsiz Öğrenme .....	35
2.1.2.1. Kümeleme .....	36
2.1.2.2. Boyut İndirgeme .....	38
2.1.3. Pekiştirmeli Öğrenme .....	38
<b>2.2. MAKİNE ÖĞRENMESİ ALGORİTMALARI .....</b>	<b>39</b>
2.2.1. k-En Yakın Komşu .....	39
2.2.2. Karar Ağaçları .....	40
2.2.3. Rastgele Orman .....	42
2.2.4. Yapay Sinir Ağları .....	43
2.2.5. Temel Bileşenler Analizi (PCA) .....	44
2.2.6. K-Ortalamalar .....	45

## ÜÇÜNCÜ BÖLÜM

### MÜŞTERİ SEGMENTASYONU UYGULAMASI

<b>3.1. KULLANILAN ARAÇLARIN TANIMLANMASI .....</b>	<b>49</b>
3.1.1. Python Programlama Dili .....	49
3.1.2. Anaconda, Anaconda Navigator ve Jupyter Notebook .....	49
3.1.3. NumPy Kütüphanesi .....	50
3.1.4. Pandas Kütüphanesi .....	50
3.1.5. Matplotlib Kütüphanesi .....	50
3.1.6. Scikit-learn Kütüphanesi .....	50
<b>3.2. VERİ KÜMESİNİN TANIMLANMASI .....</b>	<b>50</b>
3.2.1. Betimsel İstatistikler .....	54

3.2.2. Veri Görselleştirme .....	56
<b>3.3. VERİ ÖNİŞLEME .....</b>	<b>61</b>
<b>3.4. MAKİNE ÖĞRENMESİ.....</b>	<b>64</b>
<b>SONUÇ.....</b>	<b>73</b>
<b>KAYNAKÇA .....</b>	<b>76</b>
<b>EKLER.....</b>	<b>83</b>
Ek 1: Veri setinin tanımlanması .....	83
Ek 2: Veri görselleştirme.....	85
Ek 3: Veri önışleme .....	88
Ek 4: Makine öğrenmesi .....	89

## TABLÖLAR

Tablo 1: Veri kümesinin tanımlanması.....	51
Tablo 2: Veri kümesini oluşturan deęişkenler ve açıklamaları.....	52

## ŞEKİLLER

Şekil 1: E-ticaret türleri.....	13
Şekil 2: İşletme Yönetiminin Tarihsel Gelişimi.....	19
Şekil 3: DIKW Modeli .....	25
Şekil 4: Makine öğrenmesi kategorileri.....	31
Şekil 5: Örnek bir regresyon modeli .....	33
Şekil 6: Örnek bir sınıflandırma modeli.....	35
Şekil 7: Kümeleme teknikleri .....	37
Şekil 8: K-en yakın komşu yöntemi.....	40
Şekil 9: Karar ağaçlarının yapısı.....	41
Şekil 10: Karar ormanı .....	43
Şekil 11: Yapay sinir ağı.....	44
Şekil 12: Temel bileşenler analizi (PCA) algoritması ile boyut indirgeme.....	45
Şekil 13: K-ortalamlar algoritması ile veri kümeleme .....	47
Şekil 14: Veri kümesini oluşturan değerleri veren Python kodları .....	51
Şekil 15: Tip dönüştürme işlemi uygulanmadan önce veri kümesindeki değişkenler....	53
Şekil 16: Tip dönüştürme işlemi uygulandıktan sonra veri kümesindeki değişkenler ...	53
Şekil 17: Veri kümesindeki sayısal değişkenlerin betimsel istatistikleri.....	54
Şekil 18: Veri kümesindeki en çok sipariş veren kullanıcılar.....	55
Şekil 19: Veri kümesindeki tek bir siparişte en çok ürün satın alan kullanıcılar.....	56
Şekil 20: Aykırı değerlerin veri setinden çıkarılması .....	57
Şekil 21: Aykırı değerlerin çıkarılmasından önce ve sonra Skewness ve Kurtosis değerleri.....	58
Şekil 22: Aylık Satış Grafiği.....	59
Şekil 23: Fiyat Histogram Grafiği.....	60
Şekil 24: Kategorilerin Satış Oranları Grafiği.....	61
Şekil 25: Kullanıcıların kategorilere ait harcamalarının hesaplanması.....	62
Şekil 26: Kullanıcıların harcama istatistiklerinin hesaplanması.....	63
Şekil 27: Veri dönüştürme işlemi sonrası final tablo.....	64
Şekil 28: Temel Bileşen Analizi ile bileşen sayısının seçilmesi.....	65
Şekil 29: Elbow yönteminin uygulanması.....	66
Şekil 30: kElbowVisualizer metodu ile Elbow yönteminin uygulanması.....	67
Şekil 31: K-ortalamlar algoritmasının uygulanması.....	67
Şekil 32: Kümelenen veri setinin PCA yöntemi ile görselleştirilmesi .....	68
Şekil 33: Kümelere göre gruplanan değişkenlerin ortalamaları.....	69
Şekil 34: Kümelerin değişkenler bazında ortalama harcamaları.....	70

## KISALTMALAR

AR	Augmented Reality
B2B	Business-to-business
B2B2C	Business-to-business-to-consumer
B2C	Business-to-consumer
B2E	Business-to-employee
B2G	Business-to-government
C2B	Consumer-to-business
C2C	Consumer-to-consumer
CRM	Customer Relationship Management
D2C	Direct-to-consumer
DIKW	Data – Information – Knowledge – Wisdom
EDI	Electronic Data Interchange
ETBİS	Elektronik Ticaret Bilgi Sistemi
G2B	Government-to-business
G2C	Government-to-consumer
G2E	Government-to-employee
G2G	Government-to-government
E-ticaret	Elektronik ticaret
EU	European Union
F2C	Farmer-to-consumer,
KDD	Knowledge Discovery in Databases
M2M	Machine-to-machine
OECD	Organisation for Economic Co-operation and Development
PCA	Principal Component Analysis
WB	World Bank
WTO	World Trade Organization

## GİRİŞ

Günümüzde, internetin hayatın her alanına etki ettiği ve kişilerin hayat tarzını değiştirdiği görülebilmektedir. İnternet teknolojilerinin gelişmesi ile birlikte geleneksel alışveriş yöntemlerinin de dijitale taşınmasıyla günümüzdeki elektronik ticaret kavramı ortaya çıkmıştır. Kısaca e-ticaret, ürün veya hizmetlerin çevrimiçi alım satımı işlemleri olarak açıklanabilir (Mohapatra, 2013).

E-ticaretin tarihsel gelişiminde teknolojik gelişmelerin çok büyük bir etkisi olmuştur. İnternetin yaygınlaşmasıyla başlayan sürecin ardından, cep telefonlarının ve mobil uygulamaların yaygınlaşmasının e-ticaretin geniş kitlelere ulaşmasında en çok etkisi bulunan teknolojik gelişmeler olduğu söylenebilir. Günümüzde, veri bilimi ve yapay zekâ disiplinlerinde yaşanan teknolojik gelişmeler sayesinde de e-ticaret işletmeleri müşterilerini daha iyi anlamaya ve müşterilerinin ihtiyaçlarına özel ürün ve hizmetler sunmaya başlamışlardır.

E-ticaret işletmeleri, topladıkları verileri anlamlandırarak, veriye dayalı kararlar vermektedirler. Diğer bir ifadeyle işletmelerinde veri odaklı bir yaklaşım benimsemişlerdir. İşletmeler için en önemli kararlar içgüdülere dayanarak değil, verilere (gerçeklere) dayandırılarak verilmektedir. Verilere dayalı kararlar veren işletmelerin daha karlı işletmeler olduğu, müşterilerinin ihtiyaçlarını daha iyi anlayabildiği ve bu ihtiyaçlara daha iyi çözümler bulabildiği görülebilmektedir.

Müşterilere özel ürün ve hizmetler sunmanın birkaç farklı yöntemi vardır. Örneğin, müşterilerin geçmiş siparişlerine bakarak, bir kullanıcı profili çıkarılarak gelecekte alacağı ürünler tahmin edilmeye çalışılabilir. Yaygın kullanılan bir diğer yöntem ise, benzer alışveriş alışkanlıkları olan müşterileri kendi aralarında kümelere (müşteri segmentasyonu) ayırmaktır. Bu sayede, küme içerisinde bulunan kullanıcılar benzer alışveriş alışkanlıkları sergiledikleri için bu kullanıcıların satın aldığı ürünler küme içerisindeki diğer kullanıcılara önerilebilir. Ayrıca, kümelerin karakteristik özellikleri çıkarılarak o kümelere yönelik pazarlama çalışmalarında bulunulabilir. Makine öğrenmesi yöntemleri sayesinde tavsiye sistemleri kurarak bu işlemler otomatik bir şekilde gerçekleştirilebilmektedir. Özetle, bir tavsiye sistemi müşterilerini öğrenir ve

müşterilere mevcut ürünler arasından en değerli bulacağı ürünleri önerir (Schafer vd., 1999).

Bu tez çalışmasında, Hepsiburada e-ticaret platformu müşterileri verilerinden oluşan bir veri kümesi üzerinde müşteri segmentasyonu uygulaması gerçekleştirilmiştir. Çalışmada 2022 yılında Hepsiburada e-ticaret platformu müşterilerinden rastgele olarak seçilen yaklaşık 5.000 kullanıcının alışveriş verileri kullanılmıştır. Veri kümesinde, kullanıcıların alışveriş verileri 9 ana kategoriden verilen 43.000'den fazla sipariş içermektedir. Bu siparişlerde toplamda 72.863 adet ürün satın alındığı görülmektedir. Tez çalışması gereği, her bir ürüne yapılan harcamayı o ürünün ait olduğu kategori altında toplayarak, her bir kullanıcının ana kategoriler arasında ne kadarlık bir harcama yaptığı hesaplanmıştır.

Veri kümesi hakkında bilgi sahibi olabilmek için veri keşif aşaması yapılmış, veri kümesinin betimsel istatistikleri çıkarılmış ve veri kümesi hakkında bilgiler verebilecek bazı veri görselleştirmeleri yapılmıştır. Ardından, veri kümesinin makine öğrenmesi modeline uygun hale getirilebilmesi için bazı veri ön işleme yöntemlerinden yararlanılmıştır.

Denetimsiz öğrenme modellerinden olan temel bileşenler analizi yöntemi ile veri kümesinin boyutu indirgenerek yine bir diğer denetimsiz öğrenme modeli olan k-ortalamlar yöntemi ile müşteri segmentasyonu uygulaması gerçekleştirilmiştir. Yapılan analizlerde, kullanıcıların en iyi yedi farklı kümede temsil edilebileceği gözlemlenmiştir. Uygulama sonucunda, yedi farklı kümeye ayrılan kullanıcıların alışveriş alışkanlıklarını gözlemleyebilmek için kümelerin her bir değişkeninin ortalama değerleri incelenmiştir.

Bu tezin amacı, makine öğrenmesi algoritmaları ile müşteri segmentasyonu uygulaması yapmaktır. Bu amaçla, bu tezin ilk bölümünde, e-ticaret ve müşteri ilişkileri yönetimi (CRM) hakkında bilgi verilmiştir. Tezin ikinci bölümünde ise, makine öğrenmesi kategorileri ile bazı sık kullanılan makine öğrenmesi algoritmaları ve uygulamada kullanılan algoritmalar tanıtılmıştır. Son olarak üçüncü bölümde temel bileşenler analizi (PCA) ve k-ortalamlar algoritmaları ile müşteri segmentasyonu çalışması yapılmıştır.

# **BİRİNCİ BÖLÜM**

## **E-TİCARET VE**

### **MÜŞTERİ İLİŞKİLERİ YÖNETİMİ (CRM)**

#### **1.1. E-TİCARETİN TANIMI**

Hayatımıza kısaca e-ticaret olarak yerleşmiş olan elektronik ticaret kavramını birçok şekilde tanımlamak mümkündür. Avrupa Birliği; e-ticaret kavramını, mal ve hizmetlerin internet üzerinden satılması olarak tanımlarken Ekonomik Kalkınma ve İşbirliği Örgütü (OECD) ise internet üzerinden yapılan ticari işlemler olarak tanımlamaktadır. 6563 sayılı Elektronik Ticaretin Düzenlenmesi Hakkında Kanun'da ise fiziki olarak karşı karşıya gelmeden, elektronik ortamda gerçekleştirilebilen çevrimiçi iktisadi ve ticari her türlü faaliyet e-ticaret olarak ifade edilmektedir (Ticaret.gov.tr, 2019). Diğer bir deyişle geleneksel nitelikteki ticari işlemlerin internet aracılığıyla yapılmasına e-ticaret denilmektedir.

Fiziksel olarak karşı karşıya gelmeden ticaret yapabilme arzusu internet çağından çok daha önce başlamıştır. E-ticaretten önceki dönemlerde posta, faks ve telefon gibi araçlarla da ticaret yapılabilmekteydi (Deliçay, 2021). Bu yöntemlerin doğası gereği bazı sınırlılıklara sahip olacağını öngörebilmek mümkündür. Geleneksel yöntemlerle ticareti yapılan ürün ve hizmetler hakkında görsel materyallerin, daha önce bu ürün ve hizmetleri almış kişilerin değerlendirmelerinin olmaması ve alternatif seçeneklerin kolaylıkla sunulamaması gibi sınırlılıklar ticaretin gerçekleşmesini zorlaştırmaktadır. Bu ve benzeri sınırlılıklar geleneksel ticaretin potansiyeline tam olarak ulaşamamasına neden olmaktadır. Kuşkusuz, e-ticaretin tam potansiyeline ulaşabilmesi için internetin sağladığı tüm faydalardan yararlanması da gerekmektedir.

İnternet teknolojisi sayesinde firmaları geniş kitlelere ulaştırması ve satış işlemini kolaylaştırması, tüketiciler için ise uygun fiyatlarla kaliteli ürün seçebilme imkanı sağlaması e-ticaretin hızlı bir şekilde yayılmasını sağlamıştır. Bu yayılım internet teknolojisinde yaşanan baş döndürücü gelişmelere paralel olarak devam etmektedir. Buna ek olarak, e-ticaret sadece internet üzerinden ürün alıp satmak ile kısıtlı değildir. Online



fatura ödemek, etkinlik bileti almak, medya yayın hizmetlerine abone olmak ve online eğitim vb. gibi dijital içerikler almak ve online banka işlemleri gibi işlemler de birer e-ticaret işlemidir.

Diğer taraftan, e-ticaret bir proses inovasyonu olarak da varsayılmaktadır (Burt & Sparks, 2003). Diğer bir ifadeyle; e-ticaretin geleneksel ticaret yöntemlerinin ve ilişkilerinin, teknolojinin sağladığı imkanlar sayesinde yeniden yapılandırılması olduğunu söylemek yanlış olmayacaktır. Ayrıca, e-ticaretin insanlık tarihi kadar eski olan bir prosesin günümüz çağına ayak uydurması veya dijitalleşmesi olduğunu söylemek de mümkündür.

## **1.2. E-TİCARETİN TARİHİ**

İnternetin son kullanıcılar arasında hızla yayılması ve beraberinde her sektörden şirketlerin internete taşınması tarihteki en önemli dönüşümlerden biri olarak görülmektedir. Hem yıllardır geleneksel şekilde işlerini yürüten şirketler hem de internet üzerinden hizmet vermek için yeni kurulan girişimler bu teknolojiden yararlanmaya çalışmışlardır. Diğer bir ifadeyle her sektörden küçük veya büyük tüm şirketler internet ortamında bulunmak, daha görünür olmak, işlerini büyütmek, daha fazla kişiye ulaşmak ve bu sayede daha fazla satış yapmak için çaba göstermişlerdir.

Elektronik ticaretin (e-ticaretin), internetle birlikte hayatımıza giren en önemli kavramlardan biridir ve en önemli bilimsel başarılarından biri olduğu varsayılmaktadır (Qin, 2009). E-ticaretin 1970'li yıllarda kapalı bir internet sistemi kurmaya gücü yeten büyük şirketlerin çalışanları arasında dokümanlar ile bilgi ticareti yaparak başladığı savunulmaktadır. Elektronik dokümanların ve verilerin işletmeler arasında birbirlerine iletilmesi ve karşılıklı olarak değişilmesine elektronik veri değişimi (Electronic Data Interchange - EDI) denilmektedir (Uygur, 2010).

İnternet, 1990'larda özellikle iletişim kurmayı kolaylaştırmak için kullanılmaktaydı, ancak bazı şirketler internetin kendilerine daha fazla fayda sağlayabileceğinin farkına varmıştır. Bu şirketler önce internet üzerinden sağladıkları ürün ve hizmetlerini listeleterek bu ürünler hakkında bilgi vermeye daha sonra ise bu ürün ve hizmetleri direkt

olarak müşterilerine satmaya başlamışlardır (Santos vd., 2017). 20. yüzyılın sonlarında ortaya çıkan e-ticaret endüstrisi, matbaanın icadı veya endüstri devrimi kadar önemli görülmektedir. Bununla birlikte e-ticaretin gerçekleşen teknolojik devrimlerin arasında en büyüğü olduğu da düşünülmektedir (Canpolat, 2001).

İnsanların internete erişiminin kolaylaşması ve şirketlerin uluslararası pazarlara açılıp büyüme arzusu ile birlikte, e-ticaret hızla büyümeye başlamıştır. E-ticaret öncelikle gelişmiş ülkelerde hızla yaygınlaşmaya başlamış olsa da günümüzde hemen hemen her ülkede kendine bir yer edinmiştir. Günümüzde ise e-ticarette ABD ve Çin lider konumda iken ardından Avrupa ülkeleri gelmektedir.

### **1.3. E-TİCARETİN GELİŞİMİNİ ETKİLEYEN FAKTÖRLER**

Bir e-ticaret işleminin yapılabilmesi için birçok farklı unsurun bir arada çalışması gerekmektedir. En basit örnekle günümüzde sıradan bir e-ticaret işleminin gerçekleşmesi için bir e-ticaret platformu, alıcı, satıcı, ürün veya hizmet, ödeme yöntemi ve eğer ürün fiziksel ise ürün teslimatı gibi birçok unsurun bir araya gelmesi gerekmektedir. E-ticaret bu unsurlardan herhangi birinde meydana gelen teknik gelişmelerden direkt olarak etkilenmektedir. Diğer bir ifadeyle lojistik sektöründe yaşanan gelişmeler sayesinde günümüzde e-ticaret siteleri müşterilerine aynı gün teslimat özelliği sunabilmektedir. Benzer şekilde internet hizmetlerinin geçmişte bant genişliğinin artması ve halk arasında yaygınlaşması daha fazla kişinin e-ticaret hizmetlerine ulaşılabilmesinin yolunu da açmıştır. Ek olarak online ödeme yöntemlerinin bankalar ile entegrasyonu, ödeme işlemlerinin anlık sağlanması, güvenlik önlemlerinin artırılması vb. gibi gelişmeler de e-ticaretin gelişimine katkı sağlayan unsurlara örnek olarak verilebilir. Kuşkusuz, 2019'da başlayan ve tüm dünyayı etkisi altına alan pandemi (covid) salgınının da e-ticaret ve dijitalleşme üzerine önemli etkisi olmuştur. Yazılım, müzik, e-kitap, online eğitim gibi dijital ürünlerin ticarileşmesi de e-ticaretin gelişimine katkı sağlayan teknolojik gelişmeler arasında yer almaktadır.

Çevrimiçi pazaryerlerinin icadı e-ticaretin gelişimine en büyük katkı sağlayan gelişimlerden biri olmuştur. Bir markanın kendi sitesinden sadece kendi ürünlerini

satmasına kıyasla, çevrimiçi pazaryerlerinde birçok firma ve farklı markalardan ürünler satılabilmektedir. Çevrimiçi pazaryerleri hem tüketici için daha geniş bir yelpaze sunmakta hem de küçük işletmelerin baş edemeyeceği platform maliyetlerini üstlenmektedir. Aynı zamanda pazaryerleri müşterilerin en çok kullandığı e-ticaret platformları haline gelmiştir. Bu sayede küçük işletmeler pazaryerlerine gelen geniş müşteri kitlesine ürün sergileme olanağı bulmuştur.

Cep telefonlarının yaygınlaşması ve mobil uygulama teknolojilerin gelişimi sayesinde e-ticaret kullanıcılarının %70'i alışverişlerini mobil uygulamalar aracılığıyla yapmaktadır (T.C. Ticaret Bakanlığı, 2022). Günümüzde artırılmış gerçeklik (AR) teknolojisinin gelişimi sayesinde e-ticaret siteleri uygulamalarından fiziksel ürünlerin evimizde veya üzerimizde nasıl görüneceği denenebilmektedir. Yine günümüzde evlerinde üç boyutlu yazıcı olan kişiler internetten üç boyutlu modelleme programlarında daha önceden hazırlanmış ve satılan program dosyalarını bilgisayarlarına indirip kendi yazıcılarından bu ürünleri yazdırabilmektedir.

Veri bilimi ve yapay zekâ alanlarında meydana gelen gelişmeler sayesinde, e-ticaret daha akıllı bir hale gelmiştir. Hedefe yönelik pazarlama, müşterilerin ilgi alanlarına yönelik ürün önerileri, bu tezin de konusu olan müşteri segmentasyonu (müşteri bölümlendirilmesi), dinamik fiyatlandırma, tahminleme yöntemleri vb. daha birçok teknik sayesinde, e-ticaret siteleri müşterilerine daha yararlı hizmetler sunabilmektedir. Özetle; e-ticaretin tarihsel gelişiminde özellikle bilgi teknolojilerinin gelişiminin oldukça yakından etkilenmiştir (Santos vd., 2017).

#### **1.4. E-TİCARETİN AVANTAJLARI**

E-ticaret, geleneksel ticaret yöntemleriyle karşılaştırıldığında birçok avantaja sahiptir. Bu avantajlar müşteriler, işletmeler ve toplum için olmak üzere üçe ayrılabilir. Bu avantajlardan ilki müşteriler için e-ticaret sitelerinin 7 gün 24 saat ulaşılabilir olmasıdır. Müşterilerin evlerinden ayrılmadan dilediği ürünleri seçebilmesi, çevrimiçi ödeme yapabilmesi ve ürünlerin adreslerine teslim edilmesi müşteriler için büyük bir avantajdır. E-ticaret, alışveriş işleminin ilk aşamasından son aşamasına kadar online olarak

gerçekleşmesini sağlamaktadır. Bunlara ek olarak müşterinin ilgilendiği ürünleri satın alan diğer müşterilerin o ürünler hakkında yaptığı değerlendirmelere erişebilmesi de bir başka avantaj olarak karşımıza çıkmaktadır. Ayrıca yaşadığı konum veya fiziksel engeli nedeniyle, ilgilendiği ürünlere fiziksel olarak erişim imkanı olmayan müşterilerin, çevrimiçi olarak neredeyse dünyadaki tüm ürünlere zaman ve mekan kısıtı olmadan ulaşabilmesi, farklı markaları ve ürünleri kolaylıkla kıyaslayabilmesi de e-ticaretin müşterileri için sağladığı önemli avantajlar arasındadır.

İşletmeler için ise çok daha geniş bir müşteri kitlesine ulaşma imkanı en büyük avantajdır. E-ticaret sayesinde işletmeler, konum fark etmeksizin daha fazla müşteriye ulaşabilmektedir. Bu durum işletmenin bulunduğu ülke sınırlarını aşarak global pazarlara erişmesini mümkün kılmaktadır. Aynı zamanda; işletmeler, e-ticaret yöntemiyle geleneksel yöntemlere göre daha az masrafla satış yapabilme imkanına da sahip olmaktadır. Geleneksel ticaret yöntemlerinde ortaya çıkan fiziksel bir mağaza kiralama, kiralanmış mağazanın sabit giderleri, çalışan ücretleri ve stok maliyetleri gibi maliyetler, e-ticaret sayesinde önemli ölçüde azaltılabilmektedir. Diğer bir ifadeyle, e-ticaret, işletmelerin pazar payını artırır ve satış maliyetlerini azaltır. Özellikle küçük işletmeler e-ticaret platformları sayesinde ürün ve hizmetlerini daha uygun maliyetlerle geniş kitlelere tanıtılabilmekte ve pazarlayabilmektedir.

E-ticaret toplum için de önemli katkılarda bulunmaktadır. Satışlardan ödenen vergiler, sektörde oluşturduğu iş imkanları, e-ticaret firmalarının küçük işletmeleri ve girişimcileri destekleyen sosyal sorumluluk projeleri, firmalara yurt dışına satış yapabilme imkânı (e-ihracat) tanınması toplum için yaptığı katkılar arasında sayılabilir. Geleneksel ticaret yöntemlerine kıyasla e-ticaretin sahip olduğu tüm bu avantajlar ve daha fazlası e-ticaretin tarihsel gelişiminde büyük rol oynamıştır.

## **1.5. E-TİCARETİN TARAFLARI**

Geleneksel ticaret yöntemlerinin dijitalleşmiş hali olan e-ticaret işleminin gerçekleşebilmesi için uygun altyapının kurulmuş olması ve gerekmektedir. Aynı zamanda tıpkı geleneksel ticaret yöntemlerinde olduğu gibi e-ticaret işlemlerinin

gerçekleşebilmesi için tarafların bir araya gelmesi gerekmektedir. E-ticaret işlemlerinin gerçekleşmesinde rol oynayan taraflar kişiler ya da kurumlardan oluşabilir. Bu durumda e-ticaretin tarafları i) Satıcılar, ii) Müşteriler, iii) Teknoloji hizmeti sağlayıcıları, iv) Nakliye ve kargo firmaları, v) Elektronik ödeme hizmeti sağlayıcıları olmak üzere, beş başlık altında toplanabilir.

### ***1.5.1. Satıcılar***

Satıcılar, e-ticaret işlemlerindeki ürün ve hizmetlerin müşterilere satışını yapan kişi veya kurumlardır. Satıcılar, e-ticaret platformları üzerinde ürün ve hizmetleri sunarlar ve satılmaları için pazarlama çalışmaları yürütebilirler. Satıcılar, birden çok platformda ürün ve hizmetlerini satabilir ya da kendi platformlarını oluşturabilirler.

### ***1.5.2. Müşteriler***

Satıcıların sunduğu ürün ve hizmetleri satın alan kişi ve/veya kurumlardır. Diğer bir ifadeyle; müşteriler, sunulan ürün ve hizmetlerin tüketicisidirler. Günümüzde şirketlerin en çok önemsendiği kavramlardan biri müşteri odaklılık olarak karşımıza çıkmaktadır. Şirketler, müşteriyi odak noktalarına yerleştirerek yeniden yapılanmak zorunda kalmışlar ve stratejilerini de bu yönden değiştirmişlerdir (Demir & Kırdar, 2007).

### ***1.5.3. Teknoloji Hizmeti Sağlayıcıları***

Bir e-ticaret platformunun kesintisiz ve sorunsuz çalışabilmesi için birçok farklı teknolojinin uyum içinde çalışması gerekmektedir. E-ticaret platformlarının web siteleri ve mobil uygulamaları yazılımlarının geliştirilmesi ve düzenli aralıklarla güncellenmesi gerekmektedir. Web sitesinde veya mobil uygulamalarda gösterilen ürün ve hizmet kataloğunun veri tabanlarında depolanması, güncel tutulması ve her zaman erişilebilir olması gerekir. Müşterilerin faaliyetlerinin uygun bir şekilde toplanması ve analiz edilmesi de günümüzde bir ihtiyaç haline gelmiştir. Aynı zamanda, tüm bu süreçlerin güvenliğinin sağlanması, izinsiz erişimin engellenmesi ve bütünlüğün korunması da gerekmektedir.

#### ***1.5.4. Nakliye ve Kargo Firmaları***

Gerçekleşen bir e-ticaret işleminde alışverişi yapılan ürün fiziksel bir ürün ise satıcının bu ürünü müşterisine ulaştırması için bir kargo şirketiyle ulaştırması gerekmektedir. E-ticaret alışverişlerinde müşterilerin en önemsedığı konulardan biri, ürününün eline hızlı ve güvenli bir şekilde ulaşmasıdır. Günümüzde birçok e-ticaret şirketi nakliye aşamasının müşteri deneyimi üzerindeki etkisinin ne kadar büyük olduğunun farkına varmış ve süreci daha iyi yönetebilmek adına kendi dağıtım şirketlerini kurmuşlardır.

#### ***1.5.5. Elektronik Ödeme Hizmeti Sağlayıcıları***

Günümüzde e-ticaret işlemlerinde herhangi bir ürün veya hizmet satın alınırken genellikle elektronik ödeme tercih edilmektedir. Elektronik ödeme yöntemlerinin tercih edilmesi başlarda kuşkuyla karşılanmış ve alışılması zaman almış olsa da günümüzde en çok tercih edilen ödeme yöntemi olarak karşımıza çıkmaktadır. Elektronik ödeme hizmetlerini sağlayan firmaların ve bankaların işlem başı aldığı komisyon ücretleri ve müşterilerin kart bilgilerinin güvenliğinin sağlanması endişesi e-ticaret firmalarını kendi elektronik ödeme hizmetlerini geliştirmeye teşvik etmiştir.

### **1.6. E-TİCARET TÜRLERİ**

E-ticaret kavramını farklı şekillerde alt gruplara ayırmak mümkündür. Bunlardan en yaygın olduğu varsayılan gruplandırma yöntemi, e-ticaret işlemlerinde alışverişi yapan taraflar arasındaki ilişkiye dayanarak gruplara ayırmaktır. İşletmelerin tüketicilere ürün ve hizmet sunması en bilinen e-ticaret yöntemi olsa da daha birçok e-ticaret yöntemi saymak mümkündür. Buradan hareketle; e-ticaret türleri, i) İşletmeden işletmeye, ii) İşletmeden tüketiciye, iii) Tüketiciden tüketiciye, iv) Tüketiciden işletmeye, v) İşletmeden kamu kurumuna, vi) İşletmeden çalışanlarına, vii) Kamu kurumundan işletmeye, viii) Kamu kurumundan tüketiciye, ix) Kamu kurumundan çalışanlarına, x) Kamu kurumundan kamu kurumuna, xi) Çiftçiden tüketiciye, xii) İşletmeden işletmeye

ve tüketiciye, xiii) Direkt tüketiciye ve xiv) Makineden makineye şeklinde gruplandırılabilir.

### ***1.6.1. İşletmeden İşletmeye***

Ticaretin işletmeler arasında gerçekleştiği yani işletmeden işletmeye (Business-to-business, B2B) olarak da bilinen B2B, en yaygın e-ticaret yöntemlerinden biridir. Örneğin, bir otomotiv firmasının üretimde kullandığı bir malzemeyi farklı bir firmadan alması veya bir gıda firmasının başka bir firmadan ürün alması, işletmeler arası e-ticaret yöntemidir. Günümüzde, internet üzerinden yapılan yani işletmeden işletmeye alışverişlerin büyük bir payını işletmeler arasında yapılan alışverişler oluşturmaktadır (Turban vd., 2015).

