

T.C.
BEYKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI
ENDÜSTRİ MÜHENDİSLİĞİ BİLİM DALI

**MÜŞTERİ ANALİZİNDE VERİ MADENCİLİĞİ
KULLANIMI VE BİR UYGULAMA**

(Yüksek Lisans Tezi)

Tezi Hazırlayan:

Özlem HACIOSMANOĞLU

İstanbul, 2017

T.C.
BEYKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI
ENDÜSTRİ MÜHENDİSLİĞİ BİLİM DALI

**MÜŞTERİ ANALİZİNDE VERİ MADENCİLİĞİ
KULLANIMI VE BİR UYGULAMA**

(Yüksek Lisans Tezi)

Tezi Hazırlayan:

Özlem HACIOSMANOĞLU

Öğrenci No:

140792008

Danışman:

Yrd.Doç.Dr. Sabahattin Kerem AYTULUN

İstanbul, 2017

YEMİN METNİ

Yüksek Lisans tezi olarak sunduğum “**Müşteri Analizinde Veri Madenciliği Kullanımı ve Bir Uygulama**” başlıklı bu çalışmanın, bilimsel ahlak ve geleneklere uygun şekilde tarafımdan yazıldığını, yararlandığım eserlerin tamamının kaynaklarda gösterildiğini ve çalışmamın içinde kullanıldıkları her yerde bunlara atıf yapıldığını belirtir ve bunu onurumla doğrularım. 11/05/2017

Aday Özlem HACIOSMANOĞLU



T.C.
BEYKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZ SAVUNMA SINAVI SONUÇ TUTANAĞI

Beykent Üniversitesi
Fen Bilimleri Enstitüsü Müdürlüğü'ne,

Aşağıda tez adı belirtilen yüksek lisans öğrencisi 140892008 no'lu Özlem HACIOSMANOĞLU'nun 11/05/2017 tarihinde yapılan tez savunma sınavı¹ sonucunda 60 dakika süreyle sunduğu ve savunduğu tezi hakkında² oybirliğiyle, kabul kararı verilmiştir.

Bilgilerinize saygılarımızla arz ederiz.

Anabilim Dalı : Endüstri Mühendisliği
Programı : Endüstri Mühendisliği Yüksek Lisans
Tez Başlığı³ : Müşteri Analizinde Veri Madenciliği Kullanımı ve Bir Uygulama

Tez Sınav Jürisi

Öğretim Üyesi

İmza

Danışman : Yrd.Doç.Dr. Sabahattin Kerem AYTULUN
Üye : Doç.Dr.Gökhan SİLAHTAROĞLU
Üye : Yrd.Doç.Dr.Ümit TERZİ

¹ Jüri üyeleri söz konusu tezin kendilerine teslim edildiği tarihten itibaren en geç bir ay içinde toplanarak öğrenciyi tez savunma sınavına alır. Belirlenen günde yapılamayan jüri toplantısı, katılanların hazırladığı bir tutanakla enstitü yönetimine bildirilir. Bu durumda jüri en geç onbeş gün içinde toplanarak adayı tez savunma sınavına alır. Tez savunma sınav süresi en az 45 dakikadır. Yüksek lisans tez savunma sınavı, tez çalışmasının sunulması ve bunu izleyen soru-yanıt bölümlerinden oluşur ve dinleyiciye açıktır. (Beykent Lisansüstü eğitim ve Öğretim Yönetmeliği-Madde30-3)

² Tez sınavının tamamlanmasından sonra jüri, tez hakkında “kabul”, “düzeltme” veya “red” kararı verir. Jüri başkanı, jüri üyelerince imzalanmış sınav tutanağını, tez sınavını izleyen üç gün içinde ilgili enstitü yönetimine teslim eder. Tezi başarısız bulunan öğrencinin Enstitü ile ilişkisi kesilir. Tezi hakkında düzeltme kararı verilen öğrenci en geç üç ay içinde gerekli düzeltmeleri yaparak ve yönetmelikte belirtilen usullere uygun olarak tezini aynı jüri önünde yeniden savunur. Bu savunma sınavında da tezi kabul edilmeyen öğrencinin enstitü ile ilişkisi kesilir. (Beykent Lisansüstü eğitim ve Öğretim Yönetmeliği-Madde30-4)

³ İleride doğabilecek aksaklıkların engellenmesi için tezin başlığının yazılması gerekmektedir.

TEŐEKKÖR

Tez alıřmam süresince ve ders vermeye bařladıęı ilk günden itibaren, yardım ve desteęini esirgemeyen, tavsiye ve tecrübeleriyle tez alıřmama yön veren, hořgörüsünü her anımda hissettiren danıřmanım Yrd. Do. Dr. Sabahattin Kerem Aytulun'a teőekkörü bor bilirim. Tez süresince manevi desteęini esirgemeyen sevgili Kemal Demir'e ve hayatımın her adımında bana verdikleri güven, özveri ve destek için deęerli ailem; sevgili annem, babam ve kardeřlerime teőekkür ederim.



Adı ve Soyadı : Özlem HACIOSMANOĞLU
Danışmanı : Yrd. Doç. Dr. Sabahattin Kerem AYTULUN
Türü ve Tarihi : Yüksek Lisans, 2017
Alanı : Endüstri Mühendisliği
Anahtar Kelimeler :Kümeleme Analizi,Birliktelik Kuralları, Perakende, e-ticaret,
Pazarlama Stratejileri

ÖZ

MÜŞTERİ ANALİZİNDE VERİ MADENCİLİĞİ KULLANIMI VE BİR UYGULAMA

Perakende sektörü; tüketici ve üretici arasındaki hizmeti sağlayan köprüdür. Üretici yoğunluğunun arttığı bir dünyada yeni doğan her marka rekabetin içine doğmaktadır. Dolayısıyla yapılan alışverişlerin işletmeye katmadeğer olarak dönmesini sağlamak ve müşteri bağlılığını sağlamak için pazarlama yaklaşımları gerekmektedir.

Müşteri alışveriş alışkanlıkları ve ihtiyaçları dikkate alınarak oluşturulan müşteri profillerine uygun olarak müşteri memnuniyetini arttıracak pazarlama stratejileri geliştirilmelidir. Dünyaca ünlü Online perakende sitesinin satışlarından oluşan veri setinde müşterilerin ID,Ülke,alınan ürün adı,tarih ve adet bilgileri bulunmaktadır. Bu verilerin veri madenciliği yöntemleri sayesinde müşterilerin alışkanlıkları, bu alışkanlıklara göre profillenmesi gün geçtikçe önem kazanmaktadır. Yapılan veri madenciliği çalışmasının neticesinde müşterilerin rutin olarak aldıkları ürünün yanında beraber alabilecekleri ürünler, müşteri profilleri, hangi ürüne müşterinin ihtiyaç duyabileceği ve kıtasal bazda müşterilerin mevsimsel alışverişleri gözlemlenmektedir.

Name and Surname : Özlem HACIOSMANOĞLU
Thesis Advisor : Asst. Prof. Sabahattin Kerem AYTULUN
Thesis Type and Date : Master, 2017
Study Field : Industrial Engineering
Key Words :Cluster Analysis, Retail,Online Retail, Marketing Strategy

ABSTRACT

DATA MINING USAGE IN CUSTOMER ANALYSIS AND AN APPLICATION

Retail sector; is the bridge that provides service between consumer and producer. In a world where every brand new manufacturer density increases arises in the competition. Therefore, shopping and contribute to the business of added value to ensure customer loyalty marketing approach.

Considering customer shopping habits and needs, in accordance with the customer profile created will increase customer satisfaction and marketing strategies should be developed. The world famous Online retail site customer sales of data sets taken from the ID, country, product name, date and piece of information. That data into data mining an thanks to the habits of customers of methods, according to which these habits day by day is becoming increasingly important. As a result of the clients ' data mining work routinely with the product they can take together products, customer profiles, which product the customer may need and observed seasonal shopping for customers of continental basis.

İÇİNDEKİLER

ÖZ	i
ABSTRACT	ii
TABLolar LİSTESİ	v
ŞEKİLLER LİSTESİ	vi
1. GİRİŞ	1
1.1. Veri Madenciliği Kavramı.....	1
1.2. Veri Madenciliğinin Tarihsel Gelişimi.....	3
1.3. Veri Madenciliği Kullanım Alanları.....	7
1.4. Veri Madenciliğinde Kullanılan Yöntem ve Modeller.....	10
1.4.1. Değer Tahminleme Modeli.....	11
1.4.2. Bağlantı Analizi.....	12
1.4.2.1. Birliktelik Kuralları	12
1.4.2.2. Örüntü Tanıma	15
1.4.2.3. Ardışık Zaman Örüntüleri	16
1.4.3. Sınıflandırma	19
1.4.3.1. Karar Ağaçları	19
1.4.3.2. Yapay Sinir Ağları.....	23
1.4.3.3. Genetik Algoritmalar.....	24
1.4.3.4. K-En Yakın Komşu	25
1.4.3.5. Naive Bayes Sınıflandırıcısı.....	25
1.4.4. Kümeleme Analizi.....	26
1.4.4.1. Hiyerarşik Kümeleme Yöntemi.....	28
1.4.4.2. Hiyerarşik Olmayan Kümeleme Teknikleri	30
1.4.4.3. Model Tabanlı Yöntemler	31
1.4.4.4. K-means Kümeleme	32
1.4.5. Gelişmiş Veri Madenciliği Teknikleri.....	33
1.5. Literatür araştırması.....	36
2. VERİ MADENCİLİĞİNİN SÜRECİ	39
2.1. Problemin Tanımlanması.....	41
2.2. Verinin Anlaşılması.....	42
2.3. Verinin Hazırlanması.....	43
2.3.1. Veri Toplama.....	43

2.3.2. Veri Değerleme	43
2.3.3. Verileri Bir Araya Getirme ve Temizlemesi	44
2.3.4. Veri Seçimi	44
2.3.5. Veri Dönüştürme	45
2.4. Modelleme	45
2.5. Modelin Kurulması	46
2.6. Modelin Kullanılması	47
2.7. Modelin Değerlendirilmesi	47
2.8. Modelin Uygulanması	49
2.9. Veri Madenciliğinde Karşılaşılan Problemler	50
2.9.1. Veri Tabanı Boyutu	50
2.9.2. Gürültülü Veri	51
2.9.3. Eksik Veri	51
2.9.4. Artık Veri	51
2.9.5. Dinamik Veri	52
3. PERAKENDE SEKTÖRÜNDE PAZARLAMA STRATEJİLERİ	53
3.1. Pazarlama Kavramı	53
3.2. Perakende Kavramı	53
3.2.1. Perakendeciliğin Fonksiyonları	54
3.2.2. Dünya Perakendeciliğinin Tarihsel Gelişimi	55
3.3. Strateji Kavramı	56
3.4. Perakende Sektöründe Pazarlama Stratejileri	56
3.5. Perakendecilik Yönetim Kuralları	57
3.6. Perakende Sektörü İle Sunulan Hizmetler	58
4. UYGULAMA	59
4.1. Araştırmanın Amacı ve Kapsamı	59
4.2. Araştırmanın Yöntemi	59
4.3. Evren ve Örneklem	59
4.4. Verileri Anlama ve Değişkenler	59
4.5. Uygulama	61
4.6. Bulgular	61
SONUÇ	64
KAYNAKLAR	69
ÖZGEÇMİŞ	73

TABLÖLAR LİSTESİ

Tablo 1 İstatistiksel Analiz ve Veri Madenciliğinin Karşılaştırılması	2
Tablo 2 Veri Madenciliği Tarihsel Gelişimi	4
Tablo 3 Müşteri Alışveriş Tablosu	18
Tablo 4 Oluşturulan Diziler	18
Tablo 5 Karar Ağacı Beden Tablosu	22
Tablo 6 Gözlem Tablosu	28
Tablo 7 Oluşan Kümeler Tablosu	28
Tablo 8 Risk Matrisi	49
Tablo 9 Birlikte Kuralları Analiz Çıktısı	65

ŞEKİLLER LİSTESİ

Şekil 1	Hiyerarşik Tanım Ağacı	5
Şekil 2	Ağ Veri Yapısı Modeli	6
Şekil 3	İlişkisel Veri Modeli	7
Şekil 4	Birliktelik Kuralları	14
Şekil 5	Birliktelik Kuralları	14
Şekil 6	Beden Tablosu Karar ağacı.....	23
Şekil 7	Yapay Sinir Ağları.....	24
Şekil 8	Veri Madenciliği Süreci.....	41
Şekil 9	Veri Seti Görüntüsü	60
Şekil 10	Knime Analiz Ekranı	61
Şekil 11	Kıtasal bazda alışveriş dağılımı	62
Şekil 12	Kategorisel Yüzdelik Dağılım	62
Şekil 13	Mevsimsel Dağılım.....	63
Şekil 14	Kategori Bazında Sezonsal Satış	63
Şekil 15	Association Rules Algoritması	64
Şekil 16	Decision Tree Analiz Çıktısı	65
Şekil 17	Ülkesel Decision Tree Analiz Çıktısı	66
Şekil 18	Ülkesel Decision Tree Analiz Çıktısı	67
Şekil 19	Ülkesel Decision Tree Analiz Çıktısı	67

1. GİRİŞ

1.1. Veri Madenciliği Kavramı

Veri madenciliği, büyük boyutlarda bulunan verilerden, bu verilerin çalışmanın yapılma amacına stratejik olarak yön vereceği örüntü ve kurallar belirlenmesi amacıyla analiz edilmesidir. Gelişen ve değişen dünya şartlarına uyum sağlayabilmek adına teknolojik gelişimlere açık ve bu gelişimleri bünyesinde uygulamaya geçiren kurumların veritabanlarında büyük miktarlarda veriler bulunmaktadır. Ancak bu veriler işlendiği sürece kuruma katma değer sağlamaktadır. Çok büyük miktarlarda bulunan bu verilerin kullanışlı ve anlamlı veriye dönüştürülmesi ihtiyacı Veri Madenciliğini çekici kılmıştır.

Bu anlamlı bilgi rekabet ortamının mevcudiyetine bağlı olarak değer kazanmıştır. Veri analizi pazar araştırmalarından müşteri tutmaya, üretim kontrolünden bilimsel gelişimlere kadar bir çok alanda katmadeğer sağlamakta ve kurumların karar kalitesini arttırmaktadır.

Tüketici alışkanlıkları, genetik hastalıklar, kazalarda risk faktörleri, kriminal suç incelemeleri, üretim ihtiyacı, hava değişimleri gibi günlük hayatımızın içerisinde bize sıradan gelen her hareket, kurumlar tarafından kayıt altına alınmakta ve bu bilgiyi veri madenciliği yöntemleri ile analiz ederek tüketici alışkanlıkları pazarlama faaliyetleriyle arttırılmakta, genetik hastalıklar tıbbi istatistiklerle gelecek nesillerde tespit edilerek önlem alınmakta, kaza risk faktörleri önlemlerle engellenmekte ve hava değişimleri geçmiş verileri referans alarak ileri tarihler için uzmanlara bilgi vermektedir.

Sektörel bazda Veri madenciliğine ait bir çok tanım ve bir çok bakış açısı bulunmaktadır.

Veri madenciliği, yüksek boyutlarda bulunan verilerin içerisinden geleceğe yönelik tahminlemede yardımcı olacak anlamlı bağlantı ve kuralların bilgisayar programlarının yardımıyla taranması ve analizidir. Aynı zamanda veri madenciliği, büyük boyuttaki verilerin içerisindeki ilişkileri irdeleyerek aralarındaki bağlantıyı bulmayı sağlayan ve veri tabanı sistemleri içerisinde gizli kalmış bilgilerin belirlenmesini sağlayan analizi tekniğidir (1).

Veri madenciliği, veri depolarında bulunan faydalı bilginin keşfedilme sürecidir(2). Veri kümesi içerisinde farkedilmemiş örüntüleri bulmak için kullanılan tekniklerin tamamı Veri Madenciliğini işaret eder. Veri madenciliği geçmişte meydana gelen olayların sonucunu baz alarak geleceğe yön vermek adına uygulanacak adımların tahminlemesine göre karar modelleri oluşturmayı hedefler.

William Frawley ve Gregory Piatetsky-Shapiro'ya göre 1991 yılında yapmış oldukları ' ... verideki gizli, önceden bilinmeyen ve potansiyel olarak faydalı enformasyonun önemsiz olmayanlarının açığa çıkarılması...' tanımını Veri madenciliği desteklemektedir(3). Bilinen bir çok çözümleme aracı kullanılarak veri içerisinde örüntü ve birliktelikler keşfedilebilir, bunların uygulanabilir tahminler yapılması süreci Veri Madenciliğidir(Two Crows 1999)(4).

Moss and Atre 'ye göre tabanında Klasik İstatistik olan Yapay Zeka, Olasılık Kuramı, Makine öğrenimi, ve Örüntü Tanıma gibi pek çok metodun kesişimi olan yöntem Veri Madenciliğidir. Veri madenciliğin temelini Klasik İstatistik oluştursa da aşağıdaki tabloda istatistiksel analizler ve veri madenciliğinin karşılaştırması ve farklılıkları mevcuttur(5).

Tablo 1 İstatistiksel Analiz ve Veri Madenciliğinin Karşılaştırılması

İstatistiksel Analiz	Veri Madenciliği
Hipotezle başlar.	Hipoteze ihtiyaç yoktur.
Hipotezleri tenkit etmek için eşitlikleri olgunlaştırılmalıdır.	Algoritmalarında denklikler kendiliğinden gelişir.
Sayısal verileri kullanmalıdır.	Çeşitli şekilde veriyi kullanır.
Kirli veriyi çözümleme esnasında farkederek ve filtre uygular.	Kirden arındırılmış, temiz veriye dayanır.

Doç. DR. Ali Serhan Koyuncuğil'e göre Veri Madenciliği;

'Veri Madenciliği İstatistiksel Yöntemlerin evrilmiş ve gelişmiş yeni şekli olarak düşünülmesiyle birlikte; Veri Madenciliğine ait olarak tasvir edilecek özellikleri vardır. Veri Madenciliği, metotlarında güçlü teorilerle desteklense de; bu sürece ait yazılımlar incelendiğinde; temel özelliklerinden biri verinin kendiliğinden gelişerek ortaya çıkarılması, keşfi olarak karşımıza çıkmaktadır. Bir diğer özelliği tahminsel modelleme yapabiliyor olmasıdır. Bu sebeptendir ki, Veri Madenciliğini otomatik tahminsel modelleme ile bilgi keşfi şeklinde tanımlayabiliriz. Veri Madenciliğinin temel desteklerini aşağıdaki gibi yazabiliriz.

- Tahminsel modelleme sağlayan istatistiksel metodlar,
- Bu metodların kendiliğinden meydana gelmesini sağlayan algoritmalar.

Veri Madenciliği'nin tasvir edici özelliği olan "Tahmin Edici Modelleme", arkasında çalışan algoritmanın istatistik olduğunu doğrulamaktadır. Veri Madenciliğini açıklarken kullanılan; büyük veri ile;

- Veri madenciliği metodu olan İstatistiksel Analiz Yöntemlerinden çoğunun ön şartı Normal Dağılım hipotezlerini elde etmek,
- Ön çalışması yapılan ve modelde yer alacak verinin, muhtemel tutumunu tespit edebilmek, amaçlarına işaret edilmektedir.'(6).

Veri madenciliği, yazılan tanımlardan da anlaşılacağı üzere büyük şu an ve geçmişte gizli kalmış olan bilgilerin, geleceğe yönelik yeni tahminlemelerin yapılabileceği büyük ölçekteki veri tabanlarındaki bilgilerin anlamlı verilere dönüşeceği teknik analiz sürecidir.

1.2.Verit Madenciliğinin Tarihsel Gelişimi

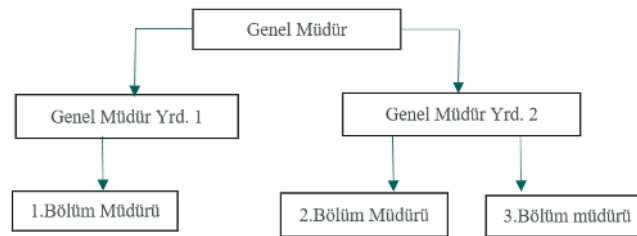
Veri madenciliğinin tarihsel gelişimine bu bölümde tablosal ve yazılı olarak ayrıntılı şekilde yer verilmiştir.

Tablo 2 Veri Madenciliği Tarihsel Gelişimi

Tarih	Basamaklar	Kullanılan Teknolojiler	İlgili Yazılımlar	Aranan Cevap
1960	Veri Toplamak, Veri tabanı Yönetim Sistemi	Kağıt dosyalar, Bilgisayar, Disk	Fortran IV	Son 10 yılda ki ihracat
1980	Veriye ulaşmak Veri sorgulamak	İlişkisel veritabanı, Hızlı bilgisayarlar, Depolama alanları	Oracle, IBM, DB, SQL	Geçen ay Almanyaya yapılan ihraacat miktarı
1990	Veri ambarları, Tahminleme Metodları, Karar Sistemleri	İlişkisel veritabanı, OLAP,Çok boyutlu veritabanı, Veri ambarları, Daha Hızlı bilgisayarlar, Daha fazla depolama alanları	OLAP,IBM Intelligent Miner,SPSS,SAS Miner	Geçen ay Almanya ve İngiltere'ye yapılan ihraacatın karşılaştırması
1990 sonu ve 2000 sonrası	Veri Madenciliği Web madenciliği	İlişkisel veritabanı, Çok boyutlu veritabanı, Veri ambarları,Gelişmiş Bilgisayar algoritmaları, Daha Hızlı bilgisayarlar, Daha fazla depolama alanları	Knime, Oracle Data Miner, SPSS Clementine, SAS Enterprise Miner	Gelecek ay Almanya'ya yapılacak olan ihraacat artışı

Bilgisayarların verimli ve pratik kullanımı bilgilerin biriktirilmesiyle birlikte harekete geçmiştir. Önceleri geliştirilen bilgisayarlar kompleks işlemleri çözümlemek için geliştirilirken kullanıcı ve kurumların ihtiyaçlarına hizmet edecek şekilde veri depolama işlemleri için de geliştirilmeye yönelindi. Bu ihtiyaç veri tabanları ortaya çıkmasını sağlamıştır. Veri tabanlarının her geçen gün büyümesi fiziksel olarak bu verilerin tutulacakları ortamların da artma ihtiyacını doğurmuştur. Veri ambarı kavramının ortaya çıkışı bu dönemlere denk gelmektedir. Kaybolması istenmeyen işletme bilgileri, ambar gibi fiziksel olarak disklerde yedeklenmekteydi. Gün geçtikçe artan veri tabanlarının büyüklükleri düzenle yapılmasını, organize olmasını ve yönetimini de gün geçtikçe zorlaştırmaktaydı. Bu durum veri modelleme kavramının ortaya çıkmasını sağlamıştır.

Geliştirilen ilk örnekler arasında Hiyerarşik ve ağ modelleri bulunur. Hiyerarşik model ağaç yapısına benzer bir veri modelidir. Bu yapının son seviyesinde kök ve kökün uzantıları mevcuttur. Ağaç yapısında her dal başka bir dala dayanaktır. Bu alt dallar sayesinde çeşitlilik artar. Modelin en önemli özelliği yapı içerisindeki her dalın sadece bir noktadan bağlanması koşuludur. 1960-1970 yıllarında yaygın olarak kullanılan bir modeldir. Birbiri ile ilişkili bir çok varlığı temsil edebilmek amacıyla, varlığa ait değişkenlerin, ilişkili olduğu tipler için ayrı ayrı tanımlanması şartı aranır. Bu iş gereksiz veri çevrimine sebep olur. Yukarıda da bahsedildiği üzere, hiyerarşik ver modeli aynı bir ağaç yapısına benzemektedir. Modele ait olan her dal, altındaki n sayıda dala bağlanırken, kendisinin üzerinde yalnızca bir dala bağlı olması gerekir. Modelin tepe noktası kök olarak adlandırılır ve bu köke ait yalnızca bağımlı dallar olabilir. Görüldüğü üzere Hiyerarşik veri modeli kısıtlı bir model olup, veri yapısına ait Hiyerarşik tanım ağacı aşağıdadır.

