

**VERİ MADENCİLİĞİ TEKNİKLERİ İLE BİR KOZMETİK
MARKANIN AYRILAN MÜŞTERİ ANALİZİ
VE MÜŞTERİ SEGMENTASYONU**

Sinem AKBULUT

**YÜKSEK LİSANS TEZİ
ENDÜSTRİ MÜHENDİSLİĞİ**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

MAYIS 2006

ANKARA

Sinem AKBULUT tarafından hazırlanan VERİ MADENCİLİĞİ TEKNİKLERİ İLE BİR KOZMETİK MARKANIN AYRILAN MÜŞTERİ ANALİZİ VE MÜŞTERİ SEGMENTASYONU adlı bu tezin Yüksek Lisans tezi olarak uygun olduğunu onaylarım.


Doç. Dr. Cevriye GENCER
Tez Yöneticisi

Bu çalışma, jürimiz tarafından Endüstri Mühendisliği Anabilim Dalında Yüksek lisans tezi olarak kabul edilmiştir.

Başkan: : Prof. Dr. Semra Oral ERBAŞ



Üye : Prof. Dr. Zülal GÜNGÖR



Üye : Prof. Dr. Hadi GÖKÇEN



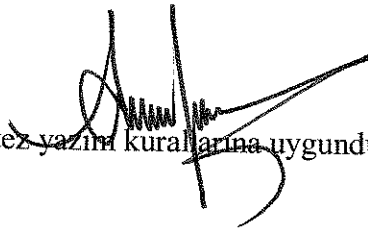
Üye : Doç. Dr. Cevriye GENCER



Üye : Öğr. Gör. Dr. Alb. Nejat BASIM



Tarih : 24/05/2006



Bu tez, Gazi Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygundur.

TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada orijinal olmayan her türlü kaynağa eksiksiz atıf yapıldığını bildiririm.



Sinem AKBULUT

**VERİ MADENCİLİĞİ TEKNİKLERİ İLE BİR KOZMETİK MARKANIN
AYRILAN MÜŞTERİ ANALİZİ VE
MÜŞTERİ SEGMENTASYONU
(Yüksek Lisans Tezi)**

Sinem AKBULUT

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
Mayıs 2006**

ÖZET

Veri madenciliği, büyük veri kümeleri içindeki anlamlı bilgiyi ortaya çıkarma sürecidir. Günümüzde veri madenciliği; pazarlama, bankacılık ve finans, tıp, imalat gibi pek çok alanda uygulanmaktadır.

Veri madenciliğinin en yaygın kullanıldığı uygulama alanlarından biri, ayrılma eğilimi gösteren müşteri profilini belirleme ve müşteri segmentasyonudur. Ayrılma eğilimi gösteren müşteri profilini belirleme, şirketlerin bu müşterilere özel pazarlama kampanyalarını geliştirmelerini sağlamaya yöneliktir. Müşteri segmentasyonu ise benzer özellikler gösteren müşterileri gruplandırarak; gruplara özel, pazarlama programlarının geliştirilmesini sağlar.

Bu tez, bir kozmetik markasının müşteri gruplarını ve ayrılma eğilimi gösteren müşteri profilini belirleyerek; bu müşterilere özel pazarlama stratejileri geliştirilmesini hedeflemektedir. Segmentasyon için kümeleme teknikleri, ayrılacak müşteri profilini belirlemek için sınıflama teknikleri kullanılmıştır.

Bilim Kodu : 906
Anahtar Kelimeler :Veri madenciliği, sınıflama, kümeleme, müşteri
Sayfa Adedi : 91
Tez Yöneticisi : Doç. Dr. Cevriye Gencer

**CHURN ANALYSIS AND CUSTOMER SEGMENTATION OF
A COSMETICS BRAND
USING DATA MINING TECHNIQUES
(M.Sc. Thesis)**

Sinem AKBULUT

**GAZI UNIVERSITY
INSTITUTE OF SCIENCE AND TECHNOLOGY
May 2006**

ABSTRACT

Data minig is the process of finding hidden and unknown patterns in huge amounts of data. Data mining has a wide application area such as marketing, banking and finance, medicine and manufacturing.

Churn analysis and customer segmentation are two common application areas of data mining. Churn modeling is predicting which customers will leave the company. This allows companies to increase customer loyalty by producing special campaigns. Customer segmentation is the process of dividing customers into mutually exclusive groups. Segmenting customers enables companies to develop customized marketing programs for them.

This thesis aims to determine the customer profile who likely to leave and the customer segments of a cosmetic brand and develop customized campaigns and marketing strategies. Clustering techniques used for segmentation and classification techniques used for determining the churners.

Science Code : 906

Key Words : Data mining, classification, clustering, customer

Page Number: 91

Adviser : Assoc. Prof. Dr. Cevriye Gencer

TEŞEKKÜR

Çalışmalarım boyunca değerli yardım ve katkılarıyla beni yönlendiren Hocam Doç. Dr. Cevriye GENCER'e ve tecrübelerinden faydalandığım Araştırma Görevlisi Emel Kızılkaya AYDOĞAN'a teşekkürü bir borç bilirim.

Ayrıca, destek ve hoşgörülerini benden esirgemeyen aileme ve Müdürüm Sefa TEKE'ye sonsuz teşekkürler...

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT	v
TEŞEKKÜR	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ	x
ŞEKİLLERİN LİSTESİ	xi
RESİMLERİN LİSTESİ	xii
1. GİRİŞ	1
2. VERİ MADENCİLİĞİ METODOLOJİSİ	3
2.1. Veri Madenciliğine Genel Bakış	3
2.2. Veri Madenciliğini Etkileyen Faktörler	5
2.3. Veri Madenciliğinin Gereksinimleri	6
2.4. Veri Madenciliğinin Uygulama Alanları	7
2.5. Veri Madenciliği Uygulamalarında Karşılaşılabilecek Problemler	12
2.6. Veri Madenciliği Yazılımları	12
2.7. Veri Madenciliği Süreci	13
2.7.1. Problemin tanımlanması	14
2.7.2. Verilerin hazırlanması	15
2.7.3. Modelin kurulması ve değerlendirilmesi	16
2.7.4. Modelin kullanılması	19
2.7.5. Modelin izlenmesi	19
2.8. Veri Madenciliği Teknikleri	20

	Sayfa
2.8.1. Sınıflama	20
2.8.2. Kümeleme	24
2.8.3. Birliktelik kuralları ve ardışık örüntüler	25
2.9. Literatürde Veri Madenciliği.....	26
3. MÜŞTERİ İLİŞKİLERİ YÖNETİMİ VE VERİ MADENCİLİĞİ.....	28
3.1. Müşteri Edinme	30
3.2. Müşteri Segmentasyonu	30
3.3. Ayrılan Müşteri Analizi ve Müşteri Tutma.....	33
3.4. Çapraz Satış.....	33
3.5. Müşteri Yaşam Ömrü Değeri	34
3.6. Pazarlama İletişimi ve Tepki Modelleme	34
3.7. Veri Madenciliğinin Pazarlama ve Müşteri İlişkileri Uygulamaları Literatürü.....	36
4. UYGULAMA	38
4.1. Problem Tanımı.....	38
4.2. Veri Madenciliği Süreci	39
4.2.1. Veri temizleme	39
4.2.2. Veri dönüştürme.....	41
4.2.3. Modelleme	47
5. SONUÇ VE ÖNERİLER	62
KAYNAKLAR	65
EKLER.....	70
EK-1 Orijinal veri tablosu.....	71

Sayfa

EK-2 ARFF uzantılı veri dosyası	72
EK-3 Naive Bayes algoritması sonuçları	75
EK-4 Lojistik regresyon algoritması sonuçları	76
EK-5 ID.3 algoritması sonuçları	78
EK-6 J.48 algoritması sonuçları	79
EK-7 JRIP algoritması sonuçları	80
EK-8 PART algoritması sonuçları	81
EK-9 Sinir ağları algoritması sonuçları	82
ÖZGEÇMİŞ	91

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 2.1. Veri madenciliğinin uygulandığı sektörler.....	11
Çizelge 2.2. Yazılım sağlayıcılar	13
Çizelge 2.3. Risk matrisi	18
Çizelge 4.1. Nitelik isimleri	45
Çizelge 4.2. Veritabanı istatistikleri.....	46
Çizelge 4.3. Bayes / Naive Bayes algoritması doğruluk değerleri	49
Çizelge 4.4. Fonksiyon / Lojistik regresyon algoritması doğruluk değerleri	50
Çizelge 4.5. Ağaçlar / ID.3 algoritması doğruluk değerleri.....	50
Çizelge 4.6. Ağaçlar / J.48 algoritması doğruluk değerleri.....	51
Çizelge 4.7. Kurallar / PART algoritması doğruluk değerleri	51
Çizelge 4.8. Kurallar / JRIP algoritması doğruluk değerleri.....	52
Çizelge 4.9. Sinir ağları / Multilayer Perceptron algoritması doğruluk değerleri	52
Çizelge 4.10. Modellemede kullanılan sınıflama algoritmaları ve doğrulukları	53
Çizelge 4.11. J.48 Karar ağacı	55
Çizelge 4.12. Simple K Means algoritması çıktısı.....	59
Çizelge 4.13. Müşteri gruplarının baskın özellikleri.....	60

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 2.1. Veri madenciliğinin veri işleme süreci içindeki yeri	5
Şekil 2.2. Veri madenciliği süreci	14
Şekil 3.1. Müşteri verilerinin kullanım amaçları	29
Şekil 3.2. Müşteri edinme faaliyetleri ve diğer pazarlama fonksiyonları ilişkisi.....	30
Şekil 3.3. Müşteri segmentasyonu	32
Şekil 4.1. Kozmetik markası veri madenciliği süreci	39
Şekil 4.2. Yaş dağılımı grafiği	42
Şekil 4.3. Müşterilerin semtlere göre dağılım grafiği	42
Şekil 4.4. Ortalama alışveriş periyodu grafiği	44
Şekil 4.5. Ortalama harcama tutarı grafiği	44
Şekil 4.7. J.48 Karar ağacı	57

RESİMLERİN LİSTESİ

Resim	Sayfa
Resim 4.1. WEKA paket programı ekran görünümü	47

1.GİRİŞ

Globalleşme ve artan rekabet, müşteri satın alma davranışlarını ve beklentilerini, büyük ölçüde değiştirmiştir. Değişen müşteri davranışlarını, işletme amaç ve hedeflerine uygun olarak yönetebilmek; değişim yönetimi ve müşteri ilişkileri yönetimi kavramlarını anlayabilmek ve doğru bir şekilde yorumlayabilmek ile mümkün olabilecektir. İşletmelerde başarılı bir müşteri ilişkileri yönetimiyle, küresel bir dünyada ve giderek artan rekabetçi piyasalarda, işletmeler için yaşamsal önem taşıyan, müşteri için değer yaratmak, müşteri sadakati sağlamak ve bu konularda kurumsallaşmayı gerçekleştirmek mümkün olabilecektir.

Müşteri ilişkileri yönetiminde pazar payı kavramından, müşteri payı kavramına geçiş söz konusu olmaktadır. Müşteri payı aynı müşteriye birden fazla ürün satabilmeyi ve müşteriyi aktif ve sadık bir müşteriye dönüştürmeyi amaçlamaktadır. Bu durum ise tek yönlü pazarlama anlayışından karşılıklı ilişkiye dayalı pazarlamaya geçişi ifade etmektedir. Günümüzde kitlesel pazarlama yöntemleri, giderek müşteriye hitap eden kişisel pazarlama yöntemlerine yönelmektedir.

Diğer taraftan bilgisayar teknolojilerindeki gelişmeler, işletmelerin çok miktarda veriyi saklayabilmesini ve işleyerek anlamlı bilgilere dönüştürmesini mümkün hale getirmiştir. Bugün, işletmeler satın alma davranışından demografik özelliklere kadar müşterileri ilgili pek çok detayı veritabanlarında tutmaktadır. Veri madenciliği tekniklerini kullanarak bu veriler içerisindeki anlamlı ve gizli örüntülerin ortaya çıkarılması mümkün olmaktadır. Veri madenciliği sonuçları müşteri odaklı birçok uygulamaya girdi teşkil etmektedir.

Veri madenciliğinin en yaygın kullanıldığı uygulama alanlarından biri, müşteri segmentasyonu ve ayrılarak rakip şirketi tercih etme eğilimi gösteren müşteri profilini belirlemedir.

Müşteri segmentasyonu, pazarın benzer karakteristikler taşıyan müşteri gruplarına bölünmesidir. Her gruptaki müşterilerin davranışlarını anlamak, şirketlerin benzer

davranışlı müşteri gruplarına özgü ürün önerileri oluşturmalarını ve pazarlama programları geliştirmelerini sağlar. Ayrılacak müşteri tahmininin temel amacı karlı müşterileri elde tutmaktır. Hangi müşterilerin ayrılma eğilimi gösterdiği tespit edildikten sonra, bu müşteriler üzerinde belirli bir pazarlama faaliyeti hedeflenerek müşteri tutmak için etkin stratejiler geliştirmek mümkündür

Bu tezin amacı bir kozmetik markasının müşteri gruplarını ve ayrılma eğilimi gösteren müşteri profilini belirleyerek; bu müşterilere özel kampanyalar ve pazarlama stratejileri geliştirmektir.

Tezin ikinci bölümünde veri madenciliğinin tanımı, adımları, uygulama alanları, veri madenciliği teknik ve algoritmaları anlatılmış; üçüncü bölümde ise veri madenciliğinin müşteri ilişkileri yönetimi ve pazarlama uygulamaları içindeki yeri belirtilmiş; son bölümde; kozmetik mağazasında yapılan uygulamanın detayları adımlar bazında ortaya konmuş ve elde edilen sonuçlar değerlendirilmiştir.

2. VERİ MADENCİLİĞİ METODOLOJİSİ

2.1. Veri Madenciliğine Genel Bakış

Veri madenciliği; verideki trendleri, ilişkileri ve profilleri belirlemek için veriyi sınıflandıran bir analitik araç ve bilgisayar yazılım paketidir. Spesifik veri madenciliği yazılımları; kümeleme, doğrusal regresyon, sinir ağları, Bayes ağları, görselleştirme ve ağaç tabanlı modeller gibi pek çok modeli içerir. Veri madenciliği uygulamalarında yıllar boyu istatistiksel yöntemler kullanılmıştır. Bununla birlikte, bugünün veri madenciliği teknolojisinde eski yöntemlerin tersine büyük veri kümelerindeki trend ve ilişkileri kısa zamanda saptayabilmek için yüksek hızlı bilgisayarlar kullanılmaktadır. Böylece veri madenciliği, gizli trendleri minimum çaba ve emekle ortaya çıkarmaktadır.

Veri ambarı, organizasyonun ihtiyaçları ile uyumlu büyük miktarlardaki verinin kolay erişilebilir bir yapıda tutulmasını sağlayan bilgisayar tabanlı depolama sistemleridir [1]. Veri ambarları organizasyonel dataya kolay bir şekilde ulaşılmasını sağlayan yapılardır. Veri ambarları 1990'lı yıllarda ortaya çıkmıştır. Veri ambarları veriyi kullanılabilir trend, ilişki ve profillerde sınıflandırmazlar, sadece potansiyel bilgiye sahip veritabanlarıdır. Veride saklı bilgiyi keşfetmeyi sağlayan ise veri madenciliği gibi tekniklerdir. Veri ambarından veriyi çekebilmek için hangi verinin gerekli olduğunu ve bu verinin nerede olduğunu tespit etmek önemlidir. Çoğunlukla gerekli veri, farklı sistemler üzerinde olup, farklı formatlardadır. Bu nedenle, ilk aşamada veri temizleme ve düzenleme işlemi gerçekleştirilmelidir. Veri ambarının yaratıcısı W.H. Immon'a göre veri ambarı verinin temizlendiği, birleştirildiği ve yeniden düzenlendiği merkezi ve entegre bir depodur [2].

Aşağıda veri madenciliği ile ilgili çeşitli tanımlar yer almaktadır:

- Jacobs (1999), veri madenciliğini, ham datanın tek başına sunamadığı bilgiyi çıkaran veri analizi süreci olarak tanımlamıştır [3].

- David (1999), veri madenciliğinin büyük hacimli datalardaki örüntüleri araştıran matematiksel algoritmaları kullandığını söylemiştir. David'e göre veri madenciliği hipotezleri keşfeder, sonuçları birleştirmek için insan yeteneğini kullanır. Veri madenciliğinin sadece bir bilim olmadığı, aynı zamanda bir sanat olduğu da söylenebilir [4].

- DuMouchel (1999), veri madenciliğinin geniş veritabanlarındaki birliktelikleri araştırdığını belirtmiştir [5].

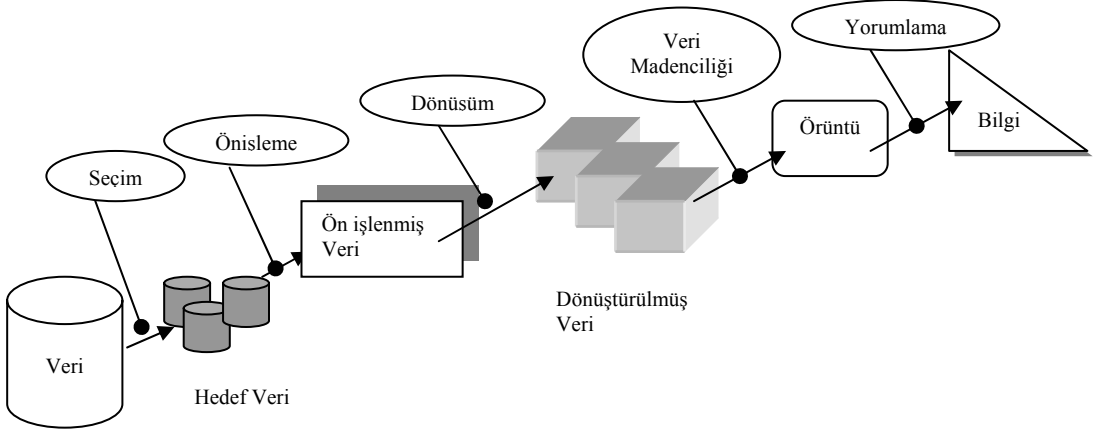
- Hand (1998), veri madenciliğini istatistik, veritabanı teknolojisi, örüntü tanıma, makine öğrenme ile etkileşimli yeni bir disiplin ve geniş veritabanlarında önceden tahmin edilemeyen ilişkilerin ikincil analizi olarak tanımlamıştır [6].

- Kitler ve Wang (1998), veri madenciliğini oldukça tahminci anahtar değişkenlerin binlerce potansiyel değişkenden izole edilmesini sağlama yeteneği olarak tanımlamışlardır [7].

- Bransten (1999), veri madenciliğinin insanın asla bulmayı hayal bile edemeyeceği trendlerin keşfedilmesini sağladığını belirtmiştir [8].

Sonuç olarak veri madenciliği, önceden bilinmeyen ilişki ve trendlerin bulunması için bugünün endüstrisinde yaratılan büyük miktarlardaki veriyi analiz eden bir yoldur. Yüksek güçlü bilgisayarlara ve gereken yazılımlara kolay ve düşük fiyatlarla ulaşılabilmesi bu teknolojinin işlemlerini olanaklı kılmıştır. İnternet ise birçok noktadaki verinin toplanmasını sağlamaktadır. Bilgisayar üzerinde çalışma aynı zamanda emek ve zaman tasarrufunu sağlamıştır.

Şekil 2.1’de veri madenciliğinin veri işleme ve veriyi bilgiye dönüştürme süreci içerisindeki yeri gösterilmektedir [9].



Şekil 2.1. Veri madenciliğinin veri işleme süreci içindeki yeri

2.2. Veri Madenciliğinin Gelişimini Etkileyen Faktörler

Temel olarak veri madenciliğini 5 ana faktör etkiler [10]:

Veri: Veri madenciliğinin bu kadar gelişmesindeki en önemli faktördür. Son yirmi yılda sayısal verinin hızla artması, veri madenciliğindeki gelişmeleri hızlandırmıştır. Bu kadar fazla veriye bilgisayar ağları üzerinden erişilmektedir. Diğer yanda bu verilerle uğraşan bilim adamları, mühendisler ve istatistikçilerin sayısı hala aynıdır. O yüzden, verileri analiz etme yöntemleri ve teknikleri geliştirilmektedir.

Donanım: Veri madenciliği, sayısal ve istatistiksel olarak büyük veri kümeleri üzerinde yoğun işlemler yapmayı gerektirir. Gelişen bellek ve işlem hızı kapasitesi sayesinde, birkaç yıl önce madencilik yapılamayan veriler üzerinde çalışmayı mümkün hale getirmiştir.

Bilgisayar ağları: Yeni nesil İnternet, çok yüksek hızları kullanmamızı sağlamaktadır. Böyle bir bilgisayar ağı ortamı oluştuktan sonra, dağıtık verileri analiz etmek ve farklı algoritmaları kullanmak mümkün olacaktır. Bundan 10 yıl

önceki bilgisayar ağları teknolojisinde hayal edemediklerimizi artık kullanabiliyoruz. Buna bağlı olarak, veri madenciliğine uygun ağların tasarımı da yapılmaktadır.

Bilimsel hesaplamalar: Günümüz bilim adamları ve mühendisleri, simülasyonu, bilimin üçüncü yolu olarak görmekteler. Veri madenciliği ve bilgi keşfi, teori, deney ve simülasyonu birbirine bağlamada önemli rol almaktadır.

Ticari eğilimler: Günümüzde, işletmeler rekabet ortamında varlıklarını koruyabilmek için daha hızlı hareket etmeli, daha yüksek kalitede hizmet sunmalı, bütün bunları yaparken de minimum maliyeti ve en az insan gücünü göz önünde bulundurmalıdır. Bu tip hedef ve kısıtların yer aldığı iş dünyasında veri madenciliği, temel teknolojilerden biri haline gelmiştir. Çünkü veri madenciliği sayesinde müşterilerin ve müşteri faaliyetlerinin yarattığı fırsatlar daha kolay tespit edilebilmekte ve riskler daha açık görülebilmektedir.

2.3. Veri Madenciliğinin Gereksinimleri

Veri madenciliğinin gereksinimleri aşağıdadır:

- Erişilebilir veri,
- Etkin erişim yöntemleri,
- Açık problem tanımı,
- Etkin algoritmalar,
- Yüksek performanslı uygulama sunucusu,
- Sonuç oluşturmada esneklik.

Veri madenciliğinin diğer bir gereksinimi temizlenmiş veridir. Veri madenciliğinde kullanılacak veri yanlış sonuçlar üretmeye yol açabilecek aykırı değerler veriden temizlenmelidir [3]. Doğru veri mevcut değilse ve verinin limitleri bilinmiyorsa; kullanılan yazılımın yanlış sonuçlar üretmesi kaçınılmazdır.

2.4. Veri Madenciliğinin Uygulama Alanları

Veri madenciliğinin asıl amacı, veri yığınlarından anlamlı bilgiler elde etmek ve bunu eyleme dönüştürecek kararlar için kullanmaktır. Örnek birkaç kullanım alanı aşağıdadır [9]:

- Bir işletme kendi müşterisiyken rakibine giden müşterilerle ilgili analizler yaparak rakiplerini tercih eden müşterilerinin özelliklerini elde edebilir ve bundan yola çıkarak gelecek dönemlerde kaybetme olasılığı olan müşterilerin kimler olabileceği yolunda tahminlerde bulunarak onları kaybetmemek, kaybettiklerini geri kazanmak için strateji geliştirebilir.
- Ürün veya hizmette hangi özelliklerin ne derecede müşteri memnuniyetini etkilediği, hangi özelliklerinden dolayı müşterinin bunları tercih ettiği ortaya çıkarılabilir.
- Müşterilerin kredi riskleri hesaplanarak hangi müşterilerin kredi riskinin yüksek olduğu, hangi müşterilerin geri ödemesini zamanında yapamayabileceği kestirilebilir.
- Kredi kartı ödemelerini aksatan, gecikmeli olarak yapan veya hiç yapmayanların özelliklerinden yola çıkılarak bundan sonra aynı duruma düşebilecek muhtemel kişiler saptanabilir.
- Ürün talebi bazında müşteri profillerini belirleyerek, müşteri segmentasyonuna gitmek ve çapraz satış olanakları yaratmakta kullanılabilir.
- Piyasada oluşabilecek değişikliklere mevcut müşteri portföyünün vereceği tepkinin firma üzerinde yaratabileceği etkinin tespitinde kullanılabilir.