### ***1.6.2. İşletmeden Tüketicie***

İşletmeden tüketiciye (Business-to-Consumer, B2C) olarak bilinen e-ticaret türü, özellikle son zamanlarda ülkemizde ve dünyada en çok bilinen çevrimiçi alışveriş yöntemlerinden biri haline gelmiştir. Örneğin, bir giyim firmasının ürünlerini tüketicilere satması veya bir kitap mağazasının direkt olarak müşterilerine kitap satması işletmeden tüketiciye e-ticaret yöntemine örnek olarak verilebilir. E-ticaret denildiğinde, çoğu kişinin aklına ilk gelen e-ticaret yöntemi, işletmeden tüketiciye (B2C) e-ticaret yöntemidir.

### ***1.6.3. Tüketiciden Tüketicie***

Tüketiciden tüketiciye (Consumer-to-consumer, C2C) e-ticaret türü, son zamanlarda giderek yaygınlaşmaktadır. Bu ticaret türünde, tüketiciler birbirleri arasında alışverişte bulunurlar. Örneğin, kişilerin kullanmadığı ikinci el ürünleri bir başka kişiye satması bu türe örnek gösterebilir. Günümüzde bu yöntem daha sıkça görülmeye başlamıştır.

#### ***1.6.4. Tüketiciden İşletmeye***

Tüketiciden işletmeye (Consumer-to-business, C2B) ürün ve hizmet satışının yapılabilmesi de mümkündür. Örneğin bağımsız bir yazılımcı bir şirketin yazılım ürünlerini geliştirip bu yazılımı talep eden bir şirkete satabilir.

#### ***1.6.5. İşletmeden Kamu Kurumuna***

İşletmeden kamu kurumuna (Business-to-government, B2G) e-ticaret türünde ise ofis malzemeleri satan bir firmanın kamu kurum ve kuruluşlarına satış yapması veya bir inşaat firmasının bir belediye binasını yapma işini üstlenmesi örnek olarak gösterilebilir.

#### ***1.6.6. İşletmeden Çalışanlarına***

İşletmeden çalışanlarına (Business-to-employee, B2E) e-ticaret türünde bir işletmenin çalışanlarına indirim ve kampanyalar sayesinde gerçekleşen bir alışveriş veya şirketin çalışanlarına özel alışveriş faaliyetleri etkinlikler örnek olarak gösterilebilir.

#### ***1.6.7. Kamu Kurumundan İşletmeye***

Kamu kurumundan işletmeye (Government-to-business, G2B) e-ticaret türünde, kamu kurum ve kuruluşlarının işletmelere yönelik hazırladığı çalışan bulma, vergi ödeme, patent alma, marka tescili gibi hizmetler sağladığı platformlar örnek olarak gösterilebilir.

#### ***1.6.8. Kamu Kurumundan Tüketicie***

Kamu kurumundan tüketiciye (Government-to-consumer, G2C) e-ticaret türünde, kamu kurum ve kuruluşlarının vatandaşlarına yönelik hazırladığı pasaport, ehliyet harçları gibi ödemeler, vergi ödeme, merkezi sınav ödemeleri örnek olarak gösterilebilir.



### ***1.6.9. Kamu Kurumundan Çalışanlarına***

Kamu kurumundan çalışanlarına (Government-to-employee, G2E) e-ticaret türünde, kamu kurum ve kuruluşlarının vatandaşlarına (G2C) yönelik sunduğu hizmetler ile işletmelerin çalışanlarına yönelik (B2E) hizmetleri benzerlik göstermektedir. Kamu kurum ve kuruluşlarının çalışanlarına yönelik eğitim hizmetleri, vergi ödemeleri, çeşitli indirimler ve kampanyalar sağlaması örnek olarak gösterilebilir.

### ***1.6.10. Kamu Kurumundan Kamu Kurumuna***

Kamu kurumundan kamu kurumuna (Government-to-government, G2G) olarak bilinen e-ticaret yöntemi, kamu kurum ve kuruluşları arasındaki mal ve hizmet alışverişlerinden oluşur. Bir kamu kurumunun bir belediye ile proje yapması kamu kurumları arası ticarete örnek olarak gösterilebilir.

### ***1.6.11. Çiftçiden Tüketicie***

Çiftçiden tüketiciye (Farmer to Consumer, F2C) e-ticaret türüne literatürde çok sık rastlanmasa da günümüzde çok sık kullanılan e-ticaret türlerinden biridir. Çiftçilerin organik ürünlerini direkt olarak müşteriye satması, çiftçiden tüketiciye (F2C) e-ticaret türüne örnek verilebilir.

### ***1.6.12. İşletmeden İşletmeye ve Tüketicie***

İşletmeden işletmeye ve tüketiciye (Business-to-business-to-consumer, B2B2C) türünde, işletmeler diğer işletmeler aracılığıyla ürün ve hizmetlerini tüketicilere ulaştırmaya çalışırlar. Giyim şirketlerinin perakende e-ticaret şirketleri üzerinden ürünlerini tüketicilere satması örnek olarak gösterilebilir.

### 1.6.13. Direkt Tüketicie

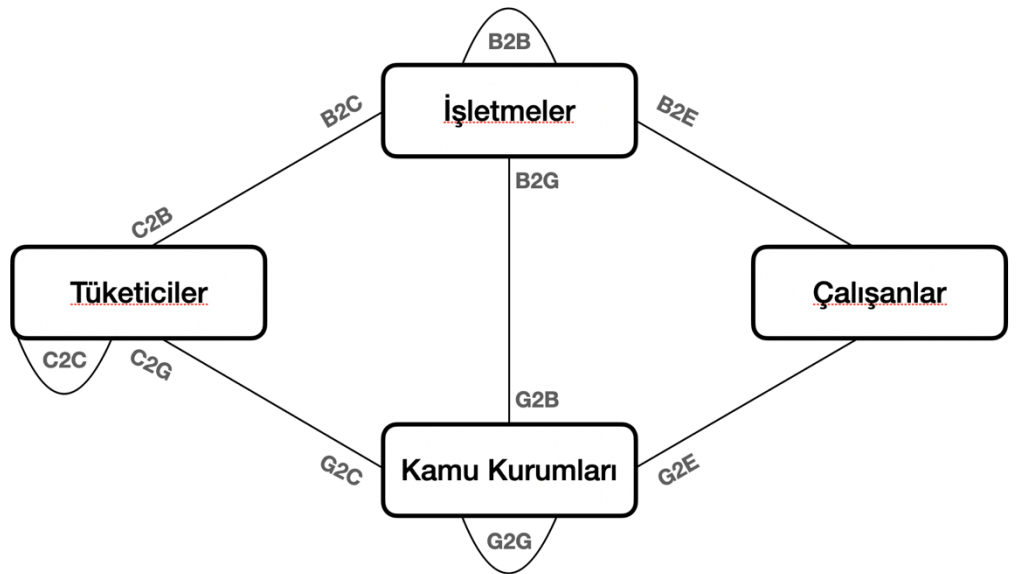
Direkt tüketiciye (Direct-to-consumer, D2C) e-ticaret türünde, satıcıların ürün ve hizmetlerini direkt olarak müşterilerine satmaktadır. Bir tasarımcının tasarımlarını müşterilerine satması veya bir zanaatkarın ürünlerini müşterilerine direkt olarak satması örnek olarak gösterilebilir.

### 1.6.14. Makineden Makineye

Makineden makineye (Machine-to-machine, M2M) e-ticaret türünde, alışveriş makineler arasında yapılır. Yapay zeka yazılımı tarafından yönetilen stok yönetim sisteminde, yapay zeka yaptığı talep tahmini sonucu ihtiyacı olduğu kadar ürün siparişi vermesi veya müşterinin aylık kullanımına göre akıllı sistemler arasında yapılan alışverişler örnek olarak gösterilebilir.

Buraya kadar bilgi verilen e-ticaret türlerinin yapısı ve birbirleriyle ilişkileri Şekil 1’de görüldüğü gibidir. Söz konusu şekil incelendiğinde, e-ticaretin kapsamında i) İşletmeler, ii) Tüketiciler, iii) Çalışanlar ve iv) Kamu kurumları olmak üzere, dört ana taraf bulunmaktadır.

Şekil 1: E-ticaret türleri



## 1.7. E-TİCARETTE KULLANILAN ARAÇLAR

E-ticaretin gelişimi teknolojinin gelişimiyle paralel olarak ilerlemiştir. Teknoloji geliştikçe e-ticarette kullanılan araçlar da değişmiştir. Günümüzde kullanımı neredeyse kalmamış olsa da geçmişte faks, telefon, e-posta ve EDI yöntemleri ile e-ticaret yapıldığı bilinmektedir (Erkan vd., 2022). Günümüzde bu yöntemlerle direkt olarak e-ticaret işlemi gerçekleşmese de pazarlama araçları olarak kullanılmaya devam edilmektedir.

Günümüzde e-ticaret platformları altyapısında birçok farklı aracın ve teknolojinin bir arada çalışması gerekmektedir. Bu teknolojilerin her birinin uyum içerisinde çalışması ve anlık olarak birbirleri ile haberleşmesi bilgi işlem ekipleri tarafından sağlanmaktadır. Günümüzde e-ticaret firmaları kendilerini teknoloji şirketi olarak konumlandırmaktadırlar.

### 1.7.1. E-ticaret Platformları

E-ticaret platformları genellikle web sitesi ve mobil uygulamalardan oluşmaktadır. Modern bir e-ticaret platformu tüm cihazlarda senkronize bir şekilde çalışmaktadır. Müşteriler web sitelerini inceleyip sepetlerine ekledikleri ürünleri cep telefonlarına geçtiğinde, ürün veya hizmetleri kaldıkları yerden devam ederek satın alabilmektedir. Kullanılacak e-ticaret platformu şirketin bilgi işlem ekipleri tarafından geliştirilebilir veya üçüncü taraf şirketlerin geliştirdiği hazır platformlar kullanılabilir. İki seçeneğin de avantajları ve dezavantajları bulunmaktadır. Firmaların ihtiyaçlarına ve bütçelerine göre kullanılan platform değişebilmektedir.

Küçük ölçekli işletmeler genellikle kısıtlı bütçelerinden dolayı üçüncü parti yazılımları tercih etmektedirler. Üçüncü taraf şirketlerin geliştirdiği e-ticaret platformu yazılımlarını firmalar kendi ihtiyaçlarına göre özelleştirerek, kısa bir zaman içerisinde kullanmaya başlayabilir. Aynı zamanda bu teknolojiler yaygın kullanım için oluşturulduğu için genellikle çok fazla teknik bilgi gerektirmez. Buna karşın bu teknolojiler firmaların her ihtiyacını birebir karşılama noktasında yetersiz kalabilir. Genel kullanım için geliştirilen bir yazılım her bir firmanın kendisine özel ihtiyaçlarını karşılayamayabilir.

Büyük ölçekli işletmeler büyük bilgi işlem ekiplerine sahiptir. Tüm süreçleri kontrol edebilmek ve ihtiyaçlarına özel geliştirmeler yapabilmek adına platformlarını kendileri geliştirirler. Bu yöntem fazlasıyla maliyetli ve zaman alıcı bir süreç olmasına karşın, fakat sürecin tamamında kontrol sahibi olabilmelerini sağlar.

### ***1.7.2. Elektronik Ödeme Sistemleri***

E-ticaret işlemlerinde müşterilere havale ve kapıda ödeme seçenekleri sunulmaktadır. Ancak, müşterilerin genellikle online ödeme yöntemlerini tercih ettikleri de bilinmektedir. Günümüzde mobil ödeme, kripto paralar ile ödeme, sanal kartlar ile ödeme gibi yöntemler de gün geçtikçe yaygınlaşmaktadır (Erkan vd., 2022). E-ticaret platformları ile benzer şekilde elektronik ödeme sistemlerinde de küçük ölçekli işletmeler, üçüncü taraf şirketler tarafından geliştirilen hizmetleri tercih ederken, büyük ölçekli şirketler ise kendi ödeme sistemlerini geliştirmeyi tercih etmektedir. E-ticaret sektöründe önde gelen şirketler kendi ödeme sistemlerini, sanal cüzdanlarını geliştirmekte ve bu sistemleri kullananlara bazı ödüller vermektedirler. Bu sayede ödeme işlemini kendi sistemleri üzerinden gerçekleştirmekte ve müşteri sadakatini arttırmaya çalışmaktadırlar.

### ***1.7.3. Veri Yönetimi ve İstatistiksel Veri Analizi Araçları***

Müşterilere daha iyi hizmet ve ürünler sunabilmek için öncelikle müşteriler hakkında bilgi sahibi olmak gerekmektedir. Müşteriler hakkında bilgi edinmenin ilk aşaması veri toplamaktan geçmektedir. Günümüzde kullanıcı verilerinin toplanması ve bu verilerin istatistiksel araçlarla analiz edilmesi belirli yasalar çerçevesinde korunmaktadır. Kişisel verilerin usul ve esaslara uygun şekilde toplanması ve istatistiksel veri analizinin yapılması gerekmektedir. Müşterilerin e-ticaret platformu ile etkileşimlerine dair veriler uygun şekilde toplanmalıdır.

Müşterilerden toplanan veriler hem şirketin geleceğini belirleyecek kararların alınmasında yardımcı olabilir hem de müşterilere daha yararlı bir alışveriş deneyimi

sunmak için kullanılabilir. Örneğin, işletme en çok satın alınan ürünlerin trendini takip ederek gelecek için bir talep tahmininde bulunabilir veya ürün kataloğunu benzer ürünlerle genişletmeye çalışabilir. Aynı zamanda benzer müşterilerin incelediği ve satın aldığı ürünlerin ortak özelliklerinden yola çıkarak, müşterilere ilgilerini çekebilecek benzer ürünler sunularak satışlar artırılabilir. Müşterilerden toplanacak verilere aşağıdaki örnekler verilebilir:

- Kullanıcıların görüntülediği sayfalar,
- Görüntülenen sayfaları oluşturan öğelerin ekranda görünüp görünmediği,
- Ekranda gösterilen öğelerin tıklanıp tıklanmadığı,
- Kullanıcıların sayfalarda geçirdiği süreler,
- Kullanıcıların platformu ne sıklıkla ziyaret ettiği,
- Çevrimiçi platformda kullanıcıların sayfalar arasında nasıl bir akış izlediği,
- Kullanıcıların sahip oldukları cihazların marka, model, işletim sistemi ve sürümü, ekran boyutu, kullandığı servis sağlayıcısı vb. gibi bilgiler ve kaydedilen veriler müşterilerden toplanacak verilerden bazılarıdır.

Müşterilerin ihtiyaçlarının anlaşılabilmesi ve beklentilerinin karşılanabilmesi için günümüzde şirketler veri odaklı yaklaşımları benimsemişlerdir. Gün geçtikçe müşteriler hakkında daha detaylı veriler toplanmaya ve daha gelişmiş analizler yapılmaya başlanmıştır. Bu durum büyük verinin (big data) depolanması ve analiz edilmesi maliyetlerini beraberinde getirmiştir.

#### ***1.7.4. Pazarlama ve İletişim Araçları***

Pazarlama kavramının farklı şekillerde tanımını yapmak mümkündür. Pazarlama piyasanın ihtiyacı olan bir ürünü karlı bir şekilde karşılamak olarak tanımlanabilir (Kotler & Keller, 2012). Pazarlamanın tanıtımdan farklı olarak insanlara ulaşıp satış yapmak anlamına gelmediği, satış öncesinden satış sonrasına devam eden bir süreç olduğu ifade edilmektedir (Güleryüz, 2019).

E-ticaret şirketleri müşterilerine ürünlerinin tanıtımını yapabilmek ve satışını arttırabilmek için çeşitli pazarlama ve iletişim araçları kullanabilirler. Sosyal medya, e-posta, kısa mesaj, anlık bildirimler gibi en bilinen araçların yanı sıra içerik pazarlama, arama motoru optimizasyonu, nüfuz (influencer) pazarlama ve satış ortaklığı (affiliate) pazarlaması yöntemleri de kullanılabilir (Kavaklı, 2018). Tanıtım ve satış yapmaya ek olarak satış sonrası verilen destek hizmetleri de iletişim araçlarıyla sürdürülmektedir.

Pazarlama için kullanılacak araç ve yöntemler müşteri profillerine göre değişebilmektedir. Şirketler, pazarlama ve iletişim yöntemlerini en karlı olacak şekilde optimize edip marka farkındalığını, müşteri sadakatini ve buna bağlı olarak satışlarını arttırmaya yönelik çalışmaktadırlar.

## **1.8. E-TİCARET VE MÜŞTERİ İLİŞKİLERİ YÖNETİMİ (CRM)**

Müşteri ilişkileri yönetimi (Customer Relationship Management, CRM) ile firmalar; yeni müşteriler kazanmayı, var olan müşterilerin sadakatini ve karlılığını arttırmayı ve kaybetme riski olan müşterileri geri kazanmayı hedeflemektedir. Diğer bir ifadeyle günümüzde işletmeler müşterilerini daha yakından tanıyabilmek, ihtiyaçlarını tahmin edebilmek ve bu ihtiyaçlara yönelik özel ürün ve hizmetler sunabilmek adına müşteri ilişkileri yönetimi tekniklerine başvurmaktadırlar.

Günümüzde çoğu işletmenin müşteri ilişkileri yönetiminin değerinin farkında olduğu görülmektedir. E-ticaret şirketlerinin müşterilerinin memnuniyeti gelirlerini doğrudan etkilediğinden, e-ticaret şirketleri için de müşteri ilişkileri yönetimi son derece önem taşımaktadır. Günümüzde önde gelen tüm e-ticaret şirketleri müşteri odaklı bir yaklaşım benimsemişlerdir.

### ***1.8.1. Müşteri İlişkileri Yönetiminin (CRM) Tanımı***

İşletmelerin sunulan ürün ve hizmetlerini bir ücret ödeyerek alan kişilere veya kurumlara müşteri denilmektedir. Günümüzde işletmeler ürün odaklı yerine müşteri odaklı bir

yaklaşım izlemektedir. Diğer bir ifadeyle, işletmelerin her şeyden önce müşteri memnuniyetini önemseydiği görülmektedir.

Günümüzde sadece bir müşterinin olumsuz bir yorumu veya şikâyeti dahi binlerce kişiye ulaşabilmektedir. Müşterilerin olumsuz bir deneyim yaşamaması veya olumsuz deneyime neden olabilecek anlaşmazlıkların ve hataların işletmeler tarafından çözülmesi gerekmektedir. İşletmeler müşteri odaklı yaklaşımlar sayesinde mutlu müşterilerin de benzer şekilde daha fazla müşteri getireceğini ve bu yaklaşımlar sayesinde uzun dönemde daha fazla kar elde edebileceklerini fark etmişlerdir. Aynı zamanda günümüzün rekabet koşullarında müşteri sadakati, işletmeler için çok önemli bir konudur. Örneğin, çoğu e-ticaret müşterisi daha önce yapmış olduğu alışverişlerde yaşadığı olumlu ya da olumsuz deneyimlerden dolayı, belirli bir markaya daha sadık hale gelebilmektedir. Müşteriler herhangi bir ürün (veya hizmet) arayacağı zaman ilk olarak kendisini daha yakın hissettiği markaya yönelme eğilimindedirler. Tüm bu nedenler, işletmeleri müşterileri hakkında daha fazla bilgi sahibi olmaya itmiş ve müşterilerinin ihtiyaçlarına yönelik çözümler sunmaya çalışmaya yöneltmiştir.

Müşteri ilişkileri yönetimi (CRM); insanların, teknolojinin ve proseslerin kombinasyonu ile mevcut müşteriyi tutma ve müşteri ile var olan ilişkiyi geliştirmeye odaklanmış entegre bir yaklaşımdır (Chen & Popovich, 2003). İşletmeler için mevcut müşterilerin müşteri memnuniyetini arttırmak ve işletmeden ayrılmamasını sağlamak, yeni müşteri elde etmeye çalışmaktan daha az maliyetlidir (Hızlı, 2019).

### ***1.8.2. Müşteri İlişkileri Yönetiminin Tarihsel Gelişimi***

Müşteri ilişkileri yönetim sistemi (CRM) kavramının ilk olarak 1990'lı yıllarda ortaya çıktığı belirtilmektedir (Söztutar, 2010). Çok disiplinli bir yaklaşım olduğu için teknolojinin gelişmesiyle birlikte müşteri ilişkileri yönetim sistemi tekniklerinin de zaman içerisinde geliştiği görülmektedir. Müşterilerle etkileşimlerin kaydedilmesi, veri ambarlarının teknolojik olarak gelişmesi ve maliyetlerinin azalması, internet üzerinden pazarlama metotlarının artması ve daha etkin hale gelmesi, istatistiksel analiz yöntemlerinin ve veri madenciliği tekniklerinin bu alanda daha sık kullanılmaya

başlanması ve makine öğrenmesi gibi teknolojilerinin gelişmesi, müşteri ilişkileri yönetim sisteminin günümüzde ulaştığı seviyeye gelmesine katkı sağlamıştır.

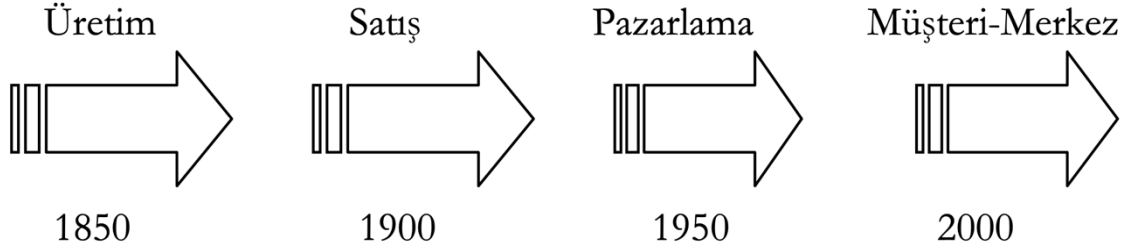
Geleneksel ticaret yöntemlerinde satıcılar genellikle müşterilerini tanımakta, her bir müşterisinin ihtiyaçları ile birebir ilgilenebilmekte ve müşterinin isteklerine yönelik özel ürün ve hizmetler sunabilmekteydi. Bu dönemlerde şirketler arası rekabet üretim kapasitesi üzerinden ilerlemekteydi. Daha fazla ürün ve hizmet sunabilen şirketler daha avantajlıydı. Sonraki dönemlerde üretim maliyetleri çok ucuzladı ve üretim kapasitesi çok arttı. Bu sayede işletmeler arası rekabet üretim kapasitesinden ziyade müşteri ilişkilerine yöneldi. İşletmelerin çok daha büyük ölçekte müşterilere ulaşması, çok yüksek sayıda ürün sunabilmesi, kitlesel pazarlama yöntemleri gibi nedenlerden dolayı işletmeler, geçmişte müşterileri ile olan özel iletişimi kaybedebilmektedir. Eski ticaret yöntemlerinde her biri hakkında birçok bilgi sahibi olunan ve özel ilgilenilen müşteriler, günümüzde veri tabanında birer satır haline gelmiştir (Chen & Popovich, 2003).

İşletmelerin müşteri ilişkileri yönetimi yaklaşımlarını benimsemesi ve bu yaklaşımları süreçlerine entegre edip doğru şekilde uygulayabilmesi zaman almıştır. Ancak işletmeler rekabetçi piyasada iş yapmayı sürdürebilmek için müşteri odaklı yaklaşımları benimsemek durumundadırlar. İşletmeler uzun dönemde başarılı olmak için potansiyel müşterileri nasıl kazanabileceklerini, var olan müşterilerin isteklerini ve kaybedilen müşterilerin nasıl geri getirilebileceğini anlamak zorundadırlar. Bu şekilde müşterilerin isteklerini daha iyi anlayıp daha doğru aksiyonlar alabilirler.

Müşteri ilişkileri yönetiminin tarihsel gelişimi dışında, işletme yönetiminin tarihsel gelişimi Şekil 1.2'deki gibidir. Şekil incelendiğinde, tarihsel gelişimin üretim, satış, pazarlama ve müşteri merkezlilik şeklinde olduğu görülecektir. Gelinek noktada, 2000'li yıllar ile birlikte müşterinin merkezde veya odakta olduğu söylenebilir. Bunun arkasında da küreselleşmeye bağlı olarak yaşanan acımasız rekabet koşullarının olduğu söylenebilir.



## Şekil 2: İşletme Yönetiminin Tarihsel Gelişimi



Kaynak: (Uysal & Aksoy, 2004)

### 1.8.3. Müşteri İlişkileri Yönetiminin Avantajları

Müşteri ilişkileri yönetim sistemi yaklaşımı işletmelerin müşterileri ile arasındaki ilişkiyi daha iyi anlamalarına ve bu ilişkiyi analiz ederek optimize edebilmelerini sağlamaktadır. Ayrıca müşteri ilişkileri yönetim sistemi, işleri otomatikleştirme yöntemiyle otomasyon imkânı sağlar ve bu doğrultuda işletmelerin verimliliğini arttırmakta ve maliyetlerini azaltmaktadır. Bu sayede, müşteri ilişkileri yönetimi sistemini kullanan işletmeler, rakiplerine kıyasla çeşitli avantajlara sahip olurlar. Aşağıda müşteri ilişkileri yönetimi sistemi kullanan şirketlerin elde ettiği avantajlara bazı örnekler verilmiştir:

- Müşteri bilgilerinin düzenli şekilde toplanması ve saklanması,
- Müşteri bilgilerinin ilgili kişilerin ortak erişimine açık olması,
- Müşteri profilleri oluşturabilme,
- Müşteri alışkanlıklarını takip edebilme,
- Daha iyi tanınan müşteriler için daha etkili pazarlama ve satış stratejileri geliştirebilme,
- Müşteri ihtiyaçlarının daha iyi anlaşılabilmesi sonucunda müşteri memnuniyetinin artırılması ve dolayısıyla müşteri sadakatinin (bağlılığının aidiyetinin) artması,
- Müşteri ilişkileri yönetimi sistemlerinin otomatikleşmesi sayesinde, verimliliğin artması ve maliyetlerin düşmesi,

- Ölçülebilir, yönetilebilir, analiz edilebilir ve iyileştirilebilir süreçler geliştirebilme avantajları sağlar.

Özetle, müşteri ilişkileri yönetim sistemi kullanan işletmeler, müşterileri hakkında daha fazla bilgi sahibi olmaktadır. Müşterilerini daha yakından tanıyarak, müşterilerinin ihtiyaçlarına ve isteklerine özel ürün ve hizmetler sunabilmektedir. İşletmeler, mevcut ürün ve hizmetlerin yetersiz kaldığı durumlarda, şirketler portföylerini geliştirme ihtiyacı olduğunu fark ederek ve müşteri ilişkileri yönetimi sistemi ile anlamış oldukları ihtiyaçlara yönelik yeni çözümler geliştirebilmektedir. Potansiyel müşterilerin isteklerinin algılanması ve bu yönde pazarlama çalışmaları yapılması, mevcut müşterilerin ihtiyaçlarına yönelik ürünler önerilmesi ve kampanyalar oluşturulması, kaybedilmek üzere olunan ve kaybedilen müşterilerin kaybedilme nedenlerinin belirlenerek önlemler alınması ve hayata geçirilmesi mümkün hale gelmektedir. Kısacası, müşteri ilişkileri yönetimini kullanan işletmeler, müşteri ilişkilerini daha etkin bir şekilde yönetilebilir, müşteri sadakatini artırılabilir, verimliliği arttırıp maliyetleri düşürebilir, daha rekabetçi hale gelebilir ve sonuçta uzun dönemde daha fazla kar edebilirler.

#### ***1.8.4. Müşteri İlişkileri Yönetiminin İlkeleri***

İşletmelerin müşteri ilişkileri yönetimi sistemini başarılı bir şekilde hayata geçirebilmesi ve sürekliliğini sağlayabilmesi için CRM'in ilkelerine uyması gerekmektedir. İşletmelerin uyması gereken bu ilkeler aşağıdaki gibi sıralanabilir (Yereli, 2001).

- İşletmenin müşteriyi teşvik eden bir rolü üstlenmesi,
- İletişim kanalları başarılı şekilde kullanabilmesi,
- İşletmenin yaşanan olumsuzluklarda sorumlulukları üstlenmesi,
- Müşteri memnuniyetinin ana hedef olarak benimsenmesi,
- Müşteri memnuniyeti için girişimci yaklaşımlarda bulunulması,
- Müşteri memnuniyetine giden sonuca odaklı çalışılması.

Özetle, işletmelerin müşteri memnuniyetini kazanmak için gerekeni yapması, ikili ilişkilerde sorumluluğu üstlenmesi ve sadece kar odaklı değil müşteri odaklı bir yaklaşım izlemesi gerekmektedir. Müşteri ile gerçekleşen etkileşimler sonucunda müşterinin memnun kalması işletmelerin ana odağı olmalıdır. Müşteri karşılıklı ilişkiden memnun değilse, bu durum işletmenin müşteri ilişkileri yönetimini başarıyla uygulayamadığını göstermektedir.

#### ***1.8.5. Müşteri İlişkileri Yönetiminin Kurulması ve Uygulanması***

Müşteri ilişkileri yönetim sisteminin organizasyondaki herkes tarafından benimsenmesi gerekmektedir. Bir işletmede müşteri ilişkileri yönetim sisteminin uygulanabilmesi için yatırımın öncelikle teknoloji yerine insan gücüne yapılması gerektiği vurgulanmaktadır (Yereli, 2001).

Doğru bir müşteri ilişkileri yönetimi sistemi kurmak için önce müşteriyi tanımak gerekmektedir. Müşteriyi doğru tanıyabilmek için öncelikle işletmelerin müşterilerin rızası doğrultusunda mümkün olduğunca her hareketini izlemesi ve kaydetmesi gerekmektedir. Bir organizasyonun veya işletmenin müşteri hakkında bilgi edinebilme becerisi, çeşitli kaynaklardan elde ettiği bilgiyi dönüştürme ve birleştirme kabiliyetine bağlıdır (Khodakarami & Chan, 2014).

Bir e-ticaret platformu için müşterilerin e-ticaret platformunu ne sıklıkla ziyaret ettikleri ve ne sıklıkla alışveriş yaptığı önemli bir bilgidir. Müşterilerin ne tarz ürünleri görüntülediği, varsa düzenli sipariş verdiği ürünler, ortalama sipariş tutarları gibi alışveriş bilgileri de kaydedilmelidir. Müşterinin her bir ziyaretinde gösterilen yüzlerce ürün arasından hangi ürünlere tıklamayı tercih ettiği, hangi kategorilerin daha çok ilgisini çektiği vb. gibi bilgiler kaydedilmelidir. Özetle müşterilerin e-ticaret platformu ile olan ilişkilerindeki her bir etkileşimi doğru bir şekilde kaydetmek ve analiz etmek müşteriyi daha yakından tanımak için gereklidir.

