

T.C.
MARMARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İLETİŞİM BİLİMLERİ ANABİLİM DALI
BİLİŞİM BİLİM DALI

**BİREYSEL İLETİŞİME DAYALI PAZARLAMADA
MÜŞTERİ BAĞIMLILIĞI: GSM OPERATÖRLERİ
ÜZERİNE BİR ÇALIŞMA**

Doktora Tezi

SABRİ SERKAN GÜLLÜOĞLU

İstanbul, 2012

T.C.
MARMARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İLETİŞİM BİLİMLERİ ANABİLİM DALI
BİLİŞİM BİLİM DALI

**BİREYSEL İLETİŞİME DAYALI PAZARLAMADA
MÜŞTERİ BAĞIMLILIĞI: GSM OPERATÖRLERİ
ÜZERİNE BİR ÇALIŞMA**

Doktora Tezi

SABRİ SERKAN GÜLLÜOĞLU

Danışman: PROF.DR. ÖZHAN TINGÖY

İstanbul, 2012

Marmara Üniversitesi
Sosyal Bilimler Enstitüsü Müdürlüğü

Tez Onay Belgesi

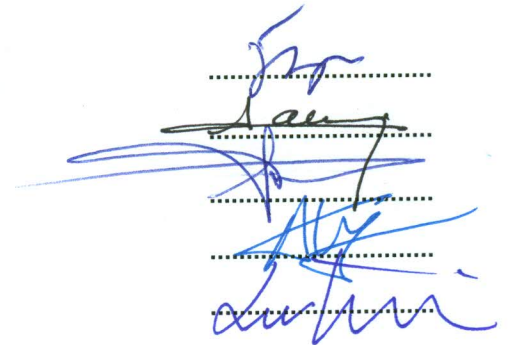
İLETİŞİM BİLİMLERİ Anabilim Dalı BİLİŞİM Bilim Dalı Doktora öğrencisi
SABRİ SERKAN GÜLLÜOĞLU'nun BİREYSEL İLETİŞİME DAYALI PAZARLAMADA
MÜŞTERİ BAĞIMLILIĞI : GSM OPERATÖRLERİ ÜZERİNE BİR ÇALIŞMA adlı tez
çalışması ,Enstitümüz Yönetim Kurulunun 04.11.2011 tarih ve 2011-23/39 sayılı
kararıyla oluşturulan jüri tarafından oybirliği/oyçokluğu ile Doktora Tezi olarak kabul
edilmiştir.

Öğretim Üyesi Adı Soyadı

İmzası

Tez Savunma Tarihi : 18.01.2012

- 1) Tez Danışmanı : PROF. DR. ÖZHAN TINGÖY
2) Jüri Üyesi : PROF. DR. TÜRKAN UĞUR DAİ
3) Jüri Üyesi : PROF. DR. SUAT GEZGİN
4) Jüri Üyesi : PROF. DR. AYKUT TOP
5) Jüri Üyesi : YRD. DOÇ. DR. LEVENT ELDENİZ



GENEL BİLGİLER

İsim ve Soyadı	:Sabri Serkan Güllüoğlu
Anabilim Dalı	:İletişim Bilimleri
Programı	:Bilişim
Tez Danışmanı	:Prof. Dr. Özhan Tıngöy
Tez Türü ve Tarihi	:Doktora – 2012
Anahtar Kelimeler	:GSM, Müşteri Kaybı, Müşteri İlişkileri Yönetimi

ÖZET

BİREYSEL İLETİŞİME DAYALI PAZARLAMADA MÜŞTERİ BAĞIMLILIĞI: GSM OPERATÖRLERİ ÜZERİNE BİR ÇALIŞMA

Günümüzde müşteriler için ürün ve hizmetlere erişmek daha kolay hale gelmiştir. Müşteriler belirli bir markayı seçmeden önce ürünün fiyatına, yeniliğine ve sunulan ek hizmetlere bakmaktadır. İşletmeler rekabet avantajı elde etmek, müşteri bağlılığını artırmak için yeni yollar aramaktadır. Bu yollardan biri de müşteriler ile uzun dönemli ilişki kurmak yoluyla ürün ve hizmetlerin farklılaştırıldığı “Müşteri İlişkileri Yönetimi”ni kullanmaktır. Müşteri bağlılığı, diğer sektörlerde olduğu gibi GSM sektörü için de önemlidir. Bunun temel nedeni, GSM sektörünün olgunluk aşamasına gelmesi sonucu yeni müşterilerin (abonelerin) kazanılmasının zor ve maliyetli olması ve GSM operatörlerinin mevcut müşterilerinin, diğer operatörlerin potansiyel müşterileri olmasıdır. Müşteri kaybına “churn” denmektedir. Önce bir GSM operatöründen diğerine geçme için kullanılırken şimdi sektör gözetmeksizin rakibe kaçan tüm müşteriler için kullanılmaya başlanmıştır. Pazar payını korumak, hatta ileriye götürmek için, mevcut müşterilere eğilimin daha verimli (az maliyetli) olduğundan yola çıkarak, müşteri kaybının engellenmesi çok önemlidir. Bu çalışmada öncelikle veri madenciliği altındaki Fuzzy(Bulanık) Kümeleme tekniklerinden, amaç fonksiyonlarına dayanan bütün kümeleme tekniklerinin temelini oluşturan Fuzzy(Bulanık) Ortalamalar tekniği incelenmiş ve bu teknik kullanılarak Türkiye'deki bir GSM Operatörünün müşteri kaybı analizi yapılarak şirket yöneticilerin mevcut müşterilerin durumlarını görebilecekleri kurallar çıkarılmıştır.

GENERAL KNOWLEDGE

Name and Surname	:Sabri Serkan Güllüoğlu
Field	:Communication Sciences
Programme	:Informatics
Supervisor	:Prof. Dr. Özhan Tıngöy
Degree Awarded and Date	:Ph.D.– 2012
Keywords	:GSM, Churn, Customer Relationship Management

ABSTRACT

CUSTOMER LOYALTY IN MARKETING WHICH BASED ON INDIVIDUAL COMMUNICATION: A STUDY ON GSM OPERATORS

It has become easier for customers to access products and services. Customers have been looking at the price, the novelty and additional services provided of a product before selecting a brand.. The businesses have been looking for new ways so as to get competition advantage increasing customers' allegiance. One of the ways to establish this relationship is by using customer relationship management in which the products and services are diversified. Client loyalty is important in GSM industry and also important as in other industries. The fundamental for this reason is that attracting new subscribers or customers is not only difficult but also costly due to the GSM industries' saturated customer portfolio. The second reason is that the current customers of GSM operators are also potential customers' of the competitor operators. The loss of customers is called "churn". Initially it was used for a customer who switches to another GSM operator but now it has been used for all the customers who moved to competitor firms, regardless of their industry. For safe guarding market and even moving it beyond, because leaning to customers is more efficient(less costly), preventing customer loss is important. Fuzzy Means, which is one of the fuzzy clustering techniques under data mining and which constitute the foundation of all clustering techniques depending on objective function, is initially examined in this study. Current customers' status is displayed to firm managers as a result of customers churn analysis of a GSM operator by using this technique in Turkey.

ÖNSÖZ

Çalışmalarım boyunca değerli yardım ve katkılarıyla beni yönlendiren ve anlayışını hiçbir zaman esirgemeyen saygıdeğer Hocam Prof. Dr. Özhan Tıngöy'e ve kıymetli tecrübelerinden faydalandığım çok değerli hocalarım sayın öğretim üyeleri başta Prof. Dr. Türkan Uğur Dai olmak üzere, Prof. Dr. Murat Özgen, Prof. Dr. Kadir Aykut Top ve Yrd. Doç. Dr. Levent Eldeniz'e teşekkürü bir borç bilirim. Doktora öğrenimim sırasında güleryüzlü ve esprilili tutumuyla beni çalışmalarımda bir kat daha teşvik eden sayın hocam Prof. Dr. Suat Gezgin'e de saygılarımı sunarım.

Bunun yanında, çalışmamın hazırlanması sürecinde bana olan inanç, güven ve motivasyonunu hiç eksik etmeyen sevgili annem Ayşe Güllüoğlu'na teşekkür ederim.

Ayrıca, destek, hoşgörü ve sabrını benden hiçbir zaman esirgemeyen sevgili ve çok kıymetli eşim Nilgün Güllüoğlu'na sonsuz şükranlarımı sunarım.

Yoğun emekler ve çabalar ile oluşturduğum bu çalışmamın bundan sonra yapılacak çalışmalara referans teşkil etmesini ve yararlı olmasını temenni ederim.

2006 yılı ekim ayında kaybettiğimiz babama atfen...

İstanbul, 2012

Sabri Serkan GÜLLÜOĞLU

İÇİNDEKİLER

	Sayfa No
ÖZET	III
ABSTRACT	IV
ÖNSÖZ	V
TABLO LİSTESİ.....	X
ŞEKİL LİSTESİ.....	XII
GRAFİKLER.....	XIV
KISALTMALAR.....	XV
1. GİRİŞ	1
2. TELEKOMÜNİKASYON	6
2.1 Türkiye'de Telekomünikasyon Sektörü	8
2.1.1 Temel Telekomünikasyon Hizmetleri	15
2.1.2 Katma Değerli Telekomünikasyon Hizmetleri	17
2.1.3 Telekomünikasyon Cihazları	17
2.2 Sektörde Türk Telekom'un Yeri ve Önemi	19
2.3 Türk Telekom'un Tarihsel Gelişimi, Hukuki Yapısı	24
3. VERİ MADENCİLİĞİ	36
3.1 Veri Madenciliği	41
3.1.1 Veritabanları	46
3.1.1.1 Veri Ambarları	49
3.1.1.2 OLTP Sistemleri	51
3.1.1.3 OLAP Sistemleri.....	51
3.2 Veri Madenciliği Aşamaları.....	61

3.2.1 Sınıflama Teknikleri	62
3.2.1.1 Karar Ağaçları.....	64
3.2.1.1.1 Karar Ağaçlarında Algoritmalar	67
3.2.1.1.2 Karar Ağaçlarında Budama İşlemleri	69
3.2.1.2 Genetik Algoritmalar	69
3.2.1.2.1 Genetik Algoritmaların Gelişimi	70
3.2.1.2.2 Genetik Algoritma Terminolojisi.....	75
3.2.1.2.3 Genetik Algoritma Operatörleri.....	85
3.2.1.2.4 GA'nın Geleneksel Optimizasyon Yöntemleri ile Karşılaştırılması .	105
3.2.1.2.5 Genetik Algoritmaların Uygulama Alanları	107
3.2.1.2.6 Genetik Algoritmaların Sınıflandırılması	109
3.2.1.3 Yapay Sinir Ağları	114
3.2.1.3.1 Biyolojik Sinir Sistemleri	116
3.2.1.3.2 Yapay Sinir Ağı Uygulama Alanları	123
3.2.1.4 K-En yakın Komşu Algoritması	124
3.2.1.5 Naive Bayes	130
3.2.1.5.1 Naive Bayes Kuralının Çalışma Yapısı	131
3.2.1.5.2 Naive Bayes Yöntemi İle İlgili Örnek Uygulamalar	134
3.2.2 Birliktelik Kuralları.....	141
3.2.2.1 Birliktelik Kurallarında Kullanılan Terimler	144
3.2.2.2 Apriori Algoritması.....	146
3.2.3 Kümeleme Analizi	152
3.2.4 Kümeleme Yaklaşımları	153
3.2.4.1 Hiyerarşik Kümeleme Yaklaşımı.....	153
3.2.4.2 Toplanmış Hiyerarşik Kümeleme Yöntemleri.....	154
3.2.4.3 Bölümlemeli Kümeleme Yaklaşımı.....	155
3.2.4.3.1 Klasik k-ortalamalar algoritması (KOA)	157
3.2.4.3.2 Klasik k-medoid algoritması (KMA).....	160
3.2.4.4 Yoğunluğa Dayalı Kümeleme Yaklaşımı	162
3.2.4.5 Fuzzy (Bulanık) Kümeleme Yaklaşımları	163
3.2.4.6 Geleneksel Fuzzy(Bulanık) Kümeleme Algoritmaları	165

3.2.4.6.1 Fuzzy(Bulanık) C-ortalamalar algoritması (BCO)	166
3.2.4.6.2 Gustafson-Kessel Algoritması (GKA)	167
3.2.4.6.3 Gath-Geva Algoritması (GGA)	167
4. MÜŞTERİ KAYIP YÖNETİMİ.....	168
4.1 Müşteri İlişkileri Yönetimi (CRM).....	168
4.1.1 Pazarlama.....	172
4.1.2 Marka Bağımlılığı.....	176
4.1.2.1 Davranış Olarak Marka Bağlılığı (Behavioral Brand Loyalty)	177
4.1.2.2 Tutum Olarak Marka Bağlılığı (Attitudinal Brand Loyalty)	178
4.2 Marka Bağlılığı Yaratma	180
4.2.1 Marka Müşteri İlişkisi.....	181
4.2.2 Müşteri Hayat Seyri Ölçütleri.....	182
4.2.3 Güven ve Marka Bağlılığı.....	185
4.2.4 Sadakat ve Marka Bağlılığı.....	186
4.2.5 Müşteri Değeri ve Marka Bağlılığı.....	189
4.2.6 Müşteri Tatmini ve Marka Bağlılığı	190
4.3 İlişkisel Pazarlama	194
5. GSM OPERATÖRLERİ ÜZERİNE BİR ÇALIŞMA.....	197
5.1 Araştırmanın Amacı ve Önemi	198
5.2 Araştırmanın Kısıtları	199
5.3 Literatür Araştırması.....	201
5.4 Materyal	205
5.4.1 Verilerin Hazırlanması ve Analizi	206
5.4.2 Veri Anlaşılabilirliği	209
5.5 Araştırma Modelinin Kurulması ve Değerlendirilmesi	210
5.5.1 Modelin İzlenmesi	211
5.5.2 Modelin Kullanılması	211
5.5.2.1 C Ortalamaları Algoritması	212
5.5.2.2 Bulanık C Ortalamaları Algoritması.....	213
5.5.2.3 ANFIS - Adaptif Ağ Yapısına Dayalı Bulanık Çıkarım Sistemi.....	214

5.6 Bulgular	215
5.6.1 Veri Analiz Aşaması	215
5.6.2 Kümeleme Aşaması	225
5.6.3 Tahmin Aşaması	231
5.7 Araştırmanın Sonucu	240
6. SONUÇ	242
EKLER	247
EK 1: TELEKOMÜNİKASYON VERİSİ	248
EK 2: RANK (SIRALAMA) TANIMI	250
EK 3: WEKA PROGRAMI HAKKINDA GENEL BİLGİLER	254
KAYNAKÇA	258
ÖZGEÇMİŞ	273

TABLO LİSTESİ

Sayfa No

Tablo 2. 1 Türkiye Telekomünikasyon Tarihsel Süreci	10
Tablo 2. 2 Türkiye Telekom Pazarı Büyüklüğü	24
Tablo 2. 3 Telekomünikasyon Sektöründe 1983-2012 Yılları Arasında Gelişmeler	26
Tablo 3. 1 Özbilgi Keşif Süreci	43
Tablo 3. 2 Oltp ve Olap Sistemlerinin Karşılaştırılması.....	57
Tablo 3. 3 Sınıflandırma Modelini Etkileyen Durumlar.....	63
Tablo 3. 4 Genetik Algoritmalar İkili Kodlama	82
Tablo 3. 5 Genetik Algoritmalar Sekizli(octal) Kodlama.....	83
Tablo 3. 6 Genetik Algoritmalar Hekzadesimal Kodlama	83
Tablo 3. 7 Genetik Algoritmalar Permutasyon Kodlama	83
Tablo 3. 8 Genetik Algoritmalar Değer Kodlama	84
Tablo 3. 9 Genetik Algoritmalar Tek Noktalı Çaprazlama	93
Tablo 3. 10 Genetik Algoritmalar Çift Noktalı Çaprazlama	94
Tablo 3. 11 Genetik Algoritmalar Uniform Çaprazlama.....	96
Tablo 3. 12 Genetik Algoritmalar Döndürme (flipping)	98
Tablo 3. 13 Genetik Algoritmalar Değiş Tokuş	99
Tablo 3. 14 Sinir Sistemi ile Yapay Sinir Ağlarının Benzerlikleri.....	118
Tablo 3. 15 Örnek Gözlem Değerleri	128
Tablo 3. 16 Örnek Gözlem Değerleri Öklit Uzaklık Değerleri	129
Tablo 3. 17 Şutların Veri Kümesi.....	135
Tablo 3. 18 Şut Şekillerine Göre Gol Durum Frekansları	135
Tablo 3. 19 Gol Durum Frekanslarına Göre Gol Olasılıkları	136
Tablo 3. 20 Adayların Kabul Durumları Veri Kümesi	138
Tablo 3. 21 Aday Bilgilerine Göre Kabul Durum Frekansları	139
Tablo 3. 22 Aday Bilgi Frekanslarına Göre Kabul Durum Olasılıkları.....	139
Tablo 3. 23 Ürünlere ait Id'ler ve Birlikte Satın Alınma Durumları.....	148
Tablo 3. 24 Müşteri İlişkileri Yönetimi Örneği	190

Tablo 5. 1 Yıllara Göre Yapılan Müşteri Kayıp Yönetimi Yayın Sayısı	202
Tablo 5. 2 Yıllara Göre Yapılan Müşteri Kayıp Tahmin Yayınlarının Sayısı.....	202
Tablo 5. 3 Yıllara Göre Yapılan Veri Madenciliği Yayınları Sayısı	203
Tablo 5. 4 Anfis Araştırma Yayınları Sayısı	203
Tablo 5. 5 Veri Kümesindeki Değişkenler	208
Tablo 5. 6 İlk 7 Özniteliğin Analizi.....	216
Tablo 5. 7 İkinci 7 Özniteliğin Analizi	216
Tablo 5. 8 Üçüncü 7 Özniteliğin Analizi.....	217
Tablo 5. 9 Dördüncü 7 Özniteliğin Analizi	217
Tablo 5. 10 Beşinci 7 Özniteliğin Analizi	218
Tablo 5. 11 Son altı Özniteliğin Analizi	219
Tablo 5. 12 Müşteri Medeni Durum Değerleri	219
Tablo 5. 13 Müşteri Cinsiyet Değerleri	220
Tablo 5. 14 Müşteri Eğitim Değerleri.....	220
Tablo 5. 15 Adres Değerleri	220
Tablo 5. 16 Uzmanlık Alanı Değerleri	221
Tablo 5. 17 Öznitelikler ve Rank(sıralama) Değerleri	225
Tablo 5. 18 Bulanık C ortalama ve K Ortalama Sınıflandırma Algoritmalarının Dağılımları	227
Tablo 5. 19 İki Farklı Doğruluk İstatistiği: K Ortalama ve Bulanık C Ortalama	227
Tablo 5. 20 K Ortalama ve Bulanık C Ortalama Algoritmalarının Doğru Pozitif Yanlış Pozitif Doğru Negatif Yanlış Negatif Değerleri	228
Tablo 5. 21 K ve Bulanık C Ortalamaları Sonuçları Arasındaki Lineer Korelasyon ...	229
Tablo 5. 22 Bulanık C Ortalaması İstatistik Bilgileri	231
Tablo 5. 23 Kullanılan Yöntem ve Eğitim Sonuçları	236
Tablo 5. 24 Olasılıksal Yapay Sinir Ağları Sonrası Oluşan Kurallar	238

ŞEKİL LİSTESİ

Sayfa No

Şekil 2. 1 Basit Yerel Ağ (local loop).....	15
Şekil 3. 1 Veritabanı Teknolojisinin Gelişimi ve Veri madenciliği	48
Şekil 3. 2 Veri Ambarı Kullanım Süreci.....	50
Şekil 3. 3 Veri Madenciliği ve Olap	54
Şekil 3. 4 Veri Madenciliği ile Diğer Disiplinlerin Gösterimi	58
Şekil 3. 5 Karar Ağaçları Gösterimi	66
Şekil 3. 6 Genetik Algoritma Döngüsü.....	73
Şekil 3. 7 Alel, Gen ve Kromozomlar	76
Şekil 3. 8 Genetik Algoritmalar Ağaç Kodlama.....	85
Şekil 3. 9 Biyolojik Sinir Sistemi	116
Şekil 3. 10 Sinir Hücresi	117
Şekil 3. 11 K en Yakın Komşu Algoritması	126
Şekil 3. 12 Bölümlemeli Kümelemenin Akış Diyagramı	156
Şekil 3. 13 Bir Nesne Grubunun Uzayda Konumlanmış Hali	159
Şekil 3. 14 Oi ve Oj İki Ayrı Kümenin Medoidleri	161
Şekil 3. 15 Müşteri Tipi Hiyerarşisi.....	174
Şekil 3. 16 Türkiye'deki Müşteri Tiplerinin Oransal Gösterimi	175
Şekil 3. 17 Müşteri Hayat Seyri.....	183
Şekil 5. 1 Bulanık C ve K Ortalamaları Karşılaştırılması İşlemi.....	226
Şekil 5. 2 K Ortalama Sınıflarının Dağılımı	229
Şekil 5. 3 Bulanık C Ortalama Sınıflarının Dağılımı.....	230
Şekil 5. 4 Eğitilmiş Verilerin Anfis Sınıflandırması I	231
Şekil 5. 5 Eğitilmiş Verilerin Anfis Sınıflandırması II	232
Şekil 5. 6 Kontrol Verilerinin Anfis Sınıflandırması.....	232
Şekil 5. 7 Bulanık Çıkarım Sisteminin Anfis Modeli	233
Şekil 5. 8 Bulanık Çıkarım Sisteminin Şeması.....	233

Şekil 5. 9 Equipment Üyelik Fonksiyonu (equipmon)	234
Şekil 5. 10 Long Distance Over Tenure Üyelik Fonksiyonu(longten)	234
Şekil 5. 11 Toll Free Over Tenure Üyelik Fonksiyonu(toll ten)	235
Şekil 5. 12 Equipment Over Tenure Üyelik Fonksiyonu (equipten)	235
Şekil 5. 13 Log Equipment Üyelik Fonksiyonu(losequi)	235
Şekil 5. 14 Karar Ağacı Tahmin Sonucu	237

GRAFİKLER

	Sayfa No
Grafik 3. 1 K en Yakın Komşu Algoritması Grafiksel Gösterimi	125
Grafik 5. 1 Bölge Gösterimi	222
Grafik 5. 2 Kullanım Süresi Değerleri	222
Grafik 5. 3 Yaş Değerleri	222
Grafik 5. 4 Medeni Durum Değerleri	223
Grafik 5. 5 Adres Değerleri	223
Grafik 5. 6 Gelir Değerleri	223
Grafik 5. 7 Eğitim Durumu Değerleri	223
Grafik 5. 8 İş Değerleri	224
Grafik 5. 9 Emeklilik Değerleri	224
Grafik 5. 10 Cinsiyet Durumu Değerleri	224
Grafik 5. 11 İkamet Değerleri	225

KISALTMALAR

ADSL	Asimetrik Sayısal Abone Hattı
A.Ş.	Anonim Şirket
ATM	Eşzamansız Aktarım Modu
Avro	Avrupa Birliği Ortak Para Birimi
Bkz.	Bakınız
Çev.	Çeviri
DPT	Devlet Planlama Teşkilatı
EEE	Verimlilik, Etkililik, Tutumluluk
GA	Genetik Algoritmalar
GMPCS	Mobil Uydu Telefonu
GPRS	Paket Anahtarlama Radyo Hizmetleri
GSM	Evrensel karasal gezgin telekomünikasyon sistemi.
IDC	İnternet veri merkezi
İP	İnternet Protokolü
ISDN	Tümleşik Hizmetler Sayısal Şebekesi
İSS	İnternet Servis Sağlayıcı
ITU	Uluslararası Telekomünikasyon Birliği
KOI	Kamu Ortaklığı İdaresi
MERNİS	Merkezi Nüfus İdaresi Sistemi
MMS	Multimedya Mesaj Servisi
No	Numero
PSTN	Anahtarlama Kamusal Telefon Şebekesi
PTT	Posta Telgraf ve Telefon İdaresi

s	Sayfa
SMS	Kısa Mesaj Servisi
TBMM	Türkiye Büyük Millet Meclisi
T.C.	Türkiye Cumhuriyeti
TT	Türk Telekomünikasyon A.Ş.
TTK	Türk Ticaret Kanunu
TÜRKEP	Türk Telekom İnternet Erişim Şebeke Altyapısı
TV	Türkiye Elektronik Posta Sistemi
vb.	Televizyon ve benzeri
VET	Verimlilik, Etkililik, Tutumluluk
VoIP	İnternet üzerinden ses iletimi
WAP	Kablosuz Uygulama Protokolü
YSA	Yapay Sinir Ağları

1. GİRİŞ

Günümüzde müşteriler için ürün ve hizmetlere erişmek daha kolay hale gelmiştir. Müşteriler belirli bir markayı seçmeden önce ürünün fiyatına, yeniliğine ve sunulan ek hizmetlere bakmaktadırlar. Artan rekabet ve bunun sonucunda müşterilere sürekli olarak sunulan daha ucuz fiyatlı daha cazip alternatifler müşterilerin ürün ve hizmetlere olan sadakatini azaltmaktadır. İşletmeler rekabet avantajı elde etmek, müşteri bağlılığını artırmak için yeni yollar aramaktadır. Bu yollardan biri de müşteriler ile uzun dönemli ilişki kurmak yoluyla ürün ve hizmetlerin farklılaştırıldığı “Müşteri İlişkileri Yönetimi”ni kullanmaktır. Müşteri kaybı her şirketin dikkat etmesi gereken konulardan biridir. Pazar payını korumak, hatta ileriye götürmek için, mevcut müşterilere eğilmenin daha verimli ve az maliyetli olduğundan yola çıkarak, müşteri kaybının engellenmesi çok önemlidir.

Dünyada müşteri bağlılığı oluşturma yönlü pazarlamaya doğru bir yönelim vardır. Bunun temel nedeni, işletmelerin faaliyetlerini sürdürmeleri ve düzenli gelir elde etmeleri için sürekli müşterilere ihtiyaç duymalarıdır. Müşterilerin tekrar tekrar aynı işletmeyi tercih etmesinin sağlanması için değişik isimlerle farklı stratejiler önerilmesine rağmen birçok işletme yöneticisi, müşteri bağlılığının sağlanması üzerinde yoğunlaşmaktadır.

İşletmelerin, müşteri bağlılığı sağlama yönlü programlara daha fazla önem vermelerinin temel nedenleri, işletmeye bağlılık duyan müşterilerin; işletmelerin daha uzun süreli müşterileri olmaları, hizmet edilme maliyetlerinin yeni müşterilere göre daha düşük olması, daha fazla harcama yapma ve satın alma eğiliminde bulunmaları, işletmenin sunduğu diğer ürünleri daha kolay satın almaları, ürünlerin fiyatlarına daha az duyarlı olup yüksek fiyat ödemeye hazır olmaları, aynı ürünü sunan rakip işletmelerin rekabet araçlarına duyarsız kalmaları nedeniyle rakiplerin pazara girmelerini zorlaştırmaları, bağlı olduğu işletmelerin ürünlerini, başkalarına tavsiye etmeleri ve işletmenin gönüllü ağızdan ağza pazarlama ve duyurum aracı olmaları, işletmenin mal ve hizmet sunumlarında yaptıkları hatalarına karşı daha toleranslı

davranması ve bağıllık duydukları işletmeleri savunma tutumu ve davranışı göstermeleri olarak sıralanabilir. Aynı şekilde işletmelerin; pazarlama faaliyetlerinde kullandığı medya ve araç türlerinin sayısının çoğalması ve çeşitlenmesi nedeniyle yeni müşterilere ulaşmasının ve kazanmasının zorlaşması, mevcut müşterilerini elde tutmasının yeni müşterileri kazanmasından daha az maliyetli olması ve müşterileri ile oluşturduğu uzun dönemli ilişkilerle karlılıklarını arttırdıkları belirlenmiştir. Bu nedenle uzun dönemli rekabet avantajı elde etmek isteyen işletmeler, müşteri bağıllığı sağlamaya ve müşteri bağıllığı programları hazırlamaya başlamışlardır.

Başarılı bir müşteri bağıllığı sağlama programı ile oluşturulan müşteri bağıllığını, müşterilerin işletmenin ürünlerine karşı olumlu bir tutum değişikliğine yol açabilmekte, rakip işletmelerin tutundurma çabalarının dikkate alınmasını engelleyebilmekte ve marka veya işletme değişikliğine karşı güçlü bir direnç göstermesine neden olabilmektedir. Böylece rakip işletmelerin müşterilerini çekmeye yönelik uygulamalarını zorlaştırmakta, katlanması gereken maliyetleri yükseltmekte ve pazara yeni girişleri caydırabilecek önemli bir engel teşkil etmektedir. Bu durum rakip işletmelerin, potansiyel müşterilerinin başka işletmelere olan bağıllıklarını kırmak için yeni yollar araması ve bağıllık duyan müşterilerin tekrar aktif bir satın alma karar sürecine dönmesi için daha yoğun ve maliyetli pazarlama çabalarında bulunmalarını gerektirmektedir.

Müşteri bağıllığının sağlanmasında rol oynayan çok sayıda değişken vardır. Bir başka ifade ile müşteri bağıllığını bağımlı değişken olarak düşünürsek, müşteri bağıllığı çok sayıda bağımsız değişkenin -müşteri memnuniyeti, ürün kalitesi, sunulan değer, güven, ödüllendirme, alışkanlıklar, değiştirme riski, satış sonrası hizmetler gibi veya pazarlama stratejilerinin ilişki pazarlaması, müşteri ilişkileri yönetimi, kulüp pazarlama gibi etkisi altındadır. Müşteri kaybına “churn” denmektedir. Önce bir GSM operatöründen diğerine geçme için kullanılırken şimdi sektör gözetmeksizin rakibe kaçan tüm müşteriler için kullanılmaya başlanmıştır. Bir kayıp azaltma çalışması, önce aktif müşteri, inaktif müşteri, kayıp müşteri gibi tanımların yapılması ile başlar. Kayıp kavramı sektörler arasında değişen bir

kavramdır. Örneğin işyerinizin yakınında ki bir lokantaya 3,4 haftadır gitmiyorsanız, muhtemelen firma sizleri kaybetmiştir. Kayıp kavramı yanında iki kavram daha taşımaktadır. Bunlar aktif ve inaktif müşteri olarak tanımlanmaktadır. Müşteri önce aktif olarak düşünülmektedir, önce mağazadan alışveriş yapıyordur, sonra inaktif durumuna düşmektedir. Mağazaya her an uğrayabileceği gibi, bir daha uğramayabilir. Bu durumda müşteri kaybedilmiş durumdadır.

GSM sektöründe pazar büyüme hızının düşmesi, fiyat rekabetinin daha da artmasına ve GSM operatörlerin hizmetlerini farklılaştırarak abone başına düşen karlılıklarını arttırmak için strateji arayışı içinde bulunmalarına neden olmuştur. Bu noktada bir GSM operatörü müşterilerinin, diğer operatörlerin de potansiyel müşterileri olması, GSM sektöründe müşteri bağlılığını etkileyen faktörlerin belirlenmesini ve buna göre stratejilerin hazırlanmasını önemli hale getirmiştir.

Müşteri kaybı, müşterilerin aldıkları GSM operatörünün hizmetlerinden memnun olmama durumlarında ayrılmaları anlamını da taşımaktadır. Müşteri kayıp durumunda müşteri aynı gsm şirketinde paket değişikliği yapmaktadır, yahut başka bir operatörü tercih etmektedir.

Telekomünikasyon hizmet sektöründe, müşteri kayıp yönetimi, değerli müşterilerin uzun ömürlü tutulmasını sağlamak anlamına gelmektedir. Benzer olarak müşteri kayıp yönetimini Berson ¹GSM operatörlerinin müşterilerini elde tutmak için gösterdikleri mücadele diye tanımlamıştır.

1990'lı yıllardan itibaren artan rekabet, ürünler (mal ve hizmetler) arasında algılanan eşitlik, artan müşteri bilinci ve ürünlerin fiyatlarına duyarlılık ve geleneksel tutundurma araçlarının maliyetlerinin artmasına karşılık etkinliğinin azalması, işletmelerin daha fazla rekabet baskısı altında bulunmalarına neden olmuştur. Bu rekabet şartları, işletmeleri müşteri bağlılığı programlarının sunabileceği fırsatları

¹ A.Berson S.SmithK., “**ThearlingBuilding data mining applications for CRM**”, McGraw-HillNew York, NY, 2000.

araştırmaya yöneltmiş ve mevcut müşterileri ile daha fazla ilgilenmeye zorlamıştır. Buna göre işletmeler, mevcut müşterilerini işletmeye ve/veya işletmenin sunduğu ürünlere, markalara veya mağazalara bağlılık duyan birer müşteri haline getirme yönlü pazarlama stratejileri hazırlamaya yönelmişlerdir.

Müşteri bağlılığı, diğer sektörlerde olduğu gibi GSM sektörü için de önemlidir. Bunun temel nedeni, GSM sektörünün olgunluk aşamasına gelmesi sonucu yeni müşterilerin (abonelerin) kazanılmasının zor ve maliyetli olması ve GSM operatörlerinin mevcut müşterilerinin, diğer operatörlerin potansiyel müşterileri olmasıdır. ABD'de yayınlanan 2005 Tüketici Raporlarına göre, 2004 yılında mobil telefon (cep telefonu) kullanıcılarının %35'i, GSM operatörlerini değiştirmeyi düşünmektedir. Bu nedenle GSM sektöründe, müşteri bağlılığı programları daha fazla gündeme gelmeye başlamış ve GSM operatörlerine bağlılığı etkileyen değişkenlerle ilgili farklı araştırmalar yapılmıştır. Ancak ülkelere göre GSM operatörlerine bağlılığı etkileyen faktörler farklılık gösterebilmektedir. Örneğin, Fransa'da sunulan GSM hizmetlerinde algılanan kalitenin ve değiştirme maliyetinin GSM operatörlerine olan bağlılığa etkisi ² Almanya'da GSM operatörlerinin müşterilerini elde tutma, memnun etme ve bağlılık düzeyleri arasındaki ilişki ³, Güney Kore'de müşteri memnuniyeti ve değiştirme maliyetinin bağlılık üzerindeki etkisi ⁴, Türkiye'de işletme imajı, algılanan hizmet kalitesi, güven ve değiştirme maliyetinin GSM operatörü müşterilerinin bağlılıklarına etkisi ⁵ ve ABD'de GSM operatörünün sunduğu hizmet kalitesi, algılanan değer ve bunların memnuniyet ve bağlılık üzerindeki etkisi ⁶ yapılan araştırmalarla analiz edilmiştir.

² LEE, J., Lee, J., Feick, L. “**The impact of switching costs on the customer satisfaction-loyalty link:mobile phone service in France**”, Journal of Services Marketing, Cilt No:15(1), 2001, ss.35-48

³ GERPOTT, T.J., Rams, W., Schindler, A. “**Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market**”, Telecommunications Policy, Cilt No:25, 2001, ss.249-269.

⁴ KIM, M., Park, M., Jeong, D. “**The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunication services**”, Telecommunications Policy, Cilt No:28, 2004, ss.145-59.

⁵ AYDIN, S., Özer, G. “**The analysis of antecedents of customer loyalty in the Turkish mobile telecommunication market**”, European Journal of Marketing, Cilt No:39(7/8), 2005, ss.910-925.

⁶ KIM, H., Widdows, R., Park, J. “**E-loyalty: winning strategies for mobile carriers**”, Journal of Consumer Marketing, Cilt No:23(4), 2006, ss.208-218.

Müşteri kayıp yönetimi işlemi veri madenciliği araçlarıyla bu çalışmada desteklenmiştir. Veri madenciliği karar verme süreçlerinde firmalar için büyük önem taşımaktadır. Yöneticiler açık olmayan veriler içinden veri madenciliği yöntemleri ile birbirleriyle ilişkili verileri ortaya çıkarırlar ve firma için operasyonel, stratejik kararlar verirler. Veri madenciliğinin asıl hedefi saklı bilgiyi geniş veri kümeleri içerisinde çıkarmaktır. Veri madenciliğinde sıklıkla kullanılan bazı teknikler yanında, diğer yöntemlerde sonuçlar kıyaslanmıştır. Çalışmada, Fuzzy(Bulanık) Kümeleme tekniklerinden, amaç fonksiyonlarına dayanan bütün kümeleme tekniklerinin temelini oluşturan Fuzzy(Bulanık) Ortalamalar(BCO) tekniği incelenmiş ve bu teknik kullanılarak Türkiye'deki bir GSM Operatörünün müşteri kaybı analizi yapılmaya çalışılmıştır. Fuzzy(Bulanık) kümeleme yöntemlerinde, diğer kümeleme yöntemlerinin aksine veri uzayındaki her bir bireyin elde bulunan bütün kümelere belirli ölçüde girmesine izin veren üyelik fonksiyonlarının kullandığı ve bu üyelik değerleri, veri setinin veya kümelerin doğasını anlama açısından çok önemli olduğu anlaşılmıştır. Hiyerarşik kümeleme gibi diğer kümeleme yöntemlerinin sonuç olarak kesin kümeler verdiği ve bireylerin bu kümelere üyelik derecelerinin ya 0 ya da 1 olduğu, oysaki bulanık kümelemede bireylerin üyelik derecelerinin 0 ile 1 arasında yer aldığı ve bireylerin bütün kümelere üyelik dereceleri toplandığında 1 elde edildiği bilgisine ulaşılmıştır.

2. TELEKOMÜNİKASYON

Telekomünikasyon (uzak iletişim), Fransızca télécommunication sözcüğünden geçmiştir. Duyum, yazı, resim, simge ya da her çeşit bilginin tel, radyo, optik ile başka elektromanyetik dizgelerle iletilmesi, bunların yayımı ya da alınması anlamına gelmektedir.

1819'da Danimarkalı fizikçi Hans Christian Oersted'in mıknatısla hareket kazandırılan bir iğnenin sinyal aracı olabileceğini bulmasının ardından, 1820'lerde Fransız fizikçi Andre Marie Ampere'nin hareketli iğneyi icat etmesi ve 1844'te Amerikalı mucit Samuel Morse'un ABD Kongresi'nde ilk elektromanyetik telgrafi sunuş gösterisi ile birlikte, dünyada modern telekomünikasyon çağı başlamıştır. Noktadan noktaya iletişim telgrafla başlamış ve daha sonra da 1870'lerde Elisha Gray ve Alexandar Graham Bell'in çalışmaları sonucu telefonun icadıyla gelişmiştir. ⁷ Her ikisi de, frekans ayırmalarını bir telgraf devresi üzerinde kullanarak çoklu mesaj gönderme imkânlarını araştırmışlardır. Gray, bu yaklaşımın bazı basit sesli mesajların gönderilmesinde kullanılma imkânını görmüş, ancak çalışmalarını dahaiyi bir telgraf cihazı geliştirme üzerinde yoğunlaştırmıştır. İBell ise, aynı sesli nakil imkânlarını görerek bu yeni konu üzerinde yoğunlaşmış ve 1876'da ilk telefonun patentini alarak Gray'in önüne geçmiştir. 1880'lere gelindiğinde, telefon iş amaçlı kullanıma girmesinin yanı sıra evlere dahi girmenin eşiğindeydi. ⁸

1900'lü yılların başında, birbiriyle irtibatı olmayan yerel şebekeler ile birlikte birçok işletmecinin aniden ortaya çıkması, telekomünikasyon açısından ilk anda pek bir sorun yaratmamıştır. Ancak; zamanla şebeke entegrasyonu yaşanıp, hizmetin bu şebekeler arasında geçiş yaparak, kullanıcıya sorunsuz sunulabilmesi gündeme geldiğinde konunun karmaşık yapısı gün yüzüne çıkmıştır. 1920'lerin teknolojik düzeyiyle bunu aşmanın tek yolu, servisin uçtan uca sağlanma şeklindeki yapılanma çerçevesinde, pazarda doyuma ulaşılmıştır. Bu süreçte, sayısal iletişimin, bireyleri

⁷ Mehmet Özçağlayan: **Yeni İletişim Teknolojileri ve Değişim**, Alfa Basım Yayım Dağıtım, İstanbul, 1998, s.79.

⁸ Nuray Keskin: **Telekomünikasyon Sektöründe Yeniden Yapılanma**, Ankara: Ankara Üniversitesi, 2001, s.2.

müşteri konumuna sokması çok önemli bir değişikliktir. Askeri bürokrasi, sivil bürokrasi, çok uluslu firmalar, orta ölçekli işletmeler ve küçük boyutlu işletmelerden sonra 1980'lerde bilgisayarların kişiselleşmesi ile sıra kişilere gelmiştir.⁹

Elektronik sanayisindeki gelişmeler santral ve tüketici cihazlarında yaşanan gelişmelerin yanı sıra, mobil haberleşme hizmetleri ve iletimde uydu sistemlerinin gelişmesine olanak sağlamıştır. Telli iletişimde de teknolojik gelişmeler olmuş, geleneksel havai hatlı iletim sistemlerinin yerine önceleri koaksiyel kablo ve daha sonra fiber optik kablolu iletim sistemleriyle iletimde hız, miktar, kalite ve güvenilirlik artmış; maliyetler daha da düşmüştür.¹⁰

1840'lı yıllardan 1990'lı yıllara kadar sektörde devre anahtarlama sistemleri kullanılırken güvenlik amaçlı geliştirilen paket anahtarlama teknolojisi telekomünikasyon alanında büyük değişimlere neden olmuştur. Bunun en yaygın örneği olan internet 1988 yılından itibaren sivil kullanıma açılmış ve yaklaşık 150 yılda oluşan telekomünikasyon şebekelerinin yapısını değiştirmeye başlamıştır. 1990'lı yılların ortalarından itibaren yaygınlaşmaya başlayan sayısal mobil telefonlarla internetin gerçek bir bileşimi olan 3. nesil şebekelerin önümüzdeki süreçte kullanılmasıyla birlikte haberleşmede önemli değişimlerin yaşanması beklenmektedir.¹¹

Günümüz koşullarında, dünyada 1 milyar civarı mobil telefon olmak üzere 2 milyardan fazla telefon abonesi bulunmakta olup, yıllık baz da Telekom sektöründe elde edilen gelirlerin toplamı yaklaşık 1,3 trilyon dolar olarak hesaplanmaktadır.¹²

Sonuçta, 1980'li yılların başından itibaren dünya ekonomisinde gözlenen serbestleştirme ve küreselleşme eğilimlerinin artması ve gelişmesiyle birlikte, bilgiye ve bilgi akışını sağlayan telekomünikasyon hizmetlerine talep de giderek artmış ve çeşitlenmiştir. Böylesine güçlü talep uyarısı, yukarıda kısaca özetlenen gelişim

⁹ Klemperer, P, **What Really Matters in Auction Design**, Şubat 2000

¹⁰ Nuray Keskin, s.4.

¹¹ SAFEL, R., **Telekomünikasyon Sektörü**, Türkiye Vakıflar Bankası T.A.O Planlama ve İktisadi Araştırmalar Grup Yönetmeliği Sektör Araştırmaları, Seri No: 25, Ekim 2001.

¹² ITU (Uluslararası Telekomünikasyon Birliği) İstatistikleri, 2010.

sürecini tetiklemiş, kişisel ve kurumsal bilgisayar teknolojisiyle birlikte, telekomünikasyon sektörü de baş döndürücü bir hızla gelişim süreci yaşamıştır. O kadar ki sektördeki gelişmeleri takip etmek dahi neredeyse imkânsızlığa bürünmüştür.¹³

Günümüzde gelinen noktada, haberleşme ve bilgisayar teknolojileri birleşmiş, bu sayede olağanüstü gelişme yaşayan telekomünikasyon sektörü, mobil ve internet iletişim ortamının temel alt yapısı halini almış, uygun altyapının kurulması halinde, dünya toplumlarına, dünyanın her noktasındaki bilgi kaynaklarına erişebilme ve bilgiyi transfer edebilme imkânı sağlar hale gelmiştir.¹⁴

2.1 Türkiye'de Telekomünikasyon Sektörü

Telekomünikasyon sektörü tüm dünyada ve Türkiye’de büyük bir hızla büyüyor. Geleceğin en parlak sektörü olarak değerlendirilen sektörün dünyadaki büyüklüğü 1.3 trilyon dolar, Türkiye’de ise 20 milyarı aşıyor. Türkiye’de son yıllarda sektörün rekabete açılması büyümede etkili oldu. Araştırmalar, Türkiye mobil iletişim pazarının 2012’ye kadar Avrupa pazarından daha hızlı büyüyerek, yüzde 21’lik artışla 10,5 milyar dolara yükseleceğini gösteriyor. 2009’da küresel krize rağmen Türkiye’nin üç GSM operatörü Turkcell, Vodafone ve Avea’nın yapacağı toplam 3 milyar 300 milyon dolarlık yatırım sektörü daha da hareketlendirecek.¹⁵

Büyük bir hızla gelişen teknoloji hayatımızın her alanını önemli ölçüde etkiliyor. Ses, görüntü ve veri aktarımının gerçekleştirildiği sektör olarak tanımlanabilen “telekomünikasyon” uluslararası alanda teknolojik gelişmelerden en fazla etkilenen sektörler arasında yer alıyor. Telekomünikasyon sektöründeki gelişmeler ve internetin yaygınlaşması dünyadaki sınırları kaldırıyor.¹⁶

¹³ OECD, **Communications Outlook 2001**, DSTI/ICCP/TISP(2000)10/CHAP4, Tablo 4.5, s.14

¹⁴ ITU, **Cellular Mobile Subscribers**, Haziran 2000

¹⁵ Süel, H., **Türk Telekomünikasyon Sektörü ve Özelleştirme**, İktisat İşletme ve Finans Dergisi, Nisan,1998.