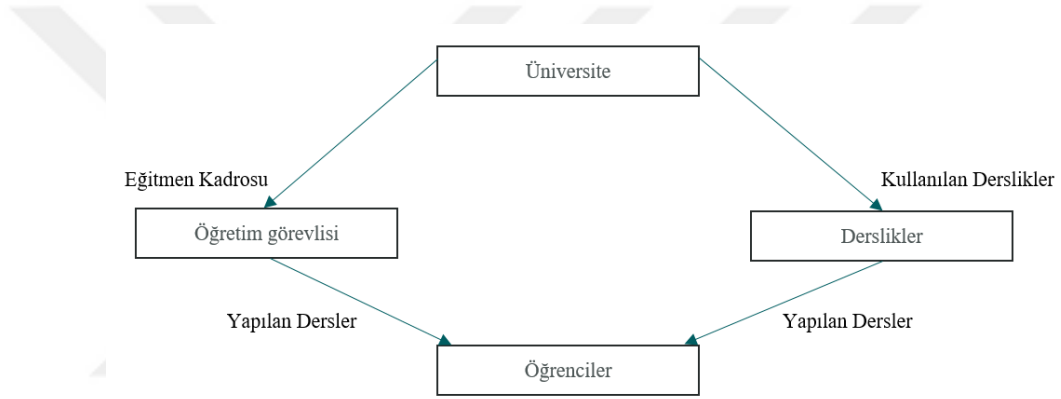


Şekil 1 Hiyerarşik Tanım Ağacı

Diğer bir örnek olan ağ modelleri yapısı içerisinde iki araç barındırır. Bunlar kayıt çeşidi ve bağlantılardır. Kayıt çeşitlerinin varlık, bağlantıların ise ilişki düzeylerini tanımladığı bir veri modelidir. Ağ veri modelinde, hiyerarşik veri modelinde olan bir dalın alt seviyesinde ki, n dalın üst seviyede bulunan 1 dala bağlanması gibi bir kısıtlama yoktur. Yani bir öge ayırım göztmeksizin başka bir ögeye bağlanabilir.

Bağlantılar tarafından belirlenmiş ilişkiler dışında kayıt tipleri arasında ilişki belirlenemez. Ağ modelinde düğümler arası çoklu ilişki kurulamadığından kısıtlı yapıdadır.

Bu veri yapısını gösteren ağ veri yapısı modeli aşağıdadır.

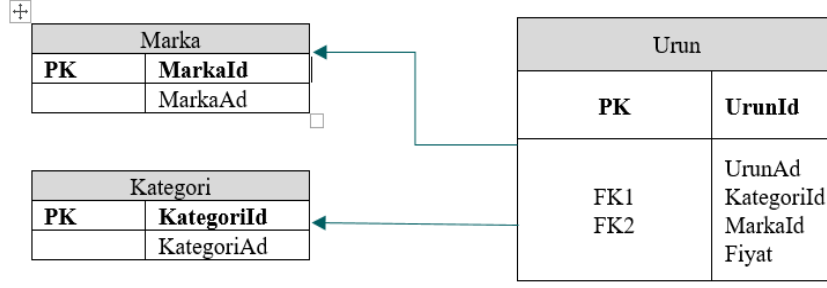


Şekil 2 Ağ Veri Yapısı Modeli

Kısıtlar ve artan ihtiyaçlar neticesinde bu iki Veri tipinin yetersiz kalması ve ihtiyaçları tam anlamıyla karşılayamaması sonucunda Geliştirilmiş Veri Modelleri ortaya çıkmıştır. Bunlar Varlık – İlişki veri modeli , İlişkisel veri modeli ve Nesne – Yönelimli veri modelleri olarak adlandırılır. Günümüzde en sık kullanılan veri modeli İlişkisel veri modelidir.

Varlık – İlişki veri modelinde kısaca tanım yapmak gerekirse varlıklar arasında n adet ilişki olabilir. Tekrar eden bağlantılar da kullanılabilir.

Bu çalışmada kullanılan veri modeli İlişkisel Veri Modelidir.



Şekil 3 İlişkisel Veri Modeli

İhtiyaçlar doğrultusunda şekillenen veri tabanları ve veri modelleme çeşitleri hızla yaygınlaşırken, donanımlar da bu sürece ayak uydurdular. Günümüzde sayısı milyarları bulan bit verilerin ufak depolama araçlarında tutulması mümkün hale gelmiştir. İhtiyaçlar teknolojiyi ciddi anlamda şekillendirerek insan hayatına ve kurumlara kolaylıklar sağlasa da yanında sorunları da getirmektedir. Verilerin saklanması, düzenlenmesi, organize edilmesi zorlu bir süreç olarak görünmese de bu kadar çok veri ile istenilen sonuca ulaşmak daha zordur.

Veri madenciliği, kavramsal olarak 1960'lı yıllarda, bilgisayarların veri analiz problemlerini çözmek için kullanılmaya başlamasıyla ortaya çıktı. O dönemlerde, bilgisayar yardımıyla, yeterince uzun bir tarama yapıldığında, istenilen verilere ulaşmanın mümkün olacağı gerçeği kabullenildi. Bu işleme veri madenciliği yerine önceleri veri taraması (data dredging), veri yakalanması (data fishing) gibi isimler verildi. 1990lı yıllara gelindiğinde veri madenciliği ismi, bilgisayar mühendisleri tarafından ortaya atıldı. Bu camianın amacı, geleneksel istatistiksel yöntemler yerine, veri analizinin algoritmik bilgisayar modülleri tarafından değerlendirmesini vurgulamaktı. Bu noktadan sonra bilimadamları veri madenciliğine çeşitli yaklaşımlar getirmeye başladılar. Bu yaklaşımların kökeninde istatistik, makine öğrenimi (machine learning), veritabanları, otomasyon, pazarlama, araştırma gibi disiplinler ve kavramlar yatmaktaydı (7).

1.3. Veri Madenciliği Kullanım Alanları

Anlamlı veriye dönüştürülecek bilginin bir araya geldiği her sektör ve alanda Veri Madenciliği kullanımı sağlanabilir. Bankacılık, astronomi, biyoloji, finans, pazarlama, sigorta, tıp ve birçok başka dalda da Veri madenciliği kendisine uygulama alanı bulmaktadır.

Veri madenciliği çalışmalarını günlük yaşantımızda farkında olmadan bilgilerle besliyor ve bilgilerle beslediğimiz veritabanlarından, veri madenciliği araçlarıyla anlamlandırılarak karşımıza çıkıyor .

Örneğin; bir parfümeriye gittiğimizde bize önerilen parfümün bir kaç denemede başarılı şekilde beğenilmesi ve satışının gerçekleşmesi. Daha önce satışlardan A parfümünü beğenerek almış olan müşteri kitlesinin fiziksel özellikleri ve giyim zevkleri, ürün içerisinde bulunan esans çeşitleri, alışveriş yapılan saat ve gün gibi bilgiler Veri tabanında biriktirilir. Bu özellikler değişken olarak belirlenebilir. Belirlenen değişkenler Veri Tabanında Veri madenciliği yöntemleriyle anlamlı veri haline getirilir. Parfümeri çalışanlarına çıkan analizler sonucu “Kumral saçlı, açık tenli, 25-30 yaş aralığında bulunana ve şık giyinimli kadınlar üst notası Sicilya mandalinası, orta notası Damask gülü ve alt notası Misk olan parfümleri önermelisiniz.” Şeklinde verilen eğitimler neticesinde satış uzmanının işi kolaylaştırılarak, potansiyelin satışa dönmesi sağlanır.

Bir başka kullanım alanı olarak spor alanında örneklendirebiliriz.

Bir basketbol maçında oyuncuların yapmış olduğu hamleler ve sayının geldiği hamle setinin veritabanında anlamlı veri setine çevrilmesi durumunda takımın oyuncularının bir sonraki karşılaşmada yapacağı hamleler tahminlenerek strateji belirlenebilir.

Risk Yönetimi ve Dolandırıcılık Saptamada ise Dolandırıcılık aşağıdaki konu başlıkları olarak karşımıza çıkabilir.

- Kredi kartı dolandırıcılığı
- Sigorta dolandırıcılığı
- Online işlemler dolandırıcılığı
- Kara-para aklama
- Telefon dolandırıcılığı
- Üyelin abonelik işlemleri dolandırıcılığı

Bu başlıklarda görülen dolandırıcılık çeşitleri teknolojik gelişmelerin tüm dünyada hızla ilerlemesi ile artış göstermekte ve çıkan teknolojilere uygun dolandırıcılık çeşitleri türemektedir.

Dr. Gökhan Silahtaroglu Kavram ve algoritmalarıyla Temek Veri Madenciliği kitabında konuyla ilgili olarak aşağıdaki bilgileri aktarmıştır.

Dolandırıcılık, tüm dünya endüstrisi için büyük bir sorun haline gelmiş ve her yıl milyarlarca dolar kayba neden olmaktadır. Master Card International 1997'deki satışlarının %7.7'sinin dolandırıcılık kapsamında gerçekleştiğini bildirmiştir.(Frenkel)

Kredi kartı veren finans kuruluşları daha dolandırıcılık meydana gelmeden dolandırıcılığı tespit etmektedir.

Bunun için de bilgi keşfi, Yapay zeka ve **Veri madenciliği** gibi yöntemler kullanılmaktadır. Küçük işletmelerde söz gelimi 25-30 yıldır o işi yapan bir kişi, gelen müşterinin dolandırıcı olabileceğini konuşma tarzından, bakışından geçmiş deneyimlerini kullanarak tahmin edebilir; ancak aynı şeyi 25-30 yaşlarında, bir bankanın kredi bölümünde çalışan bir kimseden beklememiz doğru olmayacaktır: Çünkü bu kişinin kredi isteyen kişinin daha başlangıçta dolandırıcı olabileceğini anlayabileceği kadar deneyimi yoktur. Öyleyse, burada devreye kişinin deneyimlerinden çok kurumun deneyimleri yani verileri ve bu verileri kullanacak veri madenciliği algoritmaları girmektedir (8).

Bu çalışmamızın amacına uygun olarak kullanılan alanlar ise sepet analizi, mevcut müşterilerin elde tutulması ve yeni müşterilerin kazanılması, müşteri profillemesi.

Sepet Analizi: Müşterilerin aldığı ürünlerin birliktelikleri arasında bir ilişkilendirme kurulur. Elde edilen bilgi ile promosyon stratejileri oluşturularak satış karının yükselmesi sağlanır. Müşterinin fiyat değişikliğinde alışveriş alışkanlığının belirlenmesi ve çapraz satış analizi ile ürünler arasındaki birliktelik kurallarının ve ilişkilerin belirlenmesi sağlanır.

Mevcut müşterilerin elde tutulması ve yeni müşterilerin kazanılması: Mevcut müşterilerin bağlılığının test edilip kaybedilmeye ve kazanılmaya en yakın müşterilerin belirlenmesi ve gerekli birimlere raporlanması veri madenciliği yöntemi ile yapılmaktadır. Mevcut müşterilerin elde tutulması ve yeni müşterilerin kazanılması hususunda ki yöntem ticari işlem gören tüm işletmelerde kullanılabilir. Örneğin; Telekomünikasyon şirketleri veri kayıtlarının detaylı ve yoğun oluşundan

dolayı veri madenciliğine anlamlandırılabilir dalar saęlıyor. Müşteri henüz 1 hafta sonrasında hattını deęiştireceğini bilmezken, Veri madencilięi arama ve aranma sıklığı, görüştüğü dięer operatörler, paket aşımaları vb. gibi verilerle bu kaybı önceden bildirebilir.

Müşteri Profilleme: Bir işletmenin her müşterisi aynı deęildir. Bu sebeple ihtiyaçlar ve talepler, alışkanlıklar birbirinden farklılıklar göstermektedir. Müşteri profilleme, müşterinin ortak özelliklerine göre sınırların belirlenmesine olanak saęlar. Doğru ihtiyacın belirlenip müşteriye sunulması hususunda önemli yere sahiptir.

1.4. Veri Madenciliğinde Kullanılan Yöntem ve Modeller

Veri madencilięi kullanım alana göre deęişik modellere ayrılmaktadır. Bu modeller Deęer tahmini modeli, veritabanı kümeleme modeli, baęlantı analizi ve fark sapmaları olarak dört ana başlık altında incelenebilir; bu modeller literatür aramalarında operasyon ve yöntem şeklinde isimlendirilmektedir. Bahsi geçen operasyon ve yöntemler uygulamalarında farklı amaçlara hizmet etmek için kullanılırken çeşitli teknik ve algortimadan yararlanılır. Uygulamalarda operasyon ve yöntemlerin kullandıkları teknikler ve algortimalar tahminleyici ve tamamlayıcı olarak ikiye ayrılır.

Çalışmamıza uygun olarak tahminleyici modellerin cevabını aradıęı soru: “Bu müşteriden gelecek dönemlere ait ne kadar gelir elde edilebilir?” ya da “Haziran ayında yapılacak satışların % kaç A ürüne ait olabilir?” ‘dir

Tanımlayıcı modellerin aradıęı soru ise satılan ürünler arasında ki gizli kalmış ilişkileri ortaya çıkararak ve bir sonuca ulaşır: ”Balon alan bir müşterinin, kına gecesi malzemeleri alma olasılığı %17’dir.” ya da “Yıllık 100.000 Lira deęerinde ürün ithal eden işletmeler, yıllık 1.500.000 adet ürün ithal eden işletmeler ile yıllık 50.000 Lira deęerinde ürün ithal eden işletmelerden daha çok benzer.”

Kullanılan teknik, algortimalar ve modellerin bilgisayar yazılımı olması gerçeęi unutulmamalıdır. Yazılımlar her ne kadar matematiksel ve algoritmik olarak birbirlerinden ayrılırsalar da bazı ortak özellikleri mevcuttur; bu özelliklerden biri öğrenme işlemidir. Bilgisayar yazılımları yapılan veri girişleri sayesinde bu verileri inceleyerek farklı algoritmalarla bu verilerden sonuçlar ve kurallar çıkarabilir. Bu

işlem “Öğrenme” olarak isimlendirilir. Elde edilen çıkarımlar verinin diğer kısımlarına uygulanıp denetlenir. Benzetmek gerekirse kendi sınavını yaparak öğrendiklerini sınar. Bulduğu sonuca göre yenilemesi gereken çıkarım ve kuralları yeniler. Bu sonuçları da denetlemesi gerekir, bu işleme “doğrulama” denir.

Veri madenciliğinde bir çok algoritma geliştirilmiştir, bunlar sınıflama,kümeleme, bağlantı analizinde kullanılmak üzere olan algoritmalarıdır. Geliştirilmiş algoritmalarından bir kısmı yalnızca sınıflandırma ya da kümeleme gibi konuları ilgilendirirken bazıları çeşitli sürümlerde birden fazla konuda kullanılabilir. Genetik algoritmalar, yapay sinir ağları hem sınıflandırma hem de kümeleme modellerinde kullanılabilir. Aproiri algoritması sadece birliktelik kurallarının oluşturulmasında kullanılan bir algortimadır.

1.4.1. Değer Tahminleme Modeli

Değer tahminleme yani tahminsel modelleme olarakta adlandırdığımız modeldeki öğrenme insanın öğrenme şekline benzemektedir. İnsan,gözlemleyerek hafızasının oluşturur. Benzetmek gerekirse hafızamız canlı bir veritabanıdır. Tahminsel model kendisine atanan veritabanını inceler, bu veritabanında ki verileri birbirine benzeterek ilişkilendirir ve bir sınıflandırma yapar. Örneğin hafıza, bir yiyeceğin tatlı ve tuzlu olduğu deneyimleyendiğinde şeklinden, tadından hafızada oluşturulan veritabanı sayesinde sınıflama yapabilir. Sonra ki yiyecek karşılaşmasında hafıza bu yiyecek hakkında farklılıkları ve benzerlikleri mukayese ederek bir tahmin eder. Aslında yapılan tamamen sınıflandırma, genelleme yapma işlemidir. Bu öğrenme şekline ise “Denetimli öğrenme” adı verilir.

Veri madenciliği algoritmaları da veritabanında bulunan nesnelerin niteliklerini, nesnelerin adlarıyla ilişkilendirerek bu nesnelerin birbirinden değişik ya da benzer, aynı sınıftan nesneler olduklarını yakalar ve öğrenir. Daha sonra ki süreçlerde kendisine verilen farklı nitalikleri değerlendirip bu niteliğe sahip diğer nesnelerin adını tahminler. Denetimli öğrenmenin karşısı olan öğrenme şekline ise “denetimsiz öğrenme” adı verilir. Denetimli öğrenmeden farklı olarak bir bilgi verilmeden yalnızca verilerin girilmesi ve bu verilerden muhtelif sonuçların çıkarılması sağlanır. Herhangi bir veri girilmediği için beklenen sonucun doğruluğunda nesnel bir değerlendirmesi yoktur.

Örneğin, sosyal bir platformda kişiler arkadaşlarını ekler. Sosyal platform üye kişileri arkadaşlıklarına göre sınıflandırma yaparak arkadaş grupları oluşturur ve sosyal platform kullanıcılarına “tanıyor olabileceğiniz kişiler” şeklinde öneri yapar. Sistemin sosyal platform kullanıcılarına “sadece tanıyor olabileceği kişiler arkadaşlık isteği gönderebilsin” şeklinde seçenek sunması gözetimsiz öğrenmeyi örnekler. Gözetimsiz öğrenmede veriler üzerinde bir ayırım yapmadan verilerin veri madenciliği sistemine yüklenip, çalışan algoritma ile sistemin bu bilgiyi ayıklayıp, kendisinin öğrenmesi beklenir (9).

Tahmini yaklaşımın kullanım alanları uygulama, hedef müşteri kitle belirlenmesi, pazar sepet çözümlemesi, çapraz satış, müşteri ilişkileri yönetimi olarak bilinmektedir. Tahmini yaklaşımı baz alarak geliştirilen temel iki teknik vardır; Karar ağaçları ve sınıflandırmalar. Genetik algoritmalar, yapay sinir ağları, Bayes yöntemi gibi metotlar da bu doğrultuda kullanılır.

1.4.2. Bağlantı Analizi

Tahmin modelinde yazılım girilen verilerin bulunduğu veritabanını birlik içerisinde değerlendirerek öğrenmeyi de bütünü baz alarak sunar. Halbuki bağlantı analizinde veritabanındaki kayıt ve kayıt grupları arasında bir bağlantı ve ilişki oluşturulur. Bağlantı analizi bir veritabanındaki kayıtlar ya da bir graf üzerindeki düğümler arasında çok rastlanan kuralları ortaya koyar.

Bağlantı analizinin en çok kullanıldığı alanlardan bazıları şunlardır: Çapraz satış, stok fiyat hareketleri ve hedef müşteri kitlesinin belirlenmesi. Bağlantı analizi dört ana başlık altında incelenebilir. Bunlar birliktelik kuralları, örüntü tanıma, ardışık zamana örüntüleri ve benzer zaman keşfidir. (10)

1.4.2.1. Birliktelik Kuralları

Veri madenciliğinde kullanılan ilk tekniklerden birisi de birliktelik kurallarıdır (Agrawal vd., 1993). Birliktelik kuralları verideki potansiyel ilişkileri tanımlar. Birliktelik kuralı, geçmiş verilerin analiz edilerek bu veriler içindeki birliktelik davranışlarının tespiti ile geleceğe yönelik çalışmalar yapılmasını destekleyen bir yaklaşımdır. Birliktelik kuralı madenciliğinin uygulamasına pazar sepeti analizi örnek verilebilir (Frawley vd., 1991). Birliktelik kuralındaki amaç; alışveriş esnasında müşterilerin satın aldıkları ürünler arasındaki birliktelik ilişkisini bulmak,

bu ilişki verisi doğrultusunda müşterilerin satın alma alışkanlıklarını tespit etmektir. Satıcılar, keşfedilen bu birliktelik bağıntıları ve alışkanlıklar sayesinde etkili ve kazançlı pazarlama ve satış imkanına sahip olmaktadırlar. Örneğin, bir marketten müşterilerin süt ve peynir satın alımlarının % 70’inde bu ürünler ile birlikte yoğurt da satın alınmıştır. Bu tür birliktelik örüntüsünün tespit edilebilmesi için, örüntü içinde yer alan ürünlerin birden çok satın alma hareketinde birlikte yer alması gerekir. Milyonlarca veri üzerinde veri madenciliği teknikleri uygulandığında, birliktelik sorgusu için kullanılan algoritmalar hızlı olmalıdır (Agrawal ve Srikant, 1995)(11).

Her bir veri bilgisayara nümerik değerler olarak girilir. Bu nümerik değerler arasında ilişki kurulur ve bu nümerik değerler arasında kurulan ilişki sayesinde bir nümerik değerden sonra gelecek olan nümerik değer otomatik olarak belirlenip bahsi geçen nümerik değer zarar görmesi (kaybolması,okunmaması) durumlarda bu nümerik değer yerine gelecek değer otomatik olarak atanarak sistemin sürekliliği sağlanır. Telefonlarımızda bulunan tahmin klavyeleri bu konu için iyi bir örnektir. Girdiğimiz değerler nümerik olmasa da sistem aynı model üzerinden çalışmaktadır. “Mrhab” ve “Kaybomu” yazıldığında önceden girilen bir veri olması sebebiyle “Merhaba, Merhabalar” ve “Kaybolmuş, Kaybolmuştu” gibi opsiyonel kelimeler sunmaktadır(Şekil 1,Şekil 2). Hiç bir aşamada insan müdahalesi olmadığı hatırlatılır. Dolayısıyla tahminlenen kelimeler bir arada kullanılan ve defalarca farklı kombinasyonlarda yazılıma girilen veriler sebebiyle yazılım kendinde yaptığı sorgularla tahminleme yapar. Buna bir nevi keşif diyebiliriz.

Kime:



Şekil 4 Birliktelik Kuralları

Kime:



Şekil 5 Birliktelik Kuralları

Veritabanında birliktelik kurallarının belirlenmesi veritabanı bilgi keşfinin ilk aşamasıdır (12).

Birliktelik kurallarının kullanıldığı en tipik örnek “Sepet analizi”dir. Sepet analizi ile müşterilerin yaptıkları alışverişlerdeki ürünler arasındaki birliktelikler bulunarak,

müşterilerin satın alma alışkanlıkları belirlenir. Bu tip birlikteliklerin belirlenmesi, müşterilerin hangi ürünleri bir arada aldıkları bilgisini ortaya çıkarır ve market yöneticileri de bu bilgi ışığında raf düzenlerini belirleyerek, ürün paketleri hazırlayarak satış oranlarını arttırabilir ve etkili satış stratejileri geliştirilebilir. Sepet analizinde amaç alanlar arasındaki ilişkileri bulmaktır. Bu ilişkilerin bilinmesi şirketin karını arttırmak için kullanılır. Sepet analizi yapılırken ürünler arasındaki ilişkileri ortaya çıkarmak için bir takım kurallara ihtiyaç vardır. Bu kuralların elde edilebilmesi için minimum seviyede destek ve güven değerlerinin belirlenmesi şarttır. Bir çok birliktelik kuralları algoritması “üret ve test et stratejisi” kullanır. Destek değeri elde edilen bilgilerin mevcut benzer bilgiler içindeki oranını temsil eder.

1.4.2.2. Örüntü Tanıma

“Örüntü” terimi sözlük anlamında vaka veya objelerin kararlı halde ardışır birbirini izleyerek gelişmesidir. Bu örüntüler, çoğunlukla alakalı verilerin kantitatif tanımlama şekilleridir.

Örüntü gözlenebilirliği, ölçülebilirliği olan tekrar edebilen şerik metodik yapı ya da benzerlikleri olan örnekler topluluğudur. Topluluk olarak adlandırdığımız bir kümedir. Örüntü şeklinde adlandırılan vaka, obje veya süreçler gürültü ya da düzensizlik içerebilir(13).

Yazılı bir metni tanımak ya da o metnin benzerini bulmak örüntü tanımının kapsamı altındadır. Parmak izi, kan hücrelerinin karşılaştırılması, ses, yüz tanıma, el yazılarının tespiti alanlarında uygulanır. Kısaca Örüntü; el, yüz, resim, çizim ve ses gibi varlıkların sayısal ortamda gösterdiği eylemlerdir (14).

Örüntü tanıma işlemi bir nevi sınıflandırmadır. Erişilmesi amaçlanan bir örnek mevcuttur ve biz bu örneğin benzerini ya da olasılık varsa eşini ararız. Algoritma verilen örneğe benzeyenler bir arada toplanır, çok benzerden az benzeyenlere doğru sıralaması yapılır.

Herhangi bir sebepten dolayı aranan kişilerin şehrin herhangi bir bölgesinde merkezi bir kamera sistemine takılması, kişinin kimlik tespiti için yeterlidir. Merkezi kamera sisteminde her görüntülenen yüz eldeki aranan kişinin fotoğraflarıyla

karşılaştırılmasını ve benzerliklerin yakalanmasını sağlayan çeşitli örüntü tanıma algoritmaları vardır.

Veri madenciliğinin kullanım alanlarını anlatırken yazmış olduğumuz spor örneği de Örüntü Tanıma için iyi bir açıklamadır.