### ***1.8.6. Müşteri İlişkileri Yönetiminin Süreçleri***

Müşteri ilişkileri yönetimi bir süreçtir ve müşteriyi elde etmekten başlayarak müşterinin elde tutulmasına kadar uzanan alt süreçleri kapsamaktadır. CRM'in süreçleri, i) Veri Yönetimi Aşaması (Bilgi Toplama), ii) Satış Yönetimi Aşaması (Müşteri Edinme), iii) Müşteri Hizmetleri Yönetimi (Müşterinin Sadakatini Kazanma), iv) Müşteri Değerini Artırma olmak üzere dört aşamadan oluşur.

#### ***1.8.6.1. Veri Yönetimi Aşaması (Bilgi Toplama)***

Müşteri ilişkileri yönetimi sisteminin ilk aşaması olan bilgi toplama aşamasında hedef, kitle yani aynı anlama gelmek üzere, müşteriler hakkında bilgi toplanmasıdır. Müşteriler hakkında bilgi sahibi olmak, müşteriyi daha iyi tanıyabilmek ve gruplandırabilmek için gereklidir (Demir & Kırdar, 2007). Öncelikli olarak, müşteriler hakkında düzenli veri toplanması ve bu verilerin güncel tutulması amaçlanmaktadır. Kuşkusuz, farklı müşteri gruplarının farklı ihtiyaçları ve istekleri olabilmektedir. Müşterileri harcamalarına göre ve ilgi alanlarına göre segmentlere ayırmak, her grup için farklı pazarlama kampanyaları ve ürün önerileri yapabileme imkânı sağlamaktadır.

#### ***1.8.6.2. Satış Yönetimi Aşaması (Müşteri Edinme)***

Müşteri ilişkileri yönetimi sisteminin ikinci aşaması oluşturulan müşteri segmentlerine yönelik satış stratejilerini içermektedir. Belirlenen müşteri gruplarını daha yakından tanıyarak müşterilerin ihtiyaçları ve talepleri daha doğru bir şekilde tahmin edilebilmektedir. Bu sayede hedefe yönelik pazarlama stratejileri hazırlanmaktadır. Müşteri gruplarının ihtiyaçlarına göre özelleştirilmiş ürün ve hizmet önerileri ile satış gerçekleştirilebilmektedir.

Satış yönetimi aşamasında asıl amacın müşterilere tek seferlik bir ürün satmak yerine, uzun dönemde sadık müşteriler kazanmak olduğu ifade edilir. Müşteri edinme aşaması müşteri merkezli düşünmeyi gerektirir. İşletmeler bu sayede üretim aşamasından satış sonrası sürece kadar iyileştirmeler yapabilmektedir (Gemci, 2019).

#### *1.8.6.3. Müşteri Hizmetleri Yönetimi (Müşterinin Sadakatini Kazanma)*

Müşteri ilişkileri yönetiminde mevcut müşteriyi korumanın maliyetinin yeni müşteri edinmekten daha az olduğu vurgulanmaktadır. Bu aşamada müşterinin sadakatini (aidiyetini, bağlılığını) kazanmak için müşterilerin ihtiyaçlarının karşılanarak memnuniyetlerinin artırılması hedeflenmektedir.

Müşteri ilişkileri yönetiminde işletmeler, süreçlerini müşterilerinin ihtiyaçlarına yönelik şekillendirmektedir. Örneğin; müşteriler ile yakın ilişkiler kurmak, müşterilere kendilerini özel hissettirmek, rakip işletmelere kıyasla daha uygun fiyatlarla daha kaliteli ürün ve hizmetler sunmaya çalışmak ve müşteri hizmetlerinin ortaya çıkan problemleri müşteri odaklı çözmek vb. gibi yaklaşımlar sayesinde müşterinin sadakati sağlanmaya çalışılmaktadır.

#### *1.8.6.4. Müşteri Değerini Arttırma*

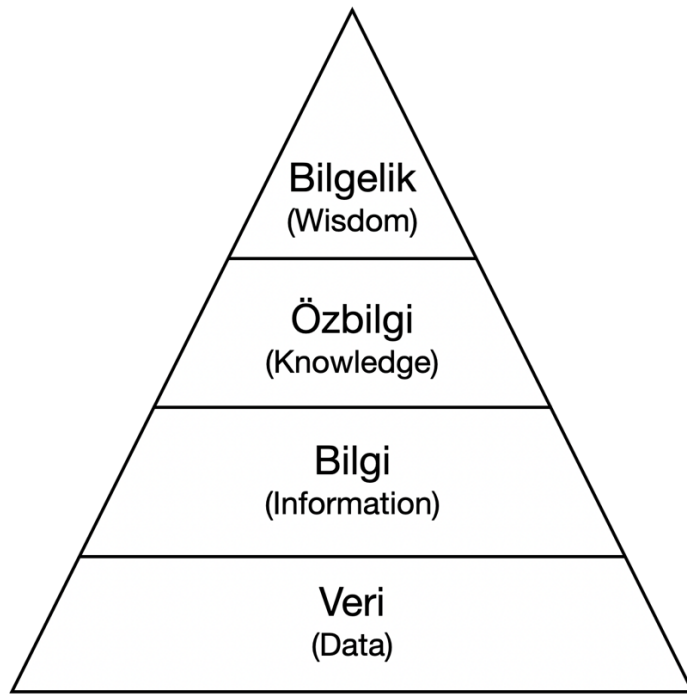
Müşteri ilişkileri yönetiminin dördüncü aşaması, müşteri yaşam boyu değerinin artırılmaya çalışılmasıdır. Müşterinin işletme ile olan ilişkisi süresince, işletmeye getireceği karın bugünkü değeri, müşteri yaşam boyu değeri olarak ifade edilmektedir. (Yapraklı & Keser, 2008). Bu aşamada, müşteriler ile olan uzun vadeli ilişkiyi faydalı bir şekilde devam ettirmek ve bu ilişkiden daha fazla kar elde etmek amaçlanmaktadır. Müşteri yaşam boyu değerini artırabilmek için müşterilerden mümkün olduğunca daha fazla kar elde etmek amaçlanmaktadır. Daha önce belirtildiği üzere, bu sırada asıl amaç yalnızca kar elde etmek değildir, müşterinin de memnun kalacağı ve ihtiyaçlarını karşılayacak yararlı öneriler sunularak, uzun dönemli ilişkiler gözetilmeye devam edilmektedir. Özetle, müşteri ilişkilerine önem vermek ve müşterileri bir değer olarak görmek işletmelerin ana odağı haline gelmektedir (Yurdakul, 2015).

#### *1.8.7. Müşteri İlişkileri Yönetiminde Kullanılan Teknolojiler*

Bir işletmenin müşteri ilişkileri yönetimini başarılı bir şekilde uygulayabilmesinin temelinde müşteriden bilgi edinmesi vardır (Peppers vd., 1999). Günümüzdeki teknolojik

imkanlar sayesinde müşterilerin kendileri hakkında ve işletme ile olan etkileşimlerinin detaylarına dair veriler toplanabilmektedir. Ancak veri tek başına yeterli değildir. İşlenmemiş veri genellikle bir anlam ifade etmez. Veriden bir yarar elde edebilmek için işlenmemiş ham veriden bir anlam çıkarılması gerekmektedir. Şekil 3'te yer alan bilgi hiyerarşisi veya bilgi piramidi olarak anılan DIKW (Data/Veri, Information/Bilgi, Knowledge/Özbilgi, Wisdom/Bilgelik) modeli, veriden bilgeliğe giden yolculuğu göstermektedir.

**Şekil 3: DIKW Modeli**



Müşterilerden elde edilen verinin anlamlandırılması ve faydalı içgörüler elde edilmesi, veri madenciliği yöntemleri ile mümkündür. Toplanan verinin analiz edilmesi ve anlamlandırılması, müşteri ilişkileri yönetimi stratejilerinin belirlenmesini mümkün kılar. İşletmeler, müşteri ilişkileri yönetimi sistemini belirli teknolojiler sayesinde uygulayabilmektedir.

Veri yönetimi teknolojileri sayesinde müşterilerden detaylı veriler toplanabilmektedir. Örneğin, müşteriler hakkında toplanan veriler arasında kullanıcıların görüntülediği

sayfalar ve görüntülenen sayfalarda geçirilen süreler gibi veriler bulunmaktadır. Toplanan müşteri etkileşimi verileri anlık veya daha sonra analiz edilmek üzere, veri ambarlarında saklanmaktadır. Anlık analizlere örnek olarak müşterinin incelediği ürünler arasındaki ilişkilerden veya benzer müşterilerin satın alma davranışlarından yola çıkarak makine öğrenmesi öneri algoritmaları sayesinde müşterinin beğenebileceğini düşündüğü ürünler öneri olarak sunulmaktadır. Ayrıca, bu verilerin daha sonra analiz edilerek iş zekası araçları ile haftalık ve aylık özet raporlar olarak gösterilmesi de mümkündür.

Verinin anlamlandırılması sayesinde şekillenen stratejilerin uygulanmasının önemli bir aşaması müşterilerle iletişimi sürdürmektir. Müşterilerle çok çeşitli teknolojik araçlar aracılığıyla iletişime geçilebilmektedir. Kısa mesaj, e-posta ve anlık bildirimler günümüzde en çok tercih edilen iletişim yöntemleri olarak karşımıza çıkmaktadır. İndirimlerin ve kampanyaların duyurulmasında ise genellikle internet ve televizyon reklamları kullanılmaktadır. Farklı müşteri gruplarına farklı iletişim yöntemleri ile hitap edilebileceği için tek bir iletişim yöntemi kullanmanın yetersiz kalacağı görülmektedir. Bu yüzden, her kitleye hitap edebilmek için çeşitli iletişim yöntemlerine başvurulmaktadır.

Gelişmiş bir e-ticaret platformunda çok sayıda müşteri ile iletişimi kolaylaştırabilmek için otomasyon araçları kullanılmaktadır. Bir müşteri ilişkileri yönetimi otomasyon aracının örnek bir çalışma akışı şu şekilde olabilir: Müşterilerin, işine yarayacak bir ürün veya hizmeti form doldurarak elde etmesi, elde edilen bilgilere göre müşterilerin gruplandırılması, form üzerinden iletişime geçilmesini onaylayan müşteriler ile iletişime geçilerek kampanya ve tekliflerin iletilmesi yaygın kullanılan bir otomasyon akışıdır.

Müşteriler bir problem yaşadıklarında karşılaştıkları problemin en hızlı şekilde çözülmesini istemektedirler. Müşteriler genellikle yardım sayfaları veya sohbet robotları yerine müşteri hizmetleri ile görüşmeyi tercih ederler. Günümüzde teknoloji sayesinde çağrı merkezleri müşteriler ile hızlı bir şekilde konuşmaya başlayabileceği ve sorununa çözüm getirebilecek donanımlara sahiptir (Gemci, 2019). Müşteri ilişkileri yönetimi süreçlerinde çağrı merkezleri önemli bir yere sahiptir.

Diğer taraftan, müşteri ilişkileri yönetimi süreçlerinin yönetiminde yararlı olabilecek bazı yazılımlar mevcuttur. İşletmeler üçüncü taraf CRM yazılımlarını kullanabilir ya da kendi süreçlerine özel CRM yazılımları geliştirmeyi tercih edebilir. Web ve mobil yazılım teknolojileri kullanılarak geliştirilen CRM yazılımları, müşteri ilişkileri yönetimi süreçlerinin tek bir platform üzerinden daha kolay takip edilebilmesini sağlamaktadır.

## **1.9. E-TİCARETİN EKONOMİYE ETKİLERİ**

E-ticaretin gelişim sürecinde, özellikle paketleme ve dağıtım gibi iş kollarının önemi artmıştır ve bu sayede yeni bir istihdam alanı oluşmuştur. Ayrıca, e-ticaretin ekonomi ve toplum üzerindeki etkisi her geçen yıl daha da artmaktadır. E-ticaret geçmişte hızlı bir yükseliş göstermiştir ve istatistiksel veriler göz önüne alındığında, bu yükselişi gelecekte de sürdüreceği tahmin edilmektedir. 2025 yılında global e-ticaret hacminin 4,4 trilyon dolar değerine ulaşacağı öngörülmektedir (Demirdöğmez, 2021). Bu gelişimin devam etmesi ve bu süreçte tarafların korunması için devletler ve uluslararası bazı organizasyonlar çalışmalar yürütmektedirler. Devletler, teşvik ve düzenlemeler ile e-ticaretin gelişimi için uygun ortamı sağlamaya yönelik çalışmalarda bulunmaktadır. E-ticaretin ülke sınırlarını aşan yapısı nedeniyle devletlerin ötesinde, Dünya Ticaret Örgütü (WTO), Dünya Bankası (WB), Ekonomik Kalkınma ve İşbirliği Örgütü (OECD), Avrupa Birliği (EU) ve daha birçok organizasyon da e-ticaretin gelişimine katkıda bulunacak çalışmalar yürütmektedirler. Bu çalışmalar özellikle e-ticaretin gelişimini desteklemek, tüketicinin haklarını korumak, küçük işletmeleri ve girişimcileri desteklemek gibi konulara odaklanmaktadır (Malawer, 2008).

2021 yılında yapılan bir çalışmaya göre, Türkiye’de e-ticaret 2019 yılında 136 milyar Türk Lirası iken %66’lık bir büyüme göstererek 2020 yılında 226 milyar Türk Lirası değerine ulaşmıştır (Deloitte Danışmanlık A.Ş., 2021, s. 11). Ülkemizde e-ticaret hacmi 2022 yılının ilk altı ayında 2021 yılının ilk altı ayına kıyasla %116 artarak 348 milyar Türk Lirası değerine ulaşmıştır (T.C. Ticaret Bakanlığı, 2022).



Ülkemizde pazar yerlerinde e-ticaret yapan ve ETBİS'e kayıtlı olan 488.706 adet işletme olduğu belirtilmiştir. 2022 yılının ilk 6 ayında e-ticaret harcamalarının 308 milyar Türk Lirası ile %88'ini yurtiçi alımların, 15,8 milyar Türk Lirası ile %5'ini diğer ülkelerin ülkemizdeki e-ticaret platformlarından yaptığı alımların ve 15,8 milyar Türk Lirası ile %7'si ise vatandaşlarımızın yurt dışından yaptığı alımların oluşturduğu görülmektedir (T.C. Ticaret Bakanlığı, 2022).

## **1.10. E-TİCARETİN GELECEĞİ**

Ortaya çıktığı yıllardan itibaren büyümesini aralıksız bir şekilde sürdüren e-ticaret sektörü özellikle son yıllarda COVID-19 pandemisinin de etkisiyle hızlanarak büyümeye devam etmiştir. Araştırmacılar tarafından e-ticaretin en yüksek büyümesinin %26,9 oranıyla pandemi döneminde gerçekleştiği belirtilmektedir (Deloitte Danışmanlık A.Ş., 2021).

Pandemi döneminde edinilen e-ticaret alışkanlıklarının pandemi sonrasında da devam edeceği tahmin edilmektedir. Tüketicilerin değişen alışveriş alışkanlıkları ile birlikte; hızlı teslimat, kolay iade, indirimli ürünler, canlı destek ve girişimciler için e-ticaret işi kurmanın çok daha kolay bir hale gelmesi vb. gibi teknolojik ve lojistik gelişmeler de dikkate alındığında, e-ticaret sektörünün gelecek yıllarda da benzer şekilde büyümeye devam edeceği öngörülmektedir.

Günümüzde cep telefonlarının kullanımının yaygınlaşması, e-ticaret şirketlerinin mobil uygulamalara daha çok önem vermelerine neden olmuştur. E-ticaret mobil uygulamalarının gelecekte ihtiyaçlarımızı karşılama konusunda daha gelişmiş sistemler olacağı öngörülmektedir. Ayrıca, günümüzde e-ticaret firmaları veri odaklı (data-driven) işletmeler haline gelmiştir. İşletmeler alınacak önemli kararları içgüdülerine göre değil verilere dayandırarak vermektedir. E-ticaret işletmelerinin veriye dayalı kararlar vermeleri daha doğru adımlar atmalarını ve daha karlı işletmeler olmalarını sağlayacaktır.

E-ticaretin yıllık büyüme hızı dünya ortalamasında %11,7 iken Türkiye'de %23,7 oranında seyrettiği görülmektedir. 2024 yılına gelindiğinde yıllık ortalama %23,7'lik

büyüme ile Türkiye’de 2020’ye kıyasla e-ticaretin 2,3 katına çıkacağı tahmin edilmektedir (Deloitte Danışmanlık A.Ş., 2021).

## İKİNCİ BÖLÜM

### MAKİNE ÖĞRENMESİ

Makine öğrenmesi terimini bugün bildiğimiz haliyle ilk tanımlayan kişilerden biri olarak kabul edilen Arthur Samuel (1959) makine öğrenmesini açıkça programlanmadan, tecrübelerden yola çıkarak bilgisayarlara öğrenme yeteneği veren bir bilgisayar bilimi alanı olarak tanımlamıştır (Samuel, 1959). Diğer bir ifadeyle makine öğrenmesi yapılacak işin direkt olarak programlanmadan istatistiksel modeller ile verideki kalıpların tespit edilebilmesidir. Bir makine öğrenmesi modelinin çalışması için belirlenmiş algoritmalar girdi olarak bir veri kümesi ile beslenir. Ardından makine, girdi olarak sağlanan veri kümesi üzerinden kendi kendine öğrenerek (self-training) bir tahmin gerçekleştirmektedir (Theobald, 2017).

Makine öğrenmesinin tecrübelerden yola çıkması ifadesinde makine öğrenmesi modelinin eğitilmesi için kullanılan veri kümesinden bahsedilmektedir. Makine öğrenmesi temelinde girdi olarak verilen veriden bir bilgi çıkarmaktadır. Veri seti ne kadar iyi ve büyük olursa makine öğrenmesinin tahminleri de o kadar iyi olacaktır (Mohri vd., 2018). Veri kümesinin doğru ve tutarlı olması, yanlış veri barındırmaması, diğer bir ifadeyle modelin eğitileceği verinin kaliteli olması gerekmektedir. Makine öğrenmesi veriyi anlamlandırabilmesinin temelinde istatistik, yapay zeka, bilgisayar bilimleri, istatistiksel öğrenme ve tahmin yöntemleri yer almaktadır (Bastem, 2021).

Tarih boyunca insanlar zor ve rutin işlerin makineler yardımıyla yapılabilmesi için çaba sarf etmişlerdir. Özellikle endüstri devrimi sonrası makineler çoğu işi yapabilecek hale geldiği halde bilişsel yetenekler gerektiren işlerin insanlar tarafından yapılması gerekmiştir (Theobald, 2017). Makine öğrenmesi ile makinelerin de veriye dayanarak karar verebilmesi ve tahminde bulunabilmesi mümkün olmuştur. Örneğin istenmeyen e-postaların tespit edilebilmesi geleneksel programlama yöntemleriyle yeterince verimli olmamaktadır. Ayrıca bir grup insanın sürekli olarak istenmeyen e-postaları tespit edebilmesi de kaynak açısından sürdürülebilir değildir. Ancak istenmeyen e-postalardan oluşan bir veri seti ile eğitilen makine öğrenmesi veriler arasındaki kalıpları tespit ederek çok daha verimli sonuçlar elde edilebilmektedir. Benzer şekilde günümüzde daha birçok

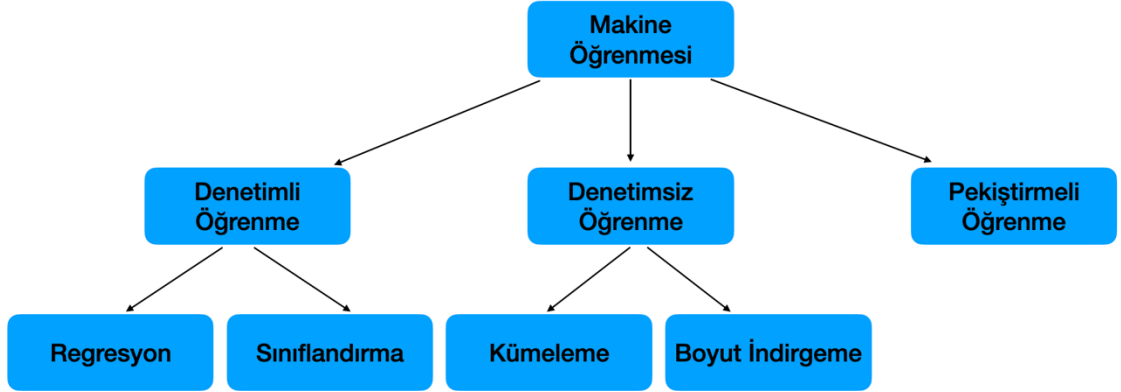
alanda makine öğrenmesinin daha yaygın bir şekilde kullanılması sayesinde sınırlı kaynakların daha verimli kullanılmasını sağlanmaktadır.

Makine öğrenmesi günümüzde en hızlı büyüyen alanlardan birisi olarak karşımıza çıkmaktadır. Özellikle otonom araçlar, sağlık, üretim, eğitim, finans ve pazarlama endüstrilerinde makine öğrenmesinin kullanımı giderek artmaktadır (Carleo vd., 2019; Jordan & Tom M. Mitchell, 2015). Gündelik hayatımızda kullandığımız neredeyse her teknolojiye bulunan makine öğrenmesi hayatlarımız üzerinde güçlü bir etkiye sahiptir. Gündelik teknolojilerde ve endüstrideki kullanımına ek olarak günümüzde doğa bilimlerinde de insanlığın kendisini ve evreni anlama çabasında makine öğrenmesinden yararlanılmaktadır (Zhou, 2021).

## **2.1. MAKİNE ÖĞRENMESİ KATEGORİLERİ**

Tüm problemler için kullanılacak tek bir makine öğrenmesi algoritması yoktur. Makine öğrenmesi farklı şekillerde kategorilere ayrılabilir. Genellikle kapsayıcı üç ana kategoriye ayrılmaktadır. Bu kategoriler denetimli öğrenme (supervised learning, gözetimli öğrenme), denetimsiz öğrenme (unsupervised learning, gözetimsiz öğrenme) ve pekiştirmeli öğrenmedir (reinforcement learning, takviyeli öğrenme). Ek olarak yarı denetimli öğrenme (semi-supervised learning), çok görevli öğrenme (multi-task learning), topluluk öğrenimi (ensemble learning), yapay sinir ağları (neural networks) ve örnek tabanlı öğrenme (instance based learning) makine öğrenmesi kategorilerinden gösterilebilmektedir (Mahesh, 2018).

**Şekil 4: Makine öğrenmesi kategorileri**



### **2.1.1. Denetimli Öğrenme**

Günümüzde denetimli öğrenme en çok kullanılan makine öğrenmesi türüdür ve bu alanda gün geçtikçe daha fazla gelişme yaşanmaktadır. Denetimli öğrenme, örnek bir veri kümesi ve halihazırda bilinen sonuçlarla eğitilerek değişkenler arasındaki ilişkileri keşfeder. Diğer bir ifadeyle denetimli öğrenme, bağımlı değişkenler (X) ile bağımsız değişken (y) arasındaki ilişkiyi keşfetmeyi amaçlamaktadır (Mirasçı Seray, 2023). Bilinen sonuçlardan çıkardığı ilişkiler sayesinde makine öğrenmesi daha önce hiç görmediği verilerin sonuçları üzerine tahminde bulunabilmektedir. Bağımlı değişkenlerin ve bağımsız değişkenin bilinmesinden dolayı kullanılan veri etiketli veri (labeled data) olarak nitelendirilebilir. Denetimli öğrenme, modelin eğitimi için kullanılan bağımlı değişkenlerden elde edilecek hedefler belirlenerek uygulanır. Bu yüzden denetimli öğrenme görev odaklı yaklaşıma sahip olduğu ifade edilmektedir (Sarker, 2021).

Denetimli öğrenme için veri kümesi belirlenen oranlarda eğitim ve test kümesi olmak üzere ikiye bölünür. Genellikle veri kümesinin %75'i eğitim verisi %25'i ise test verisi olarak tanımlanmaktadır. Eğitim için ayrılan veri kümesi ile makine öğrenmesi eğitilerek değişkenler arasındaki ilişkiler tespit edilmeye çalışırken, test için ayrılan veri kümesi ile öğrenme performansı (doğruluğu) ölçülür (Demirci, 2022). Makine öğrenmesi modelinin eğitilmesi ve test edilmesi tamamlandıktan sonra, geriye dönülerek farklı parametreler ile

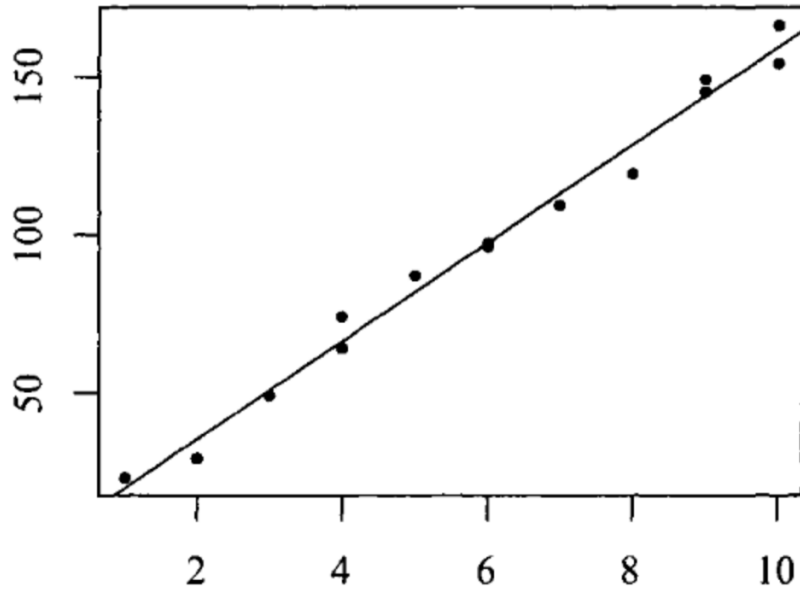
model iyileştirilebilir veya gerçek dünya verileri üzerinde tahminler yapmak üzere kullanılmaya başlanabilir.

Denetimli öğrenme, tahmin edilmeye çalışılan çıktının (y) kesikli veya sürekli olmasına bağlı olarak sınıflandırma ve regresyon olmak üzere iki ana kategoriye ayrılmaktadır (Demirel, 2023).

#### 2.1.1.1. Regresyon

Regresyon analizi, makine öğrenmesi algoritmaları arasında kullanımı en kolay olan algoritmalarından biri olmakla birlikte özellikle ekonomi ve finans alanlarında sıklıkla kullanıldığı görülmektedir. Regresyon analizi bağımlı değişkenlerin bağımsız değişken üzerindeki etkisini belirlemek için kullanılmaktadır. Diğer bir ifadeyle, regresyon analizi nicel veya nitel bağımsız değişkenler ile bağımlı değişken arasındaki ilişkiyi açıklamak için kullanılmaktadır. Regresyon analizinin sonucunda bir regresyon denklemi (model) elde edilir ve tahmin edilecek verilen bağımlı değişkenleri modelde yerine konularak bağımsız değişken tahmin edilir (Demirel, 2023).

**Şekil 5: Örnek bir regresyon modeli**



Kaynak: (Chatterjee & Hadi, 2006)

Evlerin özelliklerinden yola çıkarak fiyatının tahmin edilmesi, çalışanların performans bilgileri üzerinden maaşlarının tahmin edilmesi veya geçmiş yıllardaki hava durumu verilerine bakılarak bugün havanın kaç derece olacağı gibi örnekler regresyon analizine örnek olarak gösterilebilir.

Sıklıkla kullanılan regresyon türleri doğrusal regresyon, doğrusal olmayan regresyon ve çoklu doğrusal regresyondur (Theobald, 2017). Bunlara ek olarak kademeli regresyon, polinom regresyon, ridge regresyon, lasso regresyon, elastic-net regresyon, kantil regresyon ve sıralı regresyon gibi regresyon türleri de vardır.

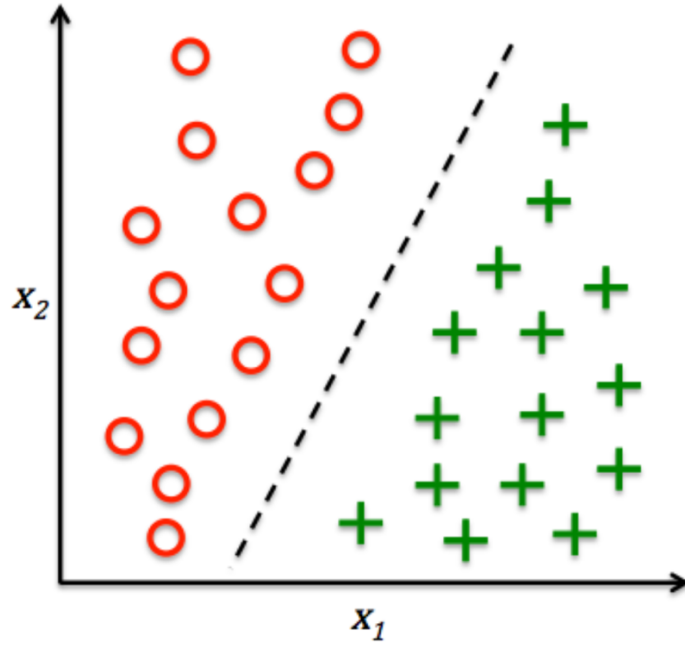
#### *2.1.1.2. Sınıflandırma*

Sınıflandırma problemi denetimli makine öğrenmesi algoritmalarının en sık kullanıldığı problemlerdendir. Sınıflandırma yapabilmek için her algoritmanın verilerin etiketli (labeled) çıktı değeri belirtilmiş olmalıdır. Sınıflandırma algoritmaları, bağımlı değişkenler ile bağımsız kategorik sınıf değişkeni arasındaki ilişkiyi kurmaya çalışır.

Etiketli eğitim verisi ile eğitilen algoritmalar test verisindeki etiketleri tahmin etmeye çalışılarak algoritmanın öğrenme performansı (doğruluğu) test edilir. Oluşturulan sınıflandırma modeli, eğitim aşamasında değişkenler arasında kurduğu ilişkiler ile uygulama aşamasında etiketlenmemiş gerçek dünya verilerinin sınıflarını tahmin etmeye çalışır (Kotsiantis, 2007).

Özetle, regresyon modelleri çıktı olarak sayısal bir değer tahmin ederken sınıflandırma modelleri ise sınırlı sayıda belirli kategorik değerlerden birini tahmin etmektedir.