¹⁶ Evren, G., **Hizmete ve Altyapıya Dayalı Rekabet ile Geçiş Stratejileri**, Ankara: Telekomünikasyon Kurumu, 2005

Ülkeler telekomünikasyon yapılarını güçlendirerek birbirleriyle rekabet ediyor. Telekomünikasyon sektörü dünyada ve ülkemizde geleceğin en parlak sektörü olarak değerlendiriliyor. Sektördeki hızlı gelişmenin temelinde, kamunun tekelinde olan sektörün rekabete açılması yatıyor. Sektöre yeni oyuncuların girmesi, rekabeti arttırıyor ve yeni teknolojilerin hızla yayılmasını sağlıyor. Bugün telekomünikasyon sektörünün dünyadaki büyüklüğü yaklaşık 1.3 trilyon dolar. 2008 yılı itibarıyla sektörün ülkemizdeki büyüklüğüse 20 milyar doları aşıyor.¹⁷

Türkiye’de telekomünikasyon sektörünün tarihi Osmanlı döneminde 1840’ta Sultan Abdülmecit tarafından “Postahane-i Amirane”nin kurulmasıyla başladı. 1855’te Telgraf Nezareti’nin kurulmasından sonra, 1871’de Posta ve Telgraf Nezaretleri birleştirilerek, Posta ve Telgraf Nezareti’ne dönüştürüldü. Türkiye’de ilk telefon sistemi İstanbul’da, 1911’de American Western Electric şirketine tanınan 30 yıllık bir imtiyaz karşılığında, bu şirket adına faaliyet göstermek amacıyla oluşturulan Dersaadet Telefon Anonim Şirketi tarafından kuruldu. Birinci Dünya Savaşı sırasında devlet tarafından el konulan şirket, 1935’e kadar İstanbul ve civarında telefon hizmetlerini sağlamaya devam etti, 1936’da satın alındı. Posta, Telgraf ve Telefon İdaresi (PTT) 1936’da kuruldu ve ülke çapında telekomünikasyon hizmetlerini sağlamakla görevlendirildi. PTT İdaresi, 1939’da Ulaştırma Bakanlığı’na bağlandı, 1953’de iktisadi devlet teşekkülü olarak yeniden yapılandırıldı. 1980’li yıllardan itibaren yapılan yatırımlarla telekomünikasyon sektöründe gelişmeler yaşandı. 1994’te telekomünikasyon hizmetleri PTT’den ayrıldı ve Türk Telekomünikasyon A.Ş. ile dinamik bir yapıya kavuştu. Aşağıdaki tablo Türkiyede Telekomünikasyon ‘un tarihsel süreciyle ilgili bilgi vermektedir.¹⁸

¹⁷ ALPTÜRK, H., **Yeni Yasa Yeni Yapılanma ve Anlayış**, Türk Telekom Dergisi, Kozan Ofset Baskı, Sayı:5, Eylül Ekim, 2000

¹⁸ Elektrik mühendisliği, 430. sayı, Nisan 2007

Tablo 2. 1 Türkiye Telekomünikasyon Tarihsel Süreci

Tarihsel Süreç	Gelişmeler
23 Ekim 1840	Bugünkü Türk Telekom Postahane-i Amirane adıyla Sultan Abdülmecit tarafından atıldı.
9 Ağustos 1847	İlk telgraf alma-çekme işleminin başarıyla gerçekleştirilmesi üzerine ilk telgraf hattının İstanbul-Edirne arasında döşenmesine başlandı.
Temmuz 1881	İstanbul Soğukçeşme'deki Posta ve Telgraf Nezareti binasıyla Yeni Cami'deki postane arasında tek telli bir telefon çekildi.
3 Mayıs 1909	İlk manuel telefon santrali, İstanbul Büyük Postane binasında 50 hatlık olarak tesis edildi.
4 Şubat 1924	406 sayılı Telefon ve Telgraf Kanunu ile yurdun her tarafında telefon tesis etme ve işletme görevi PTT Genel Müdürlüğü'ne verildi.
11 Eylül 1926	Türkiye'nin ilk otomatik telefon santrali, 2000 hatlık kapasiteyle Ankara'da hizmete verildi.
1 Eylül 1929	Tek devreli ilk şehirlerarası haberleşmesi Ankara-İstanbul arasında gerçekleştirildi.
1940	Ankara-İstanbul arasında tesis edilen 2 adet tek kanallı havai hat çoklayıcı sistemi haberleşmede eskiye göre büyük kolaylık sağladı.
Kasım 1973	İlk otomatik teleks santrali kuruldu.
6 Nisan 1976	Antalya-Catania arasında toplam 480 kanallı ilk denizaltı koaksiyel kablosunun hizmete verilmesiyle, çok kanallı yurtdışı haberleşmesi sağlandı.
23 Nisan 1979	İlk uydu haberleşme yer istasyonunun hizmete verilmesiyle INTELSAT üzerinden Atlantik bölgesi uyduları kullanılarak 13 ülke ile haberleşme sağlandı.
Mart 1982	Şehirlerarası ve milletlerarasına açık ankesörler kurulmaya başlandı.
28 Haziran 1984	Ankara, İstanbul, İzmir ve Adana illeri arasında elektronik mektup hizmeti vermeye başlandı.
5 Temmuz 1984	Bir sistem üzerinden çok sayıda köyü otomatik telefon şebekesine bağlayacak multi access özel radyo link sistemi tesis edildi.
18 Aralık 1984	Türkiye'nin ilk sayısal telefon santrali, Ankara Kavaklıdere'de hizmete verildi.
10 Nisan 1985	Haberleşmede kanal kapasitesini artıran fiber optik kablo, ilk kez Ankara(Ulus)-Gölbaşı Uydu yer merkezi arasında yeraltında döşenerek 1310 nm dalga boyunda 140 Mb/s'lik sistem hizmete verildi.
Kasım 1985	İlk sayısal radyolink sistemi Ankara-İstanbul arasında hizmete verildi.
23 Ekim 1986	Mobil telefon, Ankara ve İstanbul'da; çağrı cihazları da Ankara, İstanbul ve İzmir'de hizmete verildi.
4 Mart 1987	Avrupa'da ilk olarak uydu sistemi üzerinden video konferans ülkemizde gerçekleştirildi.

15 Temmuz 1987	İller arasına fiber optik kablo, ilk kez Aydın-Denizli arasında havai olarak döşendi.
Aralık 1988	İlk olarak Ankara, Çankaya’da Kablo TV hizmeti verilmeye başlandı.
1989	DPN modülleri kullanılarak ilk Paket Anahtarlama Data Şebekesi olan TURPAK kuruldu. Şebeke üzerinden X.25, ITI, SNA servisleri verilmeye başlandı.
21 Aralık 1990	Fransız Aerospatiale firması ile “TÜRKSAT Milli Haberleşme Uyduları” sözleşmesi imzalandı.
24 Aralık 1990	EMOS 1 projesiyle, İtalya-Yunanistan-Türkiye-Ortadoğu arasında fiber optik denizaltı kablosu üzerinden haberleşme sağlandı.
1991	Kırsal alan santrallerinin transmisyonunu sağlamak üzere düşük kapasiteli sayısal radyolink sistemleri servise erilmeye başlandı.
23 Şubat 1994	Türkiye GSM teknolojisiyle tanıştı. Haberleşmede sınır tanımayan GSM ilk kez Ankara, İstanbul ve İzmir’deki abonelerine hizmet vermeye başladı.
Nisan 1994	Özel bir santralin (PABX) sağladığı tüm özellikleri Ulusal Telefon Şebekemiz üzerinden sağlayan Centrex telefon hizmeti verilmeye başlandı.
11 Ağustos 1994	Türkiye’nin ilk uydusu TÜRKSAT uzaya fırlatıldı.
1994	TURMEOS-1 (Türkiye Marmara Ege Optik Sistemi) ve TURCYOS (Türkiye-Kuzey Kıbrıs Denizaltı Fiber Optik Kablosu) hizmete verildi.
24 Nisan 1995	PTT’deki telekomünikasyon ve posta hizmetlerinin birbirinden ayrılmasıyla Türk Telekomünikasyon A.Ş. kuruldu.
1996	Türkiye-İtalya-Ukrayna-Rusya’yı kapsayan ITUR Denizaltı Fiber Optik Kablo Sistemi servise verildi.
1996	Türk Telekom abonelerinin ödemelerini anlaşmalı bankalar, PTT merkezleri ve Türk Telekom tahsilat merkezlerinden, on-line olarak yapmalarına imkan sağlayan Merkezi Tahsilat Sistemi hizmete verildi.
1996	Passport modüllerinin ilavesiyle TURPAK şebekesi üzerinden yüksek hızlı veri iletimi sağlayabilen Frame Relay ve ATM servisleri devreye verildi.
12 Nisan 1996	Ses, veri ve görüntü iletim altyapısını teşkil eden, santraller arası bir işaretleşme sistemi olan NO 7 Türkiye’ye merhaba dedi.
10 Temmuz 1996	Türkiye’nin ikinci uydusu TÜRKSAT 1C uzaya fırlatıldı.
1996	Türkiye Ulusal İnternet Altyapı Ağı (TURNET) hizmete verildi.
7 Kasım 1996	Türksat uydularını üreten Aeorspatiale ve Türk Telekom ortaklığıyla kurulan Eurasiasat’in kuruluş

	anlaşması imzalandı.
Kasım 1996	Yüksek hızlı veri iletimine olanak sağlayan Frame Relay hizmeti devreye sokuldu.
1997	Türkiye'deki tüm üniversiteleri TURPAK şebekesi üzerinden birbirine bağlayan Ulusal Akademik Ağ (ULAKNET) projesi hayata geçirildi.
Nisan 1997	Ses frekans kablosu şebekesinin daha optimum bir şekilde kullanılması amacıyla sayısal hat çoklayıcı sistemlerin kullanımına başlandı.
1997	KAFOS (Karadeniz Fiber Optik Sistemi) ve TBL (Transbalkan Linki Karasal Fiber Optik Sistemi) hizmete verildi.
1998	Peşin ödemeli kart, ülkemizde Küresel Kart adıyla hizmete verildi.
27 Nisan 1998	GSM lisansı, 25 yıllıkına Turkcell ve Telsim şirketlerine devredildi.
28 Ağustos 1998	İnternet erişimini Türkiye geneline yaymak, hızlı ve kaliteli hizmet sunabilmek amacıyla yeni adıyla TTNetwork eski adıyla TTNNet ulusal internet altyapı ağına ilişkin sözleşme imzalandı.
Şubat 1999	Klasik telefon ve diğer darbantlı servislerin verilmesi için, santralden saha dolabına kadar tesis edilen Fiber Optik kablo üzerinden çalışan erişim çoklayıcı sistemlerin kullanımına başlandı.
Temmuz 1999	Lokal santrala kablosuz erişim sağlayan KTS (Kablosuz Telefon Sistemi) istemleri servise verilmeye başlandı.
2000	Türk Telekom operatör aracılığı ile verilmekte olan 115, 121, 122, 123,126, 131, 163 v.b. hizmetleri Sesli Yanıt Sistemi üzerinden verilmeye başladı. Kablo TV üzerinden Kablo İnternet uygulamaları başlatıldı.
29 Ocak 2000	Türk Telekom, 23948 sayılı Resmi Gazete'de yayımlanan 4502 sayılı Kanun ile KİT statüsünden çıkarılarak özel hukuk hükümlerine tabi anonim şirket oldu.
Kasım 2000	Altyapısı Türk Telekom tarafından sağlanan Türkiye çapındaki saymanlıkların günlük çalışmalarının bilgisayar yardımıyla yapılması ve devlet hesaplarının günlük olarak izlenebilmesi amacıyla oluşturulan SAY2000 Projesi ülke genelinde uygulamaya konuldu.
11 Ocak 2001	Türk Telekom ve Alcatel ortaklığı ile kurulan EURASIASAT şirketi tarafından yaptırılan TÜRKSAT 2A uydusu Güney Amerika'daki Kourou Üssü'nden uzaya fırlatıldı.
Mayıs 2001	Sabit telefonlardan CLIP (Arayan Numaranın Görülmesi) ve CLIR (Arayan Numaranın Görülmemesi) özelliği hizmete verildi.
Haziran 2001	İl Milli Eğitim Müdürlüklerinin çalışmalarının bilgisayar ortamında yapılması ve İnternet erişimlerinin

	sağlanması için İLSİS Projesi hayata geçirildi.
2002	Yalova ve Balıkesir İl Telekom Müdürlükleri'nde Toplam Kalite Yönetimi (TKY) pilot proje uygulamaları başlatıldı.
1 Şubat 2002	Türk Telekom, Uluslararası Telekomünikasyon Birliği'ne (ITU) ve işletmeci şirket olarak, sektör bazında Radyokomünikasyon ITU-R, Standardizasyon ITU-T, Geliştirme ITU-D kuruluşlarına üye oldu.
Mart 2002	Altyapısı Türk Telekom'ca sağlanan nüfus işlemlerinin bilgisayar ortamında yapılması için oluşturulan MERNİS Projesi uygulamaya konuldu.
Nisan 2003	SDH (Synchronous Digital Hierarchy) tabanlı sayısal radyo sistemleri servise vermeye başlandı.
Ekim 2003	Deneme amaçlı ilk Metro Ethernet uygulamaları başlatıldı.
Aralık 2003	42500 okulun ADSL hizmeti ile internete bağlanması için Milli Eğitim Bakanlığı ile ortak proje başlatıldı.
2004	Türk Telekom'un GSM Operatörü Aycell ile İŞ-TİM'in birleşmesi sonucu kurulan TT&TİM İletişim Hizmetleri A.Ş., "Avea İletişim Hizmetleri A.Ş." ticari ünvanı ile 19 Şubat 2004 tarihinde resmen kuruldu.
17 Mayıs 2004	Telekomünikasyon Kurumu tarafından UMTH'ye yönelik 2. Tip telekomünikasyon lisansları vermeye başlandı. UMTH konusunda lisans alan firmalar ile Türk Telekom arasında "C Tipi UMTH'ye İlişkin Arabağlantı Sözleşmesi" imzalanmaya başlandı.
18 Mayıs 2004	İstanbul Türkiye ve Katanya (İtalya) arasındaki MEDTÜRK Denizaltı Fiber Optik Kablo Sistemi hizmete verildi.
Haziran 2004	TTkart olarak adlandırılan Ön Ödemeli Kart hizmete sunuldu.
2004	Bilinmeyen Numaralar Hizmetinin yanı sıra Danışma ve Katma Değerli Hizmetlerin verilmesi planlanan Operatörlü Çağrı Merkezi hizmete sunuldu.
2004	Türk Telekom'un ulusal ve uluslararası uydu haberleşmesiyle ilgili tüm hak ve sorumlulukları, 16.06.2004 tarihli 5189 nolu yasa ile kurulan TÜRKSAT Uydu Haberleşme ve İşletme Anonim Şirketi'ne (TÜRKSAT A.Ş.) devredildi.
Haziran 2004	Yalova ve Balıkesir İl Telekom Müdürlükleri ISO 9001:2000 Kalite Yönetim Sistemi belgesi aldı.
22 Temmuz 2004	Türksat Uydu Haberleşme ve İşletme Anonim Şirketi (Türksat A.Ş.) kuruldu. Türk Telekom'dan ayrı bir şirket olarak faaliyetine başladı.

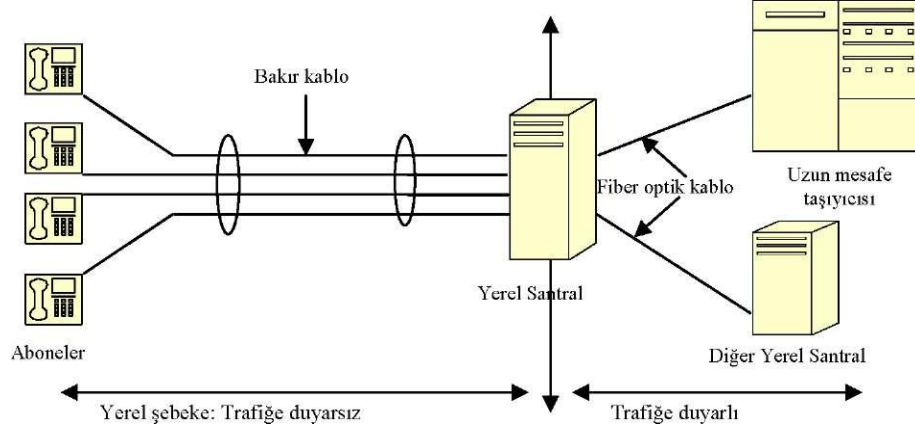
2004	Kamu ve özel kuruluşların İnternet Veri Merkezi ihtiyaçlarının karşılanması amacıyla, Web Hosting, Mail Hosting ve Co-location gibi katma değerli IDC hizmetlerinin verilebileceği TTIDC kuruldu.
2004	Resmi kurum ve kuruluşlar ile kurumsal müşterilerin yurt çapındaki birimlerinin noktadan noktaya data transferlerini sağlamak için noktadan noktaya DSL hizmetleri vermeye başlandı.
2004	TURPAK şebekesinden yüksek hızlı Frame Relay ve ATM hizmetleri tüm illerde verilir hale getirildi.
25 Ocak 2005	2005/8409 sayılı Bakanlar Kurulu Kararı ile T.C. e-Devlet Kapısı'nın kurulması görev ve sorumluluğu Türk Telekom'a verildi.
Şubat 2005	TTwinet-Kablosuz İnternet Servisi hotspot olarak isimlendirilen (oteller, havaalanları, alışveriş merkezleri, kafeler, restaurantlar, fuar merkezleri, üniversiteler v.b.) 160 yerde hizmete verildi.
14 Kasım 2005	Türk Telekom'un özelleştirilmesi çalışmaları tamamlanarak, yüzde 55'i Oger Ortak Girişim Grubu'na devredildi.
15 Eylül 2006	Türk Telekom, İş-TİM'in Avea'daki %40,56 oranındaki hissesini 500 milyon ABD doları karşılığında satın aldı ve Türk Telekom'un Avea'daki hisse oranı %81,12'ye yükseldi.
15 Mayıs 2008	Türk Telekom hisselerinin % 15'lik payı halka arz edilerek, İstanbul Menkul Kıymetler Borsası'nda işlem görmeye başladı.
2009	Avea başarılı bir şekilde 3N hizmetleri vermeye başladı.

Günümüzde telekomünikasyon sektörü genel olarak "temel hizmetler", "katma değerli hizmetler" ve "telekomünikasyon cihazları" olarak, 3 ana başlık altında sınıflandırılmaktadır.¹⁹

2.1.1 Temel Telekomünikasyon Hizmetleri

Temel hizmetler, Telefon, telgraf gibi, bir iletim hattı üzerinden sadece taşıma kapasitesi sunan hizmetler olarak tanımlanmaktadır. Bu anlamıyla temel hizmetlerin sunulduğu şebeke yapıları, telekomünikasyon sistemlerinin kendisidir. Telekomünikasyon sektörü açısından şebeke; bir veya daha fazla nokta arasında iletişim sağlamak amacıyla bu noktalar arası bağlantıları oluşturan her türlü iletim ağını ifade etmektedir.

Telefon şebekelerinin en temel iki unsurunu santral ve bağlantı (hat) adı verilen yapılar oluşturmaktadır. Abonelerin birbirleriyle görüşme yapabilmeleri için öncelikle her abonenin kendi bölgesinde yer alan santrale bağlanması gerekmekte ve



Şekil 2.1 Basit Yerel Ağ (Local Loop)

Kaynak: Laffont ve Tirole 2000, s.13.

abonelerle santraller arasındaki hatlar üzerinden ses, veri, görüntü, vs. iletimi ulaşılmak istenen aboneye ulaştırılmaktadır.²⁰ Bu şekilde abonelerle santraller

¹⁹ Recep Çakal, **Doğal Tekellerde özelleştirme ve Regülasyon**, Ankara: DPT, 1996, <http://ekutup.dpt.gov.tr/kit/cakalr/tekel.html>, (22 Mayıs 2011), s. 40-41.

²⁰ MIZRAK. E., **Türk Telekom Kalite Yolculuğunda İlk Adımı Attı**, Telekomünikasyon Mühendisleri Derneği Dergisi, Mart 2000.

arasındaki en küçük şebeke, "yerel (lokal) telefon şebekesi" olarak tanımlanır. Bu şebekeler genelde kablolu (bakır, koaksiyel veya fiber-optik kablo) veya kablosuz (uydu, hücreli radyo, mikrodalga vs) olmak üzere ikiye ayrılmaktadır. Abonelerin diğer bölgelerde, şehirlerde veya ülkelerdeki abonelerle bağlantılarının sağlanabilmesi için santraller arasında da bağlantı olması gerekmektedir. Yerel, şehirlerarası ve uluslararası hatlarda aynı anda çok sayıda görüşme yapılması gerektiğinden, daha yüksek kapasiteli iletim araçları olan fiber optik kablolar kullanılırken, yerel hatlarda ise genellikle daha sınırlı kapasiteye sahip bakır ya da koaksiyel kablolar kullanılmaktadır.²¹

Temel telekomünikasyon hizmetlerinin en temel birimini yerel şebekeler (local loop) oluşturmaktadır. Yerel şebekelerin, yerel santrale kadar olan bölümlerinde, kullanıcıların tek tek santrale kablolar yoluyla bağlanmasını gerektirmesi, bu nedenle sabit maliyetlerinin çok yüksek olması nedeniyle bu kısım halen doğal tekel olarak kabul edilmektedir.

Aboneler arasındaki bağlantı, telefon sisteminin ilk kullanılmaya başlandığı dönemlerde "manuel" olarak gerçekleştirilirken, teknolojinin gelişmesiyle önce otomatik sonra da sayısal santraller kullanılmaya başlanmıştır.²²

Kablo üzerinden sağlanan telekomünikasyon hizmetlerinde genel olarak bakır kablolar kullanılmaktadır. Ancak, bu kabloların taşıma kapasiteleri, çeşitli sıkıştırma teknikleri ile artmış olsa da, fiber optik kablolarla göre daha düşük olmaktadır. Fiber optik kablolar üzerinden ses, veri ve görüntü iletimi çok büyük hızlarda sağlanabildiğinden, fiber optik kablolarla olan talebin artması neticesini vermiş, telekomünikasyon hizmetlerinin gelişmesiyle birlikte kablosuz teknolojide de ilerlemeler beklenmektedir.

²¹ Özge İçöz, **Telekomünikasyon Sektöründe Regülasyon ve Rekabet**, Ankara: Rekabet Kurumu, 2003, s. 21.

²² MIZRAK. E., **Türk Telekom Kalite Yolculuğunda**, Türk Telekom Dergisi, Kasım Aralık 2000, 2001/1 Sayı: 6, Kozan Ofset Matbaası.

2.1.2 Katma Değerli Telekomünikasyon Hizmetleri

Katma değerli hizmetler, temel hizmeti biçim, içerik, protokol veya diğer yönleriyle bir işleme tabi tutan, bilgisayar uygulamaları ile birleştiren veya aboneye ilave, farklı veya yeniden şekillendirilmiş bilgi sunan ya da abone ile stoklanmış bilgi kaynağı arasında karşılıklı ilişkiyi sağlayan bilgi yoğun hizmetlerdir. Katma değerli hizmetler, temel hizmetler ile aynı ya da farklı bir telekomünikasyon şebekesi üzerinden sunulabilir. Farklılık daha çok, temel hizmete bilgisayar uygulamaları aracılığıyla katılan yeni özelliklerden kaynaklanmaktadır.²³

Telekomünikasyon altyapısı üzerinden sunulan katma değerli hizmetlerin çeşitliliği de teknolojik gelişmelerle birlikte gün geçtikçe artmaktadır. Elektronik posta, data hizmetleri, bankaların ATM hizmetleri, video konferans, vs. örnek olarak verilebilir.

Bunlara, günümüzde yoğun olarak kullanılan mobil telefon, çağrı cihazı, elektronik veri iletimi, etkileşimli veri hizmetleri (evden bankacılık, evden alışveriş, evde ofis vb.), VolP (internet üzerinden telefon hizmeti), sayısal TV ve interaktif video hizmetleri gibi hizmetler diğer örnekler olarak eklenebilir.

2.1.3 Telekomünikasyon Cihazları

Telekomünikasyon cihazları, şebekenin sonuna eklenen telefon cihazları, mobil telefon cihazları, fax makinaları, bilgisayar gibi cihazlardır. Uzun yıllar boyunca dikey bütünleşmiş yapı içerisinde sürdürülen cihaz üretimi, kapsam ekonomisi söz konusu olduğu düşünülerek ulusal telekom tekellerinin kontrolünde bulunmuştur. Telekomünikasyon sektörü ekonomi ve rekabet açısından iki ayrı ana başlık altında incelenebilir; Şebeke (altyapı) işletimi ve altyapı üzerinden sunulan hizmetler.²⁴

²³ Ulusoy Ali, **Telekomünikasyon Hukuku**, Turhan Kitabevi, Ankara **2002**, s.172.

²⁴ SAFEL, R., **Telekomünikasyon Sektörü**, Türkiye Vakıflar Bankası T.A.O Planlama ve İktisadi Araştırmalar Grup Yönetmenliği Sektör Araştırmaları, Seri No: 25, Ekim 2001.

Bu anlamda altyapı, iki nokta arasında iletişim kurmayı sağlayan her türlü iletim şebekesi demektir. Hizmet sunucular ise, bu şebeke üzerinden farklı nitelikte birçok hizmet sunabilmektedirler.

Telekomünikasyon şebekesi birçok santral (anahtarlama ekipmanı) ve iletim hattından oluşan karmaşık bir görünüm sergilemesine karşın temelde hiyerarşik bir yapıya sahiptir. Bu yapının en altında yerel şebeke olarak adlandırılan, en düşük kapasiteli bakır kablolarla son kullanıcıya iletilen telekomünikasyon hizmetleri yer alırken, yukarıya doğru hem iletişim kapasitesi artmakta hem de buna paralel olarak kullanılan iletişim hatlarının (kablolar veya telsiz iletişim hatları) ve santrallerin kapasiteleri artmakta, en üstte ise uluslararası iletişimi sağlayan yüksek kapasiteli fiber optik iletim hatları veya uydular yer almaktadır. Bu çerçevede, telekomünikasyon altyapısı da, bu hiyerarşik yapı içerisinde yerel sabit şebeke, yerel mobil şebeke, uzun mesafe ve uluslararası iletişim şebekesi olarak kendi içinde farklı alt şebekelere bölünebilmektedir.

Farklı ürün nitelikleri, çıktının depolanamama özelliği, zamana göre değişen talep yapısı, batık maliyetlerden kaynaklanan faaliyet riski ve kapasite sınırlamaları, dışsallıklar ve doğal tekel niteliği telekomünikasyon sektörünün genel ekonomik niteliklerin başlıcalarını teşkil etmektedir.²⁵

Bilindiği üzere, doğal tekel bir hizmetin en düşük maliyetle tek bir teşebbüs tarafından sağlanabildiği durumlarda söz konusu olabilmektedir. Burada yerel şebekenin kurulum maliyetlerinin oldukça yüksek olması dikkate alındığında, bu şebekenin birden fazla teşebbüs tarafından rekabet içerisinde kurulmasının maliyeti çok yüksek olduğundan tek bir teşebbüsün bütün bir bölgeye şebeke kurması tercih edilmektedir. Son kullanıcı açısından bakıldığında, bu şebekeye bağlı olduğu sürece son kullanıcıya bu şebeke üzerinden sunulacak hizmetler bakımından şebekenin sahibinin erişim tekeli bulunmaktadır.

Ülkemizde Türk Telekom'un özelleştirilmesi çerçevesinde Rekabet Kurumu tarafından verilen görüşte, şebekelerarası rekabetin günümüzde teknolojik olarak

²⁵ DPT, Haberleşme Özel İhtisas Komisyonu Raporu, 2005, s.76

mümkün bulunduğundan bahisle, kablo-TV şebekesi üzerinden, günümüz koşullarında sabit yerel şebekeye en yakın rakip şebeke kurulabileceği, bu sayede halihazırda kurulu fiili tekel konumundaki Türk Telekom'la rekabet ortamının oluşabileceği ifade edilmiştir. Bu çerçevede yapılan düzenlemeyle, Kablo TV şebekesi, tüm müşteri ve müşterileriyle birlikte Türksat Uydu Haberleşme ve İşletme Anonim Şirketi'ne devredilmiştir. Ancak, uygulamada bu görüşün yüksek sabit yatırım maliyeti, sermaye israfı ve faaliyet riskleriyle karşı karşıya bulunduğu bir gerçekliktir.

Yerel şebeke (local loop) olarak adlandırılan ve evlere kadar ulaşan temel telekomünikasyon şebekesi bu açıdan doğal tekel olarak nitelendirilmektedir. Ancak telekomünikasyon sektörünü diğer doğal tekel niteliği gösteren sektörlerden ayıran en önemli özellik, sektörde faaliyet gösteren veya gösterebilecek olan aktörlerin çok sayıda olması olasılığıdır. Temel telekomünikasyon hizmetlerini sağlayan sabit ve mobil telefon operatörlerinin dışında, kablo-TV, elektrik, gaz, su, demiryolu gibi şebekelerin operatörlerinin de kendi şebekeleri boyunca çekecekleri hatlar vasıtasıyla telekomünikasyon hizmetleri sağlama imkânları bulunmaktadır.

Telekomünikasyon sektöründeki en büyük giriş engeli, özellikle sabit telekomünikasyon şebekesinin hem sabit maliyetlerinin, hem de batık maliyetlerinin, faaliyet risklerinin çok yüksek olmasıdır. Evlere kadar döşenen kablunun, hizmet sunulabilmesi açısından gerekli ve zorunlu olması, ancak başka bir yere taşınmasının mümkün olmaması nedeniyle, bu kablo döşeme maliyeti batık maliyet niteliğinin yanında, ayrıca sabit telefon talebinin de doyuma ulaştığı düşünüldüğünde büyük faaliyet riski de taşımaktadır.²⁶

2.2 Sektörde Türk Telekom'un Yeri ve Önemi

Günümüzde telekomünikasyon hizmetleri sosyal yaşamın ayrılmaz bir parçası olduğu gibi ekonomik kalkınmanın da en temel altyapısıdır. Telekomünikasyonun

²⁶ SÜEL, H., **Türk Telekomünikasyon Sektörü ve Özelleştirme**, İktisat İşletme ve Finans Dergisi, Nisan,1998.

ekonomik gelişme ve kalkınmadaki önemi, 1980'li yılların başından itibaren dünyayla birlikte ülkemiz ekonomisinde de gözlenen serbestleşme ve küreselleşme eğilimleriyle artmıştır. ²⁷ Bu süreçte, bir yandan bilgiye ve bilgi akışını sağlayan telekomünikasyon hizmetlerine talebin artması ve çeşitlenmesi, aynı doğrultuda telekomünikasyon sektörünü gelişmeye zorlamış, böylelikle hem telekomünikasyon hizmetleri gelişmiş, iç içe yürüyen bu gelişme dünya ekonomisinin küreselleşmesinde önemli rol oynamıştır. Küreselleşme kavramı çerçevesinde, dünya ekonomik sisteminin geçirdiği dönüşümler telekomünikasyon teknolojilerindeki gelişmelerle ilişkilendirilmekte, uluslararası işbölümü telekomünikasyon altyapıları üzerinde yeniden şekillendirilmekte, telekomünikasyona bağımlılığın düzeyinde ve telekomünikasyon aracılığı ile girilen etkinliklerin derecesinde niteliksel bir değişim yaşanmaktadır. Telekomünikasyon hizmetlerinin 1960'lar ve 70'ler boyunca kamu hizmeti olarak sunulması ve evrensel hizmet ilkesi ile yürütülmesi, öte yandan kalkınma düşünceleri, bu doğrultuda sağlık ve eğitim sektörlerinde telekomünikasyonun önemi, ulusal bütünleşme ve güvenlik açısından stratejik bir sektör olarak kabulü bu hizmetlerin kullanımının değişmesine neden olmuş ve telekomünikasyon hizmetlerini yeniden yapılanmakta olan dünya düzeninde önem kazanan hizmet sektörünün en önemli bileşeni haline getirmiştir. ²⁸

Piyasa ekonomisinin sağlıklı işleyişi, güvenilir bilginin uygun zamanda karar vericilere ulaşmasından geçtiğinden, ekonominin diğer sektörlerinin de altyapısı niteliğini taşıyan telekomünikasyon hizmetleri, mali piyasaların, haberleşme yoğun sektörlerden turizm ve ulaştırma hizmetlerinin ayrılmaz bir parçası olduğu gibi; üretici, toptancı, perakendeci ve tüketici zincirinde bilgi akışını sağlayarak her türlü mal ve hizmet piyasalarının arz ve talebindeki değişimlere uyumlu ve etkin bir şekilde işleyebilmesi için en önemli unsurlardan biri haline gelmiştir. Böylelikle telekomünikasyon, yakın geçmişten bu yana küreselleşen dünya ekonomisinde rekabetin en önemli unsurlarından biri halini almıştır. Bu nedenle, firmaların bilgi edinme ihtiyacı doğrultusunda telekomünikasyon hizmetlerine talebinin artmasıyla,

²⁷ Elektrik mühendisliği, 430. sayı, nisan 2007

²⁸ Nuray Keskin, s.7.

bu hizmetlerin firma ihtiyalarına gre tasarlanmıř olmasının, kaliteli ve uygun bir fiyatla sunulmasının nemini artırmaktadır.²⁹

Diğerk yandan, telekomnikasyon normal vatandaşlar iin de, insanlar arası iliřkilerin geliřtirilmesinin temel bir aracıdır. Kapitalizmin yeniden yapılanmasında enformasyon, ekonomik stnlğn srdrlmesinde kilit konumuna gelmektedir. Gelecekte, toplumun sahip olduėu enformasyon kaynakları, belirleyici bir unsur olacaktır.

Telekomnikasyon teknolojilerinin "istihbarat, iletiřim, komuta, kontrol" amalı kullanımları, geliřmiř devletler erevesinde, telekomnikasyon teknolojilerinde nder olmayı, dnya sisteminde hatırı sayılır g olma ya da bu durumunu srdrme aısından, egemenlik, iliřkilerin deėiřmesi ya da korunmasında teknolojiyi kontrol etmenin nemini artırmaktadır. Bilgi, sahibine ekstra politik bir g sağlamaktadır, bu nedenle haberleřme zerinde denetim sahibi olmak modern devletlerin varolma řartlarından birisi gibi grnmektedir.³⁰

Getiėimiz yzyılın ortalarında sanayileřmenin en nemli dinamiklerini oluřturan demir-elik, tekstil, otomotiv ve petrokimya gibi sektrler, giderek nemlerini kaybederek, yerlerini, byk lde enformasyona ynelik olan telekomnikasyon, elektronik ve iletiřim sektrlerine bırakmıřlardır. Bunun etkisiyle son yıllarda sanayi toplumunun tesinde bilgi toplumu kavramı yařamımızda yer edinmiřtir Bilgi toplumu ve bilgi aėı olarak adlandırılan yařadıėımız srete hemen her řey bilgi teknolojileri kullanılarak yapılmaktadır. Bylesi bir altyapının iřletilmesi ise kaınılmaz bir biimde telekomnikasyon řirketleri tarafından saėlanmaktadır. Bu atmosferde, telekomnikasyon sektrnn stratejik nemi giderek daha da bymektedir.³¹

²⁹ ELİBOL, Semih, **Telekomnikasyon Sektrnde Deėiřim ve Trk Telekom'da Yeniden Yapılanma Sreci**, İstanbul: Marmara niversitesi, 2002

³⁰ Haluk Geray, s.43.

³¹ Christodoulou, K., Vlahos, K., "Implications of regulation for entry and investment in the local loop", **Telecommunications Policy**, 2001, s 743-757.

Bu şartlar altındaki bir sektörde, ülkemizde de 1854 yılından itibaren verilmeye başlanan telekomünikasyon hizmetlerinin odağında Türk Telekomünikasyon A.Ş. yer almaktadır. 1926 yılında Ankara'da kurulan otomatik telefon santrali, 1928'de İzmir ve 1931 yılında İstanbul'da kurulan santraller Türkiye Cumhuriyetinin kuruluşundaki ilk telekomünikasyon yatırımları olmuştur. Ancak ikinci dünya savaşının başlamasıyla birlikte bu alandaki yatırımlara ara verilmiş ve uzun yıllar Türkiye'deki telefon yoğunluğu %0,2 düzeyinde kalmıştır. Daha sonra 1950'li yıllarda yeniden başlanan telekomünikasyon yatırımlarının, esas itibariyle 1980'li yıllardan itibaren önemli artışlar gösterdiği görülmektedir.³²

Günümüze gelindiğinde ise telekomünikasyon sektöründe rol alan kurum ve kuruluşlar üç ana grupta toplanmaktadır; devlet kurumları, düzenleyici kurumlar ve müteşebbislerdir. Sektörde faaliyet gösteren müteşebbisleri üç ana grupta toplayabiliriz. Bunlar; İşletmeciler, imalat ve ithalatçılar ve tüketicilerden oluşmaktadır.

Sektörün ana müteşebbisleri Telekomünikasyon Kurumundan yetki belgesi alan "İşletmeciler" 'dir. Bunları da Sabit, mobil ve diğer telekomünikasyon hizmet işletmecisi şeklinde gruplandırabiliriz.

Ülkemizde, mobil telekomünikasyon hizmet işletmecileri, Türkcell ve Telsim GSM 900 bandında, Aycell ve Aria'nın birleşimi olan Avea ise 1800 bandında olmak üzere üç adettir.³³

Mobil işletmeciler, SMS, WAP, GPRS, MMS gibi data hizmetlerini, akıllı şebekeler üzerinden coğrafi bilgi, özel faturalama, sanal kullanıcı platformu oluşturma, farklı tarife yapıları ve seçenekleri sunma, sesli mesaj, cep-mail gibi katma değerli hizmetler sunmaktadırlar.

³² Başbakanlık Yüksek Denetleme Kurulu Türk Telekomünikasyon A.Ş. 2001 yılı Raporu

³³ YILMAZ, K., **Türk Telekomünikasyon Sektöründe Reform: Özelleştirme, Düzenlenme ve Serbestleşme**, Persembe Konferansları, Rekabet Kurumu Yayını, Ankara, 2006.

Statüsü 406 sayılı yasa ile belirlenmiş ve kanun uyarınca görev sözleşmesi imzalamış olan Türk Telekomünikasyon A.Ş., her türlü telekomünikasyon hizmetlerini yürütmeye, telekomünikasyon altyapısını işletmeye yetkili kılınmıştır. Şirket, sabit telefon, ISDN, ADSL, NMT, kiralık devreler, internet servis sağlayıcılığı gibi birçok farklı hizmet sunmaktadır. ³⁴

Tablo 2.2'den de görülebileceği üzere, 2003 yılı verilerine yaklaşık 21 milyon hat kapasitesine karşılık yaklaşık 19 milyon telefon abonesi ile dünya çapında hatırı sayılır büyüklükte operatör konumundadır. Diğer telekomünikasyon işletmecileri olarak, uydu işletmecileri, ³⁵ GMPCS ³⁶ mobil telefon hizmeti veren işletmeciler, internet servis sağlayıcılar, uzak mesafe telefon hizmeti veren işletmeciler olarak ifade edilebilir. Genel anlamda, sabit telefon gelirleri, mobil telefon gelirleri ve bilgi teknolojisi gelirleri bilişim teknolojileri sektörünün Pazar büyüklüğünü göstermektedir. ³⁷

Telekomünikasyon sektörü Pazar payları incelendiğinde, Türk Telekom, yakın geçmişe kadar doğal tekel olarak yürüttüğü Telekom işletmeciliğiyle, sadece sabit telefon gelirlerinde kaynaklanan %54'lük Pazar payıyla sektörde ilk sırada yer almaktadır.

³⁴ Türk Telekomünikasyon A.Ş. 2003 yılı Faaliyet Raporu

³⁵ Uydu işletmeciliği 2004 yılına değin Türk Telekom tarafından yürütülmekte iken 16.06.2004 tarih ve 5189 sayılı Kanunun ek 33. maddesi uyarınca kurulan Türksat Uydu Haberleşme ve İşletme Anonim Şirketi'ne tüm aktif ve pasifiyle birlikte devredilmiştir. Daha sonra, Türk Telekom'un özelleştirilmesi çerçevesinde Rekabet Kurumunca verilen görüş uyarınca, yine Türk Telekom bünyesinde bulunan Kablo TV hizmetleri de anılan şirkete devredilmiştir.

³⁶ Pozisyonu ve çalışma frekansları ITU tarafından belirlenmiş, tahsis edilmiş bulunan ve bir sabit veya mobil, geniş band veya dar band, küresel veya küresel olmayan, mevcut veya planlanan bir uydu kümesi üzerinden GMPCS MoU kapsamındaki hizmetleri kullanıcılara doğrudan veren telekomünikasyon hizmeti şeklinde tanımlanmaktadır. Faaliyet Raporu, TT, 2003.

³⁷ Nur Saygı: **Telekomünikasyon ve Bilgi teknolojileri Pazarı**, Ankara: Telekomünikasyon Kurumu, 2002, s.9.

Tablo 2. 2 Türkiye Telekom Pazarı Büyüklüğü

2010 Yılı Türkiye Telekom Pazarı Büyüklüğü	
Net Satışlar	(Milyon TL)
PSTN	6.766.931.383
GSM	4.123.970.054
Üreticiler	1.055.735.755
ISS (TT Hariç)	230.841.348
Diğer	290.583.092
Toplam	12.468.061.632

Ancak, bu gösterge dahi, Türk Telekom'un sektördeki yerini tam olarak göstermeye yetmeyecektir. Çünkü tüm telekomünikasyon alt yapısını da fiilen işlettiğinden, ara bağlantıyla verdiği transmisyon hizmetinden kaynaklanan önderliği ve fiili hâkimiyeti, Türk Telekom'u hemen her türlü haberleşmenin aktörlerinden birisi kılmaktadır.

2.3 Türk Telekom'un Tarihsel Gelişimi, Hukuki Yapısı

Türkiye'de telekomünikasyonun sektörünün geçmişi, Osmanlı dönemi, 23 Ekim 1840 tarihinde Sultan Abdülmecit tarafından Postahane-i Amirane'nin kurulması ile başlamaktadır. Telgraf Nezaretinin 1855 yılında kurulmasından sonra 1871'de Posta ve Telgraf nezaretleri birleştirilerek, Posta ve Telgraf Nezaretine dönüştürülmüştür.

Telefon hizmeti ise ilk olarak, Temmuz 1881'de İstanbul Soğukçeşme'deki Posta ve Telgraf Nezareti binasıyla Yeni Cami'deki postane arasına çekilen tek telli bir telefon ile verilmiştir.³⁸

Türkiye'de 1911 yılında İstanbul'da yabancı şirket tarafından kurulan ilk telefon sistemine I. Dünya Savaşı sırasında devlet tarafından el konulmuş, Cumhuriyetin ilanından sonra da 800.000 İngiliz Lirası bedel ile satın alınmıştır.

³⁸ ALPTÜRK, H., **Türk Telekom Radyo TV _letisim Fuarında**, Türk Telekom Dergisi, Eylül-Ekim 2000.

Cumhuriyetin ilanından sonraki dönemde ise, 21 Şubat 1924 tarihinde kabul edilen 406 sayılı Telgraf ve Telefon Kanunu ile Osmanlı devletinden devralınan ve özel şirketler eliyle yürütülen telgraf ve telefon iletişimi yeniden düzenlenmiştir. Posta Telgraf ve Telefon İdaresi (PTT) 1936'da çıkarılan Kanunla kurulmuş ve ülke çapında telekomünikasyon hizmetlerini sağlamakla görevlendirilmiştir. PTT İdaresi, 1939'da Ulaştırma Bakanlığına bağlanmış, 1953 yılında da Kamu İktisadi Teşebbüsü olarak yeniden yapılandırılmıştır.³⁹

1983 yılında işbaşına gelen hükümetin izlediği ekonomik model ve devletin ekonomideki ağırlığının azaltılması politikası çerçevesinde, kamu kesiminin imalat sanayii yatırımları kısılmış, haberleşme, enerji ve ulaştırma gibi altyapı yatırımlarına ağırlık verilmiştir.⁴⁰

Bu dönemde hız kazanan altyapı yatırımları çerçevesinde, özellikle telekomünikasyon sektöründe büyük gelişmeler yaşanmış, 1982-1992 döneminde Türkiye, ITU üyesi Avrupa ülkeleri arasında ana hat sayısı artışında en yüksek gelişme hızını gösteren ülke olmuştur.⁴¹

Telekomünikasyon şebekesinin geliştirilmesi ve genişletilmesinde finansman aracı olarak vergi gelirleri veya kamu borçlanması yerine öz kaynakların kullanılması politikası çerçevesinde, PTT'nin izlediği tekeli fiyatlandırma politikası nedeniyle gelirleri artmış ve şebekenin modernizasyonu sonucunda maliyetleri reel olarak azalmıştır. Artan verimlilik kurumun mali yapısını güçlendirmiş ve yatırımların otofinansmanı imkânını sağlamıştır.⁴² Bu şekilde yaratılan kaynaklar ile gerçekleştirilen yatırımlar sayesinde çağrı yönlendirme, çağrı bekletme, konferans, 118 bilinmeyen numaralar gibi yeni hizmetlerin sunulması sağlanmıştır.

Telekomünikasyon sektöründe 1983-2005 yılları arasında sağlanan gelişmeler tablo 2.3'de kısaca özetlenmiştir.

³⁹ T.C. Telekomünikasyon Kurumu, “**Ortak Yerleşim ve Tesis Paylaşımı Usul ve Esaslarına ilişkin Tebliğ**”, T.C. Telekomünikasyon Kurumu, Ankara, 2004, s.1-7.

⁴⁰ Recep Çakal, s.50.

⁴¹ Kamil Yılmaz: **Türk Telekomünikasyon Sektöründe Reform: Özelleştirme, Düzenleme ve Serbestleşme**, İstanbul: Koç Üniversitesi, 1999, s.1-37.

⁴² Recep Çakal, s.110.

Tablo 2. 3 Telekomünikasyon Sektöründe 1983-2012 Yılları Arasında Gelişmeler

1989 yılı itibariyle bütün köylere telefon hizmeti götürülerek 1993 yılı sonunda mezralar dâhil 46.000 kırsal yerleşim alanı telefona kavuşmuştur.
Ankesörlü telefon sayısı 1983 yılında 7 bin civarında iken 2004 yılı sonu itibariyle 76.921 adet olmuştur. Smart kartlı ankesör projesi kapsamında 70.000 adet smart kartlı ve 100.000 adet internet erişimli ankesör alımı yapılmış, hizmete sunumu çalışmaları devam etmektedir.
Telefon abone yoğunluğu % 5'ten 2004 yılı sonu itibariyle % 26.45'e çıkmış santrallerin otomatikleşme oranı yüzde yüze yaklaşmıştır.
1980'lerin başında sunulamayan veri iletim hizmetleri, mobil telefon hizmetleri, hareket halindeki araçlar arasında ve araçlarla sabit noktalar arasında haberleşme imkanı sağlayan uydu hizmetleri gibi hizmetler sunulmaya başlanmıştır.
Faks Hizmeti: 1984 yılında 4 büyük ilde başlatılan ve 1986 yılından sonra talebi artan faks hizmeti ülke geneline yayılmıştır.
Çağrı Telefon Hizmetleri: 1986 yılında hizmete başlayan çağrı telefon sistemi abone sayısı 1993 yılında 100.000'e ulaşmış, ancak mobil telefon hizmetlerinin gelişmesiyle birlikte işlerliğini kaybetmiş, 2004 yılı sonu itibariyle 2.277 aboneye hizmet verir haldedir.
Video konferans Hizmeti: 1986 yılında Ankara'da daha sonra İstanbul ve İzmir'de sabit stüdyolar kurulmuş, ancak videokonferans hizmeti yaygınlık kazanamamıştır.
Araç Telefonu: 1986 yılında analog olan NMT 450 sistemi ile başlatılmıştır. Abone sayısı 1990 yılından sonra artış kaydetmiş ve 1994 sonunda 100 bine yaklaşmıştır. Ancak, yine mobil telefon hizmetlerindeki gelişme, hizmet kalitesi ve teknolojisindeki sıkıntılar nedeniyle yaygınlık kazanamamıştır. 2004 yılı sonu itibariyle 28.364 aboneye hizmet vermektedir.
800 Kodlu Hizmet: Bedeli aranan abone tarafından ödenen 800 Kodlu hizmet 1994 yılında hizmete verilmiştir.