“Bir basketbol maçında oyuncuların yapmış olduğu hamleler ve sayının geldiği hamle setinin veritabanında anlamlı veri setine çevrilmesi durumunda takımın oyuncularının bir sonraki karşılaşmada yapacağı hamleler tahminlenerek strateji belirlenebilir.”

Örüntü tanıma için geliştirilmiş olan algoritmalar aşağıdaki gibidir.

- K-en Yakın Komşu Algoritması
- Doğrusal Sınıflayıcı
- Üstsel Sınıflayıcı

1.4.2.3. Ardışık Zaman Örüntüleri

Örüntünün sözlük anlamı dışında Veri madenciliğinde; el, yüz, resim, çizim ve ses gibi varlıkların sayısal ortamda gösterdiği eylemler olduğundan bahsetmiştik. Bunlara ek olarak örneklendirmek gerekirse alışverişte alınan cips, içecek ve kuruyemişte bir örüntüdür. Fakat bu bahsedile örnek ardışık zaman örüntüsüne örnek değildir.

Ardışık zaman örüntüsünde müşteri bugün pikap aldıysa, yarın plak alır ve sonraki süreçte tekrar plak alır ise bu alışveriş bizim için ardışık zaman örüntüsü oluşturur.

Wojciechowski ardışık örüntüleri “Ardışık örüntüler meydana gelen bir dizi olay içinde en sıklıkla rastlanan alt dizilerdir.” şeklinde tanımlamaktadır.

Wojciechowski sıklık değerini anlatırken kullanıcı tarafından belirlenecek olan destek seviyesini vurgulamaktadır. Güven aralığı ise bir kısıttır. Ardışık zaman örüntüleri alanında kullanılmak üzere geliştirilmiş algoritma ve bu algoritmaları kullanan yazılımlar, örüntülerin belirlenme işlemini basitleştirmek ve hızlandırmak için kullanıcıya bazı kısıtlar koyması için olanak tanımaktadır.

Algoritmaların daha iyi anlaşılması için, ardışık zaman örüntüleri ile ilgili matematiksel tanımların gözden geçirilmesinde faydalıdır. Bu tanımlardan bazıları aşağıdaki gibidir

$L=\{l_1, l_2, l_3, \dots, l_m\}$, m adet nesnelerden oluşan bir küme olduğunu varsayalım. Buradaki nesneler kümesi boş olamaz.

Bir dizi ise $\langle X_1, X_2, X_3, \dots, X_n \rangle$ şeklinde gösterilir ve söz konusu nesneler kümesinin sıralı listesidir. Burada herhangi bir X_i aslında nesneler kümesinin bir alt kümesidir ($X_i \subseteq L$).

D değişken uzunluğundaki diziler kümesi (veri dizileri) olsun. Burada,

$S = \langle X_1, X_2, X_3 \dots X_n \rangle$ ve her bir X_i ayrı bir zaman diliminde oluşmuştur.

Eğer zaman kısıtı yok ise şu söylenebilir: eğer $i_1 < i_2 \dots < i_n$ ise, yani;

$(X_1 \subseteq Y_{i1}), (X_2 \subseteq Y_{i2}), (X_3 \subseteq Y_{i3}) \dots (X_n \subseteq Y_{in})$ ise

$X = \langle X_1, X_2, X_3 \dots X_n \rangle$ dizisi

$Y = \langle Y_1, Y_2, Y_3 \dots Y_m \rangle$

Dizisinin içinde yer alan bir dizidir diyebiliriz.

$Y = \langle Y_{i1}, Y_{i2}, Y_{i3} \dots Y_{in} \rangle$

X'in Y içinde meydana gelişlerinden ya da oluşumlarından biridir. Verilen bir dizi içerisinde herhangi bir başka dizinin oluşumunu ararken, kullanıcı tarafından konulan zaman kısıtları da gözönüne alınmalıdır. Bu zaman kısıtları, meydana gelen olaylar arasındaki en büyük ve en küçük zaman farkından başka bir şey değildir.

$X = \langle X_1, X_2, X_3 \dots X_n \rangle$ dizisinin D veri dizisinde olma isteği ise D'deki veri dizilerinin, X içermeye yüzdesidir. Bu duruma ardışık örüntü D içindeki destek değeri kullanıcı tarafından verilmiş olan eşik değerinden büyük olan dizidir(15). (Wojciechowski, 2003)

Tablo 3 Müşteri Alışveriş Tablosu

Müşteri No	Fatura Tarihi	Ürün Kodu
1	12.03.2016	45,58,37
2	12.03.2016	37
1	13.03.2016	48,65,29
3	13.03.2016	77
2	14.03.2016	48
4	15.03.2016	65,98,75,53
2	16.03.2016	29

Yukarıdaki müşteri alışveriş tablosunda her bir müşteri bir dizi oluşturmaktadır. Örneğin 1 numaralı müşterinin sırasıyla (45),(37),(58),(48),(65),(29) kodlu ürünleri satın alması bir ardışık zaman örüntüsü oluşturur. Fakat bu örüntünün, tablo içinde bir müşteri ile benzerliği ya da tekrarı bulunmamaktadır. Eğer destek seviyesi %0 olarak belirlenirse bu örüntü olarak değerlendirilir.

Minimum destek seviyesi %25, yani en az 2 müşteri olarak belirlenirse (37),(48),(29) kodlu ürünler anlamlı bir ardışık zaman örüntü olarak düşünülebilir. (37),(48),(29) dizisi hem 1 numaralı hem de 2 numaralı müşterinin alışverişinde desteklenmektedir. 2 numaralı müşteri (37),(48),(29) kodlu ürünlerini sırasıyla satın alırken 1 numaralı müşteri (45),(58),(65) kodlu ürünlerin arasında (37),(48),(29) kodlu ürünleri satın almıştır. 1 numaralı müşterinin, bahsi geçen ürünler arasında (65) kodlu ürünü alması ardışıklığı bozmamaktadır.

Oluşturulan Diziler

Tablo 4 Oluşturulan Diziler

Müşteri Numarası	Ürün Kodu
1	45,58,37,48,65,29
2	37,48,29
3	77
4	65,98,75,53

1.4.3. Sınıflandırma

Veri madenciliğinde en çok kullanılan tekniklerden biri sınıflandırmadır. Örüntü tanıma, resim, genetik rahatsızlık tanısı, dolandırıcılık saptama, kalite kontrol çalışmaları ve pazarlama tekniği konularında sınıflandırma teknikleri bolca kullanılmaktadır. Sınıflandırma tahminleyici bir model olarak bilinmektedir. Havanın ertesi gün ne seviyelerde olacağını tahminlenmesi ya da genç kadınların küçük araba satın alıp; yaşlı, zengin erkeklerin ise büyük, lüks araba satın aldığı tahminlenmesinin yapılması sınıflama işlemidir.

Sınıflandırmanın matematiksel olarak tanımı aşağıdaki gibidir:

$D = \{t_1, t_2 \dots t_n\}$ 'i bir veritabanı olarak kabul edelim ve herbir t_i bir kaydı temsil etsin.

$C = \{C_1, C_2, \dots, C_m\}$ ise m adet sınıftan oluşan sınıflar kümesini temsil etsin.

$f : D \rightarrow C$ ve herbir t_i bir sınıfa dahil olmalıdır.

Herbir $C_j = \{t_i / f(t_i) = C_j, 1 \leq i \leq n, \text{ ve } t_i \in D\}$ 'dir.

Sınıflama işleminde elimizdeki sınıf veya istatistiksel değimiyle bağımlı değişken hem sınıfsal hem de sürekli değer taşıyabilir: bu anlamda regresyon ve çok terimli regresyona yaklaşılmaktadır.(16)

1.4.3.1. Karar Ağaçları

Karar ağaçları akış şeması benzeri yapılar olup, düğümler özellik değerleri için yapılan testleri, dallar test sonuçlarını, dalların ucundaki yapraklar ise sınıfları gösterir. Karar ağaçları kolaylıkla sınıflandırma kurallarına dönüştürülebilir.

En önemli sınıflama araçlarından biri olan karar ağaçlarında, öğrenme algoritması basittir. Ortaya konan özbilginin gösterimi kolaylıkla anlaşılabilir. Karar ağaçları yalnızca kararları göstermezler, aynı zamanda kararların açıklamasını da içerirler. Karar ağacını oluşturan eğitim süreci tümevarımdır. Bir eğitim nesneleri kümesinden karar ağacı oluşturma yordamı, ağaç tümevarımı (tree induction) olarak adlandırılır. Ağaç tümevarımı yöntemi öz bilgi keşfinin en yaygın yöntemlerinden biridir. Sınıflama veya tahmin için kullanılabilecek ağaç benzeri örüntüleri keşfetme

için bir yöntemdir. Karar ağaçları bilgi keşfi sırasında pek çok test gerçekleştirerek, hedefi tahmin etmede en iyi sırayı bulmaya çalışırlar. Her bir test karar ağacındaki dalları oluşturur ve bu dallar da diğer testlerin gerçekleşmesine neden olur. Bu durum, test işleminin bir yaprak düğümünde (leaf node) sonlanmasına kadar devam eder. Kökten hedef yaprağa kadar olan yol, hedefi sınıflandıran “kural” olarak adlandırılır. Kurallar “eğer-sonra” (if-then) yapısındadır.

Karar ağacına dayalı analizler:

- Belirli bir sınıfın olası üyesi olacak elemanların belirlenmesinde (segmentation),
- Çeşitli vakaların yüksek, orta, düşük risk grupları gibi çeşitli kategorilere ayrılmasında (stratification),
- Gelecekteki olayların tahmin edilebilmesi için kurallar oluşturulmasında,
- Parametrik modellerin kurulmasında kullanılmak üzere çok miktardaki değişken ve veri kümesinden faydalı olacakların seçilmesinde,
- Sadece belirli alt gruplara özgü olan ilişkilerin tanımlanmasında,
- Kategorilerin birleştirilmesinde ve sürekli değişkenlerin kesikli değişkene dönüştürülmesinde yaygın olarak kullanılmaktadır.

Bir karar ağacı tümevarımına, boş bir ağaç ve eğitim kümesinin tamamı ile başlanır. Tümevarım yordamı yinelemeli bir süreçtir ve her bir yineleme dört adımdan oluşur. Bu adımlar aşağıdaki gibi açıklanabilir (Zorman vd., 2001,110):
Adım 1: Eğer eğitim kümesindeki tüm eğitim nesneleri aynı sonuca sahip ise bu sonuç ile bir yaprak oluşturulur ve dördüncü adıma geçilir.

Adım 2: Bulgusal değerlendirme fonksiyonunun yardımı ile kökten mevcut düğüme kadar olan yol üzerinde henüz kullanılmamış tüm nitelikler arasından en iyi nitelik bulunur. Seçilen nitelik üzerinde bölünme işlemi ile içsel bir düğüm yaratılır. Sonra eğitim kümesi alt kümelere ayrılır.

Adım 3: Eğitim nesnelerinin her bir alt kümesi için birinci adıma gidilir.

Adım 4: Bir düzey yukarı çıkılarak adımlar tekrar edilir.

Yukarıdaki adımlardan da anlaşılacağı gibi karar ağacı algoritmalarında iki temel işlem gerçekleştirilmektedir. Bu işlemler; bölme (splitting) ve budama (pruning)

işlemleridir. Algoritmaların sonlanması ise uygulanan durdurma kriterine göre olmaktadır. Bunlar kısaca şöyle açıklanabilir.

- **Bölme:** Bu işlem, verilerin daha küçük alt kümelerine ayrılmasını içeren yinelemeli bir süreçtir. İlk yineleme tüm veriyi içeren kök düğümünü ele alır. Bundan sonraki yinelemeler, verinin alt kümelerini içeren türev düğümler üzerinde işlem yapmaktadır. Her bölme işleminde, değişkenler analiz edilir ve en iyi bölme seçilir.
- **Budama:** Bir ağaç oluşturulduktan sonra, istenmeyen alt ağaçlar veya düğümler içerebilir. Budama işlemi ile bunlar ayıklanır. Bir başka ifade ile budama işlemi bir karar ağacını daha genel yapmak için kullanılmaktadır.
- **Durdurma kriteri:** Ağaç oluşturma algoritmaları çeşitli durdurma kuralları içerirler. Bu kurallar genellikle, maksimum ağaç derinliği, bir düğümde bölme için ele alınan minimum eleman sayısı ve yeni bir düğümde olması gereken minimum eleman sayısı gibi çeşitli faktörlere dayanır. Eğitim işlemi gerçekleştikten sonra, yeni bir veri örneği için elde edilen ağacın yardımı ile ağacın en tepesinden başlanarak ve bir yaprak düğümü ile karşılaşılıncaya kadar olan dalların oluşturduğu yol izlenerek tahmin yapılabilir. İzlenecek yol, yeni örnek içindeki bağımsız değişkenlerin değerleri üzerinde elde edilen karar ağacının bölme kuralları uygulanarak belirlenir (17).

Karar ağaçlarını matematiksel açıdan tanımlayıp tablosal ve ağaç çiziminde görmek anlatımı netleştirecektir.

$D=\{t_1 \dots t_n\}$ ' bir veritabanı kabul edelim.

Buradaki her t_i , $t_i=\langle t_{i1} \dots t_{i2} \rangle$ den oluşmaktadır ve bu veritabanı $\{A_1, A_2, \dots A_n\}$ alanlarından oluşmaktadır.

Algoritmalar kullanılan metodlarda önemlidir. Örneğin; karar ağacı oluşumunda kullanılan algoritma önemlidir. Karar ağacının şekli kullanılan algoritmalara göre şekillenmektedir ve her algoritmada farklı ağaç şekli ortaya çıkabilir. Farklı ağaç yapıları, farklı sınıflandırma sonuçları oluşmasını sağlayabilir. Kök diye adlandırılan ilk düğümü oluşturan A_1 'nin farklılık göstermesi, sonra yer alan yaprak basamağına ilerlerken takip edilen yolu ve buna bağlı olarak sınıflandırmayı da değiştirecektir.

Hem kök bağı hem de bu köke bağlı bağı biçimlendirilmesinde mühim olan ölçüt o bağıdan dallanmasında veritabanında kalan kısmın eşit bölümlere ayrılması gerekmektedir. Örneğin veritabanında olumlu/olumsuz/çekimsiz şeklinde 3 farklı bulunan cevap varsa üç eşit parçaya ayrılması uygun olacaktır. Bu şekilde davranılmasının sebebi en kısa mesafeden sonuç ve sınıfa ulaşmak içindir.

Aşağıda verilen tablo d veritabanının t_i tablosudur. Cinsiyet,Kilo,Boy alanları ise A_i 'lerdir. Beden alanı sınıfları (C) oluşturmaktadır. Beden değişkeni bağımlı değişken olarakta adlandırılabilir.

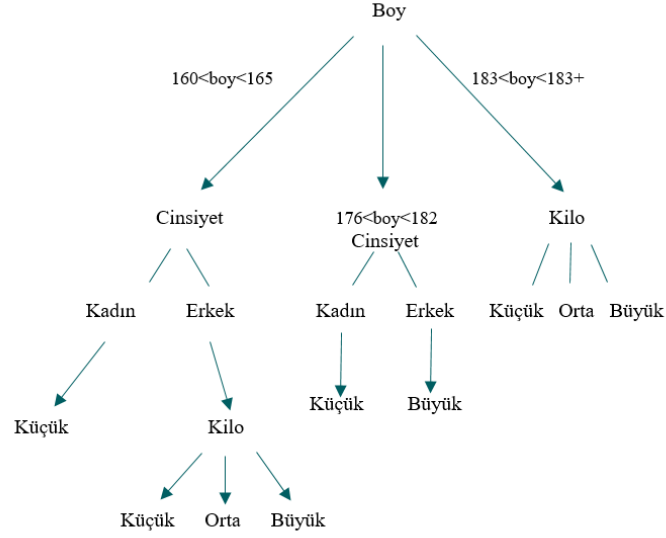
Tablo 5 Karar Ağacı Beden Tablosu

Cinsiyet	Kilo	Boy	Beden
K	48	170	Orta
K	49	151	Küçük
K	52	158	Orta
K	56	165	Orta
E	59	160	Küçük
K	61	159	Orta
E	62	162	Küçük
E	63	174	Orta
K	68	168	Orta
K	69	177	Büyük
E	72	170	Orta
E	74	165	Küçük
E	85	175	Orta
E	85	190	Büyük
E	98	190	Büyük

Karar ağacında boy,kilo ve cinsiyet değişkenleri birer düğümü temsil eder ve her düğümden sonra dallanma oluşturularak alt düğümlere geçiş sağlanacaktır.

Daha öncede belirtildiği gibi farklı algoritmalar ile farklı türde ağaçlar kurulabilir. Fakat her farklı ağaç,kendisi gibi farklı sınıflandırmalar oluşmasını sağlar. Öyle bir değişken seçilmelidir ki verilen yanıt ne olursa olsun, eldeki diğer

değişkenlerle karşılaştırıldığında veritabanı iki eşit parçaya bölünmeli. Bu sebeple düğümlerin ve düğümlerden dallara ayrılmanın kararı oldukça önemlidir.



Şekil 6 Beden Tablosu Karar ağacı

Bu analizin veri madenciliğinde kullanılma sebepleri ;

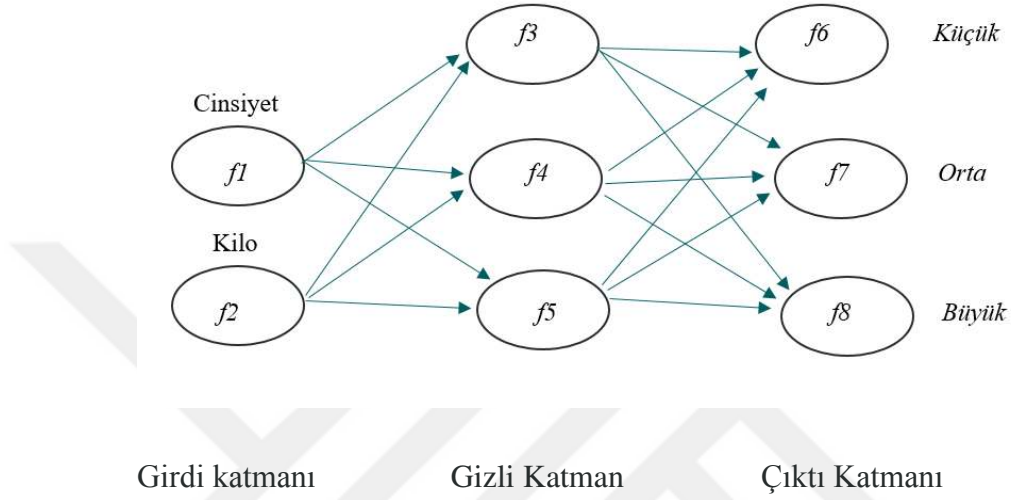
- Düşük maliyet.
- Anlaşılması ve yorumlanması kolaylığı
- Veri tabanına kolay entegre edilebilirliği
- Güvenirliği yüksek

olmasıdır.

1.4.3.2. Yapay Sinir Ağları

Yapay sinir ağları biyolojik sinir ağlarını temel alarak geliştirilmiş bir bilgi işleme metodudur. Yapay sinir ağlarının oluşma biçimi yapay sinir hücrelerinin bağlantı şekilleri birbiriyle türevli şekilde olmasıdır ve genel olarak katmanlı yapıda düzenlenmeleridir. En belirgin özellikleri bağlı nöronlar, bağlantılar arasındaki ağırlıkların belirlenmesi ve ateşleme fonksiyonudur. Sinir ağlarının katmanları iki yada üç katmandır. Bu katmanlar girdi, gizli ve çıktı şeklinde tarif edilir. İki katmandan oluşan sinir ağlarının gizli katmanları bulunmaz.

Yapay sinir ağlarını oluşturan her nöronun bir iç hali vardır. Bu iç hal ise aktivasyon veya aktivasyon seviyesi olarak adlandırılır. Bu seviye, gelen girdileri tanımlayan bir fonksiyondur. Ağ içerisindeki nöron aynı doğal nöron hücrelerinin yaptığı gibi bağlı nöronlara sinyal gönderir. Nöronlar her seferinde bir sinyal gönderir fakat bağlı olan birden fazla nörona sinyal iletilir. Giden sinyal, gönderilen nöronun giriş fonksiyonu olur(18).



Şekil 7 Yapay Sinir Ağları

Karar ağacında verilen örnek tabloya göre yapılmıştır.

1.4.3.3. Genetik Algoritmalar

Genetik algoritmalar, doğal hayatta gözlemlenen prensiplerden esinlenilmiş bir stratejidir.

Genetik algoritma, problem çözümleri için ilk basamakta rastgele başlangıç çözümleri belirler. Performansı yükselten çözümler oluşturmak için bulunan başlangıç çözümleri birbiriyle eşleştirir. Bu biçimde eşleşmeler sağlanarak yeni ve optimal çözümler bulunmaktadır. Buradaki amaç, organizma popülasyon evriminde olduğu gibi, en iyi çözümlerin bulunduğu optimal çözüm kümesini meydana getirmektir.

Çözüm uzayından rastgele seçilen noktalar üzerinde operatörler yardımı ile daha iyi noktalara ulaşmak hedeflenir. Çözüm uzayı çok büyük problemler için, bu arama sırasında iyi sonuç vermeyecek alanlar belirlenir ve gereksiz arama yapılmaması

sağlanır. Hızlı bir şekilde optimal çözüme ulaşmak açısından oldukça iyi bir tekniktir.

Genetik algoritmalar, çizelgeleme, tesis yerleşimi, hat dengeleme, atama ve optimizasyon problemlerinin çözümü ile finans, pazarlama ve üretim gibi alanlarda uygulanmaktadır(19).

1.4.3.4. K-En Yakın Komşu

En yaygın algortimalardan birisidir. İngilizce kaynaklarda K-Nearest Neighbor ya da K-NN olarak geçmektedir. T. M. Cover ve P. E.Hart’a örnek veri noktasının mevcut olduğu sınıfın ve -Nearest Neighbor’un k değeri göz önüne alınarak tayin edilen bir sınıflandırma yöntemidir.

Sınıflama yapılırken veritabanındaki kayıtların herbirini diğer kayıtlarla olan mesafesini hesaplar. Fakat bir kayıt için diğer kayıtlardan sadece k adedi gözönüne alır. Algortimanın isminden de anlaşılacağı gibi k adet kayıt, K-en Yakın Komşu algortimasının literatürde geliştirilmiş türevleri mevcuttur. Genelinde aynı davranış kullanılarak işlem yapılmaktadır.

Örnek tabanlı öğrenme algortimalarına ilk olarak K-NN algoritması verilebilir. Öğrenme prosedürü verilen eğitim veri grubundaki verilere bağlı şekilde gerçekleştirilmektedir. Eğitim setinde yer alan örnekler, yeni karşımıza çıkan örneğin kendisine benzerliğine göre sınıflandırılmasını sağlamaktadır(20).

Coğrafi bilgi sistemlerinde kullanımı fazladır; bir noktaya en yakın ilçe,istasyon gibi belirlemeler k-en yakın komşu algoritmasının temelini oluşturur.

1.4.3.5. Naive Bayes Sınıflandırıcısı

Makine öğreniminde öğreticili öğrenme alt sınıfındadır. Yani sınıflandırılması gereken sınıflar ve örnek verilerin hangi sınıflara ait olduğu bellidir. Naive Bayes sınıflandırıcısı Bayes teoreminin bağımsızlık önermesiyle basitleştirilmiş halidir.

$$P(Z|X) = \frac{P(X|Z) P(Z)}{P(X)}$$

$P(Z|X)$; X olayının tahakkuk ettiği durumda Z olayının meydana gelme olasılığıdır.

$P(X|Z)$; Z olayının tahakkuk ettiği durumda X olayının meydana gelme olasılığıdır.

$P(Z)$ ve $P(X)$; Z ve X olaylarının apriori olasılıklarıdır.

Burada apriori olasılığı Bayes teoreminine öznellik katar. Bir başka şekilde anlatmak gerekirse örneğin $P(Z)$ daha veri elde edilmeden Z vakası hakkında elde edilen bilgidir. Diğer taraftan $P(X|Z)$ ardıl olasılıktır çünkü veri toplanmasının ardından, Z olayının gerçekleşmiş olduğu durumlarda X vakasının gerçekleşme olasılığı hakkında bilgi verir.(21)

Naive Bayes, devamlı bir şekilde veri akışıyla etkili olmaz.Bu sebeple devamlı nitelikleri barındıran bağımlı veya bağımsız değişkenler sıralı kategorize edilmelidir. Naive Bayes, modelinin içeriği ele alındığında, her girdiye bağlı çıktının analiz kümesinde ne sıklıkla oluştuğunu hesaplamaktadır. Elde edilen bu değer öncelikli olasılık olarak olarak tanımlanır. Aynı zamanda Naive Bayes modelinde her bir bağımsız ve bağımlı nitelik birleşimlerinin oluşma sıklığı elde etmektedir. Bu oluşum sıklığı ilk olarak olasılıkların kombine edilmek şartıyla tahminde kullanılır.(22).