Şekil 6: Örnek bir sınıflandırma modeli



Kaynak: (Raschka, 2015)

Test sonuçlarından hastalık teşhisi konulması, bir e-postanın istenmeyen bir e-posta olup olmadığı, hesap hareketlerinde sahtekarlık riski olup olmadığı, müşterilerin firmaları terk edip etmeyeceği gibi örnekler sınıflandırma problemlerine örnek olarak gösterilebilir.

Destek vektör makineleri, bayes teoremi, en yakın komşu, karar ağaçları, rastgele orman, lojistik regresyon ve yapay sinir ağları sıklıkla kullanılan sınıflandırma modellerinden bazılarıdır.

### 2.1.2. Denetimsiz Öğrenme

Denetimli öğrenme yönteminde makine öğrenmesi, sonucu bilinen gözlem değerleri ile eğitilirken, denetimsiz öğrenme yönteminde makine öğrenmesine verilen gözlem değerlerinin beklenen bir çıktısı yoktur (Ghahramani, 2004). Diğer bir ifadeyle denetimsiz öğrenme yönteminde veriler etiketlenmemiştir ve algoritmalar verilerin beklenen çıktı değerleri ile eğitilmemektedir. Denetimsiz öğrenme, açıkça eğitilmeden büyük veri kümelerinde verilerin arasındaki kalıpları keşfetmeyi veya verileri birkaç kategoride sınıflandırmayı amaçlamaktadır (Wang, 2016).



Denetimsiz öğrenmenin en büyük avantajı, veri setinde var olduğundan haberiniz olmayan kalıpları keşfetmenizi sağlayabilmesidir (Theobald, 2017). Örneğin, denetimsiz öğrenme algoritmaları sayesinde müşteri hareketlerinden oluşan bir veri kümesinde daha önce hiç öngörülemeyen bir müşteri grubu ortaya çıkarılabilir. Gerçek dünya verileri genellikle etiketlenmemiş veriler olduğundan dolayı denetimsiz öğrenme algoritmaları için gün geçtikçe daha fazla kullanım alanı oluşmaktadır. Ancak denetimsiz öğrenme algoritmalarının bazı dezavantajları da bulunmaktadır. Denetimsiz öğrenme algoritmalarının işlem süresi uzun ve karmaşık olabilmektedir. Ayrıca denetimsiz öğrenme modellerinin çıktılarının hangi ilişkilere göre yapıldığı konusunda çok net olmayabilmektedir. Denetimli öğrenme modellerinde olduğu gibi net bir başarı kriterinin olmaması ve kavramların yeterince iyi tanımlanamaması gibi sebeplerden dolayı denetimli öğrenmeye kıyasla daha az tercih edilmektedir (Watson, 2023).

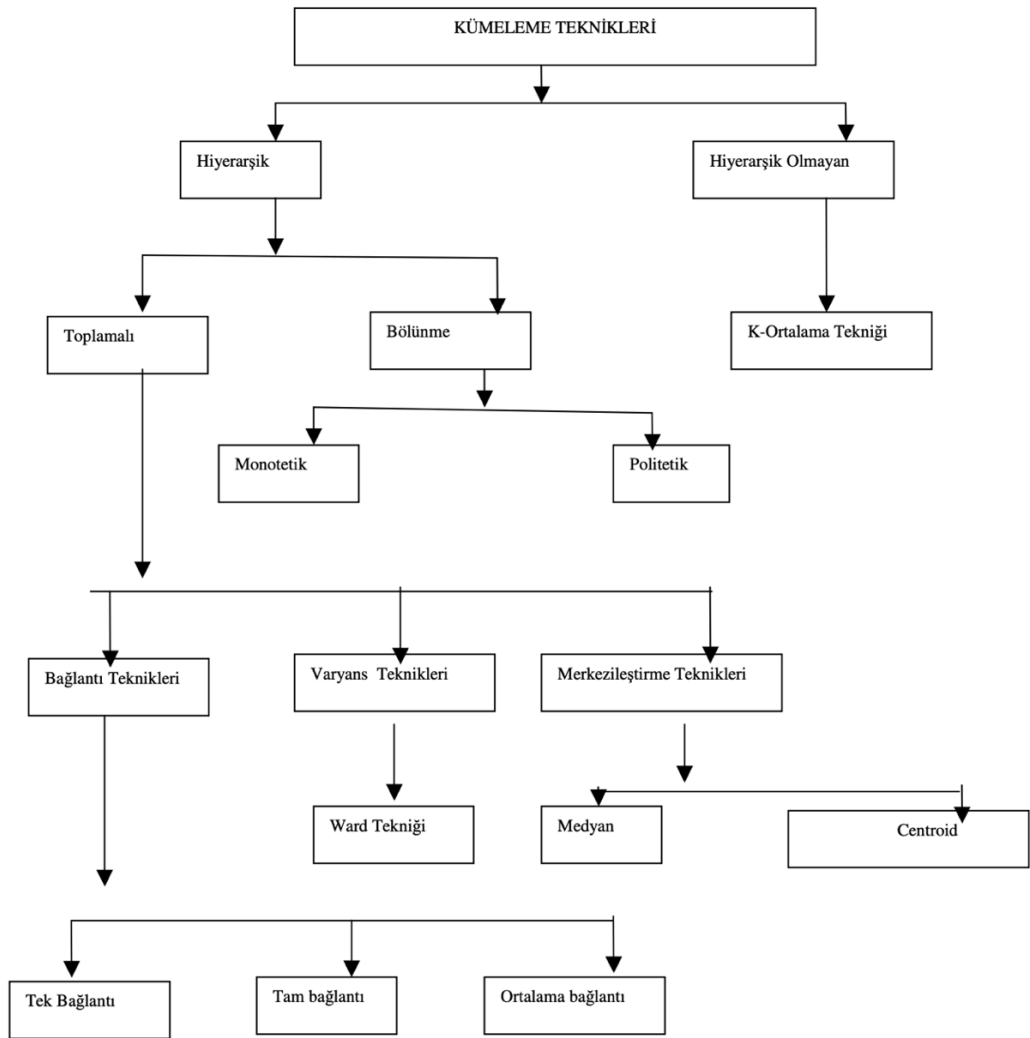
Denetimsiz öğrenme, kümeleme, birliktelik kuralları analizi, sosyal ağ analizi ve boyut indirgeme gibi kategorilere ayrılmaktadır. Bu tez çalışmasında denetimsiz öğrenmenin kümeleme ve boyut indirgeme yöntemleri kullanılmıştır.

#### *2.1.2.1. Kümeleme*

Kümeleme en yaygın kullanılan denetimsiz öğrenme yöntemidir. Kümeleme, veri kümesindeki kalıpların denetlenmeden gruplara (kümelere) sınıflandırılmasıdır (Jain vd., 1999). Denetimli öğrenme yönteminde sınıflandırma, denetimsiz öğrenme yönteminde ise kümeleme kullanılmaktadır. Sınıflandırmanın amacı tahmine dayalı iken kümelemenin amacı tanımlamaktır (Rokach & Maimon, 2006). Diğer bir ifadeyle kümeleme yöntemi etiketlenmemiş verilerden oluşan bir veri kümesindeki benzer gözlem değerlerini kendileri arasında gruplandırmaktadır. Oluşturulan kümelerdeki gözlem değerlerine birbirine yakın ve benzer özellikler sergilemekteyken, diğer kümelerdeki gözlem değerlerine kıyasla farklı özellikler sergilemektedir. Kümeleme modelinin başarısı da bu mantık üzerinden ölçülebilmektedir.

Kümeleme yöntemi bir denetimsiz öğrenme yöntemi olduğu için kümeleme ile oluşturulacak gruplar ve bu grupların özellikleri hakkında önceden bilgi sahibi olunamaz. Grupların ne şekilde oluşturulacağına model karar vermektedir. Bu tez çalışmasında hiyerarşik olmayan kümeleme tekniği k-ortalamlar kullanılmıştır. Kümeleme yöntemi kendi içerisinde hiyerarşik kümeleme teknikleri ve hiyerarşik olmayan kümeleme teknikleri olmak üzere iki ana gruba ayrılabilir (Koltan & Patır, 2011).

**Şekil 7: Kümeleme teknikleri**



Kaynak: (Koldere Akın, 2008)

### 2.1.2.2. Boyut İndirgeme

Makine öğrenmesi algoritmalarında model ne kadar fazla veri ile beslenirse, kurulan model o kadar iyi sonuç vermektedir. Ancak aynı şekilde modelin karmaşıklığı ve işlem süresi de artmaktadır. Veri kümesindeki bazı değişkenler birbirileri arasında yüksek derecede korelasyon içeriyor olabilir ve bu değişkenlerin modele beslenmesi gereksiz olabilir (Raschka, 2015).

Boyut indirgeme, yüksek boyutlu verilerin daha az boyutlu bir temsiline dönüştürülmesi işlemidir. İndirgenmiş veri kümesi, verinin gerçek boyutlu halinin özelliklerini yansıtmalıdır (Van Der Maaten vd., 2009). Diğer bir ifadeyle boyut indirgeme algoritmaları, veri kümesindeki değişken sayısını azaltarak makine öğrenmesi algoritmalarını çok boyutlu veri kümeleri ile beslemek yerine birbirinden daha kolay ayırt edilebilen daha az sayıda değişken ile beslemeyi amaçlar. Bu sayede makine öğrenmesi algoritmalarının kullandığı sınırlı depolama alanını ve işlem performansını daha verimli kullanarak daha iyi sonuçlar elde edilebilmesi amaçlanmaktadır. Boyut indirgeme yöntemi veri görselleştirmek için de kullanılabilir. Her bir gözlem değerinin üçten fazla değişkeni olan veri kümeleri görselleştirilemezken, boyut indirgeme algoritmaları sayesinde iki veya üç değişkene indirgenen veri kümeleri iki veya üç boyutlu düzlem üzerinde görselleştirilebilir (Raschka, 2015).

Boyut indirgeme yöntemleri en genel haliyle doğrusal ve doğrusal olmayan boyut indirgeme yöntemleri olarak iki ana kategoriye ayrılmaktadır (Glielmo vd., 2021). Bu tez çalışmasında boyut indirgeme yöntemi olarak doğrusal bir boyut indirgeme yöntemi olan temel bileşen analizi (PCA) yöntemi kullanılmıştır.

### 2.1.3. Pekiştirmeli Öğrenme

Pekiştirmeli öğrenme yöntemi içinde bulunduğu dinamik ortamda deneme yanılma yoluyla öğrenen bir makine öğrenmesi yöntemidir. Pekiştirmeli öğrenme algoritmaları bir görevi nasıl tamamlayacağını açıkça göstermeden belirtilen göreve yönelik aldığı aksiyonlar karşısında aldığı ödül veya cezalar ile sonuca ulaşır (Pack Kaelbling vd.,

1996). Dięer bir ifadeyle pekiřtirmeli öğrenme yönteminde algoritmalar, canlılarda olduęu gibi davranıřlardan edindięi tecrübelerden yararlanılır.

Pekiřtirmeli öğrenme algoritmaları, net bir řekilde belirlenmiř bir görevi yerine getirmek için, daha önce hiç görmedięi bir ortamın özelliklerini algılayarak bu çevreyi etkileyecek eylemlere karar verirler (Sutton & Barto, 2018). Yaptıęı eylemler sayesinde edindięi tecrübeler sonucunda aldıkları geri dönütler sayesinde sonraki kararlarını verir. Pekiřtirmeli öğrenme son yıllarda makine öğrenmesi alanında en popüler konularından biri haline gelmiřtir. Özellikle otonom araç, saęlık, finans ve oyun sektörlerinde sıklıkla kullandığı görülebilmektedir.

## **2.2. MAKİNE ÖĞRENMESİ ALGORİTMALARI**

Günümüzde, farklı ihtiyaçlar için özel olarak geliřtirilmiř birçok makine öğrenmesi algoritması bulunmaktadır. Bu algoritmaların tamamını bilmek mümkün olmasa da bazı algoritmaların literatürde dięer algoritmalara göre daha sık kullanıldıęı söylenebilir. Tezin bu bölümünde, k-en yakın komřu (k-nearest neighbors: kNN) karar ağaçları (decision tree), rastgele orman (random forest) ve yapay sinir ağları (neural networks) gibi sık kullanılan makine öğrenmesi algoritmalarına yer verilmiřtir.

Ek olarak, rastgele seçilmiř bir grup Hepsiburada kullanıcılarının alışveriř verileri üzerinden bir müşteri segmentasyonu uygulaması yapılan bu tez çalışmasında kullanılan algoritmalar da bu bölümde açıklanmıřtır. Bu algoritmalar; sıklıkla kullanılan bir boyut indirgeme algoritması olan temel bileřenler analizi algoritması ve en yaygın kullanılan kümeleme algoritması k-ortalamlardır.

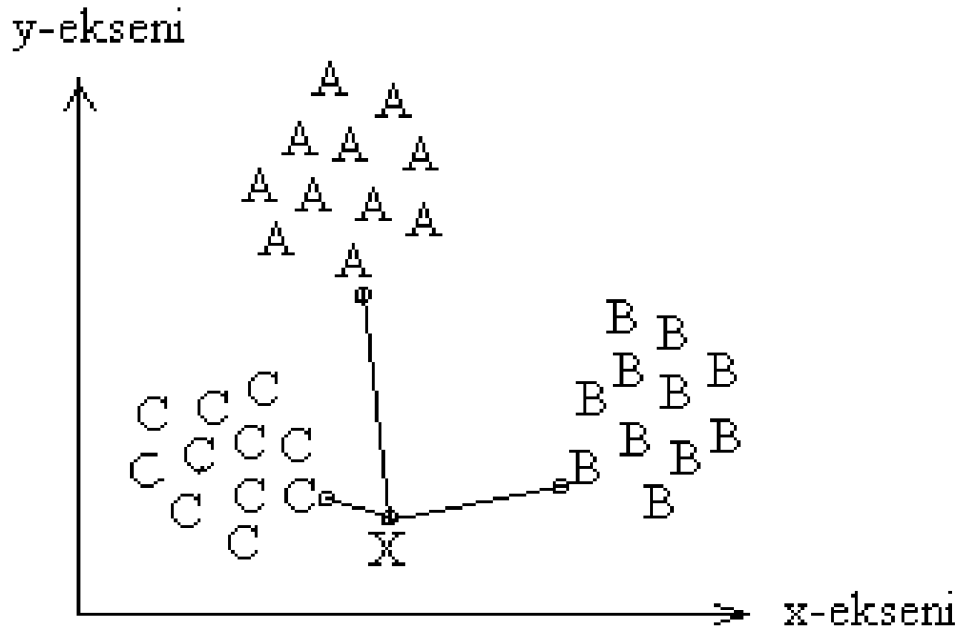
### **2.2.1. k-En Yakın Komřu**

T. M. Cover ve P. E. Hart tarafından önerilen k-en yakın komřu (kNN) algoritması bir gözlem deęerinin en yakınında bulunan k adet komřusuna göre sınıflandırıldıęı bir yöntemdir (Cover & Hart, 1967; Tařçı & Onan, 2016). K-en yakın komřu algoritmasının kullanılabilmesi için, bakılacak en yakın komřu sayısı olan k parametresinin ve aradaki

komşularla arasındaki mesafenin ölçüleceği mesafe fonksiyonunun belirtilmesi gerekmektedir.

K-en yakın komşu algoritması, anlaşılması basit ve etkili bir algoritma olmasına rağmen yavaş çalışan bir algoritmadır. Eğitim verisi üzerinden bir model oluşturmak yerine eğitim verisini ezberleyip, her yeni bir gözlem değeri tahmin edilmeye çalışıldığında gözlem değerinin eğitim verisindeki komşularına uzaklığını hesaplayarak bir tahminde bulunur. Komşular arası uzaklık öklid uzaklığı, manhattan uzaklığı, chebysev ve minkowski uzaklığı vb. gibi uzaklık ölçütleriyle hesaplanabilmektedir (Dilki & Deniz Başar, 2020; Taşçı & Onan, 2016).

**Şekil 8: K-en yakın komşu yöntemi**



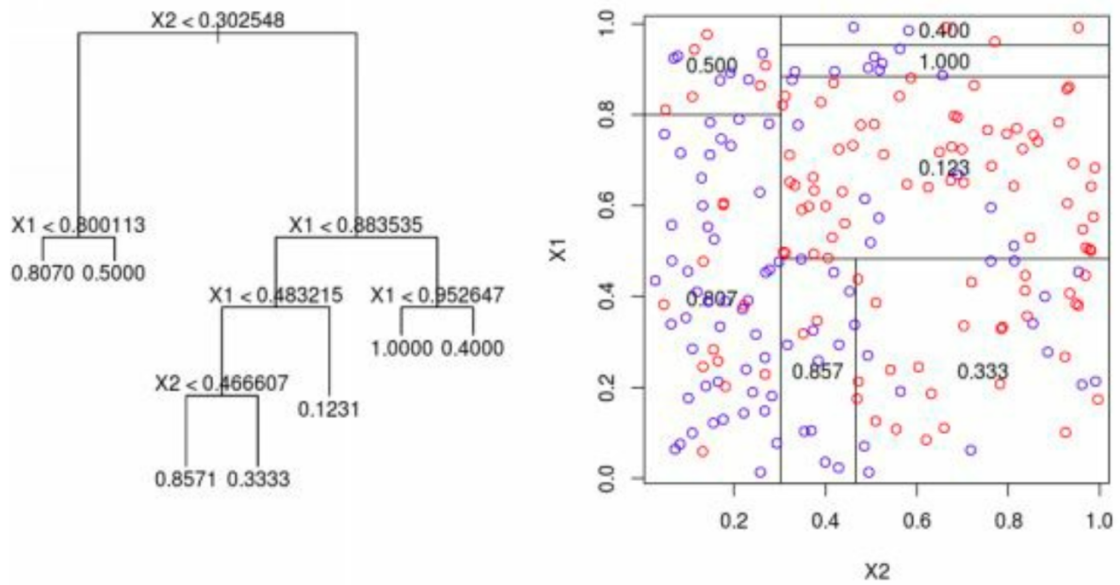
Kaynak: (Aşlıyan & Günel, 2010)

### 2.2.2. Karar Ağaçları

Denetimli bir makine öğrenmesi algoritması olan karar ağaçları, kural tabanlı bir hiyerarşik sınıflandırma algoritmasıdır. Şekil 9’da görülebileceği üzere, karar ağaçları algoritması aslında bir sınıflandırma algoritması olmasına rağmen regresyon

problemlerinde de kullanılabilir. Karar ağaçları basit testleri mantıksal olarak birleştiren sıralı modellerdir. Diğer makine öğrenmesi algoritmalarına kıyasla daha anlaşılabilir olduğu için araştırmacılar tarafından sıklıkla tercih edilmektedir (Kotsiantis, 2013). Diğer bir ifadeyle, araştırmacı karar alırken hangi unsurların göz önüne alınması gerektiğini anlayabilir ve karar kriterlerinin farklı çıktılar ile ilişkilerinin belirlemesi konularında bilgi sahibi olabilir (Emel & Taşkın, 2005).

**Şekil 9: Karar ağaçlarının yapısı**



Kaynak: (Theobald, 2017)

Karar ağaçları kök, dal (iç karar düğümleri) ve yaprak olarak ifade edilen düğümlerden oluşmaktadır. Karar ağaçları, kök düğümünden başlayarak, dallarda belirlenen kurallara bağlı olarak alt gruplara ayrılmaktadır ve bir yaprağa ulaştığında bir çıktı elde edilmektedir. Karar ağaçlarında kök ve yaprak arasında kalan düğümlere dal denir ve dallanma, modelin karışıklığına göre artabilmektedir.

Karar ağaçları algoritmaları, veri kümesindeki gözlem değerlerini homojen gruplara ayırmak için gerekli karar kriterlerini belirlemektedir. Algoritma, farklı gruplar arasındaki varyansı (entropi) en düşük tutacak karar kriterlerini bulmaya çalışır. Son olarak, karar ağaçları algoritmaları sınıflandırma problemlerinde özellikle görsel açıdan

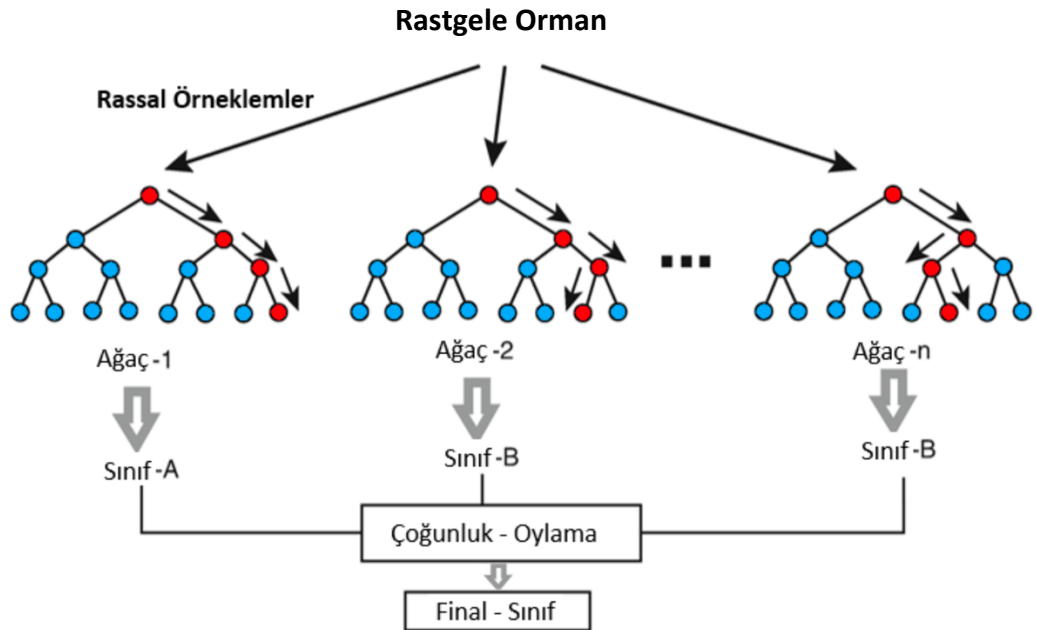
çok yararlı olabilirken, aşırı öğrenmeye duyarlı ve çok esnek olmayan bir algoritmadır (Theobald, 2017). Diğer bir ifadeyle, karar ağaçları yeni veri kümelerinde yeterince iyi sonuç veremeyebilir.

### 2.2.3. Rastgele Orman

Rastgele orman algoritması, en yaygın kullanılan toplu sınıflandırma yöntemlerinden biridir. Toplu sınıflandırma yöntemleri, tek bir sınıflandırıcı model yerine birçok sınıflandırıcı oluşturur ve oluşturulan bu sınıflandırıcıların tahminlerini birleştirerek bir çıktı elde eder (Akar & Güngör, 2012). Diğer bir ifadeyle, toplu sınıflandırma yöntemi; birden fazla algoritma ya da birden fazla ağacın bir araya getirilerek toplu bir şekilde öğrenmesi ve tahmin etmeye çalışmasıdır.

Rastgele orman, torbalama tekniği (bagging - bootstrap aggregating: yerine koymalı seçme işlemi) kullanarak rastgele örneklemeler oluşturur ve bu örneklemelerden rastgele değişkenler seçerek karar ağaçları oluşturur. Bu işlemi birçok kez tekrar eder ve Şekil 10'da görülebileceği üzere bir karar ormanı elde edilmiş olur.

Şekil 10: Karar ormanı



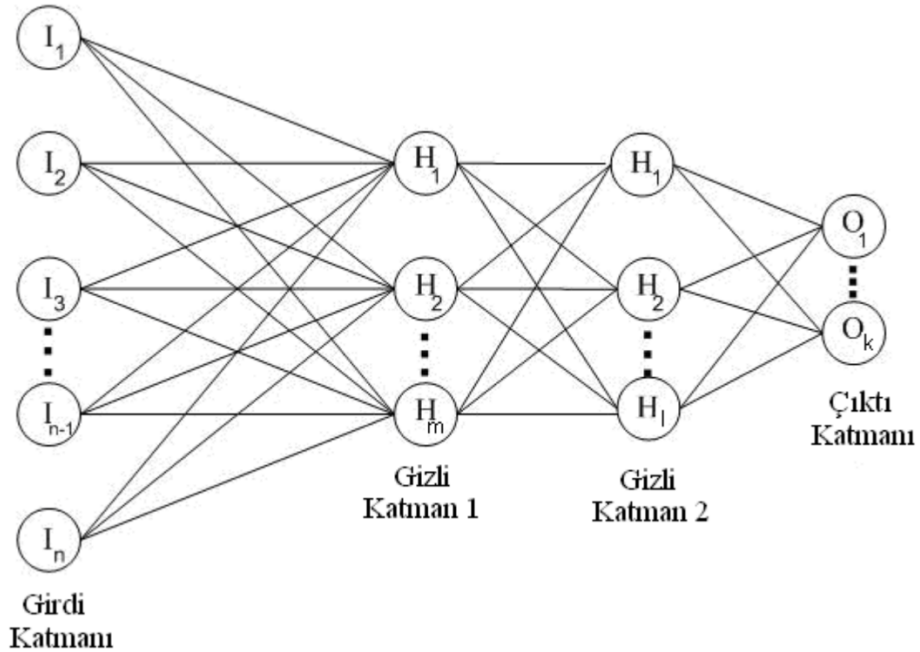
Kaynak: (Sniatala vd., 2020; Tütüncü, 2022)

Özetle, rastgele orman algoritması torbalama tekniği kullanarak birçok karar ağacı oluşturur ve yeni gelen gözlem değerini tüm karar ağaçlarından geçirerek çıkan sonuçları oylar. Bu şekilde, birçok karar ağacının sonuçları birleştirilerek bir çıktı elde edilir. Bu sayede daha iyi sonuçlar elde edilebilmektedir fakat rastgele orman algoritmaları oluşturulan rastgele ormanın büyüklüğüne göre daha yavaş çalışabilmektedir.

#### 2.2.4. Yapay Sinir Ağları

İnsan beyninin çalışma şekline esinlenen yapay sinir ağları, deneme yoluyla öğrenerek genelleştirme yapabilmekte ve geleceğe dair tahminlerde bulunabilmektedir (Hamzaçebi & Kutay, 2004). Yapay sinir ağları, herhangi bir ön bilgi verilmeksizin girdi ve çıktı değerler arasındaki ilişkilerden yola çıkarak bir fonksiyon ortaya çıkarması ve ağı eğitilmesini sağlamaktadır (Hamzaçebi & Kutay, 2004; Kaastra & Boyd, 1999). Diğer bir ifadeyle, yapay sinir ağları parametrik olmayan esnek bir modelleme aracıdır (Karaatlı vd., 2012; Tang & Chi, 2005).

Şekil 11: Yapay sinir ağı



Kaynak: (Kutlu & Badur, 2009)



Yapay sinir ađları, nöron denilen işlem birimlerinden oluşmaktadır. Nöronlar, bir önceki ve bir sonraki katmanda bulunan her nörona farklı sayısal ağırlıklarla bađlıdır (Kutlu & Badur, 2009). Şekil 11’de de görülebileceđi üzere nöronlar, katman denilen mantıksal grupları oluşturur. Yapay sinir ađı, bir girdi katmanı, gizli katman (veya katmanlar) ve bir de çıktı katmanından oluşmak üzere en az üç katmandan oluşmalıdır. Gizli katmanlar birden çok olabilmektedir.

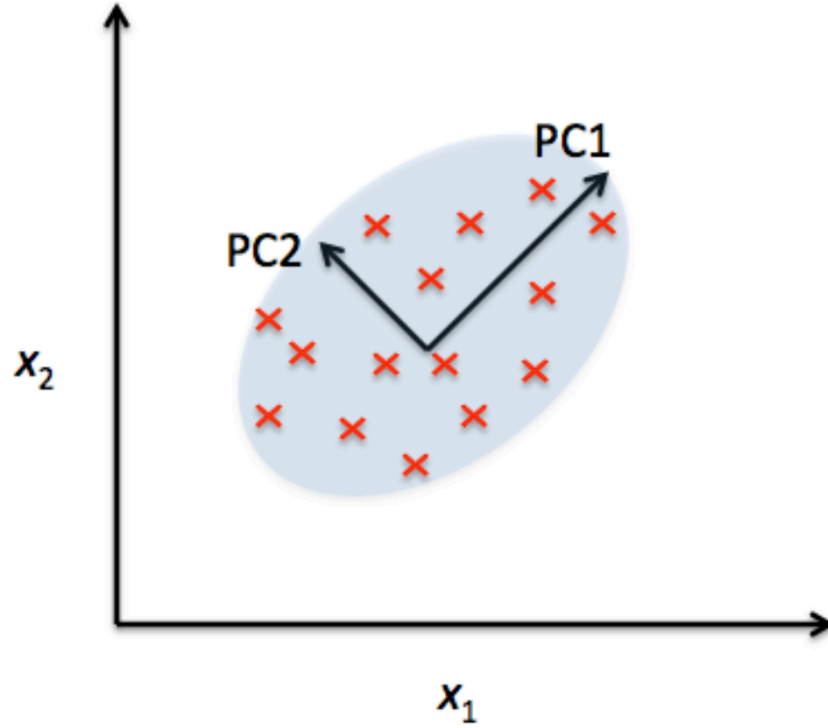
Özetle, yapay sinir ađları modelinin eğitilmesi sırasında, nöronların birbirine bađlandığı sayısal ağırlıkların ayarlanarak bir fonksiyon elde edilmesi sađlanır.

### ***2.2.5. Temel Bileşenler Analizi (PCA)***

Büyük veri kümeleri genellikle çok fazla deđişkenden meydana gelmektedir. Bazı durumlarda daha az deđişken ile veri kümesinin benzer özelliklerini temsil eden yeni veri kümeleri elde etmek mümkün olabilmektedir (Jolliffe, 2005). Temel bileşenler analizi ilk kez Pearson (1901) tarafından ortaya atılmış ve daha sonra Hotelling (1933) tarafından geliştirilmiştir (Ersungur vd., 2007; Hotelling, 1933; Pearson, 1901).

Temel bileşenler analizi, veri kümesindeki deđişkenler arasındaki varyans ve kovaryans yapısını, bu deđişkenlerin doğrusal birleşimleri aracılığıyla açıklayarak, boyut indirgenmesini sađlayan bir istatistik yöntemidir (Yıldız vd., 2010). Diđer bir ifadeyle bađımlılık yapısı gösteren deđişkenlerden doğrusal, ortogonal ve birbirinden bađımsız yeni deđişkenler oluşturulur (Ersungur vd., 2007). Bu sayede deđişkenler arası bađımlılık azalmış olur.