- En karlı mevcut müşteriler saptanarak, potansiyel müşteriler arasından en karlı olabilecekler belirlenebilir. Karlı müşteriler tespit edilerek onlara özel kampanyalar uygulanabilir. En masraflı müşteriler daha masrafsız müşteri haline dönüştürülebilir. Örneğin en çok bankacılık işlemi yapanlar ortaya çıkarılıp bunlar şube bankacılığı yerine daha masrafsız İnternet bankacılığına yönlendirilebilir.

- Bir ürün veya hizmetle ilgili bir kampanya programı oluşturmak için hedef kitlenin seçiminden başlayarak bunun hedef kitleye hangi kanallardan sunulacağı kararına kadar olan süreçte veri madenciliği kullanılabilir.

- Operasyonel süreçte oluşabilecek olası kayıpların veya suiistimallerin tespitinde kullanılabilir.

- Kurum teknik kaynaklarının optimal şekilde kullanılmasını sağlamakta kullanılabilir.

- Firmanın finansal yapısının, makro ekonomik değişimler karşısındaki duyarlılığı ve oluşabilecek risklerin tespitinde kullanabilir.

- Geçmiş ve mevcut yapı analiz edilerek geleceğe yönelik tahminlerde bulunulabilir. Özellikle ciro, karlılık, pazar payı, gibi analizlerde veri madenciliği çok rahat kullanılabilir.

Yukarıda anlatılanlar ışığında veri madenciliğinin uygulama alanları konu başlıkları itibariyle aşağıdaki gibi sınıflandırılabilir [10].

Veri madenciliğinin pazarlama alanındaki kullanım amaçları aşağıdadır:

- Müşteri segmentasyonunda,
- Müşterilerin demografik özellikleri arasındaki bağlantıların kurulmasında,
- Çeşitli pazarlama kampanyalarında,

- Mevcut müşterilerin elde tutmada,
- Pazar sepeti analizinde,
- Çapraz satış analizleri,
- Müşteri değerlendirme,
- Müşteri ilişkileri yönetiminde,
- Çeşitli müşteri analizlerinde,
- Satış tahminlerinde.

Veri madenciliğinin bankacılık alanındaki kullanım amaçları aşağıdadır:

- Farklı finansal göstergeler arasındaki gizli korelasyonların bulunmasında,
- Kredi kartı dolandırıcılıklarının tespitinde,
- Müşteri segmentasyonunda,
- Kredi taleplerinin değerlendirilmesinde,
- Usulsüzlük tespiti,
- Risk analizleri,
- Risk yönetimi,
- Sigortacılık,
- Yeni poliçe talep edecek müşterilerin tahmin edilmesinde,
- Sigorta dolandırıcılıklarının tespitinde,
- Riskli müşteri tipinin belirlenmesinde.

Veri madenciliğinin perakendecilik alanındaki kullanım amaçları aşağıdadır:

- Satış noktası veri analizleri,
- Alış-veriş sepeti analizleri,
- Tedarik ve mağaza yerleşim optimizasyonu.

Veri madenciliğinin borsa alanındaki kullanım amaçları aşağıdadır:

- Hisse senedi fiyat tahmini,
- Genel piyasa analizleri,
- Alım-satım stratejilerinin optimizasyonu.

Veri madenciliğinin telekomünikasyon alanındaki kullanım amaçları aşağıdadır:

- Kalite ve iyileştirme analizlerinde,
- Hisse tespitlerinde,
- Hatların yoğunluk tahminlerinde,
- İletişim desenlerinin belirlenmesi,
- Kaynakların daha iyi kullanılması,
- Servis kalitesinin artırılması.

Veri madenciliğinin sağlık ve ilaç alanındaki kullanım amaçları aşağıdadır:

- Test sonuçlarının tahmini,
- Ürün geliştirme,
- Tıbbi teşhis,
- Tedavi sürecinin belirlenmesinde,
- Semptomlara göre hastalık tespiti,
- Magnetik rezonans verileri ile sinir sistemi bölge ilişkilerinin belirlenmesi.

Veri madenciliğinin endüstri alanındaki kullanım amaçları aşağıdadır:

- Kalite kontrol analizlerinde,
- Lojistik,
- Üretim süreçlerinin optimizasyonunda.

Çizelge 2.1’de 2003 yılında veri madenciliğinin sektörler bazında kullanımına ilişkin bir araştırmanın sonuçları yer almaktadır [11]. Bu çizelgede araştırmaya katılan toplam 421 şirketin 51 adedinin bankacılık alanında veri madenciliğinin kullandığı görülmektedir.

Çizelge 2.1. Veri madenciliğinin uygulandığı alanlar

Son 3 yıl içinde veri madenciliğinin uygulandığı alanlar	
Bankacılık (51)	12%
Bioteknoloji / Genetik (11)	3%
Kredi skorlama (35)	8%
CRM (52)	12%
Doğrudan pazarlama (34)	8%
e-Ticaret (11)	3%
Eğlence/ Müzik (4)	1%
Sahtekarlık tespiti (31)	7%
Şans oyunu (2)	0,01 %
Kamu uygulamaları (12)	3%
Sigortacılık (24)	6%
Yatırım / Hisse senedi (5)	1%
Junk email / Anti-spam (5)	1%
Sağlık/ İK (15)	4%
İmalat (19)	5%
Tıp/ Farmakoloji (12)	3%
Perakende (25)	6%
Bilim (17)	4%
Güvenlik / Anti-terörizm(5)	1%
Telekomünikasyon (23)	5%
Seyahat (8)	2%
Web (9)	2%
Diğer (11)	3%

2.5. Veri Madenciliği Uygulamalarında Karşılaşılan Problemler

Veri madenciliği girdi olarak kullanılacak ham veriyi veritabanlarından alır. Bu da veritabanlarının dinamik, eksiksiz, geniş ve net veri içermemesi durumunda sorunlar doğurur [12].

Diğer sorunlar da verinin konu ile uyumsuzluğundan doğabilir. Sınıflandırmak gerekirse başlıca sorunlar aşağıdadır:

Sınırlı bilgi: Veritabanları genel olarak veri madenciliği dışındaki amaçlar için tasarlanmışlardır. Bu yüzden, öğrenme görevini kolaylaştıracak bazı özellikler bulunmayabilir.

Gürültü ve kayıp değerler: Veri özellikleri ya da sınıflarındaki hatalara gürültü adı verilir. Veri tabanlarındaki eksik bilgi ve bu yanlışlardan dolayı veri madenciliği amacına tam olarak ulaşmayabilir. Bu bilgi yanlışlığı, ölçüm hatalarından, ya da öznel yaklaşımdan olabilir.

Belirsizlik: Yanlışlıkların şiddeti ve verideki gürültünün derecesi ile ilgilidir. Veri tahmini bir keşif sisteminde önemli bir husustur.

Ebat, güncellemeler ve konu dışı sahalara: Veri tabanlarındaki bilgiler, veri eklendikçe ya da silindikçe değişebilir. Veri madenciliği perspektifinden bakıldığında, kuralların hala aynı kalıp kalmadığı ve istikrarlılığı problemi ortaya çıkar. Öğrenme sistemi, kimi verilerin zamanla değişmesine ve keşif sisteminin verinin zamansızlığına karşın zaman duyarlı olmalıdır.

2.6. Veri Madenciliği Yazılımları

Farklı algoritmaları bünyesinde bulunduran ve farklı işletim sistemleri üzerinde çalışabilen birçok veri madenciliği yazılımı bulunmaktadır. Çizelge 2.2’de, bu

yazılımlardan bazıları üzerinde çalıştıkları platform ve fonksiyonel özellikleri ile birlikte gösterilmektedir.

Çizelge 2.2.Yazılım sağlayıcılar

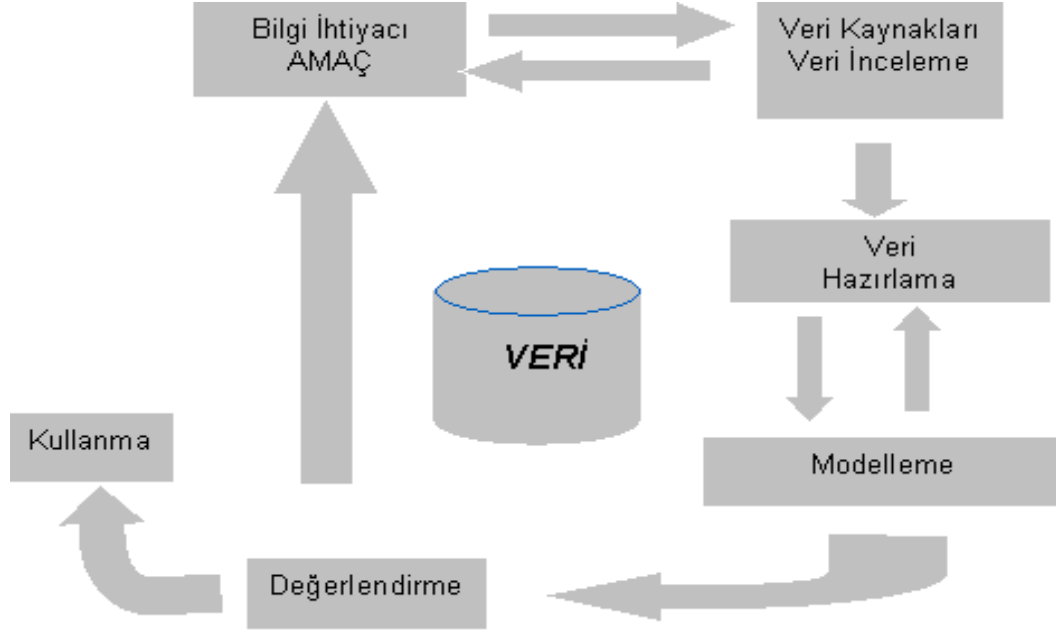
Ürün Adı	IBM Intelligent Miner	Oracle Darwin	SAS Enterprise Miner	Angoss Knowledge Seeker	NCR Teraminer Stats	WEKA
Platform	AIX 4.1, NVS, AS/400, Windows	Windows	Masintosh, Windows, Unix	Windows, Unix	Windows, Unix	Masintosh Windows Unix
Karar Ağacı	X	X	X	X		X
Sinir Ağları	X	X	X		X	
Zaman Serileri	X		X			
Tahmin	X	X	X	X		X
Kümeleme	X		X			X
Birliktelik	X		X			X
Görselleştirme	X	X	X	X	X	X

2.7. Veri Madenciliği Süreci

Ne kadar etkin olursa olsun hiç bir veri madenciliği algoritmasının üzerinde inceleme yapılan işin ve verilerin özelliklerinin bilinmemesi durumunda fayda sağlaması mümkün değildir. Bu nedenle yukarıda tanımlanan tüm aşamalardan önce, iş ve veri özelliklerinin öğrenilmesi / anlaşılması başarının ilk şartı olacaktır. Başarılı bir veri madenciliği projelerinde izlenmesi gereken adımları aşağıdadır [13];

1. Problemin tanımlanması,
2. Verilerin hazırlanması,
3. Modelin kurulması ve değerlendirilmesi,
4. Modelin kullanılması,
5. Modelin izlenmesi.

Veri madenciliği süreci Şekil 2.2’de gösterilmektedir.



Şekil 2.2. Veri madenciliği süreci

2.7.1. Problemin tanımlanması

Veri madenciliği çalışmalarında başarılı olmanın en önemli şartı, projenin hangi işletme amacı için yapılacağını ve elde edilecek sonuçların başarı düzeylerinin nasıl ölçüleceğinin tanımlanmasıdır. Ayrıca yanlış tahminlerde katlanılacak olan maliyetlere ve doğru tahminlerde kazanılacak faydalara ilişkin tahminlere de bu aşamada yer verilmelidir.

Bu aşamada mevcut iş probleminin nasıl bir sonuç üretilmesi durumunda çözüleceğinin, üretilecek olan sonucun fayda - maliyet analizinin başka bir ifadeyle üretilen bilginin işletme için değerinin doğru analiz edilmesi gerekmektedir. Analistin, işletmede üretilen sayısal verilerin boyutlarını, proje için yeterlilik düzeyini ve iş süreçlerini iyi analiz etmesi gerekmektedir.

2.7.2. Verilerin hazırlanması

Veri madenciliğinin en önemli aşamalarından biri olan verinin hazırlanması aşaması, analistin toplam zaman ve enerjisinin %50 - %85 ini harcamasına neden olmaktadır [14]. Bu aşamada firmanın mevcut bilgi sistemleri üzerinde ürettiği sayısal bilginin iyi analiz edilmesi, veriler ile mevcut iş problemi arasında ilişki olması gerektiği unutulmamalıdır. Proje kapsamında kullanılacak sayısal verilerin, hangi iş süreçleri ile yaratıldığı da bu veriler kullanılmadan analiz edilmelidir, bu sayede analist veri kalitesi hakkında fikir sahibi olabilir.

Verilerin hazırlanması aşaması kendi içerisinde toplama, birleştirme ve temizleme, dönüştürme adımlarından meydana gelmektedir.

Veri toplama

Tanımlanan problem için gerekli olduğu düşünülen verilerin ve bu verilerin toplanacağı veri kaynaklarının belirlenmesi adımdır. Verilerin toplanmasında kuruluşun kendi veri kaynaklarının dışında, nüfus sayımı, hava durumu, merkez bankası kara listesi gibi veri tabanlarından veya veri pazarlayan kuruluşların veri tabanlarından faydalanılabilir.

Bu adımda toplanan verilerde bulunan farklılıklar giderilmeye çalışılır. Hatalı veya analizin yanlış yönlendirilmesine sebep olabilecek verilerin temizlenmesine çalışılır. Genellikle yanlış veri girişinden veya bir kereye özgü bir olayın gerçekleşmesinden kaynaklanan verilerin, önemli bir uyarıcı enformasyon içerip içermediği kontrol edildikten sonra veri kümesinden atılması tercih edilir. Ancak basit yöntemlerle ve baştan savma olarak yapılacak sorun giderme işlemlerinin, ileriki aşamalarda daha büyük sorunların kaynağı olacağı unutulmamalıdır.

Veri dönüştürme

Kullanılacak model ve algoritma çerçevesinde verilerin tanımlama veya gösterim şeklinde değiştirilmesi gerekebilir. Örneğin kredi riski uygulamasında iş tiplerinin, gelir seviyesi ve yaş gibi değişkenlerin kodlanarak gruplanmasının faydalı olacağı düşünülmektedir.

2.7.3. Modelin kurulması ve değerlendirilmesi

Tanımlanan problem için en uygun modelin bulunabilmesi, olabildiğince çok sayıda modelin kurularak denenmesi ile mümkündür. Bu nedenle veri hazırlama ve model kurma aşamaları, en iyi olduğu düşünülen modele varıncaya kadar yinelenen bir süreçtir.

Model kuruluş süreci denetimli (supervised) ve denetimsiz (unsupervised) öğrenimin kullanıldığı modellere göre farklılık göstermektedir [15].

Örnekten öğrenme olarak da isimlendirilen denetimli öğrenimde, bir denetçi tarafından ilgili sınıflar önceden belirlenen bir kritere göre ayrılarak, her sınıf için çeşitli örnekler verilir. Sistemin amacı verilen örneklerden hareket ederek her bir sınıfa ilişkin özelliklerin bulunmasıdır.

Öğrenme süreci tamamlandığında, tanımlanan kurallar verilen yeni örneklerle uygulanır ve yeni örneklerin hangi sınıfa ait olduğu kurulan model tarafından belirlenir.

Denetimsiz öğrenimde, kümeleme analizinde olduğu gibi ilgili örneklerin gözlenmesi ve bu örneklerin özellikleri arasındaki benzerliklerden hareket ederek sınıfların tanımlanması amaçlanmaktadır.

Denetimli öğrenimde seçilen algoritmaya uygun olarak ilgili veriler hazırlandıktan sonra, ilk aşamada verinin bir kısmı modelin öğrenimi, diğer kısmı ise modelin

geçerliliğinin test edilmesi için ayrılır. Modelin öğrenimi öğrenim kümesi kullanılarak gerçekleştirildikten sonra, test kümesi ile modelin doğruluk derecesi (accuracy) belirlenir.

Bir modelin doğruluğunun test edilmesinde kullanılan en basit yöntem basit geçerlilik (simple validation) testidir. Bu yöntemde tipik olarak verilerin %5 ile %33 arasındaki bir kısmı test verileri olarak ayrılır ve kalan kısım üzerinde modelin öğrenimi gerçekleştirildikten sonra, bu veriler üzerinde test işlemi yapılır. Bir sınıflama modelinde yanlış olarak sınıflanan olay sayısının, tüm olay sayısına bölünmesi ile hata oranı, doğru olarak sınıflanan olay sayısının tüm olay sayısına bölünmesi ile ise doğruluk oranı hesaplanır [10].

Sınırlı miktarda veriye sahip olunması durumunda, kullanılabilecek diğer bir yöntem çapraz geçerlilik (cross validation) testidir. Bu yöntemde veri kümesi tesadüfi olarak iki eşit parçaya ayrılır. İlk aşamada a parçası üzerinde model eğitimi ve b parçası üzerinde test işlemi; ikinci aşamada ise b parçası üzerinde model eğitimi ve a parçası üzerinde test işlemi yapılarak elde edilen hata oranlarının ortalaması kullanılır.

Bir kaç bin veya daha az satırdan meydana gelen küçük veri tabanlarında, verilerin n gruba ayrıldığı n katlı çapraz geçerlilik (n-fold cross validation) testi tercih edilebilir. Verilerin örneğin 10 gruba ayrıldığı bu yöntemde, ilk aşamada birinci grup test, diğer gruplar öğrenim için kullanılır. Bu süreç her defasında bir grubun test, diğer grupların öğrenim amaçlı kullanılması ile sürdürülür. Sonuçta elde edilen on hata oranının ortalaması, kurulan modelin tahmini hata oranı olacaktır.

Bootstrapping küçük veri kümeleri için modelin hata düzeyinin tahmininde kullanılan bir başka tekniktir. Çapraz geçerlilikte olduğu gibi model bütün veri kümesi üzerine kurulur. Daha sonra en az 200, bazen binin üzerinde olmak üzere çok fazla sayıda öğrenim kümesi tekrarlı örneklemelerle veri kümesinden oluşturularak hata oranı hesaplanır.

Model kuruluşu çalışmalarının sonucuna bağlı olarak, aynı teknikle farklı parametrelerin kullanıldığı veya başka algoritma ve araçların denendiği değişik modeller kurulabilir. Model kuruluş çalışmalarına başlamazdan önce, imkansız olmasa da hangi tekniğin en uygun olduğuna karar verebilmek güçtür. Bu nedenle farklı modeller kurarak, doğruluk derecelerine göre en uygun modeli bulmak üzere sayısız deneme yapılmasında yarar bulunmaktadır.

Özellikle sınıflama problemleri için kurulan modellerin doğruluk derecelerinin değerlendirilmesinde basit ancak faydalı bir araç olan risk (yakınsaklık) matrisi kullanılmaktadır. Aşağıda bir örneği görülen bu matriste sütunlarda fiili, satırlarda ise tahmini sınıflama değerleri yer almaktadır. Örneğin fiilen B sınıfına ait olması gereken 46 elemanın, kurulan model tarafından 2'sinin A, 38'inin B, 6'sının ise C olarak sınıflandırıldığı Çizelge 2.3'deki matriste görülmektedir [12].

Çizelge 2.3. Risk matrisi

	Gerçek sınıf		
Tahmin edilmiş sınıf	A Sınıfı	B Sınıfı	C Sınıfı
A Sınıfı	45	2	3
B Sınıfı	10	38	2
C Sınıfı	4	6	40

Önemli diğer bir değerlendirme kriteri modelin anlaşılabilirliğidir. Bazı uygulamalarda doğruluk oranlarındaki küçük artışlar çok önemli olsa da, birçok işletme uygulamasında ilgili kararın niçin verildiğinin yorumlanabilmesi çok daha büyük önem taşıyabilir. Çok ender olarak yorumlanamayacak kadar karmaşıklıksalar da, genel olarak karar ağacı ve kural temelli sistemler model tahmininin altında yatan nedenleri çok iyi ortaya koyabilmektedir.

Kurulan modelin değerinin belirlenmesinde kullanılan diğer bir ölçü, model tarafından önerilen uygulamadan elde edilecek kazancın bu uygulamanın

gerçekleştirilmesi için katlanılacak maliyete bölünmesi ile edilecek olan yatırımın geri dönüş (return on investment) oranıdır.

Kurulan modelin doğruluk derecesi ne denli yüksek olursa olsun, gerçek dünyayı tam anlamı ile modellediğini garanti edebilmek mümkün değildir. Yapılan testler sonucunda geçerli bir modelin doğru olmamasındaki başlıca nedenler, model kuruluşunda kabul edilen varsayımlar ve modelde kullanılan verilerin doğru olmamasıdır. Örneğin modelin kurulması sırasında varsayılan enflasyon oranının zaman içerisinde değişmesi, bireyin satın alma davranışını belirgin olarak etkileyecektir.

2.7.4. Modelin kullanılması

Kurulan ve geçerliliği kabul edilen model doğrudan bir uygulama olabileceği gibi, bir başka uygulamanın alt parçası olarak kullanılabilir. Kurulan modeller risk analizi, kredi değerlendirme, dolandırıcılık tespiti gibi işletme uygulamalarında doğrudan kullanılabilmesi gibi, promosyon planlaması simülasyonuna entegre edilebilir veya tahmini envanter düzeyleri yeniden sipariş noktasının altına düştüğünde, otomatik olarak sipariş verilmesini sağlayacak bir uygulamanın içine gömülebilir [13].

2.7.5. Modelin izlenmesi

Zaman içerisinde bütün sistemlerin özelliklerinde ve dolayısıyla ürettikleri verilerde ortaya çıkan değişiklikler, kurulan modellerin sürekli olarak izlenmesini ve gerekiyorsa yeniden düzenlenmesini gerektirecektir. Tahmin edilen ve gözlenen değişkenler arasındaki farklılığı gösteren grafikler model sonuçlarının izlenmesinde kullanılan yararlı bir yöntemdir [13].

2.8. Veri Madenciliği Teknikleri

Veri madenciliği tekniklerini işlevlerine göre 3 temel grupta toplanır:

- Sınıflama (Classification),
- Kümeleme (Clustering),
- Birliktelik kuralları ve sıralı örüntüler (Association rules and sequential patterns).

2.8.1. Sınıflama

Sınıflama, verinin önceden belirlenen çıktılarına uygun olarak ayrıştırılmasını sağlayan bir tekniktir. Çıktılar, önceden bilindiği için sınıflama, veri kümesini denetimli (supervised) olarak öğrenir [16].

Örneğin; A finans hizmetleri şirketi; müşterilerinin yeni bir yatırım fırsatıyla ilgilenip ilgilenmediğini öğrenmek istemektedir. Daha önceden benzer bir ürün satmıştır ve geçmiş veriler hangi müşterilerin önceki teklife cevap verdiğini göstermektedir. Amaç; bu teklife cevap veren müşterilerin özelliklerini belirlemek ve böylece pazarlama ve satış çalışmalarını daha etkin yürütmektir.