1.4.4. Kümeleme Analizi

Veri madenciliği yapısını meydana getiren testlerden biri de kümeleme analizidir. Sınıflandırmayla benzer özellikler barındırmaktadır. Kümeleme analizini sınıflandırmadan soyutlayan ayırt edici özellik, kümelerin daha önceden seçilmiş olmasıdır. Sıklıkla yapay sinir ağları ve istatistiksel metodlardan kümeleme analizinden faydalanılmaktadır. Kümeleme analizinin hedefi, bir ya da birden fazla özellik bakımından aynı veya benzer olan nitelikleri ve değişkenleri belirlemektir.

Kümeleme analizinde ana hedef; küme üyelerinin benzer özelliklerini taşıyan,fakat nitelikleri bakımından ayırtıran kümeleri meydana getirmek ve veritabanındaki kayıtların farklı kümelere bölünerek ayrılmasıdır (Akpınar,2000,s. 4). Kümeleme analizi, küme üyelerinin aralarındaki benzer özelliklerden yararlanarak, bir araya getirir. Dolayısıyla, bu üyeler değişkenlik gösteren nitelikleri nedeniyle, kümelere ayrılırken,kümeler içi benzerlik ve kümeler arası değişkenlik en yüksek derecede sabitlenmiş olur. Dolayısıyla kümeleri oluşturan üyeler birbiri arasında homojen yapıyı korurken, başka kümelerin üyeleriyle heterojen yapı sağlar.

En sık faydalanan kümeleme algoritması k-ortalamlar tekniğidir.

K-yaklaşık değer tekniği 1967 yılında Mac Queen tarafından bulunmuş ve takip eden yıllar boyunca çok sayıda uygulama, araştırma, geliştirme alanında fazlasıyla tercih edilmiş bir kümeleme algoritmasıdır. K-ortalamlar tekniğinde K kadar küme meydana gelmekte ve elde edilen her küme dahilinde bulunan verilerin ağırlıklı ortalamaları neticesinde bir değer elde edilmektedir. Küme içinde elde edilen değere en yaklaşık olan değer “Küme merkezi” olarak tanımlanmaktadır. (Berkhin, 2002, s: 15). K-yaklaşık değer tekniği ilk olarak mevcut verileri, kümelerin ortalamalarına paralel gruplara ayırır ve k küme miktarı kullanıcı isteğine bağlı olarak belirlenir. Burada adı geçen ortalama (yaklaşık değer), küme merkezidir. Sonrasında dahil edilen her veri merkez noktaya en yakın olan merkez küme içerisine yerleştirilir. Dahil edildiği küme elemanlarının ağırlıklı ortalama hesapları yinelenerek yeni bir küme merkezi değeri elde edilir. Bu yeni değer, sonra ki yapılacak işlemlerde var olan kümeyi temsil eder(23).

Kullanıcının hedefleriği kümeleme analizinin ve uygulama ortamına göre kümeleme analiznin aşamaları aşağıdaki gibi sıralanabilir. (Ball,1971):

- Gerçekçi Sıfırların belirlenmesi - Model oluşturmak
- Kümelere dahaylı tahmin
- Hipotez testi - Veri araştırma (inceleme)
- Hipotez elde etmek
- Veri yalınlaştırmak

Örnek olarak araştırmacı, bir çok yerde kümeler elde etmeden kontrol altına alınamayan geniş hacimli gözlemlerle karşı karşıya kalabilir. Kümeleme yöntemi bu tip verilerin indirgenmesinde yararlanılabilir. Bu yöntemi pazarlama incelemesi ortamında, bir uygulama yapmakla örneklendirebiliriz. Pazar uygulaması için bir çok ilçe belirlenebilir. Ancak belirlenen bu ilçelerin ekonomik etkenlere dayalı olarak sayısı düşürülmelidir. Bu ilçelerden birbiriyle benzer özellikte olanları küçük kümelere ayırıp kümelendirecek olursak, her kümedeki bir ilçe, bir test Pazar olarak belirlenebilir(Everitt,1993)(24).

Kümeleme uygulamalarının bir çoğu verilere bağlı uzaklıkları kullanır. Hiyerarşik kümeleme teknikleri en benzer komşu algoritması ve benzer olmayan komşu

algortimasıdır. Hiyerarşik olmayan kümeleme tekniklerinde k-ortalamlar tekniği kabul edilebilir. Yöntemde bir çok defa gruptama tekniğinden faydalanılmaktadır.

Bu teknikler, değişkenlerin birbirine benzer özelliklerinden veya birbirinden ayrılan özelliklerinden yararlanarak kümeyi alt kümelerine ayırmakla yükümlülük gösterir. iki tekniğin beraber kullanılması daha faydalı olacaktır. Bu anlayışla sürdürülen çalışmalar hem sonuçları karşılaştırmayı hem de bahsedilen iki yöntemden hangisinin daha etkili neticeler sunduğunu karşılaştırmaya olanak sağlar. Tablo 4'e en yakın komşu algoritması uygulandığında Tablo 5 ile karşılaşırsınız.

Tablo 6 Gözlem Tablosu

Gözlem	X1	X2
1	1	1
2	2	1
3	4	5
4	7	7
5	5	7

Tablo 7 Oluşan Kümeler Tablosu

Kümeler	
Küme 1	1,2
Küme 2	4,5
Küme 3	3,4,5
Küme 4	1,2,3,4,5

1.4.4.1.Hiyerarşik Kümeleme Yöntemi

Hiyerarşik kümeleme yöntemleri,kümeleri ardı ardına bağlama aşamasıdır, bu aşamadan sonra bağlı gruplar bir daha katiyen ayrı bir halde bulunamazlar. Hiyerarşik yöntemlerin ağaç diyagramları ile şemalize edilen sonuçlarına dedogram adı verilir(25).

Hiyerarşik yöntemler bir küme ağacı oluşturur. Bu küme ağacındaki herbir düğüm oğullara sahiptir; ağaç yapraklarla sona erer(26).

Hiyerarşik kümeleme yöntemlerini aşağıdaki biçimde 2 başlıkta inceleyebiliriz.

Toplama Teknikleri

Bu tekniğin ilk adımı $\{\frac{1}{2}[a(a-1)]\}$ muhtemel gözlem çiftinde bir benzerlik ya da mesafe matrisinin hesaplanmasıdır. İlk aşamada her küme bir gözleme aittir. Benzerlik ve ya mesafe matrisinin sonucuna göre en benzeyen iki küme birleştirilir. Sonrasında küme sayısı bir indirgenerek benzerlik matrisi tekrar meydana getirilir ve a birim mertebeli olarak sırasıyla a, (a-1), (a-2),...(a-r),...3,2,1 kümeye yerleştirilir (Everitt,1971). Bu yöntem her kaydı ayrı bir küme olarak kabul eder ve bu kümelerden tek büyük bir küme oluşuncaya kadar kümeleri birleştirir. Bu tekniklerden bir kaçı konunun devamında açıklamıştır(25).

Tek Bağlantı Tekniği: Mesafe ya da benzerlik matrisinden faydalanılarak en yakın iki gözlem ya da küme birleştirilir ve bu birleştirme prosedür tekrarlanır. Tek bağlantı tekniği sonuçlarını, kümeleri yapıdaki dalların temsil ettiği bir ağaç taslağında ya da dendogramda yansıtabilir. (25).

Tam Bağlantı Tekniği: Tek bağlantı tekniğinin tersine çalışma prensibi olan bir tekniktir. Bu teknik ile ulaşılan mesafe veya benzerlik matrisinden faydalanılarak en yakın iki küme ya da gözlem bir araya getirilmektedir. Tek küme içerisindeki mevcut gözlemlerin mesafelerinin belirli bir değişkenden ufak olması halinde tüm kümelerin güvenilir bir şekilde meydana getirilmesini garanti edememektedir(25).

Ortalama Bağlantı Tekniği: Bu yöntemde, iki kümeyi birbirinden ayıran özellik, bir kümedeki öge çiftleri ile başka bir kümedeki öge çiftleri arasındaki yaklaşık fark olarak alınır.(Everitt,1981)(25). Bu yöntemin farklılık gösteren tipleri de mevcuttur. Bu tipler arasında en sık yararlanılan göslem çiftlerine bağlı uzaklığın aritmetik olarak yaklaşık ortalaması bulmayı hesaplamaktadır. Ayrıca bu yöntem sıklıkla biyoloji biliminde kullanılmakta olduğu gibi, sosyal bilimlerde de kullanımını gün geçtikçe arttırmaktadır. Genel olarak ortalama bağlantı ve tam bağlantı yöntemlerinde birbirine yakın dendogramlar meydana gelmektedir. Ancak bu bağlantı yöntemleri arasındaki uzaklık farklılık gösterdiğinden, birleştirmeler birbirinden farklı seviyelerde elde edilebilmektedir.

Ward's Tekniği: Bu yöntemde hedef, gruplara ait varyansı minimum seviyeye sabitlemektir. Bu amaçla aşağıdaki hata kareler toplamına (ESS) ilişkin formülden yararlanılır:

$$ESS = \sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2$$

Kümeleme sürecinin ilk adımında, her bir gözlem bir küme olduğundan ESS sıfırdır. Ward tekniği. ESS' de minimum artışta sonuçlanan gruplar veya gözlemleri elde edilmesi ile devam eder(25).

Ayrırma Teknikleri

Ayrırma Teknikleri, toplama tekniğinin tam tersi şekilde çalışmaktadır. Başlangıçta tüm kayıtlar tek büyük bir kümenin bir parçası olarak kabul edilir ve daha sonra bu küme iki ve ya üç alt kümeye bölünür. Her alt küme birbirine benzemeyen alt kümelerle bölünerek işlem devam eder. Ortaya çıkan kümeler içerisinde en iyi olanları tercih edilir.

1.4.4.2. Hiyerarşik Olmayan Kümeleme Teknikleri

Parametrelerdense birimlerin K adet kümede birleştirilmesi için tasarlanmış teknik, hiyerarşik yapıya sahip olmayan kümeleme yöntemidir. Küme sayısı (K) önceden belirlenebilir ya da kümeleme tekniğinin bir kısmı olarak belirlenir. Mesafe matrisinde belirlenmiş olması mecburi değildir ve asıl verinin bilgisayarın çalışması boyunca depolanması gibi bir zorunluluk yoktur. Büyük veri kümelerinde Hiyerarşik Tekniklerdense Hiyerarşik olmayan teknikler, daha uygulanabilirlik sağlar. (Johnson ve Wichern,1988).

K- ortalama tekniği Hiyerarşik olmayan kümeleme yöntemlerinde en çok kullanılan tekniklerden bir tanesidir(25).

K- Ortalama Tekniği: K-ortalama yöntemi en kısa mesafedeki değerlere sahip her öğeyi, kümelerle ayıracak algoritmayı açıklamak amaçlı Mac Quenn tarafından savunulmuştur. Bu yöntemin adımları aşağıdaki gibidir :

1. Her birim K adet kümeye ayrılır.
2. Birimler, değer açısından en yakın hangi kümeyeysen ona atanarak ilerlenilir. Sonraki adımda birimler ölçümlenerek kümeye yeni değeri belirlenir.

3. Bu adımlar çıktılarda artık atama yapılamayana kadar tekrar edilir(Norusis,1993; Atamer,1992).

Küme Sayısının Belirlenmesi:

Parametre tercihleri ve küme sayısının belirlenmesi kümeleme analizinin sağlıklı bir sonuç çıkması için önemlidir. Aşağıdaki denklem küçük örneklemelerde küme sayısının belirlenmesi için sıkça kullanılmaktadır:

$$k = \left(\frac{n}{2}\right)^{\frac{1}{2}}$$

Bir başka yöntem ise Mariot tarafından önerilmiştir;

$$M = k^2 |W|$$

En küçük M değerini veren küme sayısı gerçek küme sayısını gösterir. Grup içi kareler toplamı matrisi ise W ile ifade edilmiştir.

Calinsky ve Harabasz'ın geliştirdiği teknikte ise ;

$C = [iz(B)/k - 1] - [iz(W)(n = k)]$ eşitliğinin en büyük değeri almasını sağlayan k değeri küme sayısı olarak ifade edilir. B gruplar arası kareler toplamı matrisini simgelerken, W grup içi kareler toplamı matrisini simgeler(Atamer,1992)(25).

1.4.4.3. Model Tabanlı Yöntemler

Varolan verileri, matematiksel bir modelleme ile açıklamayı amaçlar. Yöntemde istatistiksel yaklaşım ve yapay zeka yaklaşımı olarak adlandırılan iki ana yaklaşım mevcuttur.

İstatistiksel Yaklaşım: İstatistiksel yaklaşım 2 kavramdan oluşan bir metottur. İlk basamak kümeleme olup ikinci basamak ise tanımlama adıımıdır.İçerisindeki bir çok kavram, taktik veya kümelerin oluşturulmasında olasılık araçlarından faydalanan istatistiksel yaklaşımdır.

1.4.4.4. K-means Kümeleme

K-means kümeleme tekniği, N adet veri değişkenlerinden meydana gelen bir veri grubunu başlangıç aracı olarak verilen k adet kümeye ayırmaktır. Hedef; ayırma işleminin ardından ortaya çıkan kümelerin, küme içi en benzer olma durumu ve kümeler arası benzerliklerinin en az seviyede olmasını sağlamaktır. En yaygın kullanılan yöntem olmasının yanı sıra, uygulaması da basit olan bir yöntemdir. Büyük ölçekli verilerin hızlı ve faal şekilde gruplandırılmasını sağlamaktadır. “K” gerekli olan sabit küme sayısının anlamını taşımaktadır. K-means algoritması her verinin elemanı olduğu kümeye olan uzaklıkları toplamını daha az seviyeye indirgemektedir. Aynı zamanda karesel hatayı minimuma düşürecek K adet kümeyi bulma amaçlı çalışmaktadır. K-means algortması ile grup dahilinde ki benzerliğin yüksek, grup dışı benzerliğin ise düşük seviyede olması halinde kümelemenin doğruluğundan söz edilebilir.

K-means tekniğinin işlev anlayışına göre, başta her kümeye ait merkez noktasını ya da ortalaması anlamını taşımak üzere K tane değişken gelişigüzel bir şekilde belirlenir. Kalan diğer değişkenler, kümelerin ortalama değer mesafeleri göz ardı edilmeksizin, kendisine en yakın olan kümelere eklenir. Sonrasında her kümenin ortalama değer analizi yapılarak yeni küme merkezleri elde edilir ve yinelenen değişkenlerin merkeze olan mesafeleri hesaplanır. Bir farklılık gözlemleninceye dek ve en anlam yüklü neticeye varıncaya dek algoritma yinelenerek sürdürülür.

Algoritma ana olarak 4 evreden meydana gelir:

1.Küme merkezlerinin tespiti

2.Merkez içerisinde olmayan verilerin uzaklıkları göz önünde tutularak gruplandırılması

3.Elde edilen kümelendirmeye göre, yeni merkezlerin tespiti (ya da eski merkezlerin yeni merkezlerin yerini alması)

4.Kararlı bir durum gözlemleninceye dek 2. ve 3. Evrelerin yinelenmesi(27).

1.4.5. Gelişmiş Veri Madenciliği Teknikleri

Kurumların ihtiyaçlarının gelişen teknoloji sebebiyle artması müşteri kayıtları, hastane kayıtları, kişisel bilgiler, satış pazarlama kayıtları gibi kurumsal hafızaya destek olacak, geleceğe yarar sağlayacak bilgilerin depolanması artmaktadır. Zamanla büyüyen bu veritabanları, büyük maliyetlerle gerek fiziksel gerek bulut çözümlerle saklanmaktadır. Veriler geleceğe katkı sağlamadığı sürece bu maliyetler kurumlara ancak yük olacaktır. Dolayısıyla eldeki verilerin bu tezin konusu olan Veri Madenciliği yöntemiyle işlenerek katma değer sağlıyor olması her kurum için önemlidir. Eldeki verinin işlenmesi aynı sektörde bulunan rakiplerinin önüne geçmesini, bireylerde genetik olarak aktarılan hastalıkların erken tanı ile gözetim altında tutulması, bir spor müsabakasında rakip takımın hamlelerini önceden tahminleyip galip gelebilmeleri, dolandırıcılık tespiti için önseziler yerine verilerle tahminlenmesi açısından önemlidir.

Berkeley üniversitesinde yapılan bir araştırmaya göre bir yılda ortalama bir trilyon bayt yani bir milyon terabayt veri yedeklenmekte ve bu sayı her geçen yıl daha da yukarıya çıkmaktadır. Sürekli olarak artan bu veri kitleleri dağınık ve düzensiz bir şekilde bulunmaktadır. Veri madenciliğinin amacı, kullanıcının bilgi çıkarma sürecinde olabildiğince az bilgi sağlaması ve en yüksek ölçüde otomatik gerçekleşmesidir(30).

Veri madenciliği, verilerin bütünsel ve yapısal kavramda analiz etmektedir. Yapısal özelliklerinde farklılıklar olan verilerin veri madenciliği yönteminde kullanılacak şekle evrilmesinde aracı olarak kullanılır.

Metin Madenciliği: Metin üzerinden biçimlendirilmiş olan veriye ulaşmayı amaçlamaktadır. Metinlerin kümelendirilmesi, sınıfların meydana getirilmesini, duygusal çözümleme, metin özetleme, varlık-ilişki modelinin tayin edilmesi gibi yöntem çalışmalarını amaçlar.

Örüntü tanımlı varlıkların bulunması: Metin içerisinde bulunan bazı özel bilgi alanlarının metin madenciliğine konu olması muhtemeldir. E-posta adresi, telefon numarası, adres gibi alanlar veri olarak alınabilir.

Duygu analizi: Bir konu üzerine yorumlanmış olan mesajların ve ya yazıların pozitif ya da negatif şekilde 2 sınıflandırma yapılabilir.

Enformasyon faydası: Web sayfaları, adresleri veya dosya sistemi üzerinden kullanıcı bilgileri, dosya isimleri, izin bilgileri gibi bilgilerin toplandığı aşamadır(28).

Multimedya Madenciliği: İstatistiksel bağlar ve modelleri keşfetmek için yüksek boyuttaki verilerin çözümlemesini sağlar. Veri arttığında anlamlı bağlantı keşfetmek ve analiz için yazılımlar yardımcı olur. Bu veriler, hükümet tarafından toplumun sosyal sistemlerini geliştirmek ya da pazarlamada müşteri davranışlarını analiz etmek gibi bir çok alanda kullanılmaktadır.

Ses Madenciliği: Ses sinyalinin otomatik olarak analiz edilmesi ve araştırılması tekniğidir. Örneğin; Shazam ve Siri

Video Madenciliği: Bu grupta 3 tip video madenciliği bulunmaktadır. Üretilmiş (Filmler,yeni videolar), İşlenmemiş (Trafik videoları, gözetim videoları), Medikal (Ultrason).

Yüksek düzey bilgi içeren videoları ise aşağıdaki şekilde anlatabiliriz.

- Tetikleyici olaylardan (belirlenen bölgelere araç/insan giriş çıkışı),
- Tipik ya da sıradışı insan veya cisim hareketlerinden,
- Hareketlerin isimlerle kategorize edilmesinden (yürümek,bisiklete binmek)

oluşur.

Görüntü Madenciliği: Görüntü madenciliğinin amacı geniş görüntü veritabanı içerisinde değişik görüntülerle birleştirme elde edebilmektir.

Örneğin; Medikal teşhis, uzay araştırması, uzaktan tanıma, parmak izi tanıma, yüz tanıma gibi alanlarda kullanılır.

Web Madenciliği: Veri Madenciliği, büyük miktardaki birbirinden ilgisiz görünen veriden anlamlı bilginin çıkarılması veya içerisinde gelecek ile ilgili tahminde bulunabilmemizi sağlayabilecek bağıntıların bilgisayar programı kullanarak aranmasıdır. Günümüzde birçok işlemin internet üzerinden yürütülmesi sonucu, çok büyük oranda veri yığınları WWW (World Wide Web) ortamında oluşmuş durumdadır. İnternet üzerinde bir siteye bağlanan herkes bağlantı loglarını tutan sunucularda iz bırakır (IP adresi, browser, cookie'ler, gibi). Web

üzerindeki veri yığınları çok farklı standart ve tiplerde bulunmaktadır. Bu veri yığınları aşağıdaki şekilde sıralanabilir.

- Web sayfaları
- Access Log dosyaları
- Kullanıcı kayıt bilgileri
- Oturum ve hareket bilgileri
- Site yapısı ve içeriği

Web madenciliği yukarıda sayılan çeşitli yapıdaki web sayfaları dokümanlarını ve kayıt bilgilerini incelemek, bunlardaki kalıpları keşfetmek için veri madenciliği tekniklerinin kullanılması olarak tanımlanabilir. Bir web sitesinin kullanımının artırılması için, bu siteyi kullanan kullanıcıların ilişki ve hareket eğilimleri ortaya çıkarılarak çeşitli ipuçları elde edilebilir. Bir online alışveriş sitesindeki kayıtlarından, alışveriş yapan kullanıcıların hareketleri incelenerek, elde edilen sonuçlar satışların artırılması için kullanılabilir(29).

Web madenciliğinde kullanılacak verilerin yapısına göre 3 gruba ayrılır.

- Web İçerik madenciliği
- Web yapı madenciliği
- Web kullanım madenciliği

Web İçerik Madenciliği: Web içerik madenciliği, yapay zeka ve akıllı yazılım programları, bilgi tarama teknikleri kullanarak web kaynaklarının içeriklerinden yararlı bilgiyi elde etme tekniğidir(30).

Web Yapı Madenciliği: Web yapı madenciliğinin amacı web sayfaları arasındaki linkleri takip ederek bilgi üretmektir(29).

Web Yapı Madenciliği: Web Kullanım Madenciliğinde kullanılan veriler, web üzerindeki çeşitli sunucularda tutulan kullanıcı erişim hareketlerinin yer aldığı çeşitli log dosyalarından elde edilir(29).

1.5.Literatür araştırması

Veri madenciliği, günümüzde başta pazarlama, bankacılık ve sigortacılık gibi sektörler olmak üzere bir çok farklı alanda kullanılmakta ve iş dünyasında başarılı sonuçlar edinmektedir. Pazarlama alanında, müşteri yapısının ve satın alma alışkanlıklarının analizi ile, hem mevcut müşterilerin sadakatının artırılması hem de müşteri portföyünün genişletilmesinde; bankacılık alanında ise müşterilerin kredi taleplerinin ve risklerinin değerlendirilmesi, müşteri kategorilerinin tespiti ve diğer finansal göstergeler arasındaki ilişkilerin test edilmesinde; sigortacılık alanında ise yine müşteri risklerinin değerlendirilmesi ve suçluların karakteristik özelliklerinin bulunması ve bu şekilde suç ile mücadele olacak sayılabilir. (SAS, 2009).

Literatürde veri madenciliği modelleri üç farklı başlık altında incelenmektedir (31), bunlar;

- Sınıflama (Classification) ve Regresyon (Regression),
- Kümeleme (Clustering),
- Birlikte Kuralları (Association Rules) ve Ardışık Zamanlı Örüntüler (Sequential Patterns)

Pazarlama alanında ek sık kullanılan veri madenciliği analiz yöntemi Pazar Sepet Analizi'dir (Market Basket Analysis). Bu teknik bir müşterilerin birlikte satın aldığı ürünlerin incelenmesi sonucunda, herhangi bir ürünü alan müşterinin satın alma potansiyeli yüksek diğer yardımcı ürün kategorilerinin de belirlenmesini sağlar ve tüketicilerin satın alma alışkanlıklarının belirlenmesinde etkili sonuçlar üretir. Yaygın kullanıma alanı ise birden fazla ürünün birlikte satın alındığı perakende sektörüdür.

Bu alanda çalışmaları örneklendirmek gerekir ise; daha iyi ürün konumlandırma kararları verebilmek için Brjis çalışmalarında çapraz satışları artırmak amacı ile sık olarak birlikte satın alınan ürünlerin tespiti yapılmıştır (32). Benzer şekilde ise Deckert ve Monien'in çalışmalarında farklı sinir ağı yaklaşımları ile yine satın alma alışkanlıklarına uyum açıklanmaya çalışılmıştır. Bu çalışmada ele alınan yöntemlerinin her biri yine perakende sektöründe birlikte satın alma kararlarının ölçümü için uygulama alanı bulmuştur.(33)

Sepet Analizi, birlikte satın alınan ürünler arasındaki ilişkileri inceleyerek, doğru konumlandırma ve doğru promosyon kararlarının verilmesini amaçlamaktadır. Müşterilerin fiyat esnekliğinin tespiti ve bunun sonucunda doğru promosyon stratejileri de ana ürünün rakipler karşısında cazip hale getirilerek, ürün karlılığının artırılmasını hedeflemektedir.