Şekil 12: Temel bileşenler analizi (PCA) algoritması ile boyut indirgeme



Kaynak: (Raschka, 2015)

Temel bileşenler analizi, veri kümesini makine öğrenmesi modeline beslemeden önce yapılan veri ön işleme adımlarından biridir. Literatürde en yaygın kullanılan boyut indirgeme yöntemi temel bileşenler analizidir. Ayrıca çok boyutlu veri kümelerini temel bileşenler analizi ile iki veya üç boyuta indirgeyerek görselleştirmek mümkündür.

### 2.2.6. K-Ortalamlar

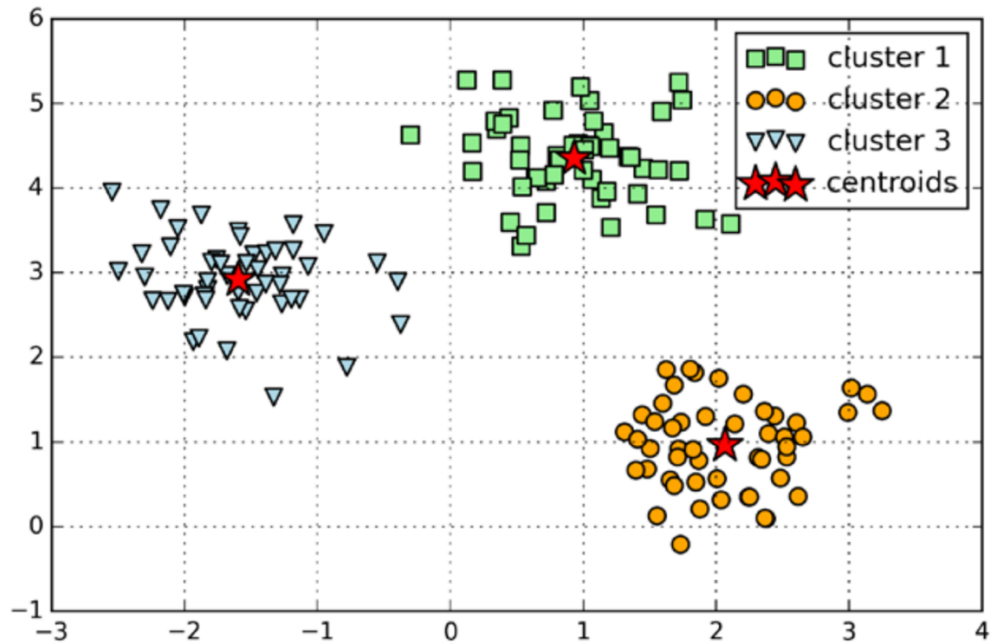
K-ortalamlar algoritması J.B. MacQueen (1967) tarafından önerilen bir kümeleme algoritmasıdır (Macqueen, 1967). K-ortalamlar (k-means), veri kümesindeki benzer özellikleri taşıyan gözlem değerlerinin bir arada gruplandırılması için kullanılan popüler bir denetimsiz öğrenme algoritmasıdır (Gürler vd., 2020; Kırmızıgül & İbrahim, 2008; Li & Wu, 2012; Sarıman, 2011). K-ortalamlar hiyerarşik olmayan kümeleme algoritmalarından biridir ve literatürde en yaygın olarak kullanılan kümeleme algoritmalarından biridir.

Bir gözlem değerinin birden fazla kümede olabildiği kümeleme yöntemlerinin aksine k-ortalamlar algoritmasında her bir gözlem değeri tek bir kümeye atanır (Bramer, 2007). K-ortalamlar algoritması ile kümeleme işleminin gerçekleştirilebilmesi için öncelikle veri kümesinin kendi içerisinde kaç kümeye ayrılacağı (k) belirlenmelidir. Dirsek yöntemi (Elbow method) veya silhouette skoru gibi yöntemlerle veri kümesinin ayrılacağı en uygun küme sayısı belirlenebilir.

K-ortalamlar algoritması şu şekilde çalışır (Bramer, 2007):

1. Küme sayısı (k) belirlenir
2. K-ortalamlar algoritması, belirlenen küme sayısı kadar küme merkezlerini (centroid) rastgele konumlara yerleştirir
3. Gözlem değerleri en yakında bulunan küme merkezinin grubuna atanır
4. Küme merkezleri tekrar hesaplanır
5. Küme merkezleri artık hareket etmeyince kadar üçüncü ve dördüncü adımlar tekrarlanır.

**Şekil 13: K-ortalamlar algoritması ile veri kümeleme**



Kaynak: (Raschka, 2015)

Özetle, k-ortalamlar algoritması gözlem değerlerinin benzer özelliklerine göre belirli kümeler oluşturur. Kümeleri en iyi şekilde temsil edene kadar küme merkezlerinin yeri ve o merkeze yakın gözlem değerlerinin kümeleri değiştirilir. En uygun sonucu bulana kadar bu döngü tekrarlanır. Eğer iyi bir kümeleme işlemi yapıldıysa, kümelerin kendi içerisinde homojen, kümeler arasında ise heterojen olması beklenmektedir (Sammut & Geoffrey I. Webb, 2011).

## ÜÇÜNCÜ BÖLÜM

### MÜŞTERİ SEGMENTASYONU UYGULAMASI

Bu çalışmada müşteri segmentasyonu uygulanacak veri, Hepsiburada firmasının veri tabanından alınmıştır. Çalışmada kullanılacak veri kümesi rastgele seçilmiş yaklaşık 5.000 kullanıcının 2022 yılı içerisinde satın aldığı ürünleri ve bu ürünlerin ait oldukları kategorileri içermektedir. Kişisel Verilerin Korunması Kanunu (KVKK) gereği, veri kümesinde kişisel bilgi bulunmamasına özen gösterilmiştir. Kullanıcılara ait çalışmada kullanılması hedeflenen hassas sayılabilecek bilgiler ise şifrelenmiştir.

Uygulamada tüm kodlar Python programlama dili ile Jupyter Notebook ortamında yazılmıştır. Veri işleme ve analiz aşamalarında Pandas, Numpy, Scipy kütüphaneleri kullanılmıştır. Veri görselleştirme aşaması için ise yaygın olarak kullanılan Seaborn ve Matplotlib kütüphaneleri kullanılmıştır. Makine öğrenmesi uygulamaları için ise Scikit-learn kütüphanesi kullanılmıştır. Uygulamada kullanılan yazılım dili, ortam ve kütüphanelerin tümü yaygın şekilde kullanılan açık kaynak kodlu teknolojilerden oluşmaktadır.

Uygulamanın ilk aşaması veri kümesinin çalışma ortamına aktarılmasını ve veri tiplerinin dönüştürülmesini içermektedir. Ardından verilerin davranışlarını tanımaya yönelik bazı betimsel istatistiklerin çıkarılmasını içermektedir. Veri tipleri ve değişkenlerin açıklamaları da bu bölümde incelenmektedir.

Uygulamanın ikinci aşamasında, çalışmanın yapılabilmesi için veri ön işleme ve veri dönüşümü işlemleri yapılmıştır. Veri bütünlüğü kontrol edilerek, aykırı değerlerin çıkarılması ve veri kümesinin makine öğrenmesi modelinin eğitilebileceği şekilde düzenlenmesi için gerekli manipülasyonlar bu aşamada uygulanmıştır.

Uygulamanın üçüncü aşamasında, makine öğrenmesi modeli kullanılarak müşteri segmentasyonu işlemi gerçekleştirilmiştir. Makine öğrenmesi modeli uygulanmadan önce verilerin standardize edilmesi, modeli eğitecek değişkenleri azaltmak için temel

bileşen analizi, optimal grup sayısının belirlenmesi için Elbow yöntemi (dirsek yöntemi) ve kümeleme tekniğinin değerlendirilmesi için Silhoutte yöntemi kullanılmaktadır.

### **3.1. KULLANILAN ARAÇLARIN TANIMLANMASI**

#### ***3.1.1. Python Programlama Dili***

Guido Van Rossum (1990) tarafından geliştirilmeye başlanan Python programlama dili, günümüzde Python Yazılım Vakfı adlı kâr amacı gütmeyen kuruluş tarafından geliştirilmektedir. Ücretsiz ve açık kaynak kodlu bir yazılım dili olan Python programlama dili son yıllarda veri analizi alanında en yaygın kullanılan programlama dili haline gelmiştir. Python programlama dilinin veri analizi için geliştirilmiş birçok kütüphanesi bulunmaktadır. Veri üzerinde düzenleme ve analiz yapmak için NumPy, Pandas ve SciPy kütüphaneleri, veri görselleştirme için Matplotlib ve Seaborn kütüphaneleri ve makine öğrenmesi uygulamaları için Scikit Learn, Tensorflow, Keras, Pytorch kütüphaneleri yaygın kullanılan kütüphanelerdendir (Arslan İ, 2019).

#### ***3.1.2. Anaconda, Anaconda Navigator ve Jupyter Notebook***

Anaconda veri bilimi alanında kullanılan paketlerin yönetimini kolaylaştırmayı amaçlayan açık kaynak kodlu bir yazılımdır. Anaconda Navigator ise geliştiricilere komut satırını kullanmak yerine grafiksel bir ara yüz sağlayarak paket yönetimi ve ortam yönetiminin kolaylaştırılmasını sağlamaktadır.

Bilimsel çalışmaların günümüzde daha yoğun veri içermesi ve daha fazla hesaplama gerektirmesi, çalışmaların anlaşılabilmesini, yorumlanmasını ve tekrar edilebilmesini zorlaştırmıştır (Randles vd., 2017). Bu ihtiyaca çözüm olarak geliştirilen Jupyter Notebook web tarayıcısı üzerinden çalıştırılabilen ve interaktif programlama yapmayı sağlayabilen bir geliştirme ortamıdır. Ayrıca Jupyter Notebook ile yapılan çalışmaların ve elde edilen sonuçlarının dağıtımı kolay bir şekilde yapılabilir. Bu sayede çalışmaların Jupyter Notebook ile yapılan çalışmaların ve sonuçlarının farklı kişiler tarafından daha kolay tekrar edilebilmesi sağlanmaktadır.

### ***3.1.3. NumPy Kütüphanesi***

Dizi işlemlerinin daha verimli bir şekilde yapılabilmesi için 1990'lı yıllarda bir grup gönüllü tarafından geliştirilen bir veri yapısıdır. N-boyutlu NumPy dizisi ve bazı matematiksel işlemleri içeren NumPy paketi akademide ve endüstride yaygın bir şekilde kullanılmaktadır (Van Der Walt vd., 2011).

### ***3.1.4. Pandas Kütüphanesi***

Pandas kütüphanesi, Python dili ile diğer bilimsel programlama dilleri ve istatistik yazılımları arasındaki farkı kapatılmak için 2008 yılında geliştirilmeye başlamıştır. Başlangıçta finansal veri analizi amacıyla geliştirilmiştir. Kütüphanenin adı, istatistik ve ekonometride karşılaşılan çok boyutlu veri kümeleri için kullanılan panel veri (panel data) teriminden gelmektedir (McKinney, 2011).

### ***3.1.5. Matplotlib Kütüphanesi***

Matplotlib iki boyutlu verileri görselleştirme imkanı sunan bir Python kütüphanesidir. Kullanımı kolay ve esnek bir yapıya sahip olan Matplotlib kütüphanesi ile interaktif grafikler üretilebilmekte ve grafiklerin farklı formatlarda çıktıları alınabilmektedir (Tosi, 2009).

### ***3.1.6. Scikit-learn Kütüphanesi***

Scikit-learn, Python programlama diline entegre edilmiş, kolay kullanım sağlamayı hedefleyen ve içerisinde en yaygın kullanılan makine öğrenmesi algoritmalarını bulunduran açık kaynak kodlu bir kütüphanedir (Pedregosa vd., 2011).

## **3.2. VERİ KÜMESİNİN TANIMLANMASI**

Pandas kütüphanesi ile .csv formatındaki veri kümesi, Jupyter Notebook çalışma ortamına aktarılmıştır. Veri kümesinin tamamı, Kişisel Verilerin Korunması Kanunu

(KVKK) gereği paylaşılmayacak olmakla birlikte ilk beş satırı örnek olarak Tablo 1’de verilmiştir.

**Tablo 1: Veri kümesinin tanımlanması**

receivedAt	order Number	userId	sku	domain	basket Amount	product Amount
2022-01-01	22070875	89105699	1158940	Fashion	80.91	80.91
2022-01-01	34633985	59129796	3189275	Fashion	334.77	53.99
2022-01-01	34633985	59129796	5653133	Fashion	334.77	54.00
2022-01-01	34633985	59129796	6653633	Fashion	334.77	226.78
2022-01-01	72585287	28841111	2391104	Home	414.00	414.00

Müşterilerin bir yıl içerisindeki alışveriş bilgilerini içeren veri kümesi 72.863 satırdan ve Tablo 1’de yer alan 7 değişkenden oluşmaktadır. Her bir satır ilgili müşterinin bir siparişinde bulunan bir ürüne denk gelmektedir. Tablo 1’de görülebileceği gibi, veri kümesinde rastgele seçilen 4.963 kullanıcının 01-01-2022 ile 31-12-2022 tarihleri arasında vermiş olduğu 43.003 siparişinde toplamda 72.863 adet ürün satın aldığı görülmektedir.

#### Şekil 14: Veri kümesini oluşturan değerleri veren Python kodları

```
print(df.shape)
```

```
(72863, 7)
```

72.863 adet satırdan ve 7 adet kolondan oluşuyor.

```
print(f"Benzersiz kullanıcı: {len(df.groupby('userId'))}")
```

```
Benzersiz kullanıcı: 4963
```

```
print(f"Benzersiz sipariş: {len(df.groupby(['userId', 'orderNumber']))}")
```

```
Benzersiz sipariş: 43003
```

```
print("tarih aralığı: " + str(min(df['receivedAt'])) + " - " + str(max(df['receivedAt'])))
```

```
tarih aralığı: 2022-01-01 05:27:37+00:00 - 2022-12-31 23:30:39+00:00
```

Veri kümesini oluşturan değişkenler ve detaylı açıklamaları ise Tablo 2’de verilmiştir.



**Tablo 2: Veri kümesini oluşturan değişkenler ve açıklamalar**

<b>Değişken</b>	<b>Açıklama</b>
<b>receivedAt</b>	Kullanıcının sipariş verdiği anda siparişin gerçekleştiği tarih ve saati belirtir.
<b>orderNumber</b>	Sipariş numarası, kullanıcının sipariş verdiği anda oluşturulan bir numaradır. Hassas bilgi içerebileceği için maskelenmiştir.
<b>userId</b>	Kullanıcı numarası, her kullanıcının kendine özel ayırt edici bir numarası vardır. Hassas bilgi içerebileceği için maskelenmiştir.
<b>sku</b>	Stok tutma birimi, e-ticaret sitesinde satılan her bir ürünün kendisine ait bir stok tutma birimi numarası vardır. Hassas bilgi içerebileceği için maskelenmiştir.
<b>domain</b>	Ürünlerin ait olduğu ana kategoriyi temsil eder. Veri kümesinde 9 farklı ana kategori bulunmaktadır.
<b>basketAmount</b>	Sepet tutarı, kullanıcının yaptığı alışverişin toplam tutarını belirtir. Her orderNumber için bir basketAmount tutarı vardır.
<b>productAmount</b>	Ürün tutarı, kullanıcının aldığı her bir ürünün fiyatını belirtir. Her stok tutma birimine (sku) karşılık gelen bir ürün tutarı (productAmount) vardır. Aynı sepet (orderNumber) içerisindeki ürün tutarlarının (productAmount) toplamı sepet tutarına (basketAmount) eşittir. Eğer sepet tutarı (basketAmount) ve ürün tutarı (productAmount) eşit ise kullanıcının sepetinde tek bir adet ürün olduğu söylenebilir.

Tablo 2’de belirtildiği üzere sipariş numarası (orderNumber), kullanıcı numarası (userId) ve stok tutma birimi (sku) bilgileri hassas veri içerebileceği için Kişisel Verilerin Korunması Kanunu gereği maskelenmiştir. Şekil 15’te maskelenen verilerin veri tipinin nesne (object) olması beklenirken sayısal (int64) tipinde olduğu gözlemlenmiştir.

**Şekil 15: Tip dönüştürme işlemi uygulanmadan önce veri kümesindeki değişkenler**

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 72863 entries, 0 to 72862
Data columns (total 7 columns):
#   Column          Non-Null Count  Dtype
---  -
0   receivedAt      72863 non-null  object
1   orderNumber     72863 non-null  int64
2   userId          72863 non-null  int64
3   sku             72863 non-null  int64
4   domain          72863 non-null  object
5   basketAmount   72863 non-null  float64
6   productAmount  72863 non-null  float64
dtypes: float64(2), int64(3), object(2)
memory usage: 3.9+ MB
```

Şekil 16’da görülebileceği gibi, bazı değişkenlerine tip dönüşümü uygulanmıştır. Sipariş tarihi (receivedAt) değişkeni nesne (object) tipinden tarih-zaman (datetime) tipine; sipariş numarası (orderNumber), kullanıcı numarası (userId) ve stok tutma birimi (sku) değişkenleri sayısal (int64) tipinden nesne (object) tipine dönüştürülmüştür. Performans iyileştirmesi için az sayıda değişken ve çok kez tekrar eden domain değişkeni ise nesne (object) tipinden kategori (category) tipine dönüştürülmüştür. Ayrıca yine şekil 16’da veri kümesinde her bir değişkenden 72.863 adet gözlem değerinin bulunduğu ve boş değer olmadığı görülmektedir.

**Şekil 16: Tip dönüştürme işlemi uygulandıktan sonra veri kümesindeki değişkenler**

```
df = df.astype({"orderNumber":'object', "userId":'object', "sku": 'object', "domain": 'category'})
df['receivedAt'] = pd.to_datetime(df['receivedAt'])

df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 72863 entries, 0 to 72862
Data columns (total 7 columns):
#   Column          Non-Null Count  Dtype
---  -
0   receivedAt      72863 non-null  datetime64[ns, UTC]
1   orderNumber     72863 non-null  object
2   userId          72863 non-null  object
3   sku             72863 non-null  object
4   domain          72863 non-null  category
5   basketAmount   72863 non-null  float64
6   productAmount  72863 non-null  float64
dtypes: category(1), datetime64[ns, UTC](1), float64(2), object(3)
memory usage: 3.4+ MB
```

Veri kümesindeki nesne (object) ve kategorik (category) değişkenleri benzersiz olarak saydırıldığında veri kümesinin 43.003 adet sipariştten, 4.963 adet kullanıcıdan, 44.868

adet birbirinden farklı üründen ve 9 adet ana kategoriden oluştuğu görülmektedir. Kategorik bir değişken olan domain değişkenine ait ürünler saydırıldığında ise Kitap ve hobi kategorisinde 25.124 adet, Süpermarket kategorisinde 17.904 adet, Oda ve yaşam tarzı kategorisinde 11.477 adet, Teknoloji kategorisinde 7.724 adet, Ev ve bahçe kategorisinde 7.038 adet, Ev aletleri kategorisinde 2.519 adet, Hepsiburada hizmetleri kategorisinde 569 adet, Mobil kategorisinde 506 adet ve Diğer kategorisinde 2 adet ürün satıldığı görülmektedir (bkz. Ek 1).

### 3.2.1. Betimsel İstatistikler

Veri seti hakkında daha detaylı bilgi sahibi olabilmek için sayısal değerler üzerinden bazı betimsel istatistikler yorumlanabilir. Şekil 17’de veri kümesinde bulunan sepet tutarı (basketAmount) ve ürün tutarı (productAmount) sayısal değişkenlerinin betimsel istatistiklerine bakıldığında toplamda 72.863 adet veri olduğu ve minimum ürün ve sepet tutarlarının 0 TL, maksimum ürün ve sepet tutarlarının 114.314,38 TL olduğu görülmektedir.

**Şekil 17: Veri kümesindeki sayısal değişkenlerin betimsel istatistikleri**

df.describe()		
	basketAmount	productAmount
<b>count</b>	72863.000000	72863.000000
<b>mean</b>	599.433931	338.604118
<b>std</b>	2069.979017	1612.866872
<b>min</b>	0.000000	0.000000
<b>25%</b>	77.000000	25.450000
<b>50%</b>	178.950000	71.020000
<b>75%</b>	414.515000	172.780000
<b>max</b>	114314.380000	114314.380000

Şekil 17’de görülebileceği üzere, ortalama sepet tutarı 599,43 TL iken, ortalama ürün tutarı ise 338,60 TL’dir. Sepet tutarında standart sapma 2.069,98 TL, ürün tutarında ise 1.612,87 TL olarak görülmektedir. Sepet tutarının birinci çeyreğinin 77 TL, ikinci çeyreğinin yani diğer bir ifadeyle medyanının 178,95 TL, üçüncü çeyreğinin ise 414,52 TL olduğu görülmektedir. Benzer şekilde, ürün tutarının ise birinci çeyreğinin 25,45 TL, ikinci çeyreğinin yani medyanının 71,02 TL ve üçüncü çeyreğinin 172,78 TL olduğu görülmektedir.

Şekil 18’de kullanıcı numarası (userId) değişkeni gruplandırılarak benzersiz sepet numarası (orderNumber) değerleri saydırılmış ve seçilen kullanıcılar arasından bir yıl içerisinde en çok sipariş veren kullanıcılar listelenmiştir. 2022 yılında rastgele seçilmiş olan kullanıcılar arasından en çok sipariş veren kullanıcının 214 adet siparişi vardır. Şekil 18’deki kodu sepet numarası (orderNumber) sadece 1 olacak şekilde filtrelediğinde ise yalnızca bir kez sipariş vermiş ve tekrar hiçbir sipariş vermemiş kullanıcı sayısı ise 1240 olarak karşımıza çıkmaktadır (Bkz. Ek 1).

### Şekil 18: Veri kümesindeki en çok sipariş veren kullanıcılar

```
temp = df.groupby(by=['userId'], as_index=False)['orderNumber'].nunique()  
temp.sort_values('orderNumber', ascending=False)
```

	userId	orderNumber
<b>4891</b>	8932083154555201028	214
<b>4134</b>	6096383531488901231	200
<b>4049</b>	5756178235851501224	193
<b>4762</b>	8411903090895221856	167
<b>1597</b>	-3387392014918255632	165
...	...	...
<b>3540</b>	4021241346941361819	1
<b>3537</b>	4009513232304776404	1
<b>288</b>	-8136459466196051341	1
<b>1714</b>	-2972262092339268389	1
<b>4962</b>	9219766512169354843	1

Şekil 19’da ise kullanıcı numarası (userId) ve sepet numarası (orderNumber) gruplandırılarak sipariş saatleri saydırılmış ve veri kümesindeki kullanıcılar arasından tek bir siparişte en çok ürün satın alan kullanıcılar listelenmiştir. İki farklı kullanıcının tek seferde 50 farklı ürün satın aldığı, üçüncü olarak ise bir kullanıcının tek seferde 41 adet ürün satın aldığı görülmektedir.

**Şekil 19: Veri kümesindeki tek bir siparişte en çok ürün satın alan kullanıcılar**

```
temp = df.groupby(by=['userId', 'orderNumber'], as_index=False)['receivedAt'].count()
product_cnt_per_basket = temp.rename(columns = {'receivedAt': 'productCnt'})
product_cnt_per_basket.sort_values('productCnt', ascending=False)
```

	userId	orderNumber	productCnt
<b>37648</b>	6974245716608969091	4843190789353525478	50
<b>33875</b>	5592608416177972138	1782207703985799107	50
<b>25749</b>	1909321156647226594	-8646454977037671575	41
<b>41655</b>	8580366716790331560	3130275092037129314	37
<b>4381</b>	-7377823875730449696	-7164073312091665009	37
...	...	...	...
<b>16870</b>	-1907438460568492980	-6775734169075984139	1
<b>16869</b>	-1907438460568492980	-7978477186061666457	1
<b>16865</b>	-1909316214901328364	4457992930150488692	1
<b>16864</b>	-1909316214901328364	3752116096664292299	1
<b>43002</b>	9219766512169354843	7043331984096424294	1

Veri kümesinin tanımlanması ve içerisindeki verilere ait bazı betimsel istatistiklerin hesaplanması sonucunda veri kümesinde aykırı değerlerin olduğu veya bir kategoride sadece 2 adet sipariş verilmiş olduğu görülmüştür. Bu tarz verilerin makine öğrenmesi modelinin eğitimine başlamadan önce temizlenmesi veya düzenlenmesi gerekmektedir. Buna ek olarak incelediğimiz ham veri kümesinin makine öğrenmesi modeli için düzenlenmesi ve bazı yeni değişkenler eklenmesi, makine öğrenmesi modelinin daha iyi sonuçlar vermesine olanak sağlayacaktır.

### 3.2.2. Veri Görselleştirme

Veri görselleştirme aşamasında daha anlamlı grafikler elde edebilmek için veri kümesinin bütünlüğünü bozan aykırı değerlerin olmadığı yeni bir veri kümesi oluşturulmuştur. Veri kümesinde boş değer olup olmadığı kontrol edilmiştir ve boş değer olmadığı

belirlenmiştir. Ancak sepet tutarı ve ürün tutarı 0 TL olan veriler bulunmaktadır. Aykırı değerler görselleştirme amacıyla veriden çıkarılmıştır, aykırı değerlerin manipüle edilmesi makine öğrenmesi modelinin tahminlerini etkileyebileceği modele beslenecek veri kümesinden çıkarılmamıştır. Aykırı değerlerin çıkarılmadığı bir grafik örneği Ek 2’de verilmiştir.

Şekil 20’de Stats kütüphanesi kullanılarak sepet tutarı ve ürün tutarı sayısal değişkenlerinin z skoru (standart z değişkeni) hesaplanmıştır. Z skorunun mutlak değeri 3’ten küçük olan veriler yeni bir değişkene aktarılarak aykırı değerler görselleştirmenin yapılacağı veri kümesinden çıkarılmıştır.

### Şekil 20: Aykırı değerlerin veri setinden çıkarılması

```
data = df[(np.abs(stats.zscore(df["productAmount"])) < 3)]
```

```
data = data[(np.abs(stats.zscore(data["basketAmount"])) < 3)]
```

```
data
```

	receivedAt	orderNumber	userId	sku	domain	basketAmount	productAmount
0	2022-01-01 05:27:37+00:00	-2207087542370126279	8910569964850783703	-1158940591953934668	Fashion & Lifestyle	80.91	80.91
1	2022-01-01 06:53:38+00:00	3463398509105352910	-5912979623820006624	-3189275990496335346	Fashion & Lifestyle	334.77	53.99
2	2022-01-01 06:53:38+00:00	3463398509105352910	-5912979623820006624	5653133865623348119	Fashion & Lifestyle	334.77	54.00
3	2022-01-01 06:53:38+00:00	3463398509105352910	-5912979623820006624	6653633759903391451	Fashion & Lifestyle	334.77	226.78
4	2022-01-01 07:44:02+00:00	-7258528727536827197	-2884111188620100873	2391104320713948662	Home & Garden	414.00	414.00
...	...	...	...	...	...	...	...
72858	2022-12-31 23:18:18+00:00	-6024306834204222552	-496073675151649235	710583592107002819	Books & Hobbies	50.00	50.00
72859	2022-12-31 23:27:36+00:00	-4878815539775359249	-496073675151649235	7345114099369013420	Books & Hobbies	23.00	22.00
72860	2022-12-31 23:27:36+00:00	-4878815539775359249	-496073675151649235	710583592107002819	Books & Hobbies	23.00	1.00
72861	2022-12-31 23:27:54+00:00	166553617047767314	4828818243572142253	-2211919226393502800	Fashion & Lifestyle	229.99	229.99
72862	2022-12-31 23:30:39+00:00	7513772089685661913	-496073675151649235	710583592107002819	Books & Hobbies	4.00	4.00

71107 rows x 7 columns

Şekil 20’de görülebileceği üzere aykırı değerler temizlendikten sonra veri setinde 71.107 adet veri kaldığı görülmektedir. Diğer bir ifadeyle 1.756 adet aykırı değer veri kümesinden silinmiştir. Şekil 21’de aykırı değerler çıkarılmadan önceki ve sonraki skewness ve kurtosis değerleri görülebilir.

## Şekil 21: Aykırı değerlerin çıkarılmasından önce ve sonra Skewness ve Kurtosis değerleri

aykırı değerler çıkarılmadan önce skewness ve kurtosis değerleri

```
print("ProductAmount Skewness: %f" % df['productAmount'].skew())
print("ProductAmount Kurtosis: %f" % df['productAmount'].kurt())

print("basketAmount Skewness: %f" % df['basketAmount'].skew())
print("basketAmount Kurtosis: %f" % df['basketAmount'].kurt())
```

```
ProductAmount Skewness: 16.726797
ProductAmount Kurtosis: 557.565278
basketAmount Skewness: 13.078179
basketAmount Kurtosis: 309.482217
```

aykırı değerler çıkarıldıktan sonra skewness ve kurtosis değerleri

```
print("ProductAmount Skewness: %f" % data['productAmount'].skew())
print("ProductAmount Kurtosis: %f" % data['productAmount'].kurt())

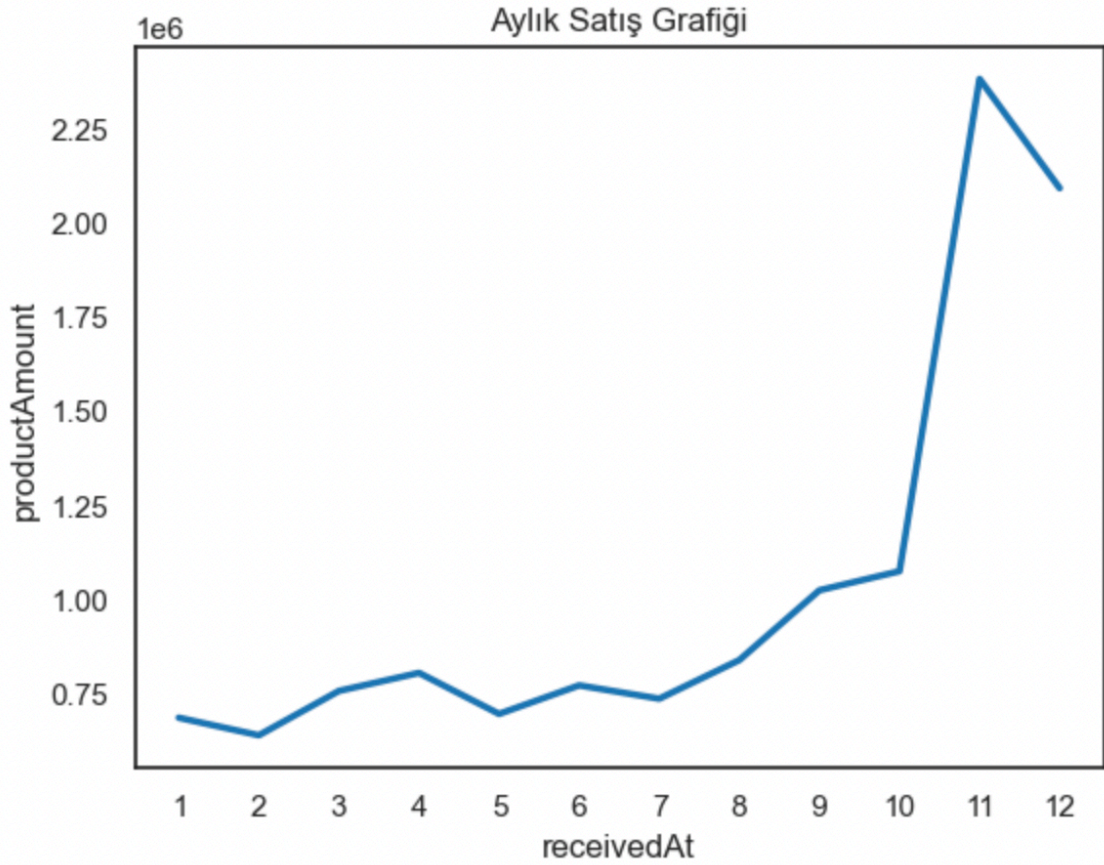
print("basketAmount Skewness: %f" % data['basketAmount'].skew())
print("basketAmount Kurtosis: %f" % data['basketAmount'].kurt())
```