Satış sonrası hizmet ve tüketici ilişkilerinin güçlendirilmesi amacıyla birkaç büyük firma tarafından kullanımına başlanmıştır.
Elektronik Posta: Türkiye Elektronik Posta Sistemi (TÜRKEP), 1993 yılında başlangıçta 5.000 kullanıcı kapasitesiyle hizmet vermeye başlamış, günümüzde internet ve veri hizmetlerinin gelişmesiyle bu alanda hizmet alan müşterilere benzer hizmetler geniş kapsamlı olarak sunulmaktadır.
Görüntülü Telefon: ISDN'e Uyumlu Dahili Abone Santralı Projesi çerçevesinde aynı bina içinde olmak kaydıyla görüntülü telefon hizmeti 1993 yılında hizmete verilebilir hale gelmiştir.
Cep Telefonu: 1994 yılında iki ayrı firmaya (Telsim ve Turkcell) gelir ortaklığına dayalı sözleşmeyle Pan Avrupa Sayısal Mobil Sistem (Group Speciale Mobile- GSM 900) mobil telefon hizmetleri başlatılmıştır.
Veri Haberleşmesi: Veri transferinde kiralık özel devrelerin kullanılmasına 1982 yılında başlanılmıştır.
Uydu Haberleşme ve Yer İstasyonları: 1993 yılında, üç uydu üzerinden hizmet veren 22 alıcı/verici özellikte uydu yer istasyonuna ulaşılmıştır. Mobil uydu yer istasyonları sayısı ise 6 adede ulaşmıştır. Türk Cumhuriyetlerinin başkentlerinde kurulan uydu yer istasyonları vasıtasıyla bu ülkelerin uluslararası telefon haberleşmelerinin Türkiye üzerinden yapılması sağlanmıştır.
Üç uluslararası uydu organizasyonundan transponder kiralananak yürütülen uydu haberleşme hizmetlerinin, yurtiçi haberleşmenin ve televizyon yayınlarının daha kaliteli yapılabilmesi ve yaygınlaştırılması amacıyla Türkiye'ye ait bir uydu (TÜRKSAT) Mayıs 1994 tarihinde yörüngeye yerleştirilmiş ve Eylül 1994 tarihinde hizmete verilmiştir. Uydu işletmeciliği 2004 yılına değin Türk Telekom tarafından yürütülmekte iken 16.06.2004 tarih ve 5189 sayılı Kanunun ek 33. maddesi uyarınca kurulan Türksat Uydu Haberleşme ve İşletme Anonim Şirketi'ne tüm aktif ve pasifyle birlikte devredilmiştir.
Kablo-TV Hizmeti: 1989 yılında başlanmıştır. 3,5 milyon abonenin öngörüldüğü, 10 ili kapsayan ve başlangıçta 20 TV, 20 radyo (daha sonra 35

TV, 40 Radyo) kanalının yayımına olanak sağlayan kablo-TV sistemi 1991 yılında 6 şirketler birliğine ihale edilmiştir. Bu hizmet, Türk Telekom'un özelleştirilmesi çerçevesinde Rekabet Kurumunca verilen görüş uyarınca, 16.06.2004 tarih ve 5189 sayılı Kanunla kurulan Türksat Uydu Haberleşme ve İşletme Anonim Şirketi'ne devredilmiştir.
1997'de Türkiye'deki tüm üniversiteleri TURPAK şebekesi üzerinden birbirine bağlayan Ulusal Akademik Ağ (ULAKNET) projesi hayata geçirildi. 1997'de KAFOS (Karadeniz Fiber Optik Sistemi) ve TBL (Transbalkan Linki Karasal Fiber Optik Sistemi) hizmete verildi.
27 Nisan 1998'de GSM lisansı, 25 yıllığına Turkcell ve Telsim şirketlerine devredildi.
1998'de İnternet erişimini Türkiye geneline yaymak, hızlı ve kaliteli hizmet sunabilmek amacıyla TTNET ulusal internet alt yapı ağına ilişkin sözleşme imzalandı.
2000'de Kablo TV üzerinden Kablo İnternet uygulamaları başlatıldı.
2001'de Türk Telekom ve Alcatel ortaklığı ile kurulan EURASIASAT şirketi tarafından yaptırılan TÜRKSAT 2A uydusu Güney Amerika'daki Kourou Üssü'nden uzaya fırlatıldı.
Mart 2002'de Altyapısı Türk Telekom tarafından sağlanan nüfus işlemlerinin bilgisayar ortamında yapılması için oluşturulan MERNİS Projesi uygulamaya konuldu.
2003'te Deneme amaçlı ilk Metro Ethernet uygulamaları başlatıldı.
2003'te 42500 okulun ADSL hizmeti ile internete bağlanması için Milli Eğitim Bakanlığı ile ortak proje başlatıldı.
2004'te Türk Telekom'un GSM Operatörü Aycell ile İŞ-TİM 'in birleşmesi sonucu kurulan TT&TİM İletişim Hizmetleri A.Ş., "Avea İletişim Hizmetleri A.Ş." ticari ünvanı ile 19 Şubat 2004 tarihinde resmen kuruldu.
2004'te İstanbul Türkiye ve Katanya (İtalya) arasındaki MEDTÜRK Denizaltı Fiber Optik Kablo Sistemi hizmete verildi.
2004'te Türk Telekom'un ulusal ve uluslararası uydu haberleşmesiyle ilgili tüm hak ve sorumlulukları, 16.06.2004 tarihli 5189 nolu yasa ile kurulan TÜRKSAT Uydu Haberleşme ve İşletme Anonim Şirketi'ne (TÜRKSAT A.Ş.) devredildi.

2004'te Kamu ve özel kuruluşların İnternet Veri Merkezi ihtiyaçlarının karşılanması amacıyla, Web Hosting, Mail Hosting ve Co-location gibi katma değerli IDC hizmetlerinin verilebileceği TTIDC (Türk Telekom İnternet Data Center) kuruldu.

2004'te Resmi kurum ve kuruluşlar ile kurumsal müşterilerin yurt çapındaki birimlerinin noktadan noktaya data transferlerini sağlamak için noktadan noktaya DSL hizmetleri verilmeye başlandı.

2005'te 2005/8409 sayılı Bakanlar Kurulu Kararı ile T.C. e-Devlet Kapısı'nın kurulması görev ve sorumluluğu Türk Telekom'a verildi.

2005'te TTwinet- Kablosuz İnternet Servisi hotspot olarak isimlendirilen (oteller, havaalanları, alışveriş merkezleri, kafeler, restaurantlar, fuar merkezleri, üniversiteler v.b) 160 yerde hizmete verildi. ⁴³

Kaynak: Sekizinci beş yıllık kalkınma planı, Haberleşme Özel İhtisas Komisyonu Raporu, Ankara 2001.

⁴³ Türk Telekom 2003 ve 2004 yılları Faaliyet Raporları, http://www.telekom.gov.tr/webtech/default.asp?sayfa_id=73, (15 Temmuz 2010).

1994 yılında posta ve telekomünikasyon hizmetlerinin ayrılmasına kadar; PTT ülkenin her tarafında açık ve kapalı mektuplar ve posta kartlarını kabul etmek, taşımak ve dağıtmak, 2813 sayılı Telsiz Kanunu Hükümleri saklı kalmak kaydıyla telgraf, telefon ve her türlü haberleşme tesislerini kurmak ve işletmek ile radyo-televizyon verici istasyonları ile program linklerini planlamak, projelendirmek, yenilemek ve işletmek görevlerini tekel olarak yürütmekle yetkili olarak faaliyet göstermiştir.⁴⁴

Bu süreç içerisinde; sektörde fiyat tarifelerini belirleme, cihaz onayı verme, diğer operatörlerin kendi şebekesine erişiminin sağlanması, bunlara uygulanacak fiyat tarifeleri gibi düzenleme görevlerini PTT, daha sonra da Türk Telekom A.Ş. olarak kendisi yerine getirmiştir.⁴⁵

Telekomünikasyon sektörünün rekabete açılan ilk bölümü olan telekomünikasyon cihazları piyasasında serbestleşme Türkiye'de birçok OECD ülkesinden önce gerçekleşmiş ve PTT iştirakleri olan cihaz üreticisi şirketleri özelleştirilerek sektör rekabete açılmıştır.

PTT'nin TELETAS'ta % 40, NETAS'ta % 49 hissesi önce Kamu Ortaklığı İdaresine (KOİ) devredilmiş, daha sonra halka arz ve blok satış yöntemleriyle özelleştirilmiştir. TELETAS'ta bulunan % 40 oranındaki kamu payının % 22'si, 29 Şubat 1988 tarihinde halka arz edilmiştir. 21 Ağustos 1993 tarihinde, TELETAS hisselerinin KOİ'ye ait % 18'lik payı, blok olarak ALCATEL'e satılmıştır. KOİ'nin elinde hisse kalmadığından Altın Hisse imtiyazı da kalkmıştır.⁴⁶

NETAS'ta ise önce kamuya ait hisseler blok olarak satılmış, daha sonra halka arz edilmiştir. KOİ'de kalan % 49'luk payın % 20'si 1 Mart 1993 tarihinde, şirkette % 31 paya sahip olan yabancı ortak Northern Telecom Ltd. (NTL) blok olarak satılmıştır. Daha sonra geriye kalan % 29'luk payın % 7,75'i çeşitli tarihlerde halka arz edilmiştir. NETAS hisselerinin KOİ'nin elinde bulunan % 15,32'lik son kısmı

⁴⁴ SARISU, A., **Dünya'da ve Türkiye'de Özelleştirme**, Genel Bir Değerlendirme, Kasım, 2003.

⁴⁵ TUNCER, B., **Türk Telekom'un Tarihçesi**, <http://www.palu.gov.tr/telekom.htm>, (10.01.2010)

⁴⁶ Özge İçöz, s.45.

1993 yılı Aralık ayında, yabancı kurumsal yatırımcılar da dâhil olmak üzere İMKB'de halka arz edilmiştir.

Bu firmaların kamunun elinden çıkması ve tip onayı alan her cihazın şebekeye bağlanabilmesi imkânının sağlanması ile piyasaya girişler artmış ve sektörde rekabet sağlanmıştır.

Türkiye'de telekomünikasyon hizmetleri 1 Temmuz 1994 tarihine kadar, T.C. Posta, Telgraf ve Telefon İşletmesi Genel Müdürlüğü (PTT) tarafından 233 sayılı Kanun Hükmünde Kararnameye tabi bir KİT şeklinde ve tekel olarak yürütülmüştür. Telekomünikasyon sektöründe özelleştirme ise, ilk olarak Morgan Guarantee Bankası tarafından hazırlanan "Özelleştirme Ana Planı"nda gündeme gelmiş, buna göre, telekomünikasyon alanındaki özelleştirme ile ilgili olarak önce posta ve telekomünikasyon hizmetlerinin ayrılması öngörülmüştür.⁴⁷ Bu ayrımın nedeni, başlangıçta kapsam ekonomisi⁴⁸ gerekçesiyle bir arada yürütülen posta ve telekomünikasyon hizmetlerinin bu niteliğinin, özellikle telekomünikasyon sektöründe yaşanan hızlı teknolojik gelişmeler sonucunda ortadan kalkması ve her iki faaliyetin giderek birbirini olumsuz yönde etkilemesi olarak ifade edilmektedir.⁴⁹

Bu çerçevede, PTT bünyesindeki bu iki faaliyetin ayrılması ve telekomünikasyon sektöründe yeniden yapılanmayı amaçlayan ilk yasal düzenleme 1993 yılında 509 sayılı KHK ile yapılmıştır. 20.8.1993 tarih ve 509 sayılı KHK ile PTT'nin telekomünikasyon ve posta hizmetleri ayrılarak yeni bir şirket kurulmuştur. Ancak 509 sayılı KHK 21.10.1993 tarihli ve 93/40-2 sayılı Anayasa Mahkemesi Kararı ile iptal edilmiştir. Anayasa Mahkemesinin gerekçeli iptal kararında özelleştirme usul ve esaslarının, satışa arz edilecek hisselerden T.C. Posta İşletmesi Genel Müdürlüğüne ayrılacak payın, işletme lisans ve ruhsatlarından alınacak

⁴⁷ Özge İçöz, s.50.

⁴⁸ Gökhan Evren: **Hizmete ve Altyapıya Dayalı Rekabet ile Geçiş Stratejileri** , Ankara: Telekomünikasyon Kurumu, 2005, s.7.

⁴⁹ Posta hizmetleri, kamu hizmeti niteliği nedeniyle uzun yıllar boyunca telekomünikasyon hizmet gelirleri ile sübvansede edilmiş, bu süreç halen PTT'nin telekomünikasyon hizmetlerinden doğan alacakların Türk Telekom adına tahsilatından % 4.5 oranında komisyon alınması yöntemiyle dolaylı olarak sürdürülmektedir.

cretler ile sermaye Őirketlerinin iŐletme hâsılatından ve hisse satıŐ hasılatından Trk Telekom'a ayrılacak payın belirlenmesi hususlarında UlaŐtırma Bakanına verilen yetkilerin bizzat yasa koyucu tarafından yasa kuralıyla yapılması gerektiĐi ve bu dzenlemelerin yetki devri anlamına geldiĐi belirtilerek, bu durumun Anayasa'nın 7'nci maddesine aykırı olduĐu gerekesiyle iptal edildiĐi belirtilmektedir.⁵⁰

Bu iptal kararı nedeniyle, 10.6.1994 tarih ve 4000 sayılı Kanun ile 406 sayılı Kanun'da yeni dzenlemeler yapılmıŐ, 406 sayılı Kanun'un 1. Maddesi deĐiŐtirilerek posta ve telgraf tesisi ve iŐletmesine iliŐkin hizmetleri iin T.C. Posta İŐletmesi (Pİ), telekomnikasyon hizmetleri iin ise 233 sayılı KHK ile Trk Ticaret Kanununa (kuruluŐ ve tesciline iliŐkin hkmler hari) ve zel hukuk hkmlerine tabi Trk Telekomnikasyon Anonim Őirketi kurulmuŐtur. Ana SzleŐmesi UlaŐtırma BakanlıĐı tarafından onaylanan Trk Telekom A.Ő. 1 Temmuz 1994'te faaliyetlerine baŐlamıŐtır.

4000 sayılı Kanun'da Trk Telekom hisselerinin % 49'unun, satılması, satıŐa arz edilecek hisselerden T.C. Posta İŐletmesi Genel MdrlĐne ve PTT emeklileri ile alıŐanlarına tanınacak ayrıcalıkları da kapsayacak Őekilde satılmasına iliŐkin usul ve esasların belirlenmesi hususlarında UlaŐtırma Bakanı yetkili kılınmıŐtır. Ayrıca, BakanlıĐa, mobil telefon, aĐrı cihazı, veri Őebekesi, akıllı Őebeke, kablo-TV, ankesrl telefon, uydu sistemleri, rehber basım ve benzeri katma deĐerli hizmetler konularında sermaye Őirketlerine tekel oluŐurtmayacak koŐulları da dikkate almak suretiyle iŐletme lisans ve ruhsatı verilmesi konusunda yetki verilmiŐtir.⁵¹

Bu erevede 4000, 4107 ve 4161 sayılı Kanunlar ve Anayasa Mahkemesi Kararları ile deĐiŐik 406 sayılı Kanun'un ilgili hkmleri ile Trk Telekomnikasyon Anonim Őirketi hisselerinin en ok % 49'unun; % 10'u T.C. Posta İŐletmesi Genel MdrlĐne bedelsiz olarak verilmek, % 5'i Őirket alıŐanlarına ve kk tasarruf sahiplerine ve % 34' ise gerek ve tzel kiŐilere satılmak suretiyle devredilebileceĐi

⁵⁰ Recep akal, s.118.

⁵¹ FALAY, N., **zelleŐtirme Yntemleri ve Sorunlar**, zelleŐtirme TartıŐmaları, BaĐlam Yayınları, Yayın No: 72

hükme bağlanmıştır. Hisselerinin değerinin tespiti ve satış işlemlerinin 4046 sayılı Kanun hükümleri uyarınca Özelleştirme İdaresi Başkanlığınca yürütülmesi ve sonuçlandırılması öngörülmüş, blok satışlarda nihai devir işlemlerinin onaylanma yetkisi Bakanlar Kuruluna verilmiştir.

1998 yılında Bakanlar Kurulu Türk Telekom hisselerinin % 100 üne tekabül eden değerinin en az 10 milyar ABD Doları olduğuna ve kanunen satışı öngörülen hisselerin % 20'sinin blok olarak stratejik ortağa, % 14'ünün ise yurtiçi ve yurtdışında halka arz yöntemiyle satışına karar vermiştir.

Türk Telekom'un daha önce stratejik ortağa % 20 hissesinin satışı yoluyla özelleştirilmesi amacıyla Haziran 2000'de blok satış için ihaleye çıkılmış, fakat teklif gelmemesi sebebiyle ihale sonuçlandırılmamıştır. Daha sonra blok satış için öngörülen miktar % 20'den % 33,5'e çıkarılarak yeniden ihaleye çıkılmışsa da ihale yine sonuca ulaşamamıştır.⁵²

4502 sayılı Kanun çerçevesinde, Türk Telekomünikasyon A.Ş., her türlü telekomünikasyon hizmetlerini yürütmeye, telekomünikasyon alt yapısını işletmeye yetkili kılınmış, telekomünikasyon şebekeleri üzerinden sunulan ulusal ve uluslararası ses iletimini ihtiva eden telefon hizmetlerini yürütmesi ve bazı istisnalar dışında tüm telekomünikasyon alt yapısının kurulması, işletilmesi ve ses iletimi 31/12/2003 tarihine kadar Türk Telekomünikasyon A.Ş.'nin tekeline bırakılmış ve Türk Telekomünikasyon A.Ş.'nin, söz konusu hak ve yükümlülüklerinin Ulaştırma Bakanlığı ile imzalayacağı "Görev Sözleşmesi" ve/veya "Görev Sözleşmelerinde" belirleneceği hükme bağlanmıştır.

Türk Telekomünikasyon A.Ş.'nin tekeline bırakılan iş ve hizmetlerin dışındaki telekomünikasyon iş ve hizmetlerinin, Ulaştırma Bakanlığınca belirlenecek esaslara

⁵² ORUÇ, vd, **Özelleştirme, Serbestleşme ve Düzenleme Etkilesimi**, Sektörel Arastırma ve Stajyerler Dairesi Başkanlığı, Telekomünikasyon Kurumu, Mayıs, 2003.

göre imtiyaz, ruhsat ve genel izin yoluyla özel kesime de yaptırılması imkânı sağlanmıştır.⁵³

Anılan Kanunla verilen görevler çerçevesinde, Türk Telekom yurt içi ve yurt dışı haberleşme ve veri transferini sağlamak amacıyla faaliyet göstermiş, bu doğrultuda yurt çapında sahip olduğu telefon santralleri, şehiriçi ve şehirlerarası sabit hat şebekesi, uydu haberleşme ağları, kablo sistemleri ve veri ağlarını işletmiştir. Türk Telekom gelir ortaklığı anlaşmaları yoluyla, analog mobil telefon, Kablo TV, elektronik veri aktarımı hizmetlerini de yerine getirmiştir.

23.05.2001 tarihinde yürürlüğe giren 4673 sayılı Kanunla Türk Telekom'un 1 adet imtiyazlı hisse haricindeki tüm hisselerinin satılabilmesi mümkün kılınmış, telekomünikasyon alanında lisans verme yetkisi Ulaştırma Bakanlığından alınarak Telekomünikasyon Kurumuna devredilmiştir.

Gelinen süreçte bu noktaya kadar yürütülen özelleştirme çalışmalarında, Türk Telekom'un blok satışı için iki kez ihaleye çıkılmış ancak, dünya telekomünikasyon sektöründe yaşanan daralma ve çeşitli hukuki sorunlar sebebiyle istenilen sonuçlar elde edilememiştir. Global telekomünikasyon piyasalarının canlandığı yakın geçmişte, 25.11.2004 tarihinden itibaren yayımlanan ilanlarla Türk Telekom ihale süreci yeniden başlatılmıştır. 13.11.2003 tarih ve 2003/6403 sayılı Bakanlar Kurulu Kararı ile şirket hisselerinin en az %51'inin tek seferde blok olarak satılması ve blok satışı müteakip, kalan hisselerin Bakanlar Kurulunca belirlenecek süreç çerçevesinde halka arz edilmesi kararlaştırılmıştır.⁵⁴

Diğer taraftan, blok olarak satışa ilişkin ihale öncesinde, potansiyel yatırımcıları, Türk Telekom'un hukuki, operasyonel ve mali durumuna ilişkin olarak aydınlatmak, satış sürecine ilişkin bilgi vermek ve kendilerinin de özelleştirme sürecine ilişkin görüşlerini almak üzere bir "Bilgilendirme Süreci" işletilmiş, toplam 11 yerli ve yabancı yatırımcının katıldığı süreç 31 Temmuz 2004 tarihinde sona

⁵³ T.C. Telekomünikasyon Kurumu, “Yerel Aga Ayrıştırılmış Erisime ilişkin Usul ve Esaslar hakkında Tebliğ”, T.C. Telekomünikasyon Kurumu, Ankara, 2004, s.5-9.

⁵⁴ Civanlar, S, **Yeni Servis Altyapısına Hazırmıyız**, Telekom Dünyası Dergisi, İstanbul, Mayıs, 2005.

ermiştir. 15.10.2004 tarihli Bakanlar Kurulu Kararı ile Şirket hisselerinin %55'nin blok olarak satılması ve bu çerçevede 31.12.2004 tarihine kadar ihale ilanına çıkılması hükme bağlanmıştır.⁵⁵

25.11.2004 tarihinde yerli ve yabancı basın organlarına verilen ilanlarla başlayan ihale sürecinin son teklif verme tarihi olan 24.06.2005 tarihinde 4 adet teklif gelmiş, 01.07.2005 tarihinde finansal teklif zarfları açılarak aynı gün yapılan pazarlık görüşmeleri sonucunda Oger Telecoms Ortak Girişim Grubu 6.550.000.000 ABD Doları, Etisalat Ortak Girişim Grubu ise 6.500.000.000 ABD Doları tutarında teklif vermişlerdir. 01.07.2005 tarihinde yapılan ihalenin sonuçlarına yönelik olarak alınan İhale Komisyonu Kararı, 1998/4 sayılı Özelleştirme Yoluyla Devralmaların Hukuki Geçerlilik Kazanabilmeleri İçin Rekabet Kurumuna Yapılacak Ön Bildirimlerde ve İzin Başvurularında Takip Edilecek Usul ve Esaslar Hakkında Tebliğ çerçevesinde Rekabet Kurumu'na iletilmiş ve Rekabet Kurumu izni ise 21.07.2005 tarih ve 05-48/681-175 sayılı Rekabet Kurulu Kararı ile verilmiştir. Türk Telekom'un %55 oranındaki hissesinin blok olarak satışına ilişkin nihai devir işlemi 25/07/2005 tarih ve 2005/9146 sayılı Bakanlar Kurulu Kararı ile onaylanarak 02.08.2005 tarih ve 25894 sayılı Resmi Gazete'de yayınlanarak yürürlüğe girmiştir. İhale sonucunun Resmi Gazete'de yayınlanmasıyla beraber ihalede kazanan teklif sahibi Oger Telecoms Ortak Girişim Grubu olarak belirlenmiş olup, bu grup ile Hisse Satış Sözleşmesi imzalanmıştır.⁵⁶

⁵⁵ Mizrake, Yıldırım, A, **Türk Telekom'da Toplam Kalite Yönetimine Doğru**, Türk Telekom Dergisi, Mart Nisan 2000, Sayı: 2, Kozan Ofset Matbaası.

⁵⁶ T.C. Telekomünikasyon Kurumu, "**Erişim ve Arabaglantı Yönetmeliği**", T.C. Telekomünikasyon Kurumu, Ankara, 2003, 1-12 .

3. VERİ MADENCİLİĞİ

Veri madenciliği dünyanın anlaşılabilirliğine önemli olcude destek olan bir kavramdır. Gelişen bilgi toplama, depolama ve işleme yetkinlikleri, giderek artan bir şekilde mevcut verilerin incelenerek anlamlı sonuçlar elde edilmesine olanak sağlamaktadır. Artık hangi genlerin hangi hastalığa neden olduğunu, hangi müşterilerin kredisini geri odayemeyeceğini, hangi koşullarda yağmur yağacağını, hangi filmin yüksek gişe hasılatı yapacağını veya müşterilerin bir sonraki alışverişlerinde hangi ürünleri alacaklarını bilebilmek çok şaşırtıcı değildir. Asıl şaşırtıcı olan bu sonuçlara ulaşabilecek kurumların, verilerini topluyor olmalarına rağmen bu verileri anlamlı bilgilere dönüştürmüyor olmalarıdır.⁵⁷

Müşterilerin, şirketin ürünlerini almaktan vazgeçerek rakip şirketin ürünlerini tercih etmeleri birçok endüstride giderek büyüyen bir sorundur. Müşterilerin bir firmadan diğer firmaya geçmesinin en önemli sebebi çoğunlukla daha iyi bir teklif almış olmalarıdır. Örneğin bankalar düşük faiz oranları ile rakiplerinin kredi kartı müşterilerinin kendi kredi kartlarını kullanmalarını sağlamaktadırlar. Kredi kartı kullanımı yaygınlaştıkça yeni kredi kartı müşterileri pazarı oldukça küçülmüştür. Bu da kredi kartı sunan bankaları yeni müşteri bulabilmek için rakiplerinin müşterilerini elde etmek zorunda bırakmaktadır. Bankalar müşterileri çekebilmek için kısa bir dönem için daha düşük kredi oranları sunmakta, oranlar normale döndüğünde müşterilerin bankada kalacaklarını düşünmektedirler. Telekomunikasyon sektöründe en önemli sorun müşteri kaybıdır. Şirketler müşterilerinin rakiplerine geçmesini engellemek için çeşitli pazarlama taktikleri uygulamakta, ürünleri sürekli yeni sunular ile çeşitlendirmektedirler.⁵⁸

Kuruluşlar hangi müşterilerini kaybedebileceklerini önceden belirleyebildikleri durumda, bu müşterilerini elde tutma amaçlı stratejiler geliştirebilirler. Kendi müşterisiyken rakibine giden müşterilerle ilgili analizler yaparak rakiplerini tercih eden

⁵⁷ Büyükakın, A., 2005. **Data Mining with Fuzzy Logic**, Master of Science Thesis, İstanbul Technical University Science Institute.

⁵⁸ Dr. Yılmaz Arguden, **Veri Madenciliği Veriden Bilgiye, Masraftan Değere**, 2002, s.68 .

müşterilerinin özelliklerini elde etmek mümkündür. Burdan yola çıkarak gelecek dönemlerde kaybetme olasılığı olan müşterilerin kimler olabileceği konusunda tahminlerde bulunarak onlara özgu ürünler ve servisler geliştirebilirler. Müşteriyi elde tutmak için yapılacak maliyet her zaman için müşteriyi geri kazanmak için yapılacak maliyetten düşük olacaktır. Kaybedilen müşterileri yeniden kazanmak çok önemli hale gelmiştir. Eski müşterileri kazanmak için kurulmuş modellerdir. Müşterilerden omur boyu elde edilecek getiri belirlenerek bu müşterilere sunulacakların maliyeti ile karşılaştırılır.

Mevcut ve potansiyel müşteriler hakkında detaylı bilgiye sahip olmak rekabetsi kalmak için önemli bir gerekliliktir. Farklı müşteri grupları için en uygun ürünleri bulmak, hangi tip müşterilerin ne tür ürün aldıklarını belirlemek, müşteri tabanını gruplara ayırmak, bu grupların karlılıklarını belirlemek ve buna göre farklı seviyelerde hizmet sunmak mümkün olabilir. Ürün veya hizmete hangi özelliklerin ne derecede müşteri memnuniyetini etkilediği, hangi özelliklerinden dolayı müşterin bunları tercih ettiği ortaya çıkarılabilir. Ürün veya hizmet sunumuna kimin yanıt vereceğini tahmin etmek aliyet düşürmek açısından önemli yöntemlerden biridir. Bir ürün veya hizmet ile ilgili bir kampanya programı oluşturmak için hedef kitlenin seçiminden başlayarak bunun hedef kitleye hangi kanallardan sunulacağı kararına kadar olan süreçte veri madenciliği kullanılabilir. Aynı grubun geçmiş davranışlarına dayandırılabilceği gibi, mantıksal bir alternatif popülasyonun davranışlarına da dayandırılabilir.⁵⁹

Düzenlenecek çeşitli kampanyalarda mevcut müşteri kitlesinin seçmek ve bu müşterilerin davranış özelliklerine yönelik yaklaşımlar geliştirmek. Bu şekilde pazarlama veya perakende kampanyalarına cevap alma oranını artırmak, müşteri ilişkileri yönetimi maliyetlerinin azaltmak hedeflenir.

Potansiyel müşteriler arasından en karlı olabilecekleri belirleyerek onlara özel kampanyalar uygulamak. En masraflı müşterileri daha masrafsız müşteriler haline

⁵⁹ Bashein, B.J., and Markus, M.L., 2000. **Data Warehouses**, More Than Just Mining, Financial Executives Resources Foundation, USA.

donuřturmek. Orneęin en cok bankacılık iřlemi yapanlar ortaya cıkarılıp bunlar řube bankacılıęı yerine internet bankacılıęına yonlendirmek.

Aynı karakteristikleri (gelir duzeyi, ilgi alanları, harcama alışkanlıkları, vb.) paylařan “model” muřteri gruplarını bulmak ve satınalma profillerine gore satıř řartlarını ve fiyatları belirlemede önemli bir alıřma olacaktır. Farklı muřteri gruplarının ihtiya duydukları ozellikleri belirtip, ihtiya duymadıkları ozellikleri urunden cıkararak muřterilerin beklentilerini karřılayacak řekilde farklı urunler geliřtirmek gibi bir uygulamada etkin olacaktır.⁶⁰

Veri Madencilięinin Perakende Sektöründe üzerindeki olumlu etkilerini ifade edebilmek için dünya pazarındaki önemli firmalardan örnekler verilebilir:

Halfords marketler zinciri, İngiltere

Halford, kapsamlı büyüme planı alıřmaları sürerken yeni mağazalar açmak için en elveriřli bölge ve yerlerin neresi olacaęı sorusuna en doęru cevabı almak, doęru yatırım planlaması yapabilmek amacı ile veri madencilięi özümünü kullanma kararı almıřtır. Halford, veri madencilięi alıřmaları sayesinde mağazaların başarısını doęru olarak tahmin edebilmiřtir.

İngiltere’de otomobil yedek para mağazalar zinciri olarak faaliyet gösteren Halford için satın alma davranıřlarındaki deęiřimler son derece önemliydi. Faaliyet alanlarına dair iř hacmi tüketici trend’lerine göre deęiřim gösterdięi ve tüketiciler merkez mağazalardan řehir dıřı mağazalara doęru kaymaya bařladıęından potansiyel yeni mağazaları belirlerken doęru gelir tahmini yapabilmeleri gerekiyordu. Yeni açılacak mağazaların başarısı aısından bu önemli bir iř problemi idi.

Halford ilk olarak gemiř satıř deęerlerini ve mağaza profil verilerini içeren veritabanı üzerinde alıřmaya bařladı. Rastgele örneklem yöntemi ile 150 tane mağaza

⁶⁰ Burak Erřahin, **Veri Madencilięi**, 2002, s.95.

seçildi, önceki yıllara ait ardışık 2 yıllık satış verisi, mağaza yeri ile ilgili anket verileri, mağazaya dair spesifik veriler, her bir mağazanın bulunduğu bölgenin demografik özelliklerini gösteren nüfus sayımı verisi birleştirildi. Halford, Veri Madenciliğini veri madenciliği çözümünü kullanarak mağazanın cirosunu tahmin etme amaçlı bir model geliştirdi. Geliştirilen model var olan bir mağaza verisi üzerinde test edildi. Halford geliştirdiği modeli daha önce bir danışmanlık firması tarafından geliştirilen regresyon modelinin sonuçları ile de karşılaştırdı. Halford Yönetimi geliştirilen modeli mağaza performanslarındaki değişim nedenlerini anlamada ve karlılığı yüksek olacak yeni mağaza yerleri belirlemede kullandı ve başarısını garantiledi.

Sofmap company, Japonya

Japonya’da en büyük kişisel bilgisayar ve yazılım perakendecisi olan Sofmap’ın ülke çapında 40 mağazası bulunmaktadır. Sofmap yöneticileri müşterilerinin bir çoğunun donanım ve yazılım alım kararlarında zorlandığını, bunun online satışları engellediğini düşünmektedir. Sofmap, Veri Madenciliğini online kayıt esnasında sisteme girilen verileri ve veritabanındaki işlem verileri üzerinden müşteri profiline uygun ürünleri öneren bir alt yapı oluşturmada kullandı. Sonuç olarak sayfa ziyaret sayısı %67 arttı, satışlar bir önceki yılın aynı dönemine göre %18 arttı.⁶¹

Herlitz ag, Almanya

Almanya’nın en büyük ofis ürünleri üreticisi olan Herlitz AG, Veri Madenciliğini ile ürünlerini satan partnerlerine kendi ürünlerini alan müşterilerinin satın alma alışkanlıklarını analiz etme ve spesifik bir ürün için en iyi satış noktalarını belirleme ve satışları arttırma amaçlı çalışmalar gerçekleştirdi.

Bütün ürünler ve satış noktalarına dair veri ayrı ayrı analiz edildi. Veri setleri ortalama 4,000 kayıt ve 30 değişken içeriyordu. Veri Madenciliği ile satışlar ortalamasının

⁶¹ Berry, M. J. A. and Linoff, G.S., 2000. **Mastering Data Mining: The Art and Science of Customer Relationship Management**, John Wiley & Sons, New York.

üstü ve ortalamanın altı olarak gruplandı. Ve karar ağacı algoritmaları kullanılarak bu grupların özellikleri belirlendi ve bir tahmin modeli oluşturuldu. Modeller ilk uygulamaya alındığında satışların ne zaman ortalamanın altında ne zaman ortalamanın üstünde olacağını %68 olasılıkla veya 6,441 üründen 4,420'si için doğru olarak tahmin edildi. Herlitz AG, ürünlerini ilgili bölgenin özelliklerini içeren alt bölgelere göre gruplandırdı.

Whitbread, İngiltere

İngiltere'nin en büyük içecek üreticilerinden biri olan Withbread yaklaşık 4.000 otel, restaurant ve pub'a ürün vermektedir. Withbread'in bu pozisyonunu koruması için hızlı değişen market koşullarına çabuk cevap vermesi ve büyüme potansiyeli olan sahaları yakalaması gerekiyordu. Bunun da ötesinde;

- Her markanın ayrı ayrı piyasa performansını ve değerini değerlendirmek,
- Yerel pazar özelliklerini iyi anlayarak, tutunmak için gerekli çıkış noktalarını yakalayabilmek,
- Firma yatırım hedeflerini belirlemek, diğer önemli hedeflerdi.

Veri Madenciliği ile daha doğru tahmin ve karar verme mekanizmaları oluşturmak üzere kolları sıvadılar ve verilerine dayalı analitik modeller geliştirdiler. Pazar araştırmaları, demografik veriler, jeolojik özellikler, pazarlama ve satıştan gelen tüm verileri detaylı olarak analiz ettiler.⁶²

Sonuç olarak:

- %85 doğrulukla ciro tahmini yapabildiler,
- 400 milyon dolar bütçesi olan pazarlama ve reklam aktivitelerini, markaların piyasadaki satış performansı ve piyasa payına göre yönlendirdiler.

⁶² Laudon, K.C and J.P. Laudon, **Management Information Systems**, Prentice-Hall Inc, New Jersey, 2000.

Marketi yerel ve ulusal boyutlarda analiz ederek, büyük müşterilerinden biri olan bir restoran zincirine özel bir hizmet programı geliştirdiler.

Direct wines, Kanada

Çok sayıda şarap kulübü olan ve dünyanın en büyük sipariş ile şarap satışı yapan Direct Wines Veri madenciliği ile veritabanı verilerini analiz etme ve müşterilerine özel öneriler geliştirme imkânı buldu. Sonuç olarak müşteri segmentlerine özel duyurular göndererek, duyurularına geri dönüş oranlarını arttırdı; müşteri odaklı çapraz satış, pazarlama ve karlılık planları ile müşteri değerini ve müşteri hizmet kalitesini arttırdı. Direct Wines'ın yıllık cirosu £132 milyon idi ve aylık posta gönderim adedi 500,000 idi.

3.1 Veri Madenciliği

Veri madenciliği, bilgi çağında en güncel teknolojilerden biridir. Günümüzde artan veri sayısı, bilgisayar ve iletişim teknolojilerindeki gelişmeler çok daha fazla verinin hızlı bir şekilde depolanmasına, işlenmesine ve bilgiye dönüştürülmesine imkan vermektedir. Bu yüzden, büyük miktardaki verilerden yararlanılmasını ve bu verilerin anlamlı hale getirilmesini sağlayan teknikler büyük önem taşımaktadır. Bu gibi durumlarda çözüm sağlayan veri madenciliği son yıllarda yoğun olarak kullanılmaya başlanmıştır.

1960'larda veri toplama sistemleri popülerken, 1970'ler de ilişkisel veri tabanları, 1980'ler de ise ilişkisel veritabanı modelleri kullanılmaktaydı. 1990'lar da adından söz edilmeye başlanan veri madenciliğinin 2000'li yıllarda dünya da oldukça önemli bir yer aldığı görülmektedir.⁶³

Veri madenciliği yakın zamanda oldukça dikkat uyandırmıştır. Ticari ve bilimsel keşifler için büyük potansiyeli ile, yeni sorunlarla uğraşan, yeni bir

⁶³ Bölükbaşı, İ., . **Defining Critical Factors Affecting Student Success Data Mining Approach**, Master of Science Thesis, İstanbul Technical University Institute of Science and Technology, 2005.

teknolojidir. Bu gelişimine rağmen veri madenciliğinin yine de bütün sorulara cevap vermesi beklenemez. Veri madenciliği yararlı araçlar sunarken, bu durum, onun kaçınılmaz olarak önemli, ilginç ve değerli sonuçlara yol açacağı anlamına gelmemelidir.⁶⁴

Veri madenciliği ile ilgili olarak birçok tanım yapılmıştır. Konunun önde gelen uzmanlarından Piatetsky-Shapiro veri madenciliğini, verilerden daha önceden bilinmeyen, muhtemelen yararlı enformasyonun (yapılandırılmış verinin yani bilginin) monoton olmayan bir süreçte çıkartılması işlemi olarak tanımlamaktadır⁶⁵.

J. Han ve M. Kamber'e göre veri madenciliği; veritabanlarında, veri ambarlarında veya diğer veri depolarındaki büyük miktarlardaki veri içindeki ilginç bilgileri keşfetme sürecidir. I.H.Written ve E.Frank'a göre ise, veri içinden önceden bilinmeyen ve yararlı bilgiyi çıkarmaktır. D.Hand, H.Mannila ve P.Smyth ise, veri madenciliğini daha önceden akla gelmeyen ilişkileri bulmak ve verileri araştırmacı için hem yararlı hem de anlaşılır olacak şekilde veriyi tıpkı bir roman gibi özetleyerek beklenmeyen ilişkileri ortaya çıkaran analiz olarak tanımlamışlardır⁶⁶.

Veri madenciliği, istatistik ve başka alanlarda farklı isimler ile birlikte birçok bilim dalında kullanılmış ve bilgisayar teknolojisindeki hızlı gelişmelerden sonra, veri arkeolojisi, veri taranması, veri balıkçılığı, özbilgi çıkarımı veya veri tabanlarında özbilgi keşfi gibi farklı isimler ile de anılmıştır. Belki de bu isimler içinde en ilginç olanı veri balıkçılığıdır. Bu isimle anılmasının nedeni ise, veri madenciliği verilerin belirli bir örüntü bulunması beklentisiyle taranması deniz dibinin taranarak kum ve çamurdan arındırıldıktan sonra yararlı deniz ürünlerinin elde edilmesine benzetilmesidir.

Bu isimlerden Veritabanlarında Özbilgi Keşfi üzerinde durmakta yarar vardır. Veritabanlarında özbilgi keşfi ve veri madenciliği arasındaki ilişki dikkate alındığında iki yaklaşım sözkonusu olmaktadır. Bunlardan birincisi veritabanlarında

⁶⁴ Hand,D., Manila, H. , Smyth,P., **Principles of Data Mining**, Cambridge, CA: MIT Press, 2001, s.24.

⁶⁵ Akpınar,H., '**Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği**', İ.Ü. İşletme Fakültesi Dergisi, Cilt.29, Sayı.1, Nisan 2000,s.1.

⁶⁶ Hand ve diğerleri, s.1.

özbilgi keşfi ile veri madenciliği bir bütün olarak kabul edilmekte iken, ikinci yaklaşım ise veri madenciliğini veritabanlarında özbilgi keşfinden bir önceki basamak olarak görmektedir⁶⁷. İkinci yaklaşıma göre özbilgi keşif süreci aşağıdaki adımları izlemektedir⁶⁸:

Tablo 3. 1 Özbilgi Keşif Süreci

Verilerin Temizlenmesi:	Gürültülü, karmaşık, tutarsız verilerin temizlenmesi.
Verilerin Birleştirilmesi:	Çok sayıdaki veri kaynağının birleştirilmesi.
Verilerin seçilmesi:	Analize uygun verilerin veritabanından seçilmesi.
Verilerin Dönüşümü:	Verilerin özetleme ve birleştirme süreçleri için uygun formlara dönüştürülmesi.
Veri Madenciliği:	Veri örüntülerinin ortaya çıkarılması için uygun yöntemlerin uygulanması.
Örüntülerin Değerlendirilmesi	: Özbilgiyi temsil eden ilginç örüntülerin bulunması.
Özbilginin Sunumu:	Ortaya çıkarılan özbilginin görselleştirilmesi ve özbilgi sunum tekniklerinin kullanımı.

Yukarıdaki süreç incelendiğinde, veri madenciliğinin veritabanlarında özbilgi keşfinin alt basamağı olarak bile temel ve önemli bir adımı oluşturduğu görülmektedir.

Büyük veritabanlarındaki veriden bilginin veya örüntünün çıkarımı olarak da tanımlanan veri madenciliği, verideki örüntüleri keşfetmek için algoritmaların bilgisayar ortamında uygulanması diye de tanımlanmaktadır⁶⁹.

⁶⁷ Zhi-Hua Zhou. 'Three Perspective of Data Mining', **Artificial Intelligence**, Vol.143, 2003, s.142.

⁶⁸ Oğuzlar, A., **Veri Madenciliğine Giriş**, Bursa, Ekin Kitabevi, 2004, s.4-5.

⁶⁹ Oan Chang Sung, 'Affordable Distributed Data Mining', Ph.D. Department of Computer Science, August 2004, s.2.

David J. Hand veri madenciliğini, 'Data Mining: Statistics and More?' adlı makalesinde, istatistik, veritabanı teknolojisi, örüntü tanıma, makine öğrenimi ve diğer alanların arayüzünde (interface) bulunan yeni bir disiplin olarak tanımlamıştır⁷⁰.

Yukarıda verilen veri madenciliği tanımlarını daha da arttırmak mümkündür. Ancak bütün tanımlamalar incelendiğinde, veri madenciliğinin genel bir tanımı aşağıdaki gibi yapılabilir:

Veri madenciliği; büyük sayıda verinin (gigabyte ve hatta terabytelarca veri) incelendiği veritabanlarındaki ilişkileri, eğilimleri ve ilginç örüntüleri keşfetmede ve veri ambarlarında oluşan açıklanmamış bilginin çıkarımında kullanılan bir süreçtir.

Veri madenciliği uygulamalarında aşağıdaki soruların cevaplanması gerekmektedir⁷¹.

- Analiz edilecek veri seti veri madenciliği için uygun mu? Veri setinin işlem yapılabilecek formatta olması gerekmektedir. Genellikle farklı kaynaklarda ve farklı formatlarda bulunan veri dosyaları, bir araya getirilerek analiz edilebilmektedir.
- Analiz edilecek olan veri seti istenen bütün faktörleri kapsıyor mu? Veri setinin istenen bütün faktörleri kapsıyor olması tahminleri doğruluk derecesinin oranını arttırmaktadır.
- Analiz edilecek veri setinin büyüklüğü yeterli mi? Çoğu zaman problem, veri setinin küçüklüğünden değil çok büyük olan veri setinin temsil gücünün zayıf olmasından kaynaklanmaktadır. Eğer veri setinin temsil gücü yüksek ise, birkaç bin gözlem veri madenciliği için yeterli olmaktadır.

⁷⁰.David J. Hand, 'Data Mining: Statistics and More?', The American Statistician, Vol.52, No.2, 1998, s.112.

⁷¹ SPSS, Kazanmanın Yeni Bilimi: Analitik Zeka, 2007, <http://www.spss.com.tr/> (10 Mart 2011), s.10.

Günümüzde veri madenciliğinin başlıca uygulama alanları aşağıdaki gibi sayılabilir⁷²;

Pazarlama alanında

- Müşteri segmentasyonu
- Müşterilerin demografik özellikleri arasındaki bağlantıların kurulması
- Mevcut müşterilerin elde tutulması
- Yeni müşterilerin kazanılması
- Pazar sepeti analizi
- Çapraz satış analizleri ve satış tahminleri
- Müşteri değerlendirme ve müşteri ilişkileri yönetimi
- Çeşitli pazarlama kampanyaları

Bankacılık alanında

- Farklı finansal göstergeler arasındaki gizli ilişkilerin bulunması
- Kredi kartı dolandırıcılıklarının tespiti
- Kredi taleplerinin değerlendirilmesi
- Usulsüzlük tespiti
- Risk analizleri Sigortacılık alanında
- Yeni poliçe talep edecek müşterilerin tahmin edilmesi
- Sigorta dolandırıcılıklarının tespiti
- Riskli müşteri tipinin belirlenmesi Perakendecilik alanında
- Satış noktası veri analizleri
- Alış-veriş sepeti analizleri
- Tedarik ve mağaza yerleşim optimizasyonu Borsa alanında
- Hisse senedi fiyat tahmini
- Genel piyasa analizleri
- Alım-satım stratejilerinin optimizasyonu Telekomünikasyon alanında

72 Eker,H., 'Veri Madenciliği veya Bilgi Keşfi', <http://www.bilgiyonetimi.org/cm/,07> Ekim 2011.

- Kalite ve iyileştirme analizleri
- Abonelik tespitleri
- Hatların yoğunluk tahminleri

Sağlık ve İlaç alanında

- Test sonuçlarının tahmini
- Ürün geliştirme
- Tıbbi teşhis
- Tedavi sürecinin belirlenmesi

Endüstri alanında

- Kalite kontrol analizleri
- Lojistik
- Üretim süreçlerinin optimizasyonu

Veri madenciliğinin uygulama alanları yukarıda da görüldüğü gibi oldukça geniştir. Bu kadar çok uygulama alanının yaygınlaşmasının nedenlerini şu şekilde sıralamak mümkündür⁷³.

