

T.C.  
İSTANBUL AYDIN ÜNİVERSİTESİ  
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ



PAZARLAMA VE MÜŞTERİ İLİŞKİLERİ YÖNETİMİNDE VERİ  
MADENCİLİĞİ UYGULAMALARI: PAZAR SEPET ANALİZİNDE  
APRİORİ ALGORİTMASININ UYGULANMASI

YÜKSEK LİSANS TEZİ

Anıl AKSOY

İşletme Anabilim Dalı  
İşletme Yönetimi Bilim Dalı

Kasım, 2019

T.C.  
İSTANBUL AYDIN ÜNİVERSİTESİ  
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ



PAZARLAMA VE MÜŞTERİ İLİŞKİLERİ YÖNETİMİNDE VERİ  
MADENCİLİĞİ UYGULAMALARI: PAZAR SEPET ANALİZİNDE  
APRİORİ ALGORİTMASININ UYGULANMASI

YÜKSEK LİSANS TEZİ

Aml AKSOY  
(Y1512.040032)

İşletme Anabilim Dalı  
İşletme Yönetimi Bilim Dalı

Tez Danışmanı: Dr. Öğretim Üyesi Burçin KAPLAN

Kasım, 2019

T.C.  
İSTANBUL AYDIN ÜNİVERSİTESİ  
SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜ



YÜKSEK LİSANS TEZ ONAY FORMU

Enstitümüz İşletme Anabilim Dalı İşletme Yönetimi Tezli Yüksek Lisans Programı Y1512.040032 numaralı öğrencisi Anıl AKSOY'un "PAZARLAMA VE MÜŞTERİ İLİŞKİLERİ YÖNETİMİNDE VERİ MADENCİLİĞİ UYGULAMALARI : PAZAR SEPET ANALİZİNDE APRİORİ ALGORİTMASININ UYGULANMASI" adlı tez çalışması Enstitümüz Yönetim Kurulunun 13.06.2019 tarih ve 2019/13 sayılı kararıyla oluşturulan jüri tarafından oybirliği/oyçokluğu ile Tezli Yüksek Lisans tezi 25.11.2019 tarihinde kabul edilmiştir.

<u>Unvan</u>	<u>Adı Soyadı</u>	<u>Üniversite</u>	<u>İmza</u>
<b>ASIL ÜYELER</b>			
<b>Danışman</b>	Dr. Öğr. Üyesi	Burçin KAPLAN	İstanbul Aydın Üniversitesi
<b>1. Üye</b>	Dr. Öğr. Üyesi	Özge EREN	İstanbul Aydın Üniversitesi
<b>2. Üye</b>	Dr. Öğr. Üyesi	Vildan GÜLPINAR DEMİRCİ	Aksaray Üniversitesi

**ONAY**

Prof. Dr. Ragıp Kutay KARACA  
Enstitü Müdürü

## YEMİN METNİ

Yüksek Lisans tezi olarak sunduđum “Pazarlama ve Müşteri İlişkileri Yönetiminde Veri Madenciliđi Uygulamaları: Pazar Sepet Analizinde Apriori Algoritmasının Uygulanması ” adlı çalışmanın, tezin proje safhasından sonuçlanmasına kadarki bütün süreçlerde bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurulmaksızın yazıldığını ve yararlandığım eserlerin Bibliyografya’da gösterilenlerden oluştuđunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve onurumla beyan ederim. (.../.../2019)

**Anıl AKSOY**



## **ÖNSÖZ**

Yüksek lisans öğrenim hayatım boyunca öncelikle değerli fikir ve önerileriyle beni yönlendiren, her konuda destek veren, gösterdiği sabır ve katkılarıyla bilgilerini her zaman benimle paylaşan, esirgemeyen danışmanım Dr. Öğretim Üyesi Burçin KAPLAN'a teşekkürlerimi sunarım. Bütün bu süreçte bana desteklerinden dolayı Emiliya NABIYEVA'ya teşekkür ederim. Son olarak hayatım boyunca beni hem maddi hem de manevi olarak destekleyen, beni bugünlere getiren aileme sonsuz teşekkür ederim.

**Kasım, 2019**

**Anıl AKSOY**



## İÇİNDEKİLER

### Sayfa

ÖNSÖZ.....	iv
KISALTMALAR .....	viii
ÇİZELGE LİSTESİ.....	ix
ŞEKİL LİSTESİ.....	x
ÖZET.....	xi
ABSTRACT .....	xii
<b>1. GİRİŞ .....</b>	<b>1</b>
<b>2. PAZARLAMA VE MÜŞTERİ İLİŞKİLERİ YÖNETİMİ .....</b>	<b>4</b>
2.1 Pazarlama ve Müşteri İlişkileri Yönetiminin Tanımı.....	4
2.2 CRM Kavramının Ortaya Çıkışı .....	7
2.3 MİY'in İşletmeler Açısından Önemi.....	9
2.4 CRM'nin Türkiyedeki Yeri .....	9
2.5 E-CRM Amaçları, Araçları ve Yararları .....	11
2.6 MİY'de Müşteri İle İlişki İçerisindeki Kavramlar .....	13
2.7 CRM Boyutları.....	13
2.7.1 Operasyonel müşteri ilişkileri yöntemi.....	14
2.7.2 Analitik müşteri yöntemi .....	14
2.7.3 İşortağı müşteri ilişkileri yönetimi .....	14
2.8 MİY Ortaya Çıkma Sebepleri.....	15
2.9 Müşteri Yaşam Döngüsü .....	15
2.9.1 Müşteri odaklılık .....	16
2.9.2 Müşteri tatmini ve değer yaratma.....	16
2.9.3 Müşteri sadakati .....	17
2.9.4 Müşteri koruma .....	18
2.9.5 Müşteriyi elde tutma .....	18
2.9.6 Müşteri büyütme ve derinleştirme.....	19
2.9.7 Müşteri geri kazanım .....	19
2.9.8 Müşteri seçimi .....	19
2.9.9 Müşteri edinme.....	20
2.9.10 Müşteri koruma .....	20
<b>3. VERİ, VERİ AMBARI VE VERİ MADENCİLİĞİ .....</b>	<b>21</b>
3.1 Veri,Enformasyon ve Bilgi.....	21
3.1.1 Veri ile bilgi arasındaki bağlantı .....	21
3.2 Veri Ambarı .....	22
3.2.1 Veri ambarının özellikleri .....	24
3.2.2 OLAP ve OLTP .....	25
3.3 Veri Madenciliği Süreci Ve Crisp-Dm Süreci.....	26
3.3.1 Crisp-Dm süreci .....	26
3.3.1.1 Amaç ve hedeflerin belirlenmesi .....	26
3.3.1.2 Veriyi anlamak .....	27
3.3.1.3 Veriyi hazırlama .....	27

3.3.1.4 Veriye model seçimi .....	27
3.3.1.5 Verinin değerlendirilmesi ve uygulanması .....	27
3.3.1.6 Konuşlandırma ve sahaya sürmek .....	28
3.4 Problemin Belirlenmesi .....	28
3.4.1 Verilerin hazırlanması .....	28
3.4.2 Model kurulması ve değerlendirilmesi .....	29
3.4.3 Model uygulaması .....	31
3.5 Veri Madenciliği ve Erdemli Döngüsü .....	31
3.5.1 Tipik Operasyonel İşletme Sistemleri ve Veri Madenciliği Sistemleri Arasındaki Farklılıklar .....	33
3.6 Veri Madenciliği Metodolojileri, Algoritmaları .....	35
3.6.1 Tanımlayıcı modeller (descriptive) .....	36
3.6.2 Tahmin edici modeller (predictive) .....	36
3.7 Metodoloji Adımları ve Farklı Teknikler .....	36
3.7.1 Sınıflama ve regresyon modelleri.....	37
3.7.2 Genetik algoritmalar .....	39
3.7.3 Kodlama süreci.....	40
3.7.4 İlk popülasyonun oluşturulması .....	40
3.7.5 Uygunluk değerinin hesaplanması .....	40
3.7.6 Çoğalma işlemi.....	41
3.7.7 Yeni kuşağın oluşması ve döngünün durdurulması .....	41
3.7.7.1 Örnek tabanlı yöntemler ( K-En yakın komşu ) .....	45
3.7.7.2 Kümeleme modeli.....	46
3.7.7.3 Ardışık zamanlı örüntüler .....	48
3.7.7.4 Yapay sinir ağları (YSA) .....	48
3.7.7.5 Karar ağaçları .....	52
3.8 Veri Madenciliği Uygulama Alanları.....	56
3.8.1 Pazarlama.....	55
3.8.2 Finans ve bankacılık .....	57
3.8.3 Sigorta sektöründe .....	58
3.8.4 Perakendecilik .....	59
3.8.5 Üretim sektöründe .....	59
3.8.6 Elektronik ticaretinde .....	59
3.8.7 Telekomünikasyon .....	59
3.8.8 Hilekarların tespitinde .....	59
3.8.9 Tıp sektöründe.....	61
3.8.10 Web sektöründe.....	61
3.9 Veri Madenciliğinin Uygulandığı Diğer Alanlar .....	61
<b>4. PAZAR SEPET ANALİZİ VE BİRLİKTELİK KURALLARI</b> .....	<b>63</b>
4.1 Market Sepet Analizi.....	63
4.2 Birliktelik Kuralları.....	64
4.3 Birliktelik Kurallarının Belirlenmesinde Kullanılan Temel Algoritmalar .....	66
4.3.1 AIS algoritması .....	66
4.3.2 SETM algoritması .....	67
4.3.3 Apriori algoritması .....	68
4.3.4 Apriori-TID algoritması.....	72
4.3.5 Apriori-hybrid algoritması .....	73
4.4 Güven (Confidence) ve Destek (Support) Ölçüleri.....	73
4.5 Konu ile İlgili Literatür Taraması.....	74

<b>5. PAZARLAMA VE MÜŞTERİ İLİŞKİLERİ YÖNETİMİNDE VERİ MADENCİLİĞİ UYGULAMALARI:PAZAR SEPETİ ANALİZİNDE APRIÖRİ ALGORİTMASININ UYGULANMASI.....</b>	<b>76</b>
5.1 Arařtırmada Verilerin Toplanması.....	76
5.2 Anakütle ve Örnekleme.....	76
5.3 Arařtırmanın Yöntem ve Algoritması .....	76
5.4 Arařtırmanın Amacı .....	77
5.5 Arařtırmanın Bulguları.....	77
<b>6. SONUÇ.....</b>	<b>87</b>
<b>KAYNAKLAR.....</b>	<b>89</b>
<b>ÖZGEÇMİŐ .....</b>	<b>95</b>





## KISALTMALAR

<b>CRISP-DM</b>	: Cross Industry Standard Process For Data Mining
<b>CRM</b>	: Customer Relationship
<b>DM</b>	: Data Mining
<b>MİY</b>	: Müşteri İlişkileri Yönetimi
<b>OLAP</b>	: On-line Analytic Process
<b>OLTP</b>	: On-line Transaction Process
<b>VM</b>	: Veri Madenciliği
<b>C</b>	: Aday öge küme
<b>Ck</b>	: k elemanlı aday öge küme
<b>L</b>	: Sık Öge küme
<b>Lk</b>	: k elemanlı sık öge küme
<b>X2</b>	: Tüm güven

## ÇİZELGE LİSTESİ

	<u>Sayfa</u>
Çizelge 4.1: Notasyonlar.....	69
Çizelge 4.2: Bağıntı Verileri .....	70
Çizelge 4.3: Ürün Güven Endeksi.....	72
Çizelge 5.1: Cinsiyet Dağılımı.....	77
Çizelge 5.2: Ödeme Şeklinin Dağılımı.....	78
Çizelge 5.3: Yaş Dağılımı.....	78
Çizelge 5.4: Cinsiyete Göre Laptop Çanta Satışı.....	78
Çizelge 5.5: Cinsiyete Göre Küçük Boy Bel Çanta Satışı.....	79
Çizelge 5.6: Yaşa Göre Deri Kredikartı Satışı .....	79
Çizelge 5.7: Veri Kesitinden Elde Edilen Birliklik Kuralları.....	82
Çizelge 5.8: Düzenlenmiş Ürünler İçin Birliklik Kuralları .....	84
Çizelge 5.9: Cinsiyete Göre Çantaların Birliklik Kuralları .....	85
Çizelge 5.10: Cinsiyete Göre Cüzdan ve Kartlık Birliklik Kuralları .....	86

## ŞEKİL LİSTESİ

### Sayfa

Şekil 2.1:	CRM Bileşenler .....	7
Şekil 2.2:	Müşterinin İlişkide Bulunduğu Kavramlar .....	13
Şekil 2.3:	Operasyonel ve Analitik CRM Arasındaki Bağlımlar .....	15
Şekil 3.1:	Veri ile Bilgi Arasındaki Bağlantı .....	222
Şekil 3.2:	OLTP ve OLAP Arasındaki Farklar .....	26
Şekil 3.3:	Modelin Kurulması .....	300
Şekil 3.4:	Modelin Değerlendirilmesi .....	300
Şekil 3.5:	Veri Madenciliği Erdem Döngüsü .....	333
Şekil 3.6:	VM Metodolojileri ve Algoritmaları .....	355
Şekil 3.7:	VM Metodolojilerinde İzlenen Adımlar .....	377
Şekil 3.8:	Naive Bayes Formülü .....	41
Şekil 3.9:	Yapay Sinir Ağı ve Biyolojik Sinir Ağının Benzerlikleri .....	49
Şekil 3.10:	Yapay Sinir Ağı Üzerinde Girdiler ve Çıktılar Arasındaki Ağırlıklar Matrisi .....	50
Şekil 3.11:	Yapay Sinir Ağlarında Girdi Katmanı İle Çıktı Katmanının İşleyiş İlişkisi .....	51
Şekil 4.1:	Apriori Algoritmasının Adımları .....	71
Şekil 4.2:	Destek Ölçüsünün Matematiksel Formülü .....	73
Şekil 4.3:	Güven Ölçütünün Matematiksel Formülü .....	73
Şekil 5.1:	Algoritmanın Aşamaları .....	80
Şekil 5.2:	Ürünlerin Hareketlerdeki Dağılımı .....	81
Şekil 5.3:	Web Grafiği İle Ürünler Arasındaki İlişkiler .....	83
Şekil 5.4:	Sıklığı Yüksek Olan İlişkiler .....	83

# PAZARLAMA VE MÜŞTERİ İLİŞKİLERİ YÖNETİMİNDE VERİ UYGULAMALARI: PAZAR SEPETİ ANALİZİNDE APRIORİ ALGORİTMASININ UYGULANMASI

## ÖZET

Günümüzde sürekli gelişen teknoloji sayesinde veri madenciliği CRM ile pazarlama arasında stratejik bir iş bağı oluşturmaktadır. Müşteri ilişkilerinde veri madenciliğinden faydalanmak için pazar sepet analizinden yararlanılabileceği düşünülmüştür. Yapılan çalışmanın da ana konusu bu analizden elde edilen sonuçlara göre müşteri gruplarına hitap etmektir.

Bu tez çalışmasında müşteri ilişkileri yönetimi, müşteri yaşam döngüsü veritabanlarında bilgi keşfi süreci, veri madenciliği ve veri madenciliğinde yer alan modeller açıklanmıştır. Tanımlayıcı veri madenciliği modellerinden biri olan Birliktelik Kuralları ve Apriori Algoritması ayrıntılı olarak incelenmiştir.

Tezin uygulama bölümünde, Türkiye’de perakende sektöründe yer alan bir işletmenin havalimanı mağazasından iki aylık satış verileri kullanılmıştır. 1650 işlem sayısı içerisinde birden fazla ürün tercihi yapan 1550 müşteri hareketinin bulunduğu iki aylık veri kümesi üzerinden 42 ürün grubu incelenerek ürün gruplarının birliktelik analizi SPSS Modeler 15.0 programı yardımı ile gerçekleştirilmiştir.

Böylece hangi ürünün beraber, hangi ürünün ayrı sergilenmesi gerektiği, hangi ürünün müşteri tarafından daha az tercih edildiği ortaya çıkarılmış. Müşteri istek ve ihtiyaçlarını bulabilmek için veri madenciliği uygulamalarından faydalanılabileceği gösterilmiştir.

**Anahtar Kelimeler:** *Veri Madenciliği, Müşteri İlişkileri Yönetimi, Müşteri İlişkileri, İletişim, Veritabanı, Apriori Algoritması, Birliktelik Kuralları*

# **DATA MINING TECHNIQUES IN MARKETING AND CUSTOMER RELATIONSHIP MANAGEMENT: APRIORI ALGORITHM IN MARKET BASKET ANALYSIS**

## **ABSTRACT**

In these days, technological improvements become continuous so data mining CRM and marketing have some common strategical issues. There is an idea had been thought that data mining can help the consumer relations and the research of these important analyses's outcomes is directly related with costumers.

In this thesis, the management of costumer relations, the timeline of cutomers life cycle research models have been explained. The theory of togetherness Apiori Algoritma is also investigated well in this research.

The aplication process of this thesis, a Turkish company's sales verications examined which constructed in airport and sale their product retail. The sales data of two months of this company has been used in this research. Inside the number of 1650 transactions, 1550 number customer movement of transactions is found who pick more than one product. In this two-month dataset 42 was examined how associated this group of products by using SPSS moduls programme.

To sum up, which product display together and which one can display separetly from other products has been emerged. This research also provided the efficent demand number of the goods to find the costumers needs and demands.

**Keywords:** *Data Mining, Customer Relationship Management, Customer Relationship, Communication, Database, Association Rules, Apriori Algorithm.*

## 1. GİRİŞ

Son zamanlarda teknolojik gelişim ile birlikte firmalar için yarış daha fazla hissedilmeye başlanmıştır. Bundan dolayı şirketler alıcıların sadakatine güvenemez duruma gelmiş ve alıcısına yakın ilgi gösterme gereksinimi duymaya başlamıştır. Bu durum müşteri ilişkileri yönetimi kavramını meydana getirmiştir. Veri madenciliği genellikle müşteri ilişkileri yönetimine ile beraber kullanılmaya başlanmıştır (Tezcanlar,2007, s. 3).

Müşteri ilişkileri yönetimi şirketlerin elinde bulunan müşterileriyle ileri düzeyde olan bağlantılarını geliştirmek, müşteri kazanmak ve müşterilerinin tamamını kaybetmemek için izlenen stratejiler sürecidir. Müşteri davranışlarını tanımlamak ve geleceğe dair varsayım uygulamak adına benzerlikler meydana getirmesi veri madenciliğinin amacını yansıtmaktadır (Şimşek,2006, s. 91).

Global firmalar teknolojinin gelişmesini bir fırsata çevirerek elde ettikleri verileri kullanarak, veri içerisinde faydalı olabilecek kısımları ortaya çıkarmaya çalışmışlardır. Bu kısımda talepleri karşılayabilecek çözüm olarak veri madenciliği ortaya çıkmıştır. Veri setinin bir kademesi olan veri madenciliği, daha önce kullanılmayan ama aslında kullanılabilme ihtimali olan verinin bulunmasını sağlar. Diğer bir deyişle, geniş bir veriyi incelenerek anlam ifade eden bir tablo ve kaidelerin bulunmasına sebebiyet verir.

Sayım bilimi ve yazılım programlarını değerlendirerek elde edilen kullanışlı bilgiyi veri tabanından çıkarma işlemi Veri Madenciliği olarak adlandırılmaktadır. Bundan dolayı veri madenciliği, veri tabanı üzerinde bilgi bulma sürecinden meydana gelmiştir. Müşterinin eğilim gösterdiği hizmetleri ve malları temin etme meyilinin olduğunun belirlenmesi, alıcının ihtiyacından fazla mal satın almasına sebep olarak firmaya yarar sağlamaktadır. Tüketicinin ürün ve hizmet satın alma davranışını tanımlayan birliktelik kuralları ve ardışık zamanlı örüntüler, pazarlama amaçlı olarak market sepet analizinde kullanılır (Döşlü, 2008, s. 15).

Günümüzde perakende yöneticilerinin teşhir planlamasında uyguladıkları yöntemler satış, karlılık ve rekabet odaklıdır. Bu yöntemler kolay ve hızlı sonuç verdikleri kadar yerinde sayan ve çokta doğru olmayan kararlar alınmasına sebep olmaktadır.

Bunun başlıca sebebi satışları arttırırken arkada kalmış ürün gruplarını daha da geriye itmesidir. Potansiyeli olan ürün gruplarının satış potansiyelini göstermeden başarısız sayılmasına yol açmaktadır. Asıl amaç, tüketicinin almayı düşünmediği ve aklında olmayan ürünleri sattırarak olmalıdır. Belirli bir satış oranına veya karlılığa takılıp keşfedilemeyen potansiyel ürünlerin ortaya çıkmasını sağlamak gerekmektedir (Dürek, 2018, s. 3).

Bu tez çalışmasının amacı veri madenciliği yöntemlerinin perakende sektöründe yer alan bir firmanın müşteri ilişkileri yönetimine yönelik kullanılmasına bağlı olarak uygulamanın gerçekleştirilmesi amaçlanmıştır.

Bu amaçla bir hazır giyim firmasının müşterilerine ait veri seti kullanılmıştır. Bu verilere Birliktelik kuralları analizi ve Apriori algoritması uygulanmıştır. Bu algoritmanın amacı müşterilerin müşterilerin satın alma alışkanlıklarının bulunması ve en çok beraber satılan ürünlerin tespitinin yapılmasıdır. Bu sonuçların yardımıyla şirketin müşterileri tanınması ve yeni pazarlama stratejileri geliştirmesi hedeflenmektedir.

Bu tez çalışması 5 bölümden oluşmaktadır:

Birinci bölümde pazarlamanın tanımı yapılmış ve tarihsel gelişim süreci ele alınmıştır. Pazarlama aşamaları ve pazarlama karması hakkında bilgiler verilmiştir.

İkinci bölümde Müşteri İlişkileri Yönetimi'nin tanımı yapılmış, MİY süreci, faydaları, tarihçesi ve müşteri yaşam döngüsü, Türkiye ve dünya geneli CRM kavramı hakkında bilgiler verilmiştir.

Üçüncü bölümde veri, veri ambarı, enformasyon hakkında bilgiler verilmiştir. Veri madenciliği için gereken alt yapıyı oluşturan veri ambarları ve OLAP, OLTP teknolojileri gözden geçirilmiştir. Veri Madenciliği kavramı, modelleri ve teknikleri anlatılmıştır. Veri Madenciliği süreci ele alınmış ve bu süreç boyunca izlenen adımlar anlatılmıştır. Veri madenciliğinde kullanılan tanımlayıcı ve tahmin edici modellerden Diskriminant Analizi, Kümeleme

Analizi, Regresyon Analizi, Lojistik Regresyon, Faktör Analizi, Karar Ağaçları, Yapay Sinir Ağları, Genetik algoritmaların kullanım alanları açıklanmıştır.

Dördüncü bölümde Birliktelik Kuralları ve farklı algoritma çeşitleri açıklanmıştır. Birliktelik kurallarının tanımı, tarihçesi anlatılmıştır. Apriori algoritmasının ortaya çıkışı ve kodlanma mantığı anlatılmıştır.

Beşinci bölümde ise araştırma bulgularına yer verilmiştir. Pazar sepet analizinde ortaya çıkan ve birbiri ile en çok ilişkisi bulunan ürünler bu ürünlerin birbirleri ile olan örüntüleri açıklanmıştır.





## **2.PAZARLAMA VE MÜŞTERİ İLİŞKİLERİ YÖNETİMİ**

### **2.1 Pazarlama ve Müşteri İlişkileri Yönetiminin Tanımı**

Pazarlama en basit tanımıyla herhangi bir işletmenin elde ettiği mal ve hizmetleri son kullanıcıya kadar ulaştırma sürecidir. Pazarlamada iki yaklaşım bulunmaktadır. Bunlardan ilki sosyo-ekonomik açıdan konuya yaklaşım gösterendir. Bu yaklaşıma göre pazarlama, kişiler arasında en kısa sürede oluşan ve kar amacı gütmeyen; insanlar arasındaki ilişkileri pozitif yönde etkileyerek alıcı-satıcı mantığından meydana gelmektedir. Pazarlama rolünü “insanlara daha iyi yaşam olanakları sunmak” olarak ele alan bu yaklaşıma göre, pazarlama; kişi ve grupların istek ve ihtiyaçlarını karşılamak amacıyla bir önem arz eden mal, hizmet ve fikirlerin oluşturulması, sunulması ve değişimini içeren beşeri faaliyetlerin tümüdür (Tek, 2001: 1).

İkinci yaklaşıma göre pazarlama sosyal bakımdan daha çok işletme açısından yaklaşmaktadır. Bu yaklaşım pazarlamanın alıcı-satıcı ilişkisi üzerinde yoğunlaşmaktadır. Alıcının satın aldığı mal ve hizmetlerle, satıcının aldığı ücretle ilgilenmektedir. Bu yaklaşımda ise pazarlama: örgütsel hedeflere ulaşmak için işletmelerce, tüketicilerin beklenti ve gereksinimlerini karşılayacak mal ve hizmetlerin yaratılması, fiyatlandırılması, tutundurulması ve dağıtılması aşamalarına ilişkin bir bütündür (Tek, 2001: 2).

Başka bir tanıma göre pazarlama: pazarlama süreci, birleştirilmiş faaliyetler dizisi olup, bunun sayesinde firmalar, alıcıları için bir kıymet yaratmakta ve bunun karşılığında müşterilerden değer görebilmek için etkili müşteri ilişkileri kurmaktadır.

CRM, en değerli müşterileri ve ilişkilerinin seçimi ve yönetimi konusunda bir iş stratejisidir (Karadeniz, 2008: 4). Günümüzde gelişen teknoloji ile birlikte pazarlama anlayışı da değişmektedir. Küreselleşen dünya ekonomisi ticarete üreticiyi odak noktasından çıkartıp müşteriye odak noktasına yerleştirmiştir.

Veri madenciliği kullanılarak CRM'ne katkı sağlamaktadır. Müşterilerinin tüketim alışkanlıklarına dair sonuçlar elde etmektedirler ve bu CRM'ne stratejik planlama fırsatını ortaya koymaktadır.

Müşteri İlişkileri Yönetimi; müşteriler ile sıkı bağlar kurulmasını, potansiyel müşteriye elde tutmayı ve yeni müşteriler kazanılmasını hedefleyen yönetim biçimidir (Suresh, 2002: 102).

Müşteri İlişkileri Yönetimi Pazarlama bilimine bağlı olarak ortaya çıkmış ve 2000'li yıllarda gelişim göstermiştir. İngilizce CRM olarak adlandırılan Müşteri İlişkileri Yönetimi aslında müşteri ve onlara ait verilerden yararlanılması üzerine geliştirilen teknolojik kökenli projelerden oluşmaktadır (Seyrek, 2006: 5).

Genel bir tanımıyla CRM satış-pazarlama faaliyetlerinde karar alma süreçlerini keskinleştiren, alınacak ve alınması gereken aksiyonlar hakkında satış/pazarlama ekibine fikir verebilen bir yazılımdır (Erol, 2013: 3). Ama maalesef ülkemizde CRM yazılım ve bilgi teknolojileri kolunu kullanmaktan daha çok felsefi olarak bir yönetsel bilim olarak kalmaktadır.

Müşteri, kısıtlı üretici ve fazla pazar olduğundan dolayı istediği alternatifi seçme hakkına sahiptir. Rekabetten dolayı ürün kalitesi ve özellikler arasında çok büyük farklar vardır. Bir rakip firmanın geliştireceği bir strateji, fiyat özelliği olarak firmaya yansımaktadır. Günümüzde rakip firmalar ile aradaki farkı belirleyen önemli etkenlerden biri CRM'dir. Müşterilerin istekleri ve talepleri doğrultusunda hareket eden işletmenin pazarlama hedeflerini daha iyi bir şekilde gerçekleştirmektedir (Bergeron, 2002: 15-19).

CRM pazarlama ve satış süreçlerinde müşteri odaklı projeleri işleme süreci olarak da bilinmektedir. Önemli olan bir diğer konu da MİY'e yalnızca bir yazılım ve otomasyon olarak ele alınmamalıdır. En başta MİY'nin bir teknolojiler bütünü olmadığını, temelinde bir süreç olduğunu unutmamak gerekir. MİY, birbiri ile ilişkisi bulunan olguların sürekli geliştirilmesi, bu olgulardan birbiriyle uyumlu sonuçlar

elde etmek için meydana getirilen kendisini tekrar eden bir süreçtir (Taşpınar, 2005: 13)

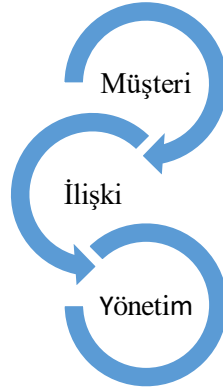
CRM, müşteri sadakati, müşteri memnuniyeti ve müşteri bağlılığı ve mevcut müşteriyi yaşam boyu elinde tutma amacıyla müşteri bilgilerini yöneten ve düzenleyen bir yöntemler dizisidir. CRM, müşteri merkezli stratejiler ile bu kararlara arka çıkabilecek; satış ve pazarlamayla bir arada, müşteri hizmetleri, muhasebe, üretim ve lojistik gibi fonksiyonları kapsayan, bu fonksiyonlardan yararlanacak tüm paydaşlar için tüm iş amaçlarının sıfırdan düzenlenmesini içeren ve bunları gerçekleştirirken de teknolojiye yararlanan bir idare etme sanatıdır (Aksoy, 2002: 16).

CRM bir iş disiplini. Müşteri ilişkilerini ilgilendiren yönetim stratejisinin haricinde bütün yönetim birimleri tarafından uygulanması gereken bir disiplin sürecidir. CRM iyi bir şekilde uygulandığında bir iş felsefesine dönüşmektedir. CRM felsefesi yönetim açısından müşteri merkezli bir politika uygulamaktadır (Karahan, 2010: 15).

Herhangi bir ürünü rastgele bir yerden satın alan kişi eğer bu davranışı, mekan ve zaman açısından tekrarlı bir biçim göstermiyorsa o kişi tüketici olarak isimlendirilir. Diğer yandan satın alma davranışına sahip kişi tükettiği ürün açısından, mekan ve zaman açısından tekrarlı bir biçimde bir davranış içerisinde ise müşteri olarak isimlendirilmektedir (Yalçın, 2008: 4).

CRM, işletmelerde pazarlamayla ilgili teknolojilerini bir araya getiren teknoloji tabanlı satış eylemidir. Bu sebeple CRM için teknolojik uygulamalar ve bilgisayar yazılımları önemlidir (Kalakota & Robinson, 2001: 142-148). CRM veri toplama, müşteri sadakati, potansiyel müşterilerinin yönetimi ve yeni müşterilerinin kazanılması gibi fonksiyonel özelliklere sahiptir.

CrM 3 bileşenden oluşmaktadır:



**Şekil 2.1:** CRM Bileşenler

**Kaynak:** Ezerçe, 2008: 33

## 2.2 CRM Kavramının Ortaya Çıkışı

CRM’i daha iyi anlayabilmek için onun tarihsel gelişim sürecine bakmalıyız:

CRM Customer Relationship Management, Müşteri İlişkileri Yönetimi literatürde kısaltması ile “MİY” kavramı 2000’li yılların sonuna doğru işletmelerde bilimsel bir iş disiplini olarak yerini almıştır (Kartajaya vd. , 2011: 37). Ama bunun gelişim süreci vardır.

II. Dünya Savaşından çıkmış, ekonomik olarak zayıflamış ülkelerin toparlanmaları uzun bir süre almış ve bu barış yılları olan 1950 yıllarda firmaların ”ürettiğini satma” fikriyle hareket etmişlerdir. Burada seri üretime önem verildiği gözükmektedir (Amirov, 2006: 28).1970’lere kadar süreçte gelişmiş ülkelerde üretici ve tüketici ilişkisine bakıldığında herhangi bir problemin olmadığı gözükmektedir (Gel, 2003: 12).

1980’lere gelindiğinde ise tüketici ilişkileri yönetimi önemli bir husus olarak algılanmış ve üzerine ciddi çalışmalar gerçekleştirilmiştir. 1990’lar ve sonrasında müşteri istek ve beklentilerinin en iyi şekilde alınması, müşteri ihtiyaçlarının her temas noktasında en üst seviyede karşılanması olarak CRM gelişmesini tamamlamak üzereydi. 90’lı yılların sonunda artan internet ve otomasyon kullanımı ile CRM firmalar ve pazarlama dünyası için vazgeçilmez bir araştırma ve uygulama başlığı olmuştur (Gökmen, 2004: 11-12).