```
ProductAmount Skewness: 5.740578
ProductAmount Kurtosis: 40.201348
basketAmount Skewness: 3.557520
basketAmount Kurtosis: 15.035434
```

Şekilde 21’te görülebileceği üzere, aykırı değerler çıkarılmadan önce ürün tutarı (productAmount) değişkeninin skewness değeri 16,73 ve kurtosis değeri ise 557,56’dır. Aykırı değerler çıkarılmadan önce sepet tutarı (basketAmount) değişkeninin ise skewness değeri 13,08 ve kurtosis değerinin 309,48 olduğu görülmektedir. Python dilinde örneklem skewness ve kurtosis değerleri kullanıldığı hesaplanmıştır, dolayısıyla 0’a göre kıyaslanmalıdır (Işığışık, 2022). Diğer bir ifadeyle, veri kümesinin aşırı sağa çarpık ve aşırı sivri bir dağılıma sahip olduğu söylenebilir. Aykırı değerler çıkarıldıktan sonra ürün tutarı (productAmount) değişkeninin skewness değeri 5,74’e ve kurtosis değeri ise 40,20’ye düştüğü görülmektedir. Ürün tutarı (basketAmount) değişkeni için ise skewness değeri 3,56 ve kurtosis değeri 15,04 olacak şekilde değişmiştir. Makine öğrenmesi modeline beslenecek veri kümesindeki verilerin, modele beslenmesinden önce standardize edilmesi gerektiği görülebilmektedir.

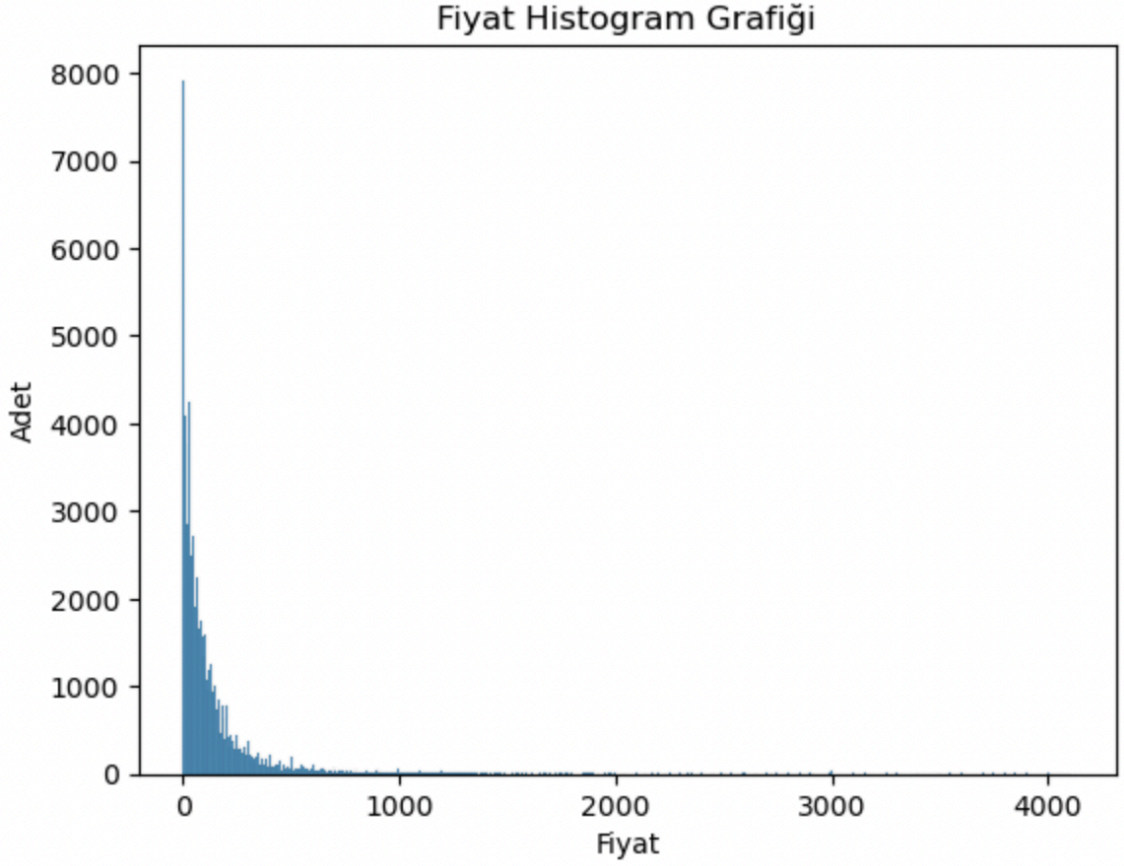
Şekil 22’de aylara göre satılan ürünlerin tutarları toplanarak aylık gelir verilmiştir. Grafikte yıl sonuna doğru satışlarda bir artış trendi olduğu gözlemlenebilmektedir. E-ticaret şirketlerinin Kasım ayında yaptığı indirimler sebebiyle Kasım ve Aralık aylarında yılın diğer aylarına göre satışlardaki artış net bir şekilde görülmektedir.

**Şekil 22: Aylık Satış Grafiği**



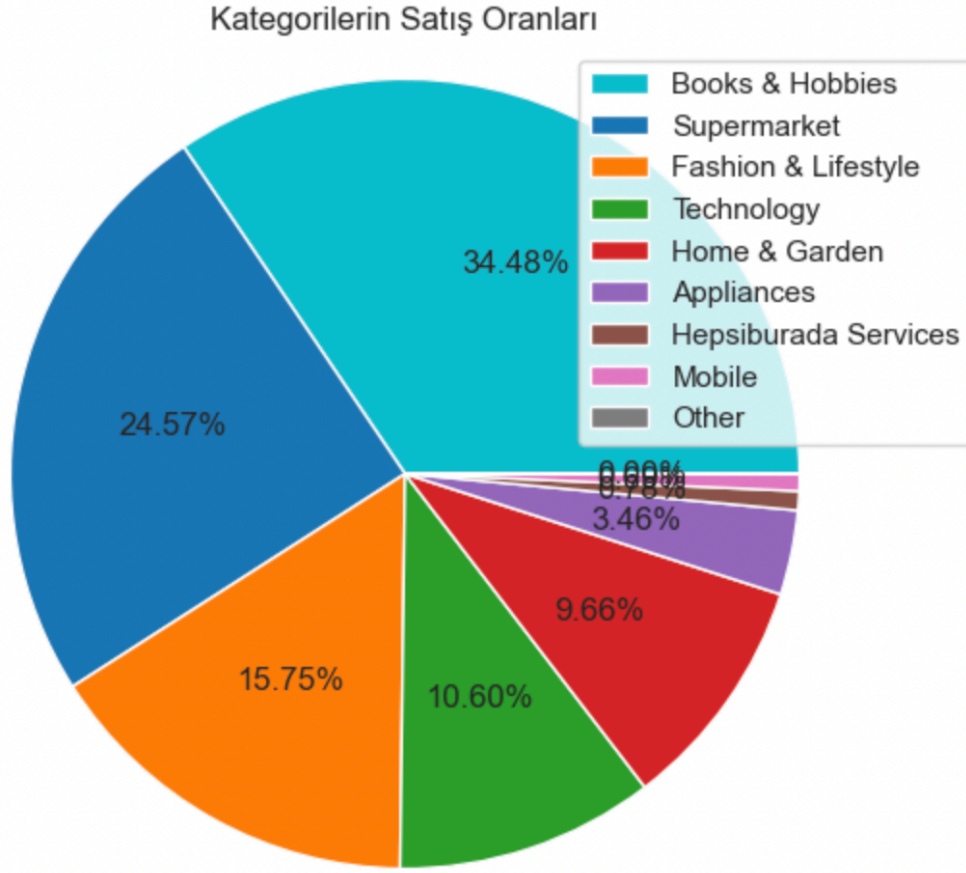


**Şekil 23: Fiyat Histogram Grafiği**



Şekil 23'teki Fiyat Histogram Grafiği incelendiğinde, satın alınan ürünlerin fiyatı (productAmount) değişkenindeki sağa çarpıklık ve sivrilik net şekilde görülmektedir. Grafikte de görülebileceği üzere, satılan ürünlerin çoğu 0-1000 TL aralığındadır. Şekil 24'e bakıldığında, en çok satış yapılan ürünlerin kitap ve hobi (books & hobbies) ile süpermarket (supermarket) kategorilerindeki ürünler olması da bu durumu doğrulamaktadır. Bu iki grafik, kullanıcıların e-ticaret siteleri üzerinden yaptıkları alışveriş alışkanlıklarını özetlemektedir.

Şekil 24: Kategorilerin Satış Oranları Grafiği



Şekil 24’de Kategorilerin Satış Oranları Grafiği incelendiğinde ise, en çok satış yapılan kategorinin %34,48 ile kitap ve hobi (books & hobbies) kategorisinin olduğu ve onu %24,57 ile süpermarket (supermarket) kategorisinin izlediği görülebilmektedir. Ardından moda ve yaşam tarzı (fashion & lifestyle), teknoloji (technology), ev ve bahçe (home & garden), ev aletleri (appliances), Hepsiburada hizmetleri (Hepsiburada services), mobil ve diğer (other) kategorilerinin geldiği görülmektedir.

### 3.3. VERİ ÖNİŞLEME

Uygulamada bazı betimsel istatistik ve veri görselleştirme yöntemleriyle veri kümesi hakkında daha detaylı bilgi sunulmasının ardından, makine öğrenmesi modeli için gerekli veri dönüşüm işlemlerine başlanmıştır. Müşteri segmentasyonunun yapılabilmesi için veri kümesindeki her satır bir kullanıcının tüm harcamalarının bir özeti olacak şekilde

düzenlenmiştir. Diğer bir ifadeyle, veri kümesindeki her bir satır ilgili kullanıcının çeşitli kategorilerdeki toplam harcamalarını içermektedir. Buna ek olarak, kullanıcıların toplamda kaç kez alışveriş yaptığı, minimum, maksimum, ortalama ve toplam harcamaları da modelin eğitilmesinde yararlı bilgiler olarak görüldüğü için veri kümesine eklenmiştir. Bu dönüşümün sonucunda makine öğrenmesi modelinin eğitilmesinde bir katkısı olmayacak veriler veri kümesinden çıkarılmıştır. Örneğin, her kullanıcıya özel ayırt edici bir değişken olan kullanıcı numarası (userId) değişkeni, modelin eğitilmesi sırasında bir katkı sağlamayacağı için veri kümesinden çıkarılmıştır. Buna benzer şekilde, sadece 2 adet ürün içeren diğer (other) kategorisi, pivot edilme işlemi sonrasında bir değişken haline gelmiştir ve boyut azaltmak için tablodan çıkarılmıştır. Diğer bir ifadeyle, veri temizleme işlemi sonrasında sadece makine öğrenmesi modeli için yararlı olan verilerden oluşan bir veri kümesi elde edilmiştir.

Her bir kullanıcının hangi kategoride toplamda ne kadarlık harcama yapabildiğini hesaplayabilmek için veri kümesinin olduğu tabloyu özetlememiz (pivot table) gerekmektedir. Şekil 25'te kullanıcıların aldığı ürünlerin fiyatları toplanarak ait olduğu kategoriye yazdırılmıştır.

### Şekil 25: Kullanıcıların kategorilere ait harcamalarının hesaplanması

```
pivoted = pd.pivot_table(data,
                        index='userId',
                        columns = 'domain',
                        values = 'productAmount',
                        aggfunc = 'sum').reset_index()
```

pivoted

domain	userId	Appliances	Books & Hobbies	Fashion & Lifestyle	Hepsiburada Services	Home & Garden	Mobile	Other	Supermarket	Technology
0	-9219215372597658510	0.00	576.82	818.99	0.0	49.99	0.0	0.0	796.79	214.82
1	-9213536368735428322	0.00	0.00	469.20	0.0	0.00	0.0	0.0	243.89	0.00
2	-9212213781084510606	0.00	955.19	802.66	0.0	0.00	0.0	0.0	1167.09	3469.55
3	-9212143531100464576	0.00	0.00	581.42	0.0	0.00	0.0	0.0	0.00	0.00
4	-9205248175333532922	4766.12	19833.71	11871.84	0.0	5182.68	0.0	0.0	11388.17	3167.15
...	...	...	...	...	...	...	...	...	...	...
4900	9197639350600673735	0.00	0.00	0.00	0.0	0.00	0.0	0.0	0.00	171.99
4901	9198648949746904109	0.00	382.62	0.00	0.0	256.29	0.0	0.0	759.80	103.79
4902	9199769750491676274	56.11	0.00	0.00	9.9	993.87	0.0	0.0	801.95	290.98
4903	9208887071934344602	0.00	0.00	0.00	0.0	0.00	0.0	0.0	0.00	77.70
4904	9219766512169354843	0.00	0.00	0.00	0.0	0.00	0.0	0.0	0.00	99.99

4905 rows x 10 columns

Şekil 25’te görüleceği üzere; ilk kullanıcı kitap ve hobi (books & hobbies) kategorisinde 576,82 TL, moda ve yaşam tarzı (fashion & lifestyle) kategorisinde 818,99 TL, ev ve bahçe (home & garden) kategorisinde 49,99 TL, süpermarket (supermarket) kategorisinde 796,79 TL ve teknoloji (technology) kategorisinde 214,82 TL harcama yapmış iken diğer kategorilerden herhangi bir ürün satın almamıştır.

Kullanıcıların minimum, maksimum ve ortalama sepet tutarları, toplamda kaç kez sipariş verdiği ve toplamda kaç TL değerinde ürün aldığı Şekil 26’da hesaplanmıştır. Kullanıcının yaptığı en düşük ve en yüksek harcama, minimum ve maksimum sepet tutarı (basketAmount) değerinden alınmıştır. Kullanıcının toplam yaptığı harcama sayısı ise sepet tutarı (basketAmount) değerlerinin benzersiz saydırılmasından elde edilmiştir. Kullanıcının yaptığı toplam harcama ise sepet tutarı (basketAmount) değeri tekrar eden bir değer olduğu için ürün tutarı (productAmount) üzerinden hesaplanmıştır. Son olarak; kullanıcının toplam harcaması, toplam sipariş sayısına bölünerek kullanıcının ortalama sipariş tutarı hesaplanmıştır.

### Şekil 26: Kullanıcıların harcama istatistiklerinin hesaplanması

```
agg_functions = {'basketAmount': ['min', 'max', 'nunique'],
                 'productAmount': ['sum']}
df_agg = data.groupby('userId').agg(agg_functions).reset_index()
```

```
df_agg.columns = ['userId', 'min', 'max', 'cnt', 'sum']
```

```
df_agg['mean'] = df_agg['sum']/df_agg['cnt']
```

```
df_agg
```

	userId	min	max	cnt	sum	mean
0	-9219215372597658510	49.99	338.94	14	2457.41	175.529286
1	-9213536368735428322	118.99	469.20	3	713.09	237.696667
2	-9212213781084510606	73.00	1888.69	12	6394.49	532.874167
3	-9212143531100464576	581.42	581.42	1	581.42	581.420000
4	-9205248175333532922	4.00	3925.64	110	56209.67	510.997000
...	...	...	...	...	...	...
4900	9197639350600673735	171.99	171.99	1	171.99	171.990000
4901	9198648949746904109	11.44	278.09	14	1502.50	107.321429
4902	9199769750491676274	9.90	462.34	13	2152.81	165.600769
4903	9208887071934344602	77.70	77.70	1	77.70	77.700000
4904	9219766512169354843	99.99	99.99	1	99.99	99.990000

Şekil 25’te hesaplanan kullanıcıların kategorilere ait harcamalarının hesaplanması tablosu ile Şekil 26’da hesaplanan kullanıcıların harcama istatistiklerinin hesaplanması tablosu birleştirilerek yeni bir tablo elde edilmiştir. Elde edilen bu tablodan kullanıcı numarası (userId) ve diğer (other) değişkenlerinin çıkarılmasının ardından, Şekil 27’de görüleceği üzere makine öğrenmesi modelinin eğitilebileceği 4905 satırdan ve 14 değişkenden oluşan bir tablo oluşturulmuştur.

**Şekil 27: Veri dönüştürme işlemi sonrası final tablo**

	Appliances	Books & Hobbies	Fashion & Lifestyle	Hepsiburada Services	Home & Garden	Mobile	Supermarket	Technology	min	max	cnt	sum	mean
0	0.00	576.82	818.99	0.0	49.99	0.0	796.79	214.82	49.99	338.94	14	2457.41	175.529286
1	0.00	0.00	469.20	0.0	0.00	0.0	243.89	0.00	118.99	469.20	3	713.09	237.696667
2	0.00	955.19	802.66	0.0	0.00	0.0	1167.09	3469.55	73.00	1888.69	12	6394.49	532.874167
3	0.00	0.00	581.42	0.0	0.00	0.0	0.00	0.00	581.42	581.42	1	581.42	581.420000
4	4766.12	19833.71	11871.84	0.0	5182.68	0.0	11388.17	3167.15	4.00	3925.64	110	56209.67	510.997000
...	...	...	...	...	...	...	...	...	...	...	...	...	...
4900	0.00	0.00	0.00	0.0	0.00	0.0	0.00	171.99	171.99	171.99	1	171.99	171.990000
4901	0.00	382.62	0.00	0.0	256.29	0.0	759.80	103.79	11.44	278.09	14	1502.50	107.321429
4902	56.11	0.00	0.00	9.9	993.87	0.0	801.95	290.98	9.90	462.34	13	2152.81	165.600769
4903	0.00	0.00	0.00	0.0	0.00	0.0	0.00	77.70	77.70	77.70	1	77.70	77.700000
4904	0.00	0.00	0.00	0.0	0.00	0.0	0.00	99.99	99.99	99.99	1	99.99	99.990000

4905 rows x 14 columns

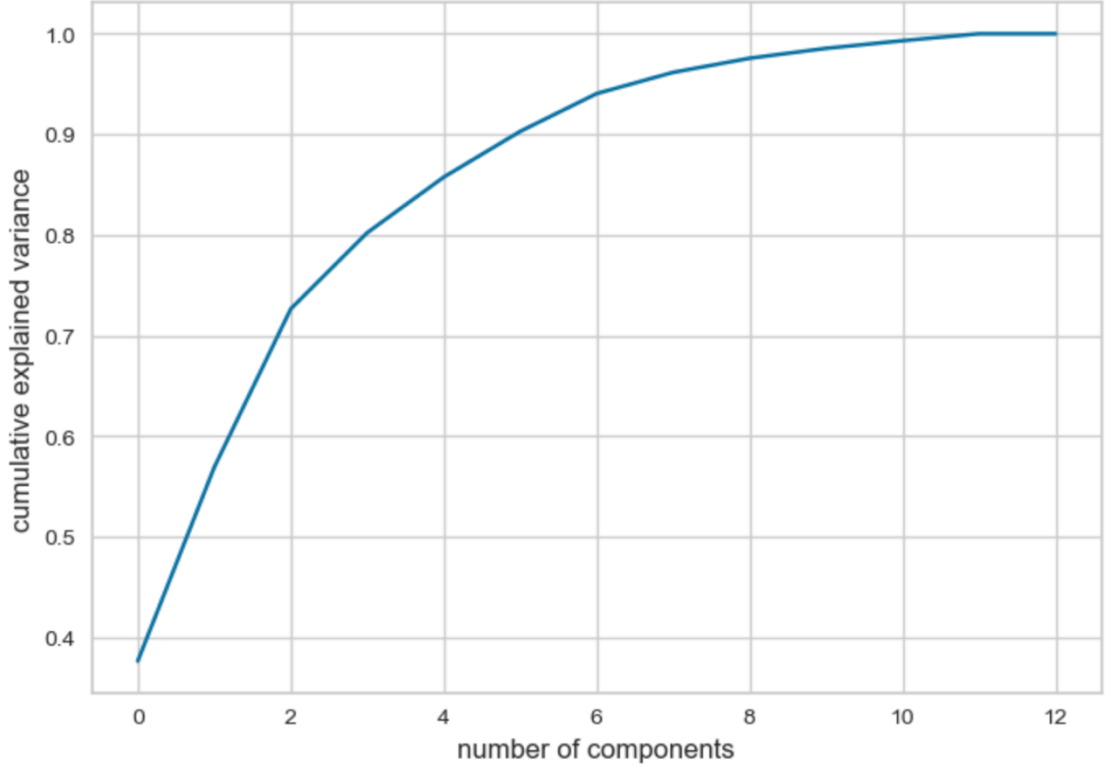
### 3.4. MAKİNE ÖĞRENMESİ

Bu bölümde yer alan 3.2. Veri Kümesinin Tanımlanması ve 3.2.1. Veri Görselleştirme başlıklarında görülebileceği üzere veri kümesi hem aşırı çarpık hem de aşırı sivri bir dağılıma sahiptir. Aynı zamanda değişkenlerin bazıları çok yüksek değerlere sahip iken bazıları çok düşük değerlere sahiptir. Bu nedenler doğrultusunda, Makine öğrenmesi modelinin eğitilmesi öncesinde veri kümesinde ölçeklendirme işlemi uygulanmıştır. Bu amaçla kullanılan MinMaxScaler fonksiyonu, verileri sıfır ile bir arasında olacak şekilde ölçeklendirir.

Veri kümesinin boyutu büyüdükçe makine öğrenmesi modelleri çok daha fazla kaynak tüketmektedir. Temel bileşenler analizi (Principal Component Analysis, PCA) boyut indirgemeyi sağlayan bir istatistiksel tekniktir. Verilerdeki genel özellikleri koruyarak çok boyutlu bir veri kümesinin boyut azaltarak daha az sayıda değişkenle temsil

edilmesini sağlar. Şekil 28’de PCA dönüşümü yapıldığında kaç adet boyutla toplam varyansın ne kadarını sağlayabileceğini göstermektedir.

**Şekil 28: Temel Bileşen Analizi ile bileşen sayısının seçilmesi**

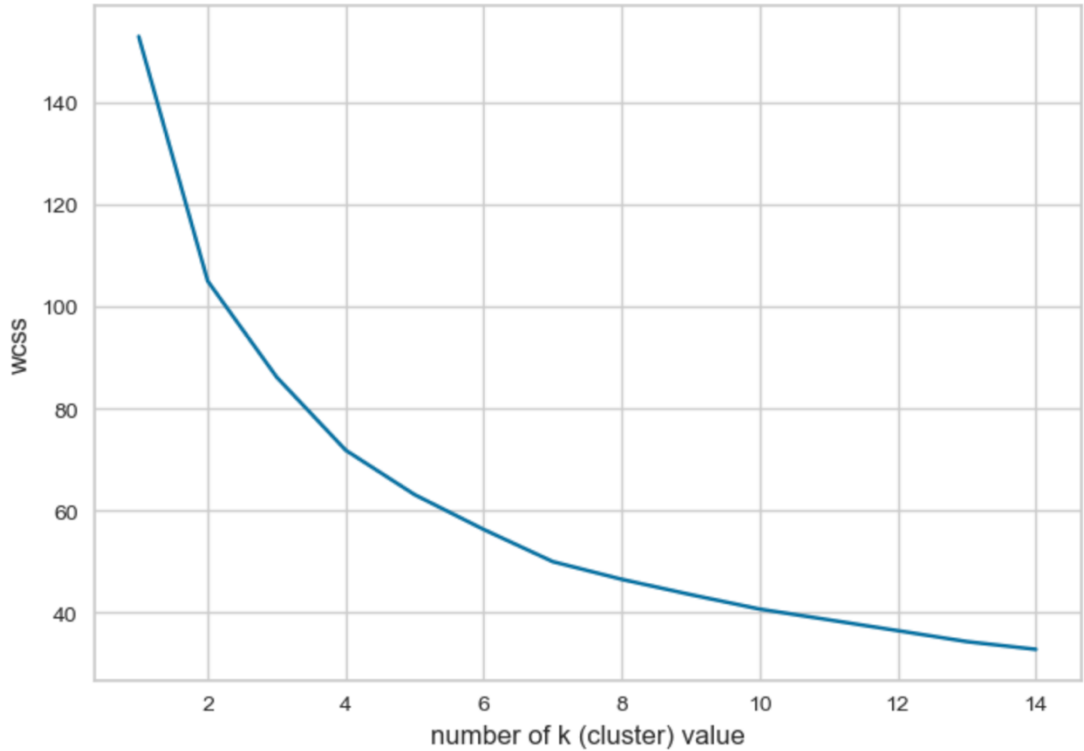


Şekil 28’de x eksenini PCA ile veri kümesindeki bileşen sayısının düşürülebileceği olası değerleri, y eksenini ise denk gelen bileşen sayısı ile ne kadarlık bir varyans yakalanabileceğini göstermektedir. Şekilde 7 bileşen ile verilerin toplam varyansının yaklaşık %94’ünü koruyabileceği görülmektedir. Daha fazla bileşen olması varyansı artırmaktadır fakat etkisi giderek azalmaktadır. Bu nedenle veri kümesini 7 ana bileşen ile temsil edecek şekilde PCA uygulanmıştır.

K-ortalamlar algoritmasının en önemli parametresi veri kümesinin bölüneceği küme sayısıdır. Makine öğrenmesi modeli tanımlanırken verinin kaç kümeye ayrılacağı belirtilmelidir. Bu nedenle, temel bileşen analizi gerçekleştirildikten sonra optimum küme sayısını belirleyebilmek için Dirsek yöntemi (elbow method) uygulanabilir. Şekil 29’da Dirsek yöntemi (elbow method) uygulandığında oluşan grafik verilmiştir.

## Şekil 29: Elbow yönteminin uygulanması

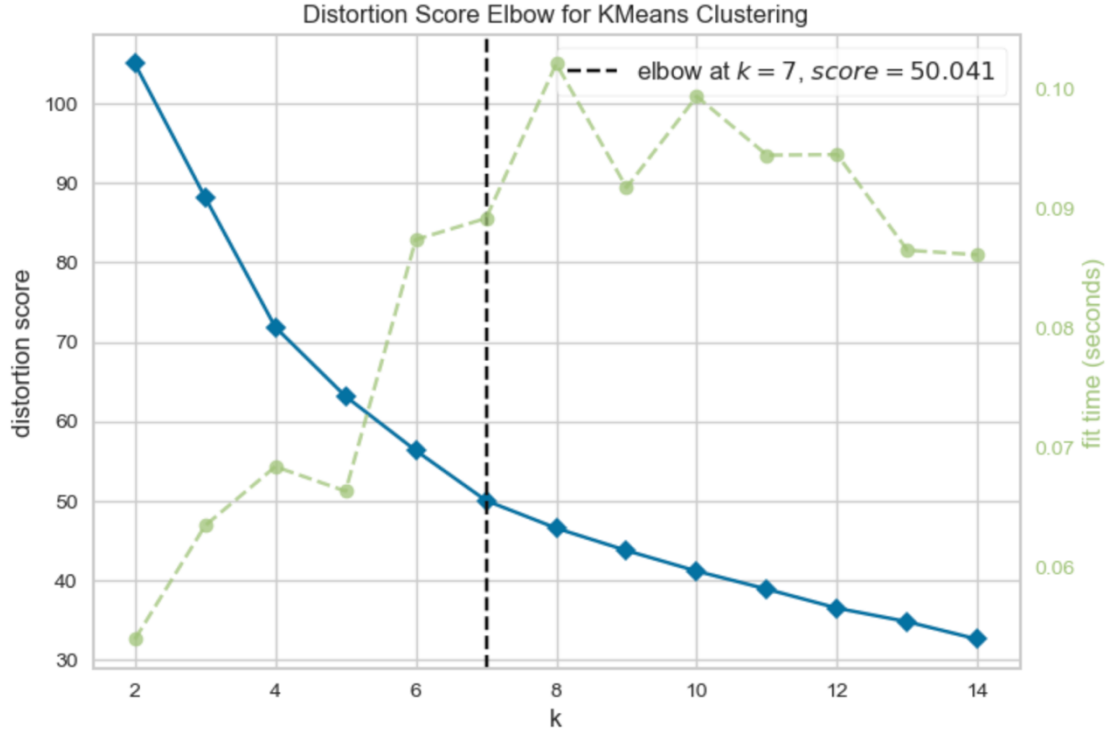
```
: wcss = []  
  
: for k in range (1,15):  
    kmeans = KMeans(n_clusters=k)  
    kmeans.fit(x_pca)  
    wcss.append(kmeans.inertia_)  
  
: plt.plot(range(1,15),wcss)  
plt.xlabel("number of k (cluster) value")  
plt.ylabel("wcss")  
plt.show()
```



Elbow yönteminin uygulanması ile grafik üzerinde bir dirsek noktası bulunması hedeflenmektedir. Şekil 29'da görülebileceği üzere grafikte net bir dirsek noktası bulunmamasıyla birlikte 7 kümeye kadar kümeler içi kareler toplamının (WCSS) sert bir şekilde düştüğü, fakat sonrasında düşüşün azaldığı görülebilir. Grafikte net bir dirsek noktası görülemediği için Yellowbrick kütüphanesinin `kElbowVisualizer` metodu kullanılmıştır. `kElbowVisualizer` şekil 29'a ek olarak makine öğrenmesi modelinin eğitilme süresini de grafiğe ekleyerek, en optimal küme sayısını düz bir çizgi ile gösterir. Şekil 30'da `kElbowVisualizer` grafiği verilmiştir.

Şekil 30: kElbowVisualizer metodu ile Elbow yönteminin uygulanması