Müşteri kayıtlarında müşterinin önceki teklife cevap verip vermediğini gösteren “evet”/ “hayır” şeklinde bir alan bulunur. Bu alan “hedef ” ya da “bağımlı” değişken (target variable) olarak adlandırılır. Amaç, müşterilerin diğer niteliklerinin (gelir düzeyi, iş türü, yaş, medeni durum, kaç yıldır müşteri olduğu, satın aldığı diğer ürün ve yatırım türleri) hedef değişken üzerindeki etkilerini analiz etmektir. Analizde yer alan diğer nitelikler “bağımsız” yada “ tahminci” değişken adını alır.

Temel sınıflama algoritmaları aşağıdadır:

- Diskriminant analizi,
- Naive Bayes,

- Karar ağaçları,
- Sinir ağları,
- Kaba kümeler,
- Genetik algoritma,
- Regresyon analizi.

Diskriminant analizi

Diskriminant analizi, bir dizi gözlemi önceden tanımlanmış sınıflara atayan bir tekniktir. Model, ait oldukları sınıf bilinen gözlem kümesi üzerine kurulur. Bu küme, öğrenme kümesi olarak da adlandırılır. Öğrenme kümesine dayalı olarak, diskriminant fonksiyonu olarak bilinen doğrusal fonksiyonların bir kümesi oluşturulur. Diskriminant fonksiyonu, yeni gözlemlerin ait olduğu sınıfı belirlemek için kullanılır. Yeni bir gözlem söz konusu olduğunda için tüm diskriminant fonksiyonları hesaplanır ve yeni gözlem diskriminant fonksiyonunun değerinin en yüksek olduğu sınıfa atanır [17].

Naive Bayes

Naive Bayes, hedef değişkenle bağımsız değişkenler arasındaki ilişkiyi analiz eden tahminci ve tanımlayıcı bir sınıflama algoritmasıdır [9].

Naive Bayes, sürekli veri ile çalışmaz. Bu nedenle sürekli değerleri içeren bağımlı ya da bağımsız değişkenler kategorik hale getirilmelidir. Örneğin; bağımsız değişkenlerden biri yaş ise, sürekli değerler “<20” “21-30”, “31-40” gibi yaş aralıklarına dönüştürülmelidir.

Naive Bayes, modelin öğrenilmesi esnasında , her çıktının öğrenme kümesinde kaç kere meydana geldiğini hesaplar. Bulunan bu değer, öncelikli olasılık olarak adlandırılır. Örneğin; bir banka kredi kartı başvurularını “iyi” ve “kötü” risk sınıflarında gruptandırmak istemektedir. İyi risk çıktısı toplam 5 vaka içinde 2 kere

meydana geldiyse iyi risk için öncelikli olasılık 0,4'tür. Bu durum, “Kredi kartı için başvuran biri hakkında hiçbir şey bilinmiyorsa, bu kişi 0,4 olasılıkla iyi risk grubundadır” olarak yorumlanır Naive Bayes aynı zamanda her bağımsız değişken / bağımlı değişken kombinasyonunun meydana gelme sıklığını bulur. Bu sıklıklar öncelikli olasılıklarla birleştirilmek suretiyle tahminde kullanılır.

Karar ağaçları

Karar ağaçları, yaygın olarak kullanılan sınıflama algoritmalarından biridir. Karar ağacı yapılarında, her düğüm bir nitelik üzerinde gerçekleştirilen testi, her dal bu testin çıktısını, her yaprak düğüm ise sınıfları temsil eder. En üstteki düğüm kök düğüm olarak adlandırılır. Karar ağaçları, kök düğümden yaprak düğüme doğru çalışır [18].

En yaygın kullanılan karar ağacı algoritmaları;

- CHAID (Chi-Squared Automatic Interaction Detector , Kass 1980),
- C&RT (Classification and Regression Trees, Breiman ve Friedman, 1984),
- ID3 (Induction of Decision Trees, Quinlan, 1986),
- C4.5 (Quinlan, 1993).

Sinir ağları

Sinir ağları, tanımlayıcı ve tahminci veri madenciliği algoritmalarındandır. İnsan beyninin fizyolojisini taklit ederler. Komplike ve belirsiz veriden bilgi üretirler. Keşfettikleri örüntü ve trendler, insanlar yada bilgisayarlarca kolay keşfedilemez. Bu tür karmaşık problemlerde birbirleriyle etkileşimli yüzlerce değişken bulunur [19].

Bu teknik, veritabanındaki örüntüleri, sınıflandırma ve tahminde kullanılmak üzere genelleştirir. Sinir ağları algoritmaları sayısal veriler üzerinde çalışırlar.

Kaba kümeler

Kaba küme teorisi 1970’li yıllarda Pawlak tarafından geliştirilmiştir. Kaba küme teorisinde bir yaklaştırma uzayı ve bir kümenin alt ve üst yaklaşımları vardır. Yaklaştırma uzayı, ilgilenilen alanı ayrı kategorilerde sınıflandırır. Alt yakınlştırma belirli bir altkümeye ait olduđu kesin olarak bilinen nesnelerin tanımıdır. Üst yakınlştırma ise alt kümeye ait olması olası nesnelerin tanımıdır. Alt ve üst sınırlar arasında tanımlanan herhangi bir nesne ise “kaba küme” olarak adlandırılır [20].

Genetik algoritma

Genetik algoritma, Darwin tarafından geliştirilen “evrim teorisini”ne dayalıdır. Algoritma ilk olarak popölasyon adı verilen bir çözüm kümesi (öğrenme veri kümesi) ile başlatılır. Bir popölasyondan alınan sonuçlar bir öncekinden daha iyi olacağı beklenen yeni bir popölasyon oluşturmak için kullanılır. Evrim süreci (yeni popölasyonlar yaratma iterasyonu) tamamlandığında bağımlılık kuralları veya sınıf modelleri ortaya konmuş olur [21].

Regresyon analizi

Regresyon analizi, bir yada daha fazla bağımsız değışken ile hedef değışken arasındaki ilişkiyi matematiksel olarak modelleyen bir yöntemdir.

Veri madenciliğinde yaygın olarak kullanılan regresyon modellerinden doğrusal regresyonda tahmin edilecek olan hedef değışken sürekli deđer alırken; lojistik regresyonda hedef değışken kesikli bir deđer almaktadır. Doğrusal regresyonda hedef değışkenin deđerı; lojistik regresyonda ise hedef değışkenin alabileceđi deđerlerden birinin gerçekleşme olasılığı tahmin edilmektedir [22].

2.8.2. Kümeleme

Kümeleme, verideki benzer kayıtların gruplandırılmasını sağlayan bir tekniktir. Kümelemede, genellikle k-ortalamalar algoritması ya da Kohonen şebekesi gibi istatistiksel yöntemler kullanılmaktadır. Hangi yöntem kullanılırsa kullanılsın süreç aynı şekilde işler. Her kayıt var olan kümelerle karşılaştırılır. Bir kayıt kendisine en yakın kümeye atanır ve bu kümeyi tanımlayan değeri değiştirir. Optimum çözüm bulununcaya kadar kayıtlar yeniden atanır ve küme merkezleri ayarlanır [22].

En yaygın kullanılan kümeleme algoritması “k ortalamalar algortması”dır.

k ortalamalar algoritması

Diğer kümeleme teknikleri ile karşılaştırıldığında k ortalamalar algoritması (k means, simple k means) büyük veritabanlarının kümelenmesinde oldukça etkin bir algoritmadır. Yeni bir vaka ortaya çıktığında; algoritma tüm veriyi inceleyerek buna en çok benzeyen vakaların bir altkümesini oluşturur ve onları çıktığı tahmin etmek için kullanır [23].

Bu algoritmanın adımları aşağıdadır:

1. Veri seti rassal olarak k adet başlangıç kümesine ayrılır.
2. Veri setinde yer alan örnekler; merkezi kendisine en yakın olan kümeye atanır.
3. Her atamanın sonunda küme merkezi (ortalama) yeniden hesaplanır.
4. Veri setindeki tüm örneklerin ataması yapılmaya kadar 2. ve 3. adımlar tekrarlanır.

Yeni bir vakanın ait olduğu kümeyi belirlemek için algoritma yeni vakanın öğrenme verisindeki her bir vakadan uzaklığını hesaplar. k değerinin ve uzaklık ölçüsünün modelin kalitesi üzerinde büyük etkisi vardır bu nedenle onları dikkatle seçmek çok önemlidir.

k ortalamalar algoritması oldukça etkin bir algoritma olmakla birlikte; sadece nümerik veri ile çalışır fakat veri madenciliği uygulamaları sıklıkla kategorik verileri de içermektedir. k ortalamalar algoritmasının geliştirilmesi ile elde edilen k modlar algoritması ise kategorik veriler üzerinde çalışabilen bir algoritmadır. k ortalamalar algoritmasında küme merkezleri, küme ortalaması alınarak hesaplanırken; k modlar algoritmasında küme merkezlerinin belirlenmesinde kümede en sık tekrarlanan değerler (mod) dikkate alınır [24].

2.8.3. Birliktelik kuralları ve ardışık örüntüler

Birliktelik analizi, bir veri kümesindeki kayıtlar arasındaki bağlantıları arayan denetimsiz (unsupervised) veri madenciliği şeklidir. Birliktelik analizi çoğu zaman perakende sektöründe süpermarket müşterilerinin satın alma davranışlarını ortaya koymak için kullanıldığından “pazar sepeti analizi” olarak da adlandırılır [25].

Birliktelik kurallarına ait örnekler aşağıda yer almaktadır:

- “Müşteriler bira satın aldıklarında %75 olasılıkla çocuk bezi de satın alırlar.”
- “Düşük yağlı peynir ve yağsız süt alan müşteriler %85 olasılıkla diyet süt alırlar.”

Ardışık analiz ise birbiriyle ilişkisi olan ancak birbirini izleyen dönemlerde gerçekleşen ilişkilerin tanımlanmasında kullanılır. Aşağıda ardışık analize ait örnekler yer almaktadır.

- “Çadır alan müşterilerin %10’u bir ay içerisinde sırt çantası almaktadır.”
- “A hissesi %15 artarsa üç gün içinde B hissesi %60 olasılıkla artacaktır.”

2.9. Literatürde Veri Madenciliği

Veri madenciliği 1990'lı yıllarda ortaya çıkmıştır. Bir online veritabanı olan Science Direct'te 1960'tan günümüze kadar bir literatür taraması yapıldığında veri madenciliği ile ilgili 1240'a yakın makale olduğu görülmektedir. Veri madenciliğinin özellikle 2000 yılından bu yana büyük bir gelişme gösterdiği göze çarpmaktadır. Aşağıda 2000 – 2006 tarihleri arasında veri madenciliği konusunda farklı alanlarda gerçekleştirilen uygulama örnekleri yer almaktadır.

Boginski ve Butensko (2006), çapraz korelasyon ve kümeleme tekniklerini kullanarak hisse senedi piyasalarının yapısal özelliklerini ortaya koymuşlardır [26]. Jiao, Zhang ve Helande (2006), Kansai haritalama tekniği ile bir karar destek sistemi tasarlamışlardır [27]. Jeng, Chen ve Liang (2006), genetik algoritma ile biyolojik sistemlerin kinetik parametrelerini belirlemişlerdir [28].

Facca ve Lanzi (2005), web kütüklerinde tutulan verileri analiz etmek için makine öğrenme algoritmalarını kullanmışlardır [29]. Hong, Park, Jon ve Rho (2005), veri madenciliği tekniklerini kullanarak bir tedarikçi seçim modeli önermişlerdir [30]. Huang, Chen ve Wu (2005), kümeleme tekniklerini kullanarak dağıtım merkezleri için bir sipariş yönetim sistemi geliştirmişlerdir [31]. Cervone, Kafatos ve Singh (2006), veri madenciliği tekniklerini kullanarak bir deprem erken uyarı sistemi geliştirmişlerdir [32].

Crespo ve Weber (2004), bulanık kümelemeye dayalı veri madenciliği metodolojisi geliştirmişlerdir [33]. Lee, Chiu, Chou ve Lu (2004), sınıflama ve regresyon tekniklerini kullanarak bir kredi derecelendirme uygulaması geliştirmişlerdir [34]. Bellazi, Larizza ve Magni (2004), veri madenciliği tekniklerini kullanarak hemodiyaliz servislerinin kalite ölçümünü geliştirmişlerdir [35]. Last ve Kandel (2004), karar ağacı algoritmasını kullanarak yarı iletken endüstrisindeki bir fabrikada üretim planlama uygulaması geliştirmişlerdir [36].

Lian, Lai, Lin ve Yao (2002), veri madenciliği tekniklerini montaj hattı uygulamalarında kullanmışlardır [37].

Lin ve McClean (2001), şirket iflaslarının tahminine yönelik veri madenciliği yaklaşımı geliştirmişlerdir [38]. Caskey (2001), genetik algoritma ve sinir ağları teknikleri ile bir fabrikadaki çalışma koşullarını ortaya koymuş ve bu koşulları iyileştirici işletme stratejileri önermiştir [39]. Cox ve Lewis (2002), çelik endüstrisindeki bir fabrikada yaptıkları uygulamada yapay sinir ağları yöntemini kullanarak ürünün istenilen kalite standartlarını sağlaması için gerekli girdi miktarını saptamışlardır [40].

Sforna (2000), veri madenciliği teknikleri ile bir elektrik şirketinin müşteri veritabanını analiz etmiştir [41]. Kusiak (2000), elektronik endüstrisinde yaptığı uygulamada imalat hatalarının tahmini için karar kurallarını kullanmıştır [42].

3. MÜŞTERİ İLİŞKİLERİ YÖNETİMİ VE VERİ MADENCİLİĞİ

Değişim, günümüz iş dünyasını derinden etkileyen kavramların başında gelmektedir. Değişimin yönlendirdiği ve değişim odaklı bir dünyada, değişime uyum sağlama gücü ve yeteneği olan işletmeler ve organizasyonlar varlıklarını sürdürüp, ayakta kalabileceklerdir.

Bu bağlamda değişim, müşteri satın alma davranışlarını ve beklentilerini, büyük ölçüde değiştirmiştir. Değişen müşteri davranışlarını, işletme amaç ve hedeflerine uygun olarak yönetebilmek; değişim yönetimi ve müşteri ilişkileri yönetimi kavramlarını anlayabilmek ve doğru bir şekilde yorumlayabilmek ile mümkün olabilecektir. İşletmelerde başarılı bir müşteri ilişkileri yönetimiyle, küresel bir dünyada ve giderek artan rekabetçi piyasalarda, işletmeler için yaşamsal önem taşıyan, müşteri için değer yaratmak, müşteri sadakati sağlamak ve bu konularda kurumsallaşmayı gerçekleştirmek mümkün olabilecektir [43].

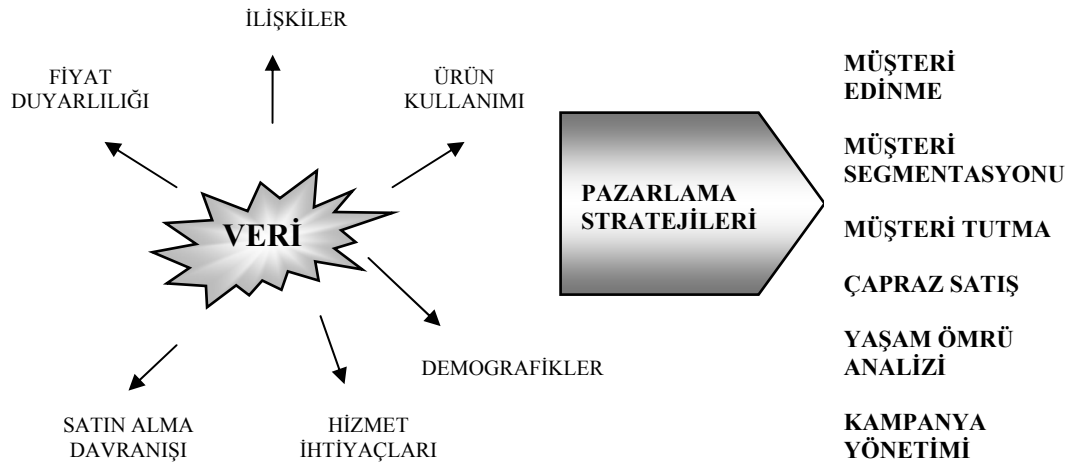
Piyasalarda ortaya çıkan artan rekabet baskısı, birbirine giderek benzeyen ürünler, ürün ve hizmet farklılaştırmasının giderek zorlaşması ve kar oranlarının azalması, işletmeleri pazarlama alanında yeniden yapılanmaya yöneltmiştir.

İşletmeler açısından uzun dönemli müşteri bağlılığı, yeni rekabet alanında önemli bir değişimi beraberinde getirmiştir. Sanayi devrimi ile ortaya çıkan kitle üretim ve kitle pazarlama, yerini merkezde müşterinin yer aldığı yaklaşımlara bırakmıştır. Eski “tasarla-yap-sat” modelinin (ürün odaklı yaklaşım) yerini “sat-yap-yeniden tasarla” modeli (müşteri odaklı yaklaşım) almıştır. Bu bağlamda müşteri sadakatini gerçekleştirmek ve sürekli kılmak için çeşitli yöntem ve tekniklerin kullanıldığı görülmektedir. Bu yöntemlerden en önemli olanı, müşteriyi merkeze alan ve müşteri memnuniyetinin ötesine geçmeyi amaçlayan müşteri ilişkileri yönetimi yaklaşımıdır. Bu yaklaşım aynı zamanda yeni pazarlama anlayışını da ortaya koymaktadır [44].

Müşteri ilişkileri yönetimi, geleneksel pazarlama anlayışından çok farklı özellikler taşımaktadır. Geleneksel pazarlama anlayışında kitlesel üretimin etkisiyle kitlesel

pazarlama anlayışı geçerli durumda idi. Oysa müşteri ilişkileri yönetiminde pazar payı kavramından, müşteri payı kavramına geçiş söz konusu olmaktadır. Müşteri payı aynı müşteriye birden fazla ürün satabilmeyi ve müşteriye aktif ve sadık bir müşteriye dönüştürmeyi amaçlamaktadır. Bu durum ise tek yönlü pazarlamadan karşılıklı ilişkiye dayalı pazarlamaya geçişi ifade etmektedir. Günümüzde kitlesel pazarlama yöntemleri, giderek müşteriye hitap eden kişisel pazarlama yöntemlerine yönelmektedir.

Diğer taraftan yazılım ve donanım teknolojilerindeki gelişmeler, işletmelerin çok miktarda veriyi saklayabilmesini ve işleyerek anlamlı bilgilere dönüştürmesini mümkün hale getirmiştir. Bugün, işletmeler satın alma davranışından demografik özelliklere kadar müşterileri ilgili pek çok detayı veritabanlarında tutmaktadır. Veri madenciliği tekniklerini kullanarak bu veriler içerisindeki anlamlı ve gizli örüntülerin ortaya çıkarılması mümkün olmaktadır. Şirketlerin veritabanlarında tuttukları veriler ve bu verilerin hangi amaçlarla kullanıldığı Şekil 3.1’de gösterilmektedir.

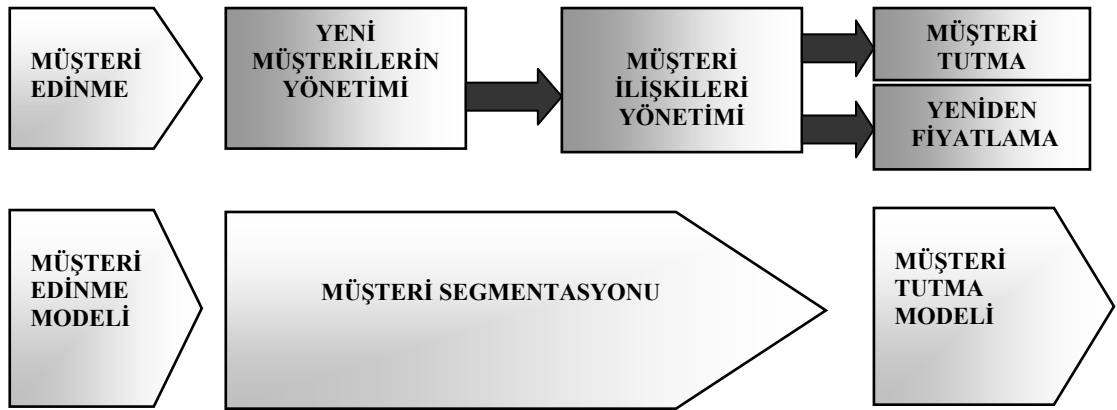


Şekil 3.1. Müşteri verilerinin kullanım amaçları

Veri madenciliği sonuçları müşteri edinme, müşteri segmentasyonu, müşteriye elde tutma, ayrılma eğilimi gösteren müşteri profilini ortaya koyma, müşteri değerlendirme, kredi derecelendirme ve pazar sepeti analizi gibi pek çok müşteri odaklı uygulamaya girdi teşkil etmektedir [45].

3.1. Müşteri Edinme

Yeni müşteriler edinme, şirketler için mevcut müşterileri elde tutma kadar kritik bir fonksiyondur. Müşteri edinme faaliyetleri ile diğer pazarlama fonksiyonları arasındaki ilişki Şekil 3.2’de gösterilmektedir.



Şekil 3.2. Müşteri edinme faaliyetleri ve diğer pazarlama fonksiyonları ilişkisi

Şirketler pazar paylarını ancak yeni müşteriler edinerek artırabilirler [46]. Burada dikkat edilmesi gereken nokta, şirkete kazandırılması gereken müşteri profilinin doğru saptanması ve bu profile uygun eylem planları hazırlanmasıdır. Örneğin; çocukları olmayan bir aileye oyuncak katalogu göndermek mantıklı bir yaklaşım değildir. Yeni müşteri edinme maliyeti astronomik olduğu için müşteri edinme çalışmalarına doğru hedef kitleyi saptayarak başlamak ve en az maliyetli yöntemi kullanmak gerekir.

3.2. Müşteri Segmentasyonu

Müşteri segmentasyonu, müşterilerin istek, ihtiyaç ve davranışlarının esas alınarak pazarın benzer karakteristikler taşıyan müşteri gruplarına bölünmesidir. Müşteri segmentasyonu ile şirketler geniş ve heterojen pazarları küçük gruplara ayırarak, müşteriye bireysel istek ve ihtiyaçlarına cevap veren ürün ve hizmet sunar, müşteri grupları bazında ürün önerileri oluşturur ve pazarlama programları geliştirirler [47].

Segmentasyonda asıl amaç, “çok karlı” müşterileri “az karlı” müşterilerden ayırmaktır. Karlı grupların karlı olmayanlardan ayrılması, pazarlama faaliyetlerinin müşteri grupları bazında farklılaştırılmasını sağlar [48].

Müşteri segmentasyonu, şirket açısından müşterilerin değerini grupları itibariyle ortaya koyarak, pazarlama faaliyetlerinin uygulanmasında en değerli müşterilere öncelik verilmesini sağlar ve böylece mevcut müşterilerden elde edilen geliri maksimize eder. Ayrıca, doğru kitleye doğru mesajın gönderilmesini sağlayarak, kampanya yönetim maliyetlerini azaltır [49].

Müşteriler; coğrafi, demografik, psikografik ve davranışa dayalı özelliklerine göre gruplandırılırlar [50].