Şirketler arası rekabet ortamı “ürün” yerine “müşteri” kökenli tanımlar yapılmasına olanak sağladı. CRM satış-pazarlama haricinde, iç organizasyonlardan başlayıp yönetim kademesine kadar ilerleyen bir iş disiplini.

Günümüzde müşteri analitiği, hareketli ve değişken bir durum izlemektedir. Bu durum müşterileri izlemeyi ve hoşnut kılmayı güçleştiren bir izlenim doğurmuştur. Bugün dünyanın birçok bölgesinde, üretici işletmelerin karşı karşıya kaldıkları en esas problem, müşteri sadakatinin çok önemli derecelerde etkisini yitirme durumudur. Bu tavrın asıl sebebi, çoğalan rekabet ve yüksek teknoloji sonucu, müşterilere arz edilen, daha mutabık maliyetli, daha alımlı değişik ürünlerdir. Müşteri sadakatinin azalması ile birlikte kar marjları da azalmaktadır. Bu bağlamda CRM, müşteri sadakatini arttırmayı amaçlayan ve kar marjlarının da artışa geçmesininin şartlarını hazırlayan bir beceridir (Kırım, 2004: 46-47).

CRM müşteri isteklerini yerine getirip aralarından en değerli müşterileri belirlemektedir. Bu değerli müşterileri müşteri portföyünde tutmak ve yeni müşterileri elde etmek amacıyla geliştirilmiştir.

MİY, müşteriler ile alakalı en detaylı malumatları elde etmek, bu müşterileri çok ince ayrımlarla bölümlendirmek (mikro-segmentasyon), bu departmanları kazançlarına uygun olarak ayrıştırmak, karlı olan müşterilere gerçekleştirilecek fazladan planmanın düzeyinin belirlemek ve her müşteriye başka ve hususi pazarlama stratejisi uygulamak olarak da belirtilebilir (Kırım, 2004: 60).

Bilişim anlamında CRM ise, bir şirket için tüm müşteri bilgilerinin belirli bir merkezde toplandığı ve bu bilgilerden elde edilen neticelere göre hali hazırda bulunan müşterilerini korumaları, onlarla alışveriş imkanlarını arttırmaları, onlara her münasebet hususunda katma değerli ve uyumlu hizmet sunmayı hedefleyen ve bu amaçla iş süreçlerini ve bilgi yönetimi sistemlerini bir araya getiren, müşteri odaklı kararlar ve düşüncülerin tümüdür (Taşpınar, 2005: 17).

CRM 3 kavramdan oluşmaktadır:

- Teknoloji Odaklı Kavram
- Müşteri Odaklı Kavram
- Strateji Odaklı Kavram

CRM'in tanımsal çerçevesi genel olarak müşteri odaklı, teknolojiden yararlanan ve strateji üzerine kurulu bir yönetim tarzı olduğu söylenebilir.

### 2.3 MİY'in İşletmeler Açısından Önemi

İşletmelerin MİY'nin tamamı ile kabullenebilmesi için MİY altyapısının takıma mensup kişiler tarafından anlaşılmalı ve bu altyapıya herkesin katılımı sağlanmalıdır. Altyapının oluşturabilmesi için teknolojiye yatırım yapılmasının gerekliliği düşünülmektedir. Çünkü yetersiz bir teknoloji ile istenilen MİY altyapısına ulaşılamayacağı düşünülmektedir. Sistemin altyapısı oluşturulduktan sonra spesifik amaçların oluşturulması gereklidir. Takımdaki herkes entegrasyona uyum sağlamalıdır. Entegrasyon işlemi tamamlandıktan sonra CRM projesini yönetebilecek bir proje yöneticisinin varlığına ihtiyaç duyulduğu öngörülmektedir.

Proje yönetimi ve CRM takımları değişime direnmemeli ve değişime açık olması istenmektedir. Yenilenmeye açık olan tüm departmanların ortak bakış açısına sahip olması ve bütün CRM takımlarının uyum içinde olması önemliliği söylenmektedir. Proje yöneticisi ve yönetim takımının en önemli görevi, kilit bir rol içeren müşteri karlılığı konusunda yükümlü hissetmeleri gerekliliği düşünülmektedir. Diğer önemli bir nokta da müşteri ilişkileri sürecinde yapılan iyileştirmelerin devamlılığının olması gerektiğidir. Başarılı bir MİY uygulama sürecini isteyen işletmelerin misyon, vizyon ve stratejiler konusunda aşağıda bahsedilen adımları izlemelerinin onların yararına olacağı düşünülmektedir (Karahana, 2010: 50);

- Müşteri merkezli bir ilişki stratejisinin oluşturulması
- Müşteri İlişkileri Yönetimi sisteminin tasarlanması
- Sistemin yaratılması ve oluşturulması: değişimin yönetimi
- Kurum içinde işlevsel görevleri yeniden belirtmek
- Süreçleri en baştan canlandırmak
- Teknoloji hakkında karar vermek
- Sürekli iyileştirme

### 2.4 CRM'nin Türkiye'deki Yeri

CRM Institute Turkey tarafından yapılan Türk şirketlerinde CRM konusu ile alakalı kişilerle yapılan araştırma şunları göstermektedir:

Türk şirketleri CRM süreçlerine aşağıda verilen dağılımda önem vermektedirler.

- Müşteri Seçimini önemseyenler %14

- Müşteri Edinmeyi önemseyenler %35
- Müşteri Korumayı önemseyenler %31
- Müşteri Sadakatini önemseyenler %20

Yukarıdaki sonuçlara bakıldığında Türk şirketlerinde “Müşteri Edinme” segmentine ağırlık verilerek satış kaygılarının yüksek olduğu kanısına varabiliriz. ”Müşteri Seçimi” segmentinin en az yüzdeyi alarak Türkiyedeki firmaların ürünlerini konumlandırmada ve temel pazarlama faaliyetlerinde yeterince ciddi davranmadıkları görülmektedir. Amerika ve Avrupada yapılan araştırmalara bakıldığında müşteri koruma ve sadakatini arttırma eğiliminde olduğu görülmektedir. Türkiyede ise bu durumun tersi yönde negatif bir eylem görülmektedir. Türk şirketleri pazar paylarını değiştirme maliyetlerini üstlenmek yerine sadık müşteriler yaratıp, bu müşteriler üzerinden ekonomik dayatma yöntemini tercih ettikleri görülmektedir. Böylece pazar paylarını değiştirme maliyetlerini başka türlü müşteri programlarına aktarmaları ile farklı ve başarılı bir strateji izledikleri görülmektedir (Güzel, 2001: 79)

CRM literatüründe 4 ana model bulunmaktadır. Bunlar aşağıdaki gibidir:

- Karlılık Modeli
- Müşteri Bilgileri
- Müşteriler İle Etkileşim
- Metot Geliştirme ve Kurumsallaşma

Araştırmada Türkiye’de şirketlere bu soru yöneltildiğinde karşımıza çıkan sonuçlar aşağıdaki gibidir:

- Karlılık Model’ine yatırımı seçenler %48
- Müşteri Bilgileri ’ne yatırımı seçenler %23
- Müşteri Etkileşimini yükseltmeyi seçenler %19
- Metot Geliştirme ve Kurumlaşma yatırımlarını seçenler %10

Bu sonuca göre Türkiyedeki firmalar yeterince kurumsallaştırmaya önem göstermedikleri ve çalışanların kendi metotları ile yollarına devam ettikleri görülmektedir. Müşteri Etkileşimi segmentinin gittikçe rağbet gördüğü ve yakın zamanda müşteri ile haberleşme kanallarında hakimiyetin artabileceği söylenebilir (Güzel, 2001: 82).

Türkiyede hedeflenen rol dağılımı

Araştırmanın Türkiyedeki rol dağılımı ile alakakalı sonuçlar aşağıdaki gibidir (Tezcanlar, 2007: 31):

- İnsan Katkısı %45
- Teknoloji Katkısı %31
- Proses Katkısı %24

Müşteri Odaklılık aşamasına tekabül eden bu araştırma sonucu üç önemli bileşenden meydana gelmektedir. Bunlar insan, proses, teknolojidir. Bu sonuçlara göre Türkiye yine Amerikada izlenen stratejilerin ters yönünde bir sonuç çıkmaktadır. Türk şirketleri CRM'i teknoloji kökenli kullanmak yerine insan olgusunun ağır bastığı bir program yönettikleri söylenebilir.

CRM araştırmaları sonuçları;

- Türkiye pazarının gelişmekte olan bir pazar olduğunu görüyoruz. Satış performansının şirketler için hala önemli bir konu teşkil ettiği görülmektedir.
- Pazarlama stratejileri biraz daha potansiyel müşteri gözüyle bakılmaktan vazgeçilmelidir.
- Gün geçtikçe artan rekabet sonrası müşteri koruma ve satış sonrası promosyonlarının önemi gittikçe artmaktadır.

## 2.5 E-CRM Amaçları, Araçları ve Yararları

- Amaçlar

E-ticaretin yaygınlaşması sayesinde bilhassa rakip firmaların bir tıklama uzaklığında olduğu “modern sanal ortamlı” işletme faaliyetlerinde tüketicilerin diğer firmalara ulaşmaları kolaylaşmıştır. Müşteri ilişkileri yönetim sürecinde firmaların proaktif ve reaktif hizmet sunmaları önem arz etmekte ve bunun içinde internet ortamına uygun teknoloji tabanlı pazarlama anlayışına ulaşmaları gerekmektedir. Yukarıda bahsedilen anlayış; E-MİY (E-CRM) olarak nitelendirilen elektronik ortamda müşteri ilişkileri yönetimi anlayışıdır (Korkmaz, 2006: 199).

Müşterilerle etkin bir iletişim ortamının oluşturulması, müşteri memnuniyetinin ölçümlerinin oluşturulması, geri dönüşüm halkasının tamamlanması ve işleyişe



konulması,operasyonel stratejilere daha fazla önem verilmesi gibi amaçların öncelikli olduğu söyleyebiliriz (Çubukçu, 2010: 46).

İkincil amaçları ise şöyle sırayabiliriz;Faaliyet maliyetlerin en aza indirgenmesi,ürün altyapısının internet ortamına aktarılması,geliştirilmesi ve yayımlanmasıdır.

- Araçlar

E-CRM kullandığı bazı araçları şöyle sırayabiliriz:

- Etkili bir iletişim ortamının oluşturulması.
- Müşteri memnuniyet ölçümlerinin meydana getirilmesi
- Müşteri geri beslemenin hızlandırılması
- Operasyonel işletme faaliyetleri ile ilgili karlılığın çoğaltılması
- Her basamaktaki ve düzeyde bulunan faaliyet maliyetlerinin en aza indirilmesi
- Müşteri orantılı mal ve hizmet portföyünün internet ortamında geliştirilmesi ve sıklaştırılması (Karadeniz, 2008: 20).

Bu araçlar çerçevesinde gerçekleşen iletişimde gerçekleşen faaliyetler bilgi güvenlik sistemi,web sitesinde çok lisanslı ve çok kurlu destek,ürün ve fiyat tanımlamaları,data analizi ve sanal düzenlemeler yer alır.

- Yararları

Verilen hizmetin yirmidört saat olmasından dolayı e-hizmetlere her zaman ve kolay ulaşılabilme imkanı sağlamaktadır. Bu hizmet sürecince gelen taleplere daha hızlı cevap verebilme, basit bir iletişim ve geri besleme şansını çoğaltmaktadır. Kampanyalar aracılığıyla yapılacak ve bir çok masraftan oluşan hizmetler e-crm ile daha düşük maliyetli hizmetler ile sağlandığını belirtebilir (Deniz, 2002: 20).

Elektronik ortamlarda online hizmet verebilme ve bu sayede veri-transfer teknolojisi kullanılmaktadır.Verit-transfer teknolojisi kullanılarak;yüksek maliyetli iletişimin yüklerinden kurtulmak hedeflenir.Eletronik ortama entegre edilmiş birimler arasında hızlı ve basit bir bilgi paylaşımının olacağından söz edilebilir (Kotorov, 2002: 218-232).

E-ticaretin meydana getirdiği geniş pazar olanaklarından yararlanarak ,müşteri tatminini arttırmak ve müşteri davranışlarını izlemek için E-CRM'in en büyük fırsatlarındandır.

CRM'nin sistematik verilerden ve analitik araçlardan yararlandığını görebilmekteyiz. Yapılan bu işlemlerden çıkan sonuçlara göre E-CRM çerçevesinde yeni bir imaj yaratılabileceği düşünülebilir.

## 2.6 MİY'de müşteri ile ilişki içerisindeki kavramlar

Müşteri İlişkileri Yönetimi odak noktası olan müşteri çerçevesi ilk olarak yönetim tarafından belirlenen politikalar ile belirlenir. Daha sonra pazarlama ve satış departmanlarının uygulamaları ile karşı karşıya gelmektedir. Bu aşamaları atlatan müşteri aldığı mal ve hizmetler karşısında işletmelere para kazandırmaktadır. Bu ücretler CRM'i finansa bağlamaktadır. Oluşturulan müşteri portföyüne ve yeni müşterilere ulaşmak için çağrı merkezleri kullanılmaktadır. Operasyon merkezleri sayesinde müşteri ile iletişim kuvvetlendirilmektedir. Bütün bu aşamalardan sonra ortaya çıkan veriler önce raporlanır ve daha sonra analiz edilir böylece daha sonraki herhangi müşteri döngüsünde bu verilerden yararlanır.



Şekil 2.2: Müşterinin İlişkide Bulunduğu Kavramlar

## 2.7 CRM Boyutları

- Operasyonel Müşteri İlişkileri Yöntemi
- Analitik Müşteri İlişkileri Yöntemi
- İş Ortağı Müşteri İlişkileri Yöntemi

### **2.7.1 Operasyonel müşteri ilişkileri yöntemi**

İş süreçlerinin yatay ve sistematik olarak birbirlerine bağlanmasına Operasyonel Müşteri İlişkileri denmektedir. Müşteriyle yapılan birebir görüşmeleri kapsamaktadır. Bunlar telefon görüşmesi, yüzyüze görüşme, posta yolu ile görüşme ve internet üzerinden müşteriye ulaşılmasıdır.

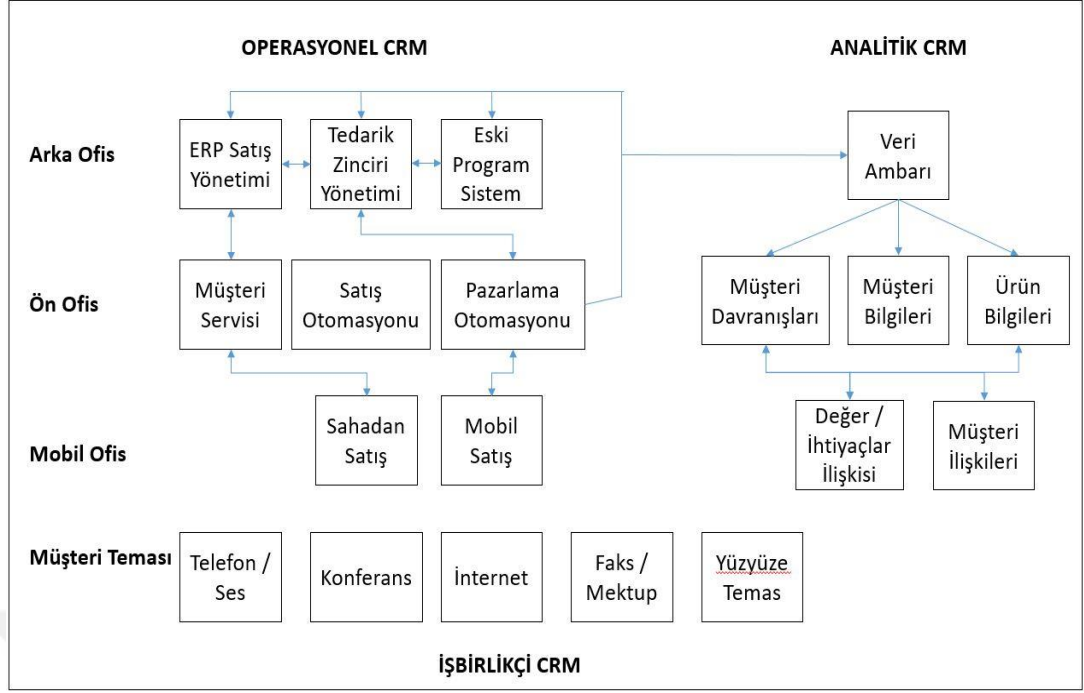
- Satış Gücü Otomasyonu (Sales Force Automation-SEA)
- Müşteri Hizmetleri ve Destek (Customer Service and Support-CSS)
- Kurumsal Pazarlama Yönetimi (Enterprise Marketing Management-EMM)

### **2.7.2 Analitik müşteri yöntemi**

Operasyonel müşteri ilişkileri yönetimi tarafından ortaya konulan verilerin analiz edilmesini “Analitik Müşteri İlişkileri Yönetimi” yapmaktadır. Analitik müşteri ilişkileri görevleri şunlardır: Müşterilere ait verilerin elde edilmesi, depolanması ve bu verilerin işlenmesi ve analiz edilerek raporlanmasıdır.

### **2.7.3 İş ortağı müşteri ilişkileri yönetimi**

İş ortağı “Müşteri İlişkileri Yönetimi” müşteri ve işletme arasında bir köprü görevi sağlar. Örneğin; müşteri ile görüşme, alıcıya e-mail gönderme, müşteriye bülten gönderme vb. uygulamaları iş ortağı müşteri ilişkilerini yönetmek için yapılır. Müşteri ile işletme arasında sağlanan koordinasyon sayesinde hangi mal ve hizmetin nasıl pazarlanacağına imkan sağlamaktadır.



**Şekil 2.3:** Operasyonel ve Analitik CRM Arasındaki Bağlılıklar

**Kaynak:** Ezerçe, 2008: 31

## 2.8 MİY Ortaya Çıkma Sebepleri

- İletişim teknolojilerinin, veritabanı yönetim sistemleri ve karar destek sistemlerinin gelişmesi
- Müşteri payının önemli hale gelmesi
- Rekabet ortamının giderek artması
- Müşteriyi elde tutmanın önemini anlaşılması
- Müşteri memnuyeti ve sadakati gibi kavramların değerinin artması
- Kitleleşen pazarlama stratejisinin gittikçe pahalıya mal olması

## 2.9 Müşteri Yaşam Döngüsü

Müşteri yaşam döngüsü müşteri ile işletmenin arasında bulunan ilişkinin derinliği ve süresi bakımından ortaya çıkan uygulama araçlarından oluşan bir döngüdür. CRM'in müşteri odaklı bir faaliyet olması ile bu döngüdeki süreçlerin işletmeler tarafından izlenmesi ve takip edilmesi gereklidir. Çünkü günümüz pazarlama alanında klasik pazarlama tekniklerinin yanında müşterinin değeride artmıştır. Rekabet piyasasında bulunan işletmeler seçici müşteriye daha tatminkar yaklaşmak zorundadır. Müşteriyi tatmin edebilecek istek ve beklentileri karşılamaya göre mal ve hizmet sunum

yapıldığında etkin müşteri memnuyeti oluşturulmuş olur. Müşteri memnuyeti sağlamakta müşteri değeri yaratmaktan geçer (Yalçın, 2008: 10).

Müşteri değeri, alınan hizmetten elde edilen yararın, hizmetten beklenen zarara oranıdır. Müşteri değerinin yüksek olması için verilen faydanın da yüksek olması gereklidir aksi takdirde müşteriler ürünün değerinin düşük olduğunu düşüneceklerdir (Söztutar, 2010: 8).

Müşteri Yaşam Döngüsünde ele alınacak kavramlar aşağıdaki gibidir:

Müşteri tatmini, müşteri sadakati, müşteri odaklılık, değer yaratma, müşteri seçimi, müşteriye elde tutma, müşteri edinme, büyüme ve derinleştirmedir.

### **2.9.1 Müşteri odaklılık**

Müşteri memnuniyeti, müşteri edinme, müşteri sadakati ve müşteri bağımlılığı gibi birçok ana başlıkta, bu kavramlarda uygulanan politikaların “müşteri odaklılığı” üzerine kurulu olduğunu söyleyebiliriz. Bilindiği gibi CRM faaliyetleri müşteriye odaklı bir dizi faaliyetler bütünüdür. Bu sayede müşteriye odaklanma, elde edilmiş bir müşterinin bağlılık ve karlılığının uzun vadede korunması ve tüketici harcamalarındaki oranın artırılması için gereken mühim hamleleri içerdiğinden bahsedebiliriz.

Müşteri odaklı bir strateji; yeni müşteriler bulmaya özen gösterip, herkesin alabileceği ürünleri satmak yerine, var olan müşterilerin mümkün olan satın alma paylarının artırılmasına odaklanmaktadır. Müşteri odaklı olmak; ticari ortaklık ve bu ortaklıkta bütün personeller olarak her faaliyet ve yöntemin sunulan mahsul ve görevlerle müşterilerin beğenisini kazanarak, mutlu edecek şekilde planlanmasına çalışmak ve sürekli tercih edilen bir sistem olabilmektir (Karadeniz, 2008: 14).

### **2.9.2 Müşteri tatmini ve değer yaratma**

Müşteri odaklılık en önemli bileşenlerinden birisi müşteri tatmini ve müşteri değeri yaratmaktır. Önce müşteri için bir değer yaratılır ve bu değer ile müşteri tatmin etmeye çalışılır. Müşterinin sadakati için müşteri tatmini önemli bir unsurdur. Tatmin olmuş bir müşterinin işletme ile olan bağı uzun süre devam edecektir. Müşteri tatmini tüketicinin tatmin olma tepkisi olarak da adlandırılmaktadır. Bir mal ve hizmetin tüketimine verdiği memnuniyet derecesi ile o mal ve hizmetlerden olan

beklentilerin karşılanıp karşılanmadığına bakar. Eğer müşteri beklentilerinin üstünde bir sonuç aldıysa müşteri tatmini sağlanmış olur (Doyle, 2003: 162-163).

Müşteri için değer yaratma, onların neye ihtiyaçları olduğunu ve bu ihtiyaçlarını aldıktan sonra ne hissettikleri ile ilgili bir konudur. Değer yaratmanın müşteri tarafında anlamı şudur; müşteri aldığı hizmet karşılığında beklentisinin üzerindeki elde edişlerinin fazla olmasını beklemektedir (Karadeniz, 2008: 27).

Müşteri için değer yaratma sonucunda elde edilebilecek müşteri tatmini, kuruluşlara şu yararları getirebilecektir:

- Fazla sayıdaki tekrarlanmış satın almaların çoğaltılması,
- Bilinçli bir yönetim ile maliyet ve giderlerin yönetilmesi ve kar marjların yükselmesi,
- Çalışanların motivasyonlarını arttırmak için yaratıcılığın desteklenmesi ve işten çıkma oranlarının düşmesi
- Değerli müşterilerin duygularını yakınlarına anlatmaları (Odabaşı, 2005: 50).

Değer yaratmayı artırmak için müşterinin ödünlerini kısıtlamak ve daha faydacı yararlar sunulmalıdır. Bu ödümler fazla ve faydacı yararlar ne kadar az ise müşteri için yaratılan değer bir anlam taşımaz. Yaratılan değer müşteri-firma arasında duygusal bir bağ kurmasını sağlar. Müşteri duygusallık ile yeniden satın almaya yönelecektir. Bu evreden sonrada ortaya müşteri sadakati çıkacaktır (Söztutar, 2010: 9-10).

### **2.9.3 Müşteri sadakati**

İçten bağlılık, güvenilir ve etkili dostluğu ifade eden sadakat, örgütsel ve kurumsal yaşamda önemli yer tutmaktadır. Müşteri sadakati ise firmadan veya hizmet alan müşterilerin o firmaya olan devamlılık düzeyleri gösteren eğilimleridir. Müşterilerin devamlılık düzeyleri bireysel özelliklere, ekonomik durumlara ve çevresel faktörlere göre değişmektedir (Söztutar, 2010: 13-14).

Müşteri sadakatini korumak için firma kaynak ve becerilerini sıkı bir mücadele ile sahaya yansıtmak zorundadır. Her iki tarafında mutlu olacağı bir durumda müşteri sadaketinden bahsetmek mümkündür. Müşteri sadakati bir müşterinin daha önceden tavsiye edilen bir malı satın alma ve yeniden satın alma tutumunda bulunmasıdır.

Müşteri sadakati davranışsal,tutumusal ve hem davranışsal hemde tutumsal olarak ele alınmaktadır (Çoroğlu, 2003: 87).

İşletmeler açısından müşteri sadakati şu açıdan önemlidir; görünüşte rekabet koşullarında sadakati kazanılmış ve muhafaza edilmiş müşteri işletmeler için rekabetçi firmalara karşı direnç yönelimidir (Yeler, 2006: 19).

Müşteri sadakati oturana kadar müşteri profilleri,potansiyel müşteri,şüpheli müşteri, belirsiz muhtemel müşteri, ilk kez faaliyet gerçekleştiren müşteri,tekrarcı müşteri, düzenli müşteri ve sadık müşteri gibi aşamaları vardır. Düzenli ve sadık bir müşteriye sahip olan firmanın müşteri sadakati konusunda bayağı yol aldığı söylenebilir.

#### **2.9.4 Müşteri koruma**

Müşteri Koruma evresi,müşteri sadakati ve müşteri memnuyeti gibi evreleride kapsamına almaktadır. Buradaki amacın müşteriyi olabildiği kadar uzun süre elde tutmak ve işletme ile olan ilişkisinin sürekliliğini arttırmak olacağından bahsedilebilir.

İyi bir müşteri koruma onları birer sadık müşteri haline getirmek ve devamlılığını sağlamak adımlarından oluşur.

Müşteri sadakatini sağlarayarak yeni müşteriler kazanmak,karlı müşterilere ulaşmak, kararsız müşteriyi elde etmenin yolları bu evrede ortaya çıkmaktadır. Diğer bir anlamda hizmet ve ürünlerin bireyselleşmesi ile sadakatini yitirmiş müşteri tekrar elde edilmeye çalışılmaktadır (Matthews & Thompson, 2000: 94).

#### **2.9.5 Müşteriyi elde tutma**

Müşteriyi elde tutma aşaması, daha önceden ürün ve hizmet satın almış bireylerin elde edilmesi ve bu bireylerden herhangi bir kayıp yaşanmaması olarak genellenebilir. Bu evrede müşteriyi kaybetmemek için yada kaybedilen müşteri yerine geçebilecek müşteriyi kazanmak için yapılan harcamaların önlenmesi hedeflenir.

Müşteri olma sürelerini anlamak amacıyla sağkalım analizi uygulanmaktadır. Sağkalım analizi müşteriyi müşteri olma sürelerine ilişkin parametreler

atayabilmektedir. Müşteri olma süresi az çıkan müşterilerin beklentileri karşılanmalı ve daha önceden tercih ettiği ürünlere yöneltilmelidir (Mocan, 2016: 50).

Müşteriyi elde tutma aşamasında dikkat edilmesi gereken bazı hususlar vardır; bunlar müşterinin firma esasında algılanması ve bilinmesi, kişisel malumatların pek çok kez sorulmasının önüne geçilmesi, müşteriye doğru zamanda ve doğruca ulaşılması olarak sıralanabilir (Erol, 2013: 11).

### **2.9.6 Müşteri büyütme ve derinleştirme**

CRM'nin son evresi, Müşteri Derinleştirme evresidir. Pazar payı anlayışı yerine, artık sadık hale getirdilen müşterinin cüzdanındaki payını arttırmak için yapılan çalışmalar da bu evreye odaklanmaktadır. Elde edilmiş bir müşterinin sadakat ve karlılığının geniş çaplı süre boyunca korunması ve müşteri harcamalarındaki oranın yükseltilmesi için gerekli işlevleri barındırır (Tezcanlar, 2007: 17).

### **2.9.7 Müşteri geri kazanım**

Bütün bu uygulamalara rağmen bazı müşterilerin gitme ihtimali bulunmaktadır. Geri kazanım bu tarz müşterilerin geri gelme olasılığını mümkün kılmaktadır. Giden bu müşterilerden bazıları değerli müşteri kategorisinde yer almaktadır. Bu değerli müşterileri geri getirmek için promosyon, fiyat indirimleri, hediye ürün ve birtakım kampanyalar gerçekleştirilmelidir.

Geri kazanım veri madenciliği kapsamından daha çok operasyon stratejik sistemlere girmektedir. Geri kazanım elde tutma işlemlerine çok benzemektedir. Fakat geri kazanımda firmadan tatmin olmayan ve sadakati kalmayan müşterilere yönelik eylemler gerçekleştirilir.

### **2.9.8 Müşteri seçimi**

Bu aşamada planlanan hedef müşterinin belirlenip, firma yahut şirket ilk defa hizmet veya ürün satın alınmasını sağlamalıdır. Bu hedefi gerçekleştirmek için reklam kampanyaları; promosyonlar, sponsorluklar, yüzyüze pazarlama teknikleri ve benzer faaliyetlerde bulunabilir. Bu aşama esnasında yüksek verimlilik almak için ise bu müşterilerin daha önceden belirlenmeleri gerekir.



Müşteri İlişkileri bu evresinde en önemli amaç “Bize en çok fayda sağlayan müşteri kimdir?” sorusunun cevabına ulaşılmak istenmesidir. Bu çerçevede şu çalışmalar yapılmaktadır (Şimşek, 2006: 99).

- Hedef Kitle Bölümlendirmesi
- Müşteri Bölümü
- Konumlandırma Çalışmaları
- Kampanya Planları
- Müşteri Planlama
- Yeni Ürün Tanıtımı ve Pazarlama

### **2.9.9 Müşteri edinme**

Müşteri edinme kısmı daha önce ürün ve hizmet almış müşterinin elde edilebilme kısmıdır. Bu kısımda müşteri kaybetmemek veya yeniden kazanılabilecek müşteri elde etmeksizin yapılacak olan işlemlerin ve maliyetlerinin önlenmesi veya en aza indirgenebilmesi hedeflenir. Bu hedef benzer için beklentili müşteri iyi takip edilmeli ve yeniden benzer ürünlere yönelmesi gerçekleştirilmelidir (Aksoy, 2002: 32).

Müşteri şirket bazında algılanması, tanınması, kişisel bilgilerin onanması ,müşteriye doğru zamanda ve doğru kanaldan ulaşılmasının sağlanması diğer önemli etkenlerdir.

### **2.9.10 Müşteri koruma**

Müşteri koruma safhasında “Bir müşteri ne kadar süreyle şirkete bağlı kalacaktır?” sorusunun cevabının arandığı safhadır. Amaç müşteriyi şirkete bağlamak, onu şirkette tutabilmek, arada oluşan ilişkisinin sürekliliğini sağlamaktır (Barnes, 2000: 11).

Bu evrede sipariş yönetimi, teslim, teslimlere geri dönüş ve problem yönetimi ile ilgili uygulamalar bulunmaktadır.

### **3. VERİ, VERİ AMBARI VE VERİ MADENCİLİĞİ**

#### **3.1 Veri,Enformasyon ve Bilgi**

Olaylar ve olgular ile alakalı kaydedilen,bir anlam barındıran bilinen ve işlenmemiş gam gerçeklerdir.Veriler kullanılarak enformatik bilgilere ulaşılır (Wikipedia, 2017).

Enformasyon karar vermek için kullanılan, örgütlenmiş ve değişmesi muhtemel olmayan bilgiler topluluğudur. Karar destek sistemleri ve bilişim sistemleri tarafından kayıt altına alınacak verilerin depolanma, sorgulanma, özetlenme işleminden sonra analiz edilmesini anlamlandıran veriler çıktısına enformasyon denir (Çağiltay, 2010: 9).

Enformasyon veri madenciliği, bilgisayar mühendisliği, telekomünikasyon mühendisliği gibi mesleklerde yaygın bir şekilde kullanılmasını enformasyon teorisine borçludur. Matematiğin bir dalı olarak ortaya çıkmış ve sayısal bilgilerin iletişim sırasında yorumlanabilmesine olanak sağladığı için sosyal bilimler tarafında benimsenmiştir. Claude Shannon tarafından 1940 yıllarda oluşturulmuştur. Günümüze kadar birçok bilim insanının geliştirmesi ile bugünkü halini almıştır (Ölçer, 2014).