Barber ve diğeri çalışmasının lazerle göz ameliyatı için yaptığı analizde olduğu gibi, pazarlama sektöründe lojistik regresyon analizi sıklıkla kullanılmaktadır. Bu analizin amaç hedef ürün için en uygun müşterilerin kimler olduğunu tespit etmektir, bu şekilde çıkartılan müşteri profili ile ürünün doğru müşterilere pazarlanması için doğru pazarlama mesaj ve stratejileri oluşturulabilecektir (34).

Daha fazla satış üç şekilde olabilir; firma pazara yeni giren müşterilere satış yapabilir, kendi müşterilerine daha fazla satış yapabilir veya rakiplerinin müşterilerine satış yapabilir. “Churn” olarak adlandırılan bu son durum yani müşterin rakip markalar arasındaki geçişleri pazarın büyümediği sektörlerde öncelikli sorunlardan biridir, bunun nedeni de pazarlama ve promosyon stratejileri ile rakip ürünlerin zaman içinde daha avantajlı hale gelmesidir.

Günümüzde web teknolojilerinin yaygınlaşması, firmalar ve ürünler arasındaki rekabeti de farklı bir boyuta taşımış olmak ile beraber, firmalar için de özellik ile müşteri memnuniyeti ve sadakatini arttırmak için yeni fırsatlar da ortaya çıkartmıştır. Özer çalışmasında bulanık küme analizi kullanarak, on-line müzik servislerinde homojen potansiyel müşteri grupları olduğunu görmüştür. Her bir homojen grup için ise müşteri profillerine uygun olarak farklı pazarlama kampanyaları ve teklifler hazırlanmıştır (35).

Sektördeki bir başka çalışmada ise Türkiye’de faaliyet gösteren en büyük perakende hazır giyim firmalarında birinin 2005 yılına ait ilk beş aylık 42.432 satır alışverişi incelenmiştir. 5370 müşteriye ait olan 17164 farklı faturadaki bu veri üzerinde müşteri profillerinin ve satın alma alışkanlıklarının incelenmesi amacı ile yukarıda bahsedilen birliktelik kuralları ve kümeleme analizi metotları uygulanmıştır. Hem müşterilerin farklı nüfus yapılarına göre gruplanması ve kümelenmesi sağlanmış hem de belirlenen her bir grubun satın alma alışkanlıklara uygun konumlandırma ve fiyat politikalarının saptanması sağlanmıştır (36).

Veri madenciliği pazarlama sürecinin başından sonuna kadar uygulama alanı bulmaktadır. Bu müşteri profillerinin belirlenmesi ve potansiyel müşterilerin seçiminden başlar, bu hedef kitlenin satınalma alışkanlıkları ve pazarlama mesajının oluşturulması ile devam eder ve hatta iletişim kanalının tespitine kadar devam eder.

- **Kampanya şartlarını düzenlemek:** Müşteriler arasından kampanya için hedef seçilen bir müşteri profili baz alınarak, sadece bu profilin satın alma alışkanlıklarına hitap eden kampanyalar düzenlenmesi.
- **Müşterilere özgü satış politikaları düzenlemek:** Benzer müşterileri çeşitli kriterlere göre gruplamak ve müşteri profiline göre uygun pazarlama karmalarını (ürün, yer, fiyat ve tutundurma) belirlemek.
- **Özel Kampanyalar düzenleme:** Farklı profillerdeki müşterilerin farklı “yeni müşteri edinme maliyet” bulunmaktadır, burada yeni müşteri maliyet düşük profillerin belirlenmesi ve onlara özel kampanya bileşenlerinin belirlenmesi maliyetlerin azaltılması.
- **Yeni ürün geliştirmek:** Yeni ürün gelişimi sadece ürünlere yeni özellik ekleyerek olmamaktadır, bazı durumlarda ürünlerden özellik çıkartılarak da ürün müşterilere daha uygun hale getirilebilir. Doğru profil için doğru pazarlama karmasının belirlenmesi ve yeni ürünlerin tasarımı.

2. VERİ MADENCİLİĞİNİN SÜRECİ

Günümüzde işletmeler, büyük miktarlarda, farklı veriye sahiptir. İşletmelerin sahip olduğu veriler analiz teknikleriyle anlamlı ve işletmeyi ileriye taşıyacak bilgilerin elde edilmesi, işletmelerin stratejik kararlar vermesini sağlayacak ve gelişen, rekabetin her geçen gün daha da arttığı coğrafyada işletmenin rekabet gücünü artırır. Veri madenciliği, matematiksel ve istatistiksel analizler yardımı ile veritabanı içindeki saklı kalmış bağlantıları belirleyerek, işletmelere karar verme sürecinde daha hızlı ve doğru adımlar atmasını sağlar(37).

Bu analizlerin sonucunun doğruluğu, yetenekleri, deneyimi ve bilgi birikimini kadar very tabanının yeterliliğine de bağlıdır. Yani analiz sonucunun başarılı olmasında, verilerin doğru depolanmış olması, doğru sınıflandırılması, düzenli işlenmesi ve doğru yorumlanması önemlidir. Fakat analiz süreçlerinin karmaşık olması, sayısal bazda daha yüksek veriye gereksinimin doğmasını sağlamış, dolayısıyla verilerin depolanan alanlarının büyüklüğü manuel olarak kontrol edilemeyecek bir safhaya gelmiştir.

Eğer veriyi genel olarak iki bölümde incelememiz gerekirse enformasyonel veri ve operasyonel veri şeklinde incelememiz mümkündür. Enformasyonel veriyi, birey odaklı, tümleşik, zamanla oluşan ve konsolide edilmiş veriler şeklinde açıklayabiliriz. Operasyonel veriyi ise, uygulama odaklı, karışık, kısa sürede meydana gelen ve mükerrerlik olan veriler şeklinde tanımlayabiliriz(38).

Veri ambarı, tüm operasyonların, gizli kalmış verilerine inebilen, analiz yapılması için özel olarak modellenen ve tarihi geçmiş olan veri depolama sistemi olarak tanımlanır. Bir veri ambarının oluşmasında kullanıla yazılım yoktur. Yani veri ambarları, verilecek olan karar sürecine hususidir ve o karar süreci için modelleme yapılır.

Organizasyonu ve veriyi anlatacak yararlı ve tutarlı bilginin kaynağına ulaşmak veri ambarından beklenmektedir. Sistem tasarlanırken dikkat edilmesi gereken bazı hususlar vardır:

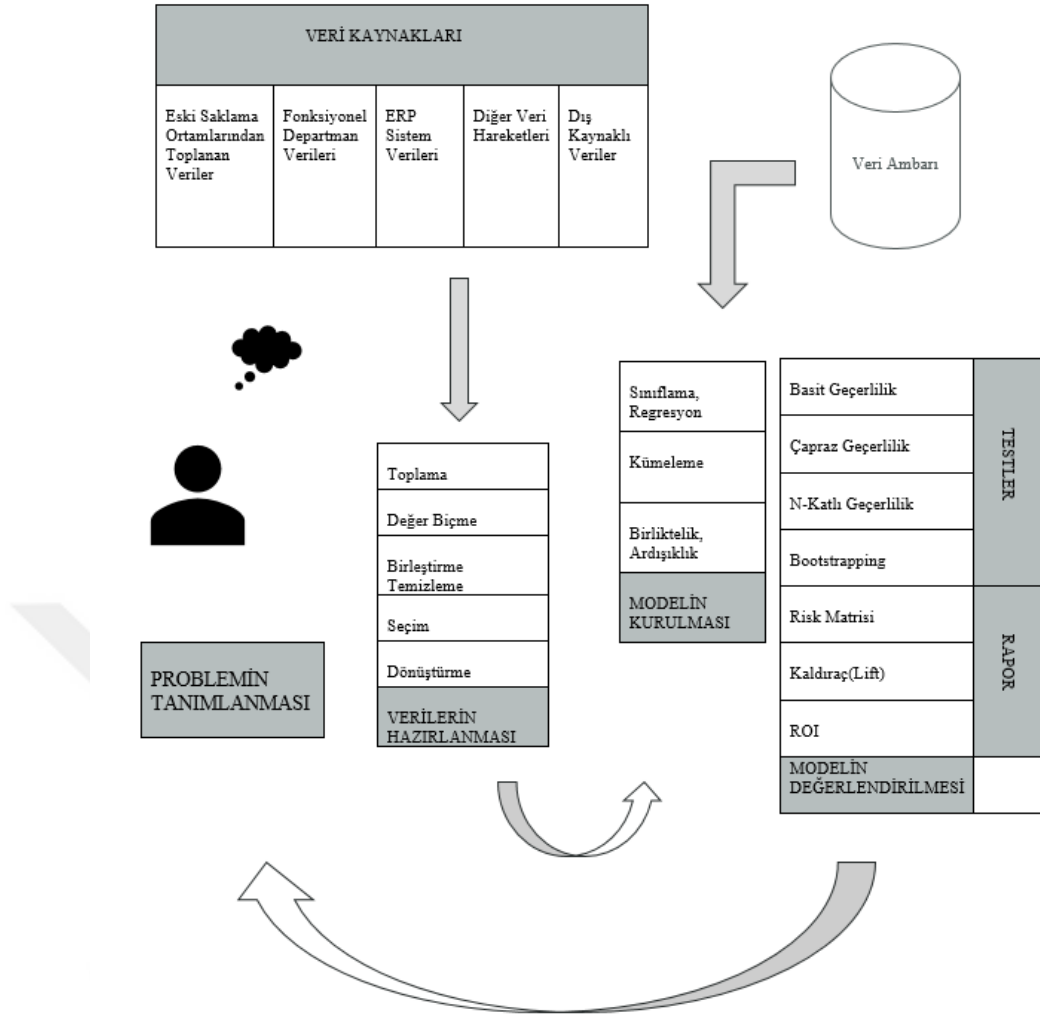
- Çözülmesi istenen problem sisteme ayrıntılı şekilde belirtilmelidir.
- Sistemden beklenen hedefler, limitler ve başarı kriterleri sıralanmalıdır.

- Sisteme ait arayüzler ve bileşenler arasındaki bağlantı ya da haberleşme ağları bilinmelidir.
- Gelecekte ait muhtemel iyileştirmeler, revizyonlar ve farklı sistemlere geçiş hakkında öngörü çalışmaları yapılmalıdır.
- Bütüne ait bir geliştirme, bakım ve destek verecek personel kaynağı oluşturulmalıdır(39).

Veri madenciliği, veritabanında yeralan çeşitli verilere dayanarak gizli kalmış, keşfedilmemiş verileri anlamlı hale getirmek, bu verileri karar verme ve gerçekleştirme sürecidir. Veri madenciliği süreci aşağıdaki aşamalardan oluşmaktadır:

- 2.1. Problemin Tanımlanması
- 2.2. Verinin Anlaşılması
- 2.3. Verinin Hazırlanması
- 2.4. Modelleme
- 2.5. Modelin Kurulması
- 2.6. Modelin Kullanılması
- 2.7. Modelin Değerlendirilmesi
- 2.8. Modelin Uygulanması
- 2.9. Veri Madenciliğinde Karşılaşılan Problemler

Veri madenciliği süreci en basit ifadesiyle Şekil 8’de gösterilmiştir. Bu şekilden de görüleceği gibi veri madenciliği süreci, verileri veri ambarından alır, bunları derler, düzenler ve yorumlar.



Şekil 8 Veri Madenciliği Süreci

2.1.Problemin Tanımlanması

Veri madenciliğinde başarılı çalışmalar için uygulamanın hangi amaç için yapılacağını temiz bir anlatımla tanımlanmasıyla gerçekleşir. Çalışmanın hizmet edeceği amacı, soruna odaklı ve açık bir dille anlatılmış şekilde olmalı, analiz sonuçlarının başarı düzeylerinin nasıl ölçüleceği tanımlanmalıdır. Sorun ile tam bağdaşmayan bir çalışma, sorunu çözmeye yetmeyeceği gibi başka problemlerin de doğmasına sebep olabilir. Alınabilecek yanlış kararlarda oluşacak olan maliyetlere ve doğru kararlarda sağlanacak faydalara ait tahminler bu başlık altında anlatılacaktır.

Bir çok türde veri madenciliği algoritmaları ile bilgi keşfi sürecinde görülen farklı türde problemler tahlil edilmektedir. Veri madenciliği görevlerini iki başlık altında inceleyebiliriz:

- Tanımlayıcı veri madenciliği verinin genel niteliklerinin ortaya koyulmasını hedefler.

- Kestirimci/Tahmin edici veri madenciliği veriden ulaşılabilecek olan tahminler sayesinde yeni stratejiler elde etmeyi hedefler(40).

Günümüzde her alanda veri madenciliği yer bulmaktadır ve aşağıda hangi sektörlerde hangi sorunlar için uygulama alanı bulduğu kısaca belirtilmiştir.

- Bankacılık: Risk çözümlemesi ve yolsuzluk belirlenmesi.
- Pazarlama: Çapraz satış analizleri, müşteri dilimleri.
- Sigortacılık:Müşteri kaybı sebeplerinin belirlenmesi, yolsuzlukların engellenmesi.
- Telekomünikasyon: Yolsuzluk tespiti, hatların yoğunluk tahminlemesi.
- Tıp: Tıbbi teşhis, uygun tedavi sürecinin belirlenmesi.
- Bilim ve Mühendislik: Deneye dayalı verilerde modeller kurularak bilimsel ve teknik sorunların analizi.
- Endüstri: Kalite kontrol, stok kontrolleri ve stok tahminlemeleri

Veri madenciliğini bir süreç olarak tanımlayacak olursak bu sürecin ana ögesi veri madenciliğinin uygulamacısıdır. İstenilen doğru sonuca ulaşmak için sürecin tüm adımları doğru olarak uygulanmalıdır. (41).

2.2.Verinin Anlaşılması

Veri Madenciliği Projelerinde; sahip olunan verinin, yapılacak olan çalışma için güvenilir ve yeterli içeriğe sahip olup-olmadığının, net bir şekilde tespit edilmesi gerekmektedir.

Veri Madenciliği Projelerinde "Verinin Anlaşılması" 5 evreden oluşur. Bunlar;

- Değişkenlerin Anlaşılması
- Veri Kalitesinin Belirlenmesi
- Görselleştirmenin Desteği
- Korelasyon Analizleri
- Olmayan Kayıt Analizleridir(42).

2.3.Verinin Hazırlanması

Modelin kurgulanması adımımda karşılaşılabacak sorunlar, bu adıma tekrar tekrar geri dönülmesine ve verilerin yeniden düzenlenmesine sebebiyet verecektir. Yaşanacak bu durum verilerin hazırlanması ve kullanılacak modelin kurgulanması aşamaları için veri keşfi sürecinde kullanılacak olan eforun % 50'si üzerinde harcanmasına neden olmaktadır.

Verilerin hazırlanması adımı veri toplama, veri değerlendirme, birleştirme ve temizleme, veri seçme ve veri dönüştürme basamaklarından meydana gelmektedir.

2.3.1. Veri Toplama

Tanımlı yapılan sorun için lüzum duyulan verilerin ve bu verilerin elde edileceği veri kaynaklarının belirlenmesi adımdır. Toplanılacak veri kaynaklarının seçimi önemli bir karar aşamasıdır. Elimizde bulunan verinin azlığı çalışmanın eksik kalmasına sebep olabileceği gibi elimizdeki veri kaynağının çokluğu kirli,gürültülü veriye sebep olabilir ve sürecin uzamasını sağlar.

Veri toplama aşamasında işletmenin sahip olduğu veri kaynakları haricinde, bir çok veri kaynağına ulaşım sağlanabilir. Nüfus sayımı, hava durumu, merkez bankası geçmiş ve gelecek döviz kuru tahminleri gibi türlü veritabanlarından ya da veri satışı yapan kurumların veri tabanlarından yararlanılmaktadır.

2.3.2. Veri Değerleme

Veri madenciliğinde amaca hizmet edecek verilerin farklı referanslardan elde edilmesi, veri bağdaşmazlıklara sebep olacaktır. Bu bağdaşmazlıkların başlıca sebepleri çeşitli zaman dilimlerine ait olmaları, güncelleme hataları, veri biçimlerinin çeşitli olması, farklı ölçü birimleri ve varsayım farklılıklarıdır. Ayrıca verilerin hangi koşullarda edildiği de önemlidir. Hatalı veri kaynak kullanımı veri madenciliğinin tüm sürecinin güvenilirliğini etkiler.

Bu sebeple, iyi çıktı sonucu veren veri madenciliği çalışmaları temiz ve düzenli verilerden sağlanacağı için, eldeki verilerin uyumlulukları bu adımda detaylıca incelenerek değerlendirilmelidir.

2.3.3. Verileri Bir Araya Getirme ve Temizlemesi

Bu adımda farklı referanslardan elde edilen verilerin içeriğinde ki ve veri değerlendirme adımında tayin edilen problem ve bağdaşmazlıklar ile ilerlenilebildiği kadar ilerlenilerek, bu verileri bir veri tabanında toplamak gerekir. Fakat bu aşamada basit metodlarla ve özensiz şekilde yapılacak sorun çözme eylemlerinin, sonraki adımlarda daha büyük sorunların kaynağı olabileceği göz önünde bulundurulmalıdır.

Veri temizleme işleminde asıl amaç, eldeki veriler içinde veri ile uyumsuz olan veya hatalı girilmiş verilerin elenmesi işlemidir. Eğer eksik alanlar var ise benzerlik ve yakınlığı diğer alanlar ile tespit edilerek doldurulması uygundur. Eksik veri çok ise bu kaydın silinmesi daha doğru olacaktır(43).

2.3.4. Veri Seçimi

Veri seçiminde kullanılacak olan modele göre veri tercih edilir. Örneğin; tahminleme metodları için, bağımlı-bağımsız parametrelerin ve modelde kullanılacak olan veri kümesinin seçilmesi anlamına gelir.

Sıra numarasından farklı olan ID parametrelerin modele girilmemesi daha anlamlı sonuçların çıkmasını sağlar. Bu tür parametreler, diğer parametrelerin modeldeki etkisinin düşmesine ve veriye ulaşım zamanının artmasına sebebiyet verir.

Verilerin grafiksel olarak gösterimine imkan sağlayan araçlar ve bu araçların sunmuş olduğu bağlantılar, bağımsız değişkenlerin seçiminde fayda sağlayabilir. Genel olarak kalitesiz veri girişlerinin ya da bir kere gerçekleşmiş bir olayın veri kümesinden çıkarılması doğru tercih olacaktır.

Veri madenciliği araştırmasında uygulanan modelde kullanılan veri tabanı eğer çok büyük ise, rastgele dağılımı değiştirmeyecek şekilde örneklem ile çalışılmalıdır. Tüm popülasyonu temsil edecek bir örneklem seçilmelidir. Kullanılan işletim sistemleri ve yazılım programları her geçen gün gelişmesine rağmen, büyük veritabanlarında bir çok modelin uygulaması zaman kısıtına takılmaktadır. Ve bu da bir çok yöntemi uygulamaktan geri bırakmaktadır. Rastgele alınan örneklem üzerinde denenen bir kaç model ile büyük bir veri tabanında bir kaç model denemek arasında büyük farklar oluşmaktadır. Yöntemler arasından en güvenilir ve

performansı yüksek olan model seçilmelidir. Yani modellerin performansları uygun bir karar yöntemi ile kontrol edilerek karar verilmelidir.(44).

2.3.5. Veri Dönüştürme

Veri dönüştürmenin amacı, ana veriyi farklı biçimlere ya da değerlere evirmektir(43).

Örneğin; Veritabanında bulunan bir mantık alanı tamsayı bir türe çevrilebilir. Nedeni kullanılmakta olan bir takım veri madenciliği algoritmalarının tamsayı bilgi türüyle matıksal veri türüne göre daha başarılı sonuçlar üretmesidir(41).

2.4.Modelleme

Veri madenciliğinin en bilinen elamanı modellemedir. Veri girdilerinin değişken vasıflara göre bir sınıflayıcı (model) tarafından sınıflara tayin edilme işlemidir. Var olan değişkenlerin bir gruba yerleştirilip yerleştirilmeyeceğinin veya gruplardan hangisine yerleştirileceğinin kararının verilmesidir. Bir başka söylemle değişkenler ve olaylar için uygun sınıf tahmininde bulunulmasıdır. Modellemeye sahip olan girdileri, her biri bir model etiketi ile tanımlanacak gözlem ya da örneklerden oluşan bir eğitim kümesidir. Modelin her bir gözlemin niteliklerine bağlı olarak atadığı model etiketi ise çıktı olarak adlandırılır.

Aynı zamanda modelleme için makine özelliklerinden yararlanılmasını öğreten vasıflardan biridir diyebiliriz.

Bütünü anlamının temel analizi eğitim örneklerinden meydana gelen bir eğitim kümesi ile analiz örneklerinden meydana gelen bir analiz kümesini kapsar. Modelleme iki kademede gerçekleşir. Bu kademeler, verilerin eğitimi ve modelin analizidir. Verilerin eğitimi, eğitim kümesinden alınarak modelin oluşturulması, model analizi ise modelin doğruluk derecesinin belirlenmesidir. Model doğruluğunun kontrolü için, analiz örneklerinin bilinen sınıfı, model içerisinde tahmin edilen sınıf ile kıyaslanır. Analiz örneklerinin model ile uyum sağlaması surumunda doğruluk derecesini elde etmiş oluruz.

Girdilerden bu çıktıları üreten model, sınıf anahtarı bilinmeyen veya kayba uğramış olan yeni gözlem ve ya örneklerin sınıf anahtarlarıyla ilgili tahminde bulunmak için kullanılır.

Temel sınıflama teknikleri olarak; diskriminant analizi, lojistik regresyon, Bayes sınıflayıcıları, karar ağaçları, diğer mantıksal formüller yapay sinir ağları, eğer-sonra kuralları (kural çıkarımı), kaba kümeler, bulanık kümeler sayılabilir.

Model ile ilgili kaynakçalara dahil olan farklı bir yaklaşımda söz konusudur. Bu yaklaşımdan sınıflama tam ve kısmi sınıflama olarak ikiye ayrılmaktadır. Tam sınıflama veri dahilindeki tüm sınıflar ve örnekleri kapsayan modeller ile ilişkilidir. Örnek olarak C&RT, C4.5, yapay sinir ağları, CHAID, C5.0 ve diğer karar ağaçları gibi akıllı teknikler ile diskriminant analizi gibi istatistiksel araçlar verilebilir.

Kısmi sınıflamada ise veri gruplarının nitelikleri söz konusudur. Fakat kısmi sınıflama, tam sınıflama tekniğinden farklı olarak tüm sınıfları ve örnekleri kapsamaz. Bu modele ait olarak birliktelik kuralları verilebilir(45).

2.5.Modelin Kurulması

Berltilen sorunu uyumlu bir hale getirebilmek için arayışı sürdürülen modelin elde edilmesi ancak bir çok sayıda modelin kurulumunun yapıp denenmesiyle mümkündür. Bu sebeple model kurulumu ve veri hazırlama evresi sürekli tekrar isteyen bir aşamadır.

Model kurulumu aşaması denetimli ve denetimsiz öğrenmeden yararlanıldığı modellere göre farklılık göstermektedir.

Bir diğer adı örnekten öğrenme olan denetimli öğrenmede, bir denetçi ile ilişkili gruplar öncesinde, öngörülen bir ölçüde bölünerek, her grup için çeşitli örnekler verilir. Sistemden beklenen hedef, verilen örneklerle dayalı her bir sınıfa sahip olan niteliklerin bulunması ve bu niteliklerin kural cümleleriyle açıklanmasını hedeflemektedir.

Öğrenim aşaması sonlandığında, tanımlanan kural cümleleri var olan başka örnekler üzerinde uygulanır ve örneklerin ait olduğu sınıflar kurulan model tarafından belirlenir.

Denetimsiz öğrenmede de ilişkili örneklerden yararlanılarak bu örneklerin gözlemi ve tabi ki neticesinde ki benzerliklerden yola çıkarak sınıfların tanımlanması hedeflenmektedir. Bu yöntem kümeleme analiziyle benzerlik göstermektedir.

Denetimli öğrenimde seçilen algoritmaya uygun veriler ilk olarak verinin bir kısmı modelin öğrenilmesi, diğer kısmı ise model geçerlilik testi amacıyla iki eşit parçaya bölünür. Öğrenim kümesi kullanılarak modelin öğrenim işlemi kazanıldıktan sonra, test kümesiyle modelin doğruluk seviyesi belirlenir.