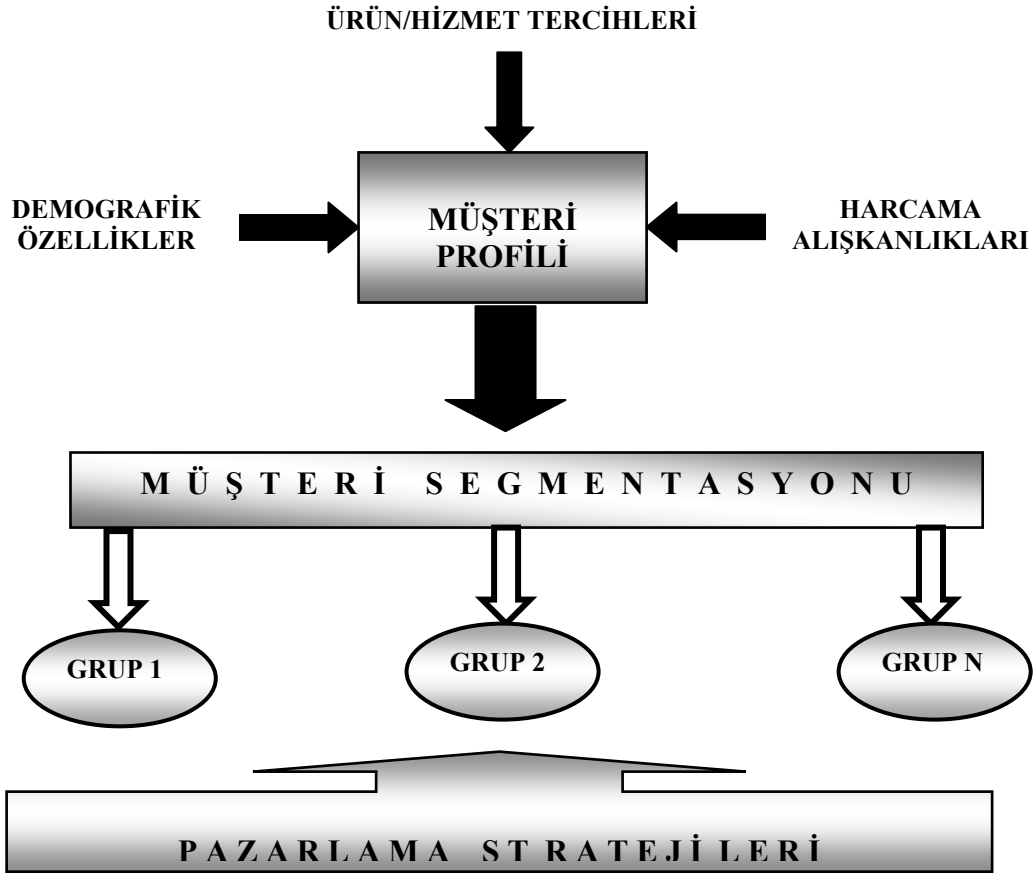
Coğrafi segmentasyon, pazarı şehir, bölge, ülke gibi coğrafik birimlere ayırır. Günümüzde birçok şirket; reklam, promosyon ve satış faaliyetlerini farklı özellikler gösteren coğrafik birimlerin ihtiyaçları ile uyumlu hale getirmeye çalışmaktadır.

Demografik segmentasyon; müşterileri yaş, cinsiyet, aile büyüklüğü, gelir, meslek, tahsil gibi kriterleri baz alarak gruplandırır.

Psikografik segmentasyon; müşterileri sosyal sınıf, yaşam tarzı ve kişilik özelliklerine dayalı olarak gruplandırır.

Davranışa dayalı segmentasyon ise müşterileri ürünü kullanma, kendisine sunulan önerilere cevap verme, satın alma sıklığı, satın alma miktarı gibi kriterlere dayalı olarak gruplandırmaktadır.

Şekil 3.3’de gösterildiği gibi müşteri segmentasyonunda ilk adım müşteri profilinin belirlenmesidir.



Şekil 3.3. Müşteri segmentasyonu

Müşteri profili belirleme sürecinde, müşterilerle ilgili şu sorulara cevap aranır:

- Temel demografik özellikleri nelerdir?
- Harcama alışkanlıkları nasıldır?
- Hangi ürün ya da hizmetleri satın alırlar?
- Müşteri hangi sıklıkla alışveriş yapar?
- Niçin sizin ürün ya da hizmetlerinizi tercih ederler?
- En değerli müşterilerinizle benzer nitelikli müşterileri nasıl elde edersiniz?

3.3. Ayrılan Müşteri Analizi ve Müşteri Tutma

Rekabetin çok yoğun olarak yaşandığı günümüzde, birçok müşteri hâlihazırda ürün ve hizmet aldığı şirketi bırakarak kendi tercih ve beklentilerine daha uygun olan rakip şirketleri tercih etmektedir.

Yeni bir müşteri edinme maliyeti, mevcut müşterileri elde tutma maliyetinin yaklaşık 5 katı, memnun olmayan müşteriyi geri kazanmanın maliyeti mevcut müşteriyi elde tutmanın yaklaşık 10 katıdır [51]. Mevcut müşterilerden sağlanan brüt gelir, yeni müşterilerden kazanılanın yaklaşık 2 katıdır. Müşteri tutmada %5 oranında bir iyileşme karı %25 artırmaktadır [52]. Bu nedenle şirketler, mevcut müşterilerini elde tutmaya yönelik bir takım sadakat programları geliştirirler.

Satış noktaları çoğaldıkça ve teknoloji ilerledikçe; müşteri sadakatini korumak daha zor hale gelmiştir. Ancak dikkatli bir planlama ile şirket değerini korumak ve müşteriyi elde tutmak mümkündür. Şirketlerin, müşteri sadakatini korumak için geliştirdikleri yöntemlerden biri de sadakat kartı uygulamalarıdır. Sadakat kartı başvuruları esnasında müşteriler; isim, meslek, cinsiyet, tahsil, gelir, doğum tarihi, adres, e-posta bilgilerinin yer aldığı formları doldururlar. Müşterinin yaptığı her alışveriş sonunda mağaza, sadakat kartını manyetik okuyucudan geçirir ve böylece müşterinin hangi ürünleri, haftanın hangi günlerinde ve günün hangi saatlerinde satın aldığı, ne kadar harcama yaptığı verisini elde eder. Başvuru formunun doldurulması sırasında elde edilen müşteri karakteristikleri ile kart üzerinde toplanan satın alma davranışı verileri birleştirilerek önemli bilgiler elde edilir ve bu bilgiler müşteriye özel indirim ve promosyonların düzenlenmesinden mağaza içi yerleşime kadar çok çeşitli alanlarda kullanılır.

3.4. Çapraz Satış

Çapraz satış, genel satın alma eğilimini analiz ederek müşteriye birbiriyle ilişkili ürün veya hizmetleri satma stratejisi olarak tanımlanabilir. Örneğin; printer alan bir müşteri büyük olasılıkla bu printer'da kullanmak üzere kağıt da alacaktır. Diğer bir

ifadeyle bu iki farklı ürün arasında bir bağ kurulacaktır. Çapraz satışın amacı müşterinin şirketle olan bağını güçlendirmek ve rakip şirketlere gidişini engellemektir. Çapraz satış birçok sektörde kullanılan etkin ve proaktif bir yaklaşım olup hem müşteri sadakatinin hem de şirket gelirlerinin artmasını sağlar. Bir şirket müşterilerinin genel satın alma alışkanlıklarını bilirse ve müşterilerine satın aldıkları ürünlerle ilişkili ek ürünleri belirli indirim paketleri içerisinde sunarsa iyi bir müşteri tutma politikası izlemiş olacaktır [53].

3.5. Müşteri Yaşam Ömrü Değeri

Müşteri yaşam ömrü değeri, müşterinin bugün, yarın ve ömrü boyunca aynı kaynaktan (marka) yapacağı alışverişler sonucu ortaya çıkacak net değeridir. Müşteri yaşam ömrü değeri analizlerinde dikkate alınan kriterler; müşterinin harcama tutarı ve alışveriş periyodudur [54]. Müşteri yaşam ömrü değeri analizi ile şirket en değerli müşterilerinin kimler olduğunu saptar ve gelecekteki değerli müşterilerinin kimler olacağını tahmin eder. Şirketler, böyle bir analizin sonuçlarını kullanarak en karlı müşterilerini, yüksek tutarlarda harcama yapmaya teşvik edecek uygulamalar geliştirebilir.

3.6. Pazarlama İletişimi ve Tepki Modelleme

Pazarlama iletişimi, doğru müşteriye, doğru mesajın, doğru kanallar kullanılarak iletilmesidir [55].

Pazarlamada kullanılan tipik iletişim kanalları; radyo, televizyon, yazılı basın, doğrudan posta ve e-postadır. Pazarlama iletişiminde televizyon, genellikle büyük bir popülasyona genel bir mesajın gönderildiği durumlarda kullanılmaktadır. Belirli özelliklere sahip bir müşteri grubu hedef kitle olarak belirlendiğinde ise yazılı basın yaygın olarak kullanılan iletişim kanalıdır. Doğrudan postalama kişiye özel bir mesajın gönderilmesini sağlayan bir iletişim kanalıdır ve müşterinin kendisine sunulan tekliflere gösterdiği tepkinin ölçülmesini sağlar. Son olarak, ihmal edilebilir

maliyetlerinden ötürü e-posta son zamanlarda birçok şirket tarafından tercih edilen bir iletişim kanalıdır [56].

Elektronik posta çok ucuz bir pazarlama kanalı olmakla birlikte; etkisi de aynı oranda düşüktür. Ayrıca şirket ve müşteri için zaman alıcıdır. Bu noktada; şirketler tüm müşterilerine e-posta göndermek yerine izinli pazarlama (permission marketing) yöntemini uygulayarak, sadece kendilerinden reklam almak isteyen müşterilerine e-posta gönderirler. Böylece hem uygulama daha başarılı olacak hem de kaynakların etkin kullanımı sağlanacaktır [57].

Seçilen iletişim kanalı ile her müşteriye mesaj göndermek yeterli değildir. İletişimin etkisini ölçmek gerekmektedir. Bu ise “tepki analizi” denilen bir yaklaşımla gerçekleştirilmektedir.

Bir müşteri, pazarlama kampanyalarına aşağıdaki şekillerde tepki (cevap) verebilir:

- Müşteri, şirketin ürün ve hizmetleri ile ilgili daha fazla bilgi edinmek ister. Bu, iyi bir başlangıç noktasıdır ve müşteri ile kurulacak uzun dönemli ilişkilerin bir işaretidir.
- Müşteri şirketin kendisine önerdiği ürünlerden bir ya da birkaçını satın alır.
- Müşteri, şirketin kendisine önerdiği ürünlerden farklı bir ürün satın alır. Lüks bir ürün önerisine karşın müşterinin standart bir ürün satın alması bile iyi bir başlangıç noktasıdır çünkü böylece şirket başka türlü elde edemeyeceği ürün/müşteri kombinasyonu bilgisini elde etmiş olur.
- Müşteri sunulan öneriye hiçbir cevap vermez.
- Müşteri, kendisine sunulan öneriye olumsuz cevap verir. Kullanılan iletişim kanalına bağlı olarak müşterinin öneri ile ilgilenip ilgilenmediğini tam olarak belirlemek mümkündür. Her ne kadar ilk bakışta yararlı bir bilgi gibi görünmese de “hayır” cevabı ileri veri analizleri için olumlu cevap kadar değerlidir.

Tepki analizi ile hangi müşterilerin hangi önerilere cevap verdiği saptanarak kampanya yönetimi maliyetleri azaltılabilir. Örneğin; bir müşteri doğrudan posta yoluyla kendisine ulaştırılan kişisel bakım hizmeti önerilerine cevap vermediyse; daha sonraki uygulamalarda bu tür müşterilere kişisel bakım hizmeti önerilerinin sunulmaması kampanya yönetim maliyetlerini ciddi oranda azaltacaktır.

3.7. Veri Madenciliğinin Pazarlama ve Müşteri İlişkileri Yönetimi Uygulamaları Literatürü

Li, Shue ve Lee (2006), ISP (İnternet Servis Sağlayıcı) endüstrisinde kullanıcıların genel ağ kullanma örüntülerini ortaya çıkarmak ve müşteri ilişkilerini güçlendirecek stratejiler geliştirmek amacıyla nitelik odaklı indirgeme yöntemini (AOI) kullanmışlardır [58].

Hsieh (2004), bir bankanın kredi kartı müşterileri için sinir ağları ve birliktelik kurallarını kullanarak bir davranışsal skorlama modeli yaratmış; müşterileri davranış ve karakteristiklerine göre farklı gruplara ayırarak her grubun karakteristiklerine uygun yönetim stratejileri önermek suretiyle müşteri sadakatini artırmayı hedeflemiştir [47].

Hwang, Jung ve Suh (2004), müşterinin geçmişte yarattığı potansiyel değeri ve rakip şirketi tercih etme olasılığını dikkate alarak, lojistik regresyon, sinir ağları ve karar ağacı tekniklerini kullanmak suretiyle bir müşteri yaşam ömrü modeli geliştirmişlerdir [59].

Kim ve Lee (2002), genetik algoritmaya dayalı çoklu sınıflayıcıları birleştirerek e ticaret müşterilerinin satın alma davranışlarının tahmini için bir model geliştirmişlerdir [60].

Liu ve Shih (2006), AHP ve birliktelik kuralı yöntemlerini birleştiren bir öneri metodolojisi geliştirerek müşterileri yaşam ömrü değerlerine göre segmentlere ayırmışlardır [61].

Chen, Chi ve Hwa (2005), perakende endüstrisinde etkin pazarlama kampanyaları düzenlemek amacıyla davranışsal müşteri verileri madenciliği çalışmasını gerçekleştirmişlerdir [62].

Wang ve Hung (2005), veri madenciliği teknikleri ile bir telekomünikasyon şirketi için müşteri karlılık yönetim sistemini geliştirmişlerdir [63].

Hsu ve Chen (2006), varyans ve entropiye dayalı kümeleme algoritması kullanarak müşteri segmentasyonu ve katalog pazarlamacılığı için bir uygulama geliştirmişlerdir [48].

Jane ve Shi (2006), birliktelik kuralları ve ardışık örüntüler madenciliğini kullanarak müşteri satın alma davranışlarını keşfetmişlerdir [64].

Lee, Chiu ve Chao (2006), sınıflama ve regresyon ağaçları (CART) ile MARS (Multivariate Adaptive Regression Splines) tekniklerinin kredi skorlama uygulamalarındaki performansını karşılaştırmışlardır [34].

Bayram (2001), veri madenciliği tekniklerinden karar ağaçları, kümeleme analizleri ve lojistik regresyon analizlerini kullanarak kablosuz iletişim endüstrisinde müşteri segmentasyonu ve ayrılacak müşterilerin tahmin modelini ortaya koymuştur [1].

Jonker, Piersma ve Poel (2003), CHAID ve Markov karar sürecini kullanarak uzun dönem karlılığı için birleşik müşteri segmentasyonu ve pazarlama politikası çalışması gerçekleştirmişlerdir [54].

4. UYGULAMA

4.1. Problem Tanımı

Bu tezin amacı, İstanbul yolu üzerinde bulunan bir kozmetik mağazasının sattığı markalardan birinin müşterilerini gruplandırarak, her grubun özelliklerine uygun pazarlama stratejileri geliştirmek ve karlı müşterileri elde tutmaya yönelik bir eylem planı hazırlamaktır.

Mağaza bünyesindeki tüm markalar mevcut müşterilerini elde tutmak için;

- Ücretsiz cilt bakımı yapmakta,
- Belirli bir tutarın üzerinde alışveriş yapan veya belli sayıda ürün alan müşterilerine hediye ürün vermekte,
- İndirim günleri düzenlemektedir.

Ücretsiz cilt bakımları ile ürün tanıtımları uygulamalı olarak gerçekleştirilmekte ve bu şekilde mağazaya çekilen müşterinin alışveriş yapması hedeflenmektedir. Alan uzmanından alınan bilgiye göre mağazaya bakım amacıyla gelen müşterilerin yaklaşık %70'ine ürün satışı gerçekleştirilmektedir.

Hediye verilirken dikkate alınan kriter, müşterinin verilen hediyein en az 2 katı tutarında alışveriş yapmış olmasıdır.

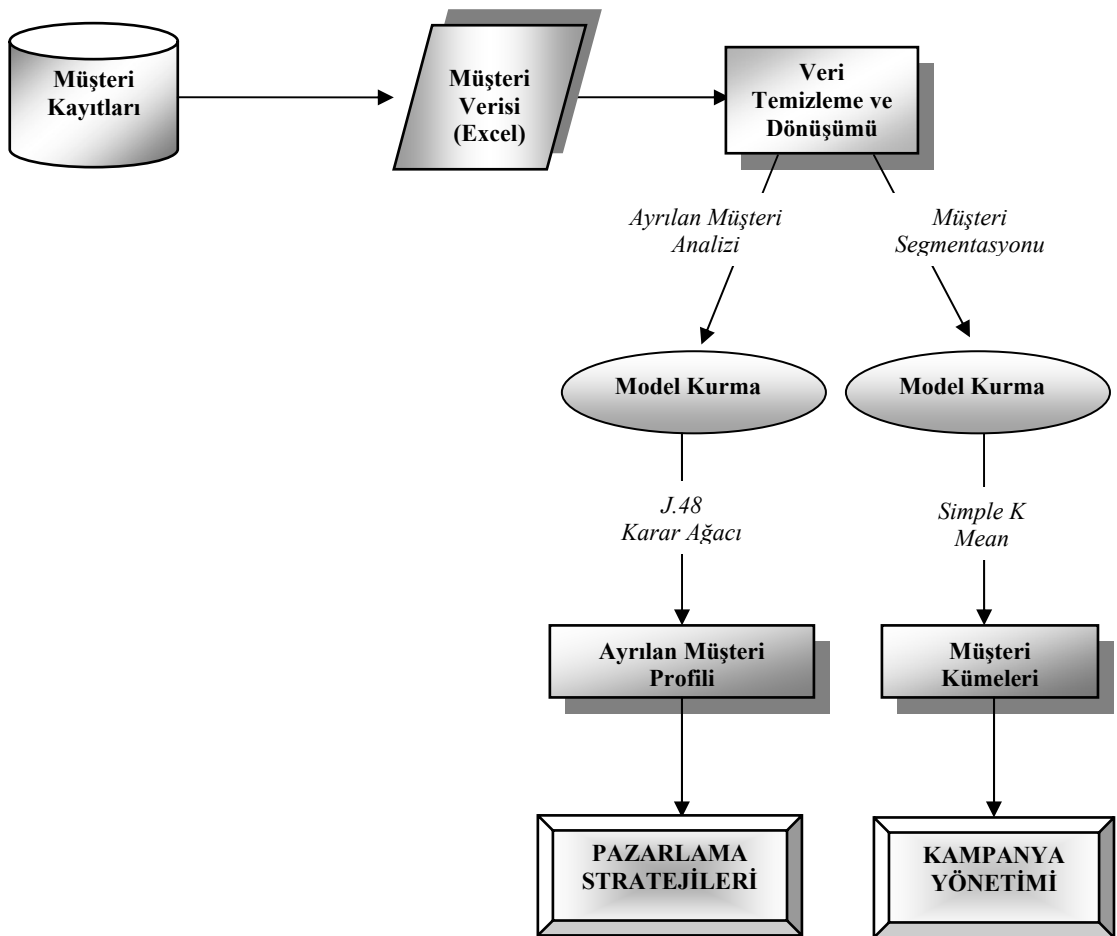
Mevcut müşterileri elde tutmak için gerçekleştirilen tüm bu uygulamalarda müşterilerin tahsil, gelir, medeni durum, meslek gibi demografik özellikleri dikkate alınmamaktadır.

Kozmetik mağazası, ürün ve hizmet tanıtımında kitlesel ve doğrudan pazarlama araçlarını kullanmaktadır. Reklâm ve promosyonlarını yazılı (gazete, dergi) ve görsel (TV) basın aracılığı ile duyurmaktadır. Mağazanın, müşterilerini ücretsiz

bakım günleri, kampanya ve promosyonlar hakkında bilgilendirmek için kullandığı iletişim kanalı telefondur. Tüm müşterilerin e-posta adresleri bilinmesine rağmen e-posta uygulamada yaygın olarak kullanılan bir iletişim kanalı değildir.

4.2. Uygulamada Veri Madenciliği Süreci

Kozmetik markasının veri madenciliği sürecinin aşamaları Şekil 4.1’de özetlenmiştir.



Şekil 4.1. Kozmetik markası veri madenciliği süreci

4.2.1. Veri temizleme

Veri temizleme, ikinci bölümde de ifade edildiği gibi veri madenciliği sürecinin en kritik ve zaman alıcı adımıdır. Müşteri demografik verisi genellikle yüksek oranlarda

kayıp değer içermektedir. Bu tür kategorik verilerdeki kayıp değer problemini çözmek için, *kayıp değeri bir değerle kodlama yöntemi* kullanılır. Bu yöntemle göre, kayıp değerli alana “u” gibi bir değer atanır. Eğer bir niteliğe (attribute) ait verilerin büyük bir çoğunluğu eksikse bu nitelik veritabanından çıkarılmalıdır. Bir niteliği veritabanından çıkarma kararı alırken sadece kayıp değerlerin toplam içindeki büyüklüğünü değil, aynı zamanda bu niteliğin kayıp değer içerme nedenini de dikkate almak gerekir [56].

Nümerik nitelikler için kayıp değer sorununu çözmek biraz daha zordur. Çünkü, çözmeye çalışmak bu niteliğe ait istatistikleri ve veri dağılımını değiştirebilir.

Nümerik veri alanlarındaki kayıp değer sorununu çözmek için çeşitli yöntemler söz konusudur:

- Kayıp değerli alanlara, o niteliğe ait diğer değerlerin ortalaması atanır. Bu yöntem, basit olmakla birlikte verinin dağılımı üzerinde azımsanmayacak bir etkiye sahiptir. Bu nedenle, sadece verinin dağılımını minimal şekilde etkileyen durumlarda kullanılmalıdır.
- Mevcut değerler kullanılarak verinin dağılımı elde edilir ve kayıp değerli alanlara bu dağılıma uygun olarak değer atanır. Bu yöntem, verinin dağılımını çok fazla değiştirmez. Fakat kurulan modelde değer atanan değişken çok önemli ise veri madenciliği sonuçlarını etkileyecektir.

Bu çalışmada kullanılmak üzere kozmetik mağazasından alınan orijinal veri alanları kategorik ve nümerik değerler içermektedir. Mağazadan alınan orijinal veri tablosunun formatı EK-1’de gösterilmiştir.

Müşteri verilerinde kategorik alanlar için kayıp değer söz konusu değilken; HARCAMA TUTARI adlı nümerik alanda kayıp değerler yer almaktadır. Kayıp değer problemini çözmek için “Structured Query Language” (SQL) komutları kullanılmıştır. Nümerik kayıp değerler için yazılan SQL sorgusuyla bilinmeyen

değerlerin yerine bu niteliğe ait bilinen değerlerin ortalaması aşağıda gösterildiği gibi atanmıştır.

update tablo_kozmetik set HARCAMA TUTARI 1=ORT_HRC1 where HARCAMA TUTARI 1 is null;

update tablo_kozmetik set HARCAMA TUTARI 2= ORT_HRC2 where HARCAMA TUTARI 2 is null;

update tablo_kozmetik set HARCAMA TUTARI n= ORT_HRCn where HARCAMA TUTARI n is null;

4.2.2. Veri dönüştürme

Veri temizlemeden sonraki adım veri dönüştürmedir. Veri dönüşümü elde edilen bilginin maksimize edilmesini sağlar.

EK-1’de gösterilen orijinal veri, kayıp değerlerden arındırıldıktan sonra aşağıdaki dönüşümler yapılmıştır.