Verilerin işlenmiş ve yorumlanmış halidir.Veriler genelde tanımlanmamış ham gerçekleri göz önünde bulundurlar (Barutçugil,2002:57). Bilgi oluşabilecek alternatifleri etkileyen işlenmiş veri olarak tutulur.Veriler bazen yenilenir, güncellenir veya özetlenir. Veriler hipotez test etmek için toplanır bundan dolayı işlenmemiş ve analize uğramamış sayılara bağlıdır. Veri analiz edildiğinde, herkes ve herşeyi içeren bilgi üzerinden hesaplar.

##### **3.1.1 Veri ile bilgi arasındaki bağlantı**

Bilgisayarların veriye ihtiyacı vardır.İnsanlar ise bilgiye ihtiyaç duyarlar.Veriler en küçük yapı taşıdır. Bu küçük yapı taşlarıyla bilgi oluşturulur.Bilgi anlamlıdır ve yorumlanabilir.Veriler bir içeriğe sahiptir.

Veri ise hamdır ve karışıktır. İnsanlar için şekillendirilmemiş olup 1 ve 0'lardan oluşan anlamsız bloklardır. Veri bilgiye dönüştükten sonra insanlar için hemen hemen bütün uygulamalarda karar verme noktasına hayati derecede öneme sahip olurlar (Günel, 2015).



**Şekil 3.1:** Veri ile Bilgi Arasındaki Bağlantı

**Kaynak:** Pieper, S. (2017). <https://www.artegic.com/blog/difference-data-information-knowledge/>.

### 3.2 Veri Ambarı

Veri madenciliği sürecinde problem çözmeye, sorgu oluşturmada ve analiz yapmak için kullanılan bir depodur. Birden fazla amaç için kullanılan aynı yönlü veritabanlarının tek çatı altında toplanmış halidir. Veri ambarları çoklu kaynaklardan gelen verilerin toplanmasıyla oluşur.

Firmaların ve kuruluşların gerçekleştirdikleri faaliyetlerden elde ettikleri veriler üzerinden sistematik ve belli analiz teknikleriyle gelecekte işine yarayabilecek verilerin oluşturabilmesi ve bu verilerin saklanabildiği bilgi havuzudur (Berry & Linoff, 1997:7)

Veri ambarının veri tabanından farkı, veri ambarının müşterinin istedikleri üzerine kuruluyor olmasıdır. Veritabanından veriler çekilirken bu göz önünde bulundurulur.

Bazı değerlerin belirli bir düzen içinde depolanması, istenildiğinde erişilebilir bir konumda yerleştirilmesi, lazım olan sıklıklarla güncellenmesi ve yazılımlar aracılığıyla bilgi elde etmek maksadıyla tutulması ve korunması verinin ambarlanması için gerekli adımlardandır (Özmen Ş. ,2003:154).

Veri ambarının taşınması gereken bazı özellikler vardır;

- Verinin zamana bağlı olmasıdır
- Verinin kalıcı olmasıdır
- Veri Ambarının amacı konu odaklı olmalıdır.
- Verinin bütünleşmiş bir özelliğe sahiptir.

Günümüzde işletmeler yoğun bir rekabet içerisinde bulunmaktadır. Bu yoğun rekabet ortamında hızlı bir reaksiyon göstermek için analitik karar alabilme becerilerinin çok yüksek olmaları gerekmektedir. Hiçbir şey en gerektiği anda elde edilen veri kadar değerli değildir. Bunun içinde veri ambarları ve bilgi sistemleri kullanılmaktadır. Gereken bilgiye en hızlı şekilde ulaşabilmek için veri ambarlarındaki karışık bilgiyi işletme anlayışına uygun olarak dönüştürülmesi, özetlenmesi ve derlenip kullanıma hazır hale getirilir (Elbaşıoğlu, 2001: 65-66).

Veri ambarları zaman içerisinde yüzlerce, binlerce, verinin biriktirdiği depolardır. Bu veri ambarlarının en büyük dezavantajlarından biridir. Eğer zaman içerisinde kullanılmayan ve geçersiz veriler üzerinde ayıklama ve temizleme yapılmadığı sürece veri ambarları karışık hale gelmektedir (Arslan & Yılmaz,2010: 13).

Veri ambarları ile veri madenciliği ilişkilerine bakacak olursak, veri ambarları verilerin saklandığı yapılarıdır. Veri madenciliği ise saklanan bu verilerinin bilgiye dönüştürülmesi için kullanılmaktadır.

Bir başka tanımda ise; Bir işletmede gerçekleşen tüm operasyonel işlemlerini en ufak yapıdaki verilere kadar inebilen, detaylı analizler yapabilmesi amacıyla hususi olarak tasarlanan, zamansal derinliği olan ve operasyonel sistemlerden fiziksel olarak ayrı ortamlardaki yapılar üzerinde gerçekleşen süreçlerin toplamına veri ambarı denir (Kılıçaslan,2003: 1).

Veri Ambarlarının çeşitli kullanım amaçları vardır. Bunlardan bazıları aşağıdaki gibidir:

- Tüketicilerin satın alma frekansları üzerinde incelemeler yaparak satın alma alışkanlıklarını ortaya çıkarmak.

- Üretim stratejilerinde profesyonel düzenlemeler için; satış performansı, yollar, coğrafik bölgeler vb. gibi değişkenlere göre karşılaştırarak ürün portföylerini yönetmek.
- İşlemlerin analizi ve yeni kar alanları oluşturmak (Akpınar Ö. , 2018: 108).

Günümüzde işletmeler çok büyük veri tabanları kullanmaktadırlar. Bu tür veri tabanları operasyonel veri tabanları olarak adlandırılmaktadır. Genellikle bu tip veri tabanları tarihsel veriyi depolamak amacıyla kullanılmaktadırlar. İşletmelerde operasyonel veri tabanları bütünlük olarak tasarlanmış haline veri ambarları denmektedir. Veri ambarları milyonlarca veriden oluşmaktadır (Öztürk & Atmaca, 2017: 201-202).

### **3.2.1 Veri ambarının özellikleri**

Veri Ambarlarının 4 ana özelliği vardır: Bunlar zamana bağlılık, kalıcılık, özde-odaklılık ve bütünlük yapısıdır.

- Zaman bilgisi değiştiğinde veri ambarının odak noktasında o anki zamana göre değişmektedir. Veri ambarı, temel veri kaynaklarının belirli periyotlardaki kopyasıdır. Tarihsel veriler bir veri ambarında saklanır. Örneğin, bir veri ambarından 3 ay, 6 ay, 12 ay, 3 yıl veya hatta eski verileri alabilirsiniz. Bu, genellikle en son verilerin tutulduğu bir işlem sistemiyle ilişkilidir. Örneğin, bir işlem sistemi bir müşterinin en son adresini tutabilir, burada bir veri ambarı bir müşteri ile ilişkili tüm adresleri tutabilir.
- Kalıcılık; Operasyonel veri tabanlarından gelen veriler güncellemeden veri ambarına giremezler, güncellenip veri ambarına girdikten sonra ise eski verinin güncellendiği anlamını taşımazlar sadece veri ambarındaki kronolojik olarak yerlerini alırlar. Eski veriler ise yerlerini muhafaza etmeye devam ederler.
- Konu-odaklılık; Veri ambarları belirli bir iş konusunu analiz etmek için oluşturulurlar. Örnek olarak müşteri, siparişler, ürünler ve dağıtım, alım-satım konuları olabilir. Belirlenen konular dışındaki veriler veri ambarına alınmamalıdır. Tek bir konuya odaklı homojen bir veri kümesi oluşmuş olur. Operasyonel veri tabanları daha çok iş süreçlerine odaklanırken veri

ambarları kuruluşun ilişki içinde bulunduğu tüketici ve ürün gibi özlere odaklanır.

- Bütünleşik yapı; birden çok sayıda olabilen operasyonel veri kaynakları ve diğer dışsal kaynaklar birleşerek veri ambarı için bütünleşik bir zemin oluştururlar. Tek bir iş konusu hakkında birleşen verilerin uyumu, beraber hareket edebilmeleri, kendi aralarında bütünleşik olmak zorundadırlar. Bir veri ambarı birden çok veri kaynağından verileri birbirlerine bütünleştirilmelidir.

### **3.2.2 OLAP ve OLTP**

OLAP: Veri tabanı üzerinde yapılan analitik sorgulama fonksiyonlarına verilen isimdir. Hiper veri küpleri yaratarak, ihtiyaç duyulan verilere çok hızlı bir şekilde erişebilme ve güçlü sorgulanmasını sağlamak için geliştirilmiş teknolojik yapıdır (Yılmaz E. , 2005: 138).

OLAP geniş zamanda uzun soluklu analizler için uygundur. OLAP karmaşıklık düzeyi ve veri boyutu yüksek durumlarda kullanılır. OLAP, OLTP tipi sistemlerden verileri alır ve gruplandırır.

İlişkisel veri tabanları üzerinde yapılan verileri düzenleme, silme ve değiştirme gibi işlemler OLTP olarak adlandırılmaktadır. Veri tabanı üzerindeki veriler genellikle ilişkisel tablolar organize edilir ve gereksiz veri yığınları azaltarak veri güncelleme hızı artırılır. Günlük hayatta kullandığımız birçok veri tabanı OLTP tarzı işlem gören veri tabanlarıdır. OLTP sistemleri günlük operasyonel kullanım için uygun bir yöntemdir (Yılmaz & Arslan, 2010: 76).

	OLTP	OLAP
<b>Kullanıcı</b>	İşlemci, BT Çalışanı	Bilgi İşçisi
<b>Fonksiyon</b>	Günlük Operasyonlar	Karar Desteği
<b>VT Tasarımı</b>	Uygulama yönelimli	Amaç yönelimli
<b>Veri</b>	Anlık, İzole	Tarihsel, bütünleşik
<b>Görünüm</b>	Detaylı, düz ilişkisel	Özet, çok boyutlu
<b>Kullanım</b>	Yapısal, tekrarlı	Ad hoc
<b>İş Birimi</b>	Kısa ve basit işlem	Karmaşık sorgu
<b>Erişim</b>	Okuma/Yazma	Okuma
<b>Operasyon</b>	İndeksler odaklı	Sıralı olasılıklarla
<b>Erişilen kayıt sayısı</b>	Onlar	Milyonlar
<b>Boyut</b>	100 MG-GB	100 GB -TB-PB
<b>Kullanıcı Sayısı</b>	Binler	Yüzler

**Şekil 3.2:** OLTP ve OLAP Arasındaki Farklar

**Kaynak:**Akça,M.(2012).”OLAP ve OLTP nedir?” from <http://mustafaakca.com/oltp-ve-olap-nedir/>

### 3.3 Veri Madenciliği Süreci ve Crisp-Dm Süreci

Veri madenciliğinde süreç çok önemli bir kavramdır. Klasik olan bu standart süreç 1996 yıllarının sonlarına doğru Cross-Industry Standart Process for Data Mining (CRISP-DM) konsorsiyumu tarafından belirlenmiştir. Bu konsorsiyumu Daimler Chrysler öncülük etmiştir, veri madenciliği platformu Clementine’i oluşturan SPSS ve NCR’dir. Bu konsorsiyumda en büyük atılımı SPSS yapmıştır. SPSS öncülüğünde birçok firmanın kullandığı bir yöntem haline gelmiştir (Şimşek, 2006: 14).

CRISP-DM bir bakıma veri ile alakalı bir süreçtir. Veri madenciliği süreci ise en küçük yapı veriden başlayıp bunların uygulandığı modelleri de içine alan uzun ve kapsamlı bir süreçtir.

#### 3.3.1 Crisp-Dm süreci

##### 3.3.1.1 Amaç ve hedeflerin belirlenmesi

İş için gerekli olan hedefleri belirlenmesidir. Mevcut durumun değerlendirilmesi için projedeki kaynaklar ve riskler incelenir. Amaçlar belirlenip proje planı oluşturulurken kaynakların hangi adımlara nasıl yönlendirileceği belirlenir (Larose, 2005: 6).

### **3.3.1.2 Veriyi anlamak**

Veriyi detaylı olarak incelenmesi sonucunda projede kullanılacak verinin kalite, kamu ve güvenilebilirliğin olarak ortaya konmasıdır. Böylece yönetilen verinin ne durumda olduğu değerlendirilir (Rokach & Maimon, 2010: 1033).

### **3.3.1.3 Veriyi hazırlama**

Verinin seçilmesi kısmında kayıt altındaki verilerden hangilerinin modelleme için kullanılacağına karar verilir. Verinin temizleme evresinde veri madenciliği sürecine engel olabilecek veriler temizlenir. Verinin inşası sırasında mevcut veriler ile yeni veri içeren alanlar inşa edilir. Verinin entegrasyonu evresinde farklı veri tabanlarındaki veriler birleştirilir. Verinin biçimlendirilmesi evresinde ise veriler modelleme için gerekli biçime getirilir (Larose, 2005: 7).

### **3.3.1.4 Veriye model seçimi**

Farklı modelleme tekniklerinden veriye uygun model seçilir. Kullanılacak teknikler belirlenir. Modelin kalitesine test ile doğrulanacağına karar verilir. Model parametreleri belirlenir. En son olarak test sonuçlarına göre değiştirilir.

### **3.3.1.5 Verinin değerlendirilmesi ve uygulanması**

Modellemedeki sonuçlar iş hedeflerine göre değerlendirilir. Modelleme olumlu bir sonuç geçtikten sonra süreç en baştan bir kalite kontrol olarak değerlendirilen çıktılar ilgili mercilere sunulur. Bu çıktılar yeni bir uygulamada veya daha sonra kullanılacak alternatif bir modelin alt başlığı da olabilir (Rokach & Maimon, 2010: 1033).

Modelin uygulanmasında, modelin teknik performansı kadar cevabı aranan iş ihtiyaçlarına verdiği tatmin edici sonuçlara da bakılır. Teknik performansı ve beklentisi yüksek olan bir model eğer yanlış bir biçimde kullanıma koyulduysa sistem dışı kalabilir (Larose, 2005: 7).

Sonuç olarak üretilen değerler gerekli veri tabanları, karar destek sistemleri ve yönetim bilişim sistemlerine ve CRM sistemlerine entegre edilmelidir.



### **3.3.1.6 Konuşlandırma ve sahaya sürmek**

Bu aşamada gerçekleştirilecek faaliyetlere geçilir. Bu süreçte model analistlere ve son kullanıcılara sunulur. Model ve sonuçların kullanılabilirliği için gerekenler planlanır. Planlama evresinde gözleme ve bakım planlaması detaylıca oluşturulur. Uzun dönemli bir takip sonucunda nelerin daha iyi yapılabileceğine karar verilir (Larose, 2005: 7).

### **3.4 Problemin Belirlenmesi**

Bir veri madenciliği modeli uygulamadan önce en önemli şey problemin belirlenmesidir. Bunun içinde incelenecek verinin doğru olması ve işletmenin hangi problemine cevap arandığının önceden tanımlanmasıdır. Sürecin oluşturulmasında yapılan yanlış tercihler sürecin uzun dönemli sonuçlarına zarar vermektedir.

Kuruluşların hangi hedeflerle veri madenciliğini değerlendireceğini ve neyi amaçladığını, nelere gereksinimi olduğunu bilmesi en önemli evredir. Ortaya konulan hedeflerin, ihtiyaçların, belgelerin detaylı bir şekilde tanımlanması veri madenciliği sürecinin en mühim hamlesidir (Esen, 2009: 22).

İşletmelerin veri madenciliği kullanmasına sebep olan sorunların ve engellerin bu basamakta belirlenebilmesi sağlıklı sonuçlar açısından çok önem arz etmektedir. Problemin belirlenmesinde en önemli etkenlerden birisi işletmenin sahip olduğu bilgidir.

Bu bilgilerin işletmelerin hangi amaç ve amaçları için hizmet edeceğinin belirlenmesi gerekmektedir. Amaçları oluşturma noktasında fiziksel ve teknik olarak veri madenciliği süreci başlamış olur, doğal döngü olarak bu safhada problemin ortaya çıkması ve belirlenmesi beklenir (Özçınar, 2006: 7).

#### **3.4.1 Verilerin hazırlanması**

Veri madenciliğinin en önemli aşamalarından bir tanesi verinin hazırlanması aşaması, analistin vaktini %50 ile %75'ini harcamasına neden olmaktadır. Bu evrede işletmenin sahip olduğu bilgi sistemleri üzerinde oluşturduğu sayısal bilginin akıllıca ayrışması, veriler ile hali hazırda bulunan iş problemlerinin arasında etkili bir iletişim olması gerektiği unutulmamalıdır. Projenin amacı kullanılacak sayısal öğelerin,

hangi görev süreçleri ile meydana getirildiği de bu veriler ele alınmadan analiz edilmelidir. Bu yardım ile analist veri kalitesi hakkında düşünce sahibi olabilir (Tezcanlar, 2007: 88).

Bu aşama verinin düzenlenmesi, ortaya çıkarılması ve bir araya getirilmesi şeklindedir. Veri yaratma evresinin bir bilirkişi kontrolünde gerçekleştirilmesi yaklaşımı Tasarlanmış Deneyimi olarak tanımlanmaktadır. Bir diğer yaklaşım ise “gözleme dayalı” yaklaşım olarak bilinmektedir ve bu yaklaşımda bilirkişi veri yaratma evresinde herhangi bir müdahalede bulunamamaktadır. VM uygulamalarında en çok gözleme dayalı modelinden yararlanılmaktadır.

Verinin hazırlanması şu adımlardan oluşmaktadır;

Toplama, Değer Biçme, Birleştirme ve Temizleme, Seçim, Dönüştürme veri hazırlanması kısmında en dikkat edilecek husus kullanılacak verinin ham veri standartlarında olması gerekmektedir. Yanlış veya eksik veri seti sistematik bir hataya yol açabilir. Veri setinin hazırlanması sürecindeki en zaman alıcı işlemdir (Balint vd. , 2006: 28-29).

### **3.4.2 Model kurulması ve değerlendirilmesi**

Model kurulma ve değerlendirme aşamasında belirlenen problem için en uygun model bulunmaya çalışılır. Birçok model kurularak test edilir. En uygun ya da uygun olan modeller test edilir.

Bir modelin doğruluğunun test edilmesinde kullanılan en basit yöntem basit geçerlilik (Simple Validation) testidir. Bu yöntemde tipik olarak verilerin %5 ile %33 arasındaki bir kısmı test verileri olarak ayrılır ve kalan kısım üzerinde modelin öğrenimi gerçekleştirildikten sonra, bu veriler üzerinde test işlemi yapılır. Bir sınıflama modelinde yanlış olarak sınıflanan olay sayısının, tüm olay sayısına bölünmesi ile hata oranı, doğru olarak sınıflanan olay sayısının tüm olay sayısına bölünmesi ile ise doğruluk oranı hesaplanır (Tezcanlar, 2007: 90).

Önemli bir değerlendirme kuralı ise model veya modellerin anlaşılabilir olmasıdır. Model değerlendirilmesinde test sonuçları doğruluk oranlarında yakın çıktığı zaman, değerlendirme olumlu yönde yapılmaktadır. Asıl önemli olan işletmede uygulamaya

dönük çalışmalar yapıldığı zaman; kararın hangi yönde verildiğinin analiz edilebilmesidir (Hegland, 2001: 313-315).

Model oluşturma ve değerlendirme aşamasında aşağıdaki süreçler izlenmektedir;

<b>Aşama</b>	Model Oluşturma
<b>Görev</b>	Modelleme Tekniğini Seçmek
<b>Çıktı</b>	Modelleme Tekniği
<b>Çıktı</b>	Modelleme Varsayımları
<b>Görev</b>	Sinama Tasarımı Üretmek
<b>Çıktı</b>	Sinama Tasarımı
<b>Görev</b>	Model Kurmak
<b>Çıktı</b>	Parametre Ayarları
<b>Çıktı</b>	Modeller
<b>Çıktı</b>	Model Tanımlaması
<b>Görev</b>	Model Değerlemek
<b>Çıktı</b>	Model Değerlemesi
<b>Görev</b>	Revize Edilmiş Parametre Ayarları

**Şekil 3.3:** Modelin Kurulması

**Kaynak:** Keskin, 2013: 23

<b>Aşama</b>	Değerlendirme
<b>Görev</b>	Sonuçları Değerlendirmek
<b>Çıktı</b>	İş Başarısı Açısından Veri Madenciliği Sonuçlarının Değerlendirilmesi
<b>Çıktı</b>	Onaylanmış Modeller
<b>Görev</b>	Süreci Gözden Geçirmek
<b>Çıktı</b>	Sürecin Gözden Geçirilmesi
<b>Görev</b>	Sonraki Adımları Belirlemek
<b>Çıktı</b>	Olası Eylemlerin Listesi
<b>Çıktı</b>	Karar

**Şekil 3.4:** Modelin Değerlendirilmesi

**Kaynak:** Keskin, 2013: 25

Hangi algoritmaların seçilen modelle uyumlu olup daha iyi bir performans sergilediği değerlendirilir. En uygun algoritma belirlendiği zaman uygulama aşamasına geçilir. Genellikle basit ve tek kullanılan modeller en çok yorumlanandır. Âmâ sonuçların çok azı doğru sonuca ulaşmaktadır. İleri veri madenciliğinde çok boyutlu modeller kullanılarak en doğru sonuca ulaşılmaya amaçlanmaktadır (Hegland, 2001: 313-315).

Analiz edilen ve yorumlanan deęişkenler arasındaki ilişkiyi aktaran grafikler model deęerlendirilmesi için önemli bir bileşkedir.

### **3.4.3 Model uygulaması**

Veri Madencilięi sürecinde model uygulama esnasında belirlenen modellerin seçilip uygulamaya geçilir. Karşımıza çıkabilecek problemlerden birisi seçilen modelin beklenen deęerlerin uygulama aşamasında gözükmemesidir (Akpınar, 2000: 28).

Tanımlanan hedefler için izlenen yolda uygulanan veri madencilięi modelinin sonuçlarını oluşması ile hareket dizisi tamamlanmaz. Özetlere deęer kazandırılması ve tefsir edilmesi sonucu açığa çıkan noksanlıklar, stratejiler, sürecin darboęazları gerçek hayatta uygulanır (Dönmez, 2008: 39).

Düzene oturan bir sistemin ürettikleri veri sebebiyle modelin uzun süreli izlenmesi gerekiyorsa yeniden CRISP-DM sürecine başlanmalıdır.

### **3.5 Veri Madencilięi ve Erdemli Döngüsü**

1990'lı yıllarda Amerika Bankası ev özkaynak kredileri portföyünü genişletmek için can atıyordu ve bunun için doğrudan posta kampanyaları düzenledi. Sonuçlar ise Amerika Bankasının hayal ettięinin tam aksine bir hayal kırıklığı idi. Ulusal Tüketici Varlıklar Grubu sorunu karşı bir mücadele başlattı. Bu mücadelede veri madencilięi kullanılmaya karar verildi. Böylece veri madencilięinde veri madencilięi erdemli döngüsünün açığa çıkmasını sağladı.

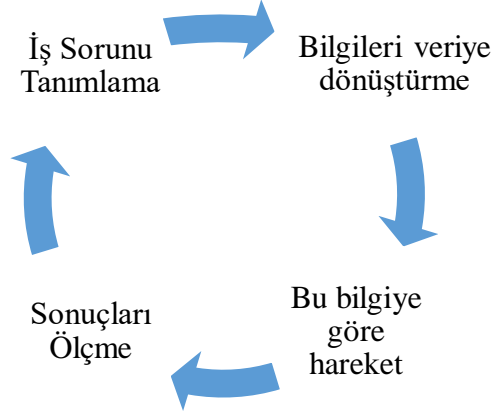
Amerikan Bankasının müşterilerine karlı daha iyi bir pazarlama yapması gerekiyordu. Sağduyulu iş danışmanları üniversite çağındaki gençlerin okul ücretleri ödeyip sahip olabilecekleri ve deęişken gelirlere sahip insanların evlerinin kurulacağı ait tepeleri ve vadileri düzenleyebilmek için özkaynak kullanım planı ile geldiler. Daha sonra Amerikan Bankası Yahoo tarafından satın alınan Hyperparallel adlı özel veri madencilięi reklam şirketi ile çalışmaya başladı. Hyperparallel ile veri madencilięi teknikleri kullanılarak problemin üstesinden gelinmeye çalışıldı (Berry & Linoff, 1997: 14-18).

Amerikan Bankası birbirine paralel durmakta olan bilgisayar ile büyük ilişkisel lokal veritabanlarından milyonlarca geniş ilişkisel verilere sahipti. 42 kayıtlı sistemden temizlenen, düzenlenen ve dönüştürülen veriler ile bir şirket veri ambarı oluşturuldu. Bu veri ambarlarında oluşturulan 250 farklı formlardan müşterilerin demografik özellikleri, çocuk sayıları, oturdukları ev tipleri yanı sıra iç veriler dahil Hyperparallel'in veri madenciliği araçları kullanılarak belirli müşteri profilleri ortaya çıkarıldı.

Daha sonra karar ağaçlarından faydalanılarak veri madenciliği modeli için kullanılabilecek en iyi model oluşturuldu. Bu model ardışık bir model olup, müşterilerin ne tür bir kredi istediklerinin olasılığını hesaplamaktaydı. Ayrıca bu modelin bir diğer amacı geçmişteki sık sık başarısızlıkla sonuçlanan kredi hilekarlıkları ve işlemlerini ortaya çıkarmaktı.

Finale kadar bütün elemeler yapılarak 14 küme kaldı ve bu kümeden bir tanesi oldukça ilgi çekiciydi. Bu küme 2 önemli özelliğe sahipti. Birinci önemli özellik kümedekilerin %39 hem iş hemde kişisel banka hesaplarına sahipti. İkinci önemli özellik ise kötü karar ağaçları sonucu kredi imkanı verilecek müşterilerinin dörtte birinden fazlasını temsil etmekteydi. Veri madenciliği mühendisleri bu insanlara ev kredilerini özkaynaklardan verebilecekleri grupta gördüler.

Bu verilerin yanında elde edilen sonuçların ve bu kadar yapılan istatistiksel işlemlerin hesap edemediği bir şey vardı. Oda insanların oyunculuğu ve duygusal tepkileriydi. İnsanların banka için dürüst yada yardımcı olmalarını gerektirecek zorunlu bir durum yoktu. Yada birikimi olan herhangi biri kapalı olan bir şube yada bozuk olan bir ATM yüzüne başka bir bankayı tercih edebilirdi. Pazarlama çabaları ne kadar yoğun olursa olsun müşterilerin odaklanmış önyargılarını kıramayabilirdi. Tüm bunlara rağmen Amerika Bankası veri madenciliği ile belirlenen yeni pazarlama kampanyası ve müşteri kitlesine ev özkaynakları %0.7'den %7'e sıçrayarak büyük bir başarı elde etmiştir (Berry & Linoff, 1997: 21-23).



**Şekil 3.5:** Veri Madenciliği Erdem Döngüsü

**Kaynak:** Berry & Linoff, 1997: 26

### 3.5.1 Tipik Operasyonel İşletme Sistemleri ve Veri Madenciliği Sistemleri Arasındaki Farklılıklar

Veri madenciliği sistemleri tipik işletme operasyonları sistemlerinden farklıdır. Tipik operasyonlarında raporlamalar ve işlemler tarihsel veriler üzerinde gerçekleştirilmektedir. Veri madenciliğinde ise tarihsel veri analizi genellikle en güncel verilere gelecekteki eylemleri belirlemek için uygulanır. Genellikle takvime bağlı olarak öngürebilir çalışmalar periyodik bir akış şeklinde devam etmektedir veri madenciliğinde ise tahmin edilemez iş akışı, işletme ve pazarlama ihtiyaçlarına bağlı olarak değişmektedir (Alsalloum, 2011: 18-19).

Tipik operasyonları sistemleri kurumsal verilerden yararlanarak iş odaklıdır, müşteriye odaklanmaz. Veri madenciliği ise daha çok ve daha fazla veriyi baz alarak müşteri,ürün ve satış bölgelerine odaklanmaktadır. Tipik sistemleri veri kayıt sistemi üzerinden açıklayıcı ve tekrarlayan veriler ile işlem yapmaktadır. Veri madenciliğinde ise veriler kopyalanarak yaratıcı işlemler yapılmaktadır. (Collins, 2015: 7-10).

Veri Madenciliği dönüşüm ile yapılan verilerin bir sonuca dayandırılması için kullanılmaktadır. Kullanılan veri asla yalın değildir çünkü o veri birçok formdan meydana gelmektedir. O formlarda iç ve dış araştırmalardan oluşmaktadır. Veri madenciliğinde hakkında genel kanı geleneksel madencilik ile bağdaşmaktadır. Madendeki oncalarca taş arasından değerli taşların ortaya çıkartılmasına benzetilir.

Buradaki ortak mantık bir çok müşteri arasından en sorumlu ve seçkin müşteriye ve onun bilgilerine ulaşmaktır. Veri madenciliğinde gerçek amaç veya aranılan cevap şudur;müşteriye ait olan işlenmiş,düzenlenmiş veri ile neler yapabileceğimiz ve bu veriyi nasıl kullanabileceğimizdir (Salame, 2015: 17-21).

Veri madenciliği, veri içindeki gizli örüntülerin keşfedilerek ispat edilmesi maksadıyla gerçekleştirilen çalışmaları kapsamaktadır. Verilerin çok sıkışık olmadığı sistemlerde bu ilişkilerin ortaya çıkarılması için çok yoğun bir çabanın harcanmasına genellikle ihtiyaç duyulmaz, bu sistemlerde ilişkiler genel sorgular aracılığı ile güçlüklerle uğramadan tanımlanabilir. Ancak uzaysal, zamanla ilintili veri kaynakları ve veri ambarı sistemleri gibi verilerin yoğun ve veriler arasındaki bağıntıların komplike bir bağıntılar aldığı sistemlerde, daha aktif araştırmaların yapılması amacıyla kullanılabilir özel yaklaşımlara gereksinim duyulur. Veri madenciliği çalışmaları, işte bu ilişkilerin keşfedilmesi amacıyla dikkate değer yöntemler ve metotlar sunar. Veri madenciliği çalışmalarının mutlaka veri ambarı üzerinden yapılması zorunluluğu bulunmamaktadır. Ancak veri ambarları da veri madenciliği çalışmalarına ihtiyaç duyulan veri modellerinden birisidir (Çağiltay, 2010: 3).

Birçok şirket için Türkiyedeki ve globalleşen dünyadaki milyonlarca insan coğrafik, yaş ve yetenek kitlesine göre ihtiyaçları doğrultusunda ürün servisi yapılmaktadır. Veri madenciliği sayesinde bu müşterilerden en uygun olabilecekler lokasyon olarak kayıt altına alınabilmektedir ve böylece akıllı bir paydaşa sahip olunmaktadır. Bu paydaşlarla iletişim kanalları üzerinden iletişime geçilmekte ve bu paydaşlar içerisinden farklı kümelerde bulunanlara uygun mesajlar yollanmaktadır.

Bazı araştırmacılar ve uzmanlar veri madenciliği metodlarını tekniksel olarak ilginç bulmaktadır. Birçok kişi bu teknikler ile sona yaklaşılacağını düşünmektedir. Teknikler sonsuzluktan veya döngüden çıkılmasını sağlamamaktadır. Bu teknikler iş bağlantılarındaki çıkışlardır. Organizasyon etrafında oluşturulan iş nesnelere veri madenciliği ile ilişkilidir.İş nesnelere ile alakalı problemler genelde müşteri ilişkilerinin bütününe kapsamaktadır.