```
kmodel=KMeans(random_state=42)
elb_visualizer=KElbowVisualizer(kmodel,k=(2,15))
elb_visualizer.fit(x_pca)
elb_visualizer.show()
```



Şekil 29 ve Şekil 30'daki grafiklerden dirsek yöntemine göre, K-ortalamlar algoritması ile veri setinin en optimal yedi kümeye ayrılacağı görülmektedir.

Şekil 31: K-ortalamlar algoritmasının uygulanması

```
kmeans = KMeans(init='k-means++', n_clusters = 7, random_state=42)
kmeans.fit(x_pca)
y = kmeans.predict(x_pca)
silhouette_avg = silhouette_score(x_pca, y)
print('Silhouette score: {:.3f}'.format(silhouette_avg))
```

Silhouette score: 0.584

Şekil 31'de 7 temel bileşenden oluşan veri kümesini 7 kümeye ayıracak k-ortalamlar algoritması çalıştırılmıştır. Kümeleme işlemi sonucunda oluşan kümelerin performansını değerlendirmek için Silhouette skora yöntemi kullanılmıştır. Diğer bir ifadeyle, Silhouette skora yöntemi ile kümelerin birbirinden ne kadar iyi ayrıldığı hesaplanmıştır. Silhouette skoru -1 ile 1 arasında değişmektedir. Şekil 31'de



görülebileceği üzere modelin Silhouette skoru 0,584'tür. Bu bulguya göre, veriler arasında makul derecede anlamlı bir ilişki bulunarak kümelere ayrıldığı söylenebilir.

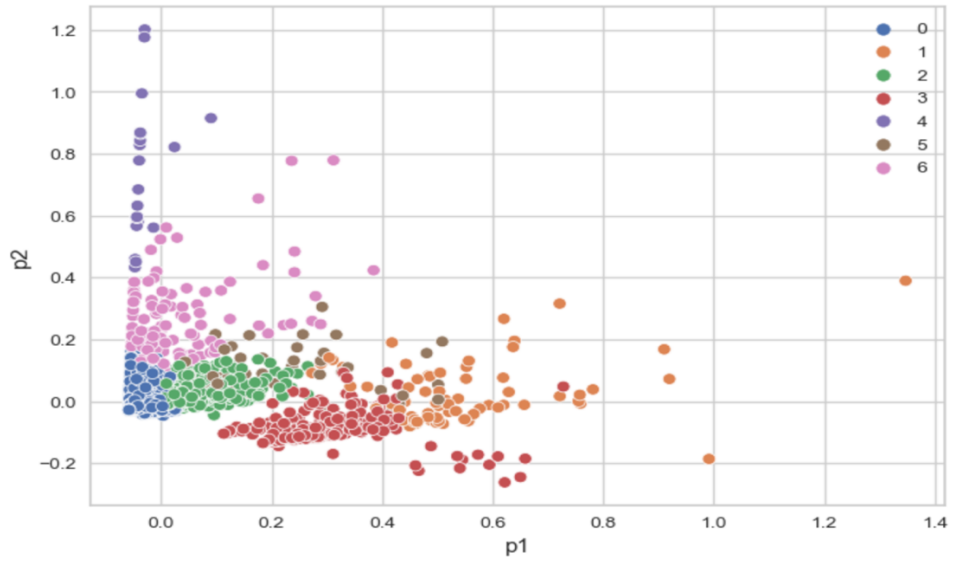
Segmentlere ayrılan veri kümesi, 7 temel bileşenden oluşmaktadır. 7 boyutlu bir veriyi görselleştirebilmek mümkün değildir. Kümelere ayrılan veriyi görselleştirebilmek için tekrar PCA yöntemi kullanılarak kümelendirilen veri, 2 temel bileşene düşürülmüştür. 2 temel bileşene düşen veri kümesine, K-ortalamlar algoritması ile oluşturulan kümeler eklenerek Şekil 32'de grafik elde edilmiştir.

**Şekil 32: Kümelendirilen veri setinin PCA yöntemi ile görselleştirilmesi**

```
pca = PCA(n_components=2)
pca_2d = pca.fit_transform(x_scaled)
pca_df = pd.DataFrame(pca_2d)
pca_df['cluster'] = pd.Series(y)

pca_df.set_axis(['p1', 'p2', 'cluster'], axis='columns', inplace=True)

sns.scatterplot(data=pca_df, x="p1", y="p2", hue="cluster", palette="deep")
plt.legend()
plt.xlabel("p1")
plt.ylabel("p2")
plt.show()
```



Şekil 32'de veri kümesinde iki boyutlu düzlemde dağılan veriler ve oluşturulan kümeler renklendirilerek gösterilmiştir.

Verilerin kümeler arası dağılımına bakıldığında, ilk kümede 3910 kullanıcı, ikinci kümede 68 kullanıcı, üçüncü kümede 514 kullanıcı, dördüncü kümede 266 kullanıcı,

beşinci kümede 22 kullanıcı, altıncı kümede 51 kullanıcı, yedinci kümede ise 132 kullanıcı bulunduğu görülebilir (Bkz. Ek 4).

Oluşturulan kümelerin özelliklerini görebilmek için veri kümesi segmentlere göre gruplandırılarak, makine öğrenmesinin eğitilmesinde kullanılan değişkenlerin ortalamaları alınmıştır. Şekil 33'te kümelere göre gruplanan değişkenlerin ortalamaları verilmiştir.

### Şekil 33: Kümelere göre gruplanan değişkenlerin ortalamaları

cluster	Appliances	Books & Hobbies	Fashion & Lifestyle	Hepsiburada Services	Home & Garden	Mobile	Supermarket	Technology	min	max	cnt
0	458.659645	245.026207	240.726212	0.032737	126.839046	221.906090	203.417882	336.940335	213.744494	1096.043179	3.913043
1	5807.736176	5734.572647	7105.994412	7.961765	2542.678382	2758.517353	4865.086912	4904.984265	8.117500	7244.365000	76.985294
2	2022.553074	1688.715117	1079.191712	0.124514	726.167393	1218.152646	1647.763716	1736.851401	48.088696	3450.509416	21.243191
3	1868.074286	1170.574774	805.922218	10.141353	461.573045	1688.442068	847.431729	1861.804737	7.115150	3539.778496	18.736842
4	587.607727	1142.548636	5.314091	0.000000	0.000000	16583.362727	0.000000	7167.307727	19845.116364	22210.299091	1.136364
5	6194.547647	1061.449804	1643.126471	1.400000	8489.810784	1760.070980	1557.206863	1632.044902	67.236078	6873.130000	25.509804
6	8873.599621	1702.024318	2882.856061	0.120455	278.809470	13480.976364	298.694545	17715.958864	1538.931591	21512.004394	8.568182

Şekil 34'te kümelerin değişkenler bazında ortalama harcamaları verilmiştir.

**Şekil 34: Kümelerin değişkenler bazında ortalama harcamaları**



Şekil 34'e bakıldığında ev aletleri (appliances) kategorisinde en yüksek harcamayı altıncı kümenin yaptığı ve ardından beşinci ve birinci kümelerin geldiği görülmektedir. Yapılan toplam harcamalarda da (sum) en yüksek harcamayı altıncı kümenin yaptığı, ardından birinci, dördüncü ve beşinci kümenin geldiği görülebilir.

Kitap ve hobi (books & hobbies), moda ve yaşam (fashion & lifestyle) ve süpermarket (supermarket) kategorilerinde açık ara farkla en yüksek harcamayı birinci kümenin yaptığı görülmektedir. Yapılan toplam alışveriş adedine (cnt) bakıldığında da aynı şekilde birinci kümenin ortalama olarak en çok alışveriş yapan küme olduğu görülebilir.

Hepsiburada hizmetleri (Hepsiburada services) kategorisinde en yüksek harcamayı üçüncü kümenin yaptığı ve ardından birinci kümenin geldiği görülmektedir. Ancak burada Şekil 33'te ilgili değişkene veya Şekil 34'teki grafiğin y eksenine bakılacak olursa bu kategorideki ortalama harcamaların 0-10 TL arasında olduğu görülebilir.

Ev ve bahçe (home & garden) kategorisinde en yüksek harcamayı beşinci kümenin yaptığı görülebilir.

Mobil (mobile) kategorisinde en yüksek harcamayı yapan kümenin dördüncü küme olduğu ve ardından altıncı kümenin geldiği görülebilir. Ayrıca minimum siparişlerin (min) ortalaması en yüksek değere sahip olanı dördüncü kümedir. Diğer bir ifadeyle dördüncü kümenin genellikle ucuz ürünler almadığı yüksek harcamalar yaptığı söylenebilir. Yapılan harcamaların ortalamalarına (mean) ve en yüksek harcamaların ortalamalarına (max) bakıldığında da dördüncü grubun en yüksek ortalamaya sahip olduğu görülebilir. En yüksek harcamalarda (max) dördüncü kümeyi altıncı küme takip etmektedir. Altıncı küme aynı zamanda teknoloji (technology) alanında en yüksek harcamayı yapan küme olarak görülmektedir.

Oluşturulan kümelerin ortalamalarından yola çıkarak, kümeler karakteristik özelliklerini yansıtabilecek şekilde isimlendirilebilir. Örneğin, neredeyse her kategoride en düşük harcamayı yapan sıfır numaralı küme “potansiyel kayıp” olarak adlandırılabilir. Kitap ve hobi (books & hobbies), moda ve yaşam (fashion & lifestyle) ve süpermarket (supermarket) kategorilerinde en çok harcamayı yapan birinci küme ise “hızlı tüketiciler” olarak isimlendirilebilir. İkinci kümenin ise neredeyse her kategoride benzer düzeyde ve ortalamanın altında alışveriş harcaması yaptığı görülmektedir. İkinci kümedeki müşterilere “dengeli tüketiciler” denilebilir. Üçüncü kümeye baktığımızda ise en çok harcamanın Hepsiburada hizmetleri (Hepsiburada services) kategorisinde yapıldığı görülmektedir. Dolayısıyla, üçüncü kümedeki kullanıcıların Hepsiburada’nın sadakat programı olan “Premium kullanıcıları” olduğu söylenebilir. Dördüncü küme ise en çok harcamayı mobil (mobile) kategorisinde yaptığı için bu küme “cep telefonu meraklıları” olarak isimlendirilebilir. Beşinci kümenin, en çok harcamayı ev ve bahçe (home & garden), ardından ev aletleri (appliances) kategorilerinde yaptığı görülmektedir. Beşinci kümede bulunan müşterilere “evini yenileyenler” denilebilir. Teknoloji (technology) ve ev aletleri (appliances) kategorilerinde en çok harcamayı yapan ve mobil (mobile) kategorisinde de en yüksek ikinci harcamaya sahip olan altıncı kümeye “teknoloji tutkunları” denilebilir.

Pazarlama bakış açısıyla bakıldığında, “cep telefonu meraklıları” ve “teknoloji tutkunları” kümelerine mobil ve teknoloji kategorisine ait yüksek fiyatlı ürünler önerilerek daha fazla kar elde edilebileceği öngörülebilir. “Hızlı tüketiciler” kümesinde bulunan müşterilere ise ev aletleri, kitap ve hobi, moda ve yaşam tarzı ve süpermarket ürünleri sunarak en çok alışveriş yapan birinci küme daha sık alışveriş yapmaya teşvik edilebilir. “Potansiyel kayıp” kümesinin müşterilerinin ise neredeyse tüm kategorilerde en düşük harcamayı yaptığı ve toplam getiri bakımından da en düşük getiriye sahip olan küme olduğu görülebilir. “Potansiyel kayıp” kümesine ait kişiler platforma yeni katılmış veya platformu terk etmiş kullanıcılar olabilir. 3910 müşteriyi içeren bu kümeye özel indirimler ve fırsatlar sunarak uzun vadede sadakatleri kazanılmaya çalışılabilir. Bu çalışmada kullanılan veri kümesindeki müşterilerin büyük bir çoğunluğu az harcama yapan “potansiyel kayıp” kümesinde olduğu için, bu kümeye daha yüksek fiyatlı ürünlerin satışı da teşvik edilmelidir. “Dengeli tüketiciler” kümesi de aynı şekilde daha yüksek fiyatlı ürünlerin satışına teşvik edilmeli ve bu kümeye özel indirimler ve fırsatlar sunulmalıdır. “Evini yenileyenler” kümesindeki müşterilere de aynı kümedeki diğer müşterilerin aldığı ürünler önerilerek ve kategorinin öne çıkan ürünleri ile daha fazla satış yapmak amaçlanabilir. Son olarak “Premium kullanıcıları” kümesine özel daha fazla avantajlar sağlanarak sadakat programına devam etmeleri ve diğer kategorilerde yapılan harcamaların artırılması sağlanmalıdır.

## SONUÇ

Bu tez çalışmasında, Hepsiburada e-ticaret platformu müşterilerinin alışveriş verileri üzerinde bir müşteri segmentasyonu uygulaması gerçekleştirilmiştir. Birinci bölümde e-ticaret ve müşteri ilişkileri yönetimi (CRM) hakkında bilgiler verilmiştir. İkinci bölümde ise genel olarak kapsayıcı makine öğrenmesi kategorileri tanımlanmış ve ardından üçüncü bölümde uygulama aşamasında kullanılacak algoritmalar anlatılmıştır. Üçüncü bölümde ise veri kümesinin tanımlayarak bazı betimsel istatistikler ve görseller verilmiştir. Ardından bazı veri ön işleme yöntemleri ile veri kümesi makine öğrenmesi algoritmaları ile kümeleme yapılabilmesi için hazır hale getirilmiştir.

Uygulama aşamasında bir denetimsiz öğrenme algoritması olan ve ikinci bölümde de tanımlanan temel bileşenler analizi algoritması ile veri kümesinde boyut indirgeme yöntemi uygulanmıştır. Temel bileşenler analizi işleminin uygulanması sonucunda veri kümesinin 14 değişken yerine, toplam varyansının %94'ünü koruyarak 7 temel değişken ile temsil edilebileceği gözlemlenmiştir.

Veri kümesinin kaç kümeye ayrılarak en iyi şekilde temsil edilebileceği dirsek yöntemi (elbow method) ile belirlenmiştir. Veri kümesi en az 7'ye kümeden (segment) oluşacak şekilde ayrılırsa kümeler içi kareler toplamının en düşük seviyelere düşeceği gözlemlenmiştir. 7 kümeden fazlasında ise kümeler içi kareler toplamının düşmeye devam edeceği fakat bu düşüşün azalara devam ettiği ve modelin işlem süresinin uzayacağı kElbowVisualizer kütüphanesi ile gösterilmiştir.

Veri kümesi üzerinde yaygın olarak kullanılan diğer bir denetimsiz öğrenme algoritması olan k-ortalamlar yöntemi ile müşteri segmentasyonu işlemi gerçekleştirilmiştir. Kümeleme işlemi sonucunda modelin oluşturduğu kümelerin performansını değerlendirmek için -1 ile 1 arasında bir değerlendirme yapan Silhouette skorlama yöntemi ile bir değerlendirme yapılmıştır. Silhouette skorunun yüksek olması, gözlem değerlerinin bulunduğu kümeye uygun olduğu ve diğer kümelerden farklı olduğunu göstermektedir. Silhouette skorunun düşük olması kümeleme işleminin kötü yapıldığını gösterir.

Bu tez çalışmasında k-ortalamlar yöntemi ile yapılan müşteri segmentasyonu uygulaması sonucu Silhouette skoru 0,584 olarak elde edilmiştir. Bu sonuçtan hareketle, veriler arasında ilişkinin makul derecede anlamlı bulunduğu ve veri kümesinin bu şekilde segmentlere ayrıldığı söylenebilir.

Oluşturulan segmenler, veri kümesi ile birleştirilmiştir. 3 boyuttan fazla boyutlu verilerin görselleştirilebilmesi mümkün olmadığı için, oluşturulan kümeleri görselleştirebilmek için tekrar bir temel bileşenler analizi algoritması uygulanarak veri kümesi 2 boyuta indirgenmiştir. Ardından her bir küme bazlı gruplama yaparak diğer değişkenlerin ortalama değerleri alınmıştır. Kümelerin ortalama değerleri sütun grafiği görselleştirilerek oluşturulan kümelerin karakteristik özellikleri hakkında bilgi almak amaçlanmıştır.

Son olarak oluşturulan kümelere ait gözlem değerlerinin ortalama alışveriş alışkanlıklarından yola çıkarak kümeler isimlendirilmiştir. Aynı zamanda bu kümelere pazarlama bakış açısıyla, karakteristik yapılarına özel ürün ve hizmet önerileri sunulmuştur.

Bu tez çalışmasında Hepsiburada kullanıcıları üzerinde müşteri segmentasyonu analizi yapılarak benzer alışveriş alışkanlıklarına sahip müşteriler kendi aralarında kümelenmiştir. Ancak, Kişisel Verilerin Korunması Kanunu (KVKK) gereği, alınan veri kümesinde kişisel bilgi bulundurulmamasına özen gösterilmiştir. Makine öğrenmesi algoritmaları genellikle daha fazla veri ile beslendikçe, bir o kadar çok yeni ilişkiler kurabilme eğilimindedir. Daha önce de bahsedildiği gibi gereksiz veriler ile beslenmesi de aynı şekilde modelin karışıklığını ve işlem süresini artırmaktadır. Ancak gelecekteki müşteri segmentasyonu çalışmaları için, daha geniş bir veri kümesi üzerinden, model için daha yararlı olabilecek değişkenleri seçerek daha iyi sonuçlar elde edilebilir.

Aynı şekilde daha fazla kullanıcı ve alışveriş bilgisinden oluşan bir veri kümesi sayesinde de daha iyi sonuçlar elde edilebilecektir. Buna ek olarak, farklı makine öğrenmesi algoritmaları kullanılarak, hibrit yaklaşımlar denenerek ve farklı veri ön işleme

yaklaşımlarının modelin sonucu üzerindeki etkileri gözlemlenerek daha iyi bir kümeleme uygulaması gerçekleştirilebilir.

Ek olarak, oluşturulan modeller gerçek dünyada uygulanarak, farklı modellerin maddi getirileri test edilebilir. Farklı yöntemlerin birbirlerine kıyasla ne kadar satışa dönüştüğü ve işletmeye ne kadar kar getirdiği, nihai olarak hangi modelin seçileceği konusunda en önemli karar kriterlerinden biridir.



## KAYNAKÇA

- Akar, Ö., & Güngör, O. (2012). Rastgele orman algoritması kullanılarak çok bantlı görüntülerin sınıflandırılması. *Journal of Geodesy and Geoinformation*, 1(2), 139-146. <https://doi.org/10.9733/jgg.241212.1t>
- Arslan İ. (2019). *Python ile Veri Bilimi* (4. bs). Pusula.
- Aşlıyan, R., & Günel, K. (2010). Metin İçerikli Türkçe Dokümanların Sınıflandırılması. *Akademik Bilişim Konferansı*, 659-665.
- Bastem, H. N. (2021). *Student Academic Performance Prediction Via Artificial Intelligence Using Machine Learning Algorithms*. Çankaya Üniversitesi.
- Bramer, M. (2007). *Principles of data mining* (third edition, C. 180). Springer. <http://www.springer.com/series/7592>
- Burt, S., & Sparks, L. (2003). E-commerce and the retail process: a review. *Journal of Retailing and Consumer Services*, 10(5), 275-286. [https://doi.org/10.1016/S0969-6989\(02\)00062-0](https://doi.org/10.1016/S0969-6989(02)00062-0)
- Canpolat, O. (2001). *E-ticaret ve Türkiye'deki gelişmeler*. Sanayi ve Ticaret Bakanlığı.
- Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., Vogt-Maranto, L., & Zdeborová, L. (2019). Machine learning and the physical sciences. *Reviews of Modern Physics*, 91(4). <https://doi.org/10.1103/RevModPhys.91.045002>
- Chatterjee, S., & Hadi, A. S. (2006). *Regression Analysis by Example* (Fourth Edition). John Wiley & Sons.
- Chen, I. J., & Popovich, K. (2003). Understanding customer relationship management (CRM): People, process and technology. *Business Process Management Journal*, 9(5), 672-688. <https://doi.org/10.1108/14637150310496758>
- Cover, T. M., & Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1), 21-27.
- Deliçay, Murat. (2021). *Perakende e-ticaretin yükselişi: fırsatlar, sorunlar ve öneriler*. Türkiye Cumhuriyeti Cumhurbaşkanlığı Strateji ve Bütçe Başkanlığı.
- Deloitte Danışmanlık A.Ş. (2021). *Türkiye'nin Hepsiburada'sı Pandemide Bir Yılı Geride Bırakırken Türkiye'de e-Ticaret Ekosistemi*. <https://www2.deloitte.com/content/dam/Deloitte/tr/Documents/consumer-business/hepsiburada-etki-analizi-calismasi.pdf>

- Demir, F. O., & Kırdar, Y. (2007). Müşteri İlişkileri Yönetimi: CRM. *Review of Social, Economic & Business Studies*, 7(8), 293-308.
- Demirci, V. G. (2022). Dijital Reklamcılıkta Makine Öğrenmesi ve Veri Gizliliği. *Kent Akademisi Dergisi*, 15, 1455-1474. <https://doi.org/10.35674/kent.1145325>
- Demirdöğmez, M. (2021). *E-Ticaret Mevcut Durum ve Gelecek Projeksiyonları*. <https://www.researchgate.net/publication/354751086>
- Demirel, B. (2023). *Bulut Temelli Bir Mühendislik Portalında Etkin Veri Girişi İçin Makine Öğrenmesi Algoritmaları*. Eskişehir Teknik Üniversitesi.
- Dilki, G., & Deniz Başar, Ö. (2020). İşletmelerin İflas Tahmininde K- En Yakın Komşu Algoritması Üzerinden Uzaklık Ölçütlerinin Karşılaştırılması. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 19(38), 224-233. <http://dergipark.gov.tr/ticaretfbd>
- Emel, G. G., & Taşkın, Ç. (2005). Veri Madenciliğinde Karar Ağaçları ve Bir Satış Analizi Uygulaması. *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, 6(2), 221-239.
- Erkan, N. (2022). *E-Ticaretin Gelişimi ve Türkiye’de E-Ticaretin Ekonomi Üzerine Etkileri*.
- Ersungur, M. Ş., Kızıltan, A., & Polat, Ö. (2007). Türkiye’de Bölgelerin Sosyo-Ekonomik Gelişmişlik Sıralaması: Temel Bileşenler Analizi. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 21(2), 55-66.
- Gemci, Ö. (2019). *Müşteri İlişkileri Yönetiminin Müşteri Sadakatine Etkisi*. Marmara Üniversitesi.
- Ghahramani, Z. (2004). Unsupervised learning. İçinde O. Bousquet, U. von Luxburg, & G. Rätsch (Ed.), *Advanced Lectures on Machine Learning* (C. 3176, ss. 72-112). Springer. <http://www.gatsby.ucl.ac.uk/~zoubin>
- Glielmo, A., Husic, B. E., Rodriguez, A., Clementi, C., Noé, F., & Laio, A. (2021). Unsupervised Learning Methods for Molecular Simulation Data. İçinde *Chemical Reviews* (C. 121, Sayı 16, ss. 9722-9758). American Chemical Society. <https://doi.org/10.1021/acs.chemrev.0c01195>
- Güleryüz, S. S. (2019). *Pazarlama, Modern Pazarlama ve Dijital Pazarlama*.

- Gürler, C., Çağlar, M., & Önay, O. (2020). Clustering Countries by K-means Method According to Causes of Death. *Alphanumeric Journal*, 8(1), 111-130. <https://doi.org/10.17093/alphanumeric.588835>
- Hamzaçebi, C., & Kutay, F. (2004). Yapay Sinir Ağları İle Türkiye Elektrik Enerjisi Tüketiminin 2010 Yılına Kadar Tahmini. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 19(3), 227-233.
- Hızlı, M. (2019). *İlaç Sektöründe Dijital Pazarlama Faaliyetlerine Yönelik CRM Uygulamalarına Dayalı Kümeleme Analizi ile Müşteri Segmentasyonu*.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6), 228-252.
- İşığışık, E. (2022). *Betimsel İstatistik* (3. baskı). Sentez Yayıncılık.
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data Clustering: A Review. *Data clustering: a review. ACM computing surveys (CSUR)*, 31(3), 264-323.
- Jolliffe, I. (2005). Principal component analysis. *Encyclopedia of statistics in behavioral science*, 3, 1580-1584.
- Jordan, M. I., & Tom M. Mitchell. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260.
- Kaastra, I., & Boyd, M. (1999). Designing a neural network for forecasting financial and economic time series. *Neurocomputing*, 10(3), 215-236.
- Karaatlı, M., Helvacioğlu, Ö. C., Ömürbek, N., & Tokgöz, G. (2012). Yapay Sinir Ağları Yöntemi İle Otomobil Satış Tahmini. *International Journal of Management Economics and Business*, 8(17). <https://doi.org/10.11122/ijmeb.2012.8.17.290>
- Kavaklı, İ. (2018). *E-ticaret Rehberi*. Kodlab Yayın Dağıtım Yazılım LTD. ŞTİ.
- Khodakarami, F., & Chan, Y. E. (2014). Exploring the role of customer relationship management (CRM) systems in customer knowledge creation. *Information and Management*, 51(1), 27-42. <https://doi.org/10.1016/j.im.2013.09.001>
- Kırmızıgül, S., & İbrahim, Ç. (2008). KxKNN: K-Means ve K en yakın Komşu Yöntemleri ile Ağlarda Nüfuz Tespiti. *EMO Yayınları*, 120-124.
- Koldere Akın, Y. (2008). *Veri Madenciliğinde Kümeleme Algoritmaları Ve Kümeleme Analizi*. Marmara Üniversitesi.

- Koltan, Ş., & Patır, S. (2011). Kümeleme Analizi ve Pazarlamada Kullanımı. *Akademik Yaklaşımlar Dergisi*, 2(1), 91-113. <https://www.researchgate.net/publication/256986444>
- Kotler, P., & Keller, K. L. (2012). *Marketing management* (14. bs).
- Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. *Emerging artificial intelligence applications in computer engineering*, 160(1), 3-24.
- Kotsiantis, S. B. (2013). Decision trees: A recent overview. *Artificial Intelligence Review*, 39(4), 261-283. <https://doi.org/10.1007/s10462-011-9272-4>
- Kutlu, B., & Badur, B. (2009). Yapay Sinir Ağları ile Borsa Endeksi Tahmini. *Yönetim Dergisi*, 20(63), 25-40.
- Li, Y., & Wu, H. (2012). A Clustering Method Based on K-Means Algorithm. *Physics Procedia*, 25, 1104-1109. <https://doi.org/10.1016/j.phpro.2012.03.206>
- Macqueen, J. (1967). Classification and Analysis of Multivariate Observations. *5th Berkeley Symp. Math. Statist. Probability*, 281-297.
- Mahesh, B. (2018). Machine Learning Algorithms - A Review. *International Journal of Science and Research (IJSR)*, 9(1), 381-386. <https://doi.org/10.21275/ART20203995>
- Malawer, S. (2008). *Global Governance of E-commerce and Internet Trade: Recent Developments*.
- Mckinney, W. (2011). pandas: a Foundational Python Library for Data Analysis and Statistics. *Python for high performance and scientific computing*, 14(9), 1-9. <http://pandas.sf.net>
- Mirasçı Seray. (2023). *Sac Malzeme Satın Alma Süreçlerinde Veri Madenciliği ve Makine Öğrenmesi Uygulamaları*. Bursa Uludağ Üniversitesi.
- Mohapatra, S. (2013). *E-Commerce Strategy*. Springer US. <https://doi.org/10.1007/978-1-4614-4142-7>
- Pack Kaelbling, L., Littman, M. L., & Moore, A. W. (1996). Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, 4, 237-285.
- Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559-572. <https://doi.org/10.1080/14786440109462720>

- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830. <http://scikit-learn.sourceforge.net>.
- Peppers, D., Rogers, M., & Dorf, B. (1999). Is your company ready for one-to-one marketing. *Harvard business review*, 77(1), 151-169. <http://www>.
- Qin, Z. (2009). Introduction to E-commerce. *Introduction to E-commerce*, 1-517. <https://doi.org/10.1007/978-3-540-49645-8/COVER>
- Randles, B. M., Pasquetto, I. V., Golshan, M. S., & Borgman, C. L. (2017). Using the Jupyter notebook as a tool for open science: An empirical study. *ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 1-2.
- Raschka, S. (2015). *Python machine learning*. Packt publishing ltd.
- Rokach, L., & Maimon, O. (2006). Clustering Methods. İçinde *Data Mining and Knowledge Discovery Handbook* (ss. 321-352). Springer-Verlag. [https://doi.org/10.1007/0-387-25465-x\\_15](https://doi.org/10.1007/0-387-25465-x_15)
- Sammut, C., & Geoffrey I. Webb (Ed.). (2011). *Encyclopedia of Machine Learning*. Springer Science & Business Media.
- Samuel, A. L. (1959). Artificial Intelligence for Games: Seminar Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3(3), 210-229.
- Santos, V. F. dos, Sabino, L. R., Morais, G. M., & Goncalves, C. A. (2017). E-Commerce: A Short History Follow-up on Possible Trends. *International Journal of Business Administration*, 8(7), 130. <https://doi.org/10.5430/ijba.v8n7p130>
- Sarman, 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.
- Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, 2(3). <https://doi.org/10.1007/s42979-021-00592-x>

- Schafer, J. Ben, Konstan, J., & Riedl, J. (1999). Recommender Systems in E-Commerce. *Proceedings of the 1st ACM conference on Electronic commerce* , 158-166. [www.reel.com](http://www.reel.com)
- Sniatala, P., Amini, M. H., & Boroojeni, K. G. (2020). *Fundamentals of Brooks–Iyengar Distributed Sensing Algorithm*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-33132-0>
- Söztutar, T. B. (2010). *Pazarlamada Müşteri İlişkileri Yönetimi ve Gıda Sektörü Üzerine Bir Uygulama*. Kadir Has Üniversitesi.
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (second edition). MIT press.
- Tang, T. C., & Chi, L. C. (2005). Neural networks analysis in business failure prediction of Chinese importers: A between-countries approach. *Expert Systems with Applications*, 29(2), 244-255. <https://doi.org/10.1016/j.eswa.2005.03.003>
- Taşçı, E., & Onan, A. (2016). K-En Yakın Komşu Algoritması Parametrelerinin Sınıflandırma Performansı Üzerine Etkisinin İncelenmesi. *Akademik Bilişim*, 1(1), 4-18.
- T.C. Ticaret Bakanlığı. (2022). *Elektronik Ticaret Bilgi Sistemi (ETBİS) 2022 Yılı İlk 6 Ay Verileri*. [https://www.eticaret.gov.tr/dnnqthgzvawtdxraybsaacxtymawm/content/FileManager/Dosyalar/Eticaret\\_Bulteni\\_2022\\_Ilk\\_6\\_Ay.pdf](https://www.eticaret.gov.tr/dnnqthgzvawtdxraybsaacxtymawm/content/FileManager/Dosyalar/Eticaret_Bulteni_2022_Ilk_6_Ay.pdf)
- Theobald, O. (2017). *Machine Learning For Absolute Beginners* (Second). Scatterplot Press.
- Ticaret.gov.tr. (2019, Temmuz 29). *Dijital Ticaret Tanım ve Kavramlar*. <https://ticaret.gov.tr/hizmet-ticareti/elektronik-ticaret/dijital-ticaret-tanim-ve-kavramlar>
- Tosi, S. (2009). *Matplotlib for Python developers*. Packt Publishing Ltd.
- Turban, E., King, D., Lee, J. K., Liang, T.-P., & Turban, D. C. (2015). *Electronic Commerce: A Managerial and Social Networks Perspective*. <https://doi.org/10.1007/978-3-319-10091-3>
- Tütüncü, T. E. (2022). *Makine Öğrenmesi Algoritmaları ile Kredi Temerrüt Riskini Tahmin Etme*. Bursa Uludağ Üniversitesi.
- Uygur, E. (2010). *E-ticaret ve Türkiye'deki durumu*.

- Uysal, F., & Aksoy, Ş. (2004). Müşteri İlişkileri Yönetimindeki Temel Boyutlar ve Tıbbi Malzeme Lojistiği Üzerine Bir Uygulama. *Akdeniz Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 4(7), 129-144.
- Van Der Maaten, L., Postma, E., & Van Den Herik, J. (2009). Dimensionality Reduction: A Comparative Review. *J Mach Learn Res*, 10, 66-71. <http://www.uvt.nl/ticc>
- Van Der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). *The NumPy array: a structure for efficient numerical computation*.
- Wang, L. (2016). Discovering Phase Transitions with Unsupervised Learning. *Physical Review B*, 94(19). <https://doi.org/10.1103/PhysRevB.94.195105>
- Watson, D. S. (2023). On the Philosophy of Unsupervised Learning. *Philosophy & Technology*, 36(28), 28. <https://doi.org/10.1007/s13347-023-00635-6>
- Yapraklı, T. Ş., & Keser, E. (2008). Müşteri Yaşam Boyu Değerinin Analizi: Bir Saha Araştırması. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 12(2), 483-503.
- Yereli, A. N. (2001). Müşteri İlişkileri Yönetimi (CRM) ve Günümüz Türkiye'sindeki Yeri. *Yönetim ve Ekonomi Dergisi*, 8(1), 29-42.
- Yıldız, K., Çamurcu, Y., & Dogan, B. (2010). Veri Madenciliğinde Temel Bileşenler Analizi ve Negatif Matris Çarpanlarına Ayırma Tekniklerinin Karşılaştırmalı Analizi. *Akademik Bilişim '10*, 207-213. <https://www.researchgate.net/publication/288499258>
- Yurdakul, M. (2015). İlişkisel pazarlama anlayışında müşteri sadakati olgusunun ayrıntılı bir şekilde analizi. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 17.
- Zhou, Z.-H. (2021). *Machine Learning*. Springer Nature. <https://doi.org/10.1007/978-981-15-1967-3>

## EKLER

### Ek 1: Veri setinin tanımlanması