2.6.Modelin Kullanılması

Yüklenen ve doğruluk derecesinden emin olunan model bir uygulama olabileceği gibi, bir başka uygulamanın içine gömülü bir yapı olarak da kullanılabilir. Yüklenen modeller risk analizi, kredi değerlendirme, dolandırıcılık tespiti gibi işletme uygulamalarında direk olarak kullanılabilirdiği gibi, promosyon planlaması simülasyonuna gömülü bir alt yapı da olabilir veya tahmin edilen nesne düzeyleri bir daha sipariş noktasının altına düştüğünde, otomatik sipariş verilmesini sağlayacak bir uygulamanın içine entegre edilebilir.

2.7. Modelin Değerlendirilmesi

Doğruluk ölçüsünün bir model üzerinde değerlendirilmesinin bilinen en kolay metodu basit geçerlilik testidir. Bu testte yaygın olarak verilerin %5 ile %33'ü aralığındaki kısım test verileri olarak ayrılır. Mevcut veriler üzerinden modelin güncelliği tamamlandıktan sonra test işlemi yapılır. Basit geçerlilik testinde, yanlış olarak belirlenen durumların, var olan tüm durumların sayısına bölünmesiyle hata oranı bulunurken,doğru olarak belirlenen durumların var olan tüm durumların sayısına bölünmesiyle doğruluk oranı bulunur. (Doğruluk Oranı = 1 - Hata Oranı)

Diğer bir yöntem çapraz geçerlilik testidir. Bu test az miktarda veriye sahip olunması halinde uygulanır. Çapraz geçerlilik testinde veri kümesini eşit iki parçaya belirli bir kriter gözetmeksizin rastgele ayırmak gerekir. Sonrasında iki aşama mevcuttur. İlk aşamada ikiye ayırdığımız veri kümesinin bir parçası üzerinde model testi yapılırken diğer parça üzerinde test işlemi yapılmalıdır.

İkinci aşamada ise birinci aşamada üzerinde test işlemi yapılan parçaya model testi uygulayıp, diğer parça üzerinde test işlemi uygulanarak elde edilen hata oranlarının ortalaması kullanılır.

Bir diğer test n katlı çapraz geçerlilik testidir. Bu yöntem bir kaç bin ya da daha fazla satırdan meydana gelen veritabanlarında verileri n gruba ayırarak uygulanan bir

bir yöntem çeşididir. Örneğin; verilerin 20 gruba sınıflandırıldığı bir yöntemde ilk olarak birinci grup anaizi yapılırken diğer gruplar eğitim için kullanılır. Bu şekilde tekrarlanarak devam eden bu yöntemde bir grup test amacıyla kullanılırken bir diğer grup eğitim için kullanılmalıdır. Neticede örnekte belirtilen 20 gruba ait 20 hata oranının ortalaması, kurulan modelin yaklaşık hata oranı olacaktır.

Küçük veri kümelerinde modelin hata oranını yaklaşık olarak bulunmasında kullanılan bir diğer yöntem “Bootstrapping” yöntemidir. Çapraz geçerlilik testine benzer bir şekilde model bütün veri grubu üzerine kurulur. Sonrasında minimum 200 olması şartıyla bazen binden fazla olmak üzere daha fazla öğrenim kümesi tekrarlanan örneklemelerle veri kümesi oluşturularak hata oranı hesaplar.

Model kuruluş yöntemlerinden anlaşılacağı üzere, aynı yöntemle başka araçların kullanıldığı veya başka algoritma ve araçların denendiği farklı tarz modeller kurulabilir. Model kuruluşuyla ilgili bir karara veya denemeye başlamadan önce, en uygun yöntemin hangisi olduğuna karar verebilmek güçtür. Bu sebepten ötürüdür ki başta da belirttiğimiz gibi farklı türde modeller kurarak hata oranı en düşük, doğruluk derecelerine en uygun modeli bulmak için sayısız tekrar yapılmalıdır.

İlk olarak sınıflama problemleri derince akla gelen modellerin doğruluk derecelerinin testinde kolay fakat faydalı bir araç olan risk matrisinden yararlanılmalıdır.

Tabloda bir örneği söz konusu olan bir matriste, sütunlarda fiili, satırlarda ise tahmini sınıflama değerleri yer almaktadır. Örneğin; Y sınıfının fiili sürunda bulunan 92 elemanın, kabul edilmiş modele bağlı 4’ünün X, 76’sının Y, 12’sinin Z olarak ayrılması matriste gözlemlenebilir.

Tablo 8 Risk Matrisi

Tahmini	Fiili		
	X Sınıfı	Y Sınıfı	Z Sınıfı
X Sınıfı	90	4	6
Y Sınıfı	20	76	24
Z Sınıfı	8	12	80

Model değerlendirme yapılırken, modelin anlaşılabilir olması da diğer bir ölçütdür. Araştırmalarda doğruluk derecelerinde az düzeyde ki artışın manası önemli olsa da, bir çok işletme veri madenciliği yönteminden elde edilen kararın neticesinin yorumlanabilmesi daha önemli bir unsurdur.

Sık sık karşılaşılmayan yorumlanması zor olsa da, karar ağacı ve kural esaslı sistemler model tahminine bağlı nedenleri en iyi şekilde ortaya çıkarmaktadır.

Kurulumu gerçekleşmiş model değerinin kısıtlandırılmasında kullanılan diğer bir değer, model tarafından tavsiye edilen uygulamadan getirisi olacak olan kazancın, bu uygulamanın sağlanması için katlanılacak olan maliyete bölünmesi ile elde edilecek olan yatırımın geri dönüşü oranıdır.

Kurulmuş olan modelin doğruluk derecesi her ne kadar yüksek olursa olsun, modelin gerçeği gösterdiğinden bahsetmek söz konusu değildir. Yapılan analiz ve ya test sonucuna bağlı olarak doğruluk yüksek ölçüde kabul edilmiş olan bir modelin gerçeği yansıtmamasının başlıca nedenleri, model kurulumunda kabul edilen varsayımlar ve modelde kullanılan verilerin tutarlı olmamasıdır.

2.8. Modelin Uygulanması

Bütün veri madenciliği modelleri birbirine benzer şekilde bir yaşam döngüsü kapsamındadır. Kimi yöntemlerde özellikler durağan yapıdadır, bu nedenle modelin güncel vaziyette bulunmasına gerek yoktur. Ancak bir çok uygulama özellikleri sıklıkla değişiklik gösterir, yani veriler kümelere dahil olur, işte bu durumlarda modelin güncel yapıda olması gerekmektedir. Yani model sıklıkla güncellenmelidir(41).

2.9. Veri Madenciliğinde Karşılaşılan Problemler

Sistemler küçük veritabanlarında çalıştığı performansla büyük veritabanlarında çalışamayabilir, davranışlar farklılık gösterebilir. Kurulan bir sistem sorunsuz çalışırken, bu veriye gürültülü veri ilave edildiğinde farkedilir bir şekilde kötüleşebilir(46).

2.9.1. Veri Tabanı Boyutu

Veri boyutlarının her geçen gün hızla arttığı kaçınılmaz bir gerçektir. Makina öğrenimi algoritmaları küçük örneklemeleri değerlendirebilecek biçimde geliştirilmiştir. Aynı algoritmaların çok daha büyük örneklemelerde kullanırken maksimum dikkat gösterilmelidir. Örneklem verisinin büyük olması, örüntülerin varlığını göstermesi açısından avantaj sağlar fakat böyle bir örneklemde elde edilebilecek muhtemel örüntü miktarı fazladır. Bu sebeple veri madenciliğinde muhtemel sorunların başında veritabanlarının yüksek miktarda veri içermesi gelir. Çözüm olarak veri madenciliği metodları örneklemi yatay ve dikey olarak yalınlaştırılmalı veya sezgisel bir yaklaşımla taranmalıdır.

Yalınlaştırma yatay olması durumunda bir çok şekilde gerçekleştirilebilir. İlk adımda, bir özelliğin alan değerleri hiyerarşik şekilde sınıflandırılır. Bu işlemin adı genelleştirme işlemidir. Bir sonra ki adım, ise hiyerarşik şekilde sınıflandırılan alanlar aşağıdan yukarıya doğru seviyelendirilir ve mükerrer kayıtlar teke indirilir. (Han, Cai ve Cercone, 1992). İkinci adımda, güvenilirliği olan örneklem kuramından yararlanılır. Büyük miktarda ve derinliği olan veri yalınlaştırılmış haliyle belirli bir güven aralığını temsil edebilir ve yalınlaştırılan veri kümesinin derinlemesine bir makine öğrenmesinde kullanılır hale gelir. (Rastogi ve Shim, 1999). Son adımda ise sürekli verilerden meydana gelen alana sahip özellik için kesikleştirme tekniği kullanılarak kategorik değerler oluşturulur. (Fayyad ve Irani, 1993). Yatay yalınlaştırma sürekli değerlerin belirli adım değerlerine çevrilmesi ile mükerrerlik gösteren kayıtlar teke indirgenerek sağlanabilir. Kesikleştirme metodu esasen verilen sürekli değişken için, değişkenin damıtılması şeklinde açıklanabilir. Kalan niteliklerin yalınlaştırılması işlemine ise Dikey yalınlaştırma denilir.(47).

2.9.2. Gürültülü Veri

Çalışma alanı geniş yer tutmuş veri tabanlarında nitelik alanları yanlış değerler ile doldurulmuş olabilir. Bu hataların sebebi veri girişinde kullanıcı hataları veya girilen değerlerin doğru ölçülmemesinden kaynaklanır. Gürültülü veri, verinin girişi sırasında ya da verinin toplanma aşamasında meydana gelen sistem dışı hatalara denir. Geliştirilen sistemleri sayesinde ticari ilişkisel veritabanları veri girişi yapılırken yapılan hataların kendiliğinden düzeltilmesinde ufakta olsa destek sağlamaktadır. Kusurlu veriler veritabanlarında işletmeler için büyük sorunlara yol açabilir. Bu tarz durumlar, veri madenciliği metodunun bahsi geçen veri kümesindeki gürültülü verilere karşı duyarlılığının az olmasını gerektirmektedir(48).

Tümevarım karar ağaçlarında yürütülen metodlar dizisinde gürültülü verinin sebep olduğu sorunlar araştırılmıştır (Quinlan, 1986b). Çalışılan veri kümesinin gürültülü olması durumunda sistem bozuk veriyi tespit etmeli ve tespit edilen bozuk veriyi görmemelidir. Quinlan, 1986b çalışmalarında gürültünün veri sınıflamasına etkisini gösteren deneyler yapmıştır. Deneyler neticesinde, sınıflandırılmalı öğrenmede etiket üzerindeki gürültü öğrenme algoritmasının etkileyerek performansının düşmesine yol açmıştır(49).

2.9.3. Eksik Veri

Eksik veriler ver tabanında ilgili alanların boş kalmasıyla oluşur. Bir veri kümesinde her alan yeterli şekilde doldurulmuş olsaydı sınıflama işlemi kolaylıkla alt kümelerden yararlanılarak yapılardı. Fakat gerçek hayatta bu şekilde ilerlemek kolay değildir. Eksik verilerin oluşmasında bir çok sebep sayılabilir. Örneğin; bir firmanın müşteri veri tabanında müşteriye ait bilgiler yer almayabilir. Verilerin toplandığı dönemde yaş aralığı bilgisi önemsiz bulunarak kayıt altına alınmamış olabilir. Müşterinin vermek istemediği bir bilgi dolayısıyla ilgili alan boş bırakılmış olabilir. Tüm bu sebeplerden dolayı eksik verilerin olduğu bir veri tabanı ortaya çıkabilir(50).

2.9.4. Artık Veri

Eldeki veri, çözümlmek istediğimiz probleme elverişli olmayan veya işe yaramayacak fazla özellik barındırabilir. Yapılan çalışmalarda bununla çok kez karşılaşabiliriz. Örneğin, çözümlmek istediğimiz sorun için gereken veriye sahip

olmak adına bağlantıları ortak özellikler aracılığıyla birleştirdiğimizde ortaya çıkan bağlantıda artık nitelikler ortaya çıkabilir. Geliştirilmiş algoritmalar ile Özellik seçimi olarak adlandırılan artık eliminasyon işlemi yapılır(51).

Özellik seçimi, eldeki veri kümesinin içerdiği kapsamı ve taşıdığı bilgiyi bozmadan daha az sayıda bir veri kümesiyle aynı bilgi ve kapsamı yansıtmalarını sağlamaktır.Özellik seçiminin veri kümesini küçültmek dışında fayda sağladığı bir konu da sınıflama işleminin performansının artmasıdır(52) .

2.9.5. Dinamik Veri

İşletmelerde kullanılan veritabanları çoğunlukla dinamiktir. Dinamik bir veri tabanı demek sürekli değişen, hareketli bir veritabanıdır. Dinamik veritabanının zorlukları arasında bilgi keşfinde karşımıza çıkmaktadır.

Domingos,Hulten ve Spencer tarafından yapılan çalışmada mevcut veritabanının çalışma anında, bilgi keşfini sağlayan yöntem bir veritabanı uygulaması şeklinde çalıştığında performansın düştüğü tespit edilmiştir. Başka bir zorluk ise, veritabanındaki verilerin kalıcı olduğu varsayıp, alınan örnek veri seti üzerinden yapılan bilgi keşfi çalışmasında dinamik olması dolayısıyla değişen verinin analiz sonucu ortaya çıkan örüntüye yansımamaktadır. Yansıma işleminin gerçekleşmesi için bilgi keşfi yönteminin bulduğu örüntüler bir süre sonra dinamik veriye göre ilgili örüntülerin yığılacağı bir günleme tekniğine sahip olmalı. (Hulten, Spencer ve Domingos, 2001).Bilgi keşif yöntemleriyle birlikte kullanılabilecek aktif veritabanlarında tetikleme düzenekleri mevcuttur. (Paton ve Diaz, 1999)(53).

3. PERAKENDE SEKTÖRÜNDE PAZARLAMA STRATEJİLERİ

3.1.Pazarlama Kavramı

Sözlükteki karşılığı, “ürünlerin, hizmetlerin satışını arttırmak hedefiyle tanıtmayı, paketlemeyi, piyasa ihtiyaçlarını belirlemek ve karşılamak adına yapılan faaliyetlerin tamamı” olarak yer alan pazarlamayı kısaca özetlemek gerekirse, bir işletmenin ürettiği bir ürün veya hizmetin maddi kâra dönüştürebilmesi için yürüttüğü sürece **pazarlama** denir(54).

İşletmeler ürünlerini hedef kitleye doğru bir şekilde ulaştırmak veya satmak için pazarlama faaliyetlerinden yararlanmaktadır. Pazarlama faaliyetlerinin neler olabileceğini incelersek, müşterilere sunulacak olan ürünün ne olabileceği, ürünün üretilmesi ve pazarlanabilir duruma gelmesi, müşterilerin bu üründen nasıl haberdar olacağı ve ürünün müşterilere nasıl ulaştırılacağına kadar pek çok kararın alınması gibi faaliyetleri kapsadığını görebiliriz.

Bir işletmenin pazarlama birimi, diğer bütün birimlerden daha fazla müşteriler ile temas halindedir. Müşteriler ile temas ve etkileşim içerisinde olan pazarlamanın en önemli amacının;

- İhtiyaç ve isteklerini karşılayacağına müşteriye ikna ederek yeni müşteri kazanımı,
- Mevcut müşteri memnuniyetini sağlayarak, rakip geçişlere engel olmak ve işletme adına değerli müşteri portföyü oluşturmak,

olduğunu söyleyebiliriz.

3.2.Perakende Kavramı

Ürünlerin tek tek ya da parça durumunda az az satışının yapılmasına dayanan satış şekline Perakende denir(54).

Perakendecilik, üreticilerin ürettiği ürünleri veya hizmetlerin bir firma aracılığıyla son tüketiciye satılmasıdır. Üreticilerin ürettiği ürün veya hizmetlere son tüketicinin ulaşabileceği birçok perakende kuruluşları (aracı firma) mevcuttur. Perakende kuruluşlarını faaliyet alanlarına göre mağazasız perakendecilik ve mağazalı perakendecilik olarak sınıflandırabiliriz.

Mağazasız perakendeciliğin içerisinde kapıdan kapıya satış, telefon ile satış, televizyon ile pazarlama, doğrudan posta ve elektronik alışveriş kuruluşlarını dâhil edebiliriz.

Mağazasız perakende içerisinde yer alan elektronik alışveriş son yıllarda popüler olan perakende kuruluşları arasında yer almaktadır. Teknolojinin gelişmesi ve tüketicilerin internet kullanım alışkanlıklarının artması ile elektronik alışveriş de artış göstermektedir. Tüketici istediği zaman, telefonundan veya bilgisayarından herhangi bir alışveriş sitesine girerek rahat bir şekilde alışveriş yapabilmektedir. Örnek verecek olursak **Amazon** adlı alışveriş sitesi dünyanın birçok ülkesine, birden fazla ürün yelpazesi ile kolay bir şekilde alışveriş yapma imkânı sağlıyor.

Mağazalı perakendeciliğin içerisinde ise bağımsız mağazalar, birleşik perakendeciler, imalatçının satış mağazası ve zincir mağazaları dâhil edebiliriz.

Perakendecilerin ana işlevleri özetle;

- Tüketicinin ihtiyacını olan doğru ürünleri, doğru yere, doğru zamanda ulaştırmak.
- Tüketicilerin taleplerindeki özellikleri ve değişimleri, doğru zamanda analiz ederek bilgileri üreticiye aktarmaktır.
- Malların talebe uygun halde, küçük miktarlarda satılması için, büyük hacimdeki ürünlerin küçük parçalara ayrılması ve daha kabul edilebilir hale gelmesini sağlamak.

3.2.1. Perakendeciliğin Fonksiyonları

Perakendeciliğin fonksiyonları, hem üretici hem de tüketici ve toptancılar açısından büyük önem taşımaktadır. Perakendeci aşağıda belirtilen hizmetlerden bir kısmını ya da tümünü müşterileri için sağlayabilir (55):

- Doğru yer.
- Belli lokasyonlara özgü ürün çeşidi tercihi.
- Toptan alınan malın bölünebilmesi ve düşük miktarlarda malın satılması.
- Satılan malın kabul görmesini sağlamak.

- Gelecek satışlar için malların stoklanması.
- Malların sistematik şekilde ulaşmasının sağlanması, (ambardan perakendeciye, müşteriye)
- Hem tüketiciye hem de tedarikçi firmalara bilgi sağlanması.
- Ürünün devamlılığının garantilenmesi.
- Satış sonrası hizmet.
- Tüketiciye şikayetyle ilgili dönüşler sağlamalı.
- Kredi ve düşük satın alma olanağı sağlanması.
- Sosyal görüşme olanağı sağlaması gibi.

3.2.2. Dünya Perakendeciliğinin Tarihsel Gelişimi

İnsanoğlu yüzyıllardır alışveriş geleneğini farklı şekillerde (pazaryeri, seyyar satıcı vb.) sürdürmektedir.

Bilinen tarihe göre Çin’de milattan öncesinde, Avrupada 16. ve 17.yüzyıllarda perakende zincirlerine rastlanmaktadır. 1670 yılında Hudson’s Bay Company mağazal zinciri yerini almıştır.

Zamanla modern mağaza zincirleri yerini almaya başlamıştır ve ilk örnekleri arasında New York’ da 1859 Great Atlantic & Pasific Tea Companyşirketi er almaktadır.

17. yüzyılda departmanlı mağazalar olarak adlandırılan mağaza zincirleri yerini almaya başlamıştır. departmanlı mağazalar ilk olarak Asya ve Avrupa’ da görülmüştür.

Zamana uyum sağlayarak 1800’lü yılların ortalarında Paris’in Bon Marché’si departmanlı mağaza sistemine geçiş yapmıştır.

20. Yüzyılların sonlarında müşterilerin her türlü ihtiyaçlarını farklı lokasyonlara gitmek zorunda kalmadan,tek seferde tüm ihtiyaçlarını toplu şekilde karşılayabilecekleri alışveriş merkezleri tasarlanmıştır.

Türkiye’de Perakende Sektörünün tarihi ;

- Osmanlı zamanında perakende adına bakkal ve esnaf gibi işletmelerin hüküm sürmesi ve Kapalıçarşıların bilinirliği

- Cumhuriyet döneminin başlangıcında geleneksel alışveriş sisteminin yanısıra semt pazarları,
- 1950'lerde Avrupa ülkelerde görünen perakende ticaret sistemine geçiş; İsviçre Migros' un gelişi, akabinde GİMA'nın kuruluşu
- 1970'lerde belediyelerin düzenlediği düzenli satış mağazaları
- Mağazaların bir kaç katlı binalarda oluşu ve şube açmaya başlaması
- 1980'lerde yurtdışından ithal edilen mallarda ki ve büyük alışveriş merkezlerindeki artış
- 1990'larda yabancı sermayenin girişiyle birlikte perakendecilerin sayısının artması
- Günümüzde büyük şehirlerde, büyük sermayeli perakende zincirlerinin yoğunluğu

şeklinde sıralanabilir.

3.3.Strateji Kavramı

Strateji kavramını, bir işletmenin bulunduğu sektörün şartlarına uyum sağlamanın yanında, istikrarını sürdürerek varlığını korumak ve bulunduğu sektörel pazarda başarı elde etmek,sermayesini güçlendirmek adına çizdiği yol olarak açıklayabiliriz(56).

3.4.Perakende Sektöründe Pazarlama Stratejileri

Perakende sektörü için temel stratejik ilke mevcuttur ;

- Hedeflenen pazarın belirlenmesi.
- Hedef belirlenen pazarın ihtiyaçlarını karşılayacak ve denetleyecek perakende biçiminin analiz edilmesi.
- Sürdürülebilir rekabeti sağlayacak üstünlüğün oluşturulması .

Hedef Pazarın Belirlenmesi; Bir pazarla görevlisi önce temas etmek istediği hedef müşterisini belirlemelidir. Bunun için öncelikli olarak pazar bölümlemesi ile başlamalıdır. Hedef müşterilerini demografik, sosyo-ekonomik, davranışsal gibi özelliklerine göre segmente etmeli. Burada dikkate alınması gereken önemli bir nokta ise ihtiyacın sürdürülebilir, ulaşılabilir ve karşılanabilir olmasıdır. Pazarlamacı bu noktaları da gözönüne alarak temas etmek istediği hedef müşterisini belirler.

Hedef belirlenen pazarın ihtiyaçlarını karşılayacak ve denetleyecek perakende biçiminin analiz edilmesi; Perakende sektöründe bulunanlar bir ve ya birden fazla amaçlanılan pazar sadakatini ve rekabet üstünlüğüne sahip olmak amacıyla çeşitli niteliklere sahip perakende şekilleri kullanır. Perakende şekilleri, kompetitif olmak adına; fiyatın, mağazaların, ürünlerin, hizmetlerin ve pazarda yer edinmenin bir çok farklı kombinasyonunu kulliyolunu tercih eder.

Sürdürülebilir rekabeti sağlayacak üstünlüğün oluşturulması; Bir perakende pazarında benzer ürünleri aynı grup müşterilere sunan işletmeciler birbirleri ile rekabet halinde olurlar. Bu rekabet sayesinde işletmeciler perakende pazarında ayakta kalabilmek için değer yaratan, sürdürülebilir stratejiler geliştirerek rekabet avantajı kazanmak için bir perde örür. Bu perde sayesinde rakiplerin müşteriler ile ilişki kurmasının önüne geçilir(57).

3.5.Perakendecilik Yönetim Kuralları

Gerekli ürünü, gerekli zamanda, gereken yerde, yeterli miktarda ve uygun fiyata müşteriyle buluşturmak, bunlar gerçekleşirken müşteri ile iyi iletişim kurmak ve sürekliliğini sağlamak önemlidir. Perakendecilik yönetiminde Müşteri İlişkileri Yönetiminde (CRM) kullanılan benzeri kurallar kullanılmaktadır:

Gerekli ürün için her şeyden müşteriye tanımak, talep ve ihtiyaçlarını doğru analiz edebilmek ve değişen nitelikleri tayin etmek önemlidir. Müşteri öncelikli olarak ihtiyacı olanı almaya odaklıdır, talepler ise ihtiyaç karşılandıktan sonra başlar. Dolayısıyla perakendeci işletmenin satacağı ürünü, derinlemesine yapacağı tahliller ardından karar vermesi gerekmektedir.

Gerekli zaman, ise müşteriye ihtiacı olan ya da olacak ürünün müşteriye istediği zamanda sunulması perakende yönetimi açısından önemlidir. Örneğin; yaz aylarında kar gözlüğü talep edecek olan müşteri sayısının “yok” denecek kadar az olması ürünün gereken zamna haricinde bulunmasının satış rakamlarını yükseltmeyecektir, promoyon br fiyat olmadığı sürece.