Ana problem iletişim halindeki potansiyel müşteriler hakkındaki az bilgidir kaynaklanmaktadır. Devamlı olan müşterilerin ilişkilerinde çoklu ürünler çoklu iletişim kanalları üzerinden yürütülmektedir. İş uygulamalarındaki açıklamalar

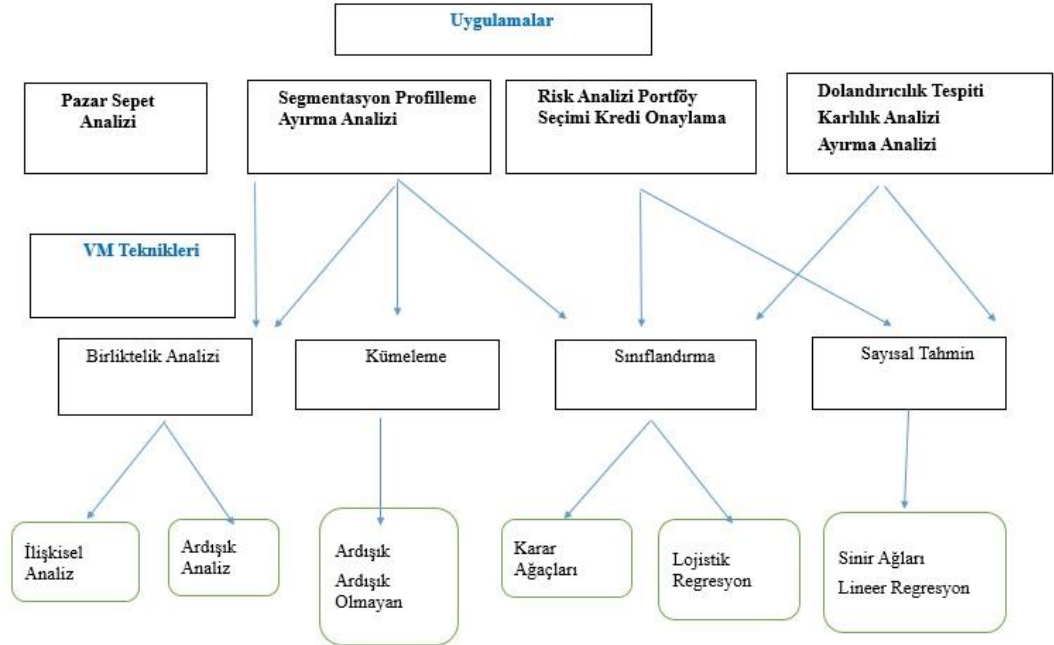
tekniksel materyal kullanma dönüştürülmesi veri madenciliğinin teknik detaylarının spesifik konularında özelliklerinden birisidir.

### 3.6 Veri Madenciliği Metodolojileri, Algoritmaları

Veri madenciliği süreci birbiri ile ilişkili örüntüler ve kurallardan oluşmaktadır. Bu kurallar veri için uygulanan bir tekniğin veya algoritmanın defalarca uygulanmasıyla belirlenir. Kurallar sonucu veri tanımlanmış sınıflara (partition) ya da sonlu sayıda kümelere ayırır (Quinlan, 1986: 81-106).

Modellemeler sonucu elde edilen bilginin kalitesi, modellemede kullanılan algoritmaların uyum ve çözebilme ölçütüne bağlıdır.

Veri madenciliğinde birbirinden farklı ve birçok teknik kullanılmaktadır. Bunların problem çözülmesinde hepsinin kendilerine ve birbirlerine avantaj ve dezavantaj bulunmaktadır. En fazla kullanılanlar karar ağaçları, sinir ağları, kümeleme analizi, k-nearest algoritmaları ve bulanık mantıktır.



Şekil 3.6: VM Metodolojileri ve Algoritmaları

**Kaynak:** Musaoğlu, C. (2003). “Veri Madenciliği İle Tüketici Finansmanı Sektöründe Müşteri Elde Etme Sadakati Modellemesi”, *Yüksek Lisans Tezi*, Boğaziçi Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.



### **3.6.1 Tanımlayıcı modeller (descriptive)**

Karar verme sürecinde veri kümelerinin yeni oluşan iç ve dış veri artıklarından etkilenmemesi için gerekli problemlerin çözümleyen ve tanımlayan modeldir.

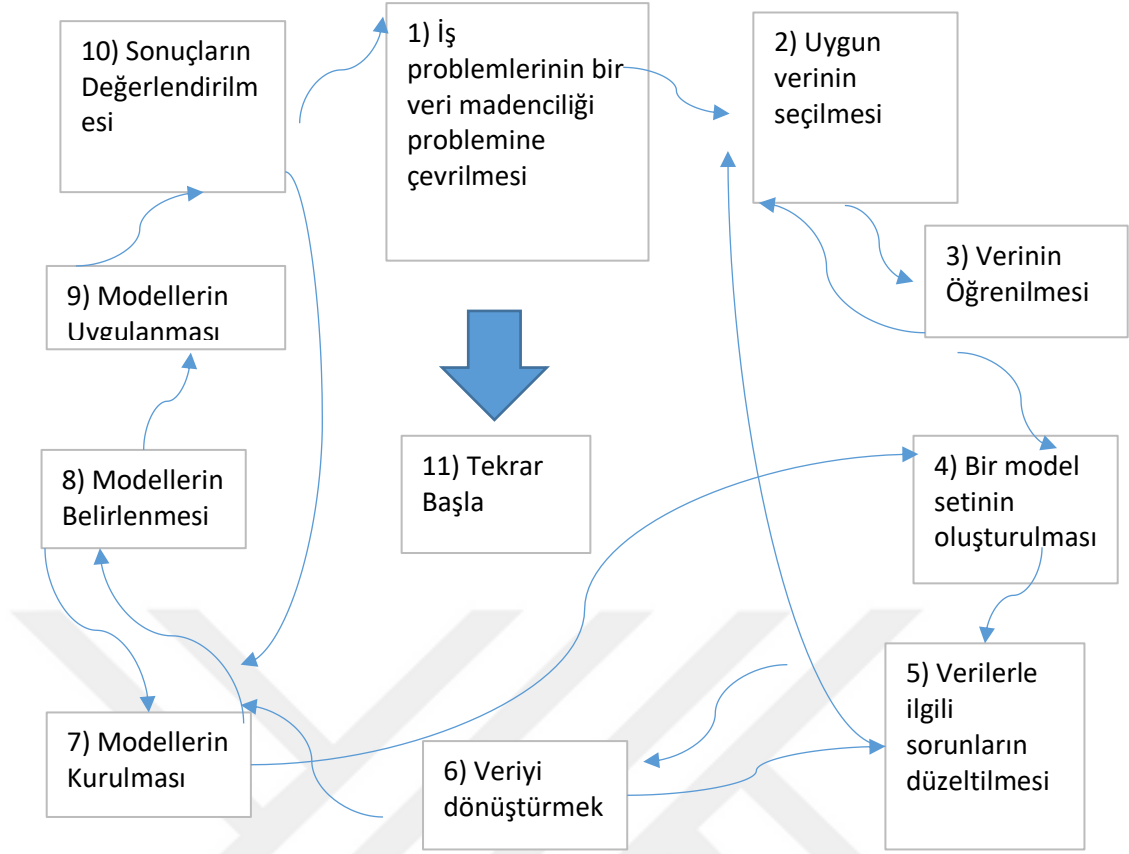
Tanımlayıcı modellerin amacı, karar vermeye kılavuzluk etmede kullanılabilecek var olan verilerdeki örüntülerin tanımlanmasıdır. X/Y eşliğinde geliri ve birden fazla arabası bulunan çocuklu aileler ile çocuğu bulunmayan ve geliri X/Y eşliğinden düşük olan ailelerin satın alma parametlerinin birbirlerine yakınlık duyduğunun belirlenmesi tanımlayıcı modellere bir örnektir (Akpınar, 2000: 6).

### **3.6.2 Tahmin edici modeller (predictive)**

Tahmin edilen modeller sonuçları bilinen verilerden hareket ederek bir model elde edilmesi ve böylece elde edilen modelden faydalanılarak sonuç tahminini amaçlamaktadır. Müşteri İlişkileri Yönetiminde kullanılan VM sürecinde yapay zekâ ve istatistik disiplinlerin etken bir yol izlenmesine olanak sağlayabilmektedir (Şimşek, 2006: 53).Veri madenciliği sürecinde birden fazla teknik bir arada kullanılabilir. Bir teknik çıktının sonuçları başka model için girdi olabilir.

## **3.7 Metodoloji Adımları ve Farklı Teknikler**

Veri madenciliği geçmişte yaptığımız hataları gelecekte tekrarlamayacağımızı göstermektedir. Geçmişte öğrenilen bilgiler için üç temel düşünce vardır. Bunlar; öğrendiğimiz bilgiler doğru değildir, öğrendiğimiz bilgiler doğru ama kullanışlı değildir veya öğrendiğimiz bilgiler kullanılabilir değildir. Öğrendiğimiz bilgileri doğruluğunu ve yanlışlığı, bilgilerin üzerine uyguladığımız metodolojiler ile kanıtlanmaktadır. Kurulan metodolojilerde aşağıdaki adımlar izlenmektedir.



**Şekil 3.7:** VM Metodolojilerinde İzlenen Adımlar

**Kaynak:** Berry & Linoff, 1997: 24

### 3.7.1 Sınıflama ve regresyon modelleri

Veri tabanında var olan verilerden yola çıkarak geleceğin tahmin edilmesinde kullanılan ve VM teknikleri içerisinde en çok tercih edilme oranına sahip olan sınıflama ve regresyon modelleri arasındaki esas ayrım, tahmin edilen bağımlı değişkenin kategorik veya süreklilik gösteren bir niteliği yapısında taşınmasıdır (Seyrek, 2006: 86).

Fakat regresyon modellemesinde özellikle çoklu regresyon modelinde değerlerin kategorize edilebilmesi sebebiyle hem sınıflama hem de regresyon modelleri birbirine oldukça benzemektedir. Bu nedenle son yıllarda aynı tekniklerden ortak olarak faydalanılmaya başlanmıştır (Seyrek, 2006: 77). En çok tercih edilen Sınıflama ve Regresyon Modelleri aşağıdaki gibidir:

- Genetik Algoritmalar
- Naive-Bayes

- Çoklu Regresyon ve Lojistik Regresyon
- Faktör ve Ayırma (Diskiriminant) Analizleri

Sınıflama Modelleri: Kullanılmamış (işlem görmemiş) yeni bir veriyi daha önceden kategorize edilmiş sınıflara atamayı amaçla. Veri tabanında yer alan çoklu örüntüler, bir sınıflama tekniği ile anlamlı alt sınıflara ayrılır. Sınıflama sorguları denetimli öğrenmeye girmektedir. Denetimli öğrenme bir defada olmasa da ikinci yahut üçüncü defada bir sınıfı diğer sınıflardan ayıran örüntüleri bulabilmektedir (Tiryaki, 2006: 16).

Veri tabanındaki yeni bir objenin niteliklerinin belirlenmesi ve önceden tanımlanmış sınıflara koyulmasıdır. Sınıflandırmalar veri tabanlarındaki gibi yaş, cinsiyet, gelir, medeni durum gibi kriterlere göre gruplandırılır (Groth, 2000: 22).

En çok banka kredisi onaylama işlemi, kredi kartı sahteciliği tespiti ve sigorta risk analizinde kullanılmaktadır. Kredi başvuruları düşük, orta ve yüksek risk grubu olarak ayrılmaktadır (Özkan, 2008: 40-45).

Regresyon analizi, herhangi bir bağımlı değişkenin bir veya birden fazla bağımsız değişkenle arasındaki bağımlı matematik bir fonksiyon biçiminde yazılması ve bu fonksiyon etkisiyle bağımlı değişkenin varabileceği değerin tahmin edilmesi işlemidir (Orhunbilge, 1996: 9).

İki unsur arasındaki ilişkiden birindeki kalitatif ve kantitatif değişimler diğerindeki bir değişimle yorumlanıyorsa, ona bağlı olarak diğer değişkende değişiyorsa ve birinci değişkene bağlı olarak yorumlanabilen değişkene bağlı değişken denir.

Örneğin; Bir bankacının gelir düzeyi ile tüketim harcamaları arasında, bir ziraatçının zeytin verimi ile ilaç düzeyi arasındaki ilişkiyi regresyon analizi ile hesaplanabilir. Aralarında doğru bir denklem bulunan bağımlı ve bağımsız değişkene uygulandığında regresyon analizi değişkenlerden herhangi birinin değeri bilindiğinde diğer değer hakkında tahmin yapılmasına olanak sağlar.

Sınıflandırmanın en etkin çözümü dataların sınıflandırılması sonucunda her sınıfta bulunan kişilerin, objelerin ve grupların özelliklerinin açığa vurulmasıdır. Herhangi bir sınıfa ait olmayan veriler bilhassa sayısal verileri tahmin etmek için Lineer Regresyon algoritması kullanılmalıdır. Müşteri İlişkileri Yönetiminde sınıflandırma

bir ürün özellikleri ile müşteri özelliklerinin eşleşmesi için kullanılır. Sonuç olarak müşteri için uygun ve kaliteli müşteri belirlenir.

Zaman Serileri Analizi: Zaman serileri analizlerinde regresyon analizden ayrı olarak yalnızca zamana bağlı olan değişkenler tahmin edilmeye çalışılmaktadır. Örneğin tatil vakitlerine yazın denizlerde meydana gelen boğulma oranları, bir önceki yılın aynı tatil vakitlerinde meydana gelmiş boğulma oranları sayısıyla belirlenmeye çalışılır (Moss, 2003: 307-310).

Tahmin edici modellemelerden lineer regresyon analizi en temel olanıdır. X bağımsız değişkeninin ile Y bağımlı değişkenlerinin arasındaki ilişki doğrusal fonksiyon olarak gösterilir. Birbiri arasındaki etkileşimden birindeki değişim diğerinde de değişmeye neden oluyorsa ve birindeki değişim diğerinin değişiminin girdisini oluşturuluyorsa buna bağımsız değişken denir (Atakurt, 1999: 191).

### **3.7.2 Genetik algoritmalar**

Genetik Algoritmanın temeli Darwin'in Evrim Teorisine dayanmaktadır. İç içe geçmiş ve birbirine karışmış çok boyutlu arama uzayında en iyi ve en güçlünün yaşama tutulabilmesi ilkesine göre bütünsel en uygun çözüme dayanmaktadır.

John Holland ve araştırmacı arkadaşları tarafından ilk kez Michigan Üniversitesinde geliştirilmiştir. Holland, çalışmalarını arama ve optimumu bulma için, doğal seçme ve genetik evrimden esinlenerek gerçekleştirmiştir. 1975 yılında yayınlanan "Adaptation in Natural and Artificial Systems" kitabında mekanik öğrenme mevzuuyla ilgili görüşler açıklamıştır, Darwin'in evrim teorisinde açıkladığı canlılarda bulunan genetik süreci bilgisayar ortamında ortaya koymayı düşünmüştür (Esen,2009: 38).

Genetik Mühendisliğinde yıllar süren araştırmalar sonucu ortaya çıkan genetik araştırmalarla amaç insan ve hayvanların sinirsel yapılarında yer alan birtakım özellikleri veri tabanı yardımıyla, çoklu ortamda taklit ederek istatistiksel kökenli bir sanal ağda modeller oluşturmaktır (Emel & Taşkın, 2002: 129).

Genetik Algoritmalar otomatik programlama, öğrenme kabiliyeti olan robotlar, iktisat, çevrebilim, planlama, sistematik üretim hattı gibi sosyal sistemlerden

evrimsel yönlerini anlamada, fiyat verme stratejileri, gelişim süreci alanlarında kullanılmaktadır.

Genetik Algoritmalar karmaşık problemlerin çözülmesinde kullanılan sezgisel bir yöntemdir. Klasik optimizasyon çözümlerine karşın genetik algoritmalar parametre ilkesi yerine kodlanmış biçimleri kullanırlar. Olasılık kurallarına göre çalışan Genetik Algoritmalar çözüm uzayının tamamının değil belirli bir kısmını tararlar. Böylece daha kapsamlı bir tarama gerçekleştirerek daha kısa sürede çözüme ulaşırlar (Goldberg, 1989: 1-7).

Finansal problemler, montaj hattı dengeleme problemleri, çizelgeleme problemleri, portföy optimizasyonu, tesis yerleşim problemleri, atama ve taşıma problemleri, gezgin satıcı problemleri gibi kapsamlı bir uygulama alanına sahip olan genetik algoritmalar; bilgisayar ağlarının tasarımı, mekanik öğrenme ve ekonomik-sosyal sistem modellerinin ilerletilmesi amacıyla da kullanılmaktadır (Esen, 2009: 41).

Genetik Algoritmalar da bulunan problemlerin çözümünde izlenecek sürece baktığımızda bu sürecin kalıtım yoluyla çözüldüğünü söyleyebiliriz. Bu süreci aşağıdaki gibi sıralanmaktadır;

### **3.7.3 Kodlama süreci**

Kodlama süreci, arama evrenindeki bütün olası detayların genetik algoritmalarla faydalanabileceği bir şekilde çeviren kodlama dizisidir (Yeniay, 2001: 38).

### **3.7.4 İlk popülasyonun oluşturulması**

Belirli bir olasılığa sahip olan çözümlerin kodlandığı çözüm kümesi yaratılır ve çözüm kümesine popülasyon (kodlarda kromozom) ismi verilir. Herhangi bir kromozomun öğelerinden her biri sonucun bir niteliğini belirtmektedir. Bu durum gen ismi ile de anılmaktadır (Goldberg, 1989: 9).

### **3.7.5 Uygunluk değerinin hesaplanması**

Uygunluk değeri, yaşama, çoğalma ve hayatta kalma şansını ilerideki kuşaklarda ki genlerin görülme oranını belirttiğini söyleyebiliriz. Çözüm kuşakları meydana geldikten sonra o çözüm kuşaklarında ki değerlerin uygunluk değerinin hesaplanması gerekir (Yeniay, 2001: 38-39).

Belirlenen detayların uygunluğunu eğer uygunsa, ne düzeyde olduğunu saptanmasına yöneliktir. Her problem için bir fonksiyonun ne şekilde kullanıldığını belirtmek gerekmektedir. Fonksiyonlar problemlerin çeşitlerine göre değişim göstermektedir.

### 3.7.6 Çoğalma işlemi

Bir sınıf serinin belirli bir olasılık oranına göre rastgele olarak seçilip uygunluk değerinin hesaplanmasından sonra, değeri en yüksek çıkan çözüm kuşağının belirlenip, problemler için yeni çözümler işlemi üretilmesidir. Bir anlamda ilk popülasyon oluşturulduktan sonra çaprazlama ve mutasyon işlemleri için daha fazla çözüm kuşaklarının üretilmesidir (Fığlalı & Engin,2002: 3).

### 3.7.7 Yeni kuşağın oluşması ve döngünün durdurulması

Çoğalma, çaprazlama ve mutasyon aşamalarından sonra yeni bir çözüm kuşağı oluşturmaktadır. Bu oluşan yeni kuşak sonradan gelecek kuşağın ebeveynleri durumundadır. Bu süreç bir sonsuz döngüdür. Bu sonsuz döngü belirli bir hedefe veya gerekli kuşak sayısına ulaştığında durdurulur (Yeo & Agyel,1996: 271).

- Naive-Bayes

Bu sınıflandırma tekniği hali hazırda sınıflandırılmış olan verilere öncelik verip yeni bir verinin hali hazırda bulunan kümelerden herhangi birine atanma olasılığını hesaplayan bir tekniktir. (Keskin, 2013: 35) dir.

$$P(C_j) | x = \frac{p(x|C_j)P(C_j)}{p(x)} \quad \longrightarrow \quad P(C_j) | x = \frac{p(x|C_j)P(C_j)}{\sum_k p(x|C_k)P(C_k)}$$

**Şekil 3.8:** Naive Bayes Formülü

**Kaynak:** Clifton, C. & Thuraisingham, B. (2001). *Emerging Standards for Data Mining*. Computer Standards & Interfaces.s. 187,194.

- $P(c | x)$ , verilen sınıfın öngörücüsünün önceden görülme olasılığıdır.
- $P(c)$  sınıfın ön olasılığıdır.
- $P(x | c)$  verilen sınıf tahmin edicisinin olasılığıdır.
- $P(x)$  öngörenin önceden görülme olasılığıdır.

Bu formülde  $C_j$  ve  $C_k$  olarak 2 ayrı belirlenen hipotez iki ayrı sınıfı temsil etmektedir.  $P(C_{j/x})$   $x$ 'in  $C$  sınıfında olma ihtimalini belirlemektedir.  $P(x)$ ,  $x$  değerinin veri tabanı kümelerinde bulunma sıklığıdır.  $P(C_j)$  ve  $P(C_k)$  ise  $C_j$  ve  $C_k$  sınıflarının veri tabanında olma sıklığıdır (Keskin, 2013: 35).

Bayes algoritması, öncelikle kendisine verilen öğrenme kümesinde  $P(C_j)$  değerini, her sınıfın verilen öğrenme kümesi içinde bulunma olasılığını hesaplar. Daha sonra  $x$ 'ler sayılarak  $P(x)$  bulunur. Benzer şekilde her bir grupta, her bir  $x$  değerinin bulunma olasılığı,  $P(x|C_j)$ ,  $C_j$  ler içinde  $x$ 'lerin sayısı ile elde edilir (Silahtaroglu, 2013: 1).

- Lojistik Regresyon Analizi ve Çoklu Regresyon Analizi

Çoklu Regresyon Analizi ile birçok ortak noktası bulunduğu için son yıllarda kullanımı gitgide artmıştır. Bu analiz normallik ve ortak kovaryansa sahip olma durumlarında diskriminant analizi ile beraber kullanılmaktadır (Esen, 2009: 37).

Lojistik regresyon, çoklu regresyondan farklı olarak doğrudan bir olayın gerçekleşme ihtimalini tahmin etmeye çalışır. Olasılık değerleri 0 ile 1 arasında herhangi bir değer alabilmektedir. Lojistik regresyon sıfır ile bir arasında anlamlı bir ilişki kurabilmek için bağımlı ve bağımsız değişkenler arasında; koordinat düzleminde büyük  $S$  olarak adlandırılan bir şeklin olduğunu varsaymaktadır. Bağımsız değişkenin çok düşük değerlerinde olasılık sıfıra yaklaşmaktadır. Bağımsız değişkenin değeri arttıkça var olan olasılık eğri üzerinde de artmaktadır. Fakat daha sonra eğim biri aşmamak kaydıyla azalmaya başlamaktadır (Kurtuluş, 1998: 479).

Yapılan incelemeler sınıfların ayrılmasında kümeleme, diskriminant ve lojistik regresyon analizleri kullanılmaktadır. Kümeleme analizinde verilerin atanacağı küme sayısı tam bilinemez ama diskriminant ve lojistik regresyon analizinde küme sayısı bilinmektedir. Ayrıca belirlenen veriler ile ayırım modeli oluşturularak inceleme ve gözlemlerin gerekli sınıflara atamaları gerçekleştirilir (Tatlidil, 1996: 225-228).

Bağımsız değişken yapısı ve dağılımı bakımından diskriminant analizinden ayrılan lojistik regresyon analizinin doğrusal regresyon analizinden üç önemli farkı vardır:

- Doğrusal regresyon analizinde tahmin edilecek olan birim süreklidir, Lojistik regresyon analizinde ise bağımlı değişken kesikli bir birimdir.

- Doğrusal regresyon analizinde bağımlı değişkenin değeri, Lojistik regresyon analizinde ise bağımlı değişkenin alabileceği değerlerden birinin gerçekleşme olasılığı tahmin edilir.
- Doğrusal regresyon analizinde bağımsız değişkenin çoklu normal dağılım göstermesi şartı aranmaktadır (Elhan, 1997: 4).

Lojistik Regresyon Analizi genellikle tıp, biyoloji, iktisat gibi alanlarda kullanılmaktadır. Örneğin; Sigaranın yaşlanma üzerindeki etkisi, kadınlarda düşük doğum ağırlığını etkileyen risk faktörleri, koroner-arter hastalığına etki eden risk faktörleri gibi klinik vakaların (Şahin, 1999:4-5) yanı sıra hastane, kimya ve biyoloji laboratuvarlarında çalışan personellerin tutum ve davranışlarını ölçme odaklı kullanıldığını söyleyebiliriz.

- Faktör Analizi ve Diskriminant Analizi

Veriler arasındaki ilişkilere dayanan, verilerin daha çok içeriğe sahip olduğu ve özet biçimde sunulmasının şartlarını hazırlayan birden fazla değişkenli bir istatistik analiz tekniğine faktör analizi denmektedir (Massy,1971: 241).

Faktör analizinin gerçek amacı değişkenler arasındaki karşılıklı bağımlılığı ortaya çıkarmaktır. Faktör analizinin yaygın bir şekilde kullanılmasında başlıca faktörlerden bazıları şunlardır; çok sayıdaki veri dizisini azaltabilmesi ve basitleştirebilmesidir. Faktör analizinin bir boyut indirgeme ve bağımlılık yapısını ortadan kaldırma tekniği olduğunu söylenebilir (Özdamar, 2004: 235).

Faktör analizinin önemli varsayımları şunlardır; veri matrisinin analizinin öncesi ölçüt ve tahmin değişkenleri alt matrislerine bölümlendirilmemesi ve değişkenler arasındaki ilişkinin doğrusal olduğudur (Kurtuluş,1998: 482).

Faktör analizi; genellikle müşteri yönelimleri, tüketici seçimleri, müşteri davranışları gibi kişisel tercihler üzerine etkili olmasının yanında pazarlama ve müşteri ilişkileri yönetiminin karşı karşıya kaldığı problemlerin çözümünde bolca tercih edilen bir yöntemdir. Bu teknik iş hayatındaki, satış temsilcisinin kabiliyetlerinin ölçülmesi, şirketin kurumsal imajının ölçülmesi ve marka değerinin belirlenmesi gibi seçeneklerde kullanılabilir (Şimşek, 2006: 29).



Faktör analizi uygulama alanının büyüklüğü bakımından psikoloji biliminde yer almıştır. Bu bilim içerisinde genellikle doğrudan ölçülemeyen ve test edilemeyen konulardaki çalışmalarda kullanılmıştır. Değişken kümelerinin belirgin özelliklerini ortaya çıkartmak için son yarım asırdır bilişim teknolojilerinin kullanımının arttığını söyleyebiliriz. Böylece sosyal bilimler olmak üzere birçok bilim dalında Faktör Analizine rastlamaktayız (Comrey & Lee, 1992: 10).

Faktör analizinin bazı dezavantajları da vardır. Örneğin; Tüm avantajlarına karşın maliyeti oldukça yüksektir, sonuçlarının kesin belli olmaması sebebiyle güvenilirliği sürekli olarak çok yüksek değildir ve yorumlanması daima zor olan bir yöntemdir. Ancak tüm bu eksik veya dikkat edilmesi gereken noktalarına rağmen, özellikle verilerin daha anlamlı ve özet şekilde ortaya çıkardığından çok faydalı sonuçlar doğurabilecek bir analizdir. Bundan ötürü faktör analizinin gerek tek başına gerekse diğer çok değişkenli analizlerle birlikte kullanılması, özellikle davranışsal faktörlerin ağırlıkta olduğu pazarlama sorunlarının çözümüne mühim faydalar sağlayabilir (Kurtuluş, 1998: 492-493).

Faktör analizi birçok bilimde kullanılmasına örnek olarak psikoloji, sosyal bilimler, eğitim bilimleri, sosyoloji ekonomi gibi alanlarda müşteri memnuniyet araştırmaları ve müşteri davranış ölçümlerinin olduğunu görebiliriz. Ancak bu bilimlerde kullanılan Faktör Analiz 'inin sonuçlarının kesin bir örnek ifade ile belirlenememesi başka bir problemdir. Bunun yanında bu teknikle ele alınan veri yapılarının ayrımları yapılırken, kategori oluşturmada karşılaşılan problemler bu analiznin dezavantajlarından (Esen, 2009: 36).

- Diskriminant Analizi

Ayırma Analizi iki veya daha çok sayıdaki grubun ayrımı ile alakalı olan çok değişkenli bir istatistik analiz tekniğidir. Diskriminant analizi ayırma veya ayırmsama analizi olarak adlandırılmaktadır. Ayırım analizi bağımlı değişkenin nominal değeri aldığı, bağımsız değişkenlerin ise metrik değerler aldığı zamanlarda kullanılan en ideal yöntemdir. Birden fazla gruplu bağımlı değişken olduğunda ayırma analizi olarak, grup sayısı üç ve daha fazla olduğunda ise analiz çoklu diskriminant analiz olarak adlandırılmaktadır (Hair vd. , 1998: 244-245).

Birden fazla deęişkeni bulunan problemin tek deęişkenli şekle dönüştürülmesi çabasına ayırma analizi denmektedir. Özetle bütün deęişkenlerin uygun ağırlıklarla katılacağı tek bir fonksiyonun elde edilmesi süreci diskriminant analizidir (Green & Carroll, 1978: 178).

Diskriminant analizlerinin çoęu kadın-erkek, yüksek-düşük, evli-bekâr gibi ikili gruptan oluşmaktadır.

Diskriminant analizi eşitliklerden oluşmaktadır. Birbirlerine en fazla benzeyen yığınları belirlemeye olanak sağlayacak şekilde grupların benzer niteliklerini belirlemek için kullanılmaktadır. Buradaki asıl amaç, gruplar arası varyansın grup içi varyansa oranını maksimum kılmaktır (Fisher, 1936: 179).

Diskriminant Analizinde hangi bağımsız deęişkenlerin grup ortalamalarını üzerine en çok ayırmsamaya sebep olanların bulunması, tanımlanan deęişkenlerin ortalama deęerlerinin arasında istatikselsel olarak bir anlamlılık olup olmadığını ve ayırmsama boyutundan geçtięi halde hangi sınıftan geldięi belli olmayan verinin en uygun sınıfa atanması şeklinde özetlenebilir (Ünsal, 2000: 19).

### **3.7.7.1 Örnek tabanlı yöntemler (K-En yakın komşu)**

Bu yöntemde sınıflama yapılırken eldeki verilerin kendi ve çevreleri arasındaki uzaklık veya benzerlik kullanılır. Uzaklığın hesaplanmasında en çok öklid bağıntısı kullanılır. Uygun uzaklıkta örnek bir ölçüm uzayı belirlenir. Birbirlerine yakın olan noktalar seçilir. Belirlenen grubun en çok rastlanıldığı aralık belirlenerek isimlendirilir. Böylece birbirlerine en yakın gruplar oluşturularak K-En yolun komşu teknięi tamamlanır. En yakın komşu algoritması bellek tabanlı(memory-based) ya da örnek tabanlı yöntemlerde(instance-based) en iyi örnektir (Keskin, 2013: 34).

K-En yakın komşu teknięi kümeleme amacıyla, kümeleme analiziyle beraber kullanılmaktadır. Bir gözlem deęerinin hangi gruba ait olduğunu tahmin etmek için Örüntü uzayını inceleyerek bilinmez gözlem deęerine en yakın bulunan K adet küme sayısı bulunur. Bilinmez deęer ile en yakın komşu içerisinde en fazla ortak noktaya sahip olduğu veri kümesine atanır. Bilinmez gözlem deęeri gerçekte var olan deęerinin tahmini içerisinde kullanılabilir (Çalışkan & Soęukpınar, s. 122).

K-Means yöntemi aşağıda belirtilen adımları kullanarak kümeleme işlemini gerçekleştirir (Özdamar, 2004).

- Elde bulunan verilerin analiz edilmesiyle ortaya çıkacak bilgilerin ışığında belirlenen ilk nokta çekirdek nokta olarak ele alınır. Belirlenen bu noktaların her birine  $p$  değişken vektörleri birer küme ortalama vektörü olduğu varsayılır. Küme ortalama vektöründen her bir birimin uzaklıkları hesaplanır.
- En yakın ortalama kümelere atama işlemi en sona kalan  $n-k$  birim tarafından yapılır. Atama işlemi tekrar edildiğinde ortaya çıkan kümenin ortalama vektörü en baştan hesaplanır. En çok benzerlik oranına sahip sınıflar bir havuzda toplanır.
- Eldeki veri, küme içi varyansın en alt değeri ve kümeler arası varyansın en üst olduğu kümeleme yapısına varana kadar  $k$  adet kümeye atanmaya devam eder.

K-En yakın komşu algoritması, gözlem değerinde oluşan bir küme için aşağıdaki adımlar izlenmektedir;

- $K$  parametresi atanır. Elde ettiğimiz parametre belirlenen herhangi noktaya en yakın bulunan komşuların sayısıdır.
- Bu yöntem koordinatı girilen bir noktaya en yakın komşuları bulacağı için, mevzu bahis nokta ile diğer tüm noktalar arasındaki mesafeler tek tek hesaplanır.
- Elde edilen mesafelere göre satırlar düzenlenir. Bunlar arasından en küçük olan  $k$  tanesi seçilir.
- Belirlenen satırların hangi gruplara ait oldukları belirlenir ve en çok tekrarlanan grup değeri seçilir.
- Tahmin edilmesi hedeflenen gözlem değeri kategorisinin, seçilen kategori olması beklenir.