1. Doğum tarihi alanı yaş aralıklarına dönüştürülmüştür. Bu işlem için aşağıdaki SQL komutları kullanılmıştır.

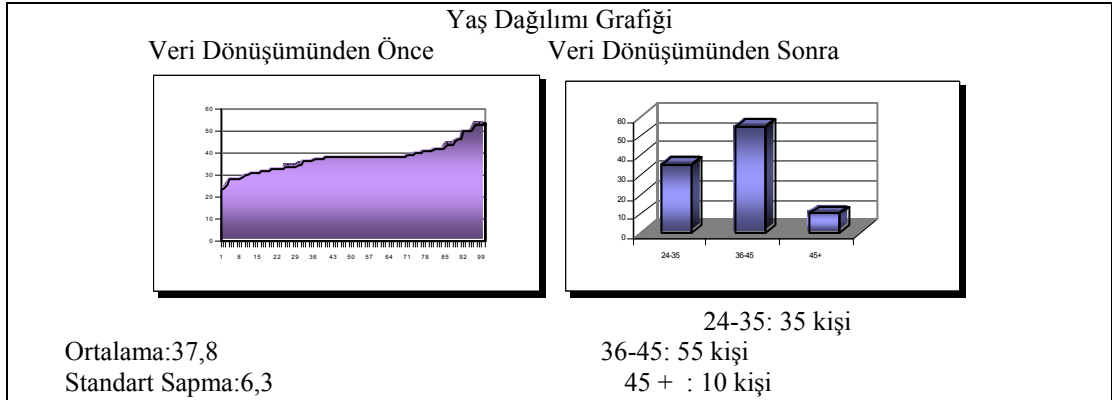
update tablo_kozmetik set YAS= datepart (year,getdate()) - datepart (year,DOG.TAR.)

update tablo_kozmetik set YAS= '24-35' where YAS<=35 and YAS >24

update tablo_kozmetik set YAS= '36-45' where YAS<=45 and YAS >35

update tablo_kozmetik set YAS= '45+' where YAS>45

Şekil 4.2’de veri dönüşümünden önce ve sonraki yaş dağılımı grafikleri gösterilmektedir.



Şekil 4.2. Yaş dağılımı grafiği

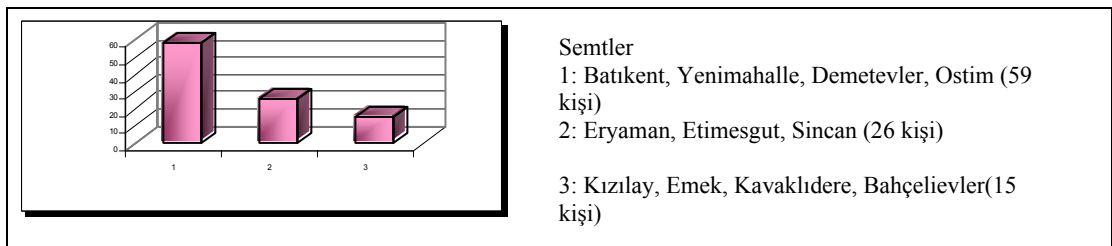
2. Müşterilerinin ikamet ettikleri 11 değişik semt, programın efektif çalışmasını ve doğru sınıflandırmalar yapmasını sağlamak amacıyla 3 temel grupta toplanmış, bunun için aşağıdaki SQL komutları kullanılmıştır.

update tablo_kozmetik set SEMT= '1' where SEMT='Batıkent' OR SEMT='Yenimahalle' OR SEMT='Ostim' OR SEMT='Demetevler'

update tablo_kozmetik set SEMT= '2' where SEMT='Eryaman' OR SEMT='Etimesgut' OR SEMT='Sincan'

update tablo_kozmetik set SEMT= '3' where SEMT='Kolej' OR SEMT='Emek' OR SEMT='Kavaklıdere' OR SEMT='Bahçelievler'

Müşterilerin semtlere göre dağılımı Şekil 4.3'de gösterilmektedir.



Şekil 4.3 Müşterilerin semtlere göre dağılım grafiği

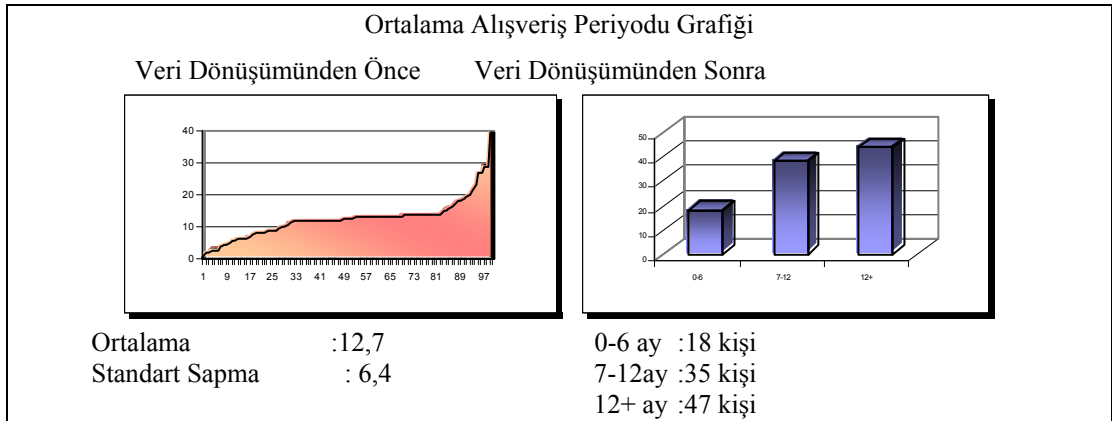
3. Alışveriş tarihi arasında geçen süreler hesaplanmak suretiyle ortalama alışveriş periyotları elde edilmiş; bu periyotlar SQL komutlarıyla aşağıdaki şekilde kategorize edilmiştir:

update tablo_kozmetik set PERİYOT= '0-6' where PERİYOT<=6 and PERİYOT >0

update tablo_kozmetik set PERİYOT= '7-12' where PERİYOT<=12 and PERİYOT >6

update tablo_kozmetik set PERİYOT= '12+' where PERİYOT>12

Şekil 4.4’de veri dönüşümünden önce ve sonraki ortalama alışveriş periyodu grafikleri gösterilmektedir.



Şekil 4.4.Ortalama alışveriş periyodu grafiği

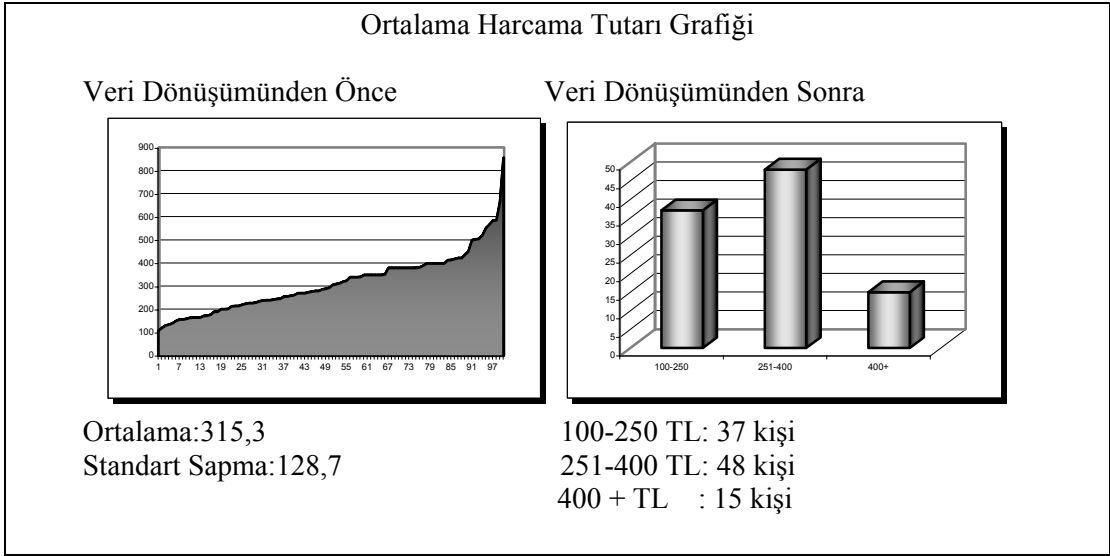
4. Müşterinin her alışverişte yaptığı harcama tutarlarının ortalaması alınarak, ortalama harcama tutarına ulaşılmış bu değer de SQL komutlarıyla aşağıdaki şekilde gruplandırılmıştır:

update tablo_kozmetik set HRC. = '100-250' where HRC. <=250 and HRC. >100

update tablo_kozmetik set HRC. = '251-400' where HRC. <=400 and HRC. >250

update tablo_kozmetik set HRC. = '400+' where HRC. >400

Şekil 4.5’de veri dönüşümünden önce ve sonraki ortalama harcama tutarı grafikleri gösterilmektedir.



Şekil 4.5. Ortalama harcama tutarı grafiği

Çizelge 4.1’de veri dönüşümünden önce ve sonraki nitelik alanları; Çizelge 4.2’de ise kozmetik markasının müşteri verilerine ait istatistikleri gösterilmektedir.

Çizelge 4.1.Nitelik isimleri

Nitelikler	
Veri Dönüşümünden Önce	Veri Dönüşümünden Sonra
Doğum Tarihi	Yaş (24-35, 36-45, 45+)
Alışveriş Tarihi	Alışveriş Periyodu (0-6 ay, 7-12 ay, 12+ ay)
Çalışma Durumu	Çalışma Durumu (Evet, Hayır)
Bakım Durumu	Bakım Durumu (Evet, Hayır, Bazen)
Cilt Tipi (Karma, Kuru, Yağlı)	Cilt Tipi (1,2,3)
Ürün Tipi (Bakım, Bakım/Makyaj, Makyaj, Bakım/Güneş, Tem/Nem, Bakım/Parfüm)	Ürün Tipi (1,2,3,4,5,6)
Semt (Batkent, Ostim, Yenimahalle, Demetevler, Eryaman, Etimesgut, Sincan, Bahçelievler, Emek, Kızılay, Kavaklıdere)	Semt (1,2,3)
Harcama Tutarı	Ortalama Harcama Tutarı (100-250, 251-400, 400+)
	Ayrılan Müşteri (Churn) (Evet, Hayır)

Çizelge 4.2 Veritabanı istatistikleri

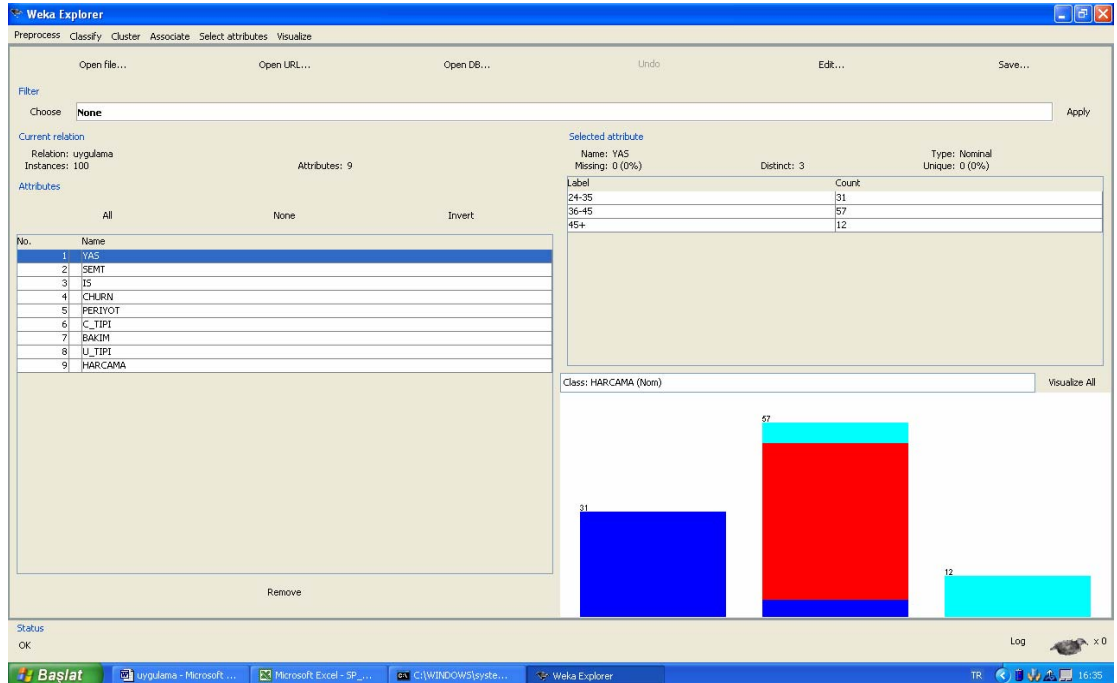
Nitelikler	Özelliği / Grubu	Kişi Sayısı
Yaş	24-35 36-45 45+	35 55 10
Cilt Tipi	1:Karma 2:Kuru 3:Yağlı	70 26 4
Harcama Tutarı (TL)	100-250 251-400 400+	37 48 15
Bakım Yaptırma	E:Evet H:Hayır B:Bazen	21 62 17
Çalışma Durumu	Evet Hayır	83 17
Ürün Tipi	1: Bakım 2: Bakım/Makyaj 3: Makyaj 4: Bakım/Güneş 5: Temizleme/Nemlendirici 6: Bakım/Parfüm	56 16 7 10 8 3
Ayrılan Müşteri	E:Evet H:Hayır	25 75
Alış Veriş Periyodu (Ay)	0-6 7-12 12+	18 35 47
Semt	1: Batıkent, Yenimahalle, Ostim, Demetevler 2: Eryaman, Etimesgut, Sincan 3: Kavaklıdere, Kızılay, Bahçelievler, Emek	59 26 15

4.2.3. Modelleme

Veri temizleme ve dönüşümünden sonraki adım modelleme adıdır. İkinci bölümde de ifade edildiği gibi bu adımda farklı modeller veri kümesi üzerinde denenerek doğruluğu en yüksek olan model seçilir.

Bu uygulamada model kurma aşamasında WEKA paket programı kullanılmıştır. WEKA, Java’da yazılmış, Windows, Linux ve Masintosh gibi farklı işletim sistemleri üzerinde çalışabilen bir programdır [15].

WEKA’da; Preprocess (önişleme), Classify (sınıflama), Cluster (kümeleme), Associate (birliktelik kuralları), Select Attribute (nitelik seçme) ve Visualize (görselleştirme) panelleri bulunmaktadır. WEKA’nın ekran görünümü Resim 4.1’de yer almaktadır.



Resim 4.1.WEKA paket programı ekran görünümü

Preprocess paneli: Bilgi keşfinin başlangıç noktasıdır. Veri dosyaları bu panelden yüklenir. WEKA'ya Excel ve Access ortamından ve herhangi bir URL (Uniform Resource Locator) adresinden dosya yüklemek mümkündür.

Bu tezde kullanılan veriler, kozmetik mağazasından Excel formatında alınmış olup; Excel ortamında yer alan verinin WEKA tarafından işlenebilmesi için ARFF (Attribute Related Format) formatına dönüştürülmesi gerekir. ARFF formatına dönüştürülmüş veri EK-2'de sunulmuştur.

Classifier paneli: Veri kümesi üzerinde WEKA'nın sınıflandırma algoritmalarının çalıştırıldığı paneldir.

Sınıflandırma paneli, karar ağaçlarından kurallara; fonksiyonlardan Bayes ağlarına birçok sınıflandırma algoritmasını içermektedir. Kurulacak modeli belirlemek için farklı algoritmalar veri üzerinde çalıştırılarak doğrulukları karşılaştırılmalıdır.

Bu panelde çapraz geçerleme (cross validation) ya da yüzde ayırma (percentage split) seçeneklerini kullanarak kurulan modellerin doğruluğunu ölçmek mümkündür. Çapraz geçerleme ikinci bölümde de ifade edildiği gibi veri kümesini n eş büyüklükte (program default olarak 10 katman kullanmaktadır) katmana ayırarak; her iterasyonda $n-1$ parçayı modelin öğrenilmesinde 1 parçayı ise modelin testinde kullanan bir yöntemdir. Yüzde ayırma yönteminde ise veri kümesinin belirli bir yüzdesi (program default olarak %66 kullanmaktadır) modelin öğrenilmesine kalan kısmı ise modelin testine ayrılmaktadır

Cluster paneli: Bu panelde veritabanı üzerinde WEKA'daki kümeleme algoritmaları çalıştırılır.

Associate paneli: Bu panelde veri kümesine birliktelik kuralları uygulanır.

Visualize paneli: Bu panel ile farklı nitelikler arasındaki ilişkileri grafik olarak görüntülemek mümkündür.

Modellemenin ilk aşamasında; kozmetik markası ürünlerini almaktan vazgeçerek rakipleri tercih eden müşterileri tespit etmek amacıyla ayrılan müşteri analizi, ikinci aşamasında; farklı müşteri grupları için taşıdıkları karakteristiklere uygun pazarlama politikaları geliştirmek amacıyla müşteri segmentasyonu gerçekleştirilmiştir.

Ayrılan müşteri analizi

Bu analizin amacı ayrılma eğilimi gösteren müşteri karakteristiklerini tanımlamak ve bu müşterileri elde tutmaya yönelik programlar ve pazarlama politikaları geliştirmektir. Ayrılma eğilimi taşıyan müşteri profilini belirlemek için sınıflandırma teknikleri kullanılmıştır. “churn” niteliği, hedef değişken (target variable) olarak belirlenmiştir. Farklı sınıflandırma algoritmaları kullanılarak, kurulan modellerin güvenilirlikleri karşılaştırılmıştır.

Kullanılan algoritmalar ve doğruluk dereceleri aşağıdaki çizelgelerde gösterilmektedir.

Çizelge 4.3. Bayes / Naive Bayes algoritması doğruluk değerleri

Correctly Classified Instances	78	78	%
Incorrectly Classified Instances	22	22	%
Kappa statistic	0.466		
Mean absolute error	0.2141		
Root mean squared error	0.371		
Relative absolute error	52.7885 %		
Root relative squared error	82.5375 %		
Total Number of Instances	100		
=== Detailed Accuracy By Class ===			
TP Rate	FP Rate	Precision	Recall
0.643	0.167	0.6	0.643
0.833	0.357	0.857	0.833
F-Measure	Class		
0.621	E		
0.845	H		
=== Confusion Matrix ===			
a	b	<-- classified as	
18	10	a = E	
12	60	b = H	

Çizelge 4.4. Fonksiyon / Lojistik regresyon algoritması doğruluk değerleri

```

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

Correctly Classified Instances      88          88      %
Incorrectly Classified Instances    12          12      %
Kappa statistic                    0.7207
Mean absolute error                 0.1444
Root mean squared error             0.3496
Relative absolute error             35.5855 %
Root relative squared error         77.7959 %
Total Number of Instances          100

=== Detailed Accuracy By Class ===

TP Rate    FP Rate    Precision    Recall    F-Measure    Class
  0.893      0.125      0.735      0.893      0.806        E
  0.875      0.107      0.955      0.875      0.913        H

=== Confusion Matrix ===

  a  b  <-- classified as
25  3  |  a = E
 9 63  |  b = H

```

Çizelge 4.5. Ağaçlar / ID.3 algoritması doğruluk değerleri

```

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

Correctly Classified Instances      91          91      %
Incorrectly Classified Instances     7           7      %
Kappa statistic                    0.827
Mean absolute error                 0.0938
Root mean squared error             0.2787
Relative absolute error             23.7101 %
Root relative squared error         62.9502 %
UnClassified Instances              2           2      %
Total Number of Instances          100

=== Detailed Accuracy By Class ===

TP Rate    FP Rate    Precision    Recall    F-Measure    Class
  0.926      0.07      0.833      0.926      0.877        E
  0.93      0.074      0.971      0.93      0.95         H

=== Confusion Matrix ===

  a  b  <-- classified as
25  2  |  a = E
 5 66  |  b = H

```

Çizelge 4.6. Ağaçlar / J.48 algoritması doğruluk değerleri

```

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

Correctly Classified Instances          93           93      %
Incorrectly Classified Instances        7           7      %
Kappa statistic                        0.8283
Mean absolute error                    0.0937
Root mean squared error                0.2426
Relative absolute error                23.0877 %
Root relative squared error            53.9703 %
Total Number of Instances              100

=== Detailed Accuracy By Class ===

TP Rate    FP Rate    Precision    Recall    F-Measure    Class
  0.893      0.056      0.862      0.893      0.877        E
  0.944      0.107      0.958      0.944      0.951        H

=== Confusion Matrix ===

  a  b  <-- classified as
25  3  |  a = E
 4 68  |  b = H

```

Çizelge 4.7. Kurallar / PART algoritması doğruluk değerleri

```

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

Correctly Classified Instances          92           92      %
Incorrectly Classified Instances        8           8      %
Kappa statistic                        0.8058
Mean absolute error                    0.1131
Root mean squared error                0.2612
Relative absolute error                27.877 %
Root relative squared error            58.1082 %
Total Number of Instances              100

=== Detailed Accuracy By Class ===

TP Rate    FP Rate    Precision    Recall    F-Measure    Class
  0.893      0.069      0.833      0.893      0.862        E
  0.931      0.107      0.957      0.931      0.944        H

=== Confusion Matrix ===

  a  b  <-- classified as
25  3  |  a = E
 5 67  |  b = H

```

Çizelge 4.8. Kurallar / JRIP algoritması doğruluk değerleri

```

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

Correctly Classified Instances          91           91      %
Incorrectly Classified Instances        9           9      %
Kappa statistic                        0.7926
Mean absolute error                    0.113
Root mean squared error                0.2583
Relative absolute error                 27.856 %
Root relative squared error            57.4725 %
Total Number of Instances              100

=== Detailed Accuracy By Class ===

TP Rate    FP Rate    Precision    Recall    F-Measure    Class
  0.964      0.111      0.771      0.964      0.857        E
  0.889      0.036      0.985      0.889      0.934        H

=== Confusion Matrix ===

  a  b  <-- classified as
27  1 |  a = E
 8 64 |  b = H

```

Çizelge 4.9. Sinir ağları / Multilayer Perceptron algoritması doğruluk değerleri

```

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

Correctly Classified Instances          91           91      %
Incorrectly Classified Instances        9           9      %
Kappa statistic                        0.7839
Mean absolute error                    0.1164
Root mean squared error                0.2866
Relative absolute error                 28.6837 %
Root relative squared error            63.7675 %
Total Number of Instances              100

=== Detailed Accuracy By Class ===

TP Rate    FP Rate    Precision    Recall    F-Measure    Class
  0.893      0.083      0.806      0.893      0.847        E
  0.917      0.107      0.957      0.917      0.936        H

=== Confusion Matrix ===

  a  b  <-- classified as
25  3 |  a = E
 6 66 |  b = H

```

Yukarıdaki çizelgelerde yalnızca doğruluk değerleri gösterilen algoritmaların sonuçları ise EK-3 – EK-8 de yer almaktadır.

Kullanılan algoritmalar ve bu algoritmaların doğruluklarının karşılaştırılması Çizelge 4.10’da gösterilmiştir.