Gereken yer, yani doğru alanda bulunmak olarak nitelendireceğimiz yönetim kuralına göre ürünün yerleşmi hem mağazaiçi hem de bölgesel anlamda önemlidir. Bir ürün satışında bölgesel nitelikler çok önemlidir. Örneğin; Karadeniz bölgesinde mısır ve lahana tüketim oranı yüksekken, Ege bölgesinde zeytinyağlı tüketilme

oranının yüksek olması, perakendecinin bu bölgelere satacağı malı lokasyona uygun ihtiyaç ve talepler yönünde belirlemesi uygun olacaktır. Mağaza içi düzenlemede de ürünlerin müşteri tarafından ulaşımı kolay bir yer ve düzende olması önemlidir.

Yeterli miktar, ürünün ihtiyaca uygun miktarlarda mağazada depolanmış olmasıdır. Bir ürün bir bölgede fazla talep görürken farklı bir bölgede o kadar yoğun talep ile karşılanmayabilir. Bu noktada stok tahmini ve yönetimi büyük önem taşımaktadır.

Uygun fiyat, ise perakendeciden müşteriye gidecek olan mal ve ya hizmeti için, perakendecinin satış maliyetlerini ve karını çıkardıktan sonra müşteriye satılmasını sağlayan fiyattır. Gerçek dünyada her şey beklenen şekilde ilerleyemeyebilir. Malın stokta kalmaması için maliyet fiyatına hatta daha altında bir değere satılması da olasıdır. Müşteri ilişkileri, perakendecilikte müşteri ilişkileri ve müşteri memnuniyeti çok önemlidir. Perakendeci eğer müşteri ilişkilerini sıcak ve iyi tutmazsa ürün satışını yapacağı kitlede kayba uğrar. Perakendecinin dikkat etmesi gereken üç ana değişken grubunun her daim dikkate alınmalıdır ve alt başlıklarının yönetiminin ciddi bir konu olduğu unutulmamalıdır(58).

- Fiziksel dağıtım
- Mal ve hizmet
- İletişim

3.6.Perakende Sektörü İle Sunulan Hizmetler

Son yıllarda teknolojinin ilerlemesi ile perakende sektöründe, özellikle elektronik alışveriş kanalları üzerinde ciddi bir artış gözlenmektedir. Tüketicilerin internet kullanım alışkanlıkları artış gösterdikçe e-ticaret sitelerine daha çok yöneldiğini söyleyebiliriz. E-ticaret sitelerinden yapılan alışverişler dışında, tüketiciler mağazaya gitmeden önce e-ticaret siteleri sayesinde mağazadaki indirimlerden, kampanyalı ürünlerden haberdar olup site üzerinden alışveriş yapmadan, mağazaya giderek alışverişlerini gerçekleştirebiliyor.

Amazon e-ticaret sitesini örnek verecek olursak eğer, dünyanın bir çok ülkesine ürün kargolama özelliğine sahip. Amazon' un bu kadar popüler olmasının sebebinin geniş bir ürün yelpazesine sahip olması ve güvenilir bir alışveriş ortamı sağladığını söyleyebiliriz.

4. UYGULAMA

4.1.Araştırmanın Amacı ve Kapsamı

“Müşteri analizinde Veri Madenciliği Kullanımı ve Bir Uygulama” konulu araştırmanın amacı, bir çok ülke ile ihraacat ağı olan Perakende devlerinden bir “A” markasının müşterilerine yönelik pazarlama stratejisi belirlemek için, müşterileri ürün alım sıklığı, işletmeye yıllık olarak kazandırdığı ciro, ithal edilen ülke, mevsimsellik özelliklerine göre çok değişkenli bir istatistik yöntemi olan Kümeleme Analizi (Cluster Analysis) sınıflandırması ve sepet analizi yapmaktır.

Çalışmanın kapsamı; global olarak faaliyet gösteren bir özel sermayeli e-ticaret şirketinin 21786 adet müşterisinin 2010,2011 yılına ait e-ticaret sitesinden yapılan alışverişlerine göre kümeleme analizi gruplamak ve sepet analizi ile müşteri alışkanlıklarında anlamlı desenler oluşturarak müşteri profillemektir..

4.2. Araştırmanın Yöntemi

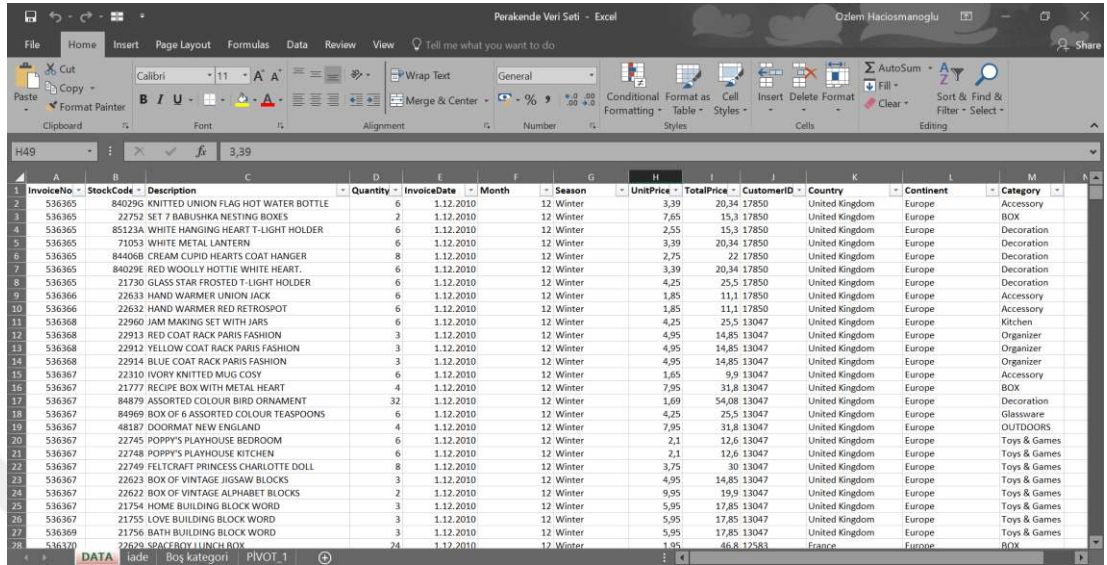
Bu çalışmada , veri madenciliği yazılımı olarak bilinen açık kaynak kodlu “Knime“ yazılımı kullanılmıştır. Araştırma için, birliktelik kuralları oluşturmak için Association rule learner(Borgelt) algoritması kullanılmıştır. Bu metod ile amaçlanan müşterilerin sepet analizi yapılarak önerilebilecek ürünlerin satışa dönüştürülmesidir. Hem işletmeye katkı sağlayacaktır hem de müşteriye farketmediği bir ihtiyacı farkettilerecektir. Bu metod ile çalışmanın amacı; düzenli olarak alışveriş yapan ve iade sürecine giren, mevcut müşterileri tanımak, ihtiyaçlarını analiz edebilmek ve analiz sonucunda müşterinin aldığı ürün ile birlikte sepetine atabileceği diğer ürün gruplarını kampanyalı bir şekilde farkettilerecektir. Ayrıca ihtiyaçlarına göre hem müşterinin memnuniyet seviyesinin artmasını hem de kendi kârlılığını yükseltecek yöntemler önerebilmeyi hedeflemektir.

4.3.Evren ve Örneklem

Toplamda 230775 satır işlem 4241 müşteri vardır. Veri seti müşteri numarası, fatura numarası, ürün stok numarası,ürün açıklaması, ürün adedi, tarih, birim fiyat ve ülkeden oluşmaktadır. Customer ID’si 14646 olan müşterinin alım yaptığı toplam ürün fiyatı 192.260,91 €’dur.

4.4. Verileri Anlama ve Değişkenler

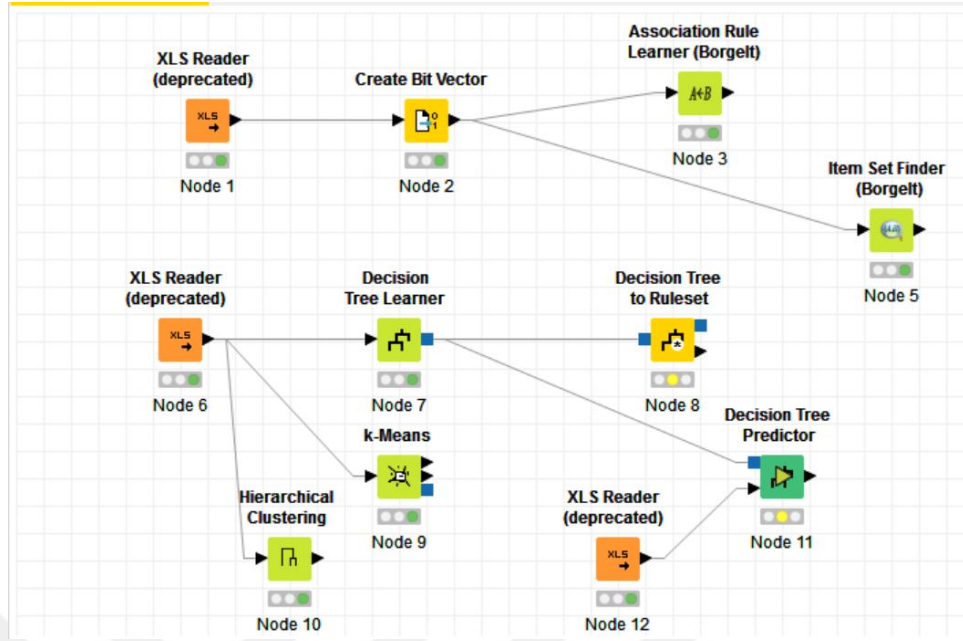
Ürün stok numaralarından veri setindeki ürün açıklamalarının tekilleştirilmesi sağlanmış ve kategorilendirilmiş,tarihten mevsimsellik üretilmiş,birim fiyattan toplam fiyata ulaşılmış,ülkeden kıtasal olarak ayrıştırılmıştır.



InvoiceNo	StockCode	Description	Quantity	InvoiceDate	Month	Season	UnitPrice	TotalPrice	CustomerID	Country	Continent	Category
536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	1.12.2010	12	Winter	3,39	20,34	17850	United Kingdom	Europe	Accessory
536365	22752	SET 7 BABUSHKA NESTING BOXES	2	1.12.2010	12	Winter	7,65	15,3	17850	United Kingdom	Europe	BOX
536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	1.12.2010	12	Winter	2,55	15,3	17850	United Kingdom	Europe	Decoration
536365	71053	WHITE METAL LANTERN	6	1.12.2010	12	Winter	3,39	20,34	17850	United Kingdom	Europe	Decoration
536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	1.12.2010	12	Winter	2,75	22	17850	United Kingdom	Europe	Decoration
536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	1.12.2010	12	Winter	3,39	20,34	17850	United Kingdom	Europe	Decoration
536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	1.12.2010	12	Winter	4,25	25,5	17850	United Kingdom	Europe	Decoration
536366	22633	HAND WARMER UNION JACK	6	1.12.2010	12	Winter	1,85	11,1	17850	United Kingdom	Europe	Accessory
536366	22632	HAND WARMER RED RETROSPOT	6	1.12.2010	12	Winter	1,85	11,1	17850	United Kingdom	Europe	Accessory
536368	22960	JAM MAKING SET WITH JARS	6	1.12.2010	12	Winter	4,25	25,5	13047	United Kingdom	Europe	Kitchen
536368	22913	RED COAT RACK PARIS FASHION	3	1.12.2010	12	Winter	4,95	14,85	13047	United Kingdom	Europe	Organizer
536368	22912	YELLOW COAT RACK PARIS FASHION	3	1.12.2010	12	Winter	4,95	14,85	13047	United Kingdom	Europe	Organizer
536368	22914	BLUE COAT RACK PARIS FASHION	3	1.12.2010	12	Winter	4,95	14,85	13047	United Kingdom	Europe	Organizer
536367	22310	IVORY KNITTED MUG COSY	6	1.12.2010	12	Winter	1,65	9,9	13047	United Kingdom	Europe	Accessory
536367	21777	RECIPE BOX WITH METAL HEART	4	1.12.2010	12	Winter	7,95	31,8	13047	United Kingdom	Europe	BOX
536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	1.12.2010	12	Winter	1,69	54,08	13047	United Kingdom	Europe	Decoration
536367	84969	BOX OF 6 ASSORTED COLOUR TEASPOONS	6	1.12.2010	12	Winter	4,25	25,5	13047	United Kingdom	Europe	Glassware
536367	48187	DOORMAT NEW ENGLAND	4	1.12.2010	12	Winter	7,95	31,8	13047	United Kingdom	Europe	OUTDOORS
536367	22745	POPPY'S PLAYHOUSE BEDROOM	6	1.12.2010	12	Winter	2,1	12,6	13047	United Kingdom	Europe	Toys & Games
536367	22748	POPPY'S PLAYHOUSE KITCHEN	6	1.12.2010	12	Winter	2,1	12,6	13047	United Kingdom	Europe	Toys & Games
536367	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8	1.12.2010	12	Winter	3,75	30	13047	United Kingdom	Europe	Toys & Games
536367	22623	BOX OF VINTAGE JIGSAW BLOCKS	3	1.12.2010	12	Winter	4,95	14,85	13047	United Kingdom	Europe	Toys & Games
536367	22622	BOX OF VINTAGE ALPHABET BLOCKS	2	1.12.2010	12	Winter	9,95	19,9	13047	United Kingdom	Europe	Toys & Games
536367	21754	HOME BUILDING BLOCK WORD	3	1.12.2010	12	Winter	5,95	17,85	13047	United Kingdom	Europe	Toys & Games
536367	21755	LOVE BUILDING BLOCK WORD	3	1.12.2010	12	Winter	5,95	17,85	13047	United Kingdom	Europe	Toys & Games
536369	21756	BATH BUILDING BLOCK WORD	3	1.12.2010	12	Winter	5,95	17,85	13047	United Kingdom	Europe	Toys & Games
536370	22629	SPARKLY LUNCH BOX	24	1.12.2010	12	Winter	1,95	46,8	12583	France	Europe	BOX

Şekil 9 Veri Seti Görüntüsü

4.5.Uygulama



Şekil 10 Knime Analiz Ekranı

Verinin birliktelik kuralları, karar ağacı sınıflandırması ve hiyerarşik sınıflandırma yapılması için Knime içerisinde bulunan algoritmalar yukarıdaki şekilde olduğu gibi uygulanmıştır.

Uygulamada kullanılan “Association Rule Learner (Borgelt)” algoritmasının ve Karar ağacı uygulamasının çalışma için verdiği analiz tabloları ve bu tablodan elde edilen sonuçlar aşağıda anlatılmıştır.

4.6. Bulgular

Toplamda 230775 satır işlem, 4241 müşterinin olduğu veri setinde 15 kategori bulunmaktadır.

Avrupa kıtasal bazda en çok alışverişin ve yüksek miktartlı alımların yapıldığı bölgedir. Aşağıdaki tabloda Kıtasal bazda yüzdesel dağılım görülmektedir.

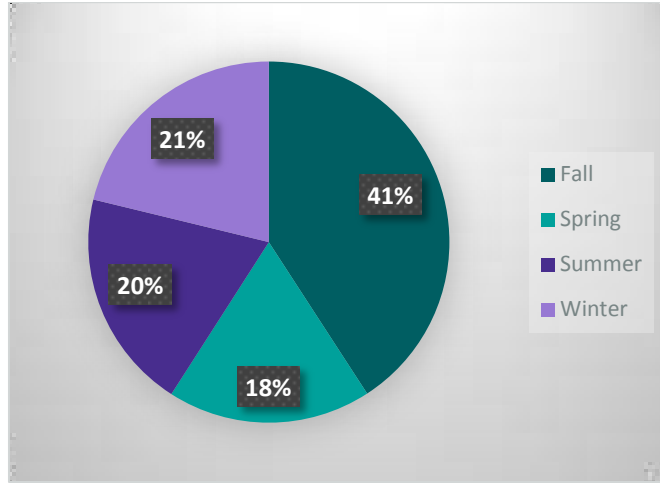
Continent	Frequency	Percent
Europe	229216	99.32
Australia	622	0.27
Asia	530	0.23
North America	204	0.09
Unknown	160	0.07
Africa	34	0.01
South America	8	0

Şekil 11 Kıtasal bazda alışveriş dağılımı

Category	Frequency	Percent
Decoration	82058	35.56
Kitchen	48530	21.03
Stationery	30748	13.32
Organizer	25334	10.98
Box	11411	4.94
Glassware	9077	3.93
Toys & Games	6972	3.02
Outdoors	6780	2.94
Accessory	4519	1.96
Party	2503	1.08
Book	1278	0.55
Gift	718	0.31
Personal Care	346	0.15
Clothing	312	0.14
Industrial	188	0.08

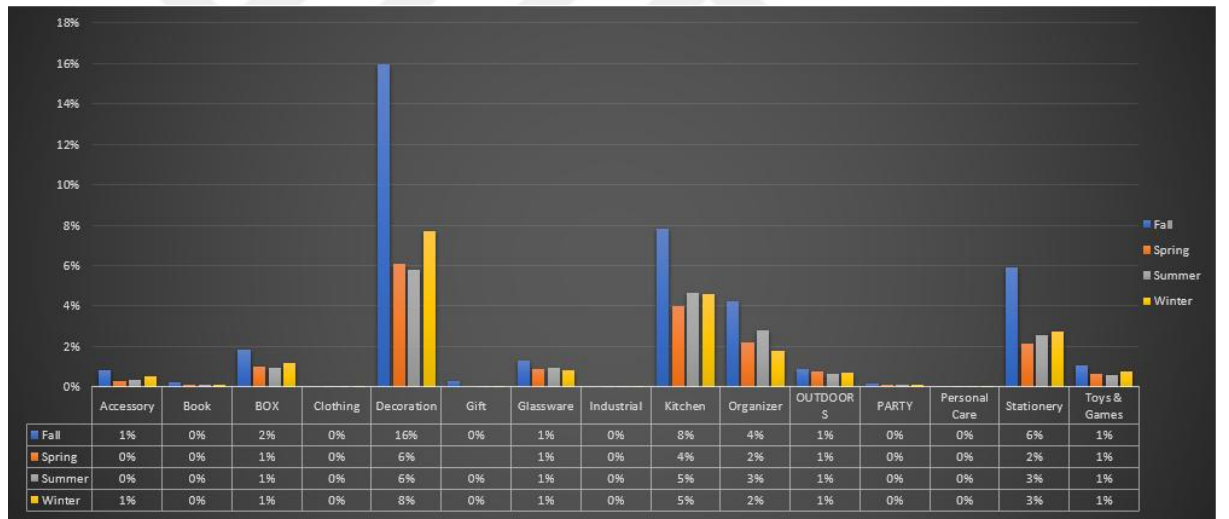
Şekil 12 Kategorisel Yüzdelik Dağılım

Kategorisel yüzdelik dağılımda dekorasyon ürünleri %35.56'lık dilimle öncü kategori olarak görülmektedir.



Şekil 13 Mevsimsel Dağılım

Grafikte mevsimler göre satılan ürünlerin yüzdesi görülmektedir. Sonbahar' da satışı gerçekleşen ürünlerin %41 olduğu gözlenmiştir. Yapılan satışların %18' i ise İlkbahar' da gerçekleşmiştir.



Şekil 14 Kategori Bazında Sezonsal Satış

Şekilde sezona göre satışı gerçekleşen ürünlerin yüzdeleri yer almaktadır. Özellikle dekorasyon ürünlerinin daha çok satıldığı gözlenmiştir. Dekorasyon ürünlerinin %16' sı sonbaharda satılmış olduğu görülmektedir. Ayrıca sonbaharda, mutfak ve kırtasiye ürünlerinin de satışlarının daha fazla olduğu gözlenmiştir. Aksesuar ve parti ürünlerinin 4 mevsimde de satışlarına baktığımızda ise oranlarının çok az olduğunu söyleyebiliriz. Kişisel bakım ürünlerinin ise neredeyse hiç satılmadığı gözlenmiştir.

SONUÇ

Kullanılan algoritma ve yöntemler program ekran görüntüleri ile birlikte verilmiş ve sonuçlar yorumlanmıştır.

Association Rules - 3:3 - Association Rule Learner (Borgelt)

File

Table "default" - Rows: 31585 Spec - Columns: 11 Properties Flow Variables

Row ID	S	Conseq...	Antecedent	ItemSe...	D Relativ...	D RuleCo...	D Absolut...	D Relativ...	D RuleLift	D RuleLift%	D Absolut...	D Relativ...
Row9171	22661.0	[23208.0,23209.0]		37	0.214	11.6	319	1.85	5.844	584.37	343	1.985
Row9172	22661.0	[20728.0,23208.0]		29	0.168	14.3	203	1.17	7.197	719.74	343	1.985
Row9173	22661.0	[20727.0,23208.0]		36	0.208	14.8	243	1.41	7.464	746.4	343	1.985
Row9174	22661.0	[22382.0,23208.0]		36	0.208	16	225	1.3	8.061	806.11	343	1.985
Row9175	22661.0	[22383.0,23208.0]		40	0.231	15.8	253	1.46	7.965	796.55	343	1.985
Row9176	22661.0	[20725.0,23208.0]		39	0.226	14.8	263	1.52	7.471	747.11	343	1.985
Row9177	23208.0	[22661.0]		58	0.336	16.9	343	1.98	4.895	489.47	597	3.455
Row9178	23307.0	[22355.0,22661.0]		18	0.104	10.1	178	1.03	3.467	346.73	504	2.917
Row9179	23307.0	[20724.0,22661.0]		20	0.116	10.6	189	1.09	3.628	362.83	504	2.917
Row9180	84992.0	[22355.0,22661.0]		18	0.104	10.1	178	1.03	3.488	348.81	501	2.899
Row9181	23201.0	[22355.0,22661.0]		27	0.156	15.2	178	1.03	3.486	348.57	752	4.352
Row9182	23207.0	[22355.0,22661.0]		35	0.203	19.7	178	1.03	5.252	525.19	647	3.744
Row9183	23199.0	[22355.0,22661.0]		32	0.185	18	178	1.03	4.327	432.69	718	4.155
Row9184	22630.0	[22355.0,22661.0]		26	0.15	14.6	178	1.03	4.098	409.77	616	3.565
Row9185	21975.0	[22355.0,22661.0]		28	0.162	15.7	178	1.03	5.139	513.87	529	3.061
Row9186	22662.0	[22355.0,22661.0]		65	0.376	36.5	178	1.03	10.553	1,055.3	598	3.46
Row9187	22629.0	[22355.0,22661.0]		25	0.145	14	178	1.03	3.448	344.76	704	4.074
Row9188	21977.0	[22355.0,22661.0]		19	0.11	10.7	178	1.03	2.757	275.73	669	3.871
Row9189	23202.0	[22355.0,22661.0]		35	0.203	19.7	178	1.03	4.402	440.15	772	4.467
Row9190	84991.0	[22355.0,22661.0]		26	0.15	14.6	178	1.03	3.842	384.2	657	3.802
Row9191	22386.0	[22355.0,22661.0]		37	0.214	20.8	178	1.03	4.124	412.41	871	5.04
Row9192	21931.0	[22355.0,22661.0]		51	0.295	28.7	178	1.03	6.397	639.7	774	4.479
Row9193	22411.0	[22355.0,22661.0]		29	0.168	16.3	178	1.03	3.564	356.39	790	4.572
Row9194	20724.0	[22355.0,22661.0]		117	0.677	65.7	178	1.03	15.371	1,537.1	739	4.276
Row9195	22355.0	[20724.0,22661.0]		117	0.677	61.9	189	1.09	18.287	1,828.7	585	3.385
Row9196	22661.0	[20724.0,22355.0]		117	0.677	39.3	298	1.72	19.781	1,978.1	343	1.985
Row9197	23206.0	[22355.0,22661.0]		48	0.278	27	178	1.03	5.381	538.11	866	5.011
Row9198	20726.0	[22355.0,22661.0]		50	0.289	28.1	178	1.03	5.92	591.98	820	4.745
Row9199	23203.0	[22355.0,22661.0]		44	0.255	24.7	178	1.03	3.955	395.53	1,080	6.25
Row9200	22384.0	[22355.0,22661.0]		55	0.318	30.9	178	1.03	5.735	573.54	931	5.387
Row9201	23209.0	[22355.0,22661.0]		39	0.226	21.9	178	1.03	3.764	376.37	1,006	5.821
Row9202	22197.0	[22355.0,22661.0]		40	0.231	22.5	178	1.03	3.752	375.2	1,035	5.989
Row9203	20728.0	[22355.0,22661.0]		56	0.324	31.5	178	1.03	5.628	562.81	966	5.59
Row9204	21212.0	[22355.0,22661.0]		34	0.197	19.1	178	1.03	3.208	320.78	1,029	5.955
Row9205	20727.0	[22355.0,22661.0]		52	0.301	29.2	178	1.03	4.799	479.88	1,052	6.088
Row9206	22382.0	[22355.0,22661.0]		59	0.341	33.1	178	1.03	5.798	579.75	988	5.717

Şekil 15 Association Rules Algoritması

Algoritma çalışmasında minimum rule confidence “10” olarak belirlenmiş olup minimum support “1” olarak belirtilmiştir.