### 3.7.7.2 Kümeleme modeli

Birbiri ile aynı olmayan verilere sahip kümelerin bulunarak,veritabanındaki kayıtların bu farklı kümelere bölünmesine kümeleme analizi denir.Kümeleme analizi ile kümelerin arasından daha önceden bilinmeyen özellikleri açığa çıkartılmasını hedeflenir.

Kümeleme modelinin kullanımını,küme elemanlarının birbiri ile benzediği durumlarda,niteliklerinin birbirinden farklı olduğu durumlardaki kümelerin

kullanılmasında ve veri havuzundaki kayıtlı kümelerin detaylandırılmasında görülmektedir. Veri havuzunda bulunan kayıtların ne tür değişkenlere göre gruplanacağı ,hangi kümelere göre onaylanacağı ,konuya hakim uzman bir kişi tarafından belirlenebileceği gibi bunu yanında günümüz teknolojisinin gelişmesiyle bilgisayar programlamaları ile yapılabilmektedir (Tiryaki, 2006: 17).

Çok değişkenli sayımlama analizi olan kümeleme analizinin birinci amacı gözlem değerlerini ortak özelliklerine göre kümelemektir. Kümeleme analizinde oluşturulan bir kümenin sahip olduğu gözlem değerleri, başlangıçta belirlenmiş bir nitelik açısından birbirine benzemektedir. Bu sebeple kümedeki gözlem birimleri homojendir (Hair vd. , 1998: 443).

Kümeleme analizi dağınık bir durumda varolan verileri benzer özelliklere göre bir araya toplayarak sınıflandırma işlemidir. Kümeleme analiz ile sınıflandırmalar arasındaki kesin fark kümeleme analizinde kümelerin daha önceden belirlenmemiş olmasıdır. Ayrıca kümeleme analizinde yapay sinir ağları ve istatistiksel metodlar kullanılmaktadır (Silahtaroglu, 2013: 157-161).

Astronomi, tıp, biyoenformatik gibi alanlarda kullanılan kümeleme analizlerinde aşağıdaki yöntemler kullanılmaktadır;

- Bölme Yöntemleri (Partitioning Methods)
- Hiyerarşik Yöntemler (Hierarchical Methods)
- Yoğun Tabanlı Yöntemler (Density-based Methods)
- Izgara Tabanlı Yöntemler (Grid-based Methods)
- Modern Tabanlı Yöntemler (Model-based Methods).

Kümeleme analizinin uygulama aşamaları aşağıdaki gibi verilebilir:

- Birimlerin doğal kümelemeleri hakkında değişmez malumatların bilinmediği popülasyonlardan seçilen n sayıdaki değerlerin p sayıda değişken ile alakalı gözlemlere sahip olunması
- Birimlerin birbirleri ile olan benzerliklerini ya da farklılıklarını belirleyen uygun bir benzerlik ölçüsü ile birimlerin/değişkenlerin birbirlerine olan mesafelerine ulaşılmaya çalışılması (benzerlik ya da farklılık matrisinin bulunması).

- Uygun kümeleme yöntemi yardımı ile benzerlik/farklılık matrislerine nazaran birimlerin uygun skordaki kümelere ayrılması.
- Elde edilen kümelerin tefsir edilmesi ve bu kümeleme sistemine bağlı olarak kurulan hipotezlerin desteklenmesi maksadıyla gerekli analitik tekniklerin uygulanması (Özdamar, 2004).

### 3.7.7.3 Ardışık zamanlı örüntüler

Alış ve satış işlemi esnasında ya da ardışık yapılan alışverişlerde müşterinin talep ettiği ürünü satın almaya ne derece istekli olduğunun anlaşılması, alıcının bu isteğinin artırılmasına yönelik yolların belirlenmesidir. Birlikte ve ardışık zamanlı örüntüler pazarlamada satın alma eğilimleri ve pazar sepet analizi için kullanılmaktadır (Christopher & Blaxton, 1998: 186).

X,Y veya Y,Z olarak nitelendirebilen olaylar sırasına göre şu şekilde açıklanabilir. X,Y veya Y,Z değişkenlerinin birbiri ile olan ilişkisi bilinmektedir. Birbirlerini izleyen dönem aralıklarında aktif olan ilişkilerin tanımlanmasına ardışık zamanlı örüntüler denir. Belirli bir süreç aralığında çoğunlukla meydana gelen olaylar kümelerini bulmaktadır (Alpaydın, 2004).

Örneğin; Bir yıl içinde Orhan Pamuk'un "Benim Adım Kırmızı" romanını satın alan insanların %70'i Buket Uzuner' in "Güneş Yiyen Çingene" adlı kitabını satın almıştır. İMKB endeksi düşerken A hisse senedinin değeri %15'den daha fazla artacak olursa, üç iş günü içerisinde B hisse senedinin değeri % 60 ihtimalle artacaktır (Tiryaki, 2006: 18).

### 3.7.7.4 Yapay sinir ağları (YSA)

1990'ların başından beri sıkça kullanılmaya başlanılan Yapay Sinir Ağları hedef fonksiyon, birbiri ile ilişkili ve etkileşimli sıradan işlemci ünitelerden meydana gelen bir ağ üzerinde yapay sınırlardan oluşmaktadır. Üniteler arasındaki bağlantı ağırlıkları veriler üzerinden öğrenme algoritmaları kullanılarak hesaplanır. İstatistik yöntemlerdeki gibi parametrik bir model kullanılmaz böylece işlem hacmi dahada genişler. İşlem hacmini genişlemesi ile birlikte yüksek işlem ve belleğe gerek duyulmaz (Tiryaki, 2006: 20).

Temeli 1800'lü yılların sonuna dayanan Yapay Sinir Ağları'nın gelişimi 1948 yılında Alan Turing'in araştırmaları ile başlamıştır. İnsan beyninin çalışma sistemine

dayanan YSA'nın mantığı Alan Turing'in "B tipli makineler" kavramının sayısal modellemelerine dayanmaktadır (Öztemel, 2012: 37).

McCulloch ve Pitts 1934 yılında ilk yapay sinir ağı modelini gerçekleştirmişlerdir. Elektrik devreleriyle basit bir sinir ağını modellemişlerdir. Bu basit sinir ağına, insan beyninin hesaplama yeteneğinden esinlenilmiştir (Şimşek, 2006: 38).

Donald Hebb 1949 yılında öğrenme ile ilgili temel teorisini ortaya atmıştır. Daha sonra Frank Rosentblatt ünlü bir laboratuvarda, birden çok vektörün skaler çarpımının aktivasyon fonksiyonuna sokulması prensibine dayanan Perceptron'u 1957 yılında geliştirmiştir. YSA 1987 yılında Elektrik Elektronik Mühendisliği Enstitüsü (IEEE) tarafından "YSA konulu gerçekleştirilen ilk uluslararası konferans" ile bir bilim dalı olarak resmîyetini kazanmıştır (Kotani & Takashi, 1997: 525-527).

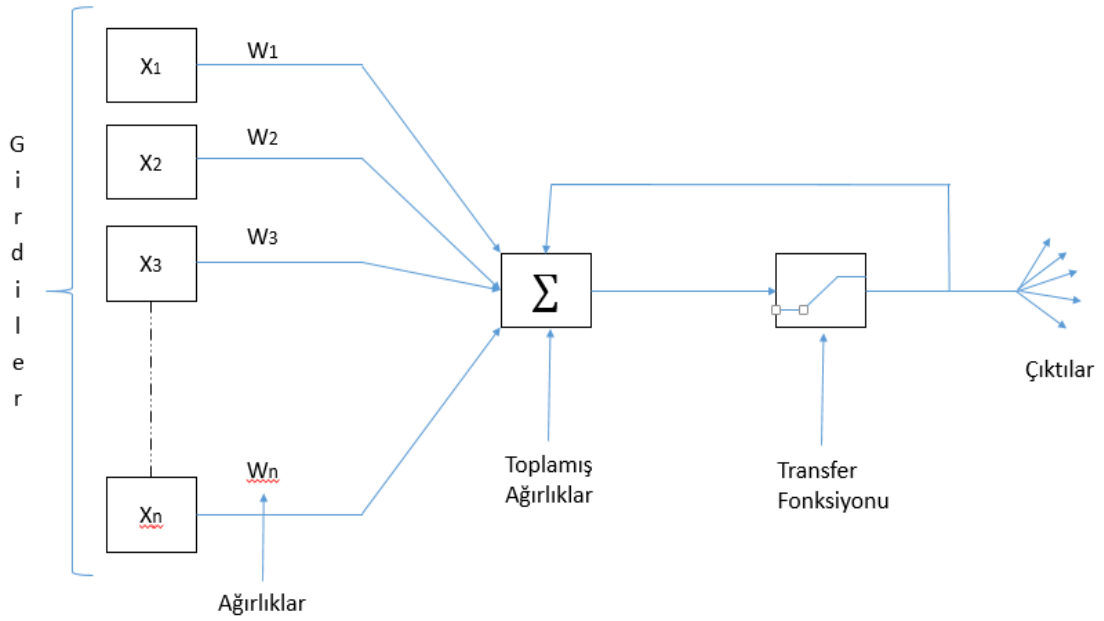
Yapay sinir ağları biyolojik sinir sistemine ait olan niteliklerden öğrenme yolu ile taze bilgiler türetme, yeni bilgiler yaratabilme, analiz edebilme gibi becerileri matematiksel ve istatistiksel bir destek olmadan otomatik bir şekilde meydana getirebilecek bilgisayar sistemleridir. Yapay sinir ağları enformatik konularla ilgilenen bir bilgisayar bilim dalıdır (Öztemel, 2012: 29).

Yapay sinir ağları işleyiş olarak insan sinir sisteminden yola çıkılarak oluşturulmuştur. İnsan sinir sistemindeki akson, dendrit, sinapsis, somadan meydana gelmektedir. Yapay nöron ağlarında ise nöron, girdi, çıktı, önem derecesinden meydana gelmektedir.

<b>Biyolojik Sinir Ağı</b>	<b>Yapay Nöron Ağı</b>
SOMA →	NÖRON
DENDRİT →	GİRİŞ
SİNAPSİS →	BAĞLANTI ARALIKLARI
AKSON →	ÇIKIŞ

**Şekil 3.9:** Yapay Sinir Ağı ve Biyolojik Sinir Ağının Benzerlikleri

**Kaynak:** Yazıcı vd., (2007). Yapay Sinir Ağlarına Genel Bakış. *Türkiye Klinikleri J Med Sci*, s.66.



**Şekil 3.10:** Yapay Sinir Ağı Üzerinde Girdiler ve Çıktılar Arasındaki Ağırlıklar Matrisi

**Kaynak:** Yazıcı vd., 2007: 66

Çıkış,  $o = f(W.X + b)$  şeklinde nöron çıkışı hesaplanır. Buradaki  $W$  ağırlıklar matrisi,  $X$  ise girişler matrisidir.  $n$  giriş sayısı olmak üzere;

$$W = w_1, w_2, w_3, \dots, w_n$$

$X = x_1, x_2, x_3, \dots, x_n$  şeklinde yazılabilir (Ezerçe, 2008: 72).

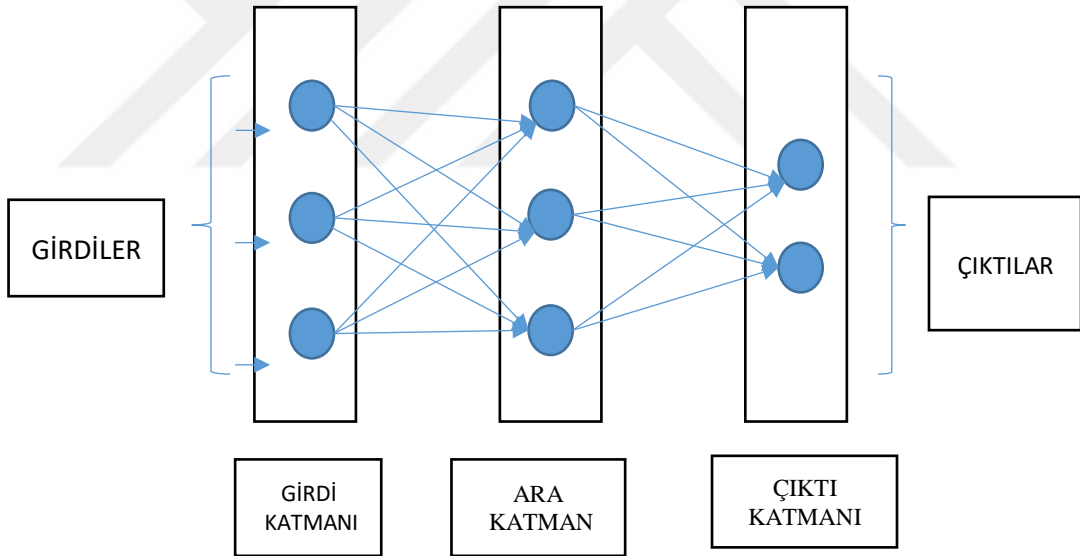
Yapay sinir hücreleri birleşerek Yapay Sinir Ağları meydana getirilir. Girdi sağlamakla yükümlü düğümler bir girdi katmanını oluştururlar. Çıktı sağlamakla yükümlü düğümler ise bir çıktı katmanını oluştururlar. Girdi katmanını ve çıktı katmanını arasında bir ara katman yer almaktadır. Ara katman girdi katmanından gelen bilgilerle işlem yaparak çıktı katmanına göndermektedir (Ataseven, 2013: 103).

Girdi ile çıktı katmanını arasındaki işleyişte en önemli unsur ağırlık setidir. Uygun ağırlık setinin bulunabilmesi için öğretim metodları uygulanmaktadır. Bu iki metod şöyledir:

**Eğiticili Öğrenme (Supervised Learning):** Yapay sinir ağının dışarıdan bir izlenim ile eğitilmesi bu metodunun ana prensibidir. Alınacak çıktıların ne tür giriş değerlerine sahip olacağı önceden bilinmektedir ve böylece yapay sinir ağı ağırlık değerleri bu korelasyona göre yeni değerler alabilmektedir. Bu öğrenme metodunun amacı

beklenen çıktı değeri ile yapay sinir ağının verdiği çıktı değeri arasındaki problemin ağırlıklara kazandırılmasıdır. Bu hata en az değerine ulaşmaya kadar ağ nöronları arasındaki ağırlıkları düzelterek bütünleşmeye devam eder.

**Eğitici Öğrenme (Unsupervised Learning):** Bu öğrenme metodunun Eğitici Öğrenme metodunun tam aksi yönde çalışmaktadır. Yani yapay sinir ağı dışarıdan herhangi bir etki olmaksızın aldığı verileri bu veri grubuna uyumlu bir değer üretecek şekilde düzenlemektedir. Yapay sinir ağı bunu yapabilmek için ilk aldığı örnek/örnekleri bir grup olarak kabul görür. Daha sonra gelecek olan tüm girdi örüntüleri o sınıfa benzetilmeye çalışılır ve böylece tüm girdi örüntüleri kendi aralarında benzerlik gösterip göstermediklerine göre ayrıştırılır. Fakat bu sınıflandırma evresinde yanlış cevaplar üretilmektedir. Yine de girdi örüntüleri sisteme birçok defa öğrenim için beslenirse elde edilen yanlışlama payı %4 gibi makul bir değere düşebilecektir (Aras, 2008: 16).



**Şekil 3.11:** Yapay Sinir Ağlarında Girdi Katmanı ile Çıktı Katmanının İşleyiş İlişkisi

**Kaynak:** Yazıcı vd., 2007: 67

Yapay Sinir Ağları öğrenme, ilişkilendirme, sınıflandırma, belirleme ve optimizasyon gibi konularda uygulandığı gibi veri madenciliği, kredi kartı dolandırıcılarının saptanması, kredi risklerinin hesaplanması, bir ürünün Pazar performansını tahmin etme, kalite kontrolü, üretim planlama, güvenlik ve parmak izi



tanıma sistemleri gibi farklı alanlarda kullanıldığını söyleyebiliriz (Öztemel, 2012: 35).

### **3.7.7.5 Karar ağaçları**

Veri madenciliğinde, veriyi sınıflandırma ve tahmin etmede en çok tercih edilen tekniklerden birisi karar ağaçlarıdır. Yapay Sinir Ağları ile Karar Ağaçları arasındaki en temel farklardan bir tanesi karar ağaçlarının ortaya koyduğu kuralların uzmanlar tarafından daha kolay bir şekilde anlaşılıyor olabilmesidir (Sezen, 2004: 5).

Öncelikle ağaç kökünde veriler oluşturulmaktadır. Karar ağaçları ağaç kökünden yaprağa doğru ilerleyen kurallar bütünü (IF-THEN rules) olarak yazılabilir. Bu şekilde gerçekleştirilen kural çıkarma (rule extraction), VM incelemesinin sonucunun kesinliğini sağlar (Seyrek, 2006: 78).

Kök ve köke bağlı bütün düğümler bir soru kalıbıyla etiketlenmektedir. İlgili soruların olası yanıtlarını ise düğümlerden ayrılan dallar belirtmektedir. Her dal düğümü de sözü edilen sıkıntının çözümüne yönelik bir tahmini temsil eder. Kural çıkarma, veri madenciliği çalışmasının sonucunu doğrulamak için kullanılır (Aras, 2008: 12).

Ortaya çıkan kurallar anlamlılık düzeyi belirlendikten sonra analistler ya da alanında uzman kişiler tarafından çalışma için kullanıldığını söyleyebiliriz.

Karar ağaçları kök, yaprak düğümleri, dallardan meydana gelmektedir. Karar ağaçları bir akış diyagramı şeklindedir. En üst kısma kök düğüm, hiçbir bağlantının bulunmadığı düğüme yaprak düğüm ve geriye kalan bütün düğümlerde ana düğüm denmektedir. Yaprak düğümü sınıf etiketlerini ara düğümler ise niteliyecilerin test değerlerini taşımaktadır (Özekes, 2003: 69).

Karar ağaçları eylemlerden oluşabilecek tüm faktörleri ve bu faktörlere değinen olasılıkları eldeki verilere göre yorumlayan ve bu yorumları geometrik semboller ile karar vericiye sunan düzenleme biçimidir.

Sınıflamada en çok tercih edilen teknik olan karar ağaçları karar verme sürecindeki fazla sayıda olan seçenek sayısını azaltmak ve ardışık aşamalı karar almalarda tercih edilmektedir.

Tahmin edici ve tanımlayıcı özelliklere sahip olan karar ağaçları veri madenciliği projelerinde,

- Karar ağaçlarının oluşturulmasının maliyetli olmaması
- Yorumlanmalarının zahmetsiz ve kolay olması
- Veri tabanı sistemleri ile kolayca uyum sağlıyor olabilmeleri
- Güvenilirliklerinin daha yüksek olması gibi sebeplerden ötürü tercih edilmektedir ve sınıflama modelleri arasında en çok kullanıma sahip olan yöntemdir (Akpınar H. , 2000: 12-13).

Karar ağaçları; arşivlemede vakit ve maliyet kaybını önlemek için arşivlenecek birimlerin gruplandırılması, tümör karakterizasyonu; ilerleyen tarihlerde olguların tahmin edilebilmesi için kuralların ortaya konulması, müşterilerin kredi geçmişlerinden yola çıkarak kredi kararlarının verilmesi, üretim hattında ürün problemlerine sebep olan problemlerin belirlenmesi, hangi ifadelerin satışları etkilediğinin bulunması gibi dallar üzerinde uygulandığında bildiğini söyleyebiliriz (Akiba & Almuallim, 2002: 72-75).

Karar verme ağaçları uygulanan bazı teknikler vardır. Bunlar CART, CHAID, C4.5, QUEST'dir.

- CART

Cart veya C&RT 1984 yılında Breiman, Freidman, Olshen ve Stone tarafından gerçekleştirilen binary (ikili) sistemli büyüyen, sınıflandırma ve regresyon için kullanılan bir algoritmadır. CART karar düğümlerinin iki alt kümeye ayıran ve devamlı olan bu süreçte önceden oluşan kümeler homojen olmak zorundadır (Rokach & Maimon, 2010: 164).

Karar ağaçları, varolan bir karar düğümünden sonra ağacın iki yan dala ayrılır. Ağaçların dallanmaları ikili şekilde gerçekleşmektedir.

C&RT algoritmasında, bir düğümde belli bir kıstas hayata geçirilerek bölünme işlemi gerçekleştirilir. Bunun için önce tüm niteliklerin var olduğu değerler göz önünde bulundurulur. Tüm eşleşmelerden sonra iki bölünme elde edilir. Bu bölünmeler üzerinde seçme işlemi uygulanarak yöntem uygulanır (Özkan, 2008: 117-118).

Ayrıca oldukça karmaşık bir algoritma yapısına sahip olan CART büyük verilerle çalışıldığında yeterli verimi vermemektedir.

- CHAID

1980 yılında Kaas tarafından geliştirilen CHAID (Chi-Squared Automatic Interaction Detector) süreci oldukça başarılı bir karar ağacı yöntemidir. Bu analizde ayırma tekniği olarak Ki-kare testi kullanır. Ki-kare testi tahmin edici değişkenin bütün birimlerini inceleyerek analiz eder. Hedef değişkeni dikkate alarak istatistik olarak aynı olan değişkenleri birleştirir ve farklı olan değişkenlerle işleme devam eder. Diğer adım ise karar ağacının ilk dalını meydana getirmek için en mükemmel tahmin edici değişkeni seçmektir. Bu süreç ağaç büyüyene kadar tekrarlanarak devam eder. Yapılan testler hedef değişkenin sınıfına göre değişkenlik göstermektedir. Sonuç olarak değişken sürekli ise F testi, kategorik ise Ki-kare testi uygulanır (Tezcanlar, 2007: 116).

İkili yani binary sistem olan CHAID algoritması, örnekleme sınıflama yöntemi ve regresyon analizinin beraber uygulandığı en popüler karar ağacı metodlarından biridir. CHAID algoritmasının her türlü değişkenler için kullanıldığından bahsedilebilir.

- C4.5

C4.5 algoritması en iyi karar ağacı metodolojisi olarak bilinmektedir. Quinlan tarafından 1993 yılında ortaya atılmıştır. Karar ağacı oluşturulurken kayıp veriler hesaba katılmaz (Tezcanlar, 2007: 104).

Kalitatif değişkenleri dikkatli bir şekilde izlenmesi C4.5 algoritmasının temel prensiplerinden birisidir. Ayrıca kayıp verileri diğer veri ve değişkenler desteği ile tahmin ederek, daha duyarlı ve daha manalı kurallar oluşturabilen bir ağaç türetebilir (Doumpos & Zopounidis, 2002: 28).

- .QUEST

Yeni bir sınıflandırma algoritması olan QUEST (Quick, Unbiased, Efficient, Statistical Tree), 1997’de Loh ve Shih tarafından bulunmuştur, İkili(binary) şeklinde genişleyen bir ağaç algoritmasıdır. Ayrı ayrı değişken seçimi ve ayırım noktası seçimi

ile ilgilenir. QUEST'deki birim deęişken ayıracı, tahmin olarak tarafsız deęişken seçimini belirlemektedir (Rokach & Maimon, 2010: 165).

### **3.8 Veri Madencilięi Uygulama Alanları**

Azalmakta olan bilgi işlem maliyeti, ham bilginin elde edilmesi ve saklanmasıdaki kolaylık, veri tabanı yönetim sistemi alanındaki gelişmeler ve kullanılabilir analitik gereçlerin giderek artması ile beraber veri madenciliğine olan ilgi gün geçtikçe yükselmektedir (Park, 2001:205).

Son otuz yıldır gelişmekte olan veri madencilięi sektör farkı gözetmeksizin, verilerin toplanıp saklanıldığı veri ambarlarına sahip bütün işletmelerde uygulanabilmektedir.

Veri madencilięi uygulamalarına pazarlama, perakende, finans, sigortacılık, astronomi, biyoloji, tıp, genetik, güvenlik, lojistik, tedarik zinciri, meteoroloji, bankacılık, telekomünikasyon gibi birçok alanda kullanılmaktadır (Rygielski vd. , 2002: 488).

Bu kadar çok alanda kullanılabilmesinin iki ana sebebi vardır. Birinci sebebi veriye sahip bütün sektörlerde veri madencilięi algoritmaları kullanılabilir. İkincisi ise müşteriyi daha iyi tanımak ve buna baęlı olarak müşterinin gereksinimlerini en muntazam şekilde karşılayarak müşteri tatminini sağlar.

#### **3.8.1 Pazarlama**

Günümüz piyasa koşullarında satış ve pazarlama faaliyetleri zaman ve piyasa verilerinin düzenlenmiş hali olan pazarlama yönetimi olarak adlandırılmaktadır. Bu düzenlenmiş bilgilerden yola çıkarak elde edilen bir karar yada kararlar silsilesi vardır. Önemli olan husus bu kararlara doğru bir şekilde varabilmek ve bu kararların hemen gerçekleştirilebilmesidir. Farklı tercihlere ve zevklere sahip müşterilerin gruplandırılması oldukça zordur. Pazarlama stratejileri açısından bireysel paketlerin müşterileri tek tek homojen gruplara bölebilmesi için veri madencilięi teknikleri uygulanmasının öneminden bahsedilebilir.

Kredi kartı işlemleri, indirim kuponları, müşteri şikâyet aramaları, yaşam stili incelemeleri, müşteri ve finans kayıtları pazar analizlerinde kullanılacak verinin kaynağını oluşturmaktadırlar. Gelir düzeyi, harcama tutumları, vb. nitelikler

açısından benzer özellikteki müşteriler için bir model belirlenmesi ve hedef pazarın saptanması; müşterinin fiyat artışı ile değişen satın alma alışkanlıklarının belirlenmesi; çapraz pazar analizleri ile ürün satışları arasındaki birlikteliklerin ve ilişkilerin tanımlanması ve bu verileri kaynak sayarak ürün satış tahminleri yapılması; müşteri profili belirleme çalışmaları kapsamında hangi kümedeki müşterilerin hangi ürünleri satın aldıklarının belirlenmesi; müşteri gereksinimlerinin belirlenmesi kapsamında farklı müşteri türleri için en doğru ürünlerin neler olduğunun belirlenmesi ve yeni müşterileri cezbetmede hangi faktörlerin etkili olacağının tahmini; çok detaylı özetleme raporları ve istatistiksel özetleme bilgileri bu alanda gerçekleştirilen çalışmalardır (Tezcanlar, 2007: 79).

Veri madenciliği pazarlama alanında daha çok Müşteri İlişkileri Yönetimi için uygulanmaktadır. İşletme ve müşterilerinin karşılıklı ilişkilerinin yönetim süreci olarak tanımlanan Müşteri İlişkileri Yönetimi ile en uygun müşteri profillerinin çıkarılması, potansiyel müşteri adaylarının çabuk ve aktif bir şekilde nitelendirilmesi ve fayda profilindeki adayların müşteri haline getirilmesi sağlanabilmektedir (Ezerçe, 2008: 61).

Veri madenciliği uygulamaları genel olarak müşteri profillerinin ortaya çıkarılmasında kullanılmaktadır. Müşteri profili, müşterilerinin ihtiyaçlarını karşılayabilmek amacıyla pazarlama ve satış amirinin stratejilerini ve taktik kararlarının uygulanıp müşteri ara yüzünün oluşturulmasıdır.

Bazı durumlarda müşteriler birden fazla ürün grubu satın alırlar. Bu ürünler arasında bir yargı ya da bağıntı olup olmadığı cevabı aranan sorular içerisinde yer almaktadır. Daha önceden alınan bir ürün ve hizmetin tekrardan alınıp alınmadığı, alınıyorsa alışlar arasında geçen süre “ne kadar gibi” soruların cevapları veri madenciliği uygulamaları ile etkin bir şekilde belirlenerek uygulanmalıdır (Kiremitçi, 2005: 31).

Satış kampanyaları ve özendirme gibi pazarlama faaliyetleri uygulanırken demografik özellikleri ve daha önceki satın alma tutumları yakınlık gösteren müşteri grupları ,veri madenciliğinin kümeleme ve bölümlendirme analizleri kullanılarak belirli bir nitelik bakımından homojen veya benzer özellikler gösteren müşteri grupları ortaya çıkartır. Bu sayede gruplar için birbirinden farklı stratejiler ortaya koyulabilir.

Veri madenciliği pazarlama alanında aşağıdaki eylemler için kullanılmaktadır (Gökmen, 2004: 38);

- Müşteri segmentasyonu çalışmaları (değer bazlı, davranış bazlı segmentasyon)
- Müşteri yaşam boyu değer analizi
- Müşteri sadakati analizleri
- Müşteri elde tutma
- Müşteri kaybı tahminleme
- Sepet analizi ve ürün öneri sistemleri
- Talep tahminleme
- Sosyal ağ analizleri
- Karlılık analizleri
- Fiyatlandırma çalışmaları
- Çapraz satış çalışmaları
- Kampanya performans analizleri
- Kanal optimizasyonu çalışmaları

### **3.8.2 Finans ve Bankacılık**

İşletmenin hedefleri ve politikaları için müşteri veri tabanlarının kullanılması finans sektöründe çok önemli bir rol oynamaktadır. Elektronik ortamda var olan müşterilere ait bütün verilerin kayıt altına alınması analiz için yeterli bilgiye ulaşmayı kolaylaştırmaktadır. Ancak veri sayısının fazla olduğu zamanlarda veri kirliliği ortaya çıkmaktadır. Bu hedefle toplanan bu veri yığınlarının analiz edilmesi, açıklanması, anlamlı raporların ortaya çıkarılabilmesi veri madenciliği ile mümkündür (Dönmez, 2008: 38).

Bankacılık sektöründe veri madenciliği teknikleri genel olarak beş amaç dâhilinde kullanılmaktadır. Bunlar; müşteri segmentasyonu ve profilleme, kredi skorlama, kredi kartı dolandırıcılıklarının tespiti, şirket analizi ve risk yönetimi ve son olarak kapasite planlama tahminidir.

Finans sektöründe veri madenciliği:

- İşletme riskinin azaltılması,

- Risk derecelendirme tahminlerinin yapılması,
- Müşteri profillerinin analiz edilerek çapraz satış tahminlerinin yapılması,
- Finansal ürünlerin veya poliçe türünün tespit işlemleri,
- Karlılık analizleri,
- Firma maliyetlerinin düşürülmesi,
- ATM'lere yirmi dört saat içerisinde kullanılacak para miktarının tespiti,
- Finansal işlemler için düzgün ve etkili karar verebilme,
- Müşteri ve çalışan memnuniyetinin artırılması için kullanılmaktadır (Özel, 2012).

Bankacılık sektöründe ise:

- Farklı finansal analizlerin arasındaki gizli korelasyonların tespiti,
- Müşteri dağılımların belirlenmesi,
- Kredi taleplerinin değerlendirilmesi,
- Kredi kartı dolandırıcılıklarının tespiti,
- Usulsüzlük tespiti,
- Risk analizleri ve Risk yönetiminde veri madenciliği uygulamaları sıklıkla kullanılmaktadır (Kalikov, 2006: 10).

### 3.8.3 Sigorta Sektöründe

Sigorta sektörü finans ve bankacılık sektöründen sonra en çok rekabet ve hilekârın olduğu sektörlerden biridir. Karlı müşterileri firmaya bağlı ve sadık hale getirmek için etkili ve istek odaklı kampanyalara ihtiyaç duyulmaktadır. Bu müşterilerinin hangi nedenlerle sigortalarını iptal ettirdikleri, diğer firmaları tercih ettikleri ve bu sebepten oluşan kayıpların nasıl en aza indirgenebileceği öğrenmek için veri madenciliği kullanılır. İlk olarak karlı müşterilerin tespit edilmesi ve davranışlarının tanınması önemlidir.

Veri madenciliği sigorta sektöründe (Dönmez,2008: 38) ;

- Yeni müşteri kümelerinin ortaya çıkarılması
- Sigorta dolandırıcılıklarının tespit edilmesi
- Tehlike arz eden müşteri kitlesinin gün yüzüne çıkartılması
- Müşteri gruplarına özel poliçe koşullarının iyileştirilmesi

- Prim ödemelerinin özelleştirilmesinde
- Talep tahmininde kullanılmaktadır.