Çizelge 4.10. Modellemede kullanılan sınıflama algoritmaları ve doğrulukları

Algoritma	Naive Bayes	Lojistik Regres.	Id.3	J.48	Jrip	Part	Sinir Ağı.
Doğru olarak sınıflandırılan birimlerin yüzdesi	0,78	0,88	0,91	0,93	0,91	0,92	0,91
Yanlış olarak sınıflandırılan birimlerin yüzdesi	0,22	0,12	0,07	0,07	0,09	0,08	0,09
Ortalama mutlak hata	0,21	0,14	0,09	0,09	0,11	0,11	0,11
Ortalama hata karekök	0,37	0,34	0,27	0,24	0,25	0,26	0,28
Görelî mutlak hata	%52,7	%35,5	%23,7	%23,08	%27,8	%27,8	%28,6
Görelî hata karekök	%82,5	%77,8	%62,9	%53,9	%57,4	%58,1	%63,7
Sınıf E (EVET)/ TP	0,64	0,89	0,92	0,89	0,96	0,89	0,89
Sınıf E (EVET)/ FP	0,16	0,12	0,07	0,05	0,11	0,07	0,08
Sınıf H (HAYIR) / TP	0,83	0,87	0,93	0,94	0,89	0,93	0,92
Sınıf H (HAYIR) / FP	0,35	0,11	0,07	0,11	0,3	0,11	0,11
Sınıf E (EVET) / Presicion	0,62	0,73	0,83	0,89	0,77	0,89	0,80
Sınıf H (HAYIR) / Presicion	0,83	0,95	0,97	0,974	0,98	0,93	0,95
Kappa istatistiği	0,46	0,72	0,82	0,83	0,79	0,80	0,78

Confusion matrix: Yakınsaklık matrisi olarak da adlandırılır. Doğru olarak sınıflandırılan örneklerin sayısı bu matrisin diyagonal elemanlarının toplamına eşittir.

True Positive (TP): Sınıflandırma algoritması tarafından A sınıfına atanan birimlerden gerçekte A sınıfına ait olanların oranını gösterir. Örneğin; Çizelge 4.3’de E sınıfı için TP oranı, 0.643 olup, bu değer yakınsaklık matrisinde gösterilen ve gerçekte E sınıfına ait olan 18 birimin; E sınıfına atanan 28 birime oranlanması ile bulunmuştur.

False Positive (FP): Sınıflandırma algoritması tarafından A sınıfına atandığı halde gerçekte A sınıfına ait olmayan birimlerin oranını gösterir. Örneğin; Çizelge 4.3’de H sınıfı için FP oranı, 0.167 olup, bu değer yakınsaklık matrisinde gösterilen ve gerçekte H sınıfına ait olmayan 12 birimin; H sınıfına atanan 70 birime oranlanması ile bulunur.

Precision: Gerçekte A sınıfına ait olan birimlerin hangi oranda sınıflandırma algoritması tarafından A sınıfına atandığı gösterir. Çizelge 4.3’de E sınıfı için Precision oranı 0.6 olup bu değer yakınsaklık matrisinde gösterilen ve E sınıfına atanan 18 birimin gerçekte E sınıfında yer alan 30 birime oranlanması ile bulunur.

Kappa istatistiği: Tahmin doğruluğunun ölçüsüdür.

Çizelge 4.10’dan da görüldüğü gibi doğruluğu en yüksek olan sınıflandırma algoritması J.48 olduğu için uygulamanın ayrılan müşteri analizi bölümünde bu algoritma esas alınacaktır.

J.48 algoritması; C4.5 karar ağacı algoritmasının WEKA’ya uyarlanmış versiyonudur. Her satır, ağaçtaki bir düğümü; alt satırlar, ilk satırın çocuk düğümlerini; düğümlerde parantezin içindeki ilk sayı veri kümesindeki kaç vakanın bu düğüm için doğru olarak sınıflandırıldığını; eğer varsa; parantezin içindeki ikinci sayı, düğüm tarafından yanlış olarak sınıflandırılan vakaların sayısını gösterir.

J.48 algoritmasının çalıştırılması sonucunda elde edilen çıktı Çizelge 4.10 'da gösterilmiştir.

Çizelge 4.11. J.48 Karar ağacı

```

=== Classifier model (full training set) ===

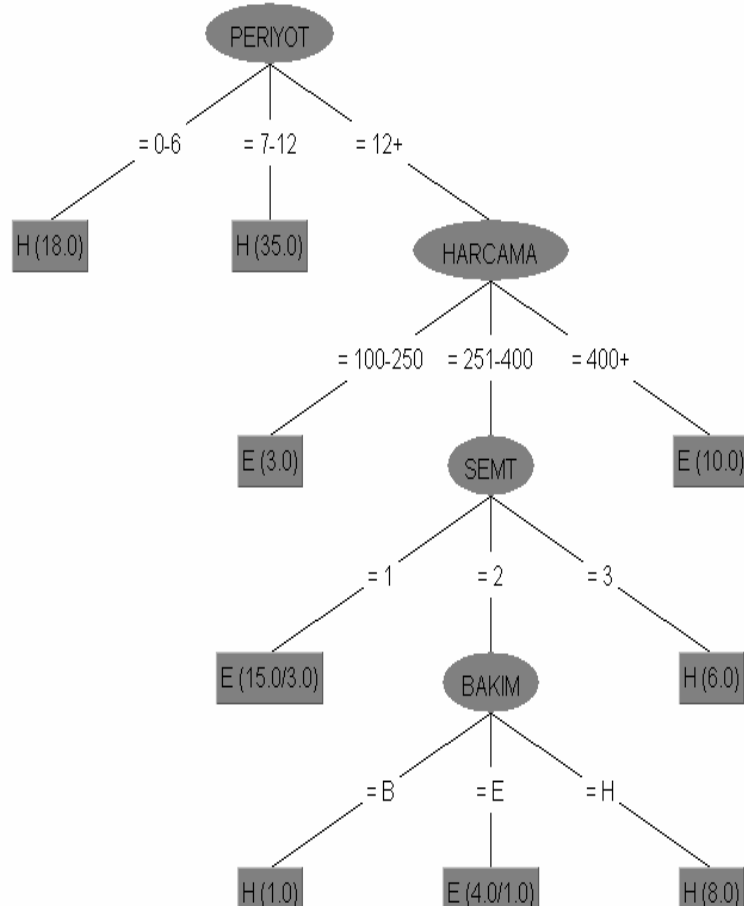
J48 pruned tree
-----

PERİYOT = 0-6: H (18.0)
PERİYOT = 7-12: H (35.0)
PERİYOT = 12+
|   HARCAMA = 100-250: E (3.0)
|   HARCAMA = 251-400
|   |   SEMT = 1: E (15.0/3.0)
|   |   SEMT = 2
|   |   |   BAKIM = B: H (1.0)
|   |   |   BAKIM = E: E (4.0/1.0)
|   |   |   BAKIM = H: H (8.0)
|   |   SEMT = 3: H (6.0)
|   HARCAMA = 400+: E (10.0)

Number of Leaves   :    9
Size of the tree   :   13

```

WEKA’da “classifier output” panelinde metin formatında sunulan karar ağacını “visualize tree “ seçeneğini kullanarak görsel olarak ifade etmek mümkündür. Bu ağaç Şekil 4.7’de gösterilmiştir



Şekil 4.7. J.48 Karar ağacı (E: Evet H: Hayır)

Şekil 4.7.'de ilk dallanma müşterilerin mağazaya geliş periyotlarını; ikinci dallanma harcama tutarlarını; üçüncü dallanma semtlerini ve son dallanma cilt bakımı yaptırıp yaptırmadığını göstermektedir.

Şekil 4.7'de görülen J.48 karar ağacı algoritmasının sonuçlarını aşağıdaki şekilde değerlendirmek mümkündür:

İlk dallanmada geliş periyodu 12 + ay olan müşteri grubunun ayrılma eğilimi taşıdığı görülmektedir.

İkinci dallanmada, harcama tutarı 100-250 TL aralığındaki müşteriler ile harcama tutarı 400 TL'nin üzerindeki müşteriler ayrılma eğilimi taşıyan müşterilerdir. Bu

müşterilerin ayrılmasını ve rakip mağazaları tercih etmelerini önlemek amacıyla geliştirilecek kampanya ve stratejiler aşağıda sıralanmıştır:

- Müşteriye doğum günü, yılbaşı gibi özel günlerde yaşına, cilt tipine ve genel alışveriş alışkanlıklarına uygun kozmetik ürünlerinden oluşan hediye paketlerinin hazırlanması ve sunulması.
- Bu müşterilerin alışveriş tarihleri ve satın aldıkları ürünün özellikleri (türü, kaç mililitre olduğu vb.) bilindiği için ürünün bitme zamanına yakın bir tarihte müşteriye telefon ve e-posta ile ulaşılarak mağazaya davet edilmesi.
- Müşterilerin her alışverişte yaptıkları harcama tutarına paralel olarak puan toplamaları ve bu puanları sonraki alışverişlerinde indirim ya da ücretsiz ürün almak üzere kullanmaları.

Üçüncü dallanmada, ortalama harcama tutarı 250 ile 400 TL arasında değişen müşterilerin rakip mağazayı tercih etmelerinde ikamet ettikleri semtin etkili olduğu görülmektedir. Batıkent hattı üzerinde (“1” ile kodlanan semtlerde) ikamet eden müşterileri elde tutmak daha zordur. Bu müşterileri elde tutmak için geliştirilecek stratejiler şunlar olabilir:

- Reklâm ve tanıtımlarda özellikle Batıkent hattı üzerinde olan billboardları ve metrodaki ilan panolarını kullanmak.
- Batıkent ve civarında ikamet eden müşterileri diğer semtlerde ikamet edenlerden ayıran sosyodemografik özellikleri saptayarak bu grubun karakteristiklerine uygun indirim ve promosyonlar düzenlemek.
- Batıkent ve çevresi lokasyonlarda müşteriye mağazaya çekmek yerine ürünü müşteriye götürmek.

Son dallanma, “2” ile kodlanan semtlerde ikamet eden müşterilerin rakibi tercih etmesindeki etkili faktörün ücretsiz cilt bakımı yaptırıp yaptırmadıkları olduğu görülmektedir. Burada ortaya çıkan sonuç son derece ilginçtir. Cilt bakımı yaptırmayan müşterilerin ayrılma eğilimi daha düşüktür. Bu durumda, mağazaya cilt

bakım programlarını yeniden gözden geçirmeleri ve ortaya çıkan bu sonucun nedenlerini araştırmaları önerilir.

Müşteri segmentasyonu

Müşteri segmentasyonu için WEKA'daki kümeleme algoritmalarından Simple K Means kullanılmıştır. Kümeleme algoritmaları, sınıflama algoritmalarının tersine modeli denetimsiz olarak öğrenirler. Sınıflama algoritmalarında veriler önceden belirlenen sınıflara atandıkları (denetimli öğrenme) için kurulan modelin doğruluğu çeşitli metriklerle ölçülürken; kümeleme algoritmalarında böyle bir değerlendirme söz konusu değildir.

Bu uygulamanın müşteri segmentasyonu bölümünde kullanılan Simple K Means (k ortalamalar) algoritmasında k, tanımlanan küme sayısını gösterir. 2. bölümde de anlatıldığı gibi k ortalamalar algoritması nümerik değerlerle; k modlar algoritması ise kategorik değerlerle çalışmaktadır. WEKA'da Simple K Means algoritması her iki algoritmayı da kullanarak hem nümerik hem de kategorik değerlerle çalışan bir algoritmadır [66]. Yapılan çalışmada kategorik değerlerle elde edilen çıktı Çizelge 4.12'de verilmektedir.

Çizelge 4.12. Simple K Means algoritması çıktısı

```

=== Run information ===

Scheme:      weka.clusterers.SimpleKMeans -N 2 -S 10
Relation:    uygulama
Instances:   100
Attributes:  9
              YAS
              SEMT
              IS
              PERİYOT
              C_TIPI
              BAKIM
              U_TIPI
              HARCAMA

Ignored:     CHURN
Test mode:   evaluate on training data

=== Model and evaluation on training set ===

kMeans
=====

Number of iterations: 3
Within cluster sum of squared errors: 231.0

Cluster centroids:

Cluster 0
  Mean/Mode:  24-35 1 E 7-12 1 H 1 100-250
Cluster 1
  Mean/Mode:  36-45 2 E 12+ 1 H 1 251-400

Clustered Instances

0      47 ( 47%)
1      53 ( 53%)

```

Simple K Means algoritmasını çalıştırarak 2 müşteri grubu elde edilmiştir. Müşteri gruplarının baskın özelliklerini gösteren bu kümeler Çizelge 4.13’de gösterilmektedir.

Çizelge 4.13. Müşteri gruplarının baskın özellikleri

Nitelikler	Grup 1 (Cluster 0)	Grup 2 (Cluster 1)
Yaş Aralığı	24-35	36-45
Semt	1	2
Çalışma Durunu	Evet	Evet
Ort. Alışveriş Periyodu	7-12 ay	12 aydan büyük
Cilt Tipi	Karma	Karma
Bakım Durumu	Hayır	Hayır
Ürün Tipi	1	1
Ort. Harcama Tutarı	100-250 YTL	251-400 YTL
Toplam Müşteriler İçindeki Oranı	%47	%53
Toplam Gelir İçindeki Payı	%21	%79

Çizelge 4.13 incelendiğinde, müşterilerin iki gruba ayrıldığı görülmektedir. Çizelgeye göre birinci grup, 24-35 yaş aralığında; 1 ile kodlanan semtlerde ikamet eden; çalışan; ortalama alış-veriş periyodu 7-12 ay arasında olan; karma ciltli; bakım yaptırmayan; 1 ile kodlu ürünü kullanan; 100-250 YTL arasında harcama yapan; toplam müşterilerin %47 sini; toplam gelirin % 21'ni oluşturan grubu ifade etmektedir.

Simple K Means algoritmasının sonuçlarını aşağıdaki şekilde değerlendirmek mümkündür:

Mağaza, gelirlerinin %21'ini birinci grup müşterilerinden elde etmektedir. Bu müşterilerin mağazaya geliş periyotları kısa ancak ortalama harcama tutarları düşüktür. Bu müşterilere daha yüksek tutarlarda alışveriş yaptırmak için uygulanacak stratejiler şunlar olabilir:

- 1. grup müşterilerinin yaş aralığı 24-35'dir. Belirli bir tutarın üzerinde harcama yapanlara, konser ve sinema bileti vermenin, bu genç grubun harcama tutarını artırmak için etkili bir uygulama olacağı düşünülmektedir.

- 1. grup müşterilerinin harcama tutarını artırmaya yönelik diğer bir öneri çapraz satışır. Çapraz satışta amaç, birbiriyle ilişkili ürünleri müşteriye bir indirim paketi içinde sunarak daha yüksek tutarlarda alışveriş yapmasını sağlamaktır. Bakım ürünlerini tercih eden 1. grup müşterilerine örneğin; bakım ve cilt temizleme ürünlerinin belirli bir indirim paketi içinde sunulması, müşterilerin daha fazla harcama yapmalarını sağlayacaktır.

Mağaza, gelirlerinin %79 gibi büyük bir bölümünü 2. grup müşterilerden sağladığı için asıl odaklanılması gereken grup; 2.gruptur. 2.gruptaki müşteriler mağazaya daha uzun periyotlarla gelmekte, fakat geldiği zaman da daha büyük tutarlarda alışveriş yapmaktadır. O halde, veri madenciliği çıktıları kampanya yönetiminde kullanılırken üzerinde durulması gereken nokta; 2.grup müşterilerini daha sık mağazaya çekmeyi sağlayacak stratejiler geliştirmektir.

- Bu grubun yaş aralığı 36-45, cilt tipleri karma ve satın aldıkları ürünler bakım kategorisinde olduğu için kampanyalarda müşteriye sunulan ürünler orta yaş grubuna hitap eden, yaşlanmayı önleyici ve geciktirici, karma cilt tipine uygun bakım ürünleri olmalıdır. Tanıtım ve kampanyalarda orta yaşlı bayanların okuduğu bir dergiyi kullanmak etkili bir yöntem olabilir.

- Bu gruptaki müşterilerin çalıştığı düşünüldüğünde ve her birinin aynı zamanda İnternet kullanıcısı olduğu varsayıldığında ürün, indirim ve promosyonlar hakkında bilgi vermek için sıfır maliyetli e-postayı kullanmak da mümkündür. Farklı gruplar için farklı ürün paketlerini içeren online kataloglar belirli periyotlarla e-posta aracılığı ile müşteriye sunulabilir. Ancak bu noktada, etkili bir uygulama için izinli pazarlama yöntemine başvurulmalı, ürünlerle ilgili e posta göndermeden önce müşterinin izni alınmalıdır.

5. SONUÇ VE ÖNERİLER

Veri madenciliğinin en yaygın uygulama alanlarından biri pazarlamadır. Şirketler, veri madenciliğini kampanya yönetimi, müşteri segmentasyonu, müşteri edinme, mevcut müşterileri elde tutma, müşteri yaşam ömrü analizi, müşteri ilişkileri yönetimi, pazar sepeti analizleri ve kredi derecelendirme gibi pek çok alanda kullanmaktadır.

Bu tezde, veri madenciliği teknikleri kullanılarak bir kozmetik markasının müşterilerini gruplandıran ve mağazadan ayrılarak rakip mağazaların müşterisi olmayı tercih eden müşteri davranışlarını ortaya koyan modeller geliştirilerek; müşteri tutmaya yönelik çeşitli kampanyalar ve pazarlama stratejileri önerilmiştir.

SQL komutları kullanarak verinin temizlenmesi ve istenilen formata dönüştürülmesi işlemlerinin ardından; WEKA'daki sınıflandırma ve kümeleme teknikleri kullanılarak veri modellenmiştir. Sınıflama tekniklerinden Naive Bayes, lojistik regresyon, J.48 ve ID.3 karar ağaçları, JRIP ve PART kuralları ile sinir ağları kullanılmış; bu algoritmaların doğruluk dereceleri karşılaştırılarak sınıflandırma için en uygun algoritmanın J.48 karar ağacı olduğu saptanmıştır. J.48 karar ağacı ile müşterinin ayrılarak rakip şirketi tercih etme (churn) davranışı modellenmiştir.

Kozmetik sektörü sirkülasyonun çok fazla olduğu dinamik bir sektördür ve müşteriler rakip mağazaların sunduğu hediye ve indirim olanaklarına kayıtsız kalmamaktadır. Ayrılma eğilimi gösteren müşteri profilini belirledikten sonra bu müşterilere özel kampanyaların düzenlenmesi ve pazarlama stratejileri belirlenmesi mümkündür.

Kozmetik mağazası, alışveriş periyodu 18 ayın üzerindeki müşterileri ayrılma eğilimi gösteren müşteri olarak tanımlamaktadır. Ortalama alışveriş tutarı 100–250 TL aralığındaki müşteriler ile 400 TL'nin üzerinde olan müşterilerin alışveriş periyodu 18 ayın üzerindedir. Önerimiz; bu müşterilere doğum günü ve yılbaşı gibi özel

günlerde taşıdıkları karakteristiklere uygun ürün ve promosyon paketleri hazırlanarak hediye edilmesidir.

Ortalama harcama tutarı 250–400 TL arasında değişen müşterilerin ise ayrılma eğilimi göstermesinde etkili olan faktörün ikamet ettikleri semt olduğu görülmektedir. Özellikle -mağaza ile aynı hat üzerinde olmasına rağmen- Batıkent hattı üzerinde ikamet eden müşterileri tutmak daha zordur. Bu konudaki önerimiz reklam ve tanıtımları lokasyon bazında özelleştirmektir. Batıkent civarındaki billboardları, metro hattındaki reklam panolarını kullanmanın etkili yöntemler olacağı düşünülmektedir.

Kümeleme algoritmalarından Simple K Means kullanılarak müşteri segmentasyonu gerçekleştirilmiş; farklı demografik özellikler ve farklı satın alma davranışı gösteren 2 müşteri grubu elde edilmiştir.

1.grup müşterileri mağazaya daha kısa periyotlarla gelmekte ve küçük tutarlarda alışveriş yapmaktadır. 2. grup müşterileri ise mağazaya geliş periyotları uzun olup büyük tutarlarda alışveriş yapmaktadır. O halde, 1. grup müşterilerinin alışveriş tutarlarını; 2.grup müşterilerinin ise mağazaya gelme sıklıklarını artırmaya yönelik stratejiler geliştirilmelidir.

Kozmetik mağazası müşteri edinme ve tutundurma faaliyetlerini yürütmek için kitle pazarlama araçlarından yazılı ve görsel basını kullanmaktadır. Önerimiz; mağazanın kitle pazarlama araçlarını veri madenciliği ile elde edilen müşteri gruplarına uygun olarak kullanmasıdır. Örneğin; 2.grup, 36-45 yaş grubundaki müşterilerden oluştuğu için bu gruba yönelik reklamlar orta yaşlı bayanların okuduğu bir dergiye verilebilir.

Mevcut müşteriye elde tutmak için doğrudan pazarlama yöntemleri kullanılmakta; satış görevlileri müşterileri telefonla aramak suretiyle indirim, hediye ve promosyonlar ile bakım günleri hakkındaki bilgi vermektedir. Müşterilerin e-posta adresleri bilinmesine rağmen e-posta aracılığı ile müşteriye ulaşma çok yaygın bir uygulama değildir. Önerimiz; farklı müşteri grupları için farklı ürün ve promosyon

paketlerini içeren online kataloglar hazırlayarak belirli periyotlarla e-posta aracılığı ile müşteriye sunmaktır. Uygulamanın etkinliği için mağaza izinli pazarlama yöntemini kullanmalı, sadece ürün/hizmetlerle ilgilenen müşterilere e-posta göndermelidir.

İleri analizler yapabilmek için daha detaylı müşteri verilerine (gelir, meslek, tahsil, medeni hal, çocuk sayısı) gereksinim duyulmaktadır. Bu konuda mağazaya önerimiz; bugün perakendeden hazır giyime pek çok sektörde kullanılan sadakat kartı uygulamasına geçmeleridir. Sadakat kartı başvuru formunun doldurulması sırasında elde edilen müşteri karakteristikleri ile kart üzerinde toplanan satın alma davranışı verileri birleştirilerek çok daha spesifik analizler yapılabilecektir.

İlerideki çalışmalarda, önerilen stratejilerin mağaza tarafından uygulamaya alınmasının ardından; satışları ve müşterilerin ayrılma oranlarını nasıl etkilediğine dayalı olarak yatırımın geri dönüş oranını (ROI) saptamak yoluyla yapılan çalışmanın etkinliği ölçülebilecektir.

KAYNAKLAR

1. Bayram E., "Customer segmentation and churn modeling in wireless communication", Yüksek Lisans Tezi, **Boğaziçi Üniversitesi Fen Bilimleri Enstitüsü**, İstanbul, 1-5 (2001).
2. Restivo K., "The drill on data mining", **Computer Dealer News**, 15 (14): 29-30 (1999).
3. Jacobs P., "Data Mining: What general managers need to know", **Harvard Management Update**, 4 (10): 8 (1999).
4. Davis B., "Data mining transformed", **Information Week**, 751: 86 (1999).
5. DuMouchel W., "Bayesian data mining in large frequency tables, with an application to the FDA spontaneous", **American Statistician**, 53 (3): 177 (1999).
6. Hand D.J., "Data mining: statistics and more ?", **The American Statistician**, 52: 112-118 (1998).
7. Kitler R., Wang W., "The emerging role of data mining", **Solid State Technology**, 42 (11): 45 (1998).
8. Bransten L., "Technology – power tools – looking for patterns: data mining enables companies to better manage the ream of statistics they collect; the goal: spot the unexpected", **Wall Street Journal**, 27 (12): 16- 20 (1999).
9. Hudairy H., "Data mining and decision making support in the governmental sector", Master Thesis, **Faculty of Graduate School of The University of Louisville**, Kentucky, 1-5 (2004).
10. Akpınar H., "Veritabanlarında bilgi keşfi ve veri madenciliği", **İstanbul Üniversitesi İşletme Fakültesi Dergisi**, 29: 1-22 (2000).
11. İnternet: Current data mining applications / percentage in different industries, http://www.kdnuggets.com/polls/2003/data_mining_applications_industries.html (2005).
12. Aydoğan F., "E-ticarette veri madenciliği yaklaşımlarıyla müşteriye hizmet sunan akıllı modüllerin tasarımı ve gerçekleştirimi", Yüksek Lisans Tezi, **Hacettepe Üniversitesi Fen Bilimleri Enstitüsü**, Ankara, 12-16 (2003).
13. Shearer C., "The Crisp-DM model: The new blueprint for data mining", **Journal of Data Warehousing**, 5 (4): 13-23 (2000).