- Veri artışı
- Veri ambarları
- Bilgisayar donanımlarındaki gelişmeler
- Başta pazarlama olmak üzere birçok alanında rekabetin hızlanması

3.1.1 Veritabanları

Veritabanı yönetim sistemleri sayesinde büyük ve karışık verilere ulaşmak oldukça kolaydır. Çünkü verilerin depolandığı dosyalar, veri miktarı büyüdükçe istenilen veriye ulaşmayı zorlaştırmaktadır. Veritabanı teknolojileri ise bu dosyaları

⁷³ Fayyad,U., Piatetsky-Shapiro G. , Smyth,P., **Knowledge Discovery and Data Mining: Towards a Unifying Framework**, Proceedings of The Second Conference on Knowledge Discovery and Data Mining, Portland, Oregon, Ağustos, 1996, s.84.

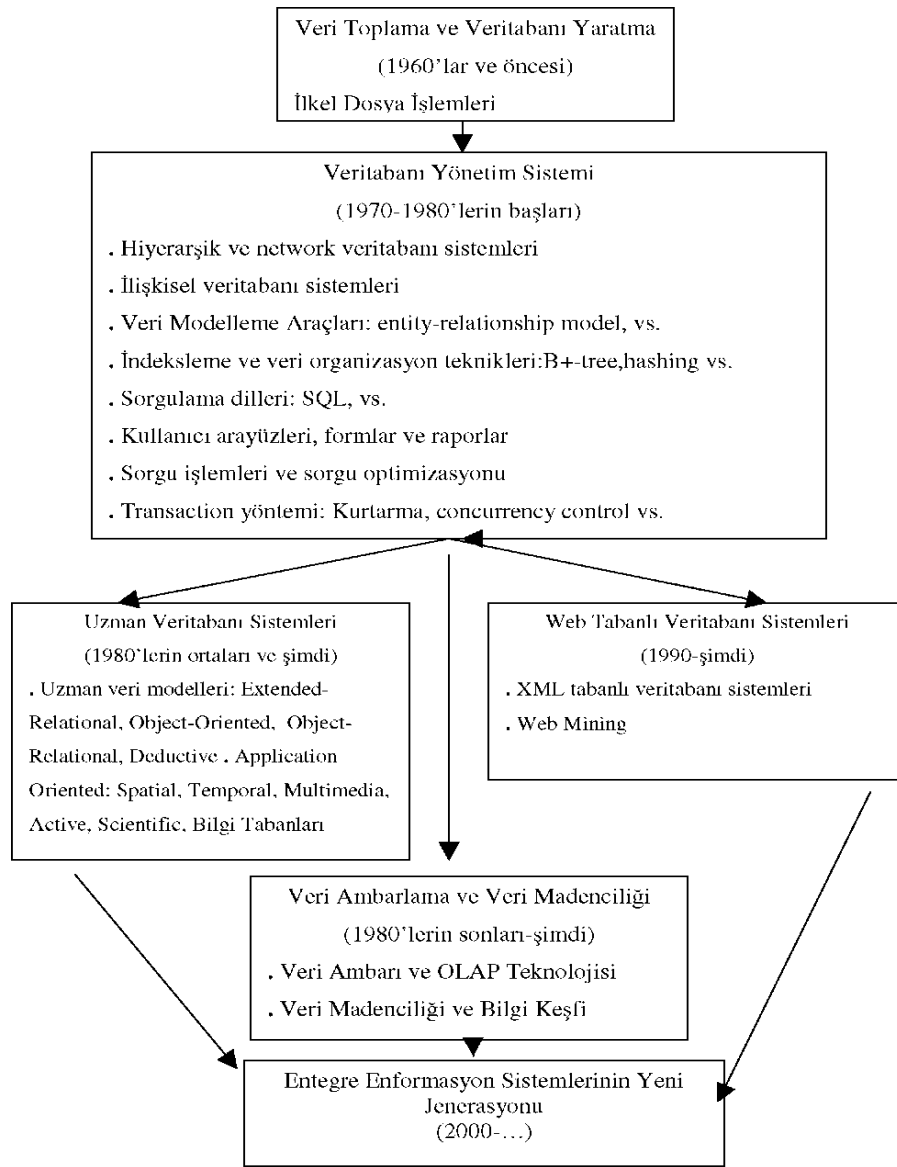
düzenleyerek verilere daha hızlı ve düzenli bir şekilde ulaşmayı sağlamaktadır. Ayrıca birden fazla kullanıcının aynı bilgileri birbirini etkilemeden kullanmasına da imkan vermektedir.

Veri madenciliği ile ilgili olarak yapılan tanımlardaki ortak nokta her bir tanımın 'büyük ölçekli veriler' den söz etmesidir. Bu büyük ölçekli verinin temeli ise veritabanlarına dayanmaktadır. Veri madenciliği disiplininin meydana gelmesinde veritabanı teknolojisindeki gelişmelerin önemi oldukça büyüktür. Veritabanı birbirleriyle ilişkili verilerin tekrara yer vermeden, çok amaçlı kullanımına olanak sağlayacak şekilde depolanması olarak tanımlanmaktadır. Kısaca, veritabanı bir veri kümesi olup, kullanıcıların ihtiyaçlarına göre sınıflandırılmalı, raporlanmalı ve analiz edilmesi gerekmektedir. Bu süreç ise veri analizi ile sağlanmaktadır⁷⁴.

Yukarıda da bahsedilen veritabanı gelişimi ve bu gelişme içerisinde veri madenciliğinin rolü aşağıdaki şekil'de gösterilmiştir⁷⁵.

⁷⁴ Babadağ,K,K., '**Veri Madenciliği Yaklaşımı ve Veri Kalitesinin Artması için Kullanılması**', TÜİK 15. İstatistik Araştırma Sempozyumu Bildiriler Kitabı, Yayın No.3062, Ankara, 2006, s.242.

⁷⁵ Han,J., Kamber,M. **Data Mining: Concept and Techniques**, USA: Morgan Kaufmann Publishers, 2001, s.39-40.



Şekil 3. 1 Veritabanı Teknolojisinin Gelişimi ve Veri Madenciliği

Kaynak: Suat Özdemir, <http://ceng.gazi.edu.tr/~ozdemir> , 2010

Şekil 3.1 incelendiğinde de veri madenciliğinin veritabanı gelişim sürecinin bir ürünü olduğu görülmektedir.

Son yıllarda veri ambarı, veri madenciliği ve OLAP (On-Line Analytical Processing) karar destek sistemleri teknolojilerindeki gelişmeler ile yapılan araştırmalar oldukça fazla önem kazanmıştır. Önceki yıllarda (1980'lerin sonlarında) veri erişimi ve veri değiştirme gibi işlemleri yapan OLTP (On-Line Transaction Processing) sistemlerini veri ambarları ve OLAP teknolojisinin gelişimi takip etmektedir. Bu nedenle aşağıda veri ambarı, OLTP ve OLAP teknolojilerinden kısaca bahsedilecektir.

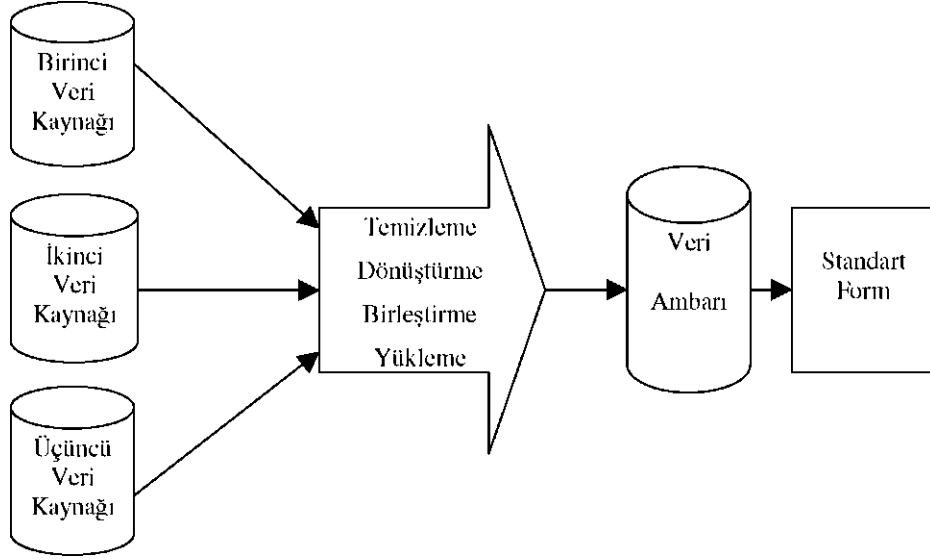
3.1.1.1 Veri Ambarları

1991 yılında ilk kez William H. Inmon tarafından ortaya atılan veri ambarı, yönetimin kararlarını desteklemek amacı ile çeşitli kaynaklardan elde ettikleri bilgileri zaman değişkeni kullanarak veri toplama olarak tanımlanmaktadır. Kısaca birçok veritabanından alınarak birleştirilen verilerin toplandığı depolardır. Veri ambarlarının özelliği kullanıcılara farklı detay düzeyleri sağlayabilmesidir. Detayın en alt düzeyi arşivlenen kayıtların kendisi ile ilgili iken, daha üst düzeyler zaman gibi daha fazla bilginin toplanması ile ilgilidir. Veri ambarları ciddi yatırımlar gerektirmekte ve uygulanması bir yıl veya daha uzun zaman almaktadır⁷⁶.

Veri ambarlarında bulunan verilerin en önemli özelliği, temiz ve dönüşümden geçmiş olmaları olduğu için veri ambarına kaydedilecek verinin öncelikle temizlenmesi, gerekli dönüşümlerin ve verilerin birleştirilmesi işlemlerinin yapılması gerekmektedir. Böylece veri ambarında saklanan veriler kullanılmaya hazır olduğundan kullanıcılar amaçlarına uygun veriye, veri ambarlarına gerekli sorgulamaları yaparak ulaşmaktadırlar. İşte bu aşamadan sonra devreye veri

⁷⁶.Murray J. Mackinnon ve Ned Glick, 'Data Mining and Knowledge Discovery in Databases- An Overview', J.Statistics., Vol.41, No.3, 1999, s.260.

madenciliği girmektedir. Bir veri ambarının kullanım süreci şekil 3.2' de gösterilmektedir⁷⁷.



Şekil 3. 2 Veri Ambarı Kullanım Süreci

Veri ambarları, verilerin üzerine yazmaya ve verilerde değişiklik yapmak için değil sadece okumaya yönelik olarak oluşturulmaktadır. Bu nedenle veri ambarında veriler, analiz yapmayı kolaylaştıran bir formatta tutulmaktadır. Burada analiz; sorgular, raporlar, karar destek sistemleri veya istatistiki hesapları kapsamaktadır. Veri ambarında tutulan bilgi şu özelliklere sahiptir:

- Özetlenmiş operasyonel veri içerir.
- Ayrı bir sistemde tutulur.
- Sadece ayıklanmış ve güvenilir veri kullanılır.
- Veriler çok boyutlu bir halde tutulur.

Veri ambarlarında 2 boyutlu operasyonel sistemlerin tersine çok boyutlu bilgi saklanabilir. Boyuttan kasıt zaman, ürün, dağıtım kanalı, marka, model, fabrika vs. olabilir. Boyutlar veri ambarının hazırlandığı işletmenin yapısına göre tayin edilir. Çok boyutlu veritabanı mantığı bir verinin her boyuta göre incelenebilmesidir.

⁷⁷ Thuraisingham,B., 'A Primer for Understanding and Applying Data Mining', IT Professional Technology Solutions for the Enterprise, Vol.2, No:6, 2000, s.55.

3.1.1.2 OLTP Sistemleri

OLTP ilk olarak ortaya çıkan karar destek sistemleri olarak bilinmektedir. Bireysel işlemleri hızlı ve etkin bir şekilde kaydetmek üzere kullanılan operasyonel veritabanları OLTP olarak bilinmektedir. Bir rezervasyon sistemi, bir muhasebe uygulaması veya sipariş giriş uygulaması OLTP sistemlerine birer örnek olarak verilebilir. Operasyonel veriler; adından da anlaşılacağı üzere çok sık daha doğrusu yapılan her işlemle güncelleşmektedir. Yani yapılan her bir rezervasyon ile veya iptal edilen her sipariş ile OLTP sistemi veritabanına yeni bilgiler eklemekte veya kayıt silmektedir. OLTP sistemlerinin bazı dezavantajları vardır. Bunlar aşağıda kısaca özetlenmiştir⁷⁸.

- Operasyonel sistemlerde her kayıt analiz açısından pek çok gereksiz bilgi içermektedir. Üstelik kullanıcı hataları gibi pek çok faktör yüzünden bu veriler zaman zaman yanlışır. İçinde tutulan veritabanı raporlama amacıyla değil günlük işlemlerin hızlı bir şekilde yapılabilmesine yönelik yapılandırılmıştır.
- Kompleks sorgular yaratmak ve işleme koymak oldukça güçtür. Gerekli bilgilere operasyonel sistemden ulaşmak için karmaşık sorgular, her bir sorgu için ayrı ayrı raporlar gerektirmektedir.
- Ayrıca bu sistemde sorguları çalıştırmak gereğinden çok fazla zaman almaktadır.

3.1.1.3 OLAP Sistemleri

E.F. Codd OLTP' nin, ilişkisel veritabanlarındaki verilerin kullanıcıya oldukça fazla olanak sağladığını savunurken ilişkisel veritabanlarına basit SQL sorguları sorulduğunda gerekli olan büyük hesaplamalar açısından yetersiz olduğunu ileri sürmüş ve bu nedenle çok boyutlu veritabanlarının kullanılmasını savunarak 1993 yılında Computerworld'deki makalesinde OLAP sistemlerini ortaya atmıştır.

⁷⁸ <http://www.kora.com.tr/turkish/veriambari1.htm> (02 Kasım 2010).

OLAP sistemlerinin temel amacı kullanıcıların gerçekleşen değişiklikleri anlamaları için çok boyutlu olarak kurumsal verileri işleyebilmesidir.

Veri ambarı için çok boyutlu analiz aracı olarak, veri ambarında ve büyük veritabanlarında on-line analitik işleme sistemlerinin kabul edilmesi gerekmektedir. OLAP, kazma, döndürme, kesme ve dilimleme gibi olanakları sağlamaktadır. Bu nedenle kullanıcı, verileri rahat bir şekilde çevirebilmekte, ilgili veri setini belirleyebilmekte, verileri farklı büyüklüklerde analiz edebilmekte ve sonuçları farklı formatlarda görüntüleyebilmektedir. Bu işlemler aynı zamanda onu açıklayıcı ve etkili bir süreç haline getirebilmek için veri çıkarımına da uygulanabilmektedir. OLAP ile birlikte veri madenciliği fonksiyonları kullanıcının daha sonra herhangi bir ilginç örüntüyü araştırabileceği şekilde bulunan bilginin genel bir görünüşünü sunabilmektedir. Çünkü OLAP işlemleri ile veri setinin büyüklüğü görece olarak daha kompakt hal alır. Böylece, OLAP sistemleri ile birlikte kullanılan veri madenciliği doğrudan ham veriden çıkarımdan daha hızlı bir şekilde sonuca ulaşılmasını sağlamaktadır.

OLAP bilgi toplamak için önemli bir araçtır. OLAP bazı ilişkilerin neden mevcut olduğunu açıklamak için kullanılmaktadır. Kullanıcı, değişkenler arasındaki ilişkiler hakkında kendi hipotezini oluşturur ve verileri izleyerek hipotezinin doğruluğunu kanıtlamaya çalışır. Örneğin, borçların neden geri ödenmediğini bulmaya çalıştığımız bir araştırma düşünelim. İlk olarak aklımıza gelen düşük gelir ve büyük borçlara sahip kişilerin yüksek risk grubunu oluşturduğudur. OLAP, değişkenler arasındaki ilişkinin çok boyutlu hiperküp olarak adlandırılan grafiğini vererek, kurduğumuz bu hipotezi kontrol etmemizi sağlamaktadır⁷⁹.

OLAP teknolojisinde veri madenciliğinin aksine araştırma hipotezi kullanıcı tarafından ileri sürülmekte ve verilerden elde edilmemektedir. Ayrıca, tüm işlemler bilgisayar tarafından yapılmakta istatistiksel modellemeler kullanılmamaktadır.

⁷⁹ Giudici, P., **Applied Data Mining Statistical Methods for Business and Industry**, New York, Chichester: Wiley, 2003, s.4.

OLAP küçük sayıda değişkene sahip veritabanları için yararlı bilgiler sağlamakta değişken sayısı arttıkça sorunlar ortaya çıkmaktadır. Böylece kurulan hipotezi test etmek amacı ile veritabanını OLAP teknolojisi ile analiz etmek oldukça zorlaşmakta ve çok zaman almaktadır.

OLAP teknolojisi ile veri madenciliği birbirlerinin tamamlayıcısıdır. OLAP veri madenciliğinin ön işleme aşamalarında verinin açıklanması, değişkenlere yoğunlaştırılması gibi ön bilgiler elde etmek için kullanılabilir. Böylece en önemli veriler üzerinde durulmasına olanak sağlar, özel durumları tanımlar. Aynı zamanda veri madenciliği sonucunda elde edilen ilişkilerin etkinliğinin incelenmesinde de OLAP teknolojisinden yararlanılabilir⁸⁰.

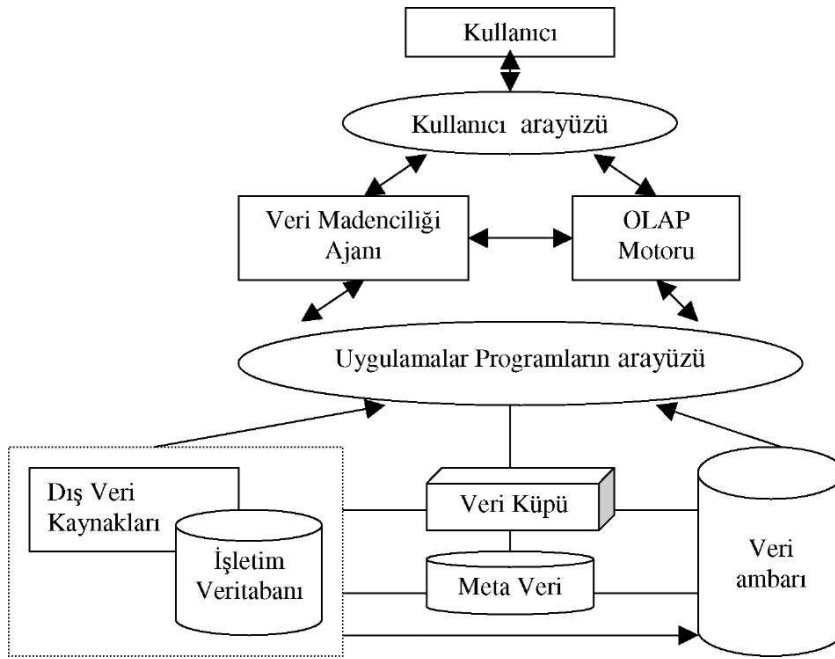
Etkili bir veri madenciliğini olumsuz olarak etkileyen en büyük neden bilgi eksikliğidir. Genellikle veritabanları veri madenciliği ile ilgisi olmayan nedenlerden dolayı oluşturulur. Bu nedenle birçok bilgi eksik olabilir. Bir diğer neden ise yanlış verilerin veritabanında yer almasıdır. Bu sorunları ortadan kaldırmak için ise veri ambarları kullanılmaktadır. Veri ambarlarındaki veriler etkili ve ölçeklenebilir veri madenciliği ile birleştğinde veriler doğru ve etkili bir şekilde kullanılmaktadır⁸¹.

OLAP teknolojisi veriler çok boyutlu işlediğinden dolayı veri ambarları ve OLAP çok boyutlu veri modeli üzerine kurulmakta ve model veriyi veri küpü biçiminde görüntülemektedir. Veri madenciliği ve OLAP teknolojisi ilişkisi şekil 3.3’de görülmektedir⁸². Şekil incelendiğinde 7 bileşenden oluştuğu görülmektedir.

⁸⁰ Babadağ, s.243.

⁸¹ Giudici, s.5.

⁸² Zhen,L. , Minyi,G. , 'A Proposal of Integrating Data Mining and On-Line Analytical Processing in Data Warehouse', 2001 International Conferences on Info-tech and Info-net Proceedings, Beijing, China, 2001, s.148-149.



Şekil 3. 3 Veri Madenciliği ve OLAP

Kaynak: Alejandro a. Vaisman, Olap, data warehousing and materialized views: a survey, University of Buenos Aires

Veri madenciliği ve OLAP ilişkisindeki bileşenler şöyle açıklanabilir:

Veri Ambarı:	On-Line Analitik veri madenciliğinin platformu
Veri Madenciliği:	OLAP sisteminin yardımcı olduğu veri küplerinde analitik veri madenciliğinin yapılması
OLAPSistemi (engine-motoru):	Çok boyutlu olarak özetlenen verilere hızlı erişimin sağlanması
Uygulama Programının Arayüzü:	Veri ambarında on-line veri madenciliği fonksiyonları, düzenlemeleri ve kurallarının toplanması
Veri Küpü:	Veri ambarları bilgilerinin toplanması
Meta Veri:	Veri ambarının oluşturulması ve bakımının kontrol edilmesine yönelik veriler
Kullanıcı Arayüzü:	Kullanıcının on-line sorularının kabul edilmesi ⁸³

⁸³ Berson, A. and Smith, S. J.. **Data Warehousing, Data Mining and OLAP**, McGraw-Hill, New York,2007.

Veri küpü on-line veri madenciliğinin odak noktasıdır. Veritabanlarının ve veri ambarlarının içeriklerinin analiz edilmesi için kullanılabilen toplu bilgiyi sağlamaktadırlar. Veritabanları ve veri ambarlarındaki simgelerin alt kümelerinden oluşmaktadır. Veri madenciliği, OLAP motorunun yardımı ile veri küplerinde analitik çıkarım yapmaktadır. Veri madenciliği ve OLAP motorunun her ikisi kullanıcı arayüzü aracılığı ile kullanıcının on-line sorularını kabul etmekte ve analizdeki uygulamalar programlama arayüzü aracılığı ile veri küpü ile çalışmaktadır. Ayrıca, veri madenciliği ajanı (agent) tanımlama, birliktelik, sınıflandırma, öngörme, kümeleme, zaman serileri analizi, vb. gibi çoklu veri madenciliği görevlerini yerine getirebilmektedir. Bu nedenle, etkin bir veri madenciliği birbiri ile etkileşim içinde olabilen birçok veri madenciliği modülünden oluştuğundan veri madenciliği ajanı OLAP motorundan daha karmaşıktır.

Veri madenciliği ajanında önemli olan sayısal boyutlar OLAP sistemlerinde kolayca bulunamadığından, modellemenin başında özel veri madenciliği modüllerinin kurulması gerekmektedir. Piyasada bulunan birçok OLAP ürününde, kurulan veri küpleri ve OLAP motorlarının direk olarak üstünde on-line analitik veri madenciliği mekanizmalarının geliştirilmesi önemlidir. Veri madenciliği ajanı OLAP analizine kıyasla daha güçlü veri küpü yapısı ve erişim araçları gerektirmesine rağmen, OLAP motoru ve veri madenciliği ajanı için gerekli veri küpü arasında temel bir fark yoktur. Veri madenciliği ajanı sıklıkla bağlantısal veritabanı sistemleri ile çalışan özel veri küpleri üzerinde veya OLAP ürünleri tarafından sağlanan veri küpleri üzerinde kurulduğundan, on-line analitik veri madenciliği sistemlerinin grup üzerine kurulması yerine mevcut OLAP ve ilişkisel veritabanı sistemleri üstüne kurulması tavsiye edilmektedir.⁸⁴

Veri ambarı nesne odaklı özet veriler sağladığından (Şekil 3.2), veri ambarı, verileri veri madenciliğinin etkinliğini geliştirmede avantajlıdır. Veri madenciliği veri ambarı üzerinden yapıldığında, ilk iki adım olan seçme ve ön işleme kabaca

⁸⁴ Chan, C. and Lewis, B., **A Basic Premier on Data Mining, Information Systems Implementations**, Morgan Kaufmann, San Diego. Management, 2002, s. 56-60

tamamlanmış olmaktadır. KDD süreci özel veri madenciliği sorunları için kullanılabilir ebatları belirlemek ve veri madenciliğini gerçekleştirmek amacı ile özel algoritmaları kullanmak için veri dönüşümünden başlamaktadır. Veri dönüşümü özel veri madenciliği görevi için yararlıdır veya özel veri madenciliği görevini büyük ölçüde etkilemektedir.

Veri madenciliği fonksiyonları genellikle basit OLAP işlemlerinden daha fazlasını yaptığından, on-line analitik veri madenciliğinde etkili uygulama ve hızlı cevap bulma temel amaçtır.⁸⁵ On-line analitik veri madenciliğinin geliştirilmesi ve standart uygulama programlama arayüzünü sağlayan veri madenciliği modülünün test edilmesi ve paylaşılması çok önemlidir. Veri madenciliği sorularına hızlı cevaplar elde etmek için, etkin ve kısıtlamaya dayanan veri madenciliği algoritmalarının uygulanması gerekmektedir⁸⁶. Kısaca OLTP sistemleri satın alma, envanter, üretim, bankacılık, bordro, kayıt ve muhasebe gibi bir kurumun günlük faaliyetlerinin çoğunu kapsamaktadır. Farklı kullanıcıların çok çeşitli ihtiyaçlarına uymak için verileri çeşitli formatlarda düzenleyen ve sunan OLAP sistemleri ise analiz ve karar alma rolüne sahip kullanıcılara hizmet vermektedir. Bu sistemler arasındaki en belirgin farklar şöyle sıralanabilir:

⁸⁵ James M. Latin, Douglas J. Carroll ve Paul E. Green, **Analyzing Multivariate Data**, United States: Thomson Brooks-Cole, 2003, s.283.

⁸⁶ Zhen ve Minyi s.148-149.

Kullanıcı ve sistem yönlendirme	OLTP sistemi müşteriye yöneliktir ve alıcı, satıcı ve bilgi teknolojisi uzmanları tarafından sorgu işlemleri için kullanılmaktadır. OLAP sistemi ise piyasaya yöneliktir ve yöneticiler, idareciler ve analistler tarafından veri analizi için kullanılmaktadır.
Veri içerikleri	OLTP sistemi, karar alma amacı ile çok ayrıntılı olan cari verileri yönetmektedir. OLAP sistemi ise büyük miktarlardaki geçmiş dönemlere ait verileri toplamakta, özetlemekte ve saklamaktadır. Bu özellikler verilerin karar alma sürecinde kullanılmasını kolaylaştırmaktadır.
Veritabanı tasarımı	OLTP sistemi genellikle kurum-ilişki (entity- relationship) veri modelini ve bir uygulamaya ilişkin veritabanı tasarımını benimsemektedir. OLAP sistemi ise bir yıldız veya kar tanesi modelini ve nesneye yönelik veritabanı tasarımını benimsemektedir.
Bakış açısı	OLTP sistemi, geçmiş dönemlere ait verilere veya farklı kurumlardaki verilere değinmeksizin bir kurum içindeki güncel veriler üzerinde odaklanmaktadır. OLAP sistemi ise veritabanına yayılmakta, farklı kurumlardan kaynaklanan bilgileri de içine almakta ve birçok veri ambarından bilgileri bir araya getirmektedir.
Erişim örüntüleri	OLTP sisteminin erişim örüntüleri kısa işlemlerden oluşmaktadır. Birçok veri ambarı güncel bilgiden çok geçmiş dönemlere ait bilgileri depoladığından OLAP sisteminde erişim örüntüleri genellikle salt okunur işlemlerdir.

Bu iki sistem arasındaki diğer farklar ise Tablo 3.2 'de özetlenmiştir⁸⁷.

Tablo 3. 2 OLTP ve OLAP Sistemlerinin Karşılaştırılması

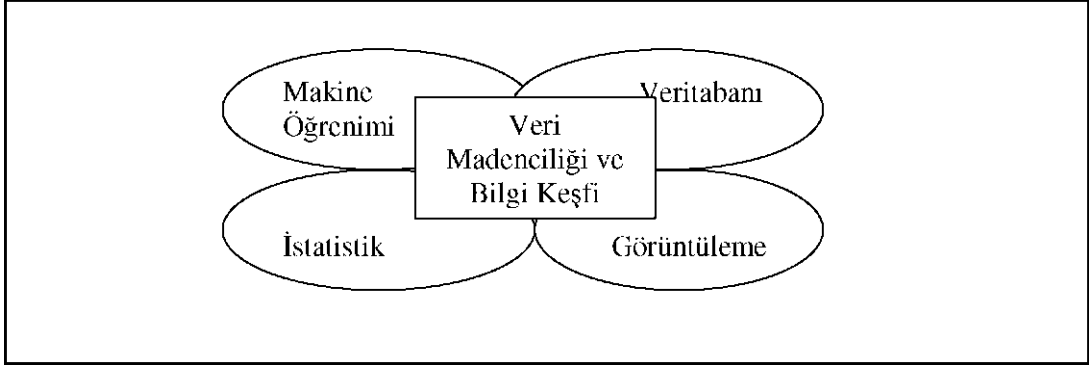
Özellikler	OLTP	OLAP
Karakteristik	Operasyonel İşlem	Bilgisel(informational) İşlem
Oryantasyon	İşlem	Analiz
Kullanıcı	Veritabanı Uzmanı	Knowledge worker (Yönetici, Analist)
Fonksiyon	Günlük İşlemler	Karar Destek Sistemleri
Veritabanı Dizaynı	Birim İlişkisel Tabanlı Uygulama Eğilimli	Star/Snowflake Konu Eğilimli
Veri	Güncellenebilir	Tarihsel(accuracy maintained over time)
Özetleme	İlkel, Detaylı	Özetlenmiş, Birleştirilmiş
Görüntü	Detaylı, Tam (flat)İlişkili	Özetlenmiş, Çok boyutlu
Çalışma Birimi	Kısa, Basit İşlem	Karmaşık Sorgu
Erişim	Okuma, Yazma	Genelde Okuma
Odak	Veri Girişi	Bilgi Çıkışı
Operasyonlar	İndeks, Hash on primary key	Çok tarama(gözden geçirme)
Ulaşılan Kayıt Sayısı	Onlarca	Milyonlarca
Kullanıcı Sayısı	Binlerce	Yüzlerce
Veritabanı Büyüklüğü	100 MB-GB	100 GB-TB
Öncelik (Önemlilik)	Yüksek Performans Yüksek Ulaşım	Yüksek esneklik, Son kullanıcıya özerk
Sayısal (Metrik)	İşlem çıktısı (throughput)	Sorgu Çıktısı. Cevap Süresi

Kaynak: G. Piatetsky-Shapiro and W. J. Frawley. Knowledge Discovery in Databases. AAAI/MIT Press, 1991

Veri madenciliği, veritabanı teknolojisi, istatistik, makine öğrenimi, örüntü tanıma, yapay sinir ağları, verilerin görselleştirilmesi, uzaysal veri analizi gibi farklı disiplinlerde yer alan tekniklerin birleşimini içermektedir. Kuşkusuz bunlar

⁸⁷ Han ve Kamber, s.42-43.

içerisinde en büyük katkıyı istatistik, makine öğrenimi ve veritabanı teknolojisi sağlamaktadır.



Şekil 3. 4 Veri Madenciliği ile Diğer Disiplinlerin Gösterimi

Kaynak: Jiawei Han and Micheline Kamber, Data Mining: Concepts and Techniques (New York: Morgan Kaufmann Publishers, 2001, s.7.

Burada, veri madenciliğinin istatistikte karşılaşılan problemlerle ilişkisine değinmek yararlı olacaktır.

Toplanan veriler, bağımsızlık, dağılımlarının aynı olması ve aralarındaki ilişkinin zayıf olması gibi geleneksel istatistik yöntemlerin varsayımlarını yerine getiremeyebilirler. Uygulamalarda veri küplerinde toplanan veriler genellikle yüksek ilişkiye sahiptir. Yukarıda belirtilen varsayımlardan herhangi birisinin ihlal edilmesi durumunda araştırmacı sağlıklı sonuçlar elde etmeye çalıştığında büyük sorunlarla karşılaşmaktadır.⁸⁸

Veri küpü içinde bütün temel eksenler ve köşegenler (diagonally) sistematik bir şekilde incelenerek ilişkiler araştırılmaktadır. Bu veri küpü içerisindeki araştırmayı, stokastik modelleme, kümeleme, sınıflandırma, regresyon, görselleştirme gibi çok değişkenli istatistiksel analiz yöntemlerini uygulayarak da yapmak mümkündür.

⁸⁸ NCSS, **User Guide**, Kaysville, NCSS Inc., 2001.

Ancak yukarıda anlatılan verilerdeki karmaşıklık bazı istatistik problemlerin ortaya çıkmasına neden olmaktadır. Bunları aşağıdaki gibi sıralamak mümkündür⁸⁹.

- Verilerdeki örüntülerin araştırılmasında, belli örüntülerin tesadüfi olarak dalgalanmaları halinde bu örüntülerin sonuçta elde edilen örüntüler olduğu sonucuna varılmasına neden olur.
- İstatistikte, örneklem hacmi küçük iken veri madenciliğinde örneklem hacmi oldukça büyüktür. Bunun sonucu olarak; veri madenciliğinde hipotez testlerinin olabilirlik oran metodları ve p-değeri en küçük etkilerin de önemli olduğu yönünde karar verme eğiliminde olacaktır. Bu nedenle Bayesian yöntemler daha çok tercih edilmektedir. Çünkü bu yöntemlerde karar genellikle sıfır hipotezinin reddedilmemesi yönünde olmaktadır.
- İlişkisel veritabanlarında veriler, hiyerarşik olmakla beraber dağınık olduğu için verileri kontrol etmek oldukça güçtür.
- Veri setlerinin büyüklüğü genellikle bilgisayarların kapasitesini geçmektedir.
- Veri madenciliğinde, veriler olasılıklı örnekleme tekniklerinden çok, kolay (convenience) örnekleme yöntemine dayanmaktadır. Bunun nedeni ise, kolay örnekleme yönteminin her türlü ön yargıyı anakütle parametresinin tahminine sokmasıdır.
- Birçok istatistiksel çıkarımda, sonuçlar nedensel değil ilişkisel (associative) olabilir.
- Veri küpünün büyüklüğü $n \times p$ (n: nesne sayısı, p: değişken sayısı) gibi modelleme problemlerine neden olmaktadır.
- Yüksek boyutlu modelleme, çok sayıdaki ortak bağımlı değişkenin davranışının incelenmesini gerektirir. Ancak bu incelemeyi yapmak oldukça zordur.
- Yüksek boyutlu olasılık dağılımlarının sayısal olarak değerlendirilmesi oldukça zordur

⁸⁹ Bozdoğan,H., **Statistical Data Mining and Knowledge Discovery**, Washington: A CRC Press Company, 2004, s.6-7.

- Çok boyutlu korelasyon matrisi istatistik modellerde kullanılabilmesine rağmen bu matrisin anlaşılması oldukça güçtür.
- Çok boyutlu olasılık tablolarını kullanarak geleneksel istatistik modellerini analiz edebilmek için şunlar gereklidir:
 - Daha az boyutlu bir tablo oluşturmak için birçok boyutu ortadan kaldırmak.
 - Çok boyutlu olasılık tablosunda bulunan verilerin Bayesian analizine izin verecek şekilde eksik verileri de içeren önceki verileri anlamlı hale getirmek.
- Her ikisinin de aynı anda uygulamak.

Yukarıda anlatılmaya çalışıldığı gibi bazı istatistiksel yöntemler ile çözümlenemeyen problemler veri madenciliği yöntemlerinin uygulanması ile çözümlenebilmektedir⁹⁰.

Berry ve Linoff (1997) veri madenciliği ile ilgili olarak iki yaklaşım üzerinde durmuşlardır. Bu yaklaşımlar, yukarıdan aşağıya (doğrulayıcı) ve aşağıdan yukarıya (keşfedici) şeklindedir. Yukarıdan aşağıya hipotezleri kabul veya reddetmeye ve kısmi olarak bilinen bir olaya ilişkin bilgileri genişletmeye çalışır. Bunu yaparken geleneksel istatistiksel yöntemleri kullanır. Aşağıdan yukarıya yaklaşımı ise kullanıcının daha önceden farkına varılmayan yararlı bilgileri ararken verileri araştırmasını sağlar. Bu yaklaşım tipik veri madenciliğidir. Aslında her iki yaklaşım birbirini tamamlamaktadır. Öyle ki önemli ilişkileri tanımlayan aşağıdan yukarıya yaklaşımında elde edilen bilgiler bu keşiflerin neden yararlı olduklarını ve ne derece geçerli olduklarını açıklayamaz. Birinci yaklaşım bu keşifleri onaylamak ve keşiflere dayanan kararların kalitesini değerlendirmek için kullanılabilir⁹¹.

İstatistik; anakütleden tesadüfi olarak alınan bir örneklem yardımı ile tümavarım yöntemini kullanarak anakütle hakkında çıkarımlar yapar. Daha çok

⁹⁰ Bozdoğan, s.6-7.

⁹¹ Giudici, , s.6.

teoriye dayalı yaklaşımlar olup bir varsayımın doğruluğunu araştırır. Makine öğrenimi daha çok sezgisel yaklaşımlarla ilgilenir ve öğrenme işleminin başarılmasını artırmaya çalışır. Veri madenciliği ve bilgi keşfi ise teori ve sezgisel yaklaşımları birleştirerek bilgi keşfinin tüm aşamalarını (veri temizleme, sonucu sunma, yorumlama gibi) gerçekleştirir ve tündengelim yöntemini kullanır.

Bu disiplinlerin aralarındaki kesin sınırları tanımlamak zordur. Ayrıca bu disiplinler ile veri madenciliği arasındaki ayrımı net olarak belirlemek de mümkün değildir⁹².

3.2 Veri Madenciliği Aşamaları

Veri madenciliği, daha önceki başlıklardada bahsedildiği üzere eldeki verilerden üstü kapalı, çok net olmayan, önceden bilinmeyen ancak potansiyel olarak kullanışlı bilginin çıkarılmasıdır. Bu da; kümeleme, veri özetleme, değişikliklerin analizi, sapmaların tespiti gibi belirli sayıda teknik yaklaşımları içerir.

Başka bir deyişle, veri madenciliği, verilerin içerisindeki desenlerin, ilişkilerin, değişimlerin, düzensizliklerin, kuralların ve istatistiksel olarak önemli olan yapıların yarı otomatik olarak keşfedilmesidir.

Temel olarak veri madenciliği, veri setleri arasındaki desenlerin ya da düzenin, verinin analizi ve yazılım tekniklerinin kullanılmasıyla ilgilidir. Veriler arasındaki ilişkiyi, kuralları ve özellikleri belirlemekten bilgisayar sorumludur. Amaç, daha önceden fark edilmemiş veri desenlerini tespit edebilmektir.

Veri madenciliğini istatistiksel bir yöntemler serisi olarak görmek mümkün olabilir. Ancak veri madenciliği, geleneksel istatistikten birkaç yönde farklılık gösterir. Veri madenciliğinde amaç, kolaylıkla mantıksal kurallara ya da görsel sunumlara çevrilebilecek nitel modellerin çıkarılmasıdır. Bu bağlamda, veri madenciliği insan merkezlidir ve bazen insan – bilgisayar arayüzü birleştirilir. Veri

⁹² Hand ve diğerleri, s.4.

madenciliği sahası, istatistik, makine bilgisi, veritabanları ve yüksek performanslı işlem gibi temelleri de içerir.

3.2.1 Sınıflama Teknikleri

Sınıflandırma, yeni bir nesnenin niteliklerini inceleme ve bu nesneyi önceden tanımlanmış bir sınıfa atamaktır. Burada önemli olan, her bir sınıfın özelliklerinin önceden net bir şekilde belirlenmiş olmasıdır.

Sınıflandırma en çok bilinen veri madenciliği tekniklerinden birisidir; resim, örüntü tanıma, hastalık tanıları, dolandırıcılık tespiti, kalite kontrol çalışmaları ve pazarlama konuları sınıflandırma tekniklerinin bolca kullanıldığı alanlardır. Sınıflandırma Tahminleyici bir modeldir; havanın bir sonraki gün nasıl olacağı ya da bir kutuda ne kadar mavi top olduğunun tahmin edilmesi aslında bir sınıflandırma işlemidir.⁹³

Sınıflama modeli, mevcut veriler içinden sınıfı tanımlanmış verilerden yola çıkarak, sınıfı belli olmayan verilerin sınıfını belirlemek üzere kullanılan veri madenciliği modelidir.⁹⁴ Sınıflandırma için öncelikle tahmin için kullanılacak bir model oluşturulur. Oluşturulan bu model sınıfı belli olmayan veriler üzerinde uygulanarak sınıflar tahmin edilir.⁹⁵

Sınıflandırma modelini etkileyen birçok durum vardır. Bunlar aşağıda kısaca açıklanmıştır :

⁹³ Dunham, 2003

⁹⁴ Han ve Kamber, 2001

⁹⁵ Han ve Kamber, 2001

Tablo 3. 3 Sınıflandırma Modelini Etkileyen Durumlar

Sınırlı Bilgi	Veri tabanları belli bir işi gerçekleştiren programların çalışması, üretilen verilerin saklanması amacıyla oluşturulurlar. Oluşturulma nedeni veri madenciliği yapmak olmadığı için yapıları genelde modeller için uygun değildir. Bu yüzden, öğrenme görevini kolaylaştıracak bazı özellikler bulunmayabilir.
Gürültü ve Eksik Değerler	Veri özellikleri ya da sınıflarındaki hatalara gürültü adı verilir. Eksik bilgi veya ilişkisel olmayan veri tabanı yapıları üzerinde yapılan veri madenciliği amacına tam olarak ulaşmayabilir.
Belirsizlik	Yanlışlıkların şiddeti ve verideki gürültünün derecesi ile ilgilidir. Veri tahmini bir keşif sisteminde önemli bir husustur.
Karmaşıklık	Verilerin çok karmaşık sistemlerde bulunuyor olması performans problemlerini de doğuracak bu problemlerinde aşılması gerekecektir.

Kaynak:Berry, M., J., A., & Linoff, G., S., Mastering data mining. New York: Wiley,2000

Sınıflandırmanın ilk adımında model oluşturulur. Veri tabanındaki kayıtların analizi yapılarak model oluşturulur. Her kayıt, sınıf etiket niteliği olarak adlandırılan nitelik tarafından tanımlanmış sınıflardan birine dâhildir. Veri tabanından rast gele seçilen kayıtlardan oluşan eğitim veri kümesindeki kayıtların analizi sonucunda bir model oluşturulur.⁹⁶ Bu adım, denetimli öğrenme (supervised learning) olarak da adlandırılabilir. Denetimli öğrenmenin denetimsiz öğrenmeden farkı, denetimli öğrenmede sınıfların önceden belirli olduğudur. Denetimsiz öğrenme modeli olan kümelemede ise sınıflar belli değildir. Sınıflamanın ikinci adımında; modelin doğruluğu hesaplanır. Bunun için veri tabanından rast gele kayıtlar (eğitim kümesinden farklı olan kayıtlar) seçilerek test veri kümeleri oluşturulur. Model bu test veri kümeleri üzerinde uygulanarak doğruluğu hesaplanır. Eğer modelin doğruluğu kabul edilebilir bir değerse, sınıfı belli olmayan kayıtların hangi sınıfa dahil olacağını tahmin etmede bu model kullanılır.⁹⁷

⁹⁶ Han ve Kamber 2001

⁹⁷ Han ve Kamber 2001, Gehrke 2002

Sınıflama modeli kullanılarak bilimsel deneyler, tıbbi teşhisler, hile analizi, kredi onayı, pazarlama gibi çeşitli alanlarda uygulama geliştirilebilir. Literatürde birçok sınıflama tekniği mevcuttur. Karar Ağaçları (Decision Trees), Genetik Algoritmalar (Genetic Algorithms), Yapay Sinir Ağları (Artificial Neural Networks), K-En Yakın Komşu (K-Nearest Neighbour), Bayesyen Sınıflandırması (Bayesian), Bulanık Küme Yaklaşımı, Birliktelik kuralları bunlardan en önemli olanlarıdır.

3.2.1.1 Karar Ağaçları

Karar ağaçları sınıflandırma problemlerinde en çok kullanılan algoritmalarından birisidir. Diğer yöntemlerle kıyaslandığında karar ağaçlarının yapılandırılması ve anlaşılması daha kolaydır denilebilir.⁹⁸ Bu teknikte sınıflandırma için bir ağaç oluşturulur; daha sonra, veritabanındaki her bir kayıt bu ağaca uygulanır ve çıkan sonuca göre de bu kayıt sınıflandırılır. Temel olarak iki adımdan oluştuğu söylenebilir: Birincisi ağacın kurulması, ikincisi de verilerin teker teker ağaca uygulanarak sınıflandırmanın gerçekleştirilmesi şeklindedir.

Karar ağacı, çok sayıda kayıt içeren bir veri kümesini, bazı kuralları uygulayarak daha küçük kümelere bölmek için kullanılan bir yapıdır. Karar ağacı kökten yapraklara doğru yinelemeli olarak veriyi bölerek kazanma yöntemine göre inşa edilir. Başlangıçta bütün veriler ağacın kökünde toplanır. Değişkenlerin seçimi bilgi kazanımı değerine göre belirlenir. Yinelemeli olan algoritmanın döngüden çıkması için o düğümdeki tüm öğelerin aynı sınıfa dahil olması şartı vardır. Eğer kalan değerler sadece bir sınıfa aitse veya sınıflandırılacak değer kalmadıysa döngüsel algoritma sonlanır ve karar ağacı oluşturulmuş olur. Sonuçta oluşan sınıflardaki her bir eleman aynı sınıfın diğer elemanları ile benzer özellikler gösterir. Ağaç yapısı heterojen yapıdaki veri kümesinin daha küçük ve homojen bir yapıya dönüşmesi için kurallar tanımlar. Ağaç inşası sonunda elde edilen ağaç büyük ağaç

⁹⁸ Agrawal, 1993

olarak adlandırılır ve öğrenme kümesindeki deney ünitelerine en uygun ağaçtır. Ancak maksimum ağaç pratikte iki dezavantaja sahiptir.⁹⁹

- Büyük ağaç öğrenme kümesini kusursuz biçimde tanımlar, çünkü eklenen her bağımsız değişken hatalı sınıflama oranını düşürür.
- Bir sınıflama ağacının karmaşıklık ölçüsü o ağacın terminal düğüm sayısına eşittir. Terminal düğüm sayıları ve dolayısıyla karmaşıklığı yüksek olan büyük ağacın anlaşılması ve yorumlanması güçtür.