#### **3.8.4 Perakendecilik**

Perakende sektöründe mağaza alışveriş üyelik kartları ve kredi kartlarından elde edilen satış bilgileri, müşteri profili, ürün nakil bilgileri, servis bilgileri gibi birçok veri, veri madenciliği tarafından izlenmektedir. Son yıllarda artan e-ticaret uygulamaları ise perakende sektörü için iyi bir veritabanı oluşturmaktadır. Bu veritabanlarının düzgün kullanılması birçok firma için kurumsal rekabet ve başarı açısından oldukça önem arz etmektedir (Nilgün, 2003: 223).

Veri madenciliği perakende sektöründe:

- Sepet Analizi
- Müşteri Kaybını Engelleme
- Müşteri Alım Profilleri
- Promosyon kampanyaları oluşturulmasında
- Stok optimizasyonunda
- Lojistik ağı performansında kullanılmaktadır (Dönmez, 2008: 39).

#### **3.8.5 Üretim Sektöründe**

- Kalite sektöründe ve kalite ile ilgili sorunların çözümünde
- Üretim hatalarının nedenlerinin bulunması
- Makine arızalarının çözümünde
- Tüketici davranışlarına bağlı olarak modellerin seçiminde
- Tedarikçi seçiminde
- Ürün gamının genişletilmesinde
- Stok maliyetleri analizinde
- Lojistik maliyetlerinin hesaplanmasında
- Ürün fiyatlandırılmasında veri madenciliği kullanılmaktadır (Dönmez, 2008: 39-40).

#### **3.8.6 Elektronik Ticaretinde**

- Web trafiği analizleri



- Müşterinin web sitesi üzerinde gerçekleştirdiği hareketlerin analizi
- Saldırıların amacına ulaşmadan tespiti ve önlenmesi çalışmaları
- Müşteriye ürün öneri sistemleri
- Sosyal ağ entegrasyonları ile müşterinin sosyal medyadaki hareketlerine göre ürün, hizmet özelleştirme çalışmaları
- Trafik optimizasyonu çalışmalarında (Gökmen, 2004: 39) veri madenciliği kullanılmaktadır.

### 3.8.7 Telekomünikasyon

Veri madenciliğinin telekomünikasyon alanındaki kullanım amaçları aşağıdaki gibidir:

- Kalite ve iyileştirme analizleri,
- Hisse tespitleri,
- Hatların yoğunluk tahminleri,
- İletişim desenlerinin belirlenmesi,
- Kaynakların daha iyi kullanılması,
- Servis kalitesinin artırılması (Akbulut, 2006: 10),
- Ses aktarımı, faks, cep telefonu hizmetleri, fotoğraf gönderimi, e-mail, internet erişimi, kullanıcı arama süresi, zamanı vb. çeşitli bilgilerin elde edilmesi,
- Müşteri hesaplarına sahte girişlerin tespit edilmesi,
- Belirli bölgelerdeki kullanımların incelenmesi, verimliliğin artırılması, müşteri sayısının artırılması, servisin kalitesini arttırmak gibi hizmetlerde veri madenciliği kullanılmaktadır.

### 3.8.8 Hilekarların Tespitinde

- Hilekârlığa meyilli olanların bulunması
- Sigorta poliçelerinde yapılan hilekârlıkların tespiti
- ATM, POS makinesi, internet bankacılığı, telefon bankacılığı hilekarlarının bulunması
- Kara para aklama
- Abonelik dolandırıcılığı

- Hilekârları engelleme politikalarının belirlenmesi

### **3.8.9 Tıp Sektöründe**

- Test sonuçlarının tahmini,
- Ürün geliştirme,
- Ameliyat prosedürlerinin belirlenmesi,
- Hastalıkların tespiti,
- Tedavi sürecinin belirlenmesi,
- Hekimlerin teşhis ve tedaviye ilişkin kararlarını etkileyen verilere erişimin sağlanması,
- Magnetik rezonans verileri ile sinir sistemi bölge ilişkilerinin ortaya çıkarılması ve
- Laboratuvar sonuçlarının, röntgen, MR gibi görüntü dosyalarının tek kayıt içerisinde düzenlenmesinde kullanılmaktadır (Seyrek,2006: 84).

### **3.8.10 Web sektöründe**

- Kullanıcı profillerini belirleyip onlara uygun reklam verme
- Kullanıcı isteklerine bağlı olarak web sitelerinin yeniden tasarlanması
- Site içeriğinin ve metninin yenilenmesi

## **3.9 Veri Madenciliğinin Uygulandığı Diğer Alanlar**

Veri madenciliği her geçen gün yeni ve farklı alanlarda kullanılmakla birlikte, günümüzde yaygın olarak kullanıldığı diğer alanlar birkaç kategoride toplanabilir (Tezcanlar, 2007: 82).

- Biyoloji, Tıp ve Genetik: Bitki türleri genetik olarak incelenmesi, gen haritasının analizi ve genetik hastalıkların tespiti, kanser hücrelerinin bulunması, bilinmeyen virüs türlerinin keşfedilmesi, fizyolojik parametrelerin analizi ve değerlendirilmesinde kullanılmaktadır.
- Kimya: Bilinmeyen kimyasal moleküllerin keşfi ve sınıflandırılması, organik yem türlerinin keşfinde kullanılmaktadır.
- Yüzey Analizi ve Coğrafi Bilgi Sistemleri: Bölgelerin coğrafi özelliklerine göre gruplandırılması, şehirlerde yerleşim yerlerini bulma, şehirlerdeki suç

orani, zenginlik-yoksulluk, şehirlere dağıtılacak posta kutusu, otomatik para makineleri, otobüs durakları gibi hizmetlerin konumlarının tespitinde kullanılmaktadır.

- Görüntü Tanıma ve Robot Görüş Sistemleri: Yapay zekâya bağlı robotların görüş sistemlerinden yola çıkarak engel tanıma, yol tanıma, yüz tanıma, parmak izi tanıma gibi tekniklerde de yararlanılmaktadır.

#### Diğerleri

- Uzay Bilimleri ve Teknolojisi: Gezegen yüzey şekillerinin ve gezegen yerleşimleri, yeni galaksiler keşfi, yıldızların konumlarına göre sınıflandırılmasında yararlanılmaktadır.
- Meteoroloji ve Atmosfer Bilimleri: Yağmur kuşağındaki bölgelerde yağış haritaları oluşturma, bölgesel ve yerel iklim tahminleri, ozon tabakası yol açan problemlerin tespiti, denizlerde ve tatlısu kaynaklarındaki belirsiz hareketlerinin açıklanmasında kullanılmaktadır.
- Sosyal Bilimler ve Davranış Bilimleri: Kamuoyu duyarlılığın incelenmesi, geniş trendleri ortaya çıkarma, seçim tahmini oluşturmada kullanılmaktadır.

## 4. PAZAR SEPET ANALİZİ VE BİRLİKTELİK KURALLARI

### 4.1 Market Sepet Analizi

Market sepet analizi birliktelik kurallarının uygulandığı en popüler analiz çeşitidir. Bu analiz, müşterilerin yaptıkları işlemlerde ürünler arasındaki birliktelikleri veya aralarındaki bağı tespit ederek müşterilerin satın alma alışkanlıklarını ve tüketim eğilimlerini analiz eder. Bunun yanında sepet analizi tekniği, çok basit şekilde hangi ürünlerin birlikte alındığını, hangi ürünlerin hangi kampanyaya girmesi gerektiğini, tüketim davranışlarını ortaya çıkarır ve ürün kombinasyonlarının birbirleri ile yakınlıklarını belirlemek için sepet verilerinin analizini yapar. Bu analiz sonucundan yola çıkarak yöneticiler müşterilere daha etkin satış stratejisi uygulamak ve promosyon teklifleri geliştirmek için plan yapıp, uygulayabilirler (Alan, 2016:46).

Market sepet analizi geçmiş tarihli işlemleri inceleyerek birbiri ile ilişkili ürünleri ve ürün gruplarını bulmak için uygulanan bir analiz çeşitidir. 2000'li yılların başına kadar satış yapıldığı anda değil, belirli bir zaman aralığında (günlük, haftalık, aylık, yıllık) gerçekleşen satış hareketlerinin bütününe ilişkin genel veriler elektronik ortamda saklanmaktaydı. Teknolojinin ilerlemesi ile barkod okuyucu sistemlerinin gelişmesi ve böylece bir harekete ait verilerin satış aksiyonlarını direk o anda toplanması ve elektronik ortama aktarılması mümkün olmuştur. Barkod sistemleri en çok süpermarketlerde uygulandıkları için ve satış odaklı mekanizmadan geldikleri için bu veri seti analizinin ismi market sepet analizi olarak kalmıştır. Market sepeti verisinde yer alan bir kayıta, fiş numarası, işlem tarihi ve satın alınan ürünlere ait ürün kodu, miktarı, fiyatı gibi bilgiler bulunmaktadır (Han & Kamber, 2000: 244).

Birliktelik kurallarının ana hedeflerinden biri sepet analizinde satış kapsamında, müşteri tarafından alınan ürünlerin arasındaki örüntüleri bulmak ve buna bağlı olarak kuralları ispatlamaktır. Örnek vericek olursak; “alkol” ürünlerini alan müşterilerin çoğunlukla “çerez” ürünlerini de aldıkları biliniyorsa ve bir müşteri “alkol” ürünlerini alıyor ve “çerez” ürünlerini almıyorsa, bu müşteri potansiyel bir “çerez” ürün müşterisi olduğundan bahsedilebilir.

Pazar sepet analizi genellikle ticari anlam barındıran verilerinden oluşan ancak veri seti üzerinde hangi ilişkilerin aranılacağı bilinmediği konumlarda bir hareket noktası olarak kullanılabilir. Böylece veri setinin içerisindeki bir takım kalıplar sayesinde satış karını arttırmak üzere farklı kampanyalardan yararlanabilir (Karagöz, 2007: 26).

#### **4.2 Birliktelik Kuralları**

Müşterilerin ürün alırken karar verme davranışlarını saptamak amacıyla farklı uygulamalar ile çözümlenmeye gidilmeye çalışılmıştır. Bu çözümlenmelerde veri havuzları ve üst üste bağıntılar kullanılmıştır. Müşterinin belirlediği ürün ve servislerin içerisinde hangilerinin daha fazla ön planda olduğunu belirlemede tercih edilen bu tip yöntemlere pazar sepet analizi denmektedir. Birliktelik kuralı olarak tanımlanan yöntemde ise müşteri tercihlerini saptamada etkin bir metot olarak kullanılmaktadır (Brijs, 2004: 7-9).

Birliktelik analizi alışveriş esnasında satışa sunulan birden fazla ürünün, hangi müşteri tarafından alınmasını belirlemek ve ölçmek için kullanılmaktadır. Birliktelik kuralları tüketici davranışlarına büyük ölçüde ışık tutmaktadır. Bu kurallar yöneticilere daha etkin bir satış geliştirme imkânı sağlar.

Birliktelik kuralları mühim hükümlere varılmasına destek sağlamaktadır. Birliktelik kuralları market satış analizleri, müşteri alım alışkanlıkları, hangi ürünlerin indirimde konacağı, katalogların nasıl tasarlanacağı, raftaki ürünlerin nasıl dizileceği vb. gibi problemlere çözüm olarak kullanılmaktadır (Esen, 2009: 33).

Sepet analizi ile varılan sonuçlar sonucunda satış işlemini planlama, katalog ve raf görselinde verilecek olan kararlar değişebilir. Örneğin; Makyaj malzemesi alan bayanların raflarının yanına diş temizliği ürünleri yerine makyaj temizleme ürünleri koyulmasına karar verilir. Çünkü sepet analizinde hesaplama yapılırken mal ve hizmetler arasındaki bağlantı, destek ve güven ölçütleri baz alınmaktadır. Destek ve güvenilirlik seviyesine göre hesaplanan Birliktelik Kuralları verinin potansiyel ilişkilerini belirtmektedir. Güçlü bir Birliktelik Kuralı büyük bir destek ve yüksek seviyede güvenilirliğe sahiptir (Aras, 2008: 27).

Birliktelik kurallarının kullanışlı olması için iki önemli husus vardır: Bunlardan biri kuralın konu ile ilgili olması bir diğeri ise anlaşılabilir olmasıdır. Anlaşılabilirlik her zaman birliktelik kurallarının güçlü yönü olmuştur bunun sebebi ise birliktelik kurallarının simgesel ve sezgisel yapıda olmasıdır. Birliktelik kurallarında, kullanıcının kuralların tipini ve sayısını kontrol edebileceği çeşitli yollar bulunmaktadır. En çok tercih edilen yöntem, eşik değerleri olarak bilinen minimum destek ve minimum güven değerlerinin belirlendiği yöntemdir. Bu yöntemde araştırmacı tarafından belirlenen eşik değerlerinden büyük olan destek ve güven değerlerine sahip kurallar bulunur ve kullanılır (Döşlü, 2008: 27).

Bu yöntemde bir işlem yani nesnenin tanımı uygulamaya bağlıdır. Müşterilerin aldığı ürünler ve gerçekleştirilen tüm işlemler ve bu işlemler sırasında beraber alınan nesnelere kümesi market sepet analizini oluşturmaktadır. Bu analiz uygulanırken kullanılan bazı önemli terimler vardır bunlar şöyledir; kuralın sol tarafını ifade eden önce (antecedent), kuralın sağ tarafını ifade eden sonuç (consequent), destek değeri, güven değeri, minimum destek olarak gösterilen minimum destek değeri, minimum güven olarak gösterilen minimum güven değeri, nesneküme, yaygın nesnekümesi ve aday nesnekümesi olarak sıralanabilir (Dolgun, 2006: 36).

Veritabanında bulunan birimlerin aynı anda ve aynı sıklıkla gerçekleşme kriterlerini göz önünde bulundurarak birimler arasındaki gizli bağlantıları keşfetmekte kullanılan Birliktelik Kuralları (Association Rules) 1993 yılında Agrawal tarafından geliştirilmiştir ve bu yöntemin veri madenciliğinde kullanılmasının en önemli ölçütü kural çıkarımıdır. Diğer veri madenciliği modellerinden farklı olarak araştırma başlangıcında herhangi bir araştırma problemi bulunmamaktadır. Bunun yerine araştırmacı tarafından belirlenen minimum destek değeri (Minimum Support Value/ MDD) ve minimum güven değeri (Minimum Confidence Value/ MGD) bulunmaktadır. Bu kısıtlar içerisinde verilerden kurallar üretilir ve sonucunda güvenilir ve geçerli kurallar oluşturulmaya çalışılır. Birliktelik kuralları içinde güvenilir bir ilişki tanımlamak için genellikle Apriori algoritmasına başvurulmaktadır (Albayrak & Koltan, 2009: 38).

Birliktelik kuralları oluşturmada kullanılan yöntem oldukça kolay ve basittir. Bu yöntem şöyle basit bir tanımlama ile açıklanabilir; Her bir sık geçen nesne kümesi X ve boş olmayan altkümeleri Ia için en düşük güven değerinden büyük ve eşit olan  $Xa \Rightarrow X - Xa$  şeklinde olası kurallar meydana getirilir. Tanımlamaya göre oluşturulan

bütün kurallar en düşük destek ve güven değerine uymak zorundadır (Gürgen, 2008: 21).

Analizin taşıdığı avantajlara rağmen bazı dezavantajları da barındırmaktadır. Bu dezavantajlar; kayıtların sayısına göre yapılacak hesaplamalar gitgide çoğalmaktadır. En uygun netice, tüm ürünlerin kayıtlar arasında hemen hemen aynı frekansa sahip olması durumunda meydana gelmektedir. Fakat yöntemin kayıtlarda nadir rastlanan ürünleri es geçmesi en iyi sonucu ortaya çıkarmasına mani olmaktadır. Analizde MDD, MGD, destek ve güven değerleri aracılığıyla kurallar oluşturulmaktadır. Eşik değerlerinin sayesinde kural sayısına sınırlama getirilebilir. Ancak, eşik değerlerinin çok düşük seviyelerde tutulması, analizin ilgilenmekte olduğu kuralları kaybetmesine sebep olmaktadır (Erdoğan, 2004: 20).

Birliktelik kurallarının uygulandığı analizlerde satıştaki ürünlerin hangilerinin birlikte tercih edileceğine yönelik tüketici davranışları belirlenip, bu bilgi satış ve pazarlama departmanlarına katkı vermektedir. Bütün işlemler sırasında müşterinin tercih ettiği tüketim gruplarını ve ürünleri inceleyerek kategorileştirmeye imkân sağlar. Hızlı tüketim sektörü çaprazlama ve kategorileştirme yöntemleri ile müşteri davranışı ve eğilimlerine büyük önem vermektedir (Arasteh & Arbab, 2016: 13).

### **4.3 Birliktelik Kurallarının Belirlenmesinde Kullanılan Temel Algoritmalar**

Birliktelik kuralları sıkça rastgelen nesne kümelerinin belirlenmesi ile ortaya çıkmaktadır. Birliktelik kuralları oluşturulurken kullanılan algoritmalar sıralı ve paralel olarak ikiye ayrılır. Yaygın nesne kümelerinin bulunması sırasında nesne kümelerinin paraleleştirilmesi işlemine paralel algoritmalar denir. Sıralı algoritmalar ise nesne ismine bağlı olarak nesne kümelerinin sözlük sıralı olarak tanımlandığı ve depolandığı bir algoritma çeşididir. Bu algoritmanın önemli bir avantajı ise nesne kümelerinin üretilmesi ve sayılması sırasında basit bir varsayım izlemesidir.

#### **4.3.1 AIS algoritması**

AIS algoritmasının amacı veri tabanında bulunan tüm kayıtların birçok kez taranması ilkesine dayanarak yaygın nesne kümelerini meydana getirmek için Agrawal, Imielinski ve Swami tarafından 1993 yılında yayınlanmış ilk algoritmadır. Karar destek sorgulamaları gerçekleştirmek için veri tabanlarının fonksiyonlarını arttırmaya odaklanmıştır.

Bu algoritmanın kodlama mantığı şu şekildedir; Sürekli bir tarama halindedir. İlk defa yapılan taramada her bir öge için destek değerleri bulunarak veri tabanında ne derece yaygın oldukları hesaplanır. En çok tekrarlanan ögeler bulunur ve ikinci taramaya geçer. İkinci taramada, ilk tarama sırasında yer alan yaygın nesne kümeleri arasındaki ortak nesne kümeleri ve tespit edilir. Keşfedilen bu ortak kümeler işlemde olan diğer ögeler ile tekrar bir kümede toplanır aday öge kümeleri oluşturulur. Nesne kümelerinden 1 ve sıralamadan olan 1'daki nesnelere büyük olan ögelerin alfabetik sıralaması ile elde edilen ögeleri genişletilir. Tahmin ve kümeleme tekniklerine başvurulup, aday nesne kümelerden lüzumsuz niceliklere sahip olanlar elenir. Elde edilen her bir aday nesne kümesinin destek değerleri hesaplanarak, değeri ile maksimum destek değeri karşılaştırılır. Bu karşılaştırma sonucunda maksimum destek değeri sahip ve büyük değerlere sahip kümelere dâhil edilir. Bu döngü daha çok ürün bulunmayana kadar devam etmektedir (Dunham vd. , 200: 6-7).

#### **4.3.2 SETM algoritması**

SETM adlı algoritma yaygın öge kümelerini kontrol etmek için SQL sorgulama dilinde Houtsma ve Swami tarafından 1993 tarihinde geliştirilmiştir. Bu algoritmanın kodlama sistemi iki sıralama ve birer birleştirme-tarama görevlerinin bir araya getirildiği tek bir döngüden meydana gelmektedir. Önce birleştirme-tarama ve daha sonra destek değerlerini belirlemek için ikinci sıralama geçilir. Birliktelik kurallarının üretilmesi için Rk değerleri üzerinde kolay bir sıralı tarama gerçekleştirilir. AIS algoritmasında olduğu gibi maksimum destek değerini sağlayamayan Rk değerleri küme içerisinde temel alınmaz. Ck değerleri kurallar için bilgi sağlamaktadır (Houtsma & Swami, 1993: 29-31).

SETM algoritması istatistiksel olarak AIS algoritmasında olduğu gibi veri tabanını birçok kez taramaktadır. SETM algoritması, AIS algoritmasından farklı olarak ise yaygın öge kümelerindeki her bir nitelik iki değerden oluşmaktadır. Bu değerler, niteliğin adı ve nitelikleri özel kılmak için kullanılan hususi kodlardır. Algoritma çalıştırıldığında bu kodlar TID olarak adlandırılmakta ve bunlar hafızada saklanmaktadır. Algoritma, küme için bekleyen öge kümeleri için bağıntısal birleştirme metotları kullanmaktadır. Bu kümeler meydana getirilirken, aday nesne kümelerini TID kodu dâhil birer tane sıralı yedeğini oluşturmaktadır. Sırası ile alfabetik sıralanarak küçük kümeler silinir. Veritabanı TID bilgisine göre



sıralanmışsa, yaygın öge kümeleri daha sonra yapılacak taramada  $C_k$ 'yı TID sırasına göre sıralanarak oluşturacaktır. Veritabanı birçok defa taranmaktadır. Bu işlem, yaygın öge kümeler bulunamayana kadar devam etmektedir. SETM algoritmasının TID değerlerini depolaması, algoritmada alan karmaşasına sebep olmaktadır (Agrawal & Srikant, 1994: 492-494).

### 4.3.3 Apriori algoritması

Agrawal ve diğerleri tarafından 1993 yılında geliştirilen Apriori algoritması, veri madenciliği tarihinde birliktelik kurallarının çıkarılması konusunda elde edilmiş büyük bir başarıdır. Birliktelik kuralları çıkarımında en çok kullanılan algoritma olmuştur. Algoritma adını yaygın nesnelere önsel bilgilerini kullanmasından yani bilgileri bir önceki adımdan almasından almaktadır. Bu teknik, yaygın bir nesnekümesinin tüm altkümeleri de yaygın olmalıdır kuralına dayanmaktadır. Diğer önemli bir husus bir nesnekümesindeki nesnelere sözlük sıralı olduğunu varsayar. Diğer algoritmalarından farklı olarak, aday nesnekümelerinin üretilme şeklinde ve sayılacak aday nesnekümelerinin seçilmesinde farklılık arzeder. Apriori algoritması bu önemli nokta üzerinde odaklanır. Apriori, önceki geçişte oluşan yaygın nesnekümelerini birleştirerek aday nesnekümelerini oluşturur. Veri havuzundaki hareketlerle ilgilenmeden, önceki geçişte oluşan altkümelerden küçük olanlarını veri havuzundan siler (Döşlü, 2008: 34).

Apriori algoritması bu önemli noktaya hitap etmektedir. Apriori, aday ürün kümelerini, bir önceki döngüde elde edilen büyük ürün kümelerini birleştirerek ve veri tabanındaki işlemlere bakmaksızın bir önceki döngüde küçük ürün kümeleri olarak elde edilen alt kümelerin silinmesiyle oluşturur. Apriori döngüsü şu şekilde oluşmaktadır;

Apriori döngüsünün çözüm şekli ise tekrarlı bir çözüm önerisi ortaya koymaktadır. Esas alınan nesne kümeleri bir alt kümelerini içerisine katarak büyümesi ile gerçekleştirilir. İlk adım olarak ilk katmanlı nesne kümeleri oluşturmak amacıyla katılma ölçütleri tanımlanır, alt kümedeki geçişlerdeki sıkça kullanılan öğeler geçtiği nesnenin adayını oluşturarak kümeleştirmek için tercih edilir. Veri setindeki önemli incelemeler yapılarak katılma ölçütleri tanımlanır böylece azalana kadar sırasıyla işlemler yinelenir (Agrawal & Srikant, 1994: 487).

Bu iteratif işlem, daha fazla büyük ürün kümesi bulunamayınca kadar sürer ve döngü tamamlandıca biter. Algoritmanın her  $i$  döngüsü veri tabanını bir kez tarar ve  $i$  büyüklüğündeki büyük ürün kümelerini belirler.  $L_i$ ,  $i$  büyüklüğündeki büyük ürün kümelerini,  $C_i$  ise  $i$  büyüklüğündeki aday ürün kümelerini ifade etmektedir.

Agrawal ve Srikant tarafından modifiye edilen yöntembilim, 1994 senesindeki yirminci uluslararası çok büyük veriler sempozyumunda anlatılmıştır (Agrawal & Srikant, 1994: 489).

Çalışmada, algoritma detayları ve kaba kodlamaları şöyle verilmiştir:

- a) Verilerin başlangıçtaki araştırılması sırasında en derin şekilde tüketim gruplarının durumunun belirlenebilmesi amacıyla bütüncül sayım gerçekleştirir.
- b) Ardışık yoklama,  $k$  sırasındaki tayin olarak iki kısımdan oluşur.
- c) Güncelleştirilmiş bağıntı tercih edilerek, bir eksiği sıradaki tarama sonucunda ürünlerin eldesindeki gruplar ile adayın ürünü kümeler oluşturulur.
- d) Ardından veri havuzunda gerekli incelemeler yapılarak aday katkısının sağlanması hedeflenir.
- e) Yavaş olmayan sayısal hedef amacıyla; giriş değerindeki verilerin hesaplamasında  $C_k$  adaylarının tayin edilmesi son derece önemlidir.

Birleşim ve azaltma işlemleri, alt küme ve esas kümelerin bulgu oluşturabilecekleri şekilde yapılmaktadır. Bu işlem basamaklarından ilki birleşim esnasında hedef çıktı kümesini tanımlayabilmek için giriş öğelerin nesne kümeleri, birincil alt kümelerle bir araya getirilir. Bu teknikte bütün seviyelerde giriş ve çıkış gruplarının isme göre alfabe esaslı sıralanmasını sağlamaktır.

**Çizelge 4.1:** Notasyonlar

<b>k-nesneküme</b>	<b>k nesne içeren bir nesneküme</b>
$L_k$	Sık geçen $k$ ögeli nesnekümeler (minimum destek değerine göre)
$C_k$	$k$ ögeli aday nesne kümeler (potansiyel sık geçen nesnekümeler)

Çizelge 4.1’de ikinci adımda  $C_k$ ’nın eleman grupları sıkça tekrarlanan ya da tekrarlanmayan durumu söz konusudur. Ama  $C_k$  tüm sıkça anılan nesne kümeleri içerir ve  $L_k$ ’nın üst kümesidir.  $C_k$ ’daki her bir adayın destek değerinin hesaplanması için yapılan taramanın sonuçları  $L_k$ ’yı meydana getirir.

$C_k$  çok büyük boyuttalarda olabilir ve çok ağır hesaplamalar gerektirebilir.  $C_k$ ’nın büyük derecedeki hesaplama boyutunu düşürmek amacıyla Apriori metodu tercih edilmektedir. Sıradan bir sık olmayan  $(k-1)$  nesne küme sık geçen  $k$ -nesne kümenin altkümesi olamaz. Bu sebepten ötürü, aday  $k$ -nesne kümenin bütün  $(k-1)$  ögesi altkümeleri  $L_{k-1}$ ’de yoksa bu aday sık geçen nesne küme olamaz ve  $C_k$ ’dan çıkarılır (Han & Kamber, 2000: 422-423).

**Çizelge 4.2:** Bağlı Verileri

TID	Nesneler
1	Ekmek, Süt
2	Ekmek, Çocuk bezi, Bira, Yumurta
3	Süt, Çocuk bezi, Bira, Kola
4	Ekmek, Süt, Çocuk bezi, Bira
5	Ekmek, Süt, Çocuk bezi, Kola

**Kaynak:** Ay, D. ve Çil İ, (2008). Migros Türk A.Ş.’de Birliklik Kurallarının Yerleşim Düzeni Planlamada Kullanılması. *Endüstri Mühendisliği Dergisi*, **21** (2) ss. 14-29.

Yukarıdaki çizelgede bir D data havuzuna ait işlemler bulunmaktadır. Data havuzunda beş basamak vardır,  $|D|=5$  olarak ifade edilmektedir.

Apriori Algoritmasının adımları aşağıda anlatılmaktadır (Dürek, 2018: 41);

1) Algoritmanın ilk basamağında her bir 1- nesne kümeleri  $C_1$  aday kümesinin elemanı olarak yer almaktadırlar. Algoritma, her ürünün destek değerini hesaplamak için tüm hareketleri basitçe tarar.

2) Destek eşik değeri 3 olarak belirlenmiştir Bu arada eşik değeri şu şekilde hesaplanmaktadır. O sütunda satın alınan ürünler / satın alınmayan ürünlerinde dahil olduğu o sütunda bulunan bütün ürünlerin toplam sayısı şeklinde özetlenebilir.  $(3/5=60\%)$  .1 öge barındıran kümeler ile sık geçen  $L_1$  kümesi oluşturulur. Bu örneğe göre  $C_1$  kümesindeki 4 adayın destek değerleri minimum destek değerinin üzerinde olduğu görülmektedir.

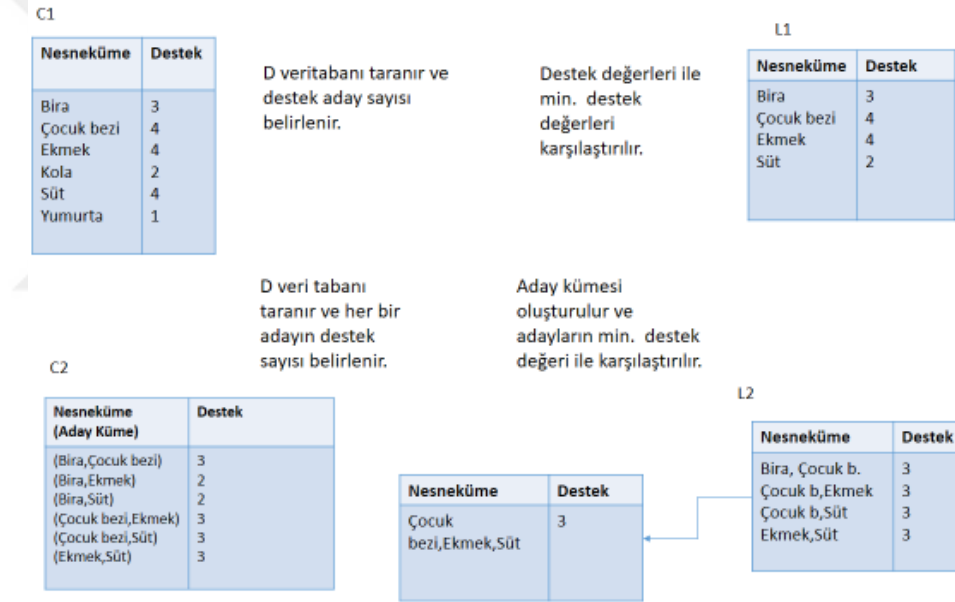
3) 2 ögeli sık geçen nesne kümeleri (L2) algoritma 2 adet L1 kümesinin birleşimini baz alır ve çift ögeli nesne kümeleri (C2) oluşturur. C2, L1'in ikili kombinasyonlarından oluşur.

4) D veri havuzundaki yapılması gereken tektikler yapıp ve C2 kümesindeki nesnelere birbirleri ile olan ilişkilerinde bağ saptanır.

5) C2 kümesindeki en az katkı miktarındaki çift ögeli nesne kümeleri L2'yi meydana getirir.

6) Üç ögeli adaylar kümesini (C3) meydana getirmek amacıyla  $L2 \times L2$  birleşimi baz alınır. Ayrıntıları belirtilen 3 ögeli adaylar kümesi  $C3 = L2 \times L2 = \{\{\text{Çocuk bezi, Ekmek, Süt}\}\}$  şeklinde elde edilir.

Apriori algoritmasının adımları ise şekil 4.1'de bir örnek olarak görülmektedir.



Şekil 4.1: Apriori Algoritmasının Adımları

**Kaynak:** Ay & Çil, 2008: 14-29

1) Bu tabloda D veri tabanını taradığımızı görmekteyiz ve C3 kümesindeki nesnelere destek değerleri belirlenir.

2) 4 ögeli adaylar kümesini (C4) oluşturmak için  $L3 \times L3$  birleşimi kullanılır. L3 kümesinde tek bir nesne küme olduğundan dolayı C4 boş kümedir. Bundan dolayı algoritma sık geçen bütün nesne kümeleri bulduğundan sonlandırılır.