14. Piramuthu S., "Evaluating feature selection methods for learning in data mining applications", *Thirty-First Annual Hawai International Conference on System Sciences*, 5: 294 (1998).
15. Witten I., Frank E., "Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations", *Morgan Kaufmann Publishers*, San Fransisco, 120-129, 267-277 (2000).
16. Giudici P. "Applied Data Mining: Statistical Methods for Business and Industry", *John Wiley & Sons Inc.*, Chichester, 85-100 (2003).
17. Huberty C., "Applied Discriminant Analysis", *John Wiley & Sons Inc.*, Canada, 25-35 (1994).
18. Wei C., Chiu T., "Turning telecommunications call details to churn prediction: a data mining approach", *Expert Systems with Applications*, 23: 103-102 (2002).
19. Aryeetey K., "Data analysis and predictive modelling using the variable precision rough set approach", Master Thesis, *Faculty of Graduate Study and Research of University of Regina*, Canada, 28-33 (2003).
20. Pawlak Z., "Rough sets, decision algorithms and bayes theorem", *European Journal of Operation Research*, 136: 181-189 (2002).
21. Shah S., Kursak A., "Data mining and genetic algorithms based gene / SNP selection", *Artificial Intelligence in Medicine*, 31: 183 -196 (2004).
22. Hui S., Jha G., "Application data mining for customer service support", *Information and Management*, 38: 1-13 (2000).
23. Anand A., "A study and comparison of data clustering techniques", Master Thesis, *Faculty of The Graduate School of The University of Texas El Paso*, Texas, 21-22 (2003).
24. San O., Huynh V., Nakamori Y., "An alternative extention of the k means algorithm for clustering categorical data", *Applied Math. and Computer Science*, 14 (2): 241-247 (2004).
25. Bland C., "The discovery of multiple level profile association rules", Doctoral Thesis, *Graduate School of Computer and Information Science of University of Missisipi*, Missisipi, 43-50 (2002).
26. Boginski V., Butenko S., Pardalos P., "Mining market data: A network approach", *Computers & Operations Research*, 33 (11): 3171-3184 (2006).
27. Jiao J., Zhang Y., Helander M., "Kansei mining system for affective design", *Expert Systems with Applications*, 30 (4): 658-673 (2006).

28. Jeng B., Chen J., Liang T., "Applying data mining to learn system dynamics in a biological model", *Expert Systems with Applications*, 30 (1): 50-58 (2006).
29. Facca F., Lanzi P., "Mining interesting knowledge from weblogs: a survey", *Data & Knowledge Engineering*, 53 (3): 225-241 (2005).
30. Hong G., Park S., Jang D., Rho H., "An effective supplier selection method for constructing a competitive supply relationship", *Expert Systems with Applications*, 28 (4): 629-639 (2005).
31. Chen M., Huang C., Chen K., Wu H., "Aggregation of orders in distribution centers using data mining", *Expert Systems with Applications*, 28 (3): 453-460 (2005).
32. Cervone G., Kafatos M., Napoletani D., Singh R., "An early warning system for coastal earthquakes", *Advances in Space Research*, 37 (4): 636-642 (2006).
33. Crespo F., Weber R., "A methodology for dynamic data mining based on fuzzy clustering", *Fuzzy Sets and Systems*, 150 (2): 267-284 (2005).
34. Lee T., Chiu C., Chou Y., Lu C., "Mining the customer credit using classification and regression tree and multivariate adaptive regression splines", *Computational Statistics & Data Analysis*, 50 (4): 1113-1130 (2006).
35. Bellazi R., Larizza C., Magni P., "Temporal data mining for the quality assessment of hemodialysis services", *Artificial Intelligence in Medicine*, 34 (1): 25-39 (2005).
36. Last M., Kandel A., "Discovering useful and understandable patterns in manufacturing data", *Robotics and Autonomous Systems*, 49 (3): 137-152 (2004).
37. Lian J., Lai M., Lin Q., Yao F., "Application of data mining and process knowledge discovery in sheet metal assembly dimensional variation diagnosis", *Journal of Materials Processing Technology*, 129 (1): 315-320 (2002).
38. Lin F., McClean S., "A data mining approach to the prediction of corporate failure", *Knowledge-Based Systems*, 14 (3): 189-195 (2001).
39. Caskey K., "A manufacturing problem solving environment combining evaluation, search, and generalisation methods", *Computers in Industry*, 44 (2): 175-187 (2001).
40. Cox I., Lewis R., Ransing R., Laszczewski H., Berni G., "Application of neural computing in basic oxygen steelmaking", *Journal of Materials Processing Technology*, 120 (1): 310-315 (2002).

41. Sforza M., "Data mining in a power company customer database", *Electric Power Systems Research*, 55 (3): 201-209 (2000).
42. Guha S., Rastogi R., Shim K., "Rock: A robust clustering algorithm for categorical attributes", *Information Systems*, 25 (5): 345-366 (2000).
43. Rud O., "Data Mining Cookbook: Modelling Data For Marketing, Risk and Customer Relationship Management", *John Wiley & Sons Inc.*, Canada, 4-11 (2001).
44. Rygielski C., Wang J., Yen D., "Data mining techniques for customer relationship management", *Technology in Society*, 24: 483-502 (2002).
45. Berry M, Linoff G., "Data Mining Techniques For Marketing, Sales and Customer Support", *John Wiley & Sons Inc.*, New Jersey, 124-128 (1997).
46. Schertel S., "Data mining and its potential use in textiles: A spinning mill", Doctoral Thesis, *Graduate Faculty of North Carolina State University*, Raleigh, 19-20 (2002).
47. Hwang H., Jung T., Suh E., "An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry", *Expert System with Applications*, 26: 181-188 (2004).
48. Hsu C., Chen Y., "Mining of mixed data with application to catalog marketing", *Expert System with Applications*, In Press (2006).
49. Kotler P., "Marketing Management, Analysis, Planning and Control", *Prentice Hall*, New Jersey, 141-158 (1976).
50. Kotler P., Armstrong G., "Principles of Marketing", *Prentice Hall*, New Jersey, 245-256 (2001).
51. Massey P., Mitzi M., "Reengineering CRM: leveraging knowledge assets in IBM", *Decision Support Systems*, 32 (2): 155-170 (2001).
52. Russel S., "A framework for CRM", *California Management Review*, 43 (4): 89-108 (2001).
53. Nath V., "Data warehousing and mining in the wireless industry", Master Thesis, *Faculty of the College of Business of Florida Atlantic University*, Florida, 5-8 (2003).
54. Jonker J., Piersma N., Poel D., "Joint optimization of customer segmentation and marketing policy to maximize long term profitability", *Working Paper of Faculty of Economics and Business Administration Ghent University*, 3: 214-239 (2003).

55. Mucuk İ., “Pazarlama İlkeleri”, *Türkmen Kitabevi*, İstanbul, 233-255 (2004).
56. İnternet: Data mining and customer relationships, <http://www.thearling.com/text/whexcerpt/whexcerpt.htm> (2006).
57. Godin S., “Permission Marketing: Turning Strangers Into Friends and Friends Into Customers”, *Simon and Shuster*, New York, 40-45 (1999).
58. Li S., Shue L., Lee S., “Enabling customer relationship management in ISP services through mining usage patterns”, *Expert Systems with Applications*, 30 (4): 621-632 (2006).
59. Hsieh N., “An integrated data mining and behavioral scoring model for analyzing bank customers”, *Expert Systems with Applications*, 27 (4): 623-633 (2004).
60. Kim E., Lee Y., “Combination of multiple classifiers for the customer's purchase behavior prediction”, *Decision Support Systems*, 34 (2): 167-175 (2003).
61. Liu D., Shih Y., “Integrating AHP and data mining for product recommendation based on customer lifetime value”, *Information & Management*, 42 (3): 387-400 (2005).
62. Chen M., Chiu A., Chang H., “Mining changes in customer behavior in retail marketing”, *Expert Systems with Applications*, 28 (4): 773-781 (2005).
63. Hung S., Yen D., Wang H., “Applying data mining to telecom churn management”, *Expert Systems with Applications*, In Press (2005).
64. Yen S., Lee Y., “An efficient data mining approach for discovering interesting knowledge from customer transactions”, *Expert Systems with Applications*, 30 (4): 650-657 (2006).
65. İnternet: Mining customer data, http://www.db2mag.com/db_area/archives/1998/q3/98fsaar.shtml (2006).
66. İnternet: K Means clustering in WEKA, <http://maya.cs.depaul.edu/~classes/ect584/WEKA/k-means.html> (2006).

EKLER

EK-1 Orjinal veri tablosu formatı

Çizelge 1.1. Orijinal veri tablosu

DOĞ. TAR	SEMT	İŞ	ALIŞ. TAR.	ALIŞ. TAR.	ALIŞ. TAR.	ALIŞ. TAR.	ALIŞ. TAR.	ALIŞ. TAR.	CİLT TİPİ	BAKIM	ÜRÜN TİPİ	HRC. (YTL)	HRC. (YTL)	HRC. (YTL)	HRC. (YTL)	HRC. (YTL)	HRC. (YTL)
19.04.50	Bahçeli	E	1.10.04	1.6.05					Karma	Hayır	Bakım	552	293				
06.02.53	Bahçeli	E	1.6.05						Kuru	Hayır	Makyaj	228					
18.04.53	Etimes.	H	1.4.05						Karma	Hayır	Bakım	270					
29.05.53	Batıkent	H	1.11.02	1.2.03	1.7.03	1.12.04	1.12.05		Kuru	Bazen	Bakım	140	270	190	440	160	
10.09.53	Batıkent	E	1.5.04	1.9.05					Kuru	Hayır	Bakım	329	574				
21.12.53	Batıkent	H							Yağlı	Hayır	Nemlendirici/ Temizleme	419					
10.01.56	Batıkent	E	1.6.04						Karma	Hayır	Bakım	290					
27.03.56	Batıkent	E	1.2.04	1.3.04	1.5.05				Karma	Bazen	Bakım/Parfüm	465	62	615			
27.03.56	Batıkent	E	1.2.04	1.3.04	1.5.05				Karma	Bazen	Bakım/Parfüm	62	407	373			

EK-2 Arff uzantılı veri dosyası

@relation uygulama

@attribute YAS{24-35,36-45,45+}
 @attribute SEMT{1,2,3}
 @attribute IS{E,H}
 @attribute CHURN{E,H}
 @attribute PERİYOT{0-6,7-12,12+}
 @attribute C_TIPI{1,2,3}
 @attribute BAKIM{B,E,H}
 @attribute U_TIPI{1,2,3,4,5,6}
 @attribute HARCAMA{100-250,251-400,400+}

@data

24-35,3,E,H,0-6,1,H,1,100-250
 24-35,3,E,H,0-6,2,H,3,100-250
 24-35,1,H,H,0-6,1,H,1,100-250
 24-35,1,H,H,0-6,2,B,1,100-250
 24-35,1,E,H,0-6,2,H,1,100-250
 24-35,1,E,H,0-6,3,H,5,100-250
 24-35,1,E,H,0-6,1,H,1,100-250
 24-35,1,E,H,0-6,1,B,6,100-250
 24-35,1,E,H,0-6,1,B,6,100-250
 24-35,1,E,H,0-6,1,H,1,100-250
 24-35,1,E,H,0-6,2,H,5,100-250
 24-35,1,E,H,0-6,2,E,1,100-250
 24-35,1,E,H,7-12,1,H,1,100-250
 36-45,1,H,H,0-6,1,H,4,100-250
 36-45,1,H,H,0-6,1,H,4,100-250
 24-35,1,E,H,7-12,2,H,1,100-250
 24-35,1,E,H,7-12,1,H,1,100-250
 24-35,1,E,H,7-12,1,H,2,100-250
 24-35,1,E,H,7-12,1,H,1,100-250
 24-35,1,E,H,7-12,2,E,1,100-250
 24-35,1,E,H,7-12,1,H,1,100-250
 24-35,1,E,H,7-12,1,B,2,100-250
 24-35,1,E,H,7-12,1,H,5,100-250
 24-35,1,E,H,7-12,1,E,1,100-250
 24-35,1,E,H,7-12,1,H,3,100-250
 24-35,1,E,H,7-12,2,H,2,100-250
 24-35,1,E,H,7-12,1,H,1,100-250
 24-35,1,E,H,7-12,1,B,5,100-250
 24-35,1,E,H,7-12,1,B,5,100-250
 24-35,1,E,H,7-12,1,H,2,100-250
 24-35,1,E,H,7-12,2,H,3,100-250
 24-35,1,E,H,7-12,2,H,4,100-250
 24-35,1,E,H,7-12,3,H,1,100-250
 36-45,1,E,H,7-12,1,H,1,100-250
 36-45,1,E,H,7-12,1,E,5,100-250
 36-45,1,E,H,7-12,1,H,1,100-250
 36-45,1,E,H,7-12,1,H,1,100-250
 36-45,1,E,H,7-12,1,H,1,100-250
 36-45,1,H,H,0-6,1,H,4,100-250
 36-45,3,H,H,7-12,1,H,2,251-400
 36-45,1,E,H,7-12,1,B,4,251-400

EK-2 (Devam) Arff uzantılı veri dosyası

36-45, 1, E, H, 7-12, 1, H, 4, 251-400
36-45, 1, H, H, 7-12, 1, E, 2, 251-400
36-45, 3, E, H, 7-12, 2, H, 4, 251-400
36-45, 3, E, H, 7-12, 1, H, 1, 251-400
36-45, 3, E, H, 7-12, H, 1, 251-400
36-45, 3, E, H, 7-12, 1, B, 2, 251-400
36-45, 2, E, H, 7-12, 2, B, 3, 251-400
36-45, 2, E, H, 7-12, 1, H, 1, 251-400
36-45, 2, H, H, 7-12, 3, E, 3, 251-400
36-45, 2, E, H, 7-12, 1, H, 2, 251-400
36-45, 2, E, H, 12+, 1, E, 1, 251-400
36-45, 2, H, H, 12+, 2, H, 1, 251-400
36-45, 2, E, H, 12+, 1, H, 1, 251-400
36-45, 2, E, H, 12+, 1, H, 1, 251-400
36-45, 2, E, H, 12+, 1, H, 1, 251-400
36-45, 2, E, H, 12+, 1, H, 1, 251-400
36-45, 2, E, H, 12+, 1, H, 1, 251-400
36-45, 2, E, H, 12+, 1, H, 3, 251-400
36-45, 2, E, H, 12+, 1, H, 1, 251-400
36-45, 1, E, H, 12+, 1, E, 1, 251-400
36-45, 3, E, H, 12+, 1, H, 3, 251-400
36-45, 3, E, H, 12+, 1, H, 1, 251-400
36-45, 3, E, H, 12+, 1, H, 1, 251-400
36-45, 2, E, H, 12+, 2, B, 1, 251-400
36-45, 3, E, H, 12+, 1, H, 1, 251-400
36-45, 3, E, H, 12+, 1, H, 4, 251-400
36-45, 1, H, H, 12+, 2, B, 2, 251-400
36-45, 3, E, H, 12+, 1, E, 5, 251-400
36-45, 1, E, H, 12+, 1, H, 1, 251-400
36-45, 1, E, E, 12+, 1, E, 2, 251-400
36-45, 1, E, E, 12+, 1, H, 1, 251-400
36-45, 2, E, E, 12+, 2, E, 6, 251-400
36-45, 2, E, E, 12+, 1, E, 1, 251-400
36-45, 2, E, E, 12+, 1, E, 2, 251-400
36-45, 1, E, E, 12+, 1, H, 1, 251-400
36-45, 1, E, E, 12+, 1, H, 1, 251-400
36-45, 1, E, E, 12+, 1, H, 1, 251-400
36-45, 1, E, E, 12+, 1, H, 1, 251-400
36-45, 1, E, E, 12+, 1, E, 1, 251-400
36-45, 1, E, E, 12+, 1, H, 1, 251-400
36-45, 1, E, E, 12+, 2, H, 1, 251-400
36-45, 1, H, E, 12+, 1, E, 1, 251-400
36-45, 1, E, E, 12+, 1, H, 251-400
36-45, 1, E, E, 12+, 1, E, 1, 251-400
24-35, 1, H, H, 0-6, 1, E, 1, 251-400
24-35, 1, H, E, 12+, 1, E, 2, 400+
36-45, 3, E, H, 7-12, 1, H, 1, 400+
36-45, 1, E, H, 7-12, 1, B, 5, 400+
36-45, 3, E, H, 0-6, 1, E, 1, 400+
45+, 1, H, H, 0-6, 2, E, 1, 400+
45+, 2, H, E, 12+, 1, E, 2, 400+
45+, 2, H, E, 12+, 2, H, 2, 400+
45+, 2, E, E, 12+, 2, H, 2, 400+
45+, 2, E, E, 12+, 2, H, 2, 400+

EK-2 (Devam) Arff uzantılı veri dosyası

45+, 2, E, E, 12+, 2, B, 4, 400+
45+, 2, E, E, 12+, 2, B, 4, 400+
45+, 2, H, E, 12+, 2, B, 400+
45+, 2, H, E, 12+, 2, B, 4, 400+
45+, 2, H, E, 12+, 2, B, 1, 400+

EK-3 Naive Bayes algoritması sonuçları

Çizelge 3.1. Naive Bayes algoritması sonuç tablosu

```

=== Run Information ===

Scheme:      weka.classifiers.bayes.NaiveBayes
Relation:    uygulama
Instances:   100
Attributes:  9
              YAS
              SEMT
              IS
              CHURN
              PERİYOT
              C_TIPI
              BAKIM
              U_TIPI
              HARCAMA
Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===

Naive Bayes Classifier

Class E: Prior probability = 0.28

YAS      : Discrete Estimator. Counts =  2 19 10  (Total = 31)
SEMT     : Discrete Estimator. Counts = 17 13  1  (Total = 31)
IS       : Discrete Estimator. Counts = 22  8      (Total = 30)
PERİYOT  : Discrete Estimator. Counts =  1  1 29  (Total = 31)
C_TIPI   : Discrete Estimator. Counts = 19 11  1  (Total = 31)
BAKIM    : Discrete Estimator. Counts =  6 10 15  (Total = 31)
U_TIPI   : Discrete Estimator. Counts = 17  8  1  5  1  2 (Total = 34)
HARCAMA  : Discrete Estimator. Counts =  4 16 11  (Total = 31)

Class H: Prior probability = 0.72

YAS      : Discrete Estimator. Counts = 35 38  2  (Total = 75)
SEMT     : Discrete Estimator. Counts = 44 15 16  (Total = 75)
IS       : Discrete Estimator. Counts = 63 11      (Total = 74)
PERİYOT  : Discrete Estimator. Counts = 19 36 20  (Total = 75)
C_TIPI   : Discrete Estimator. Counts = 53 17  5  (Total = 75)
BAKIM    : Discrete Estimator. Counts = 13 13 49  (Total = 75)
U_TIPI   : Discrete Estimator. Counts = 41 10  8  7  9  3 (Total = 78)
HARCAMA  : Discrete Estimator. Counts = 35 34  6  (Total = 75)

```

EK-4 Lojistik regresyon algoritması sonuçları

Çizelge 4.1. Lojistik regresyon algoritması sonuç tablosu

```

=== Run information ===
Scheme:      weka.classifiers.functions.Logistic -R 1.0E-8 -M -1
Relation:    uygulama
Instances:   100
Attributes:  9
              YAS
              SEMT
              IS
              CHURN
              PERİYOT
              C_TIPI
              BAKIM
              U_TIPI
              HARCAMA
Test mode:   10-fold cross-validation
=== Classifier model (full training set) ===
Logistic Regression with ridge parameter of 1.0E-8
Variable      Coeff.
   1          2.0428
   2        -10.0447
   3         22.4594
   4         12.5796
   5          8.7551
   6        -37.0784
   7         -0.294
   8        -42.2432
   9        -34.6376
  10         56.6646
  11         -3.0321
  12         -2.4283
  13         28.7487
  14        -28.4297
  15         10.518
  16          9.6201
  17        -11.1587
  18          8.2116
  19        -35.1131
  20         14.1853
  21         12.1758
  22         60.4434
  23          1.0534
  24        -10.7608
  25         19.1399
Intercept    -42.4126
Odds Ratios...
Variable      O.R.
   1          7.7119
   2           0
   3    5675163138.4016
   4    290578.3318

```

EK-4 (Devam) Lojistik regresyon algoritması sonuçları

Çizelge 4.1. (Devam) Lojistik regresyon algoritması sonuç tablosu

5	6343.1812
6	0
7	0.7453
8	0
9	0
10	4.065626493050205E24
11	0.0482
12	0.0882
13	3.0577811399642637E12
14	0
15	36975.0084
16	15064.3487
17	0
18	3683.3419
19	0
20	1447444.997
21	194044.2855
22	1.779170307311308E26
23	2.8675
24	0
25	05288846.5162

EK-5 ID.3 algoritması sonuçları

Çizelge 5.1. ID.3 algoritması sonuç tablosu

```

=== Run information ===

Scheme:      weka.classifiers.trees.Id3
Relation:    uygulama
Instances:   100
Attributes:  9
              YAS
              SEMT
              IS
              CHURN
              PERİYOT
              C_TIPI
              BAKIM
              U_TIPI
              HARCAMA
Test mode:   10-fold cross-validation
=== Classifier model (full training set) ===

Id3
PERİYOT = 0-6: H
PERİYOT = 7-12: H
PERİYOT = 12+
|  HARCAMA = 100-250: E
|  HARCAMA = 251-400
|  |  SEMT = 1
|  |  |  BAKIM = B: H
|  |  |  BAKIM = E
|  |  |  |  IS = E
|  |  |  |  |  U_TIPI = 1: E
|  |  |  |  |  U_TIPI = 2: E
|  |  |  |  |  U_TIPI = 3: null
|  |  |  |  |  U_TIPI = 4: null
|  |  |  |  |  U_TIPI = 5: null
|  |  |  |  |  U_TIPI = 6: null
|  |  |  |  IS = H: E
|  |  |  BAKIM = H
|  |  |  |  C_TIPI = 1: E
|  |  |  |  C_TIPI = 2: E
|  |  |  |  C_TIPI = 3: null
|  |  SEMT = 2
|  |  |  BAKIM = B: H
|  |  |  BAKIM = E
|  |  |  |  U_TIPI = 1: E
|  |  |  |  U_TIPI = 2: E
|  |  |  |  U_TIPI = 3: null
|  |  |  |  U_TIPI = 4: null
|  |  |  |  U_TIPI = 5: null
|  |  |  |  U_TIPI = 6: E
|  |  |  BAKIM = H: H
|  |  SEMT = 3: H
|  HARCAMA = 400+: E

```

EK-6 J.48 algoritması sonuçları

Çizelge 6.1. J.48 algoritması sonuç tablosu

```

=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    uygulama
Instances:   100
Attributes:  9
              YAS
              SEMT
              IS
              CHURN
              PERİYOT
              C_TIPI
              BAKIM
              U_TIPI
              HARCAMA
Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----

PERİYOT = 0-6: H (18.0)
PERİYOT = 7-12: H (35.0)
PERİYOT = 12+
|   HARCAMA = 100-250: E (3.0)
|   HARCAMA = 251-400
|   |   SEMT = 1: E (15.0/3.0)
|   |   SEMT = 2
|   |   |   BAKIM = B: H (1.0)
|   |   |   BAKIM = E: E (4.0/1.0)
|   |   |   BAKIM = H: H (8.0)
|   |   SEMT = 3: H (6.0)
|   HARCAMA = 400+: E (10.0)

Number of Leaves   :    9
Size of the tree   :   13