Örneğin; 22577 stok kodlu ürün ile birlikte 22578 stok kodlu ürünü alındığında 19 kez 22340 stok kodlu ürünün alınım yapıldığı gözlemlenmektedir. Bu ürünün müşterilere önerilmesi durumunda birim fiyatı 0,39 € olan ürünün edilecek minimum kar 7,41 €’dur.

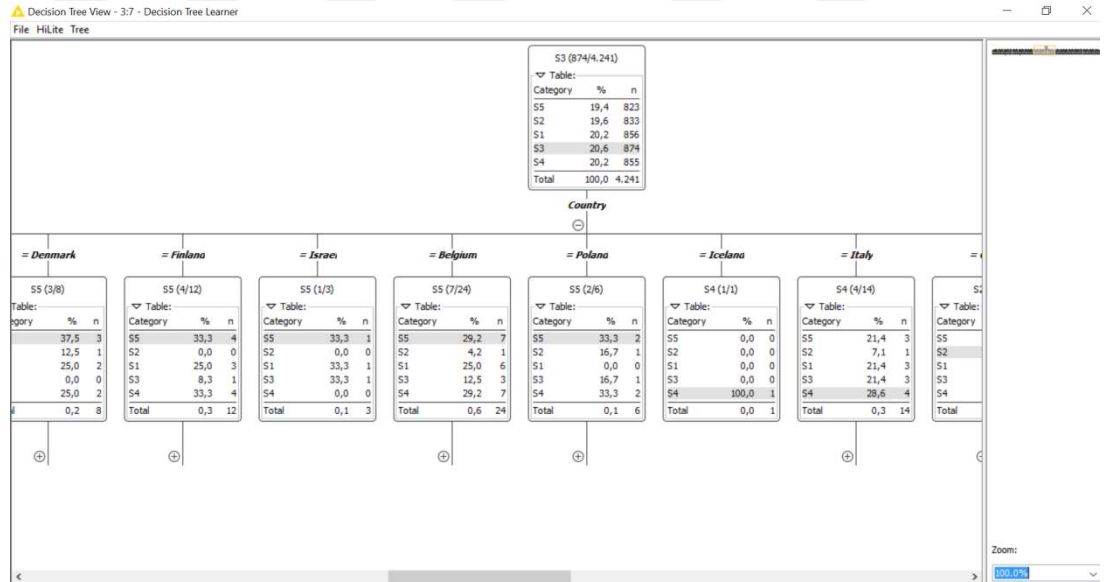
20725, 22383, 85099B stok kodlu ürünler alındığında 31 kez 85123A stok kodlu ürünün alınım yapıldığı gözlemlenmektedir. Bu ürünün müşterilere önerilmesi durumunda birim fiyatı 2,55 € olan ürünün edilecek minimum kar 79,05 €’dur.

Aşağıdaki tabloda bir kaç örneğini görebiliriz:

Tablo 9 Birlikte alınan Kuralları Analiz Çıktısı

Birlikte alınan Ürünler	Önerilebilecek Ürün	Rastlantı Sıklığı	Birim Fiyat	Ek Getirisi
[22384.0, 85099B]	20713.0	49	1,95	95,55
[22382.0, 85123A]	22960.0	20	4,25	85
[21166.0, 21181.0]	82582.0	69	2,1	144,9
[20725.0, 20726.0, 20727.0]	85099F	44	1,95	85,8
[20725.0, 22384.0, 85099B]	85099F	77	1,95	150,15
[20724.0, 22355.0]	21975.0	53	1,95	103,35
[22382.0, 22384.0]	22661.0	57	0,85	48,45

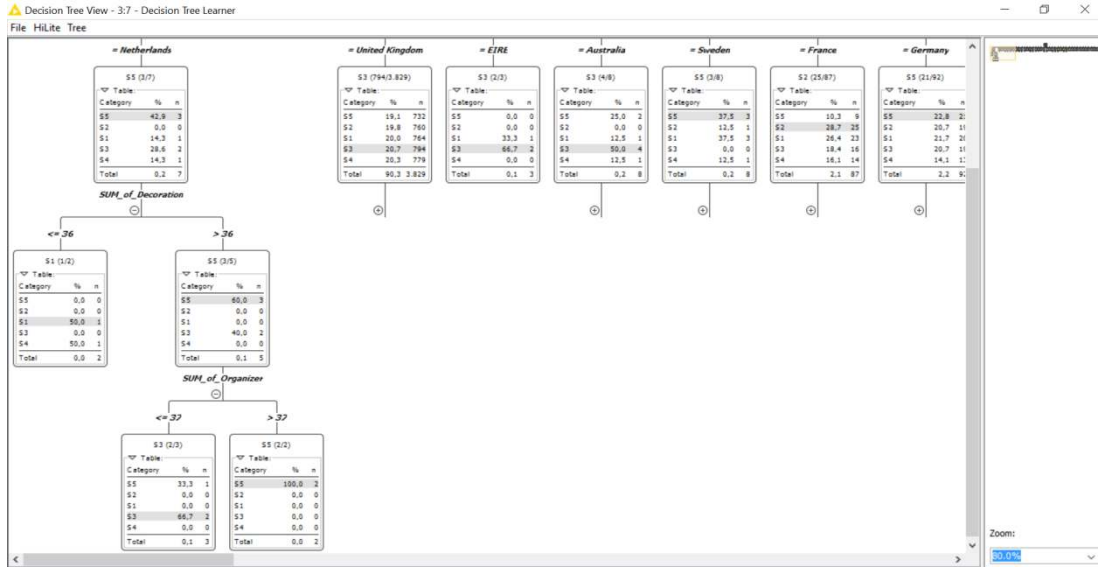
E-ticaret sitesinden alımların yapıldığı müşterilere ait bilgiler arasında ülke, müşteri numarası,kıtalar yer almaktadır. Müşterinin hangi sektörden olduğunu sınıflandırmak adına Karar ağacı kullanılarak sektörel olarak ayrılmıştır. Bu ayrım aşağıda şekil üzerinden tahmini örneklendirilmelerle anlatılmaktadır.



Şekil 16 Decision Tree Analiz Çıktısı

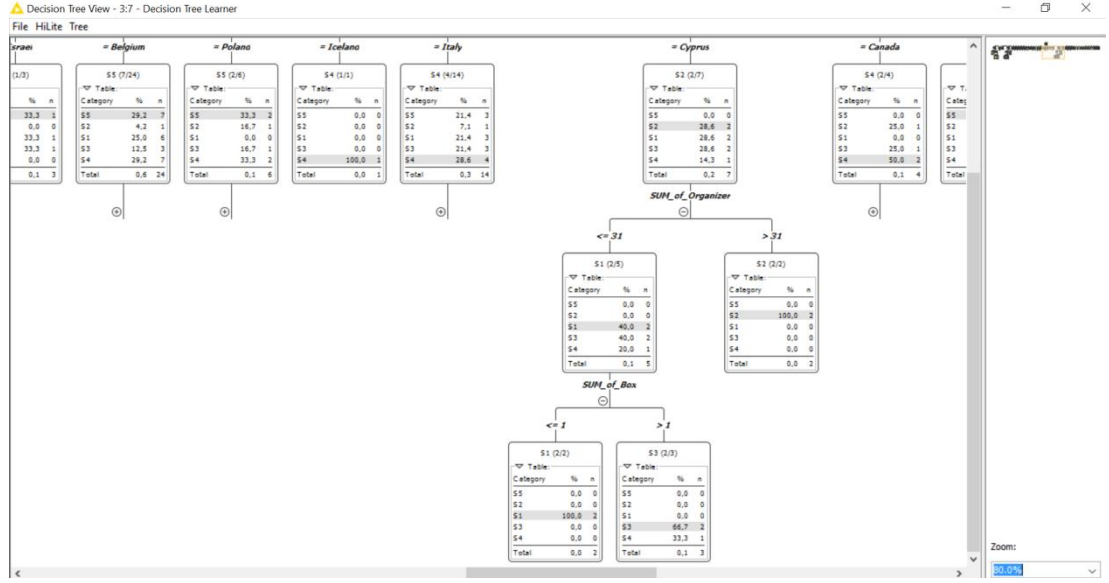
Karar ağacı bu işlemi yaparken Gini Index'ini çalıştırmaktadır.

Karar ağacından anlaşılacağı üzere müşterilerin büyük bir bölümü “S3” olarak adlandırdığımız Sektör 3’e aittir.



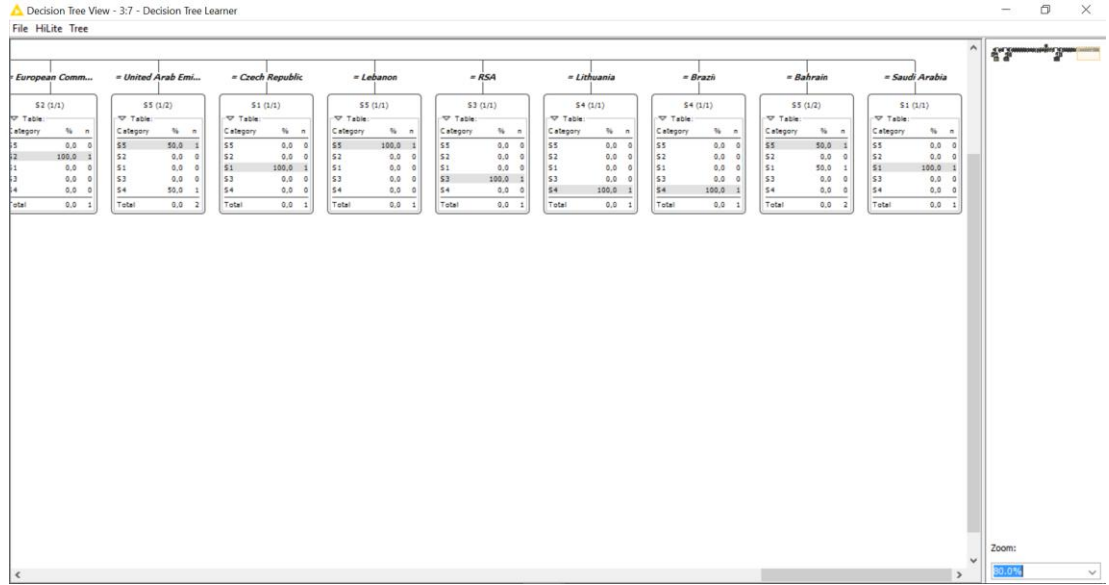
Şekil 17 Ülkesel Decision Tree Analiz Çıktısı

Gini Index'ine göre ülke bazında bir ayırım ile sektörel ayırım algoritması çalışmaktadır. Netherlands "S5" olarak görünen Sektör 5 sınıfında yer almaktadır. Alt kırımlarında ise eğer "S5" nasıl geliyor bunu görebiliyoruz. Dekorasyon ürününün 36 sınırının altında ya da üzerinde olduğunu inceleyip, üzerinde çıkması durumunda organize edici ürün grubunun özelliklerine bakılır. Organize edici ürün grubunda ise sınırı 37 alarak 37'den büyük olan sektör için "S5" 37'ye eşit ve küçük olan sektör için "S3" yani sektör 3 olabileceği sonucu çıkmıştır. Yüzdelik dilimde %100 ile analiz sonucu "S5" olmuştur ve Netherlands ülke bazında "S5" olarak sınıflandırılmaktadır.



Şekil 18 Ülkesel Decision Tree Analiz Çıktısı

Sınıflandırmada bir başka örnekte ise Cyprus ülkesinin “S2” yani Sektör 2 olarak karşımıza çıktığını görüyoruz.



Sudi Arabistan'dan yalnızca 24 adet dekorasyon ürün alımı yapılmış olduğunu gözlemlediğimizde bu alımı yapan müşterinin gerçek bir kişilik olduğu yorumunu yapabiliriz.

Yüksek alım yapan müşterilerin sepetlerini inceleyerek sektörleri hakkında yorum yapılması mümkündür.

En yüksek alımı yapan müşteri “14646” müşteri numarasına sahip müşteridir. Bu müşterinin alım yaptığı ürünler incelendiğinde “S1” sektöründe karşımıza çıkmaktadır. Sektör 1 sınıfında yer alan müşteriler için mağaza tipi bir işletme olduğu ve “Outdoor” ürün kategorisi haricinde, mevcut bütün kategori ürünlerinden barındıran ve bulunduğu lokasyonda satış gücü yüksek bir işletme olduğu yorumunu yapabiliriz.

Karar ağacı incelendiği zaman karşımıza çıkan bir diğer ayrıntı mutfak araç gereçleri ile Parti araç gereçleri arasında bir bağlantı İtalya için kurulduğunda şu sonucu vermektedir. Ülke genelinde alım yapılan ürün kategorileri ağırlıklı olarak Mutfak araç gereçleri ve parti malzemeleridir. Bu iki kategori arasında yapılan ölçümde ise Mutfak araç gereçleri yüksek oranda bulunmaktadır. Bu ürün kategorisinde bulunan ürün grupları ve alım gücü “S4”te yani Sektör 4’te bulunmaktadır.

Birlikte alınan ürünler sepet analizi ile incelendiğinde alınabilecek ek ürünler müşteriye bir kampanya kapsamında iletilmedir. Müşteri farkında olmadığı ihtiyacı farkedecek ve kendisine bu ihtiyacı farkettilirdiği için işletmeye karşı müşteri sadakati başlayacaktır.

KAYNAKLAR

- [1] Kalikov, A., (2006), Veri Madenciliği ve Bir E-Ticaret Uygulaması, Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü.
- [2] Tan P.N., Steinbach M., Kumar V., Imtroduction to Data Mining, Pearson International Edition, USA, 2006.
- [3] Berson, A., Smith, S. ve Thearling,K. 2000. Building Data Mining Applications for CRM, McGraw-Hill, USA. s. 510
- [4] Introduction to Data Mining and Knowledge Discovery, by Two Crows Corporation
- [5] Moss, L.T. ve Atre, S. 2003. Business intelligence roadmap: the complete project lifecycle for decision-support applications. Addison-Wesley Publishing, 576, USA.
- [6] Doç. DR. Ali Serhan Koyuncugil 2007. Borsa Şirketlerinin Sektörel Risk profillerinin Veri Madenciliğiyle Belirlenmesi, Sermaye Piyasası Kurulu araştırma Raporu 8 p.
- [7] Sertaç ÖĞÜT , Yeditepe Üniversitesi-İletişim Fakültesi,Veri madenciliği Kavramı ve Gelişim Süreci
- [8] Dr. Gökhan Silahtaroglu-Kavram ve algoritmalarıyla Temek Veri Madenciliği s. 13
- [9] <https://tr.wikipedia.org>
- [10] Dr. Gökhan Silahtaroglu-Kavram ve algoritmalarıyla Temek Veri Madenciliği s. 32
- [11] Birliktelik Kuralı Yöntemi için bir veri madenciliği yazılımı tasarımı ve uygulaması Feridun Cemal ÖZÇAKIR, A. Yılmaz ÇAMURCU İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi Yıl: 6 Sayı:12
- [12] The Machine Learning Network Online Information Service, Association Rule Learning <http://kiew.cs.uni-dortmund.de>
- [13] <http://emrahozkaynak.blogcu.com/oruntu-tanima-ders-notlari-1/10026042>
- [14] Dr. Gökhan Silahtaroglu-Kavram ve algoritmalarıyla Temek Veri Madenciliği s. 34
- [15] Dr. Gökhan Silahtaroglu-Kavram ve algoritmalarıyla Temek Veri Madenciliği s. 36

- [16] Dr. Gökhan Silahtaroglu-Kavram ve algoritmalarıyla Temek Veri Madenciliği s. 45
- [17] Çağatan TAŞKIN , Gül Gökay EMEL, Uludağ Üniversitesi, İİBF Veri Madenciliğinde Karar Ağaçları ve Bir Satış Analizi Uygulaması
- [18] Dr. Gökhan Silahtaroglu-Kavram ve algoritmalarıyla Temek Veri Madenciliği s. 67
- [19] Güneş Gürgen Birliktelik Kuralları ile Sepet Analizi ve Uygulaması,Marmara Üniversitesi, 2008
- [20] Mitchell, T., “Machine Learning”, McGraw Hill, New York, 1997.
- [21] <http://kod5.org/naive-bayes-algoritmasi/> adresinden 16.03.2017 tarihinde alınmıştır.
- [22] Akbulut S., Veri Madenciliği Teknikleri ile Bir Kozmetik Markanın Ayrılan Müşteri analizi ve Müşteri Segmentasyonu, Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Ankara, 2006.
- [23] Ayşegül EZERÇE,Mehpare TİMOR, U.Tuğba GÜRSOY Müşteri Profili ve Alışveriş Davranışlarını Belirlemede Kümeleme ve Birliktelik Kuralları Analizi: Perakende Sektöründe Bir Uygulama, İstanbul Üniversitesi, İşletme Fakültesi, 2011.
- [24] Gülnur KEÇE, Nevin UZGÖREN, Zeki ÇAKMAK, Kümeleme analizi Teknikleri İle İllerin Kültürel Yapılarına Göre Sınıflandırılması ve Değişimlerin İncelenmesi, Dumlupınar Üniversitesi, İİBF, İşletme Bölümü
- [25] Gülnur KEÇE, Nevin UZGÖREN, Zeki ÇAKMAK, Kümeleme analizi Teknikleri İle İllerin Kültürel Yapılarına Göre Sınıflandırılması ve Değişimlerin İncelenmesi, Dumlupınar Üniversitesi, İİBF, İşletme Bölümü
- [26] Dr. Gökhan Silahtaroglu-Kavram ve algoritmalarıyla Temek Veri Madenciliği s. 107
- [27] https://tr.wikipedia.org/wiki/K-means_kumeleme adresinden 17.03.2017 tarihinde alınmıştır.
- [28] Torunhan Okumuş, Serra Döver, <https://prezi.com/nffzsf1s-dwh/copy-of-multimedya-madenciligi/> daresinde 17.03.2017 tarihinde alınmıştır.
- [29] Abdullah Baykal, Cengiz Coşkun, Web Madenciliği Teknikleri, Dicle Üniversitesi Fen-Edebiyat Fakültesi
- [30] Hidayet Takıcı, “Kütüphane Kullanıcıların Erişim Desenlerinin Keşfi”
- [31] AKPINAR, H., 2000, "Veri tabanlarında Bilgi Keşfi ve Veri Madenciliği", I.U.İşletme Fakültesi Dergisi, 29(1), s. 1-22.

- [32] BRIJS, T., Swinnen, G., Vanhoof, K., Wets, G., 2004, "Building an Association Rules Framework to Improve Product Assortment Decisions", Data Mining and Knowledge Discovery, 8(1), s: 7-23 .
- [33] DECKER, Reinhold, K. Monien, 2003, "Market Basket Analysis with Neural Gas Networks and Self-Organising Maps ", Journal of Targeting, Measurement and Analysis for Marketing, 11, s: 373-386
- [34] BARBER, F.A., Thomas, R.K., 2001, Huang M.C., "Developing a Profile of LASIK Surgery Customers", Marketing Health Services. 21 (2), s: 32-36.
- [35] ÖZER, M., 2001, "User Segmentation of Online Music Services Using Fuzzy Clustering", Omega, 29 (2), s: 193.
- [36] Timor M., Ezerçe A., Şimşek U.T., “Müşteri Profili Ve Alışveriş Davranışlarını Belirlemede Kümeleme Ve Birliktelik Kuralları Analizi: Perakende Sektöründe Bir Uygulama”, Yönetim Dergisi, 2011,
- [37] <http://acikerisim.aku.edu.tr:8080/xmlui/handle/11630/1757> adresinden 19.03.2017 tarihinde alınmıştır.
- [38] https://prezi.com/1yj_u2hizg7r/copy-of-veri-madenciligi-wekada-snflandrma/ adresinden 19.03.2017 tarihinden alınmıştır.
- [39] www.bilisimrehber.com.tr adresinden 21.03.2017 tarihinde alınmıştır.
- [40] Ceyda DURMAZ Murat KOCAMIŞ, İstanbul Teknik Üniversitesi, Fen Edebiyat Fakültesi, Oracle Data ,Miner ile Öğrenci Kayıtları Üzerine Bir Veri Madenciliği Uygulaması, Mayıs 2008
- [41] Adem TEKEREK, Veri Madenciliği Süreci ve Açık Kaynak Kodlu Veri Madenciliği Araçları, Gazi Üniversitesi, Elektronik-Bilgisayar Eğitimi Bölümü, Ankara
- [42] <https://www.meetup.com/tr-TR/Istanbul-KNIME-Users/events/233227274/?eventId=233227274> adresinden 21.03.2017 tarihinde alınmıştır.
- [43] Tang, Z., MacLennan, J. , "Data Mining with Sql Server 2005", Wiley, 2005.
- [44] Özlem TERZİ, Ecir Uğur KÜÇÜKSİLLE, Gülşah ERGİN, Ahmet İLKER, Veri Madenciliği Süreci Kullanılarak Güneş Işınımı Tahmini, Şubat 2011
- [45] Kaynak: Bloemer, M. M. J., Brijs, T., Vanhoof, K. ve Swinnen, G. (2003); “Comparing Complete And Partial Classification For Identifying Customers At Risk”, International Journal Of Research In Marketing, Vol:20, Issue:2, s. 117-131.

- [46] Hayri Sever, Buket Oğuz, Veritabanlarında Bilgi Keşfine Formal Bir Yaklaşım Kısımı: Eşleştirme Sorguları ve Algoritmalar
- [47] Han, J., Cai, Y., ve Cercone, N. (1992). Knowledge discovery in databases: An attribute-oriented approach. Bildiri Kitabı, 18. VLDB Konferansı, Vancouver, British Columbia, Canada, s. 547-559.
- [48] Chan, K.C.C. ve Wong, A. K. C. (1991). A statistical technique for extracting classificatory knowledge from databases. G. Piatetsky-Shapiro ve W. J. Frawley (eds.), Knowledge Discovery in Databases içinde (s. 107-123). Cambridge, MA: AAAI/MIT.
- [49] Quinlan, J. R. (1986b). The effect of noise on concept learning. R. Michalski, J. Carbonell, ve T. Mitchell (eds.), Machine Learning: An Artificial Intelligence Approach, cilt. 2, s. 149-166. San Mateo, CA: Morgan Kauffmann Inc.
- [50] Uthurusamy, Fayyad ve Spangler, 1991; Thiesson, 1995; Deogun ve diğerleri, 1997; Tolun, Sever ve Uludag, 1998)
- [51] Choubey ve diğerleri, 1996
- [52] Pawlak, Slowinski ve Slowinski, 1986; Baim, 1988; Almuallim ve Dietterich, 1991; Kira ve Rendell, 1992; Deogun, Raghavan ve Sever, 1995
- [53] Hulten, G., Spencer, L. ve Domingos, P. (2001). Mining time-changing data streams. Bildiri Kitabı, 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Fransisco, CA: ACM Press.
- [54] Türk Dil Kurumu
- [55] http://pazarlamablog.blogspot.com.tr/2008/03/perakendecilik.html#_ftn1
- [56] Mintzberg, H. (1989). Mintzberg on Management: Inside Our Strange World of Organizations. Free Press
- [57] <http://www.perakende.org/sektor-haberleri/perakende-pazar-stratejisi-1197545377h.htm>
- [58] http://milasticaretlisesi.meb.k12.tr/meb_iys_dosyalar/48/08/187208/dosyalar/2014_05/07032056_pazarlamaalanbecerieitimikonular.pdf adresinden 25.03.2017 tarihinde alınmıştır.

ÖZGEÇMİŞ

13 Ocak 1990 tarihi, İstanbul doğumluyum. Lise öğrenimimi Üsküdar Anadolu Lisesi'nde tamamlamış bulunmaktayım. 2014 yılında Beykent Üniversitesi Matematik-Bilgisayar bölümü'nde lisans öğrenimimi tamamladım. 2011 yılından itibaren farklı özel IT şirketlerinde Proje Satış alanında Satış Uzmanlığı üzerine çalıştım. 2014 yılında Beykent Üniversitesi Endüstri Mühendisliği Tezli Yüksek Lisans bölümünde yüksek lisans eğitimine başladım.

Şu anda özel bir yazılım şirketinde ERP Proje Satış Uzmanı olarak görev yapmaktayım.

Özlem HACIOSMANOĞLU