**Çizelge 4.3:** Ürün Güven Endeksi

<b>Çocuk bezi ? Ekmek ? Süt</b>	<b>Güven=3/3</b>	<b>%100</b>
Çocuk bezi ? Süt ? Ekmek	Güven=3/3	%100
Ekmek ? Süt ? Çocuk bezi	Güven=3/3	%100
Çocuk bezi ? Ekmek ? Süt	Güven=3/3	%75
Ekmek ? Çocuk bezi ? Süt	Güven=3/3	%75
Süt ? Çocuk bezi ? Ekmek	Güven=3/3	%75

**Kaynak:** Ay & Çil, 2008: 14-29

Çizelge 5.3'e bakacak olursak en alt seviyedeki durum yüzde 85 seviyesinde tanımlanırsa diğer kurallar zayıf olmayan birliktelik şeklinde ifade edilmektedir.

#### **4.3.4 Apriori-TID algoritması**

Apriori algoritması destek değerlerini her geçişte veri tabanını tarar ve bu işlem gereksizdir. Bundan yola çıkarak Agrawal ve diğerleri, Apriori-TID ismiyle bir fonksiyonel kod geliştirmişlerdir. Apriori ile ortak noktası aday nesne kümelerini belirlemek ve aday üretimi işlemini gerçekleştirmektir. Apriori-TID'nin en büyük farkı ise ilk geçişten sonra destek değerleri için veri tabanını tekrar tekrar saymaz. Ck aday kümelerinin şifrelenmiş şeklini kullanır. SETM'deki gibi Ck'nın her bir üyesi <TID, Xk> biçimindedir, TID tanımlayıcı, Xk ise potansiyel yaygın k-nesnekümesidir (Srikant vd. , 1996: 495).

Apriori-TID ilk geçişten sonra kodlanmış kümenin kullanılması ve daha sonra bu kümenin gitgide küçülmesinin etkisiyle hızlı bir çözüm sunması çok makuldür. Ama nesne kümelerinin çok büyük boyutlarda olması SETM'de olduğu gibi bu algoritmada vardır. Nesne kümenin çok büyük boyutlarda olması Apriori uygun bir ortamdır. Apriori daha büyük nesne kümelerinde, Apriori-TID ise daha küçük nesne kümelerinde çözüme başarıyla ulaşmaya yatkındır.

#### **4.3.5 Apriori-hybrid algoritması**

Apriori-Hybrid algoritması verilerle alakalı tüm döngülerde aynı algoritmanın kullanılmasının zorunlu olmadığı fikrine dayanır. Agrawal ve diğerleri tarafından 1994 yılında geliştirilmiş ve ilk geçişlerde daha iyi performansa sahiptir ve sonraki geçişlerde Apriori-TID, Apriori'den daha iyidir. Apriori-Hybrid tekniği, ilk geçişlerde Apriori'yi kullanacak şekilde tasarlanmıştır. Ck kümesinin bellekte

karşılıyarak Apiori-TID'e geçiş yapar. Her geçiş sonunda Ck kümesinin tahmin edilmesi gerekir ve bu işlem uğraştırıcıdır. Bu algoritmanın performansı büyük verisetleri üzerinde denenmiş ve geçiş sonlarında herhangi bir deęişim olması durumunda Apiori'den daha iyi bir performans sergilendięi belirlenmiştir (Srikant vd. , 1996: 496).

#### 4.4 Güven (Confidence) ve Destek (Support) Ölçüleri

Birliktelik kurallarında ve market sepet analizinde satılan ürünler arasındaki ilişkiyi ortaya koymak için iki tane ölçüt kullanılmaktadır. Bunlar Destek (Support) ve Güven (Confidence) ölçütleridir. Her iki ölçütün yüzdesi ne kadar yüksekse birliktelik kurallarının da o derece güvenilir olduğuna karar verilir.

Destek ölçütü A ve B ürün gruplarının birliktelik içeren alışveriş sayısının, veri tabanındaki toplam hareketlerin sayısına yüzde cinsinden oranı olarak açıklanır. Matematiksel gösterimi aşağıdaki gibidir. N tüm yapılan tüm alışveriş hareketlerini göstermektedir.

$$\text{destek}(A \rightarrow B) = \text{sayı}(A, B) / N$$

#### Şekil 4.2: Destek Ölçüsünün Matematiksel Formülü

Bundan dolayı, eğer bir kuralın desteğinin %10 olduğunu varsayarsak, bu toplam hareketlerin &10'unun A ve B'yi kapsıyor demektir. Destek, bir birliktelik kuralının istatistiksel anlamıdır. Birliktelik kuralları için genellikle yüksek bir destek istenirken bazı özel durumlarda tahminleme ve aradaki bağıntıyı bulmak içinde daha az değerler istenebilir (Gürgen,2008: 21).

Kural ölçütü A ürününü alan müşterilerin B ürününü de alama olasılığını yüzde cinsinden gösteren ölçüt çeşididir. Matematiksel gösterimi aşağıdaki gibidir. Sayı(A) kümesi (A) ürününü alan satın alan müşteri sayısını temsil etmektedir.

$$\text{güven}(A \rightarrow B) = \text{sayı}(A, B) / \text{sayı}(A)$$

#### Şekil 4.3: Güven Ölçüsünün Matematiksel Formülü

Bundan dolayı, eğer bir kural %85 güvene sahip ise, bu A ürününü kapsayan hareketlerin %85'i B'yi de kapsıyor demektir. Bir kuralın güveni, A ve B arasındaki veri kümelerinin karşılıklı ilişkisinin derecesini gösterir. Bir birliktelik kurallında güven ne derece yüksekse birliktelik kuralının gücünde o derece yüksektir. Bir veri tabanından, birliktelik kurallarının veri madenciliği yoluyla çözülmesi, kullanıcı tarafından belirlenmiş güven ve desteği karşılayan tüm kuralların bulunmasını içerir (Gürgen, 2008: 21).

#### **4.5 Konu ile İlgili Literatür Taraması**

2013 yılında Engin Oğuz Ay tarafından yazılan "Veri Madenciliği ile geliştirilen bir akıllı buzdolabı ve market sepet analizi sistemi" başlıklı doktora tezinin amacı tahminleme modellerini kullanıp bir buzdolabı kontrol sistemi içinde bulunan kritik seviyeler için algoritmalar geliştirmesi, uygulanması ve yorumlanmasıdır. Tez çalışmasında veri madenciliği paket programlarından Weka kullanılmıştır. Bir süper marketin satış verileri üzerine Apriori algoritması ve birliktelik kuralları uygulanmıştır. Birliktelik kuralları sayesinde daha etkin analizlerin yapılması ve marketteki rafları düzenlenerek uygun bir maliyet programı izlenmesi hedeflenmiştir.

Özge Seda Seyrek 'in "Müşteri İlişkileri Yönetiminde Veri Madenciliği ve Bir Uygulama " adlı doktora tezinin amacı MİY 'in işleyiş sürecine VM'nin verimi ve önemini incelemek ve ortaya koymaktır. Tüm bu konular arasında ilişkisel bir bakış açısı sunmaktır. MİY kapsamında kampanya yönetimi ve sadakat programları ile Crisp-dm süreçlerine değinilmiştir. Koç şirketler grubundan Tanı Pazarlama A.Ş 'nin müşteri ilişkileri ve veri madenciliği çalışmaları incelenmiştir. Paro sisteminin amacı müşteri portföylerini genişletmek ve müşteri sadakat seviyelerini en üst düzeye çıkarmaktır. Birliktelik analizi kuralları ile ortaya çıkan ilişki kurallarına göre hedef kitle belirleme ve kampanya süreçlerinde bu birliktelik kurallarından faydalanılacaktır.

Diğer bir örnek ise Burcu Alan'nın 2016 yılında yazdığı "Veri Madenciliği ve Market Veritabanında Birliktelik Kurallarının Belirlenmesi" adlı çalışmasıdır. Bu çalışmada SPSS Clementine 11,1 paket programı ve veritabanı için Microsoft SQL Server 2005 kullanılmıştır. Bir marketin 7 günlük süreçte fiş hareketleri izlenmiş. Apriori algoritması sonucunda bu ürün grubu sayısı 20 adet ürün kategorisine

düşürülmüştür. Ortaya çıkan birliktelik kurallarına göre işletmeye daha etkin bir raf düzeyi ve bu raf düzeni sistemine dayanarak müşterilerin satın alma davranışlarının değiştirilebileceği öne sürülmüştür.

Ayşegül Ezerce ve Mehpere Timor tarafından Şubat 2011 yılında İstanbul Üniversitesi işletme fakültesi dergisinde yer alan“ Müşteri Profili ve Alışveriş Davranışlarını Belirlemede Kümeleme ve Birliktelik Kuralları Analizi: Perakende Sektöründe Bir Uygulama” çalışmasının verileri hazır giyim sektöründe faaliyet gösteren bir firmanın alışveriş kayıtları ile alışverişi gerçekleştiren müşterilerden elde edilmiştir. Bu veriler SPSS Clementine paket programı kullanılmıştır. Öncelikle birliktelik kuralları ile müşteri alışveriş alışkanlıkları belirlenmeye çalışılmıştır. Ortaya çıkan sonuca göre işletme için uygun Pazar stratejisi, satış, reklam ve özendirme faaliyetleri temel alınarak üretimin satış temelli olarak düzenlenmesi önerilmiştir.

Fachrul Kurniawan Bilgi Mühendisliği ve Veri Bilimi dergisinde Haziran 2018 yılında yer alan çalışması işlem verileri yoluyla müşteri davranışlarını tanımlamak için pazar sepeti analizinin kullanımınıdır. Bu çalışmanın amacı çalışmada kullanılan yöntem için veri madenciliği tekniğiyle ilgili mevcut adımları takip etmektir. Yapılan test ve analizde, BC İş Merkezinin Malang şubesinin süpermarket işlem verileri üzerinde apriori algoritması kullanılarak birliktelik kuralı üzerinde pazar sepeti analizinin geliştirilmesi ve uygulanması, iyi performans gösterebileceği sonucuna varılmıştır.

Manpreet Kaur ve Shivani Kang'ın 2016 yılında Procedia Bilgisayar Bilimi dergisinde “Pazar Sepeti analizi: Piyasa verilerinin değişen trendlerini belirlemede Birliktelik kuralında veri madenciliğini kullanma” ile alakalı bir çalışma yayınlamışlardır. Bu çalışmada veri madenciliğinde kullanılan birliktelik kurallarının uygulanarak ortaya çıkan sonuçlara göre müşteri davranışlarının incelenerek satış oranlarını arttırılması hedeflenmiştir. Ortaya çıkan manuel eşik değerleri ile büyük ölçüde çalışmayı etkilemiştir bu yüzden daha iyi eşik değerleri için değerlerin otomatikleştirilmesinin gerekli olduğu sonucuna varılmıştır. İkinci olarak aykırı değerlerin tanınarak pazara yönelik bir başka araştırma seçenekleri doğurabileceği öne sürülmüştür.



## **5. PAZARLAMA VE MÜŞTERİ İLİŞKİLERİ YÖNETİMİNDE VERİ MADENCİLİĞİ UYGULAMALARI: BİR UYGULAMA**

### **5.1 Araştırmada Verilerin Toplanması**

Araştırmada kullanılan veri seti İstanbul Yeni Havalimanı'nda perakende sektöründe faaliyet gösteren yurtdışı menşeli bir çanta ve giyim firmasının mağazalarından bir tanesinin, 2019 Mayıs ve Haziran ayları içerisindeki 61 günlük zaman diliminde toplanan alışveriş hareketlerinden oluşmaktadır. Gizlilik prensipleri nedeni ile mağaza ismi veya ürün markaları çalışmada belirtilmeyecektir.

Toplanan fişler SPSS 22 girilmiştir. Veri girişi yapılırken ürün kaydında True/False ya da 1/0 sistemi kullanılmaktadır. Ürünlerin sepete alınma durumu "1" ile alınmama durumu ise "0" ile anlatılmıştır.

### **5.2 Anakütle ve Örneklem**

Belirli bir zaman diliminde yapılan bir gözlem söz konusu olduğundan ana kütleimiz sonsuz sayıdadır. Bu süre içerisinde toplanan 1650 fiş içerisinde uygun elemeler yapılarak fiş sayısı 1550'e düşürülmüştür. 1550 fişte gözlemlenen 78 ürün grubu bulunmaktadır. Daha sonra bu ürün grupları birleştirilerek, birliktelikleri daha iyi tanımlayabilmek için 42 ürün kategorisi oluşturulmuştur.

Örneklemin uygulandığı mağazada fişlerin toplandığı dönem içerisinde ürünlerin sahip oldukları kampanya ve indirimler müşterilerinin alım tercihlerini etkileyebileceğinden bahsedebiliriz. Çalışmanın gerçekleştirildiği mağazanın bulunduğu havalimanı bulunduğu bölge ve hitap ettiği hedef kitle nedeni ile alışveriş alışkanlıkları farklılık gösterecektir.

### **5.3 Araştırmanın Yöntem ve Algoritması**

Veri setini oluşturan ürünler arasındaki ilişkinin bulunmasında bir birliktelik kuralı tekniği olan "Pazar Sepet Analizi" kullanılmıştır. Bu teknik için apriori algoritması,

bir önceki bölümde ayrıntılı olarak anlatıldığı üzere, diğer algoritmalara olan üstünlüğü ve yaygın olarak kullanılması nedeni ile tercih edilmiştir.

Analizde SPSS Clementine 13 programı kullanılacaktır. SPSS ailesinin veri madenciliği süreçleri içerisinde Birliklilik ve segmentasyon algoritmaları ile bazı Python ve Apache Spark algoritmalarında yer aldığı bir yazılımdır.

#### 5.4 Araştırmanın Amacı

Yapılan araştırmanın amacı veri toplulukları içerisinde birbirleri ile yakın ilişki içerisinde bulunan ürün kümelerini bulmaktır. Bu yakın ilişkilere dayanarak müşterilerin satın alma alışkanlıklarının ortaya çıkarılması ve böylece en çok beraber satın alınan ürünler ile tercih edilmeyen ürünlerin belirlenerek daha sonraki kampanya ve pazarlama stratejilerinin geliştirilmesi hedeflenmiştir.

Veri madenciliği modellerinden birinin uygulanmasına dayanarak ve literatür taramasındaki örneklerde baz alınarak veri madenciliği uygulamalarından Birliklilik kuralı ve Apriori algoritmasının pazarlama, perakende ve müşteri ilişkileri yönetimine olan olumlu ilişkisi ispatlanmaya çalışılmıştır.

#### 5.5 Araştırmanın Bulguları

Araştırmanın SPSS 22 üzerinde oluşturulan veri setinde demografik verilerle alakalı sonuçlar aşağıda sıralanmıştır. Cinsiyet, yaş, ödeme şekli frekans değerleri hem tablo ve histogram olarak paylaşılmıştır. Müşterilere ait ülke ve bölgesel bilgiler çalışması yürütülen mağazanın uluslararası bir pasaport gizliliğine sahip olması sebebiyle araştırmaya eklenememiştir.

Cinsiyet dağılımı aşağıdaki gibidir;

**Çizelge 5.1:** Cinsiyet Dağılımı

		Cinsiyet		
	Frekans	Yüzde	Geçerli Yüzde	Kümülatif Yüzde
Erkek	787	50,8	50,8	50,8
Kadin	763	49,2	49,2	100,0
Toplam	1550	100,0	100,0	

Katılımcıların %50,8 çok az bir çoğunlukla erkeklerden oluşmaktadır. Kadın sayısı erkeklerden 24 kişi daha azdır.

Ödeme seçeneklerinin dağılımı aşağıdaki gibidir;

**Çizelge 5.2: Ödeme Şeklinin Dağılımı**

		Ödeme Şekli		Geçerli Yüzde	Kümülatif Dağılım
		Frekans	Yüzde		
Değer	Nakit	585	37,7	37,7	37,7
	Kredi Kartı	655	42,3	42,3	80,0
	Çoklu Ödeme	310	20,0	20,0	100,0
	Toplam	1550	100,0	100,0	

Ödeme şekilleri arasında kredi kartı birinci sırada, ikinci sırada nakit ödeme ve son sırada ikisinin beraber kullanıldığı karma ödeme yöntemi vardır.

Yaş dağılımı ise aşağıdaki gibidir;

**Çizelge 5.3: Yaş Dağılımı**

		Frekans	Yaş		Kümülatif Yüzde
			Yüzde	Geçerli Yüzde	
Değer	18-25	472	30,5	30,5	30,5
	26-34	444	28,6	28,6	59,1
	35-46	192	12,4	12,4	71,5
	47-56	177	11,4	11,4	82,9
	57-65	165	10,6	10,6	93,5
	65+	100	6,5	6,5	100,0
	Toplam	1550	100,0	100,0	

Yukarıdaki grafikte gözüktüğü gibi alışveriş sıralamasında 18-25 ve 26-34 yaş grubu büyük bir pay almaktadır. Buradan firmanın gençlere hitap ettiğini söyleyebiliriz. Diğer gruplar ise çok yakın bir dağılım göstermektedirler.

**Çizelge 5.4: Cinsiyete Göre Laptop Çanta Satışı**

13" Laptop Çantası Çapraz Tablolama				
		Satılmadı	Satıldı	Toplam
Değer	Erkek	758	29	787
	Kadın	731	32	763
	Toplam	1489	61	1550

Çizelge 5.4'e bakıldığı zaman kadınların erkeklerden fazla laptop çantası aldığı görülmektedir. Değerler birbirine çok yakındır ama çalışmanın yapıldığı tarihler arasındaki laptop renklerinin erkek müşteriler için olduğu göz önünde bulundurulursa kadın müşteriler laptop çantasında erkek müşteriler için olan laptop çantalarını satın almışlardır.

**Çizelge 5.5: Cinsiyete Göre Küçük Boy Bel Satışı**

<b>Küçük Boy Bel Çantası Çapraz Tablolama</b>				
		<b>Satılmadı</b>	<b>Satıldı</b>	<b>Toplam</b>
Değer	Erkek	735	52	787
	Kadın	730	33	763
	Toplam	1465	85	1550

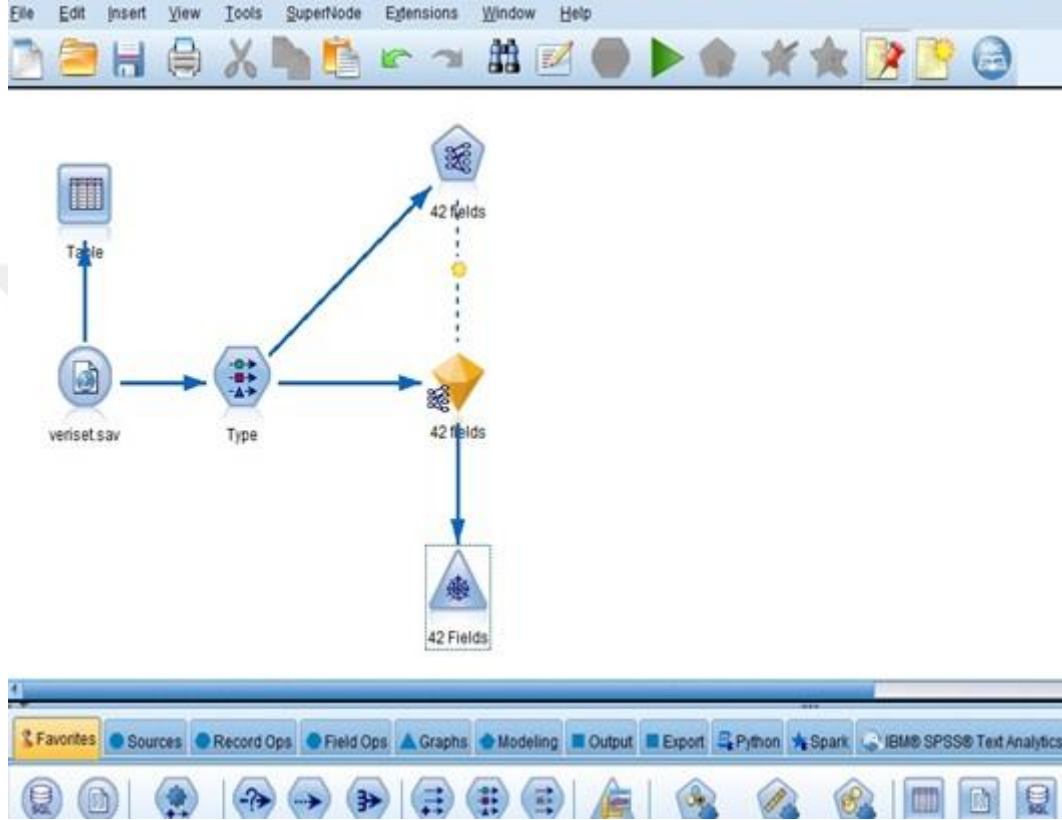
Çizelge 5.5'e bakıldığı zaman erkek müşterilerin kadın müşterilerden daha fazla bel çantası aldığı görülmüştür. Erkek müşterilerin kadınlara oranla daha geniş omuz ve bel çapına sahip oldukları tahmin edildiğinde orta boy yâda büyük boy bel çantada yüksek satış beklenmiştir. Ama burada tam tersi bir sonuç ortaya çıkmaktadır. Buradan iki sonuca varılmaktadır. Birincisi erkek müşteriler gerçekten küçük boy bel çantalarına ilgi göstermektedirler ve ikincisi ise erkek müşteriler bu çantaları hediye olarak kadınlara almaktadırlar.

**Çizelge 5.6: Yaşa Göre Deri Kredikartı Satışı**

<b>Deri Kredikartı Çapraz Tablolama</b>				
		<b>Satılmadı</b>	<b>Satıldı</b>	<b>Toplam</b>
Değer	18-25	423	49	472
	26-34	390	54	444
	35-46	165	27	192
	47-56	157	20	177
	57-65	149	16	165
	65+	91	9	100
	Toplam	1375	175	1550

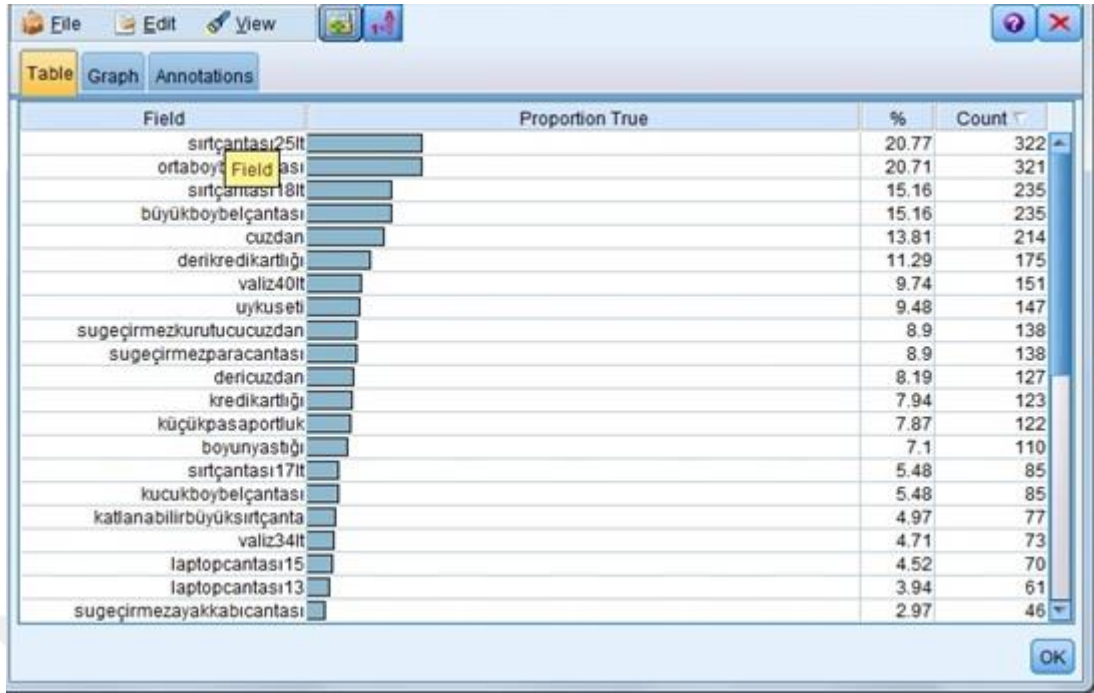
Çizelge 5,6'da ise deri kredi kartlılığının orta yaş ve üssü için daha çok atış öngörülmüştür. 18-25 yaş aralığına bakıldığı zaman 26-34 yaş aralığından sonra ikinci sırayı almaktadır. Sonuç olarak yine hediye için yüksek bir alım gerçekleştirildiği tahmin edilebilir.

Spss Clementine programında oluşturulan algoritma yapısı aşağıdaki şekilde gibidir. Veri seti dosyası SPSS 22 istatistikte hazırlanmış ve oradan kaynak alınmıştır. Veriler daha sonra çıkış kaynak olarak tabloya aktarılmıştır. Daha sonra değişkenlerin tipleri algoritmaya uygun olarak değiştirilmiştir. En son olarak Apriori algoritması uygulanmıştır. Daha sonra ürünlerin arasındaki ilişkilerin sıklık derecesini inceleyen Web grafinin butonu eklenmiştir.



**Şekil 5.1:** Algoritma Aşamaları

Öncelikle hareketlerde bulunan 42 ürün grubunun dağılımı saptanmış ve her bir ürün grubunun satın alınma oranlarını ve satın alınma frekanslarını gösteren dağılım tablosu elde edilmiştir. Aşağıdaki tablo bu dağılımı göstermektedir.



**Şekil 5.2:** Ürünlerin Hareketlerdeki Dağılımı

Şekil 6.4'de de görüldüğü gibi fişlerde en sık tekrarlanan ürünler 25 LT sırt çantası (%20,77), orta boy bel çantası(%20.71), 18 LT sırt çantası (15.16), Büyük boy bel çantası (%15,16), cüzdan (%13,81) ve deri kredikartlığı (%11,29) olarak sıralanabilir. Tüm ürün gruplarının dağılımı bu şekil üzerinde gösterilmektedir.

Araştırmanın algoritma kısmında ise Apiori Algoritması kullanılmış ve 42 ürün grubu içerisinde birbirleriyle ilişkisi olan ürünler bulunmaya çalışılmıştır. Genel çaplı bir ilişki görmek için destek değeri %3 ve güven değeri %25 tutularak birbiri ile ilişkili olan ürün ve ürün grupları gözlenmiştir. Yapılan analiz sonucunda 12 birliktelik kuralı gözlemlenmiş ve aşağıdaki çizelgede verilmiştir.

**Çizelge 5.7:** Veri Kesitinden Elde Edilen Birliktelik Kuralları

<b>Kural</b>	<b>Takip Eden (Consequent)</b>	<b>Önce Gelen (Antecedent)</b>	<b>Destek (Support) %</b>	<b>Güven (Confidence) %</b>
1	Cüzdan	Küçük Pasaportluk	7.871	99.18
2	Uyku Seti	Boyun Yastığı	7.097	98.182
3	Sugeçirmez para çantası	Sugeçirmez cüzdan	8.903	97.826
4	34 LT Valiz	13”Laptop Çantası	3.935	96.721
5	17 LT Sırt Çantası	Küçük Boy Bel Çantası	5.484	96.471
6	40 LT Valiz	15” Laptop Çantası	4.516	95.714
7	18 LT Sırt Çantası	Büyük Boy Bel Çantası	15.161	94.043
8	25 LT Sırt Çantası	Orta Boy Bel Çantası	20.71	91.589
9	Deri Cüzdan	Deri Kredi kartlığı	8.194	77.165
10	Cüzdan	Kredi kartlığı	7.935	75.61
11	Uyku Seti	Katlanabilir Büyük Sırt Çantası	4.968	55.844
12	Deri Kredi Kartlığı	40 LT Valiz	11.29	43.429

Cüzdan –Küçük Pasaportluk kuralı için;

Cüzdan ve Küçük pasaportluğun toplam fiş hareketlerinde birlikte görülme olasılıkları %7,87’dir. Cüzdan alan müşterilerin %99,18 olasılıkla Pasaportlukta aldığı söylenebilir.

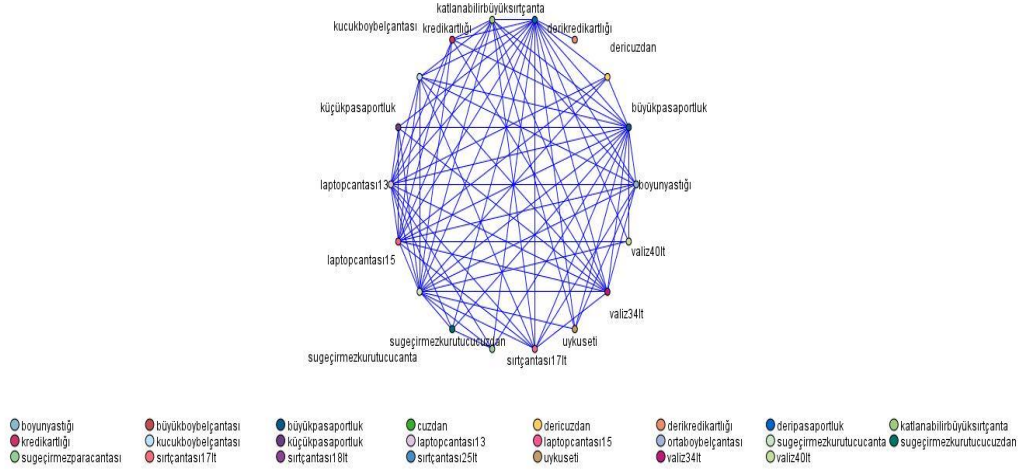
Su geçirmez para çantası ve sugeçirmez cüzdan kuralı için;

Su geçirmez para çantası ve sugeçirmez cüzdanın toplam fiş hareketlerinde birlikte görülme olasılıkları %8,903’tür. Su geçirmez para çantası alan müşterilerin %97,826 olasılıkla sugeçirmez cüzdan aldığıda söylenebilir.

18 LT sırt çantası ve büyük boy bel çantası kuralı için;

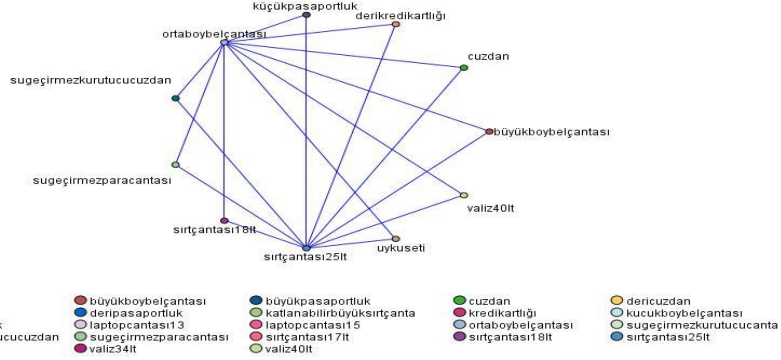
18 LT sırt çantası ve büyük boy bel çantasının toplam fiş hareketlerinde birlikte görülme olasılıkları %11,29 dur. 18 LT sırt çantası alan müşterilerin %43,429 olasılıkla büyük boy bel çantası aldığıda söylenebilir.

Birliktelik kurallarının belirlenmesi için ilk olarak Web Grafiği ile ürünler arası ilişkiler incelenmiştir ve bu sonuç aşağıda verilmiştir.



**Şekil 5.3:** Web Grafiği ile Ürünler Arasındaki İlişkiler

Ürünlerin birbirleri ile olan ilişkileri bu ağda çok yoğun olarak görülmektedir. Web grafiğinin yoğunluk değeri [100:40] 'dir. Bu yoğunluk değeri değiştirildikçe web grafiğinde gösterilen sıklıklarda değişecektir. Bir sonraki aşamada sıklığı daha yüksek olan ilişkilerin görüntülenmesi sağlanmıştır.



**Şekil 5.4:** Sıklığı Yüksek Olan İlişkiler

Yukarıdaki şekle göre Orta Boy Bel Çanta ve 25 LT'lik Sırt Çantası birlikte sık alınmaktadır. Ayrıca çantalar ile bel çantaları birlikte alm gücü yüksektir. Ürünler için elde edilen birliktelik kuralları aşağıda verilmiştir.

Yeni minimum destek değeri %5 minimum güven değeri %50 olarak belirlenmiş analiz bu kıstaslar arasında gerçekleşmiştir, bunun sebebi bir önceki tabloda birbiri



arasında birliktelik kuralları oluşturulan bazı ürün ve ürün gruplarının arasındaki ilişkisinin zayıf kalmasıdır. Aşağıdaki tabloda birliktelik kurallarına göre toplam fişler arasında Pazar sepet analizinde birbiri ile ilişkili en yoğun olan ürünler listelenmiştir.