Büyük ağacın pratikte ortaya çıkardığı bu sorunların çözümü için maksimum ağacın budanması gereklidir. Büyük ağacın budanması daha küçük ağaçlar dizisi oluşturur ve oluşturulan bu dizi içerisinde optimum ağaç seçilir. Optimum ağaç büyük ağaçtan daha az karmaşıklığa sahiptir ancak, öğrenme kümesine büyük ağaçtan daha az uyumludur ve hatalı sınıflama oranı daha yüksektir.¹⁰⁰ Karar ağacı kullanımının durumdan duruma avantaj ve dezavantajları vardır.¹⁰¹

Avantajları arasında Karar ağacı oluşturmak zahmetsizdir, yorumlamak kolaydır, anlaşılabilir kurallar oluşturulabilir, sürekli ve ayrık nitelik değerler kullanılabilir. Dezavantajları ise; Sürekli nitelik değerlerini tahmin etmekte çok başarılı değildir ve sınıf sayısı fazla ve öğrenme kümesi örnekleri sayısı az olduğunda model oluşturma konusunda iyi bir performans göstermezler.

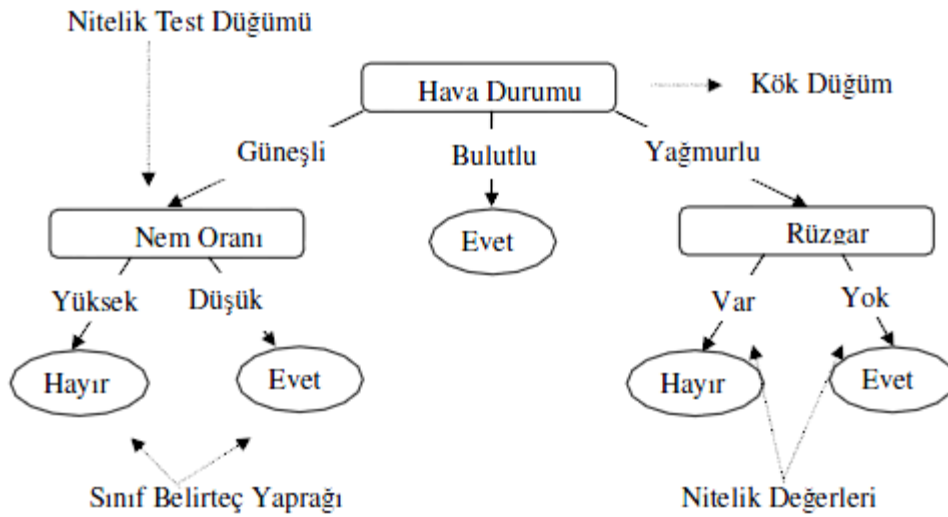
Karar ağaçları akış diyagramları gibi teknik olmayan kullanıcılar tarafından açık bir şekilde anlaşılır bir şekildedir. Herbir düğümde kendisinden sonraki yolu belirleyecek soru vardır. Bu sorunun cevabına göre ağaç dallanır ve sonuç kümesine doğru ilerler. Burada önemli olan en az soruyla en doğru sınıfa doğru gitmektir. Bütün dallar yaprak düğümle sonlanınca karar ağacı oluşturulmuş olur. Karar ağacı modellendikten sonra kolay anlaşılabilir olması için sınıflama kuralları

⁹⁹ Gülhan Örekici Temel, 2005, s.115-117.

¹⁰⁰ Gülhan Örekici Temel, 2005, s.115-117.

¹⁰¹ Tan ve Steinbach, 2003

oluşturulur.¹⁰² Sınıfı belli olmayan bir kaydın, oluşturulan bu karar ağacı kullanılarak hangi sınıfa ait olduğu belirlenebilir. Kayıt, ağacın kök düğümünden başlar ve karar düğümünde hangi yöne dallanacağı belirlenir. Her bir sınıf ağaçta tek yaprak olarak gösterilir. Bu yüzden bir sınıfa giden sadece bir yol olmalıdır. Yapraklar arasında herhangi kısa bir yol veya bağ yoktur. Dallanma işlemi yaprak düğümüne ulaşıncaya kadar devam eder.¹⁰³ Değişik yaprak düğümler aynı sınıfı temsil edebilirler fakat her bir yaprağın bu sınıflama için farklı bir nedeni vardır.¹⁰⁴ Aşağıdaki şekilde bir karar ağacı örneği gösterilmiştir.



Şekil 3. 5 Karar Ağaçları Gösterimi

Tahmin edici ve tanımlayıcı özelliklere sahip olan karar ağaçları, veri madenciliğinde hazırlanmasının ucuz olması, yorumlanmalarının kolay olması, veri tabanı sistemleri ile kolayca entegre edilebilmeleri, güvenilirliklerinin daha iyi olması nedeni ile sınıflama modelleri içerisinde en yaygın kullanılan tekniktir.¹⁰⁵ Karar ağacı oluşturmak için çeşitli algoritmalar geliştirilmiştir. Geliştirilen bu algoritmalar içerisinde CHAID (Chi-Squared Automatic Interaction Detector),

¹⁰² Murthy, 1998

¹⁰³ Berry, 2003; Utgoff 1998

¹⁰⁴ Berry 2003

¹⁰⁵ Akpınar 2000, Gehrke 2002

CART (Classification and Regression Trees)¹⁰⁶, ID3¹⁰⁷, Exhaustive CHAID, C4.5¹⁰⁸, MARS (Multivariate Adaptive Regression Splines), QUEST (Quick, Unbiased, Efficient Statistical Tree), C5.0, SLIQ (Supervised Learning in Quest)¹⁰⁹, SPRINT (Scalable Parallelizable Induction of Decision Trees)¹¹⁰ algoritmaları başlıcalarıdır.¹¹¹

3.2.1.1.1 Karar Ağaçlarında Algoritmalar

– **C4.5 Algoritması:** Quinlan (1993) tarafından ortaya atılan bir algoritmadır. Veriyi özyinelemeli olarak alt kümelerle ayırarak bir sınıflama karar ağacı oluşturur. C4.5 algoritması ID3 algoritmasının geliştirilmiş bir versiyonu olarak düşünülebilir. Algoritma sürekli değerlere sahip nitelikler için kullanılabilir, budama işlemi yapılabilmektedir ve karar üretilmesi gerçekleştirilebilir.¹¹²

– **CART Algoritması:** CART (Classification and Regression Trees), Breiman ve diğ. (1984), tarafından geliştirilen bir algoritmadır. CART algoritmasında, her aşamada ilgili kümenin, kendinden daha homojen olan iki alt kümeye ayrılması sağlanmaktadır. Ayrım işlemi kategorik bağımlı değişkenler için gini, twoing, sürekli değişkenler için en küçük kareler sapması (Least-Squared Deviation) indeks hesaplamalarına göre yapılmaktadır. Bu hesaplamalarda kar, maliyet değerleri ve değişken kategorileri arasındaki önceliklerin tanımlanabilmesi gibi sağlanan çeşitli esneklikler, CART algoritmasının günümüzde de yoğun olarak tercih edilmesine neden olmaktadır.¹¹³

– **CHAID Algoritması:** Bir başka karar ağacı algoritması da CHAID (Chi-Square Automatic Interaction Detector) algoritmasıdır. CHAID, CART algoritmasına

¹⁰⁶ Breiman, 1984

¹⁰⁷ Quinlan 1986

¹⁰⁸ Quinlan 1993

¹⁰⁹ Mehta, 1996

¹¹⁰ Shafer, 1996

¹¹¹ Joshi 1997, Akpınar 2000

¹¹² Berson 2000

¹¹³ Breiman, 1984

benzemektedir fakat, veriyi bölümlere ayırırken farklı bir yol kullanmaktadır. Optimum bölümleri seçmek için kullanılan entropy veya gini metrikleri yerine chi square testi uygulayan bir teknik kullanılır.¹¹⁴ Kategorik ve sürekli değişkenler üzerinde çalışabilmesi, ağaçta her düğümü ikiden fazla alt gruba ayırabilmesi gibi nedenlerle günümüzde de tercih edilen bir algoritmadır.

– **SLIQ Algoritması:** SLIQ (Supervised Learning In Quest), IBM Quest tarafından geliştirilmiştir. Büyük veri kümelerini bölümlere ayırarak karar ağacı oluştururken ön-sıralama tekniğini kullanır. Bu teknik her düğümdeki sıralama masrafını büyük ölçüde önlemiş olur. SLIQ her düğüm için sınıf listesi olarak adlandırılan ayrı sıralanmış bir liste tutar. Bu listedeki her eleman verideki niteliklere karşılık gelmektedir ve bir sınıf etiketine sahiptir. SLIQ, karar ağacını oluştururken genişlik öncelikli yolu kullanır. Her nitelik için uygun sıralanmış listeyi tarar ve her değer için entropy değerini hesaplar. Her nitelik için entropy hesaplandıktan sonra veriyi bölmek için bir nitelik seçilir. Bu işlem veri sınıflara ayrılana kadar yinelemeli olarak devam eder.¹¹⁵

– **SPRINT Algoritması:** SPRINT (Scalable PaRallelizable INduction of decision Trees), çok büyük veri kümeleri için çalışır ve ana bellek ile eğitim veri tabanının boyutu arasındaki tüm ilişkileri ortadan kaldırır. SPRINT, sınıf ve kayıt numarasını tutan farklı bir nitelik listesi yapısı kullanır. Bir düğüm bölündüğünde nitelik listeleri de sırasıyla bölümlenir ve oluşan çocuk düğümlere dağıtılır. Bir liste bölümlendiğinde listedeki kayıtların sıraları da yeniden düzenlenir. SPRINT, ölçeklenebilirliğini arttırmak amacıyla kolayca paralelleştirilebilecek şekilde tasarlanmıştır.¹¹⁶

– **ID3 algoritması:** J. Ross Quinlan (1986) tarafından geliştirilmiştir. Veri tabanında çok nitelik varsa ve eğitim kümesi çok fazla kayıt içeriyorsa fakat fazla hesaplama yapmadan makul bir karar ağacı oluşturulmak isteniyorsa ID3 algoritması

¹¹⁴ Berson 2000

¹¹⁵ Mehta ,1996, Joshi 1997

¹¹⁶ Shafer ,1996, Gehrke 2002

kullanılabilir. ID3 yinelemeli yapıya sahip bir algoritmadır. ID3, CLS (Concept Learning System) algoritması tabanlıdır.¹¹⁷

3.2.1.1.2 Karar Ağaçlarında Budama İşlemleri

Karar ağaçlarında karşılaşılan problemlerin başında aşırı öğrenme (overfitting) ve hatalı öğrenme (underfitting) adı verilen durumlar gelmektedir. Ağaç çok büyük ve eğitim örneklerine ait hata oranı düşük olduğu halde test verisi için sınıflandırma hatası büyük ise bu duruma aşırı öğrenme denir (Tan ve Stainbach). Karar ağacı yeteri kadar büyük olmadığı zaman model hatalı sonuçlar veriyor ise bu durum az öğrenme (model underfitting) olarak adlandırılır. Bu gibi durumları oluşumunu engellemek ve ağacın optimum hale getirilmesini sağlamak amacıyla budama adı verilen işlemler gerçekleştirilir. Budama kural olarak ağaçta yer alan ancak ağacın dengesini ve güvenilirliğini bozacak dalların ağaçtan çıkarılmasıdır. Budama için iki yöntem kullanılabilir. Bunlardan ilki ön budama olarak adlandırılan ve ağacın belli bir büyüklüğe ulaştığı zaman büyümesinin durdurulmasıyla sağlanan budamadır. Diğer yöntem ise ağaç tamamıyla oluştuğundan sonra budamadır. Ön budama işlemleri için örneklerin adedi kullanıcı tanımlı eşikten daha az ise dur komutuyla veya örneklerin sınıf dağılımı kullanılabilir özelliklerden bağımsız ise dur komutuyla sağlanabilir. Karar ağaçlarında büyüme aslında o ağacın bütünlüğü içindir. Ancak bu bütünlük iyileştirilmiş sonuç için yeterli değildir. Son budama işleminde ise düğümler aşağıdan yukarıya doğru budanır. Eğer budama sonrası genelleştirme hatası artıyorsa bir yaprak düğümü ile alt ağaç değiştirilir. Son budamaya diğer bir yöntem olarak da yaprak düğümünün sınıf etiketi alt ağaçtaki örneklerin farklı sınıflarından belirlenecek şekilde uygulanabilir.

3.2.1.2 Genetik Algoritmalar

Algoritma, problem çözmek için kullanılan adımlar serisidir. Evrimsel programlamanın (evaluationary computing) bir parçası ve bu alandaki en popüler teknik olan Genetik Algoritmalar (GA) biyolojideki evrim teorisinden esinlenilerek ortaya

¹¹⁷ Quinlan 1986

çıkıştır. Doğada da popülasyondaki bireyler her biri kısıtlı olan yiyecek, barınak vb. faktörler için birbirleri ile rekabet etmek durumundadır. Bununla birlikte aynı türler içerisinde bireyler üreme için eş seçiminde birbirleri ile de yarışmak durumundadır. Bu seleksiyon nedeni ile zayıf performans gösteren bireyler yaşamlarını sürdürmek için daha az şansa sahiptir ve daha iyi adapte olmuş veya uygun bireyler nispeten daha fazla sayıda döl ortaya koymaktadır.¹¹⁸ Üreme esnasında her bir ebeveynin iyi niteliklerinin birleşimi, uygunluğu atasından daha fazla olan yeni bir birey meydana getirebilir. Doğal evrim çevrelerine iyi adapte olmuş bireyler ortaya koyar. Benzer bir çevrenin hesaplama yapısını kontrol ederek genetik algoritmalar kromozomları (yapıları) çözümü istenen probleme daha uygun biçime dönüştürebilmektedir.¹¹⁹

3.2.1.2.1 Genetik Algoritmaların Gelişimi

Genetik algoritmalar da ilk olarak 1960'larda Holland tarafından bulunmuş ve daha sonra 60 ve 70'li yıllarda Michigan Üniversitesindeki çalışma arkadaşları ve öğrencileri ile birlikte geliştirilmiştir. Evrimsel strateji ve evrimsel programlamadan farklı olarak Holland'ın ilk amacı spesifik bir problemin çözümü için algoritma ortaya koymaktan ziyade doğada olduğu gibi bir adaptasyonu biçimsel olarak çalışmak ve doğal adaptasyon mekanizmalarının bilgisayar sistemlerine uygulanabileceği metotlar geliştirmektir.

Genetik algoritmaların arama mantığı Holland tarafından formüle edilen şema teoremi üzerine kurulmuştur. Şema, dizilerdeki tanımlanmış belirli kısımların benzerliklerini gösteren bir alt diziyi tanımlayan sabit bir şablondur. Benzer şema içeren diziler bir dereceye kadar benzer bilgi içermektedir.¹²⁰ Şema, 0,1 ve * sembollerinden oluşan bir dizidir. Şema teoreminin temel karakteristiğini yansıtan '*' sembolü, uygulamada 0 ya da 1 değeri alabilir. Burada '*', belirli uzunluktaki diziler arasındaki muhtemel tüm benzerlikleri göstermeye imkan veren bir notasyondur.

¹¹⁸ Sivanandam ve Deepa, 2008, s;31

¹¹⁹ Pelikan ve diğer. , 2006, s;40

¹²⁰ Coley, 1999, s;3

Teorem, en fazla uygun olan şema içeren dizinin bir sonraki toplumda da hayatını devam ettireceği mantığına dayalıdır. Genetik algoritma çözümü en iyi şemanın tatbiki ile en iyi kromozoma doğru yaklaşır.¹²¹

Genetik Algoritmalar doğal seçim ve doğal genetik mekanizmalarına dayalı stokastik araştırma teknikleri olarak ortaya atılmıştır¹²². 1975 yılında yayınlanan "Doğal ve Yapay Sistemlerde Adaptasyon" (Adaptation in Natural and Artificial Systems) adlı kitabında Holland ilk olarak doğal evrimin optimizasyon problemlerine nasıl uygulanacağını anlatmış ve ilk genetik algoritmayı oluşturmuştur. Holland'ın GA'sı kromozomların bir popülasy onundan yeni bir popülasyona çaprazlama, mutasyon ve inversiyonun genetikten esinlenmiş operatörleri ile bir çeşit 'doğal seleksiyon' kullanılarak geçişi sağlayan bir metottur.¹²³ Orijinal genetik algoritmanın iki ana karakteristiği sabit uzunluktaki kromozom sunumu ve oldukça yoğun kullanılan çaprazlamadır.¹²⁴

Genetik Algoritmaların popüler kılınması ise Holland'ın doktora öğrencisi olan David Goldberg'in 1989 yılında yayınladığı "Araştırma Optimizasyon ve Makine Öğrenmesinde Genetik Optimizasyon (Genetic Algorithms in Search, Optimization, and Machine Learning) adlı kitapla olmuştur. Goldberg çalışmasında teorik çalışmaları başarılı uygulamalara taşıyarak konuya büyük bir katkı sağlamıştır.¹²⁵

Genetik algoritmaların bu alanda çalışan herkes tarafından kabul görmüş kesin bir tanımı olmamasına rağmen, genetik algoritmaları bazı bileşenlerinin diğer evrimsel algoritmalarından ayırdığı ifade edilmiştir. Bu elementler kromozom popülasyonu, uygunluğa göre seçim, yeni bir döl elde etmek için çaprazlama ve rastgele mutasyon olarak sıralanabilir.¹²⁶

¹²¹ Banzhaf ve diğer., 1998, s;96-97, Şen, 2002, s;99, Spall, 2003, s;263

¹²² Sakawa, 2002, s;1

¹²³ Mitchell, 1999, s;2

¹²⁴ Banzhaf ve diğer., 1998, s;95

¹²⁵ İşçi ve Korukoğlu, 2003, s;191, Haupt ve Haupt, 2004, s;22

¹²⁶ Mitchell, 1999, s;7

Genel bir tanımla Genetik Algoritmalar, en iyinin korunumu ve doğal seçim ilkesinin benzetim yoluyla bilgisayarlara uygulanması ile elde edilen bir arama yöntemidir.¹²⁷ Doğal genetik kalıtım davranışlarını teorik ve yapay olarak gerçekleştiren stokastik bir tekniktir.¹²⁸

Genetik Algoritmalar çok sayıda bireyden oluşan bir popülasyonun tanımlı seçim kuralları altında "uygunluğu" nu (fitness) maksimize eden bir duruma dönüşümüne imkan verir.¹²⁹

Genetik algoritmalar stokastik algoritmalar; olasılık genetik algoritmalarda temel bir role sahiptir. Genetik algoritmalar daima bir çözüm popülasyonunu dikkate alır. Algoritma farklı çözümleri bir araya getirerek daha iyilerini araştırır ve popülasyondaki çeşitlilikten yarar sağlar. Genetik algoritmaları uygulamadan önce problemle ilgili özel bir gereksinimi yoktur ve dolayısıyla herhangi bir probleme uygulanabilir.

Çoğu stokastik model gibi Genetik algoritmalar da global optimal çözümü garanti etmez. Probleme kabul edilebilir derecede iyi bir çözüm hedefler.¹³⁰ Genetik algoritmalar özellikle çözüm uzayının geniş, süreksiz ve karmaşık olduğu problem tiplerinde başarılı sonuçlar vermektedir.¹³¹

Belirli bir problem çözümüne yönelik hazırlanan bir genetik algoritma aşağıdaki beş bileşeni kapsamalıdır.¹³²

- Problemin potansiyel çözümü için bir genetik gösterim,
- Potansiyel çözümün başlangıç popülasyonunu oluşturmak için bir yol,

¹²⁷ Nahiye, 2003, s;606

¹²⁸ Şen, 2004, s;65

¹²⁹ Haupt ve Haupt, 2004, s;22

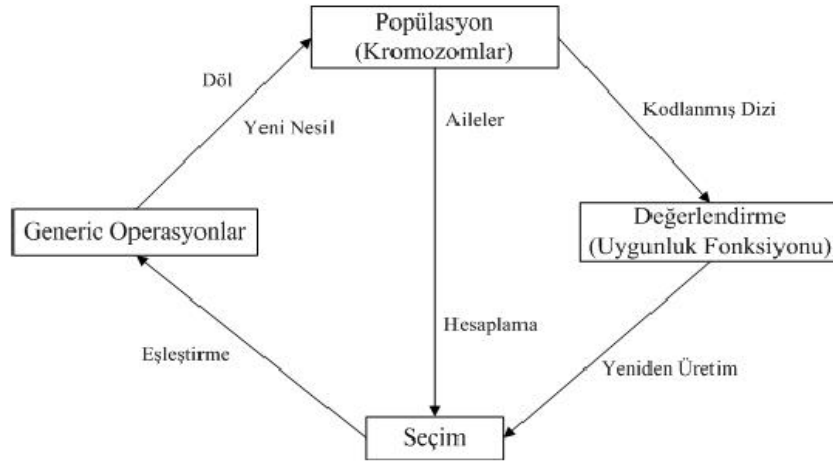
¹³⁰ Sivanandam ve Deepa, 2008, s;20

¹³¹ Emel ve Taşkın, 2002, s;139

¹³² Michalewicz, 1996, s;17, Gen ve Cheng, 2000, s;1

- Çözümleri uygunluklarına göre sıralayan bir değerlendirme fonksiyonu,
- Elde edilen çocukların bileşimini değiştiren genetik operatörler
- Genetik algoritma tarafından kullanılan (popülasyon büyüklüğü, genetik operatörlerin uygulama olasılıkları gibi) çeşitli parametre değerleri.

Genetik algoritmalar rastgele seçilen bireylerden oluşan ilk popülasyon ile başlar. Popülasyondaki her bir birey çözülmesi düşünülen problem için potansiyel bir çözüm sunar. Bu bireyler başarılı tekrarlamalar ile (iterations) yavaş yavaş gelişir. Tekrarlamalar ile yeni nesiller ortaya çıkar. Her bir nesilde popülasyondaki her bir birey bazı uygunluk ölçütleri kullanılarak değerlendirilir. Daha sonra da genetik operatörler vasıtasıyla bir sonraki nesil popülasyonu oluşturulur. Bu prosedür sonlandırma şartları sağlanıncaya kadar devam eder.¹³³ Genetik algoritma prosesindeki bu döngü Şekil 7 'de gösterilmiştir. ¹³⁴



Şekil 3. 6 Genetik Algoritma Döngüsü

Kaynak: Jourdan, L., Dhaenens, C., Talbi, E., , “A Genetic Algorithm For Feature Selection in Data-Mining For Genetics”, 4th Metaheuristics International Conference, 2001, 29-33.

Seçimin (selection) toplumdaki her bireyi karşılaştırabildiği düşünülür. Seçim işlemi uygunluk fonksiyonu kullanılarak yapılır ve her kromozom temsil ettiği çözümün uygunluğunu gösteren bir değere sahiptir. Optimal çözüm uygunluk fonksiyonunu

¹³³ Sakawa, 2002, s; 11

¹³⁴ Sivanandam ve Deepa, 2008, s;31

maksimize eden çözümdür. Seçilen kromozomlara genetik operatörlerin uygulanması ile birlikte yeni bireyler elde edilir ve yeni bir popülasyon oluşturulur. Bu döngü sonlandırma koşulları sağlanıncaya kadar sürer.¹³⁵

Genetik Algoritmaların temel prosedürleri ve işleyişi ise aşağıda gösterilmektedir.¹³⁶ Standart bir "Genetik Algoritma İşleyişi" aşağıdaki gibi verilebilir.¹³⁷

-Olası çözümlerin kodlandığı bir çözüm grubu oluşturulur. Bu çözüm grubu, biyolojideki benzerliği nedeniyle, toplum (population), çözümlerin kodları da kromozom olarak adlandırılır. Toplumda bulunacak birey sayısı için bir standart yoktur. İlk popülasyon genetik materyallerin çeşitliliğini mümkün olduğu kadar geniş yansıtmalıdır. Tipik popülasyon büyüklüğü birkaç düzineden binlere kadar değişebilir. Genel olarak önerilen 100-300 aralığında bir büyüklüktür. Büyüklük seçiminde yapılan işlemlerin karmaşıklığı ve aramanın derinliği önemlidir. Toplum bu işlemten sonra rastgele oluşturulur.

-Toplumdaki her kromozomun ne kadar iyi olduğu bulunur. Bu amaçla kullanılan fonksiyona uygunluk fonksiyonu denir. Uygunluk fonksiyonu genetik algoritmanın beynini oluşturmaktadır. Genetik algoritmada probleme özel çalışan tek kısım bu fonksiyondur. Uygunluk fonksiyonu, kromozomları problemin parametreleri haline getirerek onların bir bakıma şifresini çözmekte (decoding) ve sonra bu parametrelere göre hesaplamayı yaparak kromozomların uygunluğunu bulmaktadır. Çoğu zaman genetik algoritmanın başarısı bu fonksiyonun verimli ve hassas olmasına bağlıdır.

-Seçilen kromozomları eşleyerek yeniden kopyalama ve değiştirme operatörleri uygulanır. Sonuçta yeni bir toplum oluşturulur. Bu eşleme kromozomların uygunluk değerlerine göre yapılır. Bu seçimi yapmak için rulet tekerleği seçimi (roulette wheel selection), turnuva seçimi (tournament selection) gibi farklı seçme yöntemleri kullanılır. Yeniden kopyalama (recombination) genlerdeki genetik bilginin birinden

¹³⁵ Rutkowski, 2005, s; 271

¹³⁶ Sakawa, 2002, s;14.

¹³⁷ Jain ve Martin, 1998, s; 132, Karr ve Freeman, 1999, s;4, Nabiyeve, 2003, s;607, Güden ve diğer.2005, s;7, Sivanandam ve Deepa, 2008, s;30, Sumathi ve diğer., 2008, s;83.

iğerine geçmesi işlemine benzediğı için çaprazlama (crossover) olarak adlandırılır. Bu işlem, toplumda çeşitliliğı sağlar. Mutasyonun etkisi, yalnızca tek bir çözüm üzerinde olmaktadır.¹³⁸

-Yeni kromozomlara yer açmak için eski kromozomlar çıkartılarak sabit büyüklükte bir toplum sağlanır.

-Tüm kromozomların uygunlukları tekrar hesaplanır ve yeni toplumun başarısı bulunur.

-İşlemler tekrarlanarak verilmiş zaman içerisinde daha iyi olan yeni nesillerin oluşturulması gerçekleştirilir.

-Sonuçta toplumların hesaplanması sırasında en iyi bireyler bulunduğunda çözüm elde edilmiş olur. Popülasyon optimum çözüme yakınsadığında algoritma durdurulur.

-Yukarıda sıralanan işleyiş kapsamında genetik algoritma en az üç aşamada olasılığa dayalı adımlar içermektedir. ¹³⁹ Bu adımlar; başlangıç popülasyonunun oluşturulması, genetik operasyonlar için popülasyondan bireyler seçilmesi, çaprazlama veya mutaston gibi genetik operatörleri uygulamak için seçilen birey üzerinde bir nokta belirlenmesi olarak sıralanmaktadır.

3.2.1.2.2 Genetik Algoritma Terminolojisi

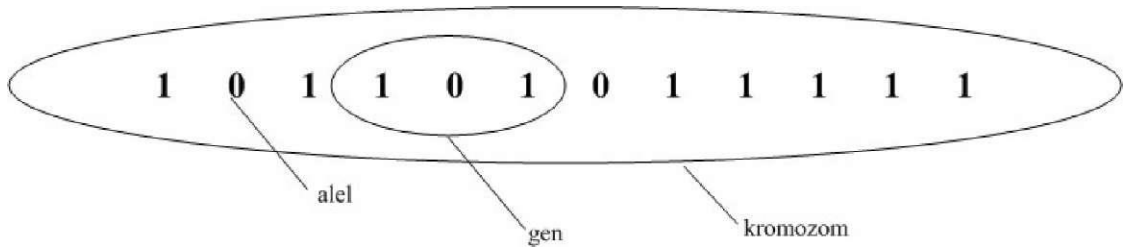
Genetik algoritmalar terminolojisinde biyolojiye ait terimler gerçek anlamları daha basite indirgenerek sıklıkla kullanılmaktadır. Bu açıdan bakıldığında kullanılan terminoloji hakkında bilgi verilmesinin yararlı olacağı düşünülmüştür.

¹³⁸ Ozcan, E., Alkan, A., “Cok Nufuslu Kararlı Hal Genetik Algoritması kullanarak Otomatik Cizelgeleme”,2005, s.76.

¹³⁹ Kinneer, 1994, s.22.

Tüm canlı organizmalar hücrelerden oluşmaktadır ve her bir hücre organizmanın bir 'plan'ını (blueprint) içeren DNA dizilerinden oluşan aynı kromozom setinin bir veya daha fazlasını taşır. Kromozomlar genlerin dizisidir. Bir kromozom kavramsal olarak genlere bölünebilir. Her bir gen farklı bir proteini kodlar. Çok genel biçimde bir genin göz rengi gibi herhangi bir özelliği kodladığı söylenebilir. Bir özellik için muhtemel farklı niteliklerin her biri (örneğin göz için mavi renk) 'alel' olarak adlandırılır. Bir bit dizisinde alel ise 0 veya 1'dir. Her bir gen de kromozom üzerinde 'lokus' adı verilen belirli bir yerde konumlanmıştır ¹⁴⁰

Organizmanın sahip olduğu genlerin ve dolayısıyla kromozomların tamamına 'genom' denilmektedir. Belirli bir bireyin genlerinin tüm bileşimine ise 'genotip' (genotype) denir. ¹⁴¹ Genotip, başka bir deyişle bir genom içerisindeki belirli bir gen setini ifade eder. Organizmanın çevresel ve genetiksel olarak belirlenen özelliklerine 'fenotip' (phenotype) adı verilir. Genel olarak bu özellikler saç rengi, boy veya zeka gibi gözlemlenebilir özelliklerdir. ¹⁴² Kromozom, gen ve alel ilişkisi aşağıdaki şekilde gösterilmiştir. ¹⁴³



Şekil 3. 7 Alel, Gen ve Kromozomlar

Kaynak: Werner, J.C., Fogarty, T.C., "Genetic Algorithm Applied in Clustering Datasets" Technical Report, London South Park University, (2001).

Genetik algoritmalarda kromozom terimi genelde bir bit dizisi olarak kodlanır ve probleme aday bir çözüme göndermede bulunur. Her bir kromozom yegane bir çözümü tanımlamalıdır, ancak bu durum her bir çözümün belirli bir kromozom

¹⁴⁰ Mitchell, 1999, s;5

¹⁴¹ Haupt ve Haupt, 2004, s;22

¹⁴² Mitchell, 1999, s;5

¹⁴³ Rothlauf, 2006, s;11

tarafından kodlandığı anlamına gelmez. Gerçekte ise tüm çözüm alanının taranabilmesi için problemin tüm aday çözümlerinin en azından muhtemel bir kromozoma uyması gerekir. Her bir genotipi onun fenotipi ile ilişkilendiren fonksiyona morfogeneze fonksiyonu denilir. Bu fonksiyon birebir değil ise temsil dejenere olmuş demektir. Optimal çözüm alanı çok geniş olsa bile küçük dejenerasyonlar önemsenmeyebilir ancak büyük dejenerasyonlar ciddi problemler doğurabilir. Farklı kromozomlar aynı fenotipi temsil edeceği için her bir gen çözümün belirli bir karakteristiğine tekabül etmez ve bu durum da genetik algoritmanın işleyişini olumsuz etkiler.

Genler ise ya bitler yada aday çözümün belirli bir elementini kodlayan bitişik bitlerin kısa bloklarıdır. Çok parametrelili bir optimizasyon fonksiyonunda belirli bir parametrenin kodunu oluşturan bitler gen olarak düşünülebilir. Genetik algoritma sunumunda bir gen, kontrol faktörünün aldığı tek bir faktör değeridir. Bu kontrol faktörünün alt ve üst sınırları olmalıdır.

Her bir genin yapısı fenotipleme parametrelerinin kayıtlarında tanımlıdır. Fenotipleme parametreleri genotip ile fenotip arasındaki eşleşmede kullanılan direktiflerdir. Kromozoma bir çözüm setini kodlarken ve kromozomun bir çözüm setine ulaşmak için kodunu çözerken (decoding) de bu direktifler geçerlidir. Genotip ve fenotip arasındaki eşleşme çözüm setini modelden Genetik algoritmanın çalışabileceği biçime dönüştürmek için ve ayrıca algoritmadaki yeni bireyleri modelin değerlendirebileceği bir forma dönüştürmek için gereklidir.¹⁴⁴

Uygunluk

Uygunluk değeri yeni popülasyona taşınacak bireylerin belirlenmesinde kullanılan bir araçtır. Bu nedenle algoritmanın her çevriminde mevcut popülasyondaki bireylerin uygunluk değerleri hesaplanır.¹⁴⁵ Genetik algoritmalarda

¹⁴⁴ Mitchell, 1999, s;5; Sivanandam ve Deepa, 2008, s.40-41.

¹⁴⁵ Dengiz ve Altıparmak, 1998, s.4.

bir bireyin uygunluđu onun fenotipinin ama fonksiyonundaki deđeridir. Uygunluđu hesaplamak iin ilk olarak kromozom kodlanmalı ve ama fonksiyonu deđerlendirilmelidir. Uygunluk fonksiyonu özölen problemin tipine bađımlıdır. Uygunluk yalnızca özöümün ne kadar iyi olduđunu göstermez; bunun yanında kromozomun optimale ne kadar yakın olduđuna da iřaret eder.¹⁴⁶

ok sayıda kriter bulunan optimizasyon durumlarında uygunluk fonksiyonunu belirlemek ok daha zordur. Bu gibi bir durumda bir özöüm diđerinden daha iyi olup olmadıđının nasıl belirleneceđine iliřkin bir sorunla karřılařılır. Bir özöüm bir kriter iin iyi sonu verirken bir diđer iin kötü ise ne yapılmalıdır? Ancak burada sorun, problemi özmek iin genetik algoritmanın nasıl uygulanacađından ziyade 'daha iyi' özöümün tanımından kaynaklanmaktadır. Eđer, farklı kriterlerin basit bir kombinasyonundan elde edilen bir uygunluk fonksiyonu iyi sonular verebiliyorsa bu kriterlerin birbirine uygun bir araya getirildiđi düşünölr.¹⁴⁷

Popölasyon

Popölasyon, test edilen bireyler, bireyleri tanımlayan fenotip parametreleri ve arařtırma alanı ile ilgili bazı bilgilerden oluřur. Genetik algoritmalarda kullanılan popölasyonun iki önemli yönü vardır:

- İlk popölasyon oluřumu
- Popölasyonun büyüklüđu

Tüm problemler iin popölasyonun büyüklüđu problemin karmařıklıđına bađlıdır. İdeal olanı, tüm arařtırma alanının arařtırılabilmesi iin ilk popölasyonun gen havuzunun mümkün olduđu kadar geniş olması gerekmektedir. Popölasyon eřitliliđi yansıtmaz ise algoritma dar bir alanı tarayacak ve muhtemelen global optimuma ulařılamayacaktır. Bundan kaçınmak iin de bařlangı popölasyonu birok durumda random olarak belirlenir. Bazı durumlarda yine de bařlangı popölasyonunu

¹⁴⁶ Rutkowski, 2005, s;268, Sivanandam ve Deepa, 2008, s.41.

¹⁴⁷ Sivanandam ve Deepa, 2008, s.41.

belirlemek için bir çeşit sezgisellik kullanılabilir. Böylece bu popülasyonun ortalama uygunluğu yüksek olur ve çözüme daha hızlı ulaşılabilir.¹⁴⁸

Seçilen popülasyon büyüklüğü de bazı problemler doğurabilmektedir. Daha geniş popülasyon araştırma alanının daha kolay taranması demektir. Ancak yakınsamak için genetik algoritma için gerekli zaman da popülasyon büyüklüğüne bağlı bir fonksiyonla belirlenir. Tüm bireylerin birbirine çok benzer ve daha ileri iyileştirmelerin yalnızca mutasyon ile mümkün olduğu durumda popülasyon yakınsamış demektir. Goldberg'de çalışmalarında genetik algoritmanın global optimuma ulaşma etkinliğinin büyük oranda popülasyon büyüklüğüne bağlı olduğunu göstermiştir. Sonuç olarak geniş popülasyon oldukça yararlıdır ancak çok daha fazla hesaplama maliyeti, hafıza ve zaman gerektirir. Pratikte genelde 100 civarında bir popülasyon büyüklüğü tercih edilmektedir ancak optimal popülasyon büyüklüğü büyük ölçüde seçilen problemin yapısına bağlıdır.¹⁴⁹

Veri Yapıları

Genetik algoritmalarındaki ana veri yapıları kromozomlar, fenotipler, amaç fonksiyon değerleri ve uygunluk değerleridir. Kromozom popülasyonunun tamamı, birey sayısı ve bireylerin genotiplerinin uzunlukları verilerek tek bir dizi içerisinde saklanabilir. Benzer biçimde bazı eşleştirmeler ile elde edilen dizayn değişkenleri veya fenotipler de tek bir dizi içinde saklanabilir. Gerçek eşleme kullanılan tersine kodlama/şifre çözme şemasına bağımlıdır. Amaç fonksiyon değerleri skaler veya vektörel olabilir. Bu durum uygunluk değerleri için de aynı olmak durumundadır. Uygunluk değerleri, amaç fonksiyonundan ölçekleme (scaling) veya sıralama (ranking) fonksiyonları kullanılarak elde edilebilir ve vektör olarak saklanabilir.¹⁵⁰

¹⁴⁸ Alataş, B., Arslan, A., , “Birlikte Kurallarının Madenciligi İçin Genetik Algoritma ve Bulanık Kume Tabanlı Yeni Bir Yaklaşım”, F. Ü. Fen ve Mühendislik Bilimleri Dergisi, 42-51 (2005).

¹⁴⁹ Kennedy ve diğer., 2001, s;153, Sivanandam ve Deepa, 2008, s.42.

¹⁵⁰ Sivanandam ve Deepa, 2008, s.43.

Arama Uzayı (Search Space)

Optimizasyon problemlerinde genellikle belirli bir çözüm seti arasından en iyi çözüm araştırılır. Tüm mümkün çözümlerin alanına arama uzayı adı verilir. Arama uzayındaki tüm noktalar olası bir çözümü temsil eder.¹⁵¹ Bundan dolayı her muhtemel çözüm, problemin tanımına bağlı olarak onun uygunluk değeri ile tanımlanır. Uygunluk değeri arama uzayındaki ilgili noktanın niteliği ile ilintilidir.¹⁵² Çözümler arasındaki farklar ölçülebildiği durumda arama uzayı kavramı geçerlidir ve her bir çözüme bir başarı yada uygunluk ölçütü tanımlanabilir (Coley, 1999, s;3).

Arama Stratejileri

Arama prosesi ilk popülasyonun belirlenmesi ve sonrasında sonlandırma koşulları sağlanıncaya kadar yeni bireylerin üremesini kapsar. Prosesin farklı amaçları olabilir. Global optimum noktanın bulunması bu amaçlardan biridir. Ancak genetik algoritmaların kullandığı tipteki modeller ile bu asla garanti edilemez. Bir sonraki iterasyonun daha iyi bir çözüm ortaya koyma olasılığı daima vardır. Aramanın diğer bir amacı daha hızlı yakınsama sağlamaktır. Amaç fonksiyonunu çalıştırmanın zor olduğu durumlarda daha hızlı yakınsama istenebilir ancak bu durumda da lokal bir yakınsama olasılığının ve düşük değerli bir optimal bulma ihtimali artmaktadır.

Aramanın başka bir amacı da iyi çözümden uzaklaşmadan çeşitliliği artırmaktır. Çözüm uzayının çok sayıda farklı optimum içerdiği durumda, ki bu durum uygunluk için de aynıdır, onların arasından seçim yapabilmek yararlıdır. Çünkü modeldeki faktör değerlerinin bazı kombinasyonları diğerlerinden daha uygun sonuçlar ortaya koyabilmektedir¹⁵³

¹⁵¹ Sivanandam ve Deepa, 2008, s.20.

¹⁵² Jain ve Martin, 1998, s.131.

¹⁵³ Sivanandam ve Deepa, 2008, s.43.

Kodlama

Genetik algoritmaları diğer arama yöntemlerinden ayıran en önemli özelliklerden biri parametrelerin kendileri yerine onları temsi eden serileri kullanmasıdır. Kodlama ayrı ayrı genleri simgeleme prosesidir. Bu proses bitleri, sayıları, ağaçları, listeleri veya başka diğer objeleri kullanarak gerçekleştirilir. Kodlama esasen problemin çözümüne bağlıdır. Genetik algoritma uygulamalarının başarısı bir ölçüde de kodlamanın iyi yapılmasına bağlıdır¹⁵⁴. Kodlama şemaları bir kodlamadan diğerine büyük ölçüde değişebilir. Hatta bazı durumlarda bir dereceye kadar bu şemalar da evrime uğrayabilmektedir. İlk uygulamalarda kodlama düzeni ardışık ikili alt dizilerden oluşan bir dizisi kromozom oluşturmak için kullanılmıştır. Son dönemde ise farklı çözüm parametreleri değişen noktalı kromozomlarla temsil edilmektedir. Bu durum birçok bilgisayar dilinde kullanılan yapılarla benzerliğinden dolayı programlamada birtakım avantajlar sağlamaktadır.¹⁵⁵

Literatürde geliştirilen bazı farklı kodlama biçimleri aşağıda sıralanmıştır:

İkili Kodlama

Genetik algoritma uygulamalarında amaç değerlerini kodlamada kullanılan en yaygın kodlama biçimidir. İkili kodlamada her bir kromozom ikili (binary) $\{0,1\}$ bir bit dizisi olarak kodlanır. Dizideki her bir bit, çözüme ait bazı karakteristikleri ortaya koyar. Her bir değişken uzunluğu kullanıcı tarafından belirlenen l_i uzunluğunda bir dizi olarak kodlanır. Daha sonra da bu alt diziler peşi sıra birleştirilir. Her bir bit dizisi bir çözüm demektir ancak bunun optimal çözüm olması gerekmez. Bit dizisinin kodlanma yolu problemten probleme değişebilmektedir.

¹⁵⁴ Dengiz ve Altıparmak, 1998, s;4, Banzhaf ve diğer., 1998, s.95.

¹⁵⁵ Karr ve Freeman, 1999, s.2.

Dizinin uzunluğu hassaslığa bağlıdır. Yüksek hassaslık gerektiren değişkenler daha geniş bir dizide kodlanır.¹⁵⁶ Çözüm için gerekli zaman ve çözümlerin sağlığı değişkenlerin temsili için kullanılacak hanelerin veya genlerin büyüklüklerine de bağlıdır. Ondalık sayılarda her bir hane için dizi üzerinde ortalama 3,3 tane haneye gereksinim olduğu düşünülürse 8 haneli bir hassasiyette 10 değişken için kromozom uzunluğu $8 \times 3,3 \times 10 = 270$ haneden oluşur. Bu açıdan problemin özelliğine göre önceden hassasiyet belirlenmelidir¹⁵⁷

İkili kodlama dizisi daha az sayıda alel ile çok sayıda muhtemel kromozom ortaya koyabilir. Diğer taraftan bu kodlama türü birçok problem için doğal değildir ve bazı durumlarda genetik operasyonlar bittikten sonra bazı düzeltmelerin yapılması gerekebilmektedir. İkili (binary) kodlama aşağıdaki biçimde gösterilir¹⁵⁸

Tablo 3. 4 Genetik Algoritmalar İkili Kodlama

Kromozom 1	1 1 0 1 0 0 0 1 1 0 1 0
Kromozom 2	0 1 1 1 1 1 1 1 1 0 0

Bu kodlamada

- Tam sayılar tam olarak (exactly) gösterilir.
- Sonlu sayıda reel sayı gösterilebilir.
- Gerçek sayıların numaralarının sunumu dizi uzunluğunu artırır.¹⁵⁹

İkili düzende kodlama çok sık kullanılan bir kodlama tipi olmasına rağmen bazı sakıncaları vardır. Örneğin, çok değişkenli bir fonksiyon optimizasyonu için değişkenlerin alt ve üst sınırlarına bağlı olarak elde edilen dizi çok uzun olabilir. Aynı zamanda gezgin satıcı, çizelgeleme, karesel atama gibi kombinatoriyal optimizasyon problemlerinde ikili düzende kodlama arama uzayını tam olarak temsil edememektedir.¹⁶⁰

¹⁵⁶ Back ve diğer., 2000b, s.5.

¹⁵⁷ Haupt ve Haupt, 2004, s; 107, Şen- 2004, s.101.

¹⁵⁸ Sivanandam ve Deepa, 2008, s.44.

¹⁵⁹ Sivanandam ve Deepa, 2008, s.43.

¹⁶⁰ Dengiz ve Altıparmak, 1998, s.4.

Sekizli (octal) Kodlama

Sekize kadar olan sayılardan (0-7) oluşturulan dizileri kullanan kodlama türüdür. Kullanım biçimi aşağıda gösterilmiştir.¹⁶¹

Tablo 3. 5 Genetik Algoritmalar Sekizli(Octal) Kodlama

Kromozom 1	0 3 4 6 7 2 1 6
Kromozom 2	1 5 7 2 3 3 1 4

Hekzadesimal Kodlama

Bu kodlama biçiminde de 10'a kadar olan rakam ve ilk altı harften (0-9, A-F) oluşturulan diziler kullanılır. Gösterim biçimi aşağıdaki gibidir.¹⁶²

Tablo 3. 6 Genetik Algoritmalar Hekzadesimal Kodlama

Kromozom 1	9 C E 7
Kromozom 2	3 D B A

Permütasyon Kodlama (Gerçek Sayı Kodlama)

Permütasyon kodlamada diziler ard arda sıralanan rakamlardan oluşmaktadır. Bazı durumlarda genetik operasyonlar tamamlandıktan sonra bazı düzeltmelerin yapılması gerekmektedir. Permütasyon kodlamada her bir kromozom rakamların ard arda sıralandığı, tamsayılı gerçek değerlerin bir dizisidir. Gösterim biçimi aşağıdaki gibidir.

Tablo 3. 7 Genetik Algoritmalar Permutasyon Kodlama

Kromozom 1	1 5 3 2 6 4 7 9 8
Kromozom 2	8 5 6 7 2 3 1 4 9

¹⁶¹ Sivanandam ve Deepa, 2008, s.44.

¹⁶² Sivanandam ve Deepa, 2008, s.44.