```
In [1]: import pandas as pd
```

```
df = pd.read_csv('data.csv')
df.head()
```

```
Out[1]:
```

	receivedAt	orderNumber	userId	sku	domain	basketAmount	productAmount
0	2022-01-01T05:27:37Z	-2207087542370126279	8910569964850783703	-1158940591953934668	Fashion & Lifestyle	80.91	80.91
1	2022-01-01T06:53:38Z	3463398509105352910	-5912979623820006624	-3189275990496335346	Fashion & Lifestyle	334.77	53.99
2	2022-01-01T06:53:38Z	3463398509105352910	-5912979623820006624	5653133865623348119	Fashion & Lifestyle	334.77	54.00
3	2022-01-01T06:53:38Z	3463398509105352910	-5912979623820006624	6653633759903391451	Fashion & Lifestyle	334.77	226.78
4	2022-01-01T07:44:02Z	-7258528727536827197	-2884111188620100873	2391104320713948662	Home & Garden	414.00	414.00

```
In [2]: print(df.shape)
```

```
(72863, 7)
```

```
In [3]: print(f"Benzersiz kullanıcı: {len(df.groupby('userId'))}")
```

```
Benzersiz kullanıcı: 4963
```

```
In [4]: print(f"Benzersiz sipariş: {len(df.groupby(['userId', 'orderNumber']))}")
```

```
Benzersiz sipariş: 43003
```

```
In [5]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 72863 entries, 0 to 72862
Data columns (total 7 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   receivedAt      72863 non-null  object
1   orderNumber     72863 non-null  int64
2   userId          72863 non-null  int64
3   sku             72863 non-null  int64
4   domain          72863 non-null  object
5   basketAmount   72863 non-null  float64
6   productAmount  72863 non-null  float64
dtypes: float64(2), int64(3), object(2)
memory usage: 3.9+ MB
```

```
In [6]: df = df.astype({"orderNumber": 'object', "userId": 'object', "sku": 'object', "domain": 'category'})
```

```
In [7]: df['receivedAt'] = pd.to_datetime(df['receivedAt'])
```

```
In [8]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 72863 entries, 0 to 72862
Data columns (total 7 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   receivedAt      72863 non-null  datetime64[ns, UTC]
1   orderNumber     72863 non-null  object
2   userId          72863 non-null  object
3   sku             72863 non-null  object
4   domain          72863 non-null  category
5   basketAmount   72863 non-null  float64
6   productAmount  72863 non-null  float64
dtypes: category(1), datetime64[ns, UTC](1), float64(2), object(3)
memory usage: 3.4+ MB
```

```
In [9]: df.describe()
```

```
Out[9]:
```

	basketAmount	productAmount
count	72863.000000	72863.000000
mean	599.433931	338.604118
std	2069.979017	1612.866872
min	0.000000	0.000000
25%	77.000000	25.450000
50%	178.950000	71.020000
75%	414.515000	172.780000
max	114314.380000	114314.380000



```
In [10]: df['domain'].value_counts()
```

```
Out[10]: Books & Hobbies      25124
Supermarket      17904
Fashion & Lifestyle  11477
Technology       7724
Home & Garden     7038
Appliances       2519
Hepsiburada Services  569
Mobile           506
Other            2
Name: domain, dtype: int64
```

```
In [11]: temp = df.groupby(by=['userId'], as_index=False)['orderNumber'].nunique()
len(temp[(temp['orderNumber']==1)])
```

```
Out[11]: 1240
```

```
In [12]: temp = df.groupby(by=['userId-'], as_index=False)['orderNumber'].nunique()
temp.sort_values('orderNumber', ascending=False)
```

```
Out[12]:
```

	userId	orderNumber
4891	8932083154555201028	214
4134	6096383531488901231	200
4049	5756178235851501224	193
4762	8411903090895221856	167
1597	-3387392014918255632	165
...	...	...
3540	4021241346941361819	1
3537	4009513232304776404	1
288	-8136459466196051341	1
1714	-2972262092339268389	1
4962	9219766512169354843	1

4963 rows x 2 columns

```
In [13]: temp = df.groupby(by=['userId', 'orderNumber'], as_index=False)['receivedAt'].count()
product_cnt_per_basket = temp.rename(columns = {'receivedAt': 'productCnt'})
product_cnt_per_basket.sort_values('productCnt', ascending=False)
```

```
Out[13]:
```

	userId	orderNumber	productCnt
37648	6974245716608969091	4843190789353525478	50
33875	5592608416177972138	1782207703985799107	50
25749	1909321156647226594	-8646454977037671575	41
41655	8580366716790331560	3130275092037129314	37
4381	-7377823875730449696	-7164073312091665009	37
...	...	...	...
16870	-1907438460568492980	-6775734169075984139	1
16869	-1907438460568492980	-7978477186061666457	1
16865	-1909316214901328364	4457992930150488692	1
16864	-1909316214901328364	3752116096664292299	1
43002	9219766512169354843	7043331984096424294	1

43003 rows x 3 columns

## Ek 2: Veri görselleştirme

```
In [14]: import numpy as np
         from scipy import stats
```

```
In [15]: data = df[(np.abs(stats.zscore(df["productAmount"])) < 3)]
```

```
In [16]: data = data[(np.abs(stats.zscore(data["basketAmount"])) < 3)]
```

```
In [17]: data
```

```
Out[17]:
```

	receivedAt	orderNumber	userid	sku	domain	basketAmount	productAmount
0	2022-01-01 05:27:37+00:00	-2207087542370126279	8910569964850783703	-1158940591953934668	Fashion & Lifestyle	80.91	80.91
1	2022-01-01 06:53:38+00:00	3463398509105352910	-5912979623820006624	-3189275990496335346	Fashion & Lifestyle	334.77	53.99
2	2022-01-01 06:53:38+00:00	3463398509105352910	-5912979623820006624	5653133865623348119	Fashion & Lifestyle	334.77	54.00
3	2022-01-01 06:53:38+00:00	3463398509105352910	-5912979623820006624	6653633759903391451	Fashion & Lifestyle	334.77	226.78
4	2022-01-01 07:44:02+00:00	-7258528727536827197	-2884111188620100873	2391104320713948662	Home & Garden	414.00	414.00
...	...	...	...	...	...	...	...
72858	2022-12-31 23:18:18+00:00	-6024306834204222552	-496073675151649235	710583592107002819	Books & Hobbies	50.00	50.00
72859	2022-12-31 23:27:36+00:00	-4878815539775359249	-496073675151649235	7345114099369013420	Books & Hobbies	23.00	22.00
72860	2022-12-31 23:27:36+00:00	-4878815539775359249	-496073675151649235	710583592107002819	Books & Hobbies	23.00	1.00
72861	2022-12-31 23:27:54+00:00	166553617047767314	4828818243572142253	-2211919226393502800	Fashion & Lifestyle	229.99	229.99
72862	2022-12-31 23:30:39+00:00	7513772089685661913	-496073675151649235	710583592107002819	Books & Hobbies	4.00	4.00

71107 rows x 7 columns

aykırı değerler çıkarılmadan önce skewness ve kurtosis değerleri

```
In [18]: print("ProductAmount Skewness: %f" % df['productAmount'].skew())
         print("ProductAmount Kurtosis: %f" % df['productAmount'].kurt())
```

```
print("basketAmount Skewness: %f" % df['basketAmount'].skew())
print("basketAmount Kurtosis: %f" % df['basketAmount'].kurt())
```

```
ProductAmount Skewness: 16.726797
ProductAmount Kurtosis: 557.565278
basketAmount Skewness: 13.078179
basketAmount Kurtosis: 309.482217
```

aykırı değerler çıkarıldıktan sonra skewness ve kurtosis değerleri

```
In [19]: print("ProductAmount Skewness: %f" % data['productAmount'].skew())
         print("ProductAmount Kurtosis: %f" % data['productAmount'].kurt())
```

```
print("basketAmount Skewness: %f" % data['basketAmount'].skew())
print("basketAmount Kurtosis: %f" % data['basketAmount'].kurt())
```

```
ProductAmount Skewness: 5.740578
ProductAmount Kurtosis: 40.201348
basketAmount Skewness: 3.557520
basketAmount Kurtosis: 15.035434
```

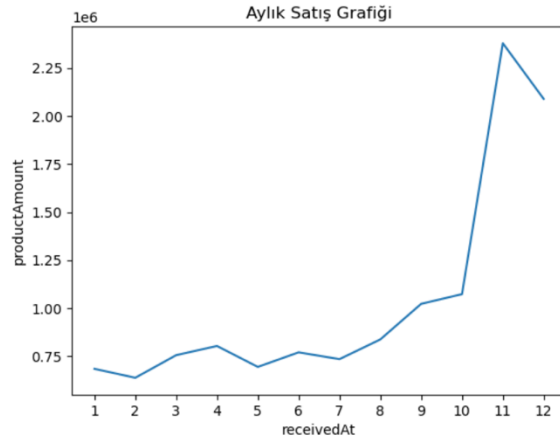
```
In [20]: import seaborn as sns
         import matplotlib.pyplot as plt
         import warnings
         warnings.filterwarnings('ignore')
```

```
In [21]: monthly_orders = data.groupby(data.receivedAt.dt.month)['productAmount'].sum()
```

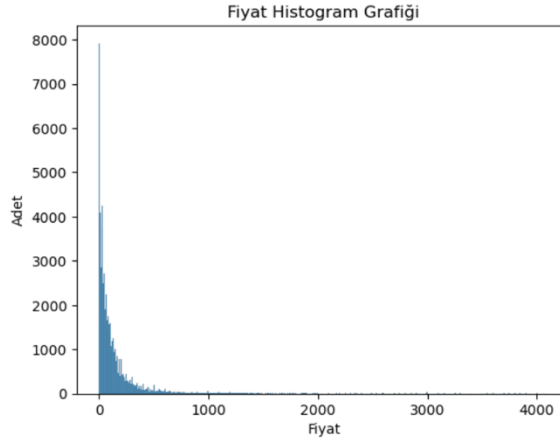
```
In [22]: monthly_orders
```

```
Out[22]: receivedAt
1      684517.24
2      637915.57
3      755577.00
4      803386.81
5      694580.18
6      770281.02
7      734881.86
8      837389.56
9      1022472.70
10     1072991.56
11     2378522.49
12     2088881.51
Name: productAmount, dtype: float64
```

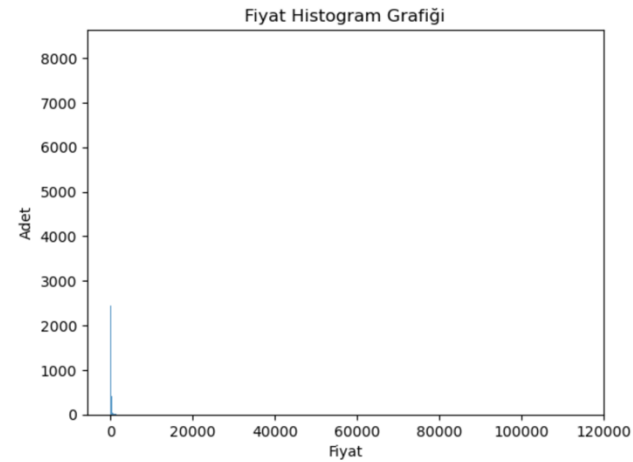
```
In [23]: sns.lineplot(monthly_orders)
plt.xticks(range(1,13));
plt.title('Aylık Satış Grafiği');
```



```
In [24]: sns.histplot(data["productAmount"]);
plt.title('Fiyat Histogram Grafiği');
plt.xlabel('Fiyat');
plt.ylabel('Adet');
```

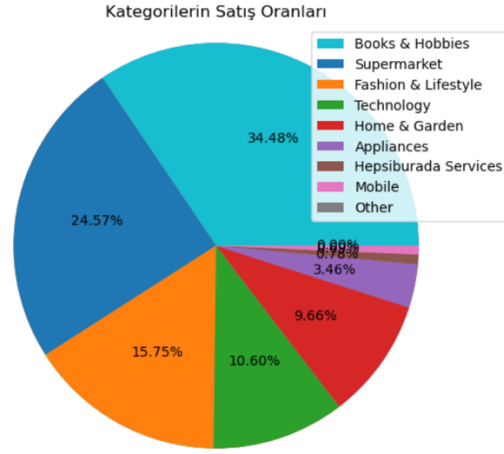


```
In [25]: #Aykırı değerler temizlenmeden önce fiyat histogram grafiği
sns.histplot(df["productAmount"]);
plt.title('Fiyat Histogram Grafiği');
plt.xlabel('Fiyat');
plt.ylabel('Adet');
```



```
In [25]: temp_s = pd.DataFrame(df['domain'].value_counts()).reset_index()
temp_s = temp_s.rename(columns={'domain': 'cnt', 'index': 'domain'})
plt.pie(temp_s["cnt"], autopct='%1.2f%%',);

patches, texts = plt.pie(temp_s["cnt"])
plt.legend(patches, temp_s["domain"], loc="upper right")
plt.axis('equal')
plt.tight_layout()
plt.title('Kategorilerin Satış Oranları');
plt.show()
```



### Ek 3: Veri önışleme

```
In [26]: pivoted = pd.pivot_table(df,
                                index='userId',
                                columns = 'domain',
                                values = 'productAmount',
                                aggfunc = 'sum').reset_index()
```

```
In [27]: pivoted
```

```
Out[27]:
```

domain	userId	Appliances	Books & Hobbies	Fashion & Lifestyle	Hepsiburada Services	Home & Garden	Mobile	Other	Supermarket	Technology
0	-9219215372597658510	0.00	576.82	818.99	0.0	49.99	0.0	0.0	796.79	214.82
1	-9213536368735428322	0.00	0.00	469.20	0.0	0.00	0.0	0.0	243.89	0.00
2	-9212213781084510606	0.00	955.19	802.66	0.0	0.00	0.0	0.0	1167.09	3469.55
3	-9212143531100464576	0.00	0.00	581.42	0.0	0.00	0.0	0.0	0.00	0.00
4	-9205248175333532922	14571.41	24817.44	11871.84	0.0	5182.68	0.0	0.0	11388.17	30636.47
...	...	...	...	...	...	...	...	...	...	...
4958	9197639350600673735	0.00	0.00	0.00	0.0	0.00	0.0	0.0	0.00	171.99
4959	9198648949746904109	0.00	382.62	0.00	0.0	256.29	0.0	0.0	759.80	103.79
4960	9199769750491676274	56.11	0.00	0.00	9.9	993.87	0.0	0.0	801.95	290.98
4961	9208887071934344602	0.00	0.00	0.00	0.0	0.00	0.0	0.0	0.00	77.70
4962	9219766512169354843	0.00	0.00	0.00	0.0	0.00	0.0	0.0	0.00	99.99

4963 rows x 10 columns

```
In [28]: agg_functions = {'basketAmount': ['min', 'max', 'nunique'],
                          'productAmount': ['sum']}
df_agg = df.groupby('userId').agg(agg_functions).reset_index()
```

```
In [29]: df_agg.columns = ['userId', 'min', 'max', 'cnt', 'sum']
```

```
In [30]: df_agg['mean'] = df_agg['sum']/df_agg['cnt']
```

```
In [31]: df_agg
```

```
Out[31]:
```

	userId	min	max	cnt	sum	mean
0	-9219215372597658510	49.99	338.94	14	2457.41	175.529286
1	-9213536368735428322	118.99	469.20	3	713.09	237.696667
2	-9212213781084510606	73.00	1888.69	12	6394.49	532.874167
3	-9212143531100464576	581.42	581.42	1	581.42	581.420000
4	-9205248175333532922	4.00	17970.28	114	98468.01	863.754474
...	...	...	...	...	...	...
4958	9197639350600673735	171.99	171.99	1	171.99	171.990000
4959	9198648949746904109	11.44	278.09	14	1502.50	107.321429
4960	9199769750491676274	9.90	462.34	13	2152.81	165.600769
4961	9208887071934344602	77.70	77.70	1	77.70	77.700000
4962	9219766512169354843	99.99	99.99	1	99.99	99.990000

4963 rows x 6 columns

```
In [32]: df_final = pd.concat([pivoted, df_agg], axis=1, join="inner")
```

```
In [33]: df_final = df_final.drop(columns=['userId', 'Other'])
```

```
In [34]: df_final
```

```
Out[34]:
```

	Appliances	Books & Hobbies	Fashion & Lifestyle	Hepsiburada Services	Home & Garden	Mobile	Supermarket	Technology	min	max	cnt	sum	mean
0	0.00	576.82	818.99	0.0	49.99	0.0	796.79	214.82	49.99	338.94	14	2457.41	175.529286
1	0.00	0.00	469.20	0.0	0.00	0.0	243.89	0.00	118.99	469.20	3	713.09	237.696667
2	0.00	955.19	802.66	0.0	0.00	0.0	1167.09	3469.55	73.00	1888.69	12	6394.49	532.874167
3	0.00	0.00	581.42	0.0	0.00	0.0	0.00	581.42	581.42	581.42	1	581.42	581.420000
4	14571.41	24817.44	11871.84	0.0	5182.68	0.0	11388.17	30636.47	4.00	17970.28	114	98468.01	863.754474
...	...	...	...	...	...	...	...	...	...	...	...	...	...
4958	0.00	0.00	0.00	0.0	0.00	0.0	0.00	171.99	171.99	171.99	1	171.99	171.990000
4959	0.00	382.62	0.00	0.0	256.29	0.0	759.80	103.79	11.44	278.09	14	1502.50	107.321429
4960	56.11	0.00	0.00	9.9	993.87	0.0	801.95	290.98	9.90	462.34	13	2152.81	165.600769
4961	0.00	0.00	0.00	0.0	0.00	0.0	0.00	77.70	77.70	77.70	1	77.70	77.700000
4962	0.00	0.00	0.00	0.0	0.00	0.0	0.00	99.99	99.99	99.99	1	99.99	99.990000

4963 rows x 13 columns

## Ek 4: Makine öğrenmesi

```
In [35]: from sklearn.preprocessing import MinMaxScaler
```

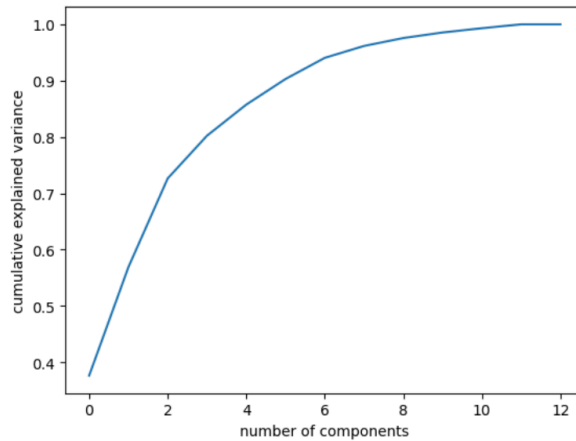
```
In [36]: # define standard scaler
scaler = MinMaxScaler()
# transform data
x_scaled = scaler.fit_transform(df_final)
print(x_scaled)

[[0.      0.01555594 0.00660783 ... 0.08074534 0.00455016 0.00516278]
 [0.      0.      0.00378563 ... 0.01242236 0.00132036 0.00699128]
 [0.      0.02576   0.00647608 ... 0.06832298 0.01184009 0.01567323]
 ...
 [0.00017451 0.      0.      ... 0.07453416 0.00398616 0.00487075]
 [0.      0.      0.      ... 0.      0.00014387 0.00228536]
 [0.      0.      0.      ... 0.      0.00018514 0.00294097]]
```

```
In [37]: from sklearn.decomposition import PCA
```

```
In [38]: import numpy as np
from scipy import stats
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
```

```
In [39]: pca = PCA().fit(x_scaled)
plt.plot(np.cumsum(pca.explained_variance_ratio_))
plt.xlabel("number of components")
plt.ylabel("cumulative explained variance");
plt.show ()
```



```
In [40]: pca = PCA(n_components=7)
pca.fit(x_scaled)
x_pca = pca.transform(x_scaled)
```

```
In [41]: print ("variance ratio: ", pca.explained_variance_ratio_)
print ("sum: ", sum(pca.explained_variance_ratio_))
```

```
variance ratio: [0.37644301 0.19261063 0.15738693 0.07578762 0.05520387 0.04549409
 0.03756749]
sum: 0.9404936290453988
```

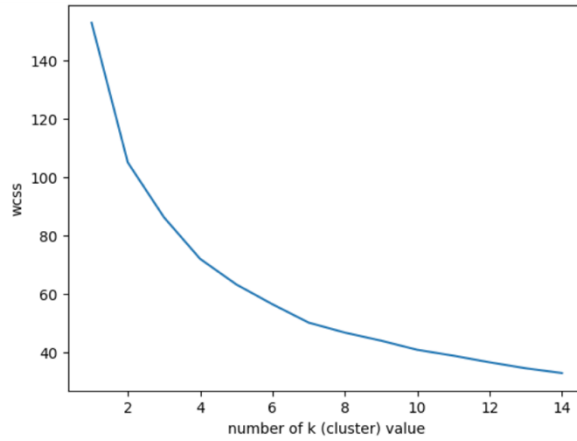
```
In [42]: from sklearn.cluster import KMeans
```

```
In [43]: import seaborn as sns
import matplotlib.pyplot as plt
import warnings
import numpy as np
warnings.filterwarnings('ignore')
```

```
In [44]: wcss = []
```

```
In [45]: for k in range(1,15):
    kmeans = KMeans(n_clusters=k)
    kmeans.fit(x_pca)
    wcss.append(kmeans.inertia_)
```

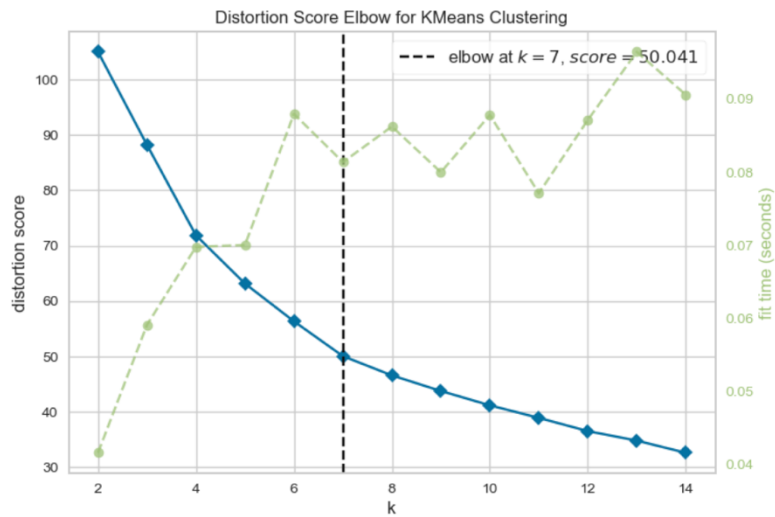
```
In [46]: plt.plot(range(1,15),wcss)
plt.xlabel("number of k (cluster) value")
plt.ylabel("wcss")
plt.show()
```



```
In [47]: !pip3 install yellowbrick
```

```
In [48]: from yellowbrick.cluster import SilhouetteVisualizer
from yellowbrick.cluster import KElbowVisualizer
```

```
In [49]: kmodel=KMeans(random_state=42)
elb_visualizer=KElbowVisualizer(kmodel,k=(2,15))
elb_visualizer.fit(x_pca)
elb_visualizer.show()
```



```
In [50]: from sklearn.cluster import KMeans
```

```
In [51]: from sklearn.metrics import silhouette_score
from yellowbrick.cluster import SilhouetteVisualizer
from yellowbrick.cluster import KElbowVisualizer
```

```
In [52]: kmeans = KMeans(init='k-means+', n_clusters = 7, random_state=42)
kmeans.fit(x_pca)
y = kmeans.predict(x_pca)
silhouette_avg = silhouette_score(x_pca, y)
print('Silhouette score: {:.3f}'.format(silhouette_avg))
```

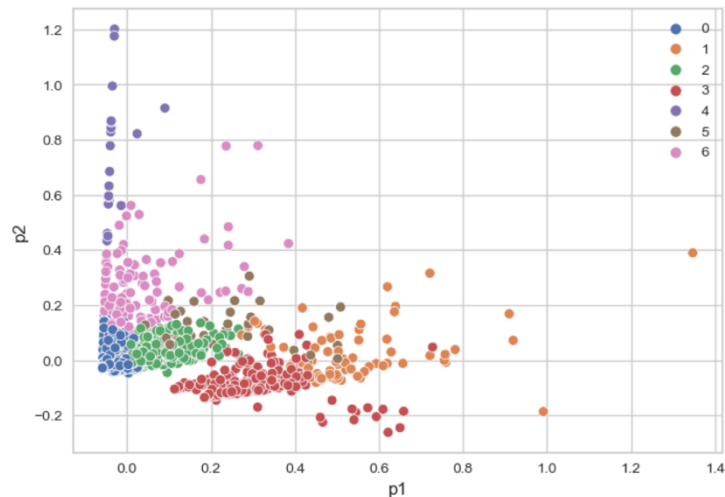
Silhouette score: 0.584

```
In [53]: pca = PCA(n_components=2)
pca_2d = pca.fit_transform(x_scaled)
pca_df = pd.DataFrame(pca_2d)
pca_df['cluster'] = pd.Series(y)
```

```
In [54]: pca_df.set_axis(['p1', 'p2', 'cluster'], axis='columns', inplace=True)
```

```
In [55]: sns.scatterplot(data=pca_df, x="p1", y="p2", hue="cluster", palette="deep")

plt.legend()
plt.xlabel("p1")
plt.ylabel("p2")
plt.show()
```



```
In [56]: df_final.loc[:, 'cluster'] = y
```

```
In [57]: df_final.groupby('cluster')['cluster'].count()
```

```
Out[57]: cluster
0      3910
1         68
2        514
3        266
4         22
5         51
6        132
Name: cluster, dtype: int64
```

```
In [58]: group_cols = ['Appliances', 'Books & Hobbies', 'Fashion & Lifestyle', 'Hepsiburada Services', 'Home & Garden', 'Mobile', 'Supermarket', 'Technology']
group_means = df_final.groupby('cluster')[group_cols].mean().reset_index()
```

```
In [59]: group_means
```

```
Out[59]:
```

cluster	Appliances	Books & Hobbies	Fashion & Lifestyle	Hepsiburada Services	Home & Garden	Mobile	Supermarket	Technology	min	max	cnt
0	458.659645	245.026207	240.726212	0.032737	126.839046	221.906090	203.417882	336.940335	213.744494	1096.043179	3.913043
1	5807.736176	5734.572647	7105.994412	7.961765	2542.678382	2758.517353	4865.086912	4904.984265	8.117500	7244.365000	76.985294
2	2022.553074	1688.715117	1079.191712	0.124514	726.167393	1218.152646	1647.763716	1736.851401	48.088696	3450.509416	21.243191
3	1868.074286	1170.574774	805.922218	10.141353	461.573045	1688.442068	847.431729	1861.804737	7.115150	3539.778496	18.736842
4	587.607727	1142.548636	5.314091	0.000000	0.000000	16583.362727	0.000000	7167.307727	19845.116364	22210.299091	1.136364
5	6194.547647	1061.449804	1643.126471	1.400000	8489.810784	1760.070980	1557.206863	1632.044902	67.236078	6873.130000	25.509804
6	8873.599621	1702.024318	2882.856061	0.120455	278.809470	13480.976364	298.694545	17715.958864	1538.931591	21512.004394	8.568182



```

In [60]: a = 5 # number of rows
         b = 3 # number of columns
         c = 1 # initialize plot counter
         fig = plt.figure(figsize = (20,15))

         for i in group_cols:
             plt.subplot(a,b,c)
             plt.ylabel(i)
             sns.barplot(data=group_means, x="cluster", y=i)
             c = c + 1

         plt.tight_layout()
         plt.show()

```