**Çizelge 5.8:** Düzenlenmiş Ürünler İçin Birliktelik Kuralları

<b>Kural</b>	<b>Takip Eden (Consequent)</b>	<b>Önce Gelen (Antecedent)</b>	<b>Destek (Support) %</b>	<b>Güven (Confidence) %</b>
1	Cüzdan	Küçük Pasaportluk	7.871	99.18
2	Uyku Seti	Boyun Yastığı	7.097	98.182
3	Sugeçirmez para çantası	Sugeçirmez cüzdan	8.903	97.826
4	17 LT Sırt Çantası	Küçük Boy Bel Çantası	5.484	96.721
5	18 LT Sırt Çantası	Büyük Boy Bel Çantası	15.161	94.043
6	25 LT Sırt Çantası	Orta Boy Bel Çantası	20.71	91.589
7	Cüzdan	Kredi kartlığı	7.935	75.61
8	Deri Cüzdan	Deri Kredi kartlığı	11.29	56.0
9	Deri Kredi kartlığı	40 LT Valiz	9.742	50.331

Bazı kurallar aşağıda belirtilmiştir.

Sırt çantası ve Büyük boy bel çantası kuralı için;

Sırt çantası ve büyük boy bel çantası toplam fiş hareketlerinde birlikte görülme olasılıkları %15,16'dır. Sırt çantası alan müşterilerin %94,043 olasılıkla bel çantasıda ürünlerindende aldığı söylenebilir.

Uyku seti ve boyun yastıkları kuralı için;

Uyku seti ve boyun yastıklarının toplam fiş hareketlerinde birlikte görülme olasılıkları %7,097'dir. Uyku seti alan müşterilerin %98,18 olasılıkla boyun yastığı ürünlerindende aldığı söylenebilir.

Cüzdan ve Kredi Kartlık kuralı için;

Cüzdan ve Kredi Kartlıklarının toplam fiş hareketlerinde birlikte görülme olasılıkları %7,94'tür. Cüzdan alan müşterilerin %56,0 olasılıkla Kredi kartı ürünlerinden de aldığı söylenebilir.

**Çizelge 5.9:** Cinsiyete Göre Çantaların Birliktelik Kuralları

<b>Dağılımlar</b>	<b>Takip Eden (Consequent)</b>	<b>Önce Gelen (Antecedent)</b>	<b>Destek %</b>	<b>Güven %</b>
Erkeğe Göre Dağılım	25 LT Sirt Çantası	Orta boy bel çantası	10.323	94.043
Kadına Göre Dağılım	Orta boy bel çantası	25 LT Sirt Çantası	10.129	93.75

Cinsiyet faktörüne bağlı olarak oluşturulan 25 Lt çanta ve orta boy birliktelik kuralı için;

İşlemleri gerçekleştiren erkekler arasında sırt çantasının ve bel çantasının destek oranı %10.32 'dir. Kadınların oranında buna çok yakın olup, bu değer %10.12 'dir.

Güven değerleride birbirine çok yakındır ancak bel çantasının genel tasarımının ve renk skalasının kadın müşteriler için ağırlıklı olduğunu düşünürsek ortaya şöyle bir sonuç çıkmaktadır. Erkek müşteriler tüm bu kadın müşteriye yönelik pazarlama faaliyetlerine rağmen çanta ile bel çantalarını almaktadırlar.

**Çizelge 5.10:** Cinsiyete Göre Cüzdan ve Kartlık Birliktelik Kuralları

<b>Dağılımlar</b>	<b>Takip Eden (Consequent)</b>	<b>Önce Gelen (Antecedent)</b>	<b>Destek %</b>	<b>Güven %</b>
Erkeğe Göre Dağılım	Deri Kredi Kartlığı	Deri Cüzdan	4.065	73.016
Kadına Göre Dağılım	Deri Kredi Kartlığı	Deri Cüzdan	4.129	81.25
Erkeğe Göre Dağılım	Cüzdan	Kredi Kartlığı	4.194	78.462
Kadına Göre Dağılım	Cüzdan	Kredi Kartlığı	3.742	72.414

Yukarıdaki tabloda ise hem kadınların hem erkeklerin Cüzdan-Kredi Kartlığı kuralı ve Deri Cüzdan-Deri Kredi Kartlığı kuralı içerisinde destek oranlarının çok yakın ve güven oranlarının oldukça yüksek olduğu görülmektedir. Önemli olan nokta ise kadın müşterilerin Deri Kredi Kartlığı-Deri Cüzdan kuralında erkek müşterilerden daha fazla güven oranına sahip olmalarıdır. Bu oran %81,25'dir. Buradaki önemli detay ise erkek müşteriler için tasarlanmış ve renk skalasında çantalara göre tam tersi şekilde erkek müşterinin zevkine göre ayarlanmasıdır.

Güven oranına bakıldığında kadın müşteriler erkeklerden daha fazla bu ürünleri beraber aynı sepet içerisinde almışlardır. Bu sonuca göre erkek müşteriler için hazırlanmış ürün görselleri kadın müşteriler için tekrardan düzenlenebilir. Böylece kadın müşterilere olan satış oranı çok daha yüksek noktalara çekilebilir.

## 6. SONUÇ

Bu tez çalışmasında günümüzde sıkça kullanılan veri madenciliği tekniklerinden olan birliktelik kurallarından bahsedilmiştir. Uygulamanın temelini oluşturan birliktelik kuralı ve birliktelik kuralı için kullanılan algoritmaya ayrıntılı olarak yer verilmiştir.

Birliktelik kuralı, Türkiye’de perakende sektöründe faaliyet gösteren bir giyim firmasından toplanan fişler üzerinde uygulanmıştır. Tezin amacı birliktelik kurallarını kullanarak ürünler arasındaki anlamlı ilişkileri bulmak ve müşterilerin satın alma alışkanlıklarını belirleyebilmektir. Bu sonuçlardan hareket ederek daha sonraki raf düzeni ve satış kampanyaları için pazarlama stratejileri geliştirebilmektir.

X firmasının havalimanı mağazasından 60 gün boyunca toplanan fişler birliktelik kuralına uygun olarak bilgisayar ortamına aktarılmış ve analizler için kullanımı çok yaygın olan ve çoğu sektör tarafından veri madenciliği programları arasında ilk sırada gösterilen SPSS Clementine paket programı tercih edilmiştir.

Ürünlerin dağılımları incelendikten sonra apriori algoritması ile veri seti taranarak birliktelik kuralları bulunmuştur. Ürün ve daha sonra ise ürün grupları için aynı işlem gerçekleştirilmiş ve elde edilen kurallar tablolar halinde sunularak ilgili açıklamalar yapılmıştır.

Çıkan sonuçlar incelendiğinde çanta ve bel çantası satışlarının diğer ürünlere göre daha fazla olduğu görülmektedir. Tabii burada araştırmamızın kısıtları göz önünde bulundurulmalıdır. Bu mağazalarda bulunan birçok ürünün diğer mağazalarda bulunmaması ve en çok tercih edilen ve iç piyasada bulunmayan modellerin sadece burada satışa koyulmasıdır.

Çalışmanın gerçekleştirildiği mağazanın bulunduğu bölge uluslararası bir arena olduğu için hitap ettiği hedef kitle nedeni ile alışveriş alışkanlıkları farklılık gösterecektir.

En yüksek güvenilirlikle elde edilen sırt çantası ve bel çantası, cüzdan kredi kartlıkları, laptop çantaları ve valiz arasındaki ilişkiler mağazaya bu dönemde gelen müşterilerin ortalama üzerinde bir gelir durumuna sahip olduklarını göstermektedir,

buradan yola çıkılarak satın alma alışkanlıklarının ve ihtiyaçlarının bu doğrultuda oluştuğunu gösterebilir.

Bu çalışmada ele alınan Pazar sepet analizi ve birliktelik kuralları çıkarımı çerçevesinde incelenmiş apriori algoritması vardır. Farklı çalışma koşulları ve yöntemleri içeren algoritmalar ile farklı veri setleri üzerinde çalışma yapılabilir. Gerçek verilerle başka birliktelik kuralları çıkarımı yapılarak sonuçlar karşılaştırılabilir. Algoritmaların iyi özellikleri birleştirilerek yeni ve daha güçlü algoritmalar elde edilebilir.

Güvenlik nedeniyle müşterilerin bölgesel faaliyetleri bilinmesede, mağazanın internet sitesinden alışveriş yapan müşteriler için birliktelik kuralları oluşturulup, müşteri profiline göre gruplandırılabilir ve farklı kampanyalar geliştirilerek müşterilerin farklı ürünleri de almaları sağlanabilir. Ya da farklı ülkelerden müşteri grupları hedef alınarak yürütülecek bir reklam kampanyası sayesinde yeni müşteri grupları oluşturulabilir.

Ayrıca ortaya çıkan sonuca göre kadın müşterilerin bazı erkek müşterilerinin ürünlerini aldığı erkek müşterilerin ise tam tersi şekilde bazı kadın müşteri ürünlerinin aldığı gözlemlenmiştir. Bu kıstaslar içinde o dönemde mağazada bulunan ürün çeşidi ve renk skalası göz önünde bulundurulmuştur. Ayrıca her iki cinsiyet faktöründe de hediye olarak bu ürünleri aldığı varsayılabilir. Buna istinaden daha sonra yapılacak olan pazarlama stratejilerinde hem kadın hem erkek müşterileri etkileyecek kararlar alınmalıdır.

Çalışma sonucunda elde edilen bilgiler, daha etkin bir raf düzenlemesi veya farklı ürün kampanyaları düzenlemede kullanılabilir. Özellikle satışı çok düşük bulunan ürünleri satışının artırılması için gerekli raf düzeni ve görsel anlatıma dikkat edilebilir. Özellikle erkek ve kadın müşteriler için düzenlenmiş raf düzeninin, daha işlevsel bir raf düzenine dönüştürülmesi önerilir. Böylece değişen görsel düzen ve gerekli ürünlerin birlikte sergilenmesi ile müşterilerin satın alma davranışlarının değişmesi beklenmektedir.

## KAYNAKLAR

- Agrawal, R. & Srikant, R.** (1994). Fast Algorithms For Mining Association Rules, *Proceedings of the 20th VLDB Conference*, Santiago, Syf. 487-499.
- Akbulut, S.** (2006). “Veri Madenciliği Teknikleri İle Bir Kozmetik Markanın Ayrılan Müşteri Analizi ve Müşteri Segmentasyonu”, *Yüksek Lisans Tezi*, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Akiba, Y. & Almuallim, H.** (2002). Development and Applications of Decision Trees. *Expert Systems*, 72-75.
- Akpınar, H.** (2000). Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği. *İ.Ü.İşletme Fakültesi Dergisi*, 6.
- Akpınar, Ö.** (2018). Sigorta Sektöründe Veri Madenciliği ve Kullanım Alanları. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 103-119.
- Aksoy, N.** (2002). *CRM El Kitabı*. İstanbul: Dohaş Yayınları.
- Alan, B.** (2016). “Veri Madenciliği ve Market Veritabanında Birliktelik Kurallarının Belirlenmesi”, *Yüksek Lisans Tezi*, Recep Tayyip Üniversitesi Sosyal Bilimler Enstitüsü, Rize.
- Albayrak, A. S. & Koltan, Ş.** (2009). Veri Madenciliği: Karar Ağacı Algoritmaları ve İMKB Verileri Üzerine Bir Uygulama. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 14(1), 31-52.
- Amirov, M.** (2006). “Bankacılık Sektöründe Müşteri İlişkileri Yönetimi, Türkiye Değerlemesi ve Bir Uygulama”, *Yüksek Lisans Tezi*, Marmara Üniversitesi Bankacılık ve Sigortacılık Enstitüsü, İstanbul.
- Aras, Ü.** (2008). “Finansal Veri Madenciliği”, *Yüksek Lisans Tezi*, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Arasteh, F. & Arbab, F.** (2016). *Market Sepeti Analizine Dayalı Rafların Yeniden Düzenlenmesi ve Promosyonlarının Sonucu Perakende Satış Mağazası Satışları*.
- Arslan, V. & Yılmaz, G.** (2010). Karar Destek Amaçlı Bir Raporlama Aracı. *Bilşim Teknolojileri Dergisi*, 11-15.
- Atakurt, Y.** (1999). Lojistik Regresyon Analizi ve Tıp Alanında Kullanımına İlişkin Bir Uygulama. *Ankara Üniversitesi Tıp Fakültesi Mecmuası*, 4(52), 191-192.
- Ataseven, B.** (2013, Ocak). Yapay Sinir Ağları İle Öngörü Modellemesi. *İstanbul Kültür Üniversitesi İ.İ.B.F Dergisi*, p. 103.
- Ay, D. & Çil, İ.** (2008). Migros Türk A.Ş.’de Birliktelik Kurallarının Yerleşim Düzeni Planlamada Kullanılması. *Endüstri Mühendisliği Dergisi*, 21 (2) ss. 14-29.
- Balint, R. & Guatam, N.** (2006). *Preparing and Mining Data with Microsoft SQL Server 2000*. Microsoft Online Books.
- Barnes, J.** (2000). *Secrets of Customer Relationship Management: It's All About How You Make Them Feel*. New York: McGraw-Hill.
- Barutçugil, İ.** (2002). *Bilgi Yönetimi* (1 ed.). İstanbul: Kariyer Yayıncılık İletişim Eğitim Hiz. Ltd. Şti.
- Bergeron, B.** (2002). *Essentials of CRM*. U.S.A: Wiley.

- Berry, M. J. & Linoff, G.** (1997). *Data Mining Techniques: For Marketing Sales and Customer Relationship Management*. WILEY.
- Birinkulu, Ö. S.** (2006). “Müşteri İlişkileri Yönetiminde Veri Madenciliği ve Bir Uygulama”, *Yüksek Lisans Tezi*, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Brijs, T., Swinnen, G., Vanhoof, K., Wets, G.,** (2004), “Building an Association Rules Framework to Improve Product Assortment Decisions”, *Data Mining and Knowledge Discovery*. Danver: World Scientific Publishing.
- Christopher, W., & Blaxton, T.** (1998). *Data Mining Solutions: Methods and Tools for Solving Real-World Problems*. New Jersey: John Wiley & Sons.
- Clifton, C. & Thuraisingham, B.** (2001). *Emerging Standards for Data Mining*. Computer Standards & Interfaces.
- Comrey, A. L. & Lee, H. B.** (1992). *A First Course in Factor Analysis*. New York: Academic Press.
- Çağltay, N. E.** (2010). *İş Zekası ve Veri Ambarları Sistemleri*. Ankara: ODTÜ Geliştirme Vakfı Yayıncılık.
- Çalışkan, S. K. & Soğukpınar, İ.** (2008). *KxKNN: K-Means ve K En Yakın Komşu Yöntemleri ile Ağlarda Nüfuz Tespiti*. 121-122.
- Çoroğlu, Ç.** (2003). *Modern İşletmelerde Pazarlama ve Satış Yönetimi*. İstanbul: 2003.
- Çubukçu, M. İ.** (2010). Konaklama İşletmeleri Web Site İçeriklerinin Değerlendirilmesi. *İnternet Uygulamaları ve Yönetimi Dergisi*, 40-59.
- Deniz, R. B.** (2002). Müşteri İlişkileri Yönetimi (CRM) ve Elektronik Ortamda Müşteri İlişkileri Yönetimi (E-CRM) Uygulamalarının İşletmelere Sağladığı Yararlar. *Pazarlama Dünyası*, 16-22.
- Doumpos, M. & Zopounidis, C.** (2002). *Multi-Criteria Decision Aid Classification Methods*. Amsterdam: Academic Publishers.
- Doyle, P.** (2003). *Değer Temelli Pazarlama*. Ankara: Kapital Hizmetleri.
- Dönmez, Z. S.** (2008). “Bayi Performansı Değerlendirmesinde Bir Veri Madenciliği Uygulaması”, *Yüksek Lisans Tezi*, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Döşlü, A.** (2008). “Veri Madenciliğinde Market Sepet Analizi ve Birliktelik Kurallarının Belirlenmesi”, *Yüksek Lisans Tezi*, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Dunham, M. H. Xiao, Y. Gruenwald L. ve Hossain, Z.** (2000). “A Survey of Association Rules”, Southern Methodist University, Department of Computer Science, TR00-CSE-8.
- Dürek, N. C.** (2018). “Hızlı Tüketim Sektöründe Tüketici Davranışlarının Birliktelik Kurallarına Göre İncelenmesi ve Sepet Analizi”, *Yüksek Lisans Tezi*, Bahçeşehir Üniversitesi Sosyal Bilimler Enstitüsü.
- Elbaşıoğlu, E.** (2001). “Müşteri İlişkileri Yönetimi; Teknoloji Kullanımı ve Bir Uygulama”, *Yüksek Lisans Tezi*, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Elhan, A. H.** (1997). *Lojistik Regresyon Analizinin Güncellenmesi ve Tıpta Bir Uygulaması*. Ankara, Türkiye.
- Emel, G. & Taşkın, Ç.** (2002). Genetik Algoritmalar ve Uygulama Alanları. *Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi*, syf. 129-150.
- Erdoğan, Ş. Z.** (2004). “Veri Madenciliği Ve Veri Madenciliğinde Kullanılan K-Means Algoritmasının Öğrenci Veri Tabanında Uygulanması”, *Yüksek Lisans Tezi*, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.

- Erol, B.** (2013). “Müşteri İlişkileri Yönetimi İçin Veri Madenciliği Kullanılması ve Sigortacılık Sektörü Üzerine Bir Uygulama”, *Yüksek Lisans Tezi*, Marmara Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Esen, F.** (2009). “Veritabalarından Bilgi Keşfi:Veri Madenciliği ve Bir Sağlık Uygulaması”, *Yüksek Lisans Tezi*, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Ezerçe, A.** (2008). “Müşteri İlişkileri Yönetimi ve Veri Madenciliği:Tekstil Sektöründe Bir Uygulama”, *Yüksek Lisans Tezi*, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Fıglalı, A. & Engin, O.** (2002). Genetik Algoritmalarla Akış Tipi Çizelgelemede Üreme Yöntemi Optimizasyonu. *İstanbul Teknik Üniversitesi Dergisi*, 1-6.
- Fisher, R. A.** (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7, 179-188.
- Gel, O.** (2003). *CRM Yolculuğu*. İstanbul: Sistem Yayıncılık.
- Goldberg, D. E.** (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Boston: Addison-Wesley Longman Publishing Company.
- Gökmen, Ş.** (2004). “MİY'de Bir Araç Olarak Veri Madenciliği ve Perakende Sektöründe Bir Uygulama”, *Yüksek Lisans Tezi*, Yıldız Teknik Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Green, P. E. & Carroll, J. D.** (1978). *Analyzing Multivariate Data*. Illinois, U.S.A: The Dryden Press.
- Groth, R.** (2000). *Data Mining: Building Competitive Advantage*. Prentice Hall PTR.
- Gürgen G.,** (2008). “Birliktelik kuralları ile sepet analizi ve uygulaması”, *Yüksek Lisans Tezi*, Marmara Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Güzel, M.** (2001, Haziran). “Müşteri İlişkileri Yönetimi ve Türkiye Değerlendirmesi”, *Yüksek Lisans Tezi*, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- Hair, J. F., Anderson, R., Tathma, R. & Black, W.** (1998). *Multivariate*. New Jersey: Pearson Prentice Hall.
- Han, J. & Kamber, M.,** (2000).*Data Mining: Concepts and Techniques*. Burbany: Morgan Kaufmann Publishers.
- Hegland, M.** (2001). *Data Mining Techniques*. Cambridge University Press.
- Kalakota, R. & Robinson, M.** (2001). *e-Business 2.0: Roadmap for Success* (2nd ed.). New Jersey: Addison Wesley.
- Kalıkov, A.** (2006, Mayıs). *Veri Madenciliği ve Bir E-Ticaret Uygulaması*. Ankara.
- Karadeniz, N.** (2008). “Müşteri İlişkileri Yönetimi Açısından Veri Madenciliği Yönetimi ve Hizmet Sektörü Üzerine Bir Uygulama”, *Yüksek Lisans Tezi*, Marmara Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Karagöz, N. E.** (2007). “Market Veri Tabanında Veri Madenciliği Uygulaması”, *Yüksek Lisans Tezi*, İstanbul Ticaret Üniversitesi Sosyal Bilimler Enstitüsü.
- Karahan, M. O.** (2010). “CRM ve Ulaşım Sektöründe Bir Uygulama:Firma Çalışanlarının CRM'ye ve Müşterilerine Bakış Açılarının Belirlenmesi”, *Yüksek Lisans Tezi*, Dumlupınar Üniversitesi Sosyal Bilimler Enstitüsü, Kütahya.
- Kartajaya, H., Kotler, P. & Setiawan, I.** (2011). *Marketing 3.0* (1. ed.). (S. Talay, Trans.) İstanbul: Optimist Yayın Dağıtım.



- Keskin, M.** (2013). “Spor ve Yaşam Merkezleri Üzerine Veri Madenciliği Çalışması”, *Yüksek Lisans Tezi*, Okan Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Kiremitçi, B.** (2005). “Veri Ambarlarında Veri Madenciliği ve Ulaştırma-Lojistik Sektöründe Bir Uygulama”, *Yüksek Lisans Tezi*, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Kırım, A.** (2004). *Strateji ve Bire-Bir Pazarlama*. İstanbul: Sistem Yayıncılık.
- Korkmaz, S.** (2006). Müşteri İlişkileri Yönetiminde İnternet Kullanımı: Seyahat Acentelerinde Bir Uygulama. *Fırat Üniversitesi Sosyal Bilimler Dergisi*, 193-213.
- Kotani, M. & Takashi, M.** (1997, Ocak). Neural Computation and Learning Strategy for Manipulator Position Control. *Proceeding of the IEEE First Annual International Conference on Neural Networks*, 4, 525-527.
- Kotler, P.** (2005). *A'dan Z'ye Pazarlama*. İstanbul: MediaCat Yayınları.
- Kotorov, R.** (2002). Ubiquitous Organization: Organizational Design For E-CRM. *Business Process Management Journal*, 8(2), 218-232.
- Kurtuluş, K.** (1998). *Pazarlama Araştırmaları*. İstanbul: Avcıol Basım.
- Larose, D. T.** (2005). *Discovering Knowledge In Data: An Introduction to Data Mining*. New Jersey: A JOHN WILEY & SONS.
- Massy, W. F.** (1971). *What is Factor Analysis?* Belmont: Wadsworth Publishing Company.
- Matthews, K. & Thompson, J.** (2000). *The Economics Of Banking*. New York: John Wiley & Sons Ltd.
- Mocan, G.** (2016). “Perakendecilikte Veri Madenciliği Uygulamaları ve Sorunları”, *Yüksek Lisans Tezi*, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Moss, L. T.** (2003). *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision- Support Applications*. Boston: Addison Wesley.
- Musaoğlu, C.** (2003). “Veri Madenciliği ile tüketici finansmanı sektöründe müşteri elde etme sadakati modellemesi”, *Yüksek Lisans Tezi*, Boğaziçi Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Nilgün, B.** (2003). Veri Tabanlı Pazarlamanın Pazarlama İletişimine Katkısı. *Türkiye Manas Üniversitesi Sosyal Bilimler Dergisi*.
- Odabaşı, Y.** (2005). *Müşteri İlişkileri Yönetimi*. İstanbul: Sistem Yayıncılık.
- Orhunbilge, N.** (1996). Uygulamalı Regresyon ve Korelasyon Analizi. *İ.Ü. İşletme Fakültesi Yayınları*, 9.
- Özçınar, H.** (2006). “KPSS Sonuçlarının Veri Madenciliği Yöntemleriyle Tahmin Edilmesi”, *Yüksek Lisans Tezi*, Pamukkale Üniversitesi Fen Bilimleri Enstitüsü, Denizli.
- Özdamar, K.** (2004). *Paket Programlar ile İstatistiksel Veri Analizi*. Eskişehir: Kaan Kitapevi.
- Özekes, S.** (2003). Veri Madenciliği Modelleri ve Uygulama Alanları. *İstanbul Ticaret Üniversitesi Dergisi*, 69.
- Özel, T.** (2012, Ocak 27). Finans Sektöründe İstatistik ve Veri Madenciliği Uygulamaları. *SPSS Sunumu*.
- Özkan, Y.** (2008). *Veri Madenciliği Yöntemleri*. Papatya Yayıncılık.
- Özmen, Ş.** (2003). *Ağ Ekonomisinde Yeni Ticaret Yolu*. İstanbul: Bilgi Üniversitesi Yayınları.
- Öztemel, E.** (2012). *Yapay Sinir Ağları* (3 ed.). İstanbul: Papatya Yayıncılık.

- Öztürk, S. & Atmaca, H. E.** (2017). İlişkisel ve İlişkisel Olmayan (NoSQL) Veri Tabanı Sistemleri Mimari Performansının Yönetim Bilişim Sistemleri Kapsamında İncelenmesi. *Bilişim Teknolojileri Dergisi*, 199-209.
- Park, S. C., Piramuthu, S. & Shaw, M. J.-P.** (2001). Dynamic rule refinement in knowledge-based data mining systems. *Decision Support Systems*, 31, 205-206.
- Quinlan, J. R.** (1986). Induction of decision trees. In J. R. Quinlan, *Readings in Machine Learning* (pp. 81-106). Morgan Kaufmann.
- Rokach, L. & Maimon, O.** (2010). Decision Trees. In *Data Mining and Knowledge Discovery Handbook* (2 ed., pp. 164-1033). Springer.
- Rygielski, C., Wang, J.-C. & Ten, D. C.** (2002). Data Mining Techniques for Customer Relationship Management. *Technology in Society*, 24(4), 488.
- Sezen, H. K.** (2004). *Yöneylem Araştırması*. Bursa: Ekin Kitapevi.
- Silahtaroglu, G.** (2013). *Veri Madenciliği Kavram ve Algoritmaları*. Papatya Yayıncılık.
- Söztutar, T. B.** (2010.). “Pazarlamada Müşteri İlişkileri Yönetimi ve Gıda Sektörü Üzerine Bir Uygulama”, *Yüksek Lisans Tezi*, Kadir Has Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Suresh, H.** (2002). *Customer Relationship Management – An Opportunity for Competitive Advantage*. PSG Institute of Management.
- Şahin, M.** (1999). “Lojistik Regresyon ve Biyolojik Alanlarda Kullanımı”, *Yüksek Lisans Tezi*, Kahramanmaraş Sütçü İmam Üniversitesi Fen Bilimleri Enstitüsü, Kahramanmaraş.
- Şimşek, U. T.** (2006).” Veri Madenciliği ve Müşteri İlişkileri Yönetiminde Bir Uygulama”, *Doktora Tezi*, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Taşpınar, H.** (2005). *Bilişim Altyapısında CRM Teknik Altyapısı ve İşlevsellikleri*. Ankara: Seçkin Yayınevi.
- Tatlıdil, H.** (1996). *Çok Değişkenli İstatistik Analiz*. Ankara.
- Tek, Ö. B.** (2001). Tek Müşteri Memnuniyetinin Sağlamasında Modern ve İlişkisel Pazarlama:Tüketici Hakları, Kalite Bilinci ve Bilinçli Müşteri Olmak. *İnteraktive Dergisi*.syf. 1.
- Tezcanlar, P.** (2007). “Müşteri İlişkileri Yönetimi ve Veri Madenciliği ve Bir Uygulama” , *Yüksek Lisans Tezi*, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Tiryaki, S.** (2006). “Lojistik Alanında Bir Veri Madenciliği Uygulaması”, *Yüksek Lisans Tezi*, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Ünsal, A.** (2000). Diskriminant Analizi ve Uygulaması Üzerine Bir Örnek. *Gazi Üni. İ.İ.B.F Dergisi*, 2(3), 19.
- Yalçın, D.** (2008). “CRM ve CRM Algısı”, *Yüksek Lisans Tezi*, Beykent Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Yazıcı, A. C., Ögüş, E., Ankaralı, S., Canan, S., Ankaralı, H. & Akkuş, Z.** (2007). Yapay Sinir Ağlarına Genel Bakış. *Türkiye Klinikleri J Med Sci*, 65-71.
- Yeler, M.** (2006). “Müşteri İlişkileri Yönetimi(CRM) ve Türkiye'de Bankacılık Sektöründe CRM Uygulamalarında Eğitimin Rolü”, *Yüksek Lisans Tezi*, Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Yeniay, Ö.** (2001). “An Overview of Genetic Algorithms”. *Anadolu Üniversitesi Bilm ve Teknoloji Dergisi*, 37-41.

- Yeo, M. & Agyel, E.** (1996). Optimising Engineering Problems Using Genetic Algorithms. *Engineering Computations*, 15(2), 271.
- Yılmaz, E.** (2005). İş Zekası Araçları ve Ormançılık. *İstanbul Üniversitesi Orman Fakültesi Dergisi*, 135-146.
- Yılmaz, G. & Arslan, V.** (2010). Karar Destek Sistemlerinin Kullanımı İçin Uygun Bir Model Geliştirilmesi. *Havacılık ve Uzay Teknolojileri Dergisi*, 75-82.

#### İNTERNET KAYNAKLAR

- Alpaydın, E.**, “Zeki Veri Madenciliği: Ham Veriden Altın Bilgiye Ulaşma Yöntemleri“, <https://www.cmpe.boun.edu.tr/~ethem/>, (Erişim Tarihi:22 Aralık 2018)
- Alsalloum,O.**, “Data Mining Knowing The Unknown”, <https://www.slideshare.net/DrOthmanAlsalloum/data-mining-knowing-the-unknown>, (Erişim Tarihi: 16 Ekim 2017)
- Akça, M.** “OLAP ve OLTP nedir?” <http://mustafaakca.com/oltp-ve-olap-nedir/>, (Erişim Tarihi: 25.11.2019)
- Collins, H.**, “Chapter 2-Data Mining Virtuous Cycle”, <https://slideplayer.com/slide/7633316/>, (Erişim Tarihi:15 Ağustos 2018)
- Günel, S.**, “Veri ve Bilgi Arasındaki Fark”, <https://limenya.com/veri-ve-bilgi-arasindaki-fark/>, (Erişim Tarihi: 9 Mart 2015)
- Kılıçaslan, G.** [http://www.academia.edu/7394199/KDSIs\\_Zekas%C4%B1](http://www.academia.edu/7394199/KDSIs_Zekas%C4%B1)(Erişim tarihi:18 Ekim 2017)
- Ölçer, S.**, “Bilim ve Gelecek“, <https://bilimvegelecek.com.tr/index.php/2014/08/01/claude-shannon-ve-bilisim-caginin-dogusu-2/>,(Erişim Tarihi:1 Ağustos 2018)
- Salame,K.**,“(Data Mining-The Big Picture”., <https://www.slideshare.net/KhalidSalam2/data-mining-the-big-picture> , (Erişim Tarihi:7 Ekim 2016)
- Pieper,S.**” What is the difference between data, information and knowledge? ” , <https://www.artegic.com/blog/difference-data-information-knowledge/>,(Erişim Tarihi:29 Mart 2017)
- URL-1:**”Ekonomi nedir?”, <https://tr.wikipedia.org/wiki/%C3%9Cr%C3%BCn>, (Erişim tarihi:22 Ağustos 2017)
- URL-2:** ”Veri Nedir?” <https://tr.wikipedia.org/wiki/Veri>,(Erişim tarihi:2017)

## ÖZGEÇMİŞ

### Kişisel Bilgiler

**Ad-Soyad:** Anıl AKSOY

**Doğum Tarihi:** 08.07.1993

**E-Posta:** aksoyanil1@gmail.com

### Eğitim Bilgileri:

**Lisans:** 2015, Trakya Üniversitesi, İşletme Bilgi Yönetimi

**Yüksek Lisans:** İstanbul Aydın Üniversitesi, İşletme Yönetimi, MBA

### Programlama Bilgileri

C#, MS SQL, SPSS.

### Dil Bilgisi

İngilizce

### Yayınları

Aksoy, Anıl (2017). Pazarlama ve Müşteri İlişkileri Yönetiminde Veri Madenciliği Uygulamaları: Genç Bilim İnsanları Sempozyumu, İstanbul Aydın Üniversitesi Yayınları. syf. 229-236, İstanbul.