Permütasyon kodlama yalnızca problemleri sıralamada yarar sağlar. Bu durumda bile kromozom uyumluluğundan kaçınmak için bazı çaprazlama ve mutasyon düzeltmelerinin yapılması zorunlu olmaktadır.¹⁶³ Bu kodlama fonksiyon optimizasyon problemlerinin çözümünde ikili ve gri kodlamaya göre daha etkin olmaktadır¹⁶⁴

Değer Kodlama

Değer kodlamada her bir kromozom bir değerler dizisinden oluşu ve bu değerler de problem ile bağlantılı rakamlar, reel sayılar veya probleme özgü bazı özel karakterler vb. herhangi bir biçim olabilir. Bazı özel problemler için bu kodlama türü en iyi sonuçları ortaya koymaktadır. Diğer açıdan bakıldığında ise sıklıkla probleme özgü genetik operatörler geliştirilmesini gerekli kılmaktadır. Bu kodlama reel sayılar gibi karmaşık verilerin kullanıldığı problemler için kullanılabilir. Bu tür problemlerde ikili kodlama kullanımı oldukça zordur. Gösterim biçimi aşağıdaki gibidir¹⁶⁵

Tablo 3. 8 Genetik Algoritmalar Değer Kodlama

Kromozom A	1.2324 5.3243 0.4556 2.3293 2.4545
Kromozom B	ABDJEIFJDHDIERJFDLDFLFEGT
Kromozom C	(geri), (geri), (sağ), (ileri), (sol)

Ağaç Kodlama

Bu kodlama biçimi çoğunlukla genetik programlama için program ifadelerini dönüştürmede kullanılır. Burada her bir kromozom programlama dilinin komutları veya fonksiyonları gibi bazı objelerden ağacı şeklindedir.¹⁶⁶ Ağaç kodlamada her kromozom, adından da anlaşıldığı gibi nesneler ve nesneler arası işlemleri içeren bir ağaç yapısından oluşmaktadır. Ağaç kodlama, program geliştirmek için uygundur.

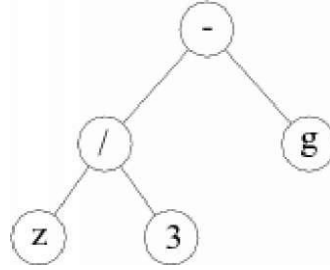
¹⁶³ Sivanandam ve Deepa, 2008, s.45.

¹⁶⁴ Gen ve Cheng, 2000, s.3.

¹⁶⁵ Sivanandam ve Deepa, 2008, s.45.

¹⁶⁶ Sivanandam ve Deepa-2008, s.45.

Örneğin, LISP ve Prolog gibi programlama dillerinde ağaçtan yapısal olarak sık bir biçimde kullanılmaktadır. Ağaç kodlamada çaprazlama ve mutasyon çok kolay bir şekilde uygulanabilir. Gösterim biçimi aşağıdaki gibidir.¹⁶⁷



Şekil 3. 8 Genetik Algoritmalar Ağaç Kodlama

Kaynak: David A. Huffman, 1952

3.2.1.2.3 Genetik Algoritma Operatörleri

Genetik algoritmalarda bir sonraki nesli elde etmek için üç ana genetik operatör kullanılır: Yeniden üretme (reproduction), çaprazlama (crossover) ve mutasyon (mutation).¹⁶⁸ Genetik algoritma uygulamalarında yapılan işlemlerin %90-95'i seçim ve çaprazlama operatörleri ile ilişkilidir. Mutasyon için genellikle düşük bir olasılık değeri tercih edilmektedir.¹⁶⁹

Yeniden Üretme/Seçim

Genetik algoritma uygulamalarında daha yüksek uygunluğa sahip bireylerin daha fazla çoğalacağı temeline dayalı doğal seleksiyon fikri yeniden üretme veya seleksiyon/seçim olarak adlandırılır. Yeniden üretim veya seçim operatörü yeni bir birey yaratmak için bireylerin popülasyondan nasıl seçileceği ve bunlardan hangi sayıda yeni birey oluşturulacağı konusyla ilgilidir.¹⁷⁰

¹⁶⁷ Nabiyev, 2003, s.610.

¹⁶⁸ Rao, 1996, s;810, Sakawa, 2002, s.13.

¹⁶⁹ Banzhaf ve diğer., 1998, s.95.

¹⁷⁰ Rao, 1996, s;810, Dengiz ve Altıparmak, 1998, s;4, Sakawa, 2002, s.19.

Üreme prosesi genetik algoritmaların kalbidir. Bu süreçte arama prosesi ile yeni ve muhtemelen daha uygun bireyler ortaya konulmaktadır. Genetik algoritmalarda üreme prosesi üç aşamadan oluşur.¹⁷¹

- Ailelerin seçimi
- Yeni bireyler (çocuk veya döl) oluşturulması için ailelerin çiftleştirilmesi ve dönüştürülmesi
- Popülasyondaki eski bireylerin yenileri ile değiştirilmesi

Üreme prosesinin ilk adımı olan seçim, eşleştirme içim popülasyondan iki aile belirleme işlemidir. Kodlama biçimine karar verdikten sonraki aşama seçimin nasıl yapılacağıının belirlenmesidir. Bu seçim, gelecek nesil için yeni bireyler oluşturacak bireylerin popülasyondan nasıl seçileceği; her birinden kaç döl elde edileceği gibi konuları kapsar. Seçimin amacı onlardan daha uygun dölleri elde etmek umudu ile popülasyondaki nispeten daha uygun bireylerin belirlenerek öne çıkarılmasıdır. Kromozomlar yeniden üretim için aile oluşturmak üzere başlangıç popülasyonundan seçilirler. Bu noktada problem kromozomların nasıl seçileceğidir¹⁷²

Seçim, değerlendirme fonksiyonu değerlerine göre popülasyondaki kromozomların seçilip ayrılması metodudur.¹⁷³ Daha yüksek uygunluk değeri birey için daha fazla seçilme şansı anlamına gelmektedir. Seçim baskısı, daha iyi bireylerin seçildiği bir seviyeyi ifade eder. Daha yüksek seçim baskısında daha da iyi bireyler belirlenir. Seçim baskısı genetik algoritmayı daha iyi nesiller için popülasyon uygunluğunu artırmaya zorlar.

Genetik algoritmanın yaklaşma oranı büyük ölçüde seçim baskısının büyüklüğü tarafından belirlenir. Daha yüksek seçim baskısı daha yüksek yaklaşma oranı sonucunu doğurur. Genetik algoritmaların akılcı bir seçim baskısı aralığında optimal veya optimale yakın çözümler ortaya koyması beklenir. Düşük seçilen seçim baskısı ise daha yavaş yakınsama ve daha uzun çözüm süresi demektir. Seçim baskısı

¹⁷¹ Sivanandam ve Deepa, 2008, s.46.

¹⁷² Back ve diğer., 2000a, s.56.

¹⁷³ Sivanandam ve Deepa, 2008, s.46.

çok yüksek ise genetik algoritmadaki artan değişim yanlış bir çözüme erken yakınsamaya yol açar ¹⁷⁴ Genel olarak uygulamalarda yüksek seçim baskısı problemleri ile karşılaşmaktadır. ¹⁷⁵ Seçim metotları karşılaştırıldığında oransal seçimden sıralama ve turnuva seçimlerine geçerken seçim baskısı da artmaktadır. Sonuç olarak belirlenen seçim baskı şemasının popülasyon çeşitliliğini koruması ve erken (premature) yakınsamadan kaçınması gerekmektedir. ¹⁷⁶

Genel olarak iki tip seçim şeması tanımlanabilir: oransal seçim ve sıra tabanlı seçim. Oran tabanlı seçim, bireyleri, popülasyondaki diğer bireylerin uygunlukları ile göreceli olan uygunluklarına göre seçer. Sıra tabanlı seçim şeması ise bireyleri onların ham uygunluk değerlerine göre değil bu değerlerin popülasyon içerisindeki sırasına göre seçer. Bu durum seçim baskısının popülasyonun uygunluk dağılımından bağımsız olmasını ve yalnızca popülasyonun nispi sıralamasına dayalı olmasını gerektirir.

Seçim baskısını adapte edebilmek için popülasyonun uygunluk alanını yaymada bir ölçekleme fonksiyonu kullanılabilir. Örneğin tüm çözümlerin uygunluk aralığı [999,1000] ise oransal metot kullanılarak daha iyi bireyler elde etme olasılığı önemsiz sayılabilir. Eğer her bir bireyin uygunluğu [0,1]'e eşit bir aralığa getirilirse kötü birey yerine iyi birey seçme olasılığı önemli olacaktır. Genetik algoritmalarda kullanılan çeşitli seçim metotları aşağıda sıralanmıştır ¹⁷⁷

Rulet Tekerleği Seçimi

Rulet seçiminin ana prensibi, bireylerin uygunluk değerlerine oranlı olarak ağırlıklandırılmış dilimlerden oluşan bir rulet tekerleği ile doğrusal bir arama yapılmasıdır. Hedef değer, popülasyondaki uygunluklar toplamının olasılıklı (random) oranından oluşan bir settir. Hedef değere ulaşıncaya kadar popülasyon

¹⁷⁴ Chambers, 1995, s;72, Sivanandam ve Deepa, 2008, s;46, Back ve diğer., 2000a, s.56.

¹⁷⁵ Olariu ve Zomaya, 2006, s. 13-196.

¹⁷⁶ Back, 1994, s.62.

¹⁷⁷ Sivanandam ve Deepa, 2008, s.47.

ilerletilir. Bu metot ortalama güçlü bir seçim tekniğidir, çünkü uygun bireylerin seçilme garantisi yoktur ancak seçilme olasılıkları yüksektir.

Bir bireyin beklenen değeri popülasyonun gerçek uygunluğuna bölünerek bulunan uygunluk değeridir. Her bir bireye oransal uygunluğu ile orantılı olarak rulet tekerleğinde bir dilim ayrılır. Tekerlek N defa döndürülür. Burada N popülasyondaki birey sayısıdır. Her bir turda tekerleğin işaretçisi altında kalan birey gelecek nesil için eşleşme havuzuna seçilir.

Bu metot aşağıdaki biçimde uygulanır ¹⁷⁸

- Popülasyondaki bireylerin toplam beklenen değerlerini topla. Bu değer T olsun
- N defa tekrar et:
- 0 ve T arasında rastgele bir 'r' sayısı seç.
- Popülasyondaki bireyleri döndürmeye başla. Toplam 'r' ye eşit veya onu geçene kadar beklenen değerleri topla. Beklenen değeri bu limiti geçen bireyi seç.

Rulet tekerleği seçiminin uygulaması kolaydır. Evrim oranı popülasyondaki uygunluğun varyansına bağlıdır.

Sıralama (Rank) Seçimi

Uygunluk değerlerinin çok fazla farklılık sergilediği durumda rulet tekerleği seçimini uygulamak zor olmaktadır. Eğer en iyi kromozomun uygunluğu %90 ise tekerlekte de %90 yer kaplayacaktır ve bu durumda diğer kromozomların seçilme şansı çok az olmaktadır.

¹⁷⁸ Sivanandam ve Deepa, 2008, s.48.

Sıralama seçiminde seçilme olasılığı mevcut popülasyondaki bireylerin uygunluğunun sıralamadaki yerine göre belirlenmektedir. Sıralama seçimi popülasyonu en iyiden en kötü bireye doğru sıralar. Her bir bireyin beklenen değeri onun mutlak uygunluk değeri ile değil sıralamadaki yerine göre belirlenmektedir. Sıralamadaki yerlerine göre bireyler ayrı bir uygunluk değeri alır. Uygunluklar arasındaki mutlak farklılıklar dikkate alınmadığı için sıralama seçiminde uygunlukları derecelendirme gereksinimi de bulunmamaktadır.¹⁷⁹

Bu seçim yakınsamayı yavaşlatır ancak çok hızlı yakınsamayı da önler. Uygunluk varyansının düşük olduğunda da seçim baskısını yüksek tutar. Çeşitliliği korur ve daha başarılı arama sağlar. Uygulamada potansiyel aileler seçilir ve hangi ailenin bir sonraki nesil seçileceğine yönelik bir turnuva düzenlenir.¹⁸⁰

Sıralama, amaç fonksiyonun tam olarak tanımlanamadığı problemler için doğal bir seçenek olabilir. Bu gibi problemlerde tüm dikkati amaç fonksiyonunu alacağı değere atfetmek mantıklı değildir çünkü bazı durumlarda bu durum mümkün de olmayabilir¹⁸¹

Farklı şekillerde yapılabilir olmakla birlikte iki yöntem aşağıda sıralanmaktadır¹⁸² Birey çiftlerini rastgele seç. 0 ile 1 arasında 'R' rassal sayısını üret. Eğer $R < r$ ise ilk bireyi aile olarak seç. $R \geq r$ ise ikinci bireyi seç. Bu ikinci ailenin seçimi için de aynen tekrarlanır. 'r' değeri bu metot için bir parametredir.

Rastgele iki farklı birey seç. Daha yüksek değerlendirme değeri olan bireyi aile için seç. İkinci aileyi bulmak için de bu adımı tekrarla.

¹⁷⁹ Sakawa, 2002, s.21.

¹⁸⁰ Mitchell, 1999, s;127, Sivanandam ve Deepa, 2008, s.47.

¹⁸¹ Back ve diğer., 2000a, s.188.

¹⁸² Sivanandam ve Deepa, 2008, s.47.

Turnuva Seçimi

Genetik algoritma arama performansının en iyi olması için seçim baskısı ve popülasyonun çeşitliliğinin ayarlanabildiği bir seçim yöntemi idealdir. Turnuva seçimi seçim baskısı açısından sıralama seçimi ile benzerdir ancak ona göre hesaplaması daha etkin ve paralel uygulamalar için daha uygundur ¹⁸³

Rulet tekerleri seçiminden farklı olarak turnuva seçimi N_u birey arasında bir turnuva düzenleyerek seçim baskısı sağlar. Kazanan birey uygunluk değerine göre belirlenir. En yüksek uygunluk değerine sahip birey genellikle deterministik olarak seçilir. Turnuva, eşleşme havuzu yeni nesil oluşturmak için doluncaya kadar tekrarlanır. Genelde turnuva iki birey arasında düzenlenir ancak keyfi olarak seçilen bir sayıdaki birey arasında da uygulanabilir. Turnuva gruplardan bağımsız olarak düzenlenir ancak grup sayısı arttıkça seçim baskısı da artacaktır. Uygulamada 6-10 grup önerilmektedir (Back ve diğer., 2000a, s; 181, Olariu ve Zomaya, 2006, s;13-196). Turnuva kazananlardan oluşan eşleşme havuzu yüksek bir ortalama popülasyon uygunluğuna sahiptir. Bu uygunluk farkı seçim baskısı sağlar ve bu durum da, genetik algoritmayı başarılı genlerin uygunluğunu artırmaya zorlar (Sivanandam ve Deepa, 2008, s;49).

Boltzmann Seçimi

Boltzmann seçimi tavlama benzetimi (simulated annealing) prensiplerini kullanarak ve termodinamik olarak seçim baskısını kontrol eden seçim mekanizmasıdır. Tavlama benzetimi algoritması bir yerel arama algoritmasıdır ve pek çok değişkene sahip fonksiyonların en büyük veya en küçük değerlerinin bulunması ve özellikle pek çok yerel en küçük değere sahip doğrusal olmayan fonksiyonların en küçük değerlerinin bulunması için tasarlanmıştır ¹⁸⁴Boltzmann seçiminde önceden belirlenmiş tabloya göre sürekli değişen sıcaklık seçim oranını

¹⁸³ Mitchell, 1999, s.127.

¹⁸⁴ Lai ve Chan, 1997, s.115.

kontrol eder. Sıcaklık düşük seçim baskısı anlamına gelecek biçimde yüksek olarak tutulur. Daha sonra seçim baskısı kademeli olarak yükseltilir ve böylelikle uygun bir seviyedeki çeşitlilik korunurken genetik algoritmanın arama uzayını en iyi kısma yakın olacak biçimde küçültmesi sağlanır.¹⁸⁵

Boltzmann seçim mekanizmasının temelinde mevcut i çözümü ve alternatif j çözümü arasındaki rekabet yatar. Burada i 'nin kazanma olasılığı 1 formülü ile hesaplanmaktadır (Back, Fogel ve Zbigniew-2000b, s; 195). Burada T sıcaklığı, f_i ve f_j mevcut ve alternatif çözüm maliyet, amaç fonksiyon değeri veya uygunluk değerini göstermektedir. $T=T_0(1-a)^k$ ve $k=(1+100*g/G)$ dir. g ; mevcut jenerasyon sayısı, G ; maksimum g değeridir. a , $[0,1]$ aralığından, T_0 ise $[5,100]$ aralığından seçilmektedir. T , 0 değerine ulaştığında global çözüme ulaşılmış olmaktadır.¹⁸⁶

Seçim mekanizmasının seçimi diğer yönlerinden bağımsız olarak yapılamaz. Değerlendirme fonksiyonu çok karmaşık ise en fazla artırımsal modeller problemlidir. Karmaşıklık durumunda seçim baskısı azaltılmalıdır. Karmaşıklık yoksa artırımsal modeller yeni iyi bireyler önemli artışlar sağlayabilir. Bireylerin silinmesi de dikkatli olunmasını gerektirmektedir. En kötü bireylerin ortadan kaldırılması büyüme oranını artırmaktadır ancak bu durum daha düşük kontrol demektir.

Seçkinlik (Elitism)

Seçkinlik, evrimsel algoritmalarda bir popülasyonun maksimum uygunluk değerinin bir diğerine geçildiğinde azalmamasını temin eden özelliktir (Back ve diğer., 2000b, s,25.) Çaprazlama ve mutasyonla yeni popülasyon üretildiği zaman, en iyi kromozom kaybedilecektir. Seçkinlik, önce en iyi kromozomu (yada birkaç en iyi kromozomu) yeni popülasyona kopyalar. Geri kalanı diğer yöntemlerle devam ettirilir. Seçkinlik, bulunan en iyi sonucun kaybını önlediğinden GA'nın performansını oldukça hızlı bir şekilde artırır.¹⁸⁷

¹⁸⁵ Mitchell, 1999, s;126, Sivanandam ve Deepa, 2008, s.49.

¹⁸⁶ Sivanandam ve Deepa, 2008, s.49.

¹⁸⁷ Mitchell, 1999, s;126, Nabiyeve, 2003, s.616.

Çaprazlama

Çaprazlama, genetik algoritmaların benzer metotlar arasında en ayırt edici özelliği ve en önemli operatörüdür. Genetik çözümün performansı büyük ölçüde bu operatöre bağlıdır. Çaprazlama, farklı çözümler arasında bilgi değişimini sağlayarak arama uzayının benzer ancak araştırılmamış bölgelerine ulaşmayı sağlayan bir arama operatörüdür.¹⁸⁸ Çaprazlama mevcut popülasyondan iki ebeveyni alıp bunlardan bir birey üretme prosesidir. Bu operatörün asıl işlevi popülasyondaki farklı bireylerden gelen parça bilgileri yada genetik materyali bir araya getirmektir.¹⁸⁹ Çaprazlama ile seçim prosesinden sonra daha iyi bireyler ile popülasyon zenginleştirilir. Yeniden üretme/seçim (reproduction) ile iyi diziler klonlanır ancak yenileri oluşturulmaz. Çaprazlama operatörü daha iyi bireylerin ortaya konulması beklentisi ile eşleşme havuzuna uygulanır.

Çaprazlama operatörleri çaprazlamaya konu olan ebeveyn sayısına göre üç ana kategoriye bölünebilir (Engelbrecht, 2007, s.144.).

- Yeni birey tek bir ebeveynden üretilecek ise aseksüel,
- Bir veya iki yeni birey iki ebeveynden üretilecek ise seksüel,
- İkiden fazla ebeveynden bir veya daha fazla yeni birey üretilmesi durumunda da çoklu-birleşim (muti-recombination)

Bunun dışında kodlama biçimine özgü (ikili-tam sayılı gibi) veya operatöre özgü (floating point gibi) sınıflandırmalara da rastlanılmaktadır.

Çaprazlama, üç adımdan oluşan bir yeniden birleştirme operatörüdür (Sivanandam ve Deepa, 2008, s;50):

- Yeniden üretme operatörü eşleşme için rassal olarak bir çift birey dizisi seçer,
- Dizi boyunca rassal olarak bir çaprazlama kısmı belirlenir.

¹⁸⁸ Gen ve Cheng, 2000, s;8, Kennedy ve diğerleri, 2001, s.157, Dengiz ve Altıparmak, 1998, s.8.

¹⁸⁹ Sakawa, 2002, s.22.

- Son olarak bu kısımdaki değerler karşı dizidekiler ile yer değiştirilir. Bir nesil içerisinde çaprazlamaya konu olacak ebeveyn dizisinin oranı çaprazlama oranı tarafından belirlenir. p_c e $[0,1]$ aralığında değer alacak bu oran çaprazlama operatörünün ne kadar sıklıkla kullanılacağını göstermektedir.¹⁹⁰

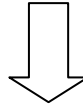
Çeşitli çaprazlama teknikleri aşağıda sıralanmaktadır:

Tek noktalı çaprazlama

Geleneksel olarak genetik algoritmalar tek noktalı çaprazlamayı kullanır. Bazı kaynaklarda bu yöntem 'basit çaprazlama' olarak adlandırılmaktadır. Bu metotta kromozomlar karşılıklı olarak belli bir noktadan kesilir ve kesilen noktadan sonraki kısımlar değiştirilir. Çaprazlama noktası dizi içerisinde rassal olarak belirlenir ve bir sonraki bit'ten başlayarak değişim gerçekleştirilir. Eğer uygun bir nokta seçilmiş ise iyi ebeveynleri eşleyerek daha iyi çocuklar elde edilebilir. Aksi durumda ise dizinin kalitesi oldukça kötü etkilenecektir.¹⁹¹ Tek noktalı çaprazlama aşağıdaki örnekte gösterilmiştir.

Tablo 3. 9 Genetik Algoritmalar Tek Noktalı Çaprazlama

Ebeveyn 1	1 0 1 1 0	0 1 0
Ebeveyn 2	1 0 1 0 1	1 1 1



Çocuk 1	1 0 1 1 0	1 1 1
Çocuk 2	1 0 1 0 1	0 1 0

¹⁹⁰ Sakawa, 2002, s.22.

¹⁹¹ Sakawa, 2002, s.24. , Sivanandam ve Deepa, 2008, s.51.

Tek noktalı çaprazlama konumsal eğiliminden dolayı kusurlu olabilmektedir. Bit indisi /'ye (uzunluk) doğru artarken değişim olasılığı da bire yaklaştığı için değişim olasılığının bit pozisyonuna büyük ölçüde bağlı olduğu söylenebilir.¹⁹²

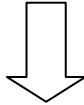
Çift noktalı çaprazlama

Birden fazla kesim noktası içeren çok farklı çaprazlama algoritmaları geliştirilmiştir. Çaprazlama noktası eklenmesi genetik algoritmanın performansını düşürür. İlave çaprazlama noktaları eklendiğinde ortaya çıkan problem ebeveynler arasında değiştirilecek blokların birbirine çok fazla benzemesidir. Bununla birlikte daha fazla çaprazlama noktasına sahip olmanın problem yüzeyini tam tarama gibi bir avantajı vardır.

İki noktalı çaprazlamada iki farklı çaprazlama noktası seçilir ve eşleşen iki ebeveynin bu iki nokta arasında kalan kısımları birbiri ile değiştirilir

Tablo 3. 10 Genetik Algoritmalar Çift Noktalı Çaprazlama

Ebeveyn 1	101	110	10
Ebeveyn 2	011	011	00



Çocuk 1	101	011	10
Çocuk 2	011	110	00

Tek noktalı çaprazlamada dizinin baş ve kuyruk ya da son kısmı birlikte döl'e kopyalanamaz. Hem baş hem de kuyruk kısmı iyi genetik bilgi içeriyor ise bu bilgi tek noktalı çaprazlama ile döl'e aktarılamaz. İki noktalı çaprazlama bu problemi ortadan kaldırır ve genellikle tek noktalı çaprazlamaya göre daha iyi sonuçlar

¹⁹² Back, 1996, s.114.

verir.¹⁹³Çaprazlamada kullanılacak nokta sayısı üzerinde yapılan çalışmalar sonucunda standart olarak iki noktalı çaprazlama kullanılmaktadır.¹⁹⁴

Kromozom üzerinde birbirine yakın olan genlerin çok noktalı bir çaprazlamada döllere birlikte geçmesi olasılığı daha fazladır. Bu durum da yan yana olan genler arasında istenmeyen bir bağlantı oluşturur. Sonuç olarak N noktalı bir çaprazlamanın etkinliği genlerin kromozomdaki yerlerine bağlıdır. Genetik bir sunumda çözümün birbirine bağımlı karakteristiklerini kodlayan genler birbirine yakın olmalıdır. Genin yerleşimi ile ilgili tüm bu problemlerden kaçınılabilmenin yolu uniform çaprazlama kullanılmasıdır¹⁹⁵

Çok noktalı çaprazlama

Çok noktalı çaprazlamada iki seçenek söz konusudur. Biri çift sayılı, diğeri ise tek sayılı kısım çaprazlamadır. Çift sayılı seçilmesi durumunda çaprazlanacak kısımlar bir daire etrafında rassal olarak seçilir ve bilgiler değiştirilir. Tek sayılı melez alan seçildiğinde dizinin başında farklı bir çaprazlama noktası farz edilir.

Çaprazlama noktası sayısı tek olduğunda değişim olasılığı asimetric biçimde bit pozisyonuna bağlı olarak değişir. Bir çok durumda simetri nedeni ile küçük çift sayılı çaprazlama nokta sayısı belirlenmesi önerilmektedir.¹⁹⁶

Uniform çaprazlama

Uniform çaprazlama n noktalı çaprazlamadan oldukça farklıdır. Döl üzerindeki her bir gen, kromozomlarla aynı uzunlukta ve tesadüfi oluşturulan ikili bir çaprazlama maskesine göre seçilen ebeveynden gelir. Çaprazlama maskesinde 1 olduğu yerde gen ilk ebeveynden, 0 olduğu yerde ise ikinci ebeveynden kopyalanır. Her bir ebeveyn çifti için yeni bir çaprazlama maskesi rassal olarak oluşturulur.

¹⁹³ Sivanandam ve Deepa, 2008, s.51.

¹⁹⁴ Back, 1996, s.115.

¹⁹⁵ Sivanandam ve Deepa, 2008, s.52.

¹⁹⁶ Back, 1996, s.114.

Bundan dolayı, döllər her bir ebeveynəndən gələn bir gen karışımını taşır. Uniform çaprazlama biçimi aşağıdaki şekilde gösterilmektedir. Örnekte ilk çocuk üretiminde maskede 1 olduğunda gen birinci ebeveynəndən seçilmiş, ikinci çocukta ise tersi uygulanmıştır.¹⁹⁷

Tablo 3. 11 Genetik Algoritmalar Uniform Çaprazlama

Ebeveyn 1	1 0 1 1 0 0 1 1
Ebeveyn 2	0 0 0 1 1 0 1 0
Maske	1 1 0 1 0 1 1 0
Çocuk 1	1 0 0 1 1 0 1 0
Çocuk 2	0 0 1 1 0 0 1 1

Üç ebeveynli çaprazlama

Bu çaprazlama tekniğinde tesadüfi olarak üç farklı ebeveyn seçilir. İlk ebeveyndeki her bir bit ikinci ile karşılaştırılır. Eğer her ikisi aynı ise bu gen döle için seçilir; eğer farklı ise üçüncü ebeveynin geni seçilir.

Karıştırmalı (shuffle) çaprazlama

Karıştırmalı çaprazlamada ilk olarak tek noktalı çaprazlamada olduğu gibi tek bir çaprazlama noktası seçilir. Ancak değişkenler karşılıklı olarak değiştirilmeden önce her iki ebeveynde rakamlar tesadüfi olarak karıştırılır. Daha sonra değişim gerçekleştirilir. Her çaprazlama yapıldığında değişkenler random olarak yeniden atandığı için bu metotla konumsal eğilimler engellenmiş olur (Sivanandam ve Deepa, 2008, s;52). Bu teknik herhangi birçok noktalı çaprazlama ile birlikte kullanılabilir. Standart operatörlerin pozisyona bağlı eğilimlerini azaltabilmek için her iki ebeveynin aynı biçimde bit pozisyonları karıştırılır ve daha sonra çok noktalı operatör uygulanır. Son aşamada ise karıştırılan bitler eski yerlerine yerleştirilir.¹⁹⁸

¹⁹⁷ Sakawa, 2002, s;24, Sivanandam ve Deepa, 2008, s.52.

¹⁹⁸ Back, 1996, s.116.

Yukarıda sıralananların dışında literatürde çok sayıda çaprazlama metodu geliştirilmiştir. Bunlardan en fazla kullanılan diğerleri Üstünlük korumalı çaprazlama, Ordered çaprazlama, ve kısmi eşlenmiş çaprazlamadır.

Mutasyon

Çaprazlama mevcut gen potansiyellerini araştırmak üzere kullanılır. Fakat popülasyon gerekli tüm kodlanmış bilgiyi içermez ise, çaprazlama tatmin edici bir çözüm üretmez. Bu durum, mevcut kromozomlardan yeni kromozomlar üretme yeteneğine sahip bir operatör gerekmektedir. Bu görevi mutasyon operatörü ile gerçekleştirilir.¹⁹⁹ Mutasyonun amacı mevcut bir bireyden yeni genetik materyaller ortaya koymak ve popülasyonun karakteristiğine çeşitlilik kazandırmaktır. Mutasyon, her bir gen üzerindeki tüm alellere ulaşılabilmesini sağlamak üzere çaprazlamaya yardımcı olarak kullanılır.²⁰⁰

Mutasyon gendeki bir alel'in tesadüfi olarak bir diğeri ile yer değiştirmesi ile uygulanır. Başka bir ifadeyle önceki potansiyel çözümün sunumunda yapılan küçük tesadüfi değişiklikler ile yeni bir potansiyel çözüm elde edilmektedir.²⁰¹ Genetik algoritmaların çalışmasında ikinci dereceden rol oynayan mutasyon, küçük bir olasılıkla bir dizideki bir veya birkaç değeri rassal olarak değiştirerek yeni dizilerin yada arama uzayında yeni çözüm noktalarının elde edilmesine çalışır²⁰²

Mutasyon popülasyonu lokal minimuma takılmaktan uzak tutar. Tesadüfi işlemlerle kaybedilen genetik bilgiler gibi kayıp genetik materyalin geri kazanılmasında rol alır. Aynı zamanda genetik materyalin geri dönülemez kaybına karşı da bir sigorta işlevi görür.

¹⁹⁹ Emel ve Taşkın, 2002, s.134.

²⁰⁰ Engerbrecht, 2007, s.153.

²⁰¹ Back ve diğer., 2000a, s. 6.

²⁰² Dengiz ve Altıparmak, 1998, s.6.

Mutasyon geleneksel olarak basit bir arama operatörü olarak düşünülmüştür. Eğer çaprazlamanın daha iyi çözüme ulaşmak için mevcut çözümü kullandığı düşünülürse; mutasyonun da tüm arama uzayının keşfi için yardımcı olacağı düşünülür. Mutasyon popülasyondaki çeşitliliği koruyan temel operatördür.

Farklı sunumlar için çok farklı mutasyon biçimleri mevcuttur. İkili kodlamada basit bir mutasyon, küçük bir olasılıkla her bir genin değerini tersine çevirmekle yapılabilir. Burada bahsedilen olasılık değeri, L kromozom uzunluğu olmak üzere, $1/L$ olarak alınır. Yalnızca çözümün kalitesini artırdığı durumda mutasyon yapılan tepe- tırmanma (hill-climbing) mutasyon operatörlerinin de kullanılması mümkündür. Bu gibi bir operatör aramayı hızlandırabilir ancak popülasyondaki çeşitliliği azaltması ve algoritmayı bazı yerel optimumlara yakınlaştırması konularında da dikkatli olunmalıdır.²⁰³

Döndürme (flipping)

Bir bit'i döndürme, oluşturulan bir mutasyon kromozomuna göre 0'ı 1'e ve 1'i 0'a çevirme işlemidir. İlk olarak bir ebeveyn seçilir ve tesadüfi olarak bir mutasyon kromozomu oluşturulur. Mutasyon kromozomundaki bir 1'e tekabül eden ebeveyn üzerindeki bit '0' ise '1'e; '1' ise 0'a çevrilerek yeni birey oluşturulur. Mutasyon kromozomunda 0 olan bitlere tekabül eden ebeveyn bitleri yeni bireye dönüştürülmeden aynen kopyalanır. Şekil 3.x te bu işleyiş gösterilmiştir.²⁰⁴

Tablo 3. 12 Genetik Algoritmalar Döndürme (Flipping)

Ebeveyn	1 0 1 1 0 1 0 1
Mutasyon Kromozumu	1 0 0 0 1 0 0 1
Çocuk	0 0 1 1 1 1 0 0

²⁰³ Sivanandam ve Deepa, 2008, s;56

²⁰⁴ Sivanandam ve Deepa, 2008, s;57

Değiş-tokuş

Dizi üzerinde tesadüfi olarak iki farklı nokta belirlenir ve bu noktalardaki bitler değiş tokuş edilerek yeni nesil kromozom üretilir. Şekil 3.x te bu işleyiş gösterilmiştir.²⁰⁵

Tablo 3. 13 Genetik Algoritmalar Değiş Tokuş

Ebeveyn	1 0 1 1 0 1 0 1
Çocuk	1 1 1 1 0 0 0 1

Gaussian mutasyon

Her bir kromozom için Poison dağılımına göre mutasyona uğratılacak gen belirlenir. Seçilen bu genler dönüştürülür.

Yenileme/Yer Değiştirme (Replacement)

Yenileme üreme döngüsünün son aşamasıdır. Sabit büyüklükteki popülasyondan iki ebeveyn seçilir ve bunlar da iki yavru yeni birey üretir; ancak bu dört bireyin dördü de popülasyona geri dönmez. Dolayısıyla ikisinin değiştirilmesi gerekmektedir. Hangi bireyin popülasyonda kalacağı ve hangisinin değiştirileceğine karar vermede kullanılan teknik, seçim kadar yakınsama sürecinde etkilidir. Temelde popülasyonun devamı için iki farklı metot vardır: kuşaksal güncelleme ve sabit durum güncellemesi.²⁰⁶

Temel kuşaksal güncelleme şeması, gelecek nesli oluşturmak için N büyüklüğündeki bir popülasyondan yine N sayıda çocuk üretimini ve bu yeni çocuk popülasyonunun tamamen ebeveynleri ile yer değiştirmesini kapsamaktadır. Bu metotta bir birey yalnızca aynı nesilden başka bir birey ile yeni bir birey oluşturabilir.

²⁰⁵ Sivanandam ve Deepa, 2008, s.57.

²⁰⁶ Sarac, O.S., Alan, O., Bahceci, E., Leblebicioğlu, K., “**Genetik Algoritmayla Çoklu Dizi Hizalama**”, Sinyal İşleme ve Uygulama Kurultayı, 2003.

Kuşaksal güncellemenin $(X+p)$ güncellemesi ve (X,p) güncellemesi gibi türetilmiş biçimleri de kullanılmaktadır. Popülasyon büyüklüğü p , ve üretilen yeni birey sayısı da p 'dür. Ya oluşturulan yeni nesil popülasyonundan yada bütünleştirilen ebeveyn ve çocuk popülasyonundan gelen p sayıda en iyi birey gelecek nesli oluşturur.

Sabit durum güncellemesinde ise yeni bireyler, kuşaksal güncellemenin tersine, oluşturulur oluşturulmaz popülasyona dahil edilir. Yeni birey eklenirken popülasyondaki ya en kötü birey ya da en eski birey silinir. Sabit durum güncellemesi metodu hem seçim hem de güncelleştirmede turnuva metodu gibi sıra temelli metotları kullanır. Turnuva yenilemesi turnuva seçimine oldukça benzerdir.

207

Tesadüfi yenileme (Random replacement)

Bu yöntemde çocuk yada yeni bireyler popülasyonda tesadüfi olarak seçilen bireylerle yer değiştirir. Ebeveynler de bu seçimde aday konumundadır. Küçük popülasyonlarda aramayı sürdürmek için yararlı olabilir çünkü popülasyona zayıf bireyler girebilir.

Zayıf ebeveyn yenileme

Bu yenileme/yer değiştirme metodunda zayıf ebeveyn güçlü çocuk ile değiştirilir. Üremeden sonra oluşan dört bireyden yalnızca en uygun ikisi popülasyona geri dönebilir. Bu süreç, çaprazlama için hem uygun hem de zayıf ebeveynleri kullanan bir seçim tekniği ile birlikte kullanıldığında popülasyonun genel uygunluğunu artırır.

Her iki ebeveyni yenileme

²⁰⁷ Sivanandam ve Deepa, 2008, s.58.

Bu yöntemde çocuklar ebeveynlerinin yerine geçer. Her birey yalnızca bir kez üremeye katılır. Popülasyon ve genetik materyal sürekli değişir. Ancak ancak uygun ebeveynleri öne çıkaran seçim yöntemleri ile düşünüldüğünde problemler yaşanmaktadır.

Yakınsama Kriterleri (Arama Sonlandırma)

Yerel bir optimuma ulaşıldığında basit yakın komşuluk arama metotlarından farklı olarak genetik algoritma prensip olarak sonsuza kadar çalışır. Bunun için pratikte bir sonlandırmam kriteri gerekmektedir. Çeşitli durdurma koşulları aşağıda sıralanmıştır²⁰⁸

Maksimum Nesil	Tanımlanan sayıda nesil üretildiği durumda genetik algoritma durdurulur.
Geçen Zaman	Önceden tanımlanan süre dolduğunda genetik proses sonlandırılır. Nesil sayısı kriteri ile birlikte kullanıldığında önceden tanımlanan nesil sayısına süreden önce ulaşırsa algoritma durur.
Değişmeyen uygunluk	Tanımlanan sayıdaki nesil için popülasyonun en uygun değeri değişmiyor ise genetik proses sonlanır.

Bazı sonlandırma teknikleri aşağıda sıralanmıştır:

En iyi birey ve En kötü birey

Bu kriter popülasyondaki en düşük uygunluğu yakınsama değerinin altına düştüğünde arama durdurulur. Bu metot, en azından bir adet iyi çözümü garanti ederek aramayı daha hızlı sonlandırır.

En kötü birey kriteri kullanıldığında popülasyondaki en az uygun bireyin yakınsama kriterinden daha düşük bir uygunluğa sahip olduğu durumda arama sonlandırılır. Bu durumda tüm popülasyonun belli bir minimum standardın üstünde

²⁰⁸ Reeves ve Rowe, 2002, s;30, Sivanandam ve Deepa, 2008, s.59.

olması sağlanır. Ancak en iyi bireyin en kötünden önemli ölçüde farklı olmayabilir. Bu durumda, aşılabacak belli bir maksimuma göre tanımlanan yüksek bir yakınsama değeri hiçbir zaman sağlanamayabilir.²⁰⁹

Uygunluk Toplamı

Bu sonlandırma şemasında, popülasyondaki uygunluğun toplamının popülasyon kaydındaki (record) yakınsama değerine eşit veya daha düşükse aramanın, yakınsamayı sağladığı düşünülür. Bu durum hemen hemen tüm bireylerin belli bir uygunluk aralığında olmasını garanti eder. Ancak bu yaklaşımın kriterini en zayıf gen değişimi ile birlikte düşünmek daha iyi sonuç verir.²¹⁰ Aksi durumda popülasyondaki yeni birkaç düşük uygunluklu birey toplam uygunluğu bozabilir. Yakınsama değeri belirlenirken popülasyon büyüklüğü dikkate alınmalıdır. Ortak uygunlukta yaklaşımında sonlandırma için bireylerin en azından yarısının belirlenen yakınsama kriterinden eşit ya da daha iyi olması gerekir.

Genetik Algoritmalarda Parametre Seçimi

Genetik algoritma uygulamasının önemli bir adımı da çeşitli parametrelerin belirlenmesidir. Parametreler genetik algoritma performansı üzerinde çok önemli bir etkiye sahiptir. Kontrol parametreleri olarak da adlandırılan bu parametreler popülasyon büyüklüğü, çaprazlama olasılığı, mutasyon olasılığı, kuşak aralığı, seçim stratejisi, fonksiyon ölçeklemesini kapsamaktadır.²¹¹

Parametreler birbirlerini doğrusal olmayan biçimde etkilediği için aynı anda optimize edilemezler. Literatürde parametre seçimi ve adaptasyonu üzerinde oldukça fazla sayıda çalışma yapılmış olmakla birlikte tüm problemlerde kullanılabilecek en iyi parametreler üzerinde bir uzlaşma söz konusu değildir.²¹²

²⁰⁹ Baker B. M. ve Ayechev M. A. (2002), “**A Genetic Algorithm For The Vehicle Routing Problem**”, Computers & Operations Research, s. 1-14.

²¹⁰ Bhattacharyya S. (1999), “**Direct Marketing Performance Modeling Using Genetic Algorithms**”, Journal on Computing, Volume: 11, Issue: 3, s. 248-267

²¹¹ Dengiz ve Altıparmak, 1998, s;9, Mitchell, 1999, s;130, Emel ve Taşkın, 2002, s.135.

²¹² Mitchell, 1999, s.131.

Popülasyon büyüklüğü

Popülasyon büyüklüğü genetik algoritma kullanıcısı tarafından verilen en önemli kararlardan biridir. Bu değer küçük olarak seçildiğinde algoritmanın yerel optimuma takılması olasılığı vardır. Büyük bir değer ise çözüm uzayının daha iyi tarama olasılığına karşı çözüm için gerekli zaman ve kaynakları artırmaktadır.²¹³

Popülasyon büyüklüğü diğer genetik algoritma parametrelerinden tamamen ayrı düşünülemez. Belirlenecek büyüklük, hesaplama etkinliği, yeterli örnekleme sağlayabilme durumu, popülasyonun çeşitliliği, çaprazlama ve seçim baskısı gibi parametrelerin dengelemesine katkı sağlamalıdır²¹⁴

Çaprazlama olasılığı

Çaprazlama operatöründe temel parametre çaprazlama olasılığıdır (P_c). Çaprazlama olasılığı, çaprazlamanın ne kadar sıklıkla yapılacağını tanımlayan bir parametredir. Çaprazlama olmadığı durumda döl, ebeveynlerin direkt kopyası olur. Ancak yeni neslin önceki ile aynı olduğu anlamına gelmez. Çaprazlamanın olduğu durumda döl, ebeveynlerin her ikisinin kromozomlarının parçalarından oluşur. Çaprazlama, yeni kromozomların eskilerinin iyi parçalarını taşıyacağı beklentisi ile yapılır²¹⁵

Çaprazlama olasılığı eşleşme havuzuna girecek kromozomların sayısını belirler. Bu oranın çok yüksek olması durumunda iyi yapı bloklarının tek bir kromozomda birikme olasılığı oldukça azalır. Düşük bir oran belirlenmesi durumunda ise yeterli sayıda yeni çocuk/döl ortaya konulamaz.²¹⁶

²¹³ Emel ve Taşkın, 2002, s;135, Reeves ve Rowe, 2002, s.25.

²¹⁴ Back ve diğer., 2000b, s.134.

²¹⁵ Sivanandam ve Deepa, 2008, s.56.

²¹⁶ Haupt ve Haupt, 2004, s.117.

Mutasyon olasılığı

Mutasyon tekniğindeki önemli parametre, kromozomda ne kadar bir değişiklik olacağını gösteren mutasyon olasılığıdır. Eğer mutasyon yok ise yeni birey çaprazlamadan hemen sonra oluşturulur demektir. Mutasyon olasılığının %100 olması ise kromozomun tamamen değiştirildiği anlamına gelir.²¹⁷ Mutasyonun temel amacı popülasyondaki çeşitliliği korumaktır. Mutasyon olasılığının artması genetik aramayı raslantısal bir aramaya dönüştürür; ancak diğer taraftan bu durum kayıp genetik materyali tekrar bulmada yardımcı olmaktadır²¹⁸

Bazı çalışmalar büyük popülasyonlar için ($p > 200$) büyük mutasyon olasılığı ($P_m > 0.05$) kullanıldığında; benzer şekilde küçük popülasyonlar ($p < 20$) içinse daha küçük mutasyon oranı ($P_m < 0.002$) kullanıldığında performansın arttığını ortaya koymuştur. Orta büyüklükteki bir popülasyon büyüklüğü için $P_m = 0.001$ oranı seçilebilir²¹⁹

Her kuşaktaki yeni kromozom oranına kuşak aralığı denir. Yüksek kuşak aralığı çok sayıda kromozomun değiştiği anlamına gelmektedir²²⁰

Seçim Stratejisi

Eski kuşağı yenilemenin çeşitli yöntemleri mevcuttur. Kuşaksal stratejide, mevcut popülasyondaki kromozomlar tamamen yavrular ile yer değiştirir ancak popülasyonun en iyi kromozomu da yenilendiğinden dolayı bir sonraki kuşağa aktarılamaz. İyi niteliklerin bir sonraki kuşağa aktarımı için bu strateji en uygun (elitist) stratejisiyle beraber kullanılmaktadır. En uygun stratejisinde, popülasyondaki en iyi kromozomlar hiçbir zaman yenilenmemektedir. Sabit durum (steady state)

²¹⁷ Sivanandam ve Deepa, 2008, s.57.

²¹⁸ Emel ve Taşkın, 2002, s.136.

²¹⁹ Back, 1996, s.123.

²²⁰ Mitchell, 1999, s;128, Emel ve Taşkın, 2002, s.136.

stratejisinde ise, her kuşakta yalnızca birkaç kromozom yenilenmektedir. Genellikle, yeni kromozomlar popülasyona katıldığında en kötü kromozomlar yenilenir.²²¹

Doğrusal ölçekleme, üstsel ölçekleme gibi yöntemler mevcuttur. Probleme göre en uygun ölçekleme yönteminin seçilmesi genetik algoritmanın etkin işlemesi açısından önem taşımaktadır.²²²

3.2.1.2.4 GA'nın Geleneksel Optimizasyon Yöntemleri ile Karşılaştırılması

Genetik algoritmaların temel prensibi basittir: genetik ve doğal seleksiyonu bilgisayar yazılımları kullanarak taklit etmek anlamındadır. Problem parametreleri de DNA'daki gibi bir doğrusal veri yapısı, bir vektör veya bir dizi gibi en doğal biçimi ile kodlanır. Genetik algoritmalar geleneksel optimizasyon tekniklerinden aşağıdaki yönlerden farklıdır²²³

- Genetik algoritmalar problem parametrelerinin kendilerinden ziyade kodlanmış versiyonları ile çalışırlar.
- Hemen tüm geleneksel optimizasyon teknikleri tek bir noktadan hareket ederek araştırır ancak Genetik algoritmalar ise tek bir çözüm yerine çözümlerin popülasyonunu kullanır. Bu durum GA'nın global optimuma ulaşma olasılığını artırır ve yerel sabit bir noktaya takılmasından kaçınılmasını sağlar.
- Genetik algoritmalar değerlendirme için türevlerden ziyade uygunluk fonksiyonunu kullanır. Bu fonksiyonlar da sürekli yada kesikli her tür optimizasyon problemine uygulanabilir.
- Geleneksel yöntemler sürekli optimizasyonlara deterministik dönüşüm uygularken genetik algoritmalar olasılıklı dönüşüm uygular.