```

EK-7 JRIP algoritması sonuçları

Çizelge 7.1. JRIP algoritması sonuç tablosu

```

=== Run information ===

Scheme:      weka.classifiers.rules.JRip -F 3 -N 2.0 -O 2 -S 1
Relation:    uygulama
Instances:   100
Attributes:  9
              YAS
              SEMT
              IS
              CHURN
              PERİYOT
              C_TIPI
              BAKIM
              U_TIPI
              HARCAMA
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

JRIP rules:
=====

(PERİYOT = 12+) and (HARCAMA = 400+) => CHURN=E (10.0/0.0)
(PERİYOT = 12+) and (SEMT = 1) => CHURN=E (18.0/3.0)
(BAKIM = E) and (SEMT = 2) => CHURN=E (5.0/2.0)
=> CHURN=H (67.0/0.0)

Number of Rules : 4

```

EK-8 PART algoritması sonuçları

Çizelge 8.1. PART algoritması sonuç tablosu

```

=== Run information ===

Scheme:      weka.classifiers.rules.PART -M 2 -C 0.25 -N 3 -Q 1
Relation:    uygulama
Instances:   100
Attributes:  9
              YAS
              SEMT
              IS
              CHURN
              PERIYOT
              C_TIPI
              BAKIM
              U_TIPI
              HARCAMA
Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===

PART decision list
-----

PERIYOT = 7-12: H (38.0)

PERIYOT = 0-6: H (18.0)

HARCAMA = 251-400 AND
SEMT = 1: E (15.0/3.0)

HARCAMA = 251-400: H (19.0/3.0)

: E (10.0)

Number of Rules : 5

```

EK-9 Sinir ağıları algoritması sonuçları

Çizelge 9.1. Sinir ağıları algoritması sonuç tablosu

```

=== Run information ===

Scheme:      weka.classifiers.functions.MultilayerPerceptron -L 0.3
-M 0.2 -N 500 -V 0 -S 0 -E 20 -H a
Relation:     uygulama
Instances:    100
Attributes:   9
              YAS
              SEMT
              IS
              CHURN
              PERİYOT
              C_TIPI
              BAKIM
              U_TIPI
              HARCAMA
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

Sigmoid Node 0
  Inputs      Weights
  Threshold   0.30406620050095984
  Node 2      -1.9457572546043367
  Node 3      0.9056530286201546
  Node 4      1.8363826263677057
  Node 5      -1.3142673143728163
  Node 6      -1.080649913401407
  Node 7      -1.0054113955272617
  Node 8      -1.4930625135264395
  Node 9      -2.564342500218616
  Node 10     -1.5541912075129076
  Node 11     -1.1016623490355641
  Node 12     1.9884467642861827
  Node 13     -1.7427303911639909
  Node 14     -2.633018644901748

Sigmoid Node 1
  Inputs      Weights
  Threshold   -0.3212247281186247
  Node 2      1.9731647746719587
  Node 3      -0.900772655443711
  Node 4      -1.8117190484180556
  Node 5      1.2997033137278047
  Node 6      1.1157847659720745
  Node 7      1.0300768232799697
  Node 8      1.4630415197255247
  Node 9      2.5761304376989425
  Node 10     1.4994634283579715
  Node 11     1.0972290618229457
  Node 12     -1.9903501569870632
  Node 13     1.779688139333916

```

EK-9 (Devam) Sinir ağıları algoritması sonuçları

Çizelge 9.1. (Devam) Sinir ağıları algoritması sonuç tablosu

Sigmoid Node 2	
Inputs	Weights
Threshold	-0.06945343791474078
Attrib YAS=24-35	0.43494843465859284
Attrib YAS=36-45	-0.17192002872652595
Attrib YAS=45+	-0.21883812671715222
Attrib SEMT=1	-0.8417536998499339
Attrib SEMT=2	0.07784672149105412
Attrib SEMT=3	0.8254991855032132
Attrib IS	-1.9660555015699899
Attrib PERİYOT=0-6	0.9349932371327012
Attrib PERİYOT=7-12	0.6860498981667753
Attrib PERİYOT=12+	-1.538508044942881
Attrib C_TIPI=1	0.42715753162619124
Attrib C_TIPI=2	-0.4176995235708546
Attrib C_TIPI=3	0.13293712737376107
Attrib BAKIM=B	0.0967629893803287
Attrib BAKIM=E	1.1279936828487274
Attrib BAKIM=H	-1.2017184893714046
Attrib U_TIPI=1	1.3703094338991015
Attrib U_TIPI=2	-1.4782695218906634
Attrib U_TIPI=3	0.6955785430253935
Attrib U_TIPI=4	-0.05209021936266501
Attrib U_TIPI=5	0.4617855852033837
Attrib U_TIPI=6	-0.7967018832673836
Attrib HARCAMA=100-250	-0.19705809378352263
Attrib HARCAMA=251-400	0.3904528247819342
Attrib HARCAMA=400+	-0.14446288436530375
Sigmoid Node 3	
Inputs	Weights
Threshold	-0.10477748795938827
Attrib YAS=24-35	0.02831092064601354
Attrib YAS=36-45	-0.3689118379184079
Attrib YAS=45+	0.4895863044995404
Attrib SEMT=1	-0.15447835474378557
Attrib SEMT=2	0.441885417462031
Attrib SEMT=3	-0.10446005790626955
Attrib IS	-0.05495345551296555
Attrib PERİYOT=0-6	-0.2749885087490871
Attrib PERİYOT=7-12	-0.09777257967335373
Attrib PERİYOT=12+	0.4926783335202455
Attrib C_TIPI=1	-0.2475739987215151
Attrib C_TIPI=2	0.20656466091897194
Attrib C_TIPI=3	0.1327380301297564
Attrib BAKIM=B	-0.162421828219382
Attrib BAKIM=E	0.10027253472434841
Attrib BAKIM=H	0.15434145842258315
Attrib U_TIPI=1	-0.6074604373432143
Attrib U_TIPI=2	0.36834245673482235
Attrib U_TIPI=3	-0.12925433059134195
Attrib U_TIPI=4	0.27824789704823816

EK-9 (Devam) Sinir ağıları algoritması sonuçları

Çizelge 9.1. (Devam) Sinir ağıları algoritması sonuç tablosu

Attrib U_TIPI=5	-0.08585805751159785
Attrib U_TIPI=6	0.6577657125075492
Attrib HARCAMA=100-250	0.3870716978529605
Attrib HARCAMA=251-400	-0.7408344631269036
Attrib HARCAMA=400+	0.4387670332922161
Sigmoid Node 4	
Inputs	Weights
Threshold	-0.10754722384883839
Attrib YAS=24-35	-0.23544190347602423
Attrib YAS=36-45	-0.09440101330385108
Attrib YAS=45+	0.45716383671612526
Attrib SEMT=1	0.12276052243088312
Attrib SEMT=2	0.7582692043519205
Attrib SEMT=3	-0.737349043260959
Attrib IS	-0.42262003755641747
Attrib PERİYOT=0-6	-0.9583877153904055
Attrib PERİYOT=7-12	-0.5145064388531398
Attrib PERİYOT=12+	1.5390138393406636
Attrib C_TIPI=1	0.5210266765368082
Attrib C_TIPI=2	-0.4473061333148283
Attrib C_TIPI=3	0.13052653623862204
Attrib BAKIM=B	-0.7399197572204943
Attrib BAKIM=E	0.25060938981905473
Attrib BAKIM=H	0.6656657263389043
Attrib U_TIPI=1	0.5053632358002299
Attrib U_TIPI=2	-0.25228535122071216
Attrib U_TIPI=3	-0.3909635229832687
Attrib U_TIPI=4	0.42853571185286526
Attrib U_TIPI=5	-0.4895459464985533
Attrib U_TIPI=6	0.7249446730356398
Attrib HARCAMA=100-250	-0.05510620906497867
Attrib HARCAMA=251-400	-0.2942849022284508
Attrib HARCAMA=400+	0.4247594074817989
Sigmoid Node 5	
Inputs	Weights
Threshold	0.20086254721837077
Attrib YAS=24-35	0.28461405648077626
Attrib YAS=36-45	0.12381851636597951
Attrib YAS=45+	-0.5479415996081907
Attrib SEMT=1	-0.7409126202656928
Attrib SEMT=2	-0.16679290035495734
Attrib SEMT=3	0.7560064115942355
Attrib IS	0.5733392798993487
Attrib PERİYOT=0-6	0.8093466820014106
Attrib PERİYOT=7-12	0.5219117840250559
Attrib PERİYOT=12+	-1.5312275144348657
Attrib C_TIPI=1	-0.2777426193525099
Attrib C_TIPI=2	0.23247465602860987
Attrib C_TIPI=3	-0.16530005545522727
Attrib BAKIM=B	0.404436201033162

EK-9 (Devam) Sinir ağıları algoritması sonuçları

Çizelge 9.1. (Devam) Sinir ağıları algoritması sonuç tablosu

Attrib BAKIM=E	-0.7599030874467222
Attrib BAKIM=H	0.16914523053524744
Attrib U_TIPI=1	-0.30290528605780914
Attrib U_TIPI=2	-0.10111373577017779
Attrib U_TIPI=3	0.3528826136875734
Attrib U_TIPI=4	-0.33933244523667183
Attrib U_TIPI=5	0.3397711855378113
Attrib U_TIPI=6	-0.7455898801483912
Attrib HARCAMA=100-250	-0.10028195081219603
Attrib HARCAMA=251-400	0.35913388169686433
Attrib HARCAMA=400+	-0.49275942800724376
Sigmoid Node 6	
Inputs	Weights
Threshold	-0.0781391106984504
Attrib YAS=24-35	0.3220074701843403
Attrib YAS=36-45	-0.04496432979838012
Attrib YAS=45+	-0.25210980463270366
Attrib SEMT=1	0.206817173318735
Attrib SEMT=2	-0.7574407780017443
Attrib SEMT=3	0.5459755345300148
Attrib IS	-0.1107456389043889
Attrib PERİYOT=0-6	0.7012803467043399
Attrib PERİYOT=7-12	0.4431645866214929
Attrib PERİYOT=12+	-1.11485468095828
Attrib C_TIPI=1	-0.4222422630680193
Attrib C_TIPI=2	0.3467902141969399
Attrib C_TIPI=3	0.11334359831479511
Attrib BAKIM=B	0.809379385177149
Attrib BAKIM=E	0.018328044134646396
Attrib BAKIM=H	-0.7740588799104897
Attrib U_TIPI=1	-0.2459485003994646
Attrib U_TIPI=2	0.18008191523443814
Attrib U_TIPI=3	0.3041502863660148
Attrib U_TIPI=4	-0.11898763120842973
Attrib U_TIPI=5	0.3815823729685208
Attrib U_TIPI=6	-0.358617691204637
Attrib HARCAMA=100-250	0.09603362243615969
Attrib HARCAMA=251-400	0.12031715226407917
Attrib HARCAMA=400+	-0.24460745205968468
Sigmoid Node 7	
Inputs	Weights
Threshold	0.017196702252009755
Attrib YAS=24-35	0.3168135781591262
Attrib YAS=36-45	-0.056433802322102
Attrib YAS=45+	-0.21141419734550437
Attrib SEMT=1	0.1513532123239624
Attrib SEMT=2	-0.646868089009625
Attrib SEMT=3	0.5543621149060187
Attrib IS	-0.03577435700782824
Attrib PERİYOT=0-6	0.7071070004305756

EK-9 (Devam) Sinir ağıları algoritması sonuçları

Çizelge 9.1. (Devam) Sinir ağıları algoritması sonuç tablosu

Attrib PERİYOT=7-12	0.44783614596034055
Attrib PERİYOT=12+	-1.114033674388799
Attrib C_TIPI=1	-0.33314026852548934
Attrib C_TIPI=2	0.36602816169967595
Attrib C_TIPI=3	0.0930432416966726
Attrib BAKIM=B	0.7339649379256014
Attrib BAKIM=E	-0.050024795887116676
Attrib BAKIM=H	-0.6532015290918723
Attrib U_TIPI=1	-0.23145447095273264
Attrib U_TIPI=2	0.16816949694673966
Attrib U_TIPI=3	0.2733446033696895
Attrib U_TIPI=4	-0.16872647695551363
Attrib U_TIPI=5	0.41450624101129024
Attrib U_TIPI=6	-0.4124301828735091
Attrib HARCAMA=100-250	0.09311606263247556
Attrib HARCAMA=251-400	0.18998798837724293
Attrib HARCAMA=400+	-0.28198408331946373
Sigmoid Node 8	
Inputs	Weights
Threshold	0.010745149196610009
Attrib YAS=24-35	0.30115912644395476
Attrib YAS=36-45	-0.10894136928420191
Attrib YAS=45+	-0.2822908848764186
Attrib SEMT=1	-0.46225943578866685
Attrib SEMT=2	-0.171236116788469
Attrib SEMT=3	0.6267964592411784
Attrib IS	-1.6088651500162992
Attrib PERİYOT=0-6	0.8257580910999345
Attrib PERİYOT=7-12	0.5009754294534571
Attrib PERİYOT=12+	-1.3609347803000507
Attrib C_TIPI=1	0.4647160565150992
Attrib C_TIPI=2	-0.5794463416257138
Attrib C_TIPI=3	0.05311946228605292
Attrib BAKIM=B	0.14948169029091254
Attrib BAKIM=E	0.3198352848208517
Attrib BAKIM=H	-0.4146968796223247
Attrib U_TIPI=1	1.1347146655133946
Attrib U_TIPI=2	-1.1361129546223476
Attrib U_TIPI=3	0.49005658747759784
Attrib U_TIPI=4	-0.08908185545231628
Attrib U_TIPI=5	0.32536503236686654
Attrib U_TIPI=6	-0.7195007591477666
Attrib HARCAMA=100-250	-0.3648448087626752
Attrib HARCAMA=251-400	0.5685117637177916
Attrib HARCAMA=400+	-0.20406444814478056
Sigmoid Node 9	
Inputs	Weights
Threshold	-0.009456453647046871
Attrib YAS=24-35	0.593775513859011
Attrib YAS=36-45	0.07918371541736571
Attrib YAS=45+	-0.6340672008662751

EK-9 (Devam) Sinir ağıları algoritması sonuçları

Çizelge 9.1. (Devam) Sinir ağıları algoritması sonuç tablosu

Attrib SEMT=1	-1.9360975332026114
Attrib SEMT=2	0.818900098354295
Attrib SEMT=3	1.0106717658337092
Attrib IS	0.3626211211224931
Attrib PERİYOT=0-6	0.8117675386337435
Attrib PERİYOT=7-12	0.8240660183947743
Attrib PERİYOT=12+	-1.6264996957664077
Attrib C_TIPI=1	0.07761393217205828
Attrib C_TIPI=2	-0.1550441839812447
Attrib C_TIPI=3	0.1254241798682948
Attrib BAKIM=B	0.05081052467536997
Attrib BAKIM=E	-1.648712294559106
Attrib BAKIM=H	1.555791875696148
Attrib U_TIPI=1	0.01266128200710541
Attrib U_TIPI=2	-0.7041497810771614
Attrib U_TIPI=3	0.6696394391576104
Attrib U_TIPI=4	-0.07296237652148675
Attrib U_TIPI=5	0.6369158193452935
Attrib U_TIPI=6	-0.47233402373297795
Attrib HARCAMA=100-250	0.10199159165404796
Attrib HARCAMA=251-400	0.4111670576913264
Attrib HARCAMA=400+	-0.500486961961303
Sigmoid Node 10	
Inputs	Weights
Threshold	-0.023289487597000023
Attrib YAS=24-35	0.2839487906270984
Attrib YAS=36-45	-0.007642879228938684
Attrib YAS=45+	-0.2965935325387297
Attrib SEMT=1	0.1454629990558151
Attrib SEMT=2	-0.8545383337035122
Attrib SEMT=3	0.6500086455710103
Attrib IS	-0.05808133223624697
Attrib PERİYOT=0-6	0.7882727509416837
Attrib PERİYOT=7-12	0.5016063160427146
Attrib PERİYOT=12+	-1.3157297650930422
Attrib C_TIPI=1	-0.49300456025317707
Attrib C_TIPI=2	0.4164291468322475
Attrib C_TIPI=3	0.03406118665446778
Attrib BAKIM=B	0.8625637131674405
Attrib BAKIM=E	-0.024404473704008424
Attrib BAKIM=H	-0.8475952062678299
Attrib U_TIPI=1	-0.37055441793657123
Attrib U_TIPI=2	0.21639157099107365
Attrib U_TIPI=3	0.31805003894915146
Attrib U_TIPI=4	-0.22674693918883973
Attrib U_TIPI=5	0.45129755422999374
Attrib U_TIPI=6	-0.4969917757268229
Attrib HARCAMA=100-250	0.0946106105413272
Attrib HARCAMA=251-400	0.1328651069110946
Attrib HARCAMA=400+	-0.2658979128691288

EK-9 (Devam) Sinir ağıları algoritması sonuçları

Çizelge 9.1. (Devam) Sinir ağıları algoritması sonuç tablosu

Sigmoid Node 11	
Inputs	Weights
Threshold	0.03568507689473812
Attrib YAS=24-35	0.39992986908170874
Attrib YAS=36-45	-0.09498545507523218
Attrib YAS=45+	-0.40670408787547285
Attrib SEMT=1	-1.960046401836919
Attrib SEMT=2	0.9927364165248326
Attrib SEMT=3	0.9676158279517911
Attrib IS	-0.6261590316285829
Attrib PERİYOT=0-6	1.1050850382544224
Attrib PERİYOT=7-12	0.7057828695170869
Attrib PERİYOT=12+	-1.8760425245539838
Attrib C_TIPI=1	0.13585886033769362
Attrib C_TIPI=2	-0.25620358925130954
Attrib C_TIPI=3	0.017865459591917564
Attrib BAKIM=B	-0.25022985863703623
Attrib BAKIM=E	0.2542311916087276
Attrib BAKIM=H	-0.14764122997917783
Attrib U_TIPI=1	1.1877749651967804
Attrib U_TIPI=2	-1.5573967492120167
Attrib U_TIPI=3	0.6807713047267309
Attrib U_TIPI=4	-0.13200886759027877
Attrib U_TIPI=5	0.544291889681948
Attrib U_TIPI=6	-0.946482293051673
Attrib HARCAMA=100-250	-0.0708129988571006
Attrib HARCAMA=251-400	0.22479739140595953
Attrib HARCAMA=400+	-0.3301587642678828
Sigmoid Node 12	
Inputs	Weights
Threshold	-0.19887800170375575
Attrib YAS=24-35	-0.23128551234816647
Attrib YAS=36-45	-0.20806738452856943
Attrib YAS=45+	0.6413904550212581
Attrib SEMT=1	0.6043339778583592
Attrib SEMT=2	0.41468118914734986
Attrib SEMT=3	-0.8397399908298709
Attrib IS	-0.7386103605684599
Attrib PERİYOT=0-6	-1.0417475685645354
Attrib PERİYOT=7-12	-0.4435151693703761
Attrib PERİYOT=12+	1.7488868303444596
Attrib C_TIPI=1	0.5347518826474738
Attrib C_TIPI=2	-0.4797261957777027
Attrib C_TIPI=3	0.2200776576537443
Attrib BAKIM=B	-0.6861485104173313
Attrib BAKIM=E	0.561993794050338
Attrib BAKIM=H	0.3260564874261092
Attrib U_TIPI=1	0.5312470181665276
Attrib U_TIPI=2	-0.07301647111465259

EK-9 (Devam) Sinir ağıları algoritması sonuçları

Çizelge 9.1. (Devam) Sinir ağıları algoritması sonuç tablosu

Attrib U_TIPI=4	0.5422293846632631
Attrib U_TIPI=3	-0.5302878343008685
Attrib U_TIPI=5	-0.48615645555034126
Attrib U_TIPI=6	0.9657700839255797
Attrib HARCAMA=100-250	0.07704279011042153
Attrib HARCAMA=251-400	-0.4589709114409424
Attrib HARCAMA=400+	0.6537975554014381
Sigmoid Node 13	
Inputs	Weights
Threshold	0.25782443033323704
Attrib YAS=24-35	0.22574984673324436
Attrib YAS=36-45	0.237269651267747
Attrib YAS=45+	-0.7173632658853958
Attrib SEMT=1	-0.9511534951852055
Attrib SEMT=2	-0.14954269220365188
Attrib SEMT=3	0.8149289252811391
Attrib IS	0.7340713738170167
Attrib PERİYOT=0-6	0.9821486759484764
Attrib PERİYOT=7-12	0.5485005330706066
Attrib PERİYOT=12+	-1.7574907399918123
Attrib C_TIPI=1	-0.2871635125865973
Attrib C_TIPI=2	0.22825446625885001
Attrib C_TIPI=3	-0.23274050415975814
Attrib BAKIM=B	0.5590461073084758
Attrib BAKIM=E	-1.0322596880755448
Attrib BAKIM=H	0.14899142388372624
Attrib U_TIPI=1	-0.5080391158124082
Attrib U_TIPI=2	-0.03386291627459638
Attrib U_TIPI=3	0.47771002286862707
Attrib U_TIPI=4	-0.45506741547719876
Attrib U_TIPI=5	0.3499026011696095
Attrib U_TIPI=6	-0.9213400737000731
Attrib HARCAMA=100-250	-0.11109850827757065
Attrib HARCAMA=251-400	0.5197685481735496
Attrib HARCAMA=400+	-0.6242861665407085
Sigmoid Node 14	
Inputs	Weights
Threshold	-0.012033876163292068
Attrib YAS=24-35	0.568188348458061
Attrib YAS=36-45	0.09252095877155496
Attrib YAS=45+	-0.6961663535017795
Attrib SEMT=1	-1.9403688250721398
Attrib SEMT=2	0.8606743588076436
Attrib SEMT=3	1.0796957328861083
Attrib IS	0.3515139097499401
Attrib PERİYOT=0-6	0.7159999729713775
Attrib PERİYOT=7-12	0.8686491201415835
Attrib PERİYOT=12+	-1.4730665545548376
Attrib C_TIPI=1	0.04895563142601774
Attrib C_TIPI=2	-0.1843978683980602

EK-9 (Devam) Sinir ağıları algoritması sonuçları

Çizelge 9.1. (Devam) Sinir ağıları algoritması sonuç tablosu

Attrib C_TIPI=3	0.1302863337764201
Attrib BAKIM=B	0.16688073922604052
Attrib BAKIM=E	-1.6823601747188814
Attrib BAKIM=H	1.5634665901017517
Attrib U_TIPI=1	-0.026401144632142007
Attrib U_TIPI=2	-0.6725032585120351
Attrib U_TIPI=3	0.6867915899450306
Attrib U_TIPI=4	-0.08422749034062581
Attrib U_TIPI=5	0.7011943334154113
Attrib U_TIPI=6	-0.4340664609671582
Attrib HARCAMA=100-250	0.12581345949337827
Attrib HARCAMA=251-400	0.4433304032025088
Attrib HARCAMA=400+	-0.5304361435347345

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : AKBULUT, Sinem
Uyruğu : T.C.
Doğum tarihi ve yeri : 17.04.1979 Samsun
Medeni hali : Bekar
Telefon : 0 (312) 436 79 56
e-mail : sinem.akbulut@gmail.com

Eğitim

Derece

Lisans

Lise

Eğitim Birimi

Gazi Üniversitesi/ Endüstri Müh., Ankara

Tülay Başaran Lisesi , Samsun

Mezuniyet tarihi

2002

1997

İş Deneyimi

Yıl

2003-2006

Yer

Türk Telekomünikasyon A.Ş

Görev

T.Uzman Yrd.

Yabancı Dil

İngilizce

Hobiler

Bilgisayar teknolojileri, Kişisel gelişim