²²¹ Emel ve Taşkın, 2002, s.136.

²²² Emel ve Taşkın, 2002, s.136.

²²³ Rao, 1996, s; 808, Sakawa, 2002, s;15, Sivanandam ve Deepa, 2008, s.34.

Genetik Algoritmaların geleneksel yaklaşımların başarısız olduğu durumlarda çarpıcı sonuçlar ortaya koymasını sağlayan diğer bazı farkları ve avantajları aşağıdaki gibi sıralanabilir.²²⁴

- Genetik algoritmalar problem parametrelerinin kendilerinden ziyade kodlanmış versiyonları ile çalışırlar.
- Hemen tüm geleneksel optimizasyon teknikleri tek bir noktadan hareket ederek araştırır ancak Genetik algoritmalar ise tek bir çözüm yerine çözümlerin popülasyonunu kullanır. Bu durum GA'nın global optimuma ulaşma olasılığını artırır ve yerel sabit bir noktaya takılmasından kaçınılmasını sağlar.
- Genetik algoritmalar değerlendirme için türevlerden ziyade uygunluk fonksiyonunu kullanır. Bu fonksiyonlar da sürekli yada kesikli her tür optimizasyon problemine uygulanabilir.
- Geleneksel yöntemler sürekli optimizasyonlara deterministik dönüşüm uygularken genetik algoritmalar olasılıklı dönüşüm uygular.
- Sürekli ve kesikli değişkenleri optimize edebilir.
- Türev alma işlemine gereksinim duymaz.
- Çözüm alanında daha geniş örneklemeyi eş zamanlı araştırır. Global optimumu daha kolay belirler. Klasik yöntemlerde ise bu durum tek nokta ve yön olarak karşımıza çıkmaktadır.
- Çok sayıda değişken ile çalışabilir.
- Farklı problemlere kolaylıkla modifiye edilebilir.
- Paralel bilgisayarlarla hesaplamalar için oldukça uygundur.
- Oldukça karmaşık çözüm yüzeylerinde yerel minimumlara takılmadan değişkenleri optimize edebilir.
- Sadece tek bir çözüm yerine optimum değişkenlerin bir listesini verebilir.
- Kodlanmış değişkenler ile optimizasyon sağlayabilmek için değişkenleri

²²⁴ Haupt ve Haupt, 2004, s.23, Şen, 2004, s.62, ve Sivanandam ve Deepa, 2008, s.34, Sumathi ve diğer., 2008, s.122.

kodlayabilir.

- Nümerik olarak oluşturulan verilerle, deneysel verilerle ve analitik fonksiyonlar ile çalışabilir.
- Kendi iç kuralları ile çalıştığı için problemin kurallarını bilmesi gerekmez.
- Problemin doğrusallaştırılmasını gerektirmemektedir.²²⁵

Bununla birlikte Genetik algoritmaların her tür problemin çözümünde en iyi metot olduğu da söylenemez. Yalnızca birkaç değişkeni olan düzgün bir analitik fonksiyon geleneksel yöntemlerle kolayca çözülebilir. Bu gibi problemlerde Genetik Algoritma başlangıç popülasyonu için maliyetleri hesaplarken geleneksel yöntemler hesaplamayı tamamlamış olabilir. Genetik algoritmaların sınırlılıkları ve dezavantajları ise şu şekilde sıralanabilir²²⁶

- Uygunluk fonksiyonunu tanımlama problemi
- Vaktinden önce yakınsama olabilmektedir.
- Popülasyon büyüklüğü, mutasyon oranı, çaprazlama oranı, seçim metodu gibi çeşitli parametrelerin belirlenmesi problem oluşturabilir.
- Probleme özgü bilgiler kolaylıkla dahil edilemez
- Lokal optimumları tanımlamada iyi bir metot değildir.
- İyi bir sonlandırıcı değildir. Harici yardım gerekebilir.
- Kesin global optimumu bulmada sorun yaşayabilir.
- Çok sayıda uygunluk fonksiyonu değerlendirmesine gereksinim duyar.

3.2.1.2.5 Genetik Algoritmaların Uygulama Alanları

Sezgisel bir teknik olarak Genetik algoritmaların Yöneylem Araştırması problemlerinde kullanımı her geçen gün artmaktadır. Genetik algoritmalar temelde

²²⁵ Romao, W., Freitas, A. A., S. Gimenès, I. M., “**Discovering Interesting Knowledge From A Science And Technology Database With A Genetic Algorithm**”, Science Direct, 2004, s. 121-137.

²²⁶ Sivanandam ve Deepa, 2008, s.34.

iki farklı amaç için kullanılmaktadır.²²⁷ Bu amaçlardan ilki bir sistem performansını optimize etmek için parametre seçimidir. Bu gibi sistemler muhtemelen kısıtlar ile birlikte büyük ölçüde sistemi kuran ya da işletenlerin belirlediği karar parametrelerine bağlıdır. Genetik algoritma uygulamalarının çoğunluğu sistemleri optimize etmek veya en azından onları optimuma yakınlaştırmaktır.²²⁸

Genetik algoritmaların ikinci potansiyel ana kullanım alanı ise sayısal modelleri test etmek ve uygunluğunu sağlamaktır. Literatürde karşımıza daha az çıkan bu uygulamalarda öncelikle açıklayıcı ya da tanımlayıcı bir model kurulur, veri toplanır ve model test edilir. Çelişkiler tespit edildiğine model modifiye edilir. Performans optimizasyonunu hedefleyen algoritmaların aksine burada model ile veriler arasındaki uyumsuzluğun minimizasyonu hedeflenir.

Genetik algoritmaların mühendislik problemlerinde optimizasyon amaçlı olarak kullanılmaya başlanmıştır. Özellikle mekanizma tasarımında çok iyi sonuçlar ortaya konulmuştur. Bunların yanında otomatik programlama, öğrenme kabiliyetli makinalar, ekonomi, ekoloji, planlama, üretim hattı yerleşimi gibi çok farklı alanlarda da başarılı uygulamalar yapılmaktadır. Ayrıca dijital resim işleme tekniğinde de çokça uygulama alanı bulmuştur. Bu problemlerin hemen hemen hepsi çok geniş bir çözüm havzasının taranmasını gerektirmektedir. Bu çözüm havzasının geleneksel yöntemlerle taranması çok uzun sürmekte, genetik algoritmayla ise kısa bir sürede kabul edilebilir bir sonuç alınabilmektedir.²²⁹

Uygulama alanlarının bazıları aşağıdaki gibi sıralanabilir.²³⁰

- Birleşik optimizasyon (Combinatorial Optimization) problemleri: Gezgin satıcı problemi, araç rotalama problemi, Çinli postacı problemi, iş atölyesi

²²⁷ Chambers, 2001, s.1-2.

²²⁸ Brown C. E. ve Sumichrast R.T. (2001), “CF-GGA: A Grouping Genetic Algorithm For The Cell Formation Problem”, **International Journal of Production Research**, Volume: 39, Number: 16, s. 3651-3669.

²²⁹ Kurt ve Cumali, 2001, s.1.

²³⁰ Coley, 1999, s.3, Chambers, 2001, s. 1-2, Kurt ve Cumali, 2001, s.8-15, Emel ve Taşkın, 2002, s.138, Güden ve diğer., 2005, s.7, Sivanandam ve Deepa, 2008, s.35

çizelgeleme problemi, atama problemi, yerleşim tasarımı problemi ve sırt çantası problemi gibi problem uygulamaları

- Bilgisayar ağları tasarımı
- Nonlineer dinamik sistem tahmini, veri analizleri
- Robot yörüngesi planlama
- Görüntü işleme
- LISP programlarının geliştirilmesi
- Stratejik planlama
- Protein moleküllerinin şeklinin bulunması
- Gaz boru hattı uygulamaları
- Mekanik öğrenme,
- Yapay sinir ağlarının tasarımı,
- Sinyal işleme, filtre dizaynı
- Portföy optimizasyonu,
- Hisse senedi fiyatlarındaki değişim kalıplarını tahmini
- Montaj hattı dengeleme problemleri
- Taşıma problemi
- İmalat planlama, yerleşim planlama ve kaynak ataması

3.2.1.2.6 Genetik Algoritmaların Sınıflandırılması

Genetik algoritmalar problem çözümünde adım adım (step by step) bir prosedür ortaya koyar. Etkin bir arama tekniği olarak kendini ispatladığı için birçok işletme, fen bilimleri ve mühendislik konularında problem çözümünde kullanılmıştır. Tüm bu özellikleri araştırmacıların genetik algoritmalarla farklı yaklaşımlar geliştirmesine de olanak tanımıştır.²³¹

Farklı genetik algoritma yaklaşımları aşağıda açıklanmaktadır²³²

²³¹ Davis L. ve Coombs S. (1987), “**Genetic Algorithms And Communication Link Speed Design: Theoretical Considerations**”, Grefenstette, s. 252-256

²³² Sivanandam ve Deepa, 2008, s.105, Sumathi ve diğer.2008, s.109.

Basit genetik algoritma

Basit genetik algoritmanın mekanizması dizileri kopyalamak ve belirli kısımlarını değiştirmek dışında karmaşık işlemler içermemektedir. Birçok problemin çözümünde başarılı sonuçlar elde eden bu yaklaşım üç işlemten oluşmaktadır:

- Yeniden birleştirme (reproduction)
- Çaprazlama
- Mutasyon

Yeniden birleştirme aşamasında birey dizileri uygunluk değerlerine göre kopyalanır. En sık kullanılan yöntem rult tekerleği seçimidir. Basit genetik algoritma kullanımının faydalı olduğu durumlar aşağıda sıralanmıştır:

- Arama yüzeyi geniş veya karmaşık ise,
- Alan bilgisi veya uzmanlığın arama uzayını kodlamada yetersiz kaldığı durum mevcut ise,
- Herhangi bir matematiksel analiz mevcut değil ise
- Geleneksel metotlar denemesi başarısız olmuş ise

Basit genetik algoritmanın avantajı keyfi kısıt ve amaçlar ile kolayca kullanılabilmesidir. Bununla birlikte kullanımda bazı dezavantajları mevcuttur:

- Genellikle etkin değildir,
- Değerlendirmeler çok zaman almaktadır,
- Çok sayıda problemde popülasyonun büyük tutulması gerekmektedir.²³³

Paralel genetik algoritma

²³³ Romero, C., Ventura, S., Bra, P., Castro, C., “**Discovering Prediction Rules in AHA Courses**”, Johnstown, PA, USA,2003, s. 25-34.

Çeşitli basit genetik algoritmaların paralel olarak uygulanması 'paralel genetik algoritma' olarak tanımlanmaktadır. Bu yaklaşım basit genetik algoritmaların geniş bir arama yüzeyinde optimale yakın çözümü bulmak için harcadığı sürenin kısaltılmasına yönelik geliştirilmişlerdir. Genetik algoritmaların paralel olarak çalıştırılabilmesi aşağıdaki elementlere bağlıdır:

- Uygunluğun değerlendirme ve mutasyonun uygulanma biçimi,
- Seçimin yerel yada global uygulanması,
- Tek veya çoklu alt popülasyonların kullanılması,
- Çoklu popülasyonlar kullanılıyor ise bireylerin değiştirilme biçimi

Probleme alternatif çözümler ortaya koyabilmesi, farklı noktalardan hareketle paralel arama yapılmasını sağlaması, daha hızlı sonuç vermesi ve diğer arama prosedürleri ile kolayca bütünleştirilmesi bu yaklaşımın avantajları olarak sıralanabilir. Paralel genetik algoritma modelinin önemli bir diğer avantajı geleneksel ya da basit genetik algoritmaların erken yakınlaşmasını zorlaştırmasıdır. Birçok araştırmacı tarafından paralel genetik algoritmanın daha gerçekçi bir model olduğu ileri sürülmektedir.²³⁴

Paralel genetik algoritmalar da uygulama metoduna göre kendi içerisinde bağımsız paralel genetik algoritmalar, göçebe paralel genetik algoritması, bölümlendirme paralel genetik algoritması, segmentasyon paralel genetik algoritması gibi kısımlara ayrılmaktadır.

Hibrit genetik algoritma

Hibrit genetik algoritma çaprazlama tarafından üretilen yeni bireylerin gelişimi için sezgiselliğin kullanımını sağlamaya dönük geliştirilmiştir. Bir sezgisellik başlangıç popülasyonu oluşturmada, diğer ikisi de ya çaprazlama veya

²³⁴ Chen, T., Hsu, T., (2006) , “A GA Based Approach For Mining Breast Cancer Pattern”, Expert Systems With Applications 30, 674-681, (2006).

kariřtırmadan gelen bireylere uygulanmaktadır. Son iki sezgisellik ise loksl minimuma takılınmasını önlemede kullanılmaktadır.²³⁵

Hibrit genetik algoritması adımları ařağıda sıralanmaktadır:

Adım 1:

- Bařlangıç Sezgisellik algoritmasını kullanarak ilk popölasyon kısmını oluřtur,
- Popölasyonun kalanını tesadüfi olarak belirle

Adım2:

- Bařlangıç popölasyonundaki her bir yerleřim için RemoveSharp algoritmasını uygula
- Bařlangıç popölasyonundaki her bir yerleřim için LocalOpt algoritmasını uygula

Adım 3:

- Tesadüfi olarak iki ebeveyn seç,
- Seçilen ebeveynlere çaprazlama uygula ve yeni birey oluřtur,
- Yeni bireye RemoveSharp algoritmasını uygula,
- Yeni bireye LocalOpt algoritmasını uygula,
- Yeni bireyin uygunluęu herhangi bir ebeveyninden büyük ise onunla yeni bireyi deęiřtir.

Adım 4:

- Popölasyondan seçilen herhangi bir bireyi mutasyona tabi tut,

Adım 5:

- Adım 3 ve 4'ü belirlenen iterasyon sayısınca tekrar et.

²³⁵ Soke, A., “Genetik Algoritma ve Benzetilmiş Tavla ile İki Boyutlu Giyotinsiz Kesme Problemlerine Olasılıksal Yaklařım”, Yüksek Lisans Tezi, Kocaeli Üniversitesi Fen Bilimleri Enstitüsü, İzmit, (2003).

Adaptif genetik algoritma

Popülasyon büyüklüğü, çaprazlama olasılığı veya mutasyon olasılığı genetik algoritmanın çalıştırılması sırasında değişen genetik algoritmalar adaptif genetik algoritmalar olarak adlandırılmaktadır.²³⁶

Çok optimumlu problemlerin optimizasyonunda genetik algoritmaların iki önemli karakteristiğinin olması zaruridir. Bu özelliklerin ilki optimuma yakınsama kapasitesi; ikincisi ise çözüm uzayında global optimumu ararken yeni bölgeleri keşfetme kapasitesidir. Bu iki karakteristik arasındaki denge mutasyon olasılığı, çaprazlama olasılığı ve kullanılan çaprazlama yöntemi tarafından belirlenmektedir.²³⁷

Adaptif genetik algoritma yaklaşımında keşif ve sömürü (exploitation) arasındaki dönüşüm, çözümün uygunluk değerine göre mutasyon olasılığı ve çaprazlama olasılığının değiştirilmesi ile yapılmaktadır. Popülasyonun yerel minimuma takılma eğilimi ortaya çıktığında her iki olasılık da artırılmakta; çözüm uzayına yayılımı artınca da azaltılmaktadır.²³⁸

Hızlı dağınık (messy) genetik algoritma

Dağınık genetik algoritma, ikili, stokastik, değişken dizi uzunluklu ve popülasyon temelli bir problem çözme yaklaşımıdır. Diğer yaklaşımlardan en önemli farkı, en iyi çözüme ulaşmak için genetik materyalin yapı bloklarını (building blocks) açıkça manipüle edebilmesidir. Dağınık genetik algoritma üç işlem aşaması içermektedir.

- Başlangıç aşaması,

²³⁶ Niimi, A., Tazaki, E., , “**Rule Discovery Technique Using Genetic Programming Combined with Apriori Algorithm**”, Springer Verlag Berlin,2000, s 273-278.

²³⁷ Alataş, B., Arslan, A., , “**Birliktelik Kurallarının Madenciliği İçin Genetik Algoritma ve Bulanık Kume Tabanlı Yeni Bir Yaklaşım**”, F. Ü. Fen ve Mühendislik Bilimleri Dergisi, 42-51 (2005).

²³⁸ Mitchell, M., ”**An Introduction To Genetic Algorithms** “, The MIT Press, (1998).

- İlksel (primordial) aşama,
- Bitişiklik durumu (juxtapositional) aşaması.

Başlangıç aşamasında, bir sonraki blok filtreleme aşamasında ortaya çıkabilecek gürültünün (noise) üstesinden gelebilmek için yeter derecede büyük bir popülasyon elde etmek için popülasyon boyutu belirleme denklemi kullanılır. Popülasyon büyüklüğü belirlendikten sonra başlangıç popülasyonu üyeleri tesadüfi olarak belirlenir ve seçilen uygulamaya göre onların uygunluk değerleri hesaplanır. Yapı bloğu filtreleme şeması kullanılarak başlangıç aşamasında belirlenen bu popülasyon üyeleri sistematik olarak, kullanıcı tarafından belirlenen yapı bloğu büyüklüğüne düşürülür. Yapı bloğu filtrelemede her bir popülasyon üyesinin belirli sayıda bit'i silinir. Bir sonraki jenerasyon için en iyi kısmi dizilerin saklanabilmesi için bu silme işlemini turnuva seçimi takip eder.²³⁹

Son aşamada ise filtreleme aşamasında gelen yapı blokları alınır ve bir 'kes ve dilimle' (cut-and-slice) operatörü ile birleştirilir. Bu prosesi de turnuva seçimi izlemektedir.²⁴⁰

3.2.1.3 Yapay Sinir Ağları

İnsanların beynin çalışmasını sayısal bilgisayarlar üzerinde taklit etmek istemesi sonucunda yapay sinir ağları ortaya çıkmıştır. Temelde beynin çalışma yapısı, insanlar gibi düşünen ve öğrenen sistemler elde etme fikrinin olması, çalışmaları insan beynini oluşturan hücrelerin incelenmesi üzerine yoğunlaştırmıştır. Bu çalışmalar esnasında her bir nöronun diğer nöronlar ile ilişkili olduğu ve aldığı bazı girdileri çıktıya dönüştürdüğü gözlemlenmiştir. Günümüzde yapay sinir ağları olarak adlandırılan bu alan, birçok nöronun belirli kurallar çerçevesinde bir araya

²³⁹ Deb K., “Genetic Algorithm in Search and Optimization: The Technique and Applications”, s. 1-29.

²⁴⁰ Emel, G.G., Taşkın, C., “Genetik Algoritmalar ve Uygulama Alanları”, Uludağ Üniversitesi İktisadi ve İdari Bilimler Dergisi, 129-152, (2002).

gelerek bir işlevin gerçekleştirilmesi esnasındaki yapısal, matematiksel ve felsefi sorunlara cevap bulmaya çalışan bir bilim dalı olmuştur.²⁴¹

İnsan beyninin taklit edilmesi ile ortaya çıkan bu yapılar her ne kadar nanosaniyeler içerisinde işlemleri yerine getirebilme yetisine sahip olsa da insan beyninin milisaniyeler mertebesindeki işlevsel hızına ulaşamamıştır. Buna rağmen karmaşık sistemlerdeki eşleştirmelerdeki hassaslığı ve yapısal esnekliği ile bu alan artan bir ilgi görmektedir.²⁴² Her ne kadar insan beyni kadar hızlı bir işlevsel yapıya sahip olmasa da sağlamış oldukları hız ile bilgisayar bilimleri, mühendislik, ekonomi ve psikoloji gibi birçok alanda farklı kullanımları bulunmaktadır.²⁴³

Yapay sinir ağı, beynin belirli bir işi veya fonksiyonu yerine getirme yöntemini modelleyen bir makinedir. Ağın yapısı, elektronik sistemler veya bilgisayar yazılımları ile tanımlanmaktadır. Tanımlanan bu model, sinir hücresi veya işlem birimi adı verilen hücreler arasındaki bağlantıyı kullanmakta ve bu işlemler esnasında öğrenme adı verilen bir süreç ile performansını artırabilmektedir. Bu noktada yapay sinir ağı kavramını deneysel bilgi saklama ve kullanıma hazır hale getirme yeteneğine sahip basit işleme birimlerinden oluşan, çok yoğun, paralel ve dağıtılmış düzende çalışan bir işlemci olarak tanımlamak mümkündür. İnsan beyni ile benzerliği ise bilgiyi öğrenme yoluyla elde etmesi ve bilginin depolanması için sinir hücreleri arası bağı kullanmasıdır.²⁴⁴

Öğrenme işlemleri öğrenme algoritmaları olarak adlandırılan ve ağın istenen modele ulaşmasını sağlayan fonksiyonlardan oluşmaktadır.²⁴⁵ Bilgisayarlardaki doğrusal programlamalardan farklı olarak yapay sinir ağları sinir hücreleri arasındaki bağlantıları, eşik değerleri ve ağ yapısı gibi parametreleri kullanarak farklı ayarlar yapıp eğitim gerçekleştirir. Yapay sinir ağlarında bellek ve işlemci paralel olarak

²⁴¹ Şahin, 2001.

²⁴² Efe ve Kaynak, 2000

²⁴³ Kuo ve diğ., 2001

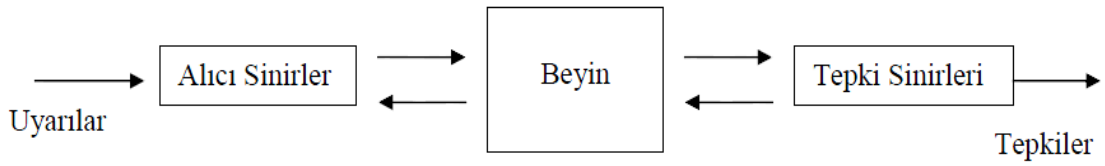
²⁴⁴ Engin O. (2001), **Akış Tipi Çizelgeleme Problemlerinin Genetik Algoritma ile Çözüm Performansının Arttırılmasında Parametre Optimizasyonu**, İTÜ, Arttırılmasında Parametre Optimizasyonu, İTÜ, Fen Bilimleri Enstitüsü, Yayınlanmamış Doktora Tezi.

²⁴⁵ Haykin, 1999.

çalışır. Bilgi sinir hücreleri arasındaki bağlarda da saklanabilir veya değiştirilebilir.²⁴⁶

3.2.1.3.1 Biyolojik Sinir Sistemleri

Biyolojik sinir sistemi verinin alınması, yorumlanması ve karar üretilmesi gibi işlevlerin yürütüldüğü beyni merkezde bulunduran 3 katmanlı bir yapıdır. Uyarılar alıcı sinirler ile birlikte elektriksel sinyallere dönüştürülerek beyne iletilir. Beynin oluşturduğu çıktılar ise, tepki sinirleri tarafından belirli tepkilere dönüştürülür.



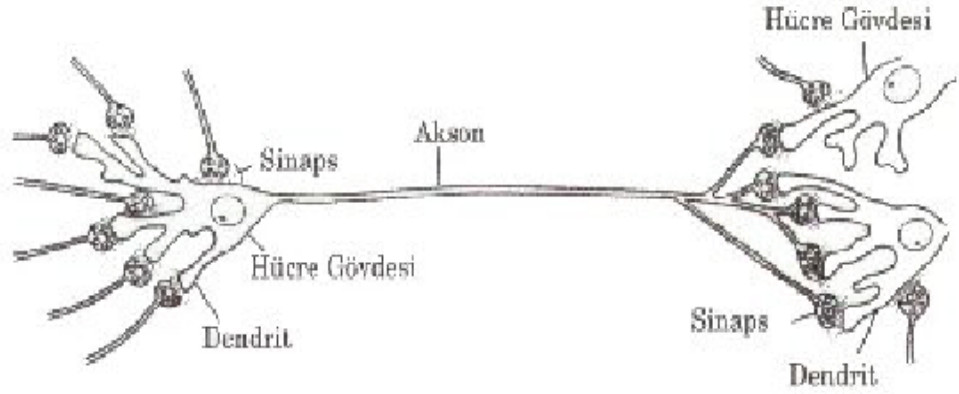
Şekil 3. 9 Biyolojik Sinir Sistemi

Kaynak: Dr. John A. Bullinari, <http://www.cs.bham.ac.uk/~jxb/>, 2004.

Sinir ağı yapısında bilgilerin alıcı ve tepki sinirleri arasında ileri ve geri beslemeli olarak değerlendirilmesi ve sonucunda tepkilerin oluşması, kapalı bir çevrim sürecine benzemektedir. Temel işlem elemanı sinir hücreleridir. İnsan beyinde yaklaşık 10 milyar sinir hücresi olduğu tahmin edilmektedir. Sinir hücreleri hücre gövdesi, gövdeye giren alıcı lifler (dendrit) ve gövdeden çıkan sinyal iletici lifler (akson) olmak üzere 3 temel bileşenden meydana gelir.²⁴⁷

²⁴⁶ Bolat, 2003.

²⁴⁷ Ergezer, H., Dikmen, M. ve Özdemir, E., **Yapay sinir ağları ve tanıma sistemleri**. PiVOLKA,, s 14-17. 2003.



Şekil 3. 10 Sinir Hücresi

Kaynak: Dr. John A. Bullinari, <http://www.cs.bham.ac.uk/~jxb/>, 2004.

Dentritler aracılığı ile bilgiler diğer hücrelerden hücre gövdesine iletilir. Hücrelerde oluşan çıktılar ise akson yardımı ile bir diğer hücreye aktarılır. Aktarımın gerçekleştiği bu noktada aksonlarda ince yollara ayrılabilmekte ve diğer hücrenin dentritlerini oluşturmaktadırlar. Akson-dentrit bağıntısını oluşturduğu bu noktalara sinaps adı verilir.

Sinapsa ulaşan ve dentritler tarafından alınan bilgiler genellikle elektriksel darbelerdir, fakat bu bilgiler sinapstaki kimyasal ileticilerden etkilenirler. Hücrenin tepki oluşturması için bu tepkilerin belirli bir sürede belirli seviyeye ulaşması gerekmektedir. Bu değer eşik değeri olarak adlandırılır.

Tipik bir nöron, hücre gövdesi ve dentritleri üzerine dış kaynaklardan gelen elektrik darbelerinden üç şekilde etkilenir. Gelen darbelerden bazıları nöronu uyarır, bazıları bastırır, geri kalanı da davranışında değişikliğe yol açar. Nöron yeterince uyarıldığında çıkış kablosundan (aksonundan) aşağı bir elektriksel işaret göndererek tepkisini gösterir. Genellikle bu tek akson üzerinde çok sayıda dallar olur. Aksondan inmekte olan elektrik işareti dallara ve alt dallara ve sonunda başka nöronlara ulaşarak onların davranışını etkiler. Nöron, çok sayıda başka nöronlardan genellikle elektrik darbesi biçiminde gelen verileri alır. Yaptığı iş bu girdilerin

karmaşık ve dinamik bir toplamını yapmak ve bu bilgiyi aksonundan aşağı göndererek bir dizi elektrik darbesi biçiminde çok sayıda başka nörona iletmektir. Nöron, bu etkinlikleri sürdürmek ve molekül sentezlemek için de enerji kullanır fakat başlıca işlevi işaret alıp işaret göndermek, yani bilgi alışverişidir.²⁴⁸

Bir sinir hücresinin tepki hızı günümüz bilgisayarlarına göre oldukça yavaş olmasına rağmen duyuşal bilgilerin değerlendirilmesi son derece hızlıdır. Bu insan beyninin öğrenme, birleştirme, uyarılma ve genelleştirme yeteneği nedeniyle son derece karmaşık, doğrusal olmayan ve paralel dağılmış bir bilgi işleme sistemi olduğunu göstermektedir.

Yapay sinir ağlarının oluşturulmasında da sinir sistemi model alınmış ve benzer tanımlamalar yapılmıştır. Bu tanımlamalar aşağıdaki tabloda görülebilir.²⁴⁹

Tablo 3. 14 Sinir Sistemi ile Yapay Sinir Ağlarının Benzerlikleri

Sinir Sistemi	Yapay Sinir Ağı
Nöron	İşlem Elemanı
Dentrit	Transfer Fonksiyonu
Hücre Gövdesi	Hareket Fonksiyonu
Akson	Eleman Çıkışı
Sinaps	Ağırlıklar

Yapay sinir ağlarının insan beynini modellemesi, dijital bilgisayarların modellenmesi gerçeğinde de var olan insan beyni gibi işleyen makineler oluşturma isteği ile ortaya çıkmıştır. 1943 yılında bir nörobiyolojist olan Warren McCulloch ve bir istatistikçi olan Walter Pitts, "Sinir Aktivitesinde Düşüncelere Ait Bir Mantıksal Hesap" başlıklı bir makale ile ilk dijital bilgisayarlara ışık tutmuştur. 1949 yılında, Hebb, daha sonra kendi adı ile anılan öğrenme kuramını, yapay sinir ağına uygulamıştır. Bu kuram "A hücresinin bir aksonu, B hücresini uyaracak ve tekrarlı

²⁴⁸ <http://bagem.tripod.com/bagem/yz2.html>, 10.05.2009

²⁴⁹ <http://batitrakya.dostweb.com/yapaysiniraglari/yapaysiniraglari.htm>, 18.07.2010

veya sürekli olarak tetikleyecek kadar yakında ise, hücrelerde B'yi tetikleyen A'nın etkinliğini artıracak bir büyüme işlemi veya metabolik değişiklik olur" şeklindedir.

Marvin Minsky, Hebb kuramı ile birlikte makroskobik zeka kavramını ortaya atmış ve uzman sistemlerin doğmasını sağlamıştır. Frank Rosenblatt, Bronx Yüksek Bilim Okulu'nda gözün hesaplamaları ile ilgilenmiştir. Bu bilim adamları, öğrenme ve zeka konusunu bilgisayar simülasyonlarında nasıl kullanabileceklerini 1956'da düzenlenen ilk yapay zeka konferansında tartışmışlardır.²⁵⁰

1959'da Stanford üniversitesinden Bernard Widrow, "ADALİNE" (Adaptive Linear Neuron) olarak adlandırılan bir adaptif lineer elemanı geliştirmiştir. Adaline ve iki tabakalı biçimi olan "MADALİNE" (Multiple Adaline) ses tanıma, hava tahmini, karakter tanıma ve adaptif kontrol gibi çok çeşitli uygulamalar için kullanılmıştır. Widrow, telefon hatları üzerinde yansımaları elemine etmeye yarayan adaptif filtreleri geliştirmede, adaptif lineer eleman algoritmasını kullanmıştır. Bu çalışma ile birlikte yapay sinir ağları ilk defa gerçek bir probleme uygulanmıştır.

1970'lerin ilk başlarında Helsinki Teknik Üniversitesi'nden Teuvo Kohonen, öğrenme ve birleşik hafızalar üzerine çalışmalar yapmış bu çalışmalar danışmansız öğrenme metotlarının gelişmesine ışık tutmuştur. Yapay sinir ağları ile ilgili 1969-1982 yılları karanlık bir dönem olarak görülmektedir. 1969 yılında Minsky ve Papert'in yapmış olduğu çalışma ile yapının XOR lojik problemlerine çözüm getiremediği ispatlanmıştır. Bu gelişme yapay sinir ağlarına olan ilgiyi azaltmış ve karanlık dönemin başlangıcı olmuştur.

1974 yılında Werbos, Geri Yayılmalı Algoritmayı tanıtmıştır. Willshaw ve Malsburg 1976 yılında Özdüzenlemeli Harita adında kümeleme algoritmasını kurmuştur. 1982 yılında Kohonen, özdüzenlemeli harita kuramını beyindeki oluşumların karşılaştırmalı haritasını çıkarabilmek için ortaya koymuştur.

²⁵⁰ M. Nikravesh, V. Loia, B. Azvine, , **Fuzzy Logic and the Internet (FLINT): Internet, World Wide Web, and Search Engines**, 2000.

Hopfield,1982 yılında moleküler biyolojiden beyin kuramcılığına geçiş yapan bir model geliştirmiştir.

Günümüzde kendi adıyla anılan bir ağ yapısı mevcuttur ve bu yapı birçok alanda uygulanmıştır. 1986 yılında, Rumelhart, 1974'te Werbos tarafından bulunan çok katmanlı algılayıcı tipli geri yayılmalı algoritma denen bir eğitim algoritması geliştirmiştir.²⁵¹ Yine aynı tarihlerde ilk yapay sinir ağları sempozyumu yapılmış ve sonrasında uygulamalar artmaya başlamıştır. Günümüzde bu alanda birçok araştırma grubu bulunmakta ve farklı alanlardan farklı uygulamacılar bu konuda çalışmaktadır.²⁵²

Günümüze kadar uygulanmış yapay sinir ağları modellerinin ortak özelliklerini yansıtan tanımlar aşağıdaki şekilde ele alınabilir.²⁵³ Ağ, hücre denilen yapı taşlarının birbirine belirli bir geometri ile bağlanmasından oluşan bir sistemdir. Hücreler, genelde çok girişli ve tek çıkışlı, yüksek dereceden lineer olmayan dinamik alt devrelerdir. Tek olan çıkış işareti, çoğullanarak dışarıya çıkış olarak alınabilir veya diğer hücrelere giriş olarak kullanılabilir. Bir hücre dış girişleri, komşu hücre çıkışlarını ve bir eşik değerini giriş olarak kabul eder. Hücrenin lineer dinamik birimi herhangi bir dereceden olabilir. Hücre çıkışındaki lineer olmayan cebrik fonksiyon tek giriş-çıkışlıdır. Hücre parametreleri hücreler arasında farklılık gösterebilir. Ağ, genelde çok boyutlu katmanların ileri veya geri beslemeli olacak şekilde bağlanmasından oluşur.

Ağdaki her katman, hücrelerin çok boyutlu bir dizisidir. Katman içinde, her bir hücre komşularına aynı şekilde bağlıdır. Böylece katman içi bağlantı geometrisi düzgündür. Katmanlar arası bağlantı, bir metrik ile tanımlı düzgün bir geometriye sahiptir. Ağda yalnızca bağlantı ağırlıkları değişebilir. Bağlantı ağırlıkları, ya eğitici ya da eğitici olmayan bir öğrenme kuralı ile değiştirilir ya da önceden belirlenen değerlerde sabit tutulur.

²⁵¹ Güzelış, 1993

²⁵² Bolat, 2003 Haykır, 1999

²⁵³ Güzelış, 1993

Literatürde çok farklı ağ yapıları yer almasına rağmen genel bir yapay sinir ağın'da:

Girdi Katmanı: Girdi katmanı en az bir girdi elemanının bulunduğu bölümdür. Girdi katmanında veriler herhangi bir işleme tabi tutulmadan girdileri ile aynı değerde çıktı üretirler.²⁵⁴

Çıktı Katmanı: Çıktı katmanı en az bir çıktıdan oluşur ve çıktı ağ yapısında bulunan fonksiyona bağlıdır. Bu birimlerde girdi katmanında olanın aksine işlem gerçekleştirilir ve birimler kendi çıktılarını üretirler.

İşlem Katmanı: Genellikle "Kara Kutu" olarak adlandırılır. Girdi birimlerinin belirli işlemlere tabi tutulduğu bölgedir. Seçilen ağ yapısına göre işlem katmanının yapısı ve fonksiyonu da değişebilir. Tek bir katmandan oluşabileceği gibi birden fazla katmandan da oluşabilir²⁵⁵.

Yapay sinir ağlarının günümüzde kullanılan diğer bilgi işlem yöntemlerinden farklılıkları aşağıdaki şekilde sıralanabilir.²⁵⁶

Paralellik: Alışlagelmiş bilgi işlem yöntemlerinin çoğunda işlemler seri bir düzen içerisinde. Bu düzen özellikle hız sorununa yol açmaktadır. Daha öncede vurgulandığı gibi bilgisayarla beyne göre çok hızlı çalışmasına rağmen beynin toplam hızı bilgisayara göre çok yüksektir. Yapay sinir ağlarında işlemler doğrusal değildir ve bu bütün ağa yayılmış durumdadır. Aynı katmanlar arasında zaman bağımlılığı da bulunmamaktadır. Bu, tüm sistemin eş zamanlı çalışabilmesine olanak vermekte ve hızı çok artırmaktadır. Bu sayede doğrusal olmayan karmaşık problemlerin de çözülmesi

²⁵⁴ Hawley ve Johnson, 1994

²⁵⁵ Chaudhury ve diğ., 1994

²⁵⁶ Seven, 1993 Simpson, 1990

mümkündür.

Öğrenebilirlik:

Mevcut bilgi işlem sistemleri belirli bir algoritma kapsamında çalışmakta ve kendisi ağırlık ya da verileri yenileyememektedir. Bu durumda tam tanımlı olmayan problemler de çözülememektedir. Yapay sinir ağları ile sisteme önceden girilen örnekler kullanılarak ağırlıkların saptanmasını sağlamakta ve her yeni çalışmada bu öğrenme işlemi yenilenebilmektedir. Amaca uygun olacak bir çalışmada önceden ağırlıkların ve bağlantıların verilememesi bir sorun oluşturmakta iken yapay sinir ağlarının örneklerle kendini eğitmesi ve gerekli verileri oluşturması bu sorunu ortadan kaldırmaktadır.

Hata Toleransı:

Geleneksel yöntemlerde herhangi bir elemanı yerinden almak o sistemin çalışmasını imkansız kılabilir. Yapay sinir ağlarındaki paralel yapı, ağına sahip olduğu bilginin tüm bağlantılara yayılmasını sağlamaktadır. Bu sayede bazı bağlantıların veya hücrelerin etkisiz hale gelmesi ağına doğru bilgiyi üretmesini önemli derecede etkilemez ve bununla birlikte ağların geleneksel yöntemlere göre hatayı tolere etme yetenekleri yüksektir.

Uyarlanabilirlik:

Yapay sinir ağlarında ağırlıkların yeniden yapılandırılabilir olması belirli bir problemi çözmek için eğitilen yapay sinir ağının, problemdeki değişikliklere göre yeniden eğitilebilmesi ve farklı koşullara uyarlanabilmesini sağlamaktadır. Bu özellik yapay sinir ağlarının örnek tanıma, sinyal işleme, sistem tanılama ve denetim gibi alanlarda etkin olarak kullanılmasını sağlamıştır.

Genelleme:

Yapay sinir ağları eğitimi sonrasında, eğitim sırasında karşılaşılmayan test örnekleri de değerlendirip, arzu edilen tepkiler üretebilir. Örneğin, karakter tanımlamada bozuk karakterlerin girişlerinde doğru karakterlerin elde edilmesi mümkündür.

Yerel Bilgi İşleme:	Yapay sinir ağlarında problemin tümü ile ilgilenmek yerine parçaları ile ilgilenilebilmesi ve görev paylaşımı sayesinde çok karmaşık ve zor problemler çözülebilmektedir.
Gerçekleme Kolaylığı:	Karışık fonksiyonların yerine basit işlemlerin kullanılması gerçekleştirme kolaylığı sağlamaktadır.
Donanım ve Hız:	Yapay sinir ağları, paralel yapısı sayesinde entegre devre teknolojisi ile gerçekleştirilebilir. Bu özellik, hızlı bilgi işleme yeteneğini artırmaktadır.

3.2.1.3.2 Yapay Sinir Ağı Uygulama Alanları

Yapay sinir ağları karmaşık ve ekonomik olmayan çok farklı alanlardaki problemlerin çözümünde başarı ile uygulanmıştır.²⁵⁷ Genel olarak uygulama alanları aşağıdaki şekildedir.²⁵⁸

Sınıflandırma:	Sinir ağından bir örnek geçirilir. Buna karşılık çıkış olarak bir örnek sınıfı belirtilir.
Tahmin:	Sinir ağlarına verilen değerlerden bir sonuç elde edilmesi için kullanılır. Gelecek tahminleri, talep tahminleri gibi alanlarda kullanılabilir.
Örnek Tamamlama:	Eksik bir örnek ağı girildiğinde örneğin eksik kalan kısımları ağ tarafından tamamlanır ve tam örnek çıktı olarak alınır.
Optimizasyon:	Ağı bir optimizasyon problemi girilir ve ilk değerleri ifade eden örnek kümesinin girilmesinden sonra çözüme ilişkin değişkenler kümesi çıktı olarak alınır.
Örnek Eşleştirme:	Ağı girilen örneğin tipine en uygun örnek tipi belirlenir.
Yapay Zeka:	Ses ve görüntü tanımlama gibi alanlarda kullanılabilir.

²⁵⁷ May, 1994

²⁵⁸ Haykin, 1999

Gürültü	Gürültülü bir veri ağı girildiğinde gereksiz veriler elenir ve yeni bir veri kümesi elde edilir.
Uzaklaştırma:	
Finansman ve Yatırım:	Kredi analizi, sigorta riskleri, opsiyon ve gelecek trend, stok yatırım analizleri gibi olayların modellenmesinde kullanılır.
Üretim:	Kalite kontrol ve analizler ile model kurma ve geliştirme ve mevcut otomasyon sistemini geliştirme mümkündür.
Kontrol:	Girdiye göre üretilen sonuçlar ile mevcut duruma ait sonuçlar arasında karşılaştırma yapılır.
Tıp:	Hastalık nedenlerinin belirlenmesi ve sınıflandırılması, genetik ve kan örnekleme haritalarının oluşturulması gibi işlemlerde kullanılmaktadır.
Bilim ve Mühendislik:	Karmaşık modellerin kurulması, doğrusal olmayan modeller, çok değişkenli eğri uyumları, iklim modellemesi gibi konularda uygulamalar yer almaktadır. Geleneksel istatistik ve regresyon testlerinin yetersiz kaldığı bir çok konuda kullanılabilir.

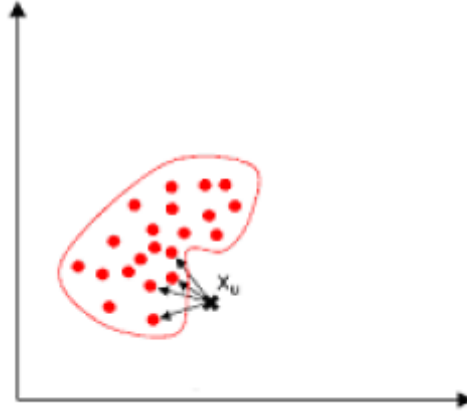
3.2.1.4 K-En yakın Komşu Algoritması

K en yakın komşu algoritması, sınıflandırma problemini çözen denetimli öğrenme (sınıflandırma için öğrenme kümesi kullanır) algoritmalarından biridir. Sınıflandırma, yeni bir nesnenin özniteliklerini inceleme ve bu nesneyi önceden tanımlanmış bir sınıfa atamaktır. Burada önemli olan, her bir sınıfın özelliklerinin önceden net bir şekilde belirlenmiş olmasıdır.²⁵⁹

K en yakın komşu yönteminde; sınıflandırma yapılacak verilerin öğrenme kümesindeki normal davranış verilerine benzerlikleri hesaplanarak; en yakın olduğu düşünülen k verinin ortalamasıyla, belirlenen eşik değere göre sınıflara atamaları yapılır.²⁶⁰

²⁵⁹ Pietraszek, T. and Tanner, A., 2005, **Data Mining and Machine Learning Towards Reducing False Positives in Intrusion Detection**, Information Security Technical Report, Elsevier, 10.3, 169-83

²⁶⁰ Nikulin, V., 2005, **Threshold-Based Clustering with Merging and Regularization in Application to Network Intrusion Detection**, Computational Statistics & Data Analysis, Elsevier, 51.2, 1184-1196



Grafik 3. 1 K En Yakın Komşu Algoritması Grafiksel Gösterimi

Kaynak: Bishop, C. M. Neural Networks for Pattern Recognition. Oxford, England: Oxford University Press, 1995.

KNN algoritmasının adımları;²⁶¹

- Test kümesindeki her verinin $X = \{x_1, x_2, x_3, \dots, x_n\}$, öğrenme kümesindeki verilere $D = \{d_1, d_2, d_3, d_4, \dots, d_m\}$ yakınlığı hesaplanır.

$$\text{sim}(x_i, d_l) = \frac{x_i \bullet d_l}{\|x_i\| \|d_l\|}$$

$$(i = \{1, 2, 3, \dots, n\}, l = \{1, 2, 3, \dots, m\})$$

- Her verinin öğrenme kümesindeki verilere olan yakınlıkları sıralanıp ilk "k" tanesi alınarak ortalamaları hesaplanır.

$$\text{sim_avg}(x_i) = \frac{\max \left(\sum_{l=1}^k \text{sim}(x_i, d_l) \right)}{k}$$

- Ortalama değerleri, belirlenen eşik değerinden büyük olanlar normal, küçük olanlar ise anormal olarak sınıflandırılır.

²⁶¹ Cover TM, Hart PE (1967). "Nearest neighbor pattern classification". IEEE Transactions on Information Theory 13 (1): 21–27.doi:10.1109/TIT.1967.105396

```

Öğrenme kümesi D belirlenir  $D = \{d_1, d_2, d_3, \dots, d_m\}$ ;
for test kümesi(X) içindeki her veri  $x_i$  için do
    if  $x_i$  bilinmeyen bir sistem çağırısı ise then
         $x_i = \text{anormal}$ ;
    else
        for öğrenme kümesi(D) içindeki her veri  $d_i$  için do
            benzerlik hesapla  $\text{sim}(x_i, d_i)$ ;
            if  $\text{sim}(x_i, d_i) = 1$  then
                 $x_i = \text{normal}$ ; exit;
        En büyük k tane  $\text{sim}(x_i, d_i)$  benzerliği bul;
        En yakın k komşu için benzerlik ortalaması( $\text{sim\_avg}$ ) hesapla;
        if  $\text{sim\_avg}(x_i) > \text{eşik değeri}$  then
             $x_i = \text{normal}$ ;
        else
             $x_i = \text{anormal}$ ;

```

Şekil 3. 11 K En Yakın Komşu Algoritması

Kaynak: Bishop, C. M. Neural Networks for Pattern Recognition. Oxford, England: Oxford University Press, 1995.

Algoritmanın performansını etkileyen kriterler;²⁶²

- k en yakın komşu sayısı (benzerlik ölçümü için seçilecek komşu sayısı: k)
- eşik değeri (etiketleme işlemi için verinin k en yakın komşuya olan benzerliklerinden hesaplanan ortalama değerinin kıyaslanmasında kullanılır)
- benzerlik ölçümü
- öğrenme kümesindeki normal davranışların yeterli sayıda olması (öğrenme kümesi yeterli çeşitlilikte ve sayıda normal davranış verisi içermiyorsa, test kümesinde yer alan yeni normal davranış verileri anormal olarak algılanabilir.)

²⁶² Toussaint GT (April 2005). "Geometric proximity graphs for improving nearest neighbor methods in instance-based learning and datamining". International Journal of Computational Geometry and Applications 15 (2): 101–150. doi:10.1142/S021819590500162

Bu yöntemde sınıfları belli olan bir örnek kümesindeki gözlem değerlerinden yararlanarak, örneğe katılacak yeni gözlemin hangi sınıfa ait olduğunu belirlemek amacıyla kullanılır. Örnek kümedeki gözlemlerin her birinin, sonradan belirlenen bir gözlem değerine olan uzaklıklarının hesaplanması ve en küçük uzaklığa sahip k sayıda gözlemin seçilmesi esasına dayanmaktadır. Uzaklıkların hesaplanmasında, i ve j noktaları için Öklid uzaklık formülü kullanılabilir:²⁶³

$$\text{karekök}((x_2-x_1)^2 + (y_2-y_1)^2 + (z_2-z_1)^2)$$

Algoritmanın uygulaması aşağıdaki gibidir.

- K parametresi belirlenir. Bu parametre verilen bir noktaya en yakın komşuların sayısıdır.
- Söz konusu nokta ile diğer tüm noktalar arasındaki uzaklıklar tek tek hesaplanır.
- Adım 2’de hesaplanan uzaklıklara göre satırlar sıralanır ve bunlar arasından en küçük olan k tanesi seçilir.
- Seçilen satırların hangi kategoriye ait oldukları belirlenir ve en çok tekrarlanan kategori değeri seçilir.
- Seçilen kategori, tahmin edilmesi beklenen gözlem değerinin kategorisi olarak kabul edilir.

²⁶³ Bremner D, Demaine E, Erickson J, Iacono J, Langerman S, Morin P, Toussaint G (2005). "**Output-sensitive algorithms for computing nearest-neighbor decision boundaries**". Discrete and Computational Geometry 33 (4): 593–604. doi:10.1007/s00454-004-1152-0

Elimizde 10 adet gözlem verisi olduğu düşünülürse:

Tablo 3. 15 Örnek Gözlem Değerleri

X1	X2	Y
2	4	BAD
3	6	GOOD
3	4	GOOD
4	10	BAD
5	8	BAD
6	3	GOOD
7	9	GOOD
9	7	BAD
11	7	BAD
10	2	BAD

Yukarıdaki gözlem değerlerine bakılarak yeni bir gözlem olan $X1=8$ ve $X2=4$ değerlerinin yani (8,4) gözleminin hangi sınıfa dahil olduğunu k-en yakın komşu yöntemi ile bulmak istiyoruz.

-K'nın belirlenmesi: $k=4$ olarak alıyoruz.

-Uzaklıkların hesaplanması: (8,4) noktası ile gözlem değerlerinin herbirisi arasındaki uzaklıklar hesaplanır. Bunun için Öklid uzaklık bağıntısı kullanılır.

$$\begin{aligned}
d(i,j) &= \sqrt{(2-8)^2 + (4-4)^2} = 6.00 \\
d(i,j) &= \sqrt{(3-8)^2 + (6-4)^2} = 5.39 \\
d(i,j) &= \sqrt{(3-8)^2 + (4-4)^2} = 5.00 \\
d(i,j) &= \sqrt{(4-8)^2 + (10-4)^2} = 7.21 \\
d(i,j) &= \sqrt{(5-8)^2 + (8-4)^2} = 5.00 \\
d(i,j) &= \sqrt{(6-8)^2 + (3-4)^2} = 2.24 \\
d(i,j) &= \sqrt{(7-8)^2 + (9-4)^2} = 5.10 \\
d(i,j) &= \sqrt{(10-8)^2 + (2-4)^2} = 2.83 \\
d(i,j) &= \sqrt{(3-8)^2 + (4-4)^2} = 5.00 \\
d(i,j) &= \sqrt{(3-8)^2 + (4-4)^2} = 5.00
\end{aligned}$$

Tablo 3. 16 Örnek Gözlem Değerleri Öklit Uzaklık Değerleri

X1	X2	DİSTANCE
2	4	6.00
3	6	5.39
3	4	5.00
4	10	7.21
5	8	5.00
6	3	2.24
7	9	5.10
9	7	3.16
11	7	4.24
10	2	2.83

- **En küçük uzaklıkların belirlenmesi:** Satırlar küçükten büyüğe doğru ya da büyükten küçüğe doğru sıralanarak en küçük k=4 tanesi belirlenir. Bu k adet nokta

verilen (8,4) noktasına en yakın gözlem değeridir. Buna göre 2.24, 2.83, 3.16, 4.24 sonuçlarının olduğu noktalar en yakın sonuçlardır.

- **Seçilen satırlara ilişkin sınıfların belirlenmesi:** (8,4) noktasına en yakın olan gözlem değerlerinin Y sınıfları göz önüne alınır ve hangisinin daha fazla olduğuna bakılır. Bu 4 sonuç içerisinde 1 adet GOOD, 3 adet BAD sonucu vardır. Buna göre verilen (8,4) noktasının sınıfı BAD olarak belirlenir.

3.2.1.5 Naive Bayes

Belirsizlik insan hayatında büyük bir alan kaplamıştır. Bu belirsizlik bilgi eksikliğine, unutilan geçmişe veya henüz gerçekleşmemiş durumlara bağlı olabilir. Günlük yaşantıda sürekli olarak kullanılan düşünceler, sezgiler bilimsel amaçlar için kullanılabilir. Bu tür oluşumların bilimsel olarak kullanılmasında Bayesci yaklaşım alternatif olarak ortaya çıkmaktadır.²⁶⁴

Bir olasılıksal sınıflandırma olan Navie Bayes sınıflandırmanın ana fikri, bir belgenin sınıfının olasılığını tahmin etmek için verilen bir kelimenin sınıfının koşullu olasılıklarını kullanmaktır. Belge sınıflandırma gibi bazı öğrenme problemlerinde yaygın olarak kullanılan en pratik yaklaşımdır. Bu yaklaşımın "Naive" kısmı içindeki kelime bağımsızlığı varsayımından kaynaklanmaktadır. Çünkü kelime kombinasyonlarının olasılıklarını tahminci olarak kullanmaz. Bu yüzden Karar Ağacı (Decision Tree) gibi algoritmaların üstel karmaşıklığından öte verimli bir yaklaşımdır ve performansı Yapay Sinir Ağları (Neural Networks) ve Karar Ağacı ile karşılaştırılabilir.²⁶⁵

Naive Bayes'de Artımlı (Incremental) olarak tabir edilen online bir öğrenme durumu vardır; her bir talim örneği artımlı olarak bir hipotezin doğru olma olasılığını arttırır veya azaltır. Öncül bilgi gözlemlenen verilerle birleştirilebilir.

²⁶⁴ Gürsaka N., "Bayesci İstatistik", **Uludag Üniversitesi Basımevi**, Bursa,1992, 68.

²⁶⁵ Zdziarski, Jonathan, A., "Ending Spam: Bayesian Content Filtering and the Art of Statistical Language Classification", **No Starch Pres Inc.**,2005, 10.

3.2.1.5.1 Naive Bayes Kuralının Çalışma Yapısı

Klasik yaklaşım, gözlemlenmiş veriler yardımı ile hakkında bilgi sahibi olunması istenen kitleye ait sonuçlar çıkarılmasıdır. Hipotez testleri gibi istatistiksel analizler, klasik yaklaşım kullanılarak yapılan çalışmalardır. Samuel Butler, "Hayat, yetersiz önsellerden yeterli sonuçlar çıkarma sanatıdır." sözüyle kitleye ilişkin parametrelere ait çıkarsamalara yeni bir boyut getirmiştir. Thomas Bayes, bu mantıkla örtüşen biçimde, nedenlerden sonuçlara doğru olan mantık izleyişini sonuçlardan nedenlere doğru değiştirmiştir.²⁶⁶

Naive Bayes yaklaşımı, genellikle sonrasal olasılıkları hesaplamakta kullanılan ve iki rastgele olayın koşullu ve marjinal olasılıklarını ilişkilendiren bir teoremdir. Maksimum Olabilirlik ilkesi üzerine kurulu bir teoremdir. Bu durumda Bayes Teoremi, konulan olasılıkların doğruluk oranını hesaplamak için kullanılabilir.²⁶⁷

A ve B rastgele olaylar olsun;

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$

$P(A)$: A olayının bağımsız olasılığı

$P(B)$: B olayının bağımsız olasılığı

$P(B | A)$: A olayının olduğu bilindiğinde B olayının olasılığı

(Likelihood, Şartlı Olasılık)

$P(A | B)$: B olayının olduğu bilindiğinde A olayının olasılığı

(Posterior, Artçıl Olasılık)

²⁶⁶ Blake, C. L., & Merz, C. J. (1998). **UCI repository of machine learning databases**. <http://www.ics.uci.edu/~mllearn/MLRepository.html>

²⁶⁷ Breese, J. S., Heckerman, D., & Kadie, C. (1998). **Empirical analysis of predictive algorithms for collaborative filtering**. Proc. UAI-98 (pp. 43–52)

Naive Bayes kuralına dayanarak $P(A | B)$ yi maksimum yapan durumlar hesaplanabilir.

E" A olayının bütün durumlarının kümesi;

$$\begin{aligned} A_{MAX} &= \arg \max_{A \in E} P(A | B) \\ &= \arg \max_{A \in E} \frac{P(B | A)P(A)}{P(B)} \\ &= \arg \max_{A \in E} P(B | A)P(A) \end{aligned}$$

Burada $P(B)$ sabit olarak göz ardı edilebilir. Sınıflandırıcının görevi yeni bir örnek için doğru sınıfı tahmin etmek olacaktır.²⁶⁸

Sınıflandırma problemi öncül (a-priori) olasılıklar kullanılarak şu şekilde formüle edilebilir:

$P(v | X)$ = örnek durum için olasılık

$X = (a_1, \dots, a_k)$ v sınıfının örnekleri

$P(\text{sınıf} = \text{tahmin edilen sonuç} | \text{parametre1, parametre2, parametre3, ...})$

X örneğine $P(v | X)$ olasılığını maksimum yapan sınıf etiketi atanır.

Bayes Teoremi:

$$P(v | X) = P(X | v)P(v)/P(X)$$

$P(X)$ bütün sınıflar için sabittir

$P(v)$ = sınıfın bütün örneklerle karşı göreceli frekansı

$P(v | X)$ maksimum = $P(X | v)P(v)$ maksimum

$P(a_1, \dots, a_k | v) = P(a_1 | v) \dots P(a_k | v)$

²⁶⁸ Cheeseman, P., & Stutz, J. (1996). **Bayesian classification (AutoClass): Theory and results.** In **Advances in knowledge discovery and data mining**, 153–180. Menlo Park, CA: AAAI Press

Eğer i . Özellik kesikli değerde ise $P(a_i | v)$, v sınıfında i . Özellik olarak a_i değerine sahip örneklerin göreceli frekansı olarak tahmin edilir.

Eğer i .ci özellik sürekli ise;

- Kesikli hale getirilebilir
- veya $P(a_i | v)$ Gauss (Normal) dağılımına sahip olduğu varsayılarak tahmin edilir, sadece ortalama varyans gereklidir.

İki durumda da hesaplanabilir.

Bayes Teoreminin Değişik Şekilleri

Bayes teoremi çok kere daha ek kavramlar eklenerek, sanki daha da genişletilerek, ifade de edilir. Bunun için önce şu ifade kullanılır: ²⁶⁹

$$P(B) = P(A \cap B) + P(A^c \cap B) = P(B|A)P(A) + P(B|A^c)P(A^c)$$

Burada A^c (çok kere A olmayan olarak ifade edilen) A olayının tamamlayıcısı olur. Bu Bayes teoremi formülüne konulunca Bayes teoremi için yeni alternatif bir formül elde edilir:

$$P(A | B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

Daha genel olarak , $\{A_i\}$ olay uzayının bir bölüntüsünü oluşturduğu göz önüne alınca, bu bölüntü içinde bulunan herhangi bir A_i için şu ifade elde edilir:

$$P(A_i | B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)}$$

²⁶⁹ D. G. Terrell; D. W. Scott (1992). "Variable kernel density estimation". Annals of Statistics 20: 1236–1265. doi:10.1214/aos/1176348768

Bayes Teoreminin Kapsamının Genişletilmesi

$$P(A|B, C) = \frac{P(A)P(B|A)P(C|A, B)}{P(B)P(C|B)}$$

Bu Bayes teoreminin ve koşullu olasılık tanımlamasının üzerine birkaç işlem yaparak genişletilmesi sağlanabilir.

$$\begin{aligned} P(A|B, C) &= \frac{P(A, B, C)}{P(B, C)} = \frac{P(A, B, C)}{P(B)P(C|B)} = \\ &= \frac{P(C|A, B)P(A, B)}{P(B)P(C|B)} = \frac{P(A)P(B|A)P(C|A, B)}{P(B)P(C|B)} \end{aligned}$$

Bu çalışmalar için uygulanacak genel strateji ortak olasılık için parçalama ile çalışmaya başlayıp ilgili değişkenleri entegrasyon ile ayırmaktır. Uygulanan parçalama şekline göre, bazı entegrallerin 1'e eşit olup parçalama ifadelerinden düşmesi sağlanabilir, bu şekilde hesaplamalarda önemli şekilde azalmalar görülür.²⁷⁰

3.2.1.5.2 Naive Bayes Yöntemi İle İlgili Örnek Uygulamalar

Örnek 1;

Kaleye sürekli şut çeken bir futbolcunun atacağı şutun şekline göre, topun gol olma olasılığı hesaplanabilir. Bunun için atış şekillerinden faydalanılmaktadır (Topu Attığı Yön, Topa Vuruş Sertliği, Topun Yüksekliği gibi).

Aşağıda ki tabloda görüldüğü gibi şut şekillerine göre daha önce yaptığı vuruşlar ve gol durumları listelenmiştir.

²⁷⁰ Chickering, D. M. (2002). **The WinMine Toolkit** (Technical Report MSR-TR-2002-103). Microsoft, Redmond, WA

Tablo 3. 17 Şutların Veri Kümesi

Şut	Topu Attığı Yön	Topa Vuruş Sertliği	Topun Yüksekliği	Gol Durumu
1.Şut	Sağ	Sert	Yerden	Evet
2.Şut	Sol	Sert	Havadan	Evet
3.Şut	Sağ	Yumuşak	Havadan	Hayır
4.Şut	Orta	Sert	Yerden	Hayır
5.Şut	Sağ	Yumuşak	Havadan	Evet
6.Şut	Orta	Yumuşak	Havadan	Evet

Şut	Topu Attığı Yön	Topa Vuruş Sertliği	Topun Yüksekliği	Gol Durumu
7.Şut	Sol	Sert	Yerden	Hayır
8.Şut	Sol	Yumuşak	Havadan	?

Tabloda bulunan değerler, şut şekillerine göre gol olma frekanslarını gösterir. Bu frekanslar daha önceden verilen şut veri kümesine göre oluşturulmuştur. Buna göre futbolcu, topu sağ tarafa attığı 3 şuttan 2' sini, ortaya attığı 2 şuttan 1' ini, sola attığı 2 şuttan 1' ini, sert yaptığı 4 vuruştan 2' sini, yumuşak yaptığı 3 vuruştan 2' sini, yerden atılan 3 toptan 1' ini, havadan atılan 4 şuttan da 3' ünü gol ile sonuçlanmıştır. Ayrıca atılan 7 şuttan 4' ü gol ile sonuçlanmıştır.

Tablo 3. 18 Şut Şekillerine Göre Gol Durum Frekansları

Topu Attığı Yön	EVET	HAYIR	Topa Vuruş Sertliği	EVET	HAYIR	Topun Yüksekliği	EVET	HAYIR	Gol Durumu	
Sağ	2	1	Sert	2	2	Yerden	1	2	Evet	Hayır
Orta	1	1	Yumuşak	2	1	Havadan	3	1	4	3
Sol	1	1								

Aşağıdaki tablo toplanan bu frekanslara göre oluşturulmuş gol olasılıklarını içermektedir. Bu olasılıklar, Topu Attığı Yönüne göre çıkan dağılıma göre sağ taraftan

atılanlar için 2/4, ortadan atılanların için 1/4 ve sol taraftan atılanların için 1/4 şeklindedir. Topa Vuruş Sertliğine göre çıkan dağılıma bakıldığında, sert yapılan vuruşlarda 2/4, yumuşak yapılan vuruşlarda ise 2/4 şeklindedir. Topun Yüksekliğine göre bakıldığında ise bu olasılık oranları, yerden yapılan vuruşlarda 1/4, havadan yapılan vuruşlarda ise 3/4 değerlerini vermektedir.²⁷¹

Tablo 3. 19 Gol Durum Frekanslarına Göre Gol Olasılıkları

Topu Attığı Yön	EVET	HAYIR	Topa Vuruş Sertliği	EVET	HAYIR	Topun Yüksekliği	EVET	HAYIR	Gol Durumu	
Sağ	2/4	1/3	Sert	2/4	2/3	Yerden	1/4	2/3	Evet	Hayır
Orta	1/4	1/3	Yumuşak	2/4	1/3	Havadan	3/4	1/3	4/7	3/7
Sol	1/4	1/3								

Bu frekans ve olasılıklar eşliğinde atılacak yeni şutun tipine göre atışın gol olma olasılığı hesaplanabilir. Futbolcu topu sağdan, yumuşak ve havadan olacak şekilde attığında gol olasılığını hesaplanır.

Topu Attığı Yön	Topa Vuruş Sertliği	Topun Yüksekliği	Gol
Sol	Yumuşak	Havadan	?

Talim kümesindeki olasılık dağılımına dayanarak her bir sınıf için olasılık hesaplanır. Önce her bir özelliğin olasılığı hesaba katılır, bütün özelliklere eşdeğer önemde ele alınır ve olasılıklar çarpılır [6].

$$P_{(EVET)} = 1/4 \cdot 2/4 \cdot 1/4 = 0,0313$$

$$P_{(HAYIR)} = 1/3 \cdot 2/3 \cdot 1/3 = 0,0370$$

²⁷¹ Domingos, P., & Pazzani, M. (1997). **On the optimality of the simple Bayesian classifier under zero-one loss**. Machine Learning, 29, 103–13

Yukarıda çıkan sonuçlar, atılacak olan şutun durumlarına göre gol olma olasılıklarının çarpılmasıyla hesaplamıştır. Burada P(EVET) oluşturulurken alınan çarpım değerleri, yapılacak olan atışın durumuna göre gol olan olasılık değerlerinden oluşurken, P(HAYIR) oluşturulurken alınan çarpım değerleri ise yapılacak atışın durumuna göre gol olmayan olasılık değerlerinden meydana gelmektedir.²⁷²

Bayes teoremine göre sadece her bir özelliğin çarpılmasıyla oluşan olasılık kendi başına yeterli olmamaktadır. Bu olasılıklar verilen bir sınıfın toplam olasılığı da hesaba katılarak çarpılır;

$$P(\text{EVET}) = 0.0313 \cdot 4/7 = 0,0179$$

$$P(\text{HAYIR}) = 0.0370 \cdot 3/7 = 0,0159$$

Çıkan değerler dikkate alındığında bir sonuç çıkarılacak olursa, bu olasılığı maksimum yapan sınıfı seçmek yeterli olmuştur, buna göre $P(\text{EVET}) > P(\text{HAYIR})$ olduğu için yeni durum "EVET" olarak sınıflandırılmış, bu da olağan verilere bağlı olarak atılan bir şutun gol olacağı sonucunu vermiştir.

Örnek 2;

Bir iş için yapılan başvurular arasından seçmeler yapılmış ve bir eğitim veri kümesi oluşmuştur. Bunun için adayların Eğitim, Yaş ve Cinsiyet bilgileri dikkate alınmıştır.²⁷³

Aşağıda ki örnek veri kümesine göre adayların Eğitim, Yaş ve Cinsiyet bilgilerine göre kabul durumları verilmiştir.

²⁷² Friedman, N. (1998). **The Bayesian structural EM algorithm**. Proc. UAI-98 (pp. 129–138).

²⁷³ Goldberg, K., Roeder, T., Gupta, D., & Perkins, C. (2001). **Eigentaste: Atime collaborative filtering algorithm**. Information Retrieval, 4(2), 133–151

Tablo 3. 20 Adayların Kabul Durumları Veri Kümesi

Aday	Eğitim	Yaş	Cinsiyet	Kabul Durumu
1. Aday	Ortaöğretim	Yaşlı	Erkek	Evet
2. Aday	İlköğretim	Genç	Erkek	Hayır
3. Aday	Yüksek Öğretim	Orta	Kadın	Hayır
4. Aday	Ortaöğretim	Orta	Erkek	Evet
5. Aday	İlköğretim	Orta	Erkek	Evet
6. Aday	Yüksek Öğretim	Yaşlı	Kadın	Evet
7. Aday	İlköğretim	Genç	Kadın	Hayır

Aday	Eğitim	Yaş	Cinsiyet	Kabul Durumu
8. Aday	Ortaöğretim	Orta	Kadın	Evet
9. Aday	Yüksek Öğretim	Orta	Kadın	?

Tabloda bulunan değerler, aday bilgilerine göre kabul edilme durum frekanslarını gösterir. Bu frekanslar daha önceden verilen aday kabul durum veri kümesine göre oluşturulmuştur. Buna göre adaylardan, eğitim durumu İlköğretim olanların 1'i, Ortaöğretim olanların 3'ü ve Yüksek Öğretim olanların 1'i kabul edilmiştir. Yaş durumların bakıldığında, yaşı Orta olanların 3'ü, Yaşlı olanların 2'si kabul edilirken, Genç adaylardan kabul işlemi olmamıştır. Cinsiyet durumlarına göre kabul durumlarına bakıldığında ise, Erkek adayların 3'ü kabul edilirken, Bayan adayların 2' si kabul edilmiştir. Ayrıca yapılan 8 başvurudan 5' işe kabul edilmiştir.

Tablo 3. 21 Aday Bilgilerine Göre Kabul Durum Frekansları

Eğitim	EVET	HAYIR	Yaş	EVET	HAYIR	Cinsiyet	EVET	HAYIR	Kabul Durumu	
İlköğretim	1	2	Genç	0	2	Erkek	3	1	Evet	Hayır
Ortaöğretim	3	0	Orta	3	1	Kadın	2	2	5	3
Yüksek Öğretim	1	1	Yaşlı	2	0					

Aşağıdaki tablo toplanan bu frekanslara göre oluşturulmuş kabul durum olasılıklarını içermektedir. Bu olasılıklar, Eğitim Durumundan çıkan dağılıma göre İlköğretim için 1/5, Ortaöğretim için 3/5 ve Yüksek Öğretim için 1/5 şeklindedir. Yaşa göre çıkan dağılıma bakıldığında, Genç adaylar için 0/5, Orta yaşta ki adaylar için 3/5 ve Yaşlı adaylar için ise 2/5 şeklindedir. Cinsiyete göre bakıldığında ise bu olasılık oranları, Erkek adaylar için 3/5, Kadın adaylar için ise 2/5 değerlerini vermektedir.

Tablo 3. 22 Aday Bilgi Frekanslarına Göre Kabul Durum Olasılıkları

Eğitim	EVET	HAYIR	Yaş	EVET	HAYIR	Cinsiyet	EVET	HAYIR	Kabul Durumu	
İlköğretim	1/5	2/3	Genç	0/5	2/3	Erkek	3/5	1/3	Evet	Hayır
Ortaöğretim	3/5	0/3	Orta	3/5	1/3	Kadın	2/5	2/3	5/8	3/8
Yüksek Öğretim	1/5	1/3	Yaşlı	2/5	0/3					

Bu frekans ve olasılıklar eşliğinde yeni adayın aday bilgilerine göre kabul edilip edilmeyeceği olasılığı hesaplanabilir. Yeni adayın eğitim durumu "Yüksek Öğretim", yaşı "Orta" ve cinsiyeti "Kadın" şeklindedir. Bu bilgilere göre yeni adayın kabul durumu hesaplanmıştır.

Eğitim	Yaş	Cinsiyet	Kabul Durumu
Yüksek Öğretim	Orta	Kadın	?

Talim kümesindeki olasılık dağılımına dayanarak (Çizelge 3.6) her bir sınıf için olasılık hesaplanır. Önce her bir özelliğin olasılığı hesaba katılır, bütün özelliklere eşdeğer önemde ele alınır ve olasılıklar çarpılır [6].

$$P(\text{EVET}) = 1/5 \cdot 3/5 \cdot 2/5 = 0,048$$

$$P(\text{HAYIR}) = 1/3 \cdot 1/3 \cdot 2/3 = 0,074$$

Yukarıda çıkan sonuçlar, yeni adayın bilgilerine göre kabul edilme olasılıklarının çarpılmasıyla hesaplamıştır. Burada P(EVET) oluşturulurken alınan çarpım değerleri, adayın bilgilerine göre kabul edilme değerlerinden oluşurken, P(HAYIR) oluşturulurken alınan çarpım değerleri ise aday bilgilerine kabul edilmeme olasılık değerlerinden meydana gelmektedir.²⁷⁴

Bayes teoremine göre sadece her bir özelliğin çarpılmasıyla oluşan olasılık kendi başına yeterli olmamaktadır. Bu olasılıklar verilen bir sınıfın toplam olasılığı da hesaba katılarak çarpılır;

$$P(\text{EVET}) = 0,048 \cdot 5/8 = 0,03$$

$$P(\text{HAYIR}) = 0,074 \cdot 3/8 = 0,028$$

Çıkan değerler dikkate alındığında bir sonuç çıkarılacak olursa, bu olasılığı maksimum yapan sınıfı seçmek yeterli olmuştur, buna göre $P(\text{EVET}) > P(\text{HAYIR})$ olduğu için yeni durum "EVET" olarak sınıflandırılmış, bu da olağan verilere bağlı olarak başvuru yapan adayın işe kabul edileceği sonucunu vermiştir.

²⁷⁴ Jourdan, L., Dhaenens, C., Talbi, E., , "A Genetic Algorithm For Feature Selection in Data-Mining For Genetics", 4th Metaheuristics International Conference, 2001, s 29-33

3.2.2 Birliktelik Kuralları

Birliktelik kuralları, büyük veri kümeleri arasında birliktelik ilişkileri bulurlar.²⁷⁵ Eldeki verilerin her geçen gün artmasından dolayı veri sahipleri bu veriler içindeki ikili ilişkileri ortaya çıkarmak istemektedirler. Birliktelik kuralları ilk kez Agrawal tarafından (1993) ortaya atılmıştır. Birliktelik kurallarının kullanıldığı en tipik örnek sepet analizi uygulamasıdır. Bu işlem, müşterilerin yaptıkları alışverişlerdeki ürünler arasındaki birliktelikleri bularak müşterilerin satın alma alışkanlıklarını analiz eder. Bu tip birlikteliklerin keşfedilmesi, müşterilerin hangi ürünleri bir arada aldıkları bilgisini ortaya çıkarır ve market yöneticileri de bu bilgi ışığında daha etkili satış stratejileri geliştirebilirler.²⁷⁶ Örneğin bir müşteri makarna satın alıyorsa, aynı alışverişte makarna yanında ketçap veya yoğurt alma olasılığı biliniyorsa raf dizimi buna göre belirlenebilir. Bu tip bir bilgi ışığında rafları düzenleyen market yöneticileri ürünlerindeki satış oranını arttırabilirler. Örneğin bir A ürününü satın alan müşteriler aynı zamanda B ürününü de satın alıyorsa, bu durum aşağıdaki birliktelik kuralı ile gösterilir.²⁷⁷

$$A \Rightarrow B [\text{destek} = \%3, \text{güven} = \%55]$$

Destek ve güven ifadeleri keşfedilen bilginin doğruluğunu ve güvenilirliğini dolayısıyla kullanılabilirliğini gösterirler. %3 oranındaki bir destek değeri, analiz edilen tüm alışverişlerden %3'ünde A ile B ürünlerinin birlikte satıldığını gösterir. %55 oranındaki güven değeri ise A ürününü satın alan müşterilerinin %55'inin aynı alışverişte B ürününü de satın aldığını ortaya koyar. Kullanıcı tarafından minimum destek eşik değeri ve minimum güven eşik değeri belirlenir ve bu değerleri aşan birliktelik kuralları dikkate alınır. Büyük veri tabanlarında birliktelik kuralları bulunurken, şu iki işlem basamağı takip edilir :

²⁷⁵ Joshi 1997

²⁷⁶ Heckerman, D., Geiger, D., & Chickering, D. M. (1995). **Learning Bayesian networks: The combination of knowledge and statist. data.** Machine Learning, 20, 197–243

²⁷⁷ Defalce, I., Cioppa, A.D., Tarantino, E., , “**Discovering Interesting lassification Rules with Genetic Programming**”, Applied Soft Computing, V-1,N-4, 257-269, www.elsevier.com, www.sciencedirect.com , 2001.

- Sık tekrarlanan öğeler bulunur: Bu öğelerin her biri en az, önceden belirlenen minimum destek sayısı kadar sık tekrarlanırlar.
- Sık tekrarlanan öğelerden güçlü birliktelik kuralları oluşturulur: Bu kurallar minimum destek ve minimum güven değerlerini karşılamalıdır.

Ardışık zamanlı örüntüler ise birbirleri ile ilişkisi olan ancak birbirini izleyen dönemlerde gerçekleşen ilişkilerin tanımlanmasında kullanılır. Örneğin "X ameliyatı yapıldığında, onbeş gün içinde %45 ihtimalle Y enfeksiyonu oluşacaktır" ilişkisi ardışık zamanlı örüntüye bir örnektir.

Birliktelik kuralları, veri madenciliğinin en çok araştırma yapılan alanlarından birisidir. Birliktelik kurallarının en çok kullanıldığı sepet analizi çalışmaları ile pazarlama ve perakende sektöründeki yararları kanıtlanmıştır.²⁷⁸ Birliktelik kuralları madenciliği ilk olarak (Agrawal et al 1993)'te öne sürülmüştür. Birliktelik kuralları, büyük veri yığınları arasındaki ilginç birliktelikleri ya da birliktelik ilişkilerini keşfeden bir veri madenciliği tekniğidir.²⁷⁹

Birliktelik kuralları, büyük veri kümeleri arasında birliktelik ilişkilerini bulurlar. Birliktelik kurallarının kullanıldığı en tipik örnek market sepeti uygulamasıdır. Bu işlem, müşterilerin yaptıkları alışverişlerdeki ürünler arasındaki birliktelikleri bularak müşterilerin satın alma alışkanlıklarını analiz eder. Bu tip birlikteliklerin keşfedilmesi, müşterilerin hangi ürünleri bir arada aldıkları bilgisini ortaya çıkarır ve market yöneticileri de bu bilgi ışığında daha etkili satış stratejileri geliştirebilirler.

Veriler arasındaki ilişkilerin ortaya çıkarılması karar destek sistemlerin oluşturulmasında çok önemlidir. Karar verme işlemlerinde verilerin ilişkileri ortaya konulursa daha etkin sonuçların çıkması sağlanmış olur. Reyonların diziliminin pazara uygun hale getirilmesi, kampanyaların planlanması, pazarlama stratejilerinin

²⁷⁸ Dunham et al 2000

²⁷⁹ Han ve Kamber 2001

geliştirilmesi, genetik biliminin daha hızlı gelişmesi hep bu kuralların tespiti ile doğru orantılıdır. Pazarlama stratejileri belirlenirken, birlikte satılan ürünlere ait sonuçlardan yararlanılabilir.²⁸⁰

Birliktelik kurallarının en sık uygulandığı alan pazarlama alanı, en çok kullanıldığı çalışma ise alışveriş sepeti analizidir. Aynı alışveriş sepeti içinde birlikte satılan ürünlerin tespiti o ürünlerin reyonlarının yan yana konması için önemlidir. Bu şekilde diğer alışverişlerin içinde bu satışların artması hedeflenir.

Müşterilerin alışveriş alışkanlıklarının öğrenilmesi şirketlerin kar oranını arttırması için çok önemlidir. Bu birliktelikler dönem dönem değişebilir. Alışkanlıkların dönemsel farklılıklarının tespiti dönemsel kampanyaların planlanmasından üretime kadar bir çok alanda etkili olur. Birliktelikleri tespit edilen ürünlerin yakın raflarda bulunması onlardan herhangi birini alan bir müşterinin diğerini alma ihtimali yükseltir. Çünkü müşteri raf boyunca başka ürünlere bakarak ilerler ve böylece bunları satın alma olasılığı doğar (Han ve Kamber 2001).

Örneğin büyük bir alışveriş merkezinin alışveriş işlemlerinin bulunduğu veritabanına sahip olduğumuzu varsayalım. Bu veritabanında da Muz ^ Süt şeklinde bir birliktelik kuralının olduğunu kabul edelim. Muz ve Süt müşterilerin aldıkları ürünlerdir. Birliktelik kuralı, "Muz satın alan müşteriler %3 olasılıkla süt de satın alırlar ve tüm satış hareketlerinin %2'si muz ve süt içermektedir" anlamına gelmektedir. Buradaki "%2" değeri desteği (support), "%3" değeri ise güveni (confidence) temsil etmektedir. Başka bir örnek verecek olursak , kural olarak "Sigara alanların %80'i kibrit de almaktadır" verilebilir. Birliktelik kuralları, "X ne ile beraber en çok satılır?" şeklindeki sorulara cevap bulunabilmesini sağlar.

²⁸⁰ Han ve Kamber2001

3.2.2.1 Birliktelik Kurallarında Kullanılan Terimler

Birliktelik kuralları probleminin biçimsel tanımlanması aşağıdaki gibi yapılabilir.²⁸¹

Tanım 1 : $I = \{I_1, I_2, \dots, I_m\}$ veri tabanındaki farklı elemanların kümesi olarak gösterilir. D, herbiri birincil bir anahtara sahip olan (TID) ve T olarak adlandırılan kayıtların oluşturduğu veri tabanı olup $T \subseteq I$ olacak şekilde eleman kümelerini içerir. X ve Y'nin $X \subseteq I$, $Y \subseteq I$ ve $X \cup Y = f$ koşullarını sağlayacak şekilde eleman kümeleri olduğu varsayılırsa birliktelik kuralı $X \wedge Y$ şeklinde gösterilebilir.

Tanım 2 : Bir birliktelik kuralının destek değeri (s), X u Y'yi birlikte içeren kayıt sayısının veri tabanındaki toplam kayıt sayısına oranıdır. Örneğin bir kuralın desteğinin %5 olması, veri tabanındaki tüm kayıtların %5'inin X u Y'yi içermesi anlamına gelir. Destek değeri bir kuralın istatistiksel anlamını belirtir. Genellikle birliktelik kuralları için yüksek destek değerleri cazip olsa da bu her zaman mümkün olmamaktadır. Örneğin iletişim ağlarındaki bir hatayı tahmin etmeye yönelik bir uygulamada, hata öncesi gerçekleşen olaylar ve bunların birliktelikleri kısıtlı sayıda olsa bile, destek değeri düşük olan bu birliktelikler hatayı bulabilmek için göz ardı edilmemelidir. Destek aşağıdaki şekilde formülize edilebilir :

$$\text{Destek } (X \wedge Y) = P(X \cup Y)$$

Tanım 3 : Bir birliktelik kuralının güven değeri (c), X u Y'yi birlikte içeren kayıt sayısının X'i içeren kayıt sayısına oranıdır. Örneğin bir birliktelik kuralının güven değerinin %85 olması X içeren kayıtlarının %85'inin aynı zamanda Y'yi de içermesi anlamına gelir. Güven değeri veri kümesi içerisinde X ve Y arasındaki bağıntının derecesine karşılık gelir. Güven değeri kuralın güçlülüğünün ölçüsüdür. Birliktelik kuralları için genellikle yüksek güven değeri gereklidir. Eğer bir ürün

²⁸¹ Agrawal, R. and Srikant, R., "Fast Algorithms for Mining Association Rules in Large Databases", Proceedings of the Twentieth International Conference on Very Large Databases, 1994., Cheung et al 1996, Han ve Kamber 2001

nadiren başka bir ürünle beraber satılıyorsa, yani güven değeri düşükse, bu kural yararlı bir kural değildir.

Güven değeri aşağıdaki gibi formülize edilebilir :

$$\text{Güven } (X \wedge Y) = P(Y / X)$$

Veri tabanlarından birliktelik kuralları elde etme işlemi, kullanıcı tarafından tanımlanmış minimum destek ve minimum güven olarak adlandırılan eşik destek ve güven değerini sağlayan tüm kuralların bulunmasını içerir. Eşik değerlerini sağlayan kurallara güçlü kurallar denir. Genel olarak bu değerler sayısal 0-1 aralığından ziyade %0 - %100 yüzde aralığı olarak ifade edilmektedir.

Birliktelik Kuralları Madenciliği süreci iki alt sınıfa ayrılabilir.²⁸²

Tüm sık geçen nesne kümelerinin bulunması

Bu aşamada tanımlı olan destek değerini kullanarak bu değeri sağlayan nesne kümelerinin bulunması gerçekleştirilir. Eşik değerini aşan nesne kümeleri "sık geçen" ya da "büyük" nesne kümeleri, eşik değerinin altında kalan nesne kümeleri ise "küçük" nesne kümeleri olarak adlandırılırlar. Eğer X veri tabanından elde edilmiş küçük bir nesne kümesi ise X'i içeren üst kümeler de küçük nesne kümeleridir denilebilir. Bunun tam tersi de doğrudur. Yani X büyük bir nesne kümesi ise X'in alt kümeleri de göz önüne alınması gereken büyük nesne kümeleri olabilir.

Sık geçen nesne kümelerinden güçlü birliktelik kurallarının bulunması

Bu aşamada, birinci aşamada bulunan büyük nesne kümelerini ve önceden tanımlı güven değerini kullanarak güçlü birliktelik kuralları bulunur.

Büyük ölçekli veri tabanlarında sık geçen nesne kümeleri bulunması işlemi, birliktelik kuralları bulma sürecinin en zahmetli ve masraflı bölümü olduğundan çoğu araştırma, sürecin birinci aşamasını çözmeye yönelik verimli algoritmalar

²⁸² Agrawal ve Srikant 1994, Han ve Kamber 2001

geliştirmeye odaklanmıştır. Bu kısımda algoritmaların başlıcaları ve bu tezde kullanılan algoritma olan Apriori algoritması açıklanmıştır.²⁸³

3.2.2.2 Apriori Algoritması

Apriori algoritması 1994 yılında ilk kez Agrawal ve Srikant tarafından öne sürülmüştür ve en bilinen birliktelik kuralı algoritmasıdır. Algoritmanın ismi, bilgileri bir önceki adımdan alması sebebiyle, İngilizce "prior" (önceki) kelimesinden gelmektedir.²⁸⁴ Bu algoritma, sık geçen bir nesne kümesinin herhangi bir alt kümesinin de sık geçen bir nesne kümesi olacağı esasını teknik olarak kullanır. Apriori algoritması, birliktelik kuralları oluşturmak için veri tabanı üzerinde bir çok kez tarama gerçekleştirir.

Veri tabanı üzerindeki ilk taramada, sadece bir elemanlı nesne kümeleri sayılır. İlk taramada elde edilen sık geçen nesne kümeleri, ikinci tarama için aday nesne kümeleri oluşturma amacıyla kullanılır. Aday nesne kümeleri bulununca, veri tabanı taranarak, iki elemanlı sık geçen nesne kümeleri bulmak için, bu kümelerin destek değerleri hesaplanır. Üçüncü taramada, ikinci taramadan elde edilen sık geçen nesne kümeleri bu tarama için aday nesne kümeleri olarak kullanılır. Bu iteratif süreç, hiç sık geçen nesne kümesi bulunamayınca kadar sürer. Her i . taramada algoritma veri tabanını dolaşır ve i - elemanlı sık geçen nesne kümeleri belirler. L_i , i - elemanlı sık geçen nesne kümelerini, C_i ise i -elemanlı aday nesne kümelerini temsil eder. Apriori algoritmasının diğer birliktelik kuralı algoritmalarından olan AIS ve SETM algoritmalarından farkı, aday nesne kümeleri oluşturma ve aday nesne kümelerini saymak için seçme yöntemlerinde görülür. AIS ve SETM algoritmalarında, bulunan ortak nesne kümeleri aday nesne kümeleri oluşturmak üzere yeni bir hareketteki (transaction) farklı nesnelerin herbiriyle genişletilir. Ancak, bu işlem sırasında, genişletilme yapılırken kullanılan nesnelerin sık geçip geçmediği göz önüne alınmaz. Sık geçen bir nesne kümesi ile sık geçmeyen bir nesne kümesinin birleşiminden oluşan bir kümenin sık geçmeyen bir nesne kümesi

²⁸³ Agrawal ve Srikant 1994, Cheung et al 1996, Han ve Kamber 2001

²⁸⁴ Han ve Kamber 2001, Dunham et al 2000

olduğu kuralına dayanarak AIS ve SETM algoritmalarının, bu kuralı göz önüne almayan yapıları nedeniyle, aslında sık geçmeyen bir çok aday nesne kümesi oluşturduğu söylenebilir. Apriori algoritması, bu sorunu çözen bir mimariye sahiptir. Algoritma, aday nesne kümeleri oluştururken, bir önceki adımın sık geçen nesne kümelerini kendi aralarında birleştirip, veri tabanındaki hareketleri dikkate almadan, sık geçmeyen alt kümeleri silen bir yapı kullanır. Sadece bir önceki adımda bulunan nesne kümelerinden sık geçenler göz önüne alınarak, sık geçen aday nesne kümelerinin sayısı önemli ölçüde azalır.

Aşağıda Apriori algoritmasının aşamaları ve ürettiği sonuçlar bir örnek üzerinde incelenmiştir.²⁸⁵

Birleştirme Adımı :

L_k kümesini bulmak için L_{k-1} kendi arasında birleştirilerek, C_k adı verilen k -elemanlı aday nesne kümeleri üretilir. l_1 ve l_2 , L_{k-1} deki iki nesne olsun. $l_i[j]$ notasyonu, l_i 'deki j . nesneyi temsil eder. Birleştirme işlemi $L_{(k-1)} \wedge L_{(k-1)}$, ancak $L_{(k-1)}$ kümesinin elemanlarının ilk $(k-2)$ nesneleri ortak olması durumunda gerçekleştirilebilir.

Budama Adımı :

C_k , L_k kümesinin bir kapsayan kümesidir, öyle ki bu aday kümenin elemanları sık geçen olabilirler veya olmayabilirler ancak yine de k -elemanlı nesne kümelerinin hepsi C_k 'da yer alır. C_k kümesindeki her elemanın veri tabanında geçiş sayısının bulunması ile L_k tespit edilir. Veri tabanında aday kümedeki elemanlardan minimum destek değerine eşit veya destek değerinden daha fazla sayıda bulunanlar kavramsal olarak "sık geçen" kümelerdir ve bu yüzden L_k kümesine aittir. Ancak C_k çok büyük boyutlarda olabilir ve taranması yüksek maliyetli işlemler gerektirebilir. Bu yüzden C_k nesne kümesinin boyutu yeniden düzenlenmelidir. Bunun için, " k -

²⁸⁵ Han ve Kamber 2001

elemanlı aday kümenin $k-1$ elemanlı alt kümeleri L_{k-1} kümesinde mevcut değilse, bu eleman kümesi sık geçen olamaz ve C_k kümesinden çıkarılabilir" özelliği (apriori özelliği) kullanılır. Bu işlem sonucunda büyük boyuttaki C_k kümesi daha küçük ve işlem yapılması daha az maliyetli bir nesne kümesine dönüşür.

Tablo 3.23 'de Apriori algoritmasının uygulanacağı test verileri bulunmaktadır. Tablo 3.23' deki verilere göre Apriori algoritmasının işleyişi adım adım açıklanmıştır.²⁸⁶

Tablo 3. 23 Ürünlere ait ID'ler ve Birlikte Satın Alınma Durumları

TID (Hareket ID si)	Eleman Kümeleri
T100	I1,I2,I5
T200	I2,I4
T300	I2,I3
T400	I1,I2,I4
T500	I1,I3
T600	I2,I3
T700	I1,I3
T800	I1,I2,I3,I5
T900	I1,I2,I3

Kaynak: R. Agrawal, R. Srikant: "Fast Algorithms for Mining Association Rules", Proc. of the 20th Int'l Conference on Very Large Databases, Santiago, Chile, Sept. 1994. Expanded version available as IBM Research Report RJ9839, June 1994.

²⁸⁶ Han ve Kamber 2001

- Algoritmanın 1. adımında, her nesne 1-elemanlı aday kümesi olan C_1 kümesinin bir elemanıdır. Algoritma her nesnenin tüm hareketlerde kaç defa geçtiğini bulmak için, tüm tabloyu tarar.
- Minimum destek değerinin 2 olarak belirlendiği varsayılın. Bu durumda yüzde olarak minimum destek değeri $2/9$ 'dan %22 olarak bulunur. Bundan sonra 1-elemanlı sık geçen nesneleri içeren L_1 kümesi belirlenebilir. Minimum destek değerine eşit veya destek değerinden yüksek desteğe sahip olan eleman kümeleri L_1 kümesini oluştururlar.
- Sık geçen 2-elemanlı nesne kümelerini içeren L_2 kümesini belirlemek üzere algoritma $L_1 \wedge L_1$ kartezyen çarpımını kullanarak yeni bir aday küme oluşturur. C_2 olarak adlandırılan bu aday küme L_1 'deki elemanların ikili kombinasyonlarından oluşur.
- Daha sonraki adımda, veri tabanındaki her hareket taranarak C_2 kümesindeki her nesne kümesinin destek değeri hesaplanır ve minimum destek değerinden yüksek olan değerlere sahip adaylardan L_2 kümesi elde edilir.
- C_3 aday kümesinin elde edilmesi şu şekilde gerçekleşir:
- $C_3 = L_2 \times L_2 = \{\{I1, I2, I3\}, \{I1, I2, I5\}, \{I1, I3, I5\}, \{I2, I3, I4\}, \{I2, I3, I5\}, \{I2, I4, I5\}\}$
- Ancak C_3 kümesi, yukarıdaki gibi 6 elemandan değil, yalnızca 2 elemandan oluşmaktadır. Bunun sebebi, kümedeki 3-elemanlı kümelerden son 4 tanesinin ikili alt kümelerinin hepsinin L_2 kümesinde bulunmaması, bu yüzden apriori özelliğine göre bu kümelerin C_3 aday kümesinden atılmasıdır. Burada dikkat edilmesi gereken bir başka durum ise şudur : C_3 aday kümesine ait alt kümeler incelenirken sadece 2- elemanlı altkümelere bakılır. O halde, k. seviyedeki aday kümelerdeki nesne alt kümeleri elenirken, o aday kümeye ait k-1 elemanlı alt kümelere bakılır. Bunun sebebi Apriori algoritmasının seviye tabanlı bir algoritma olmasıdır. Bu işlemin detaylı açıklaması aşağıda açıklanmıştır.

– $L_2 \propto L_2 = \{ \{I1, I2\}, \{I1, I3\}, \{I1, I5\}, \{I2, I3\}, \{I2, I4\}, \{I2, I5\} \propto$

$\{I1, I2\}, \{I1, I3\}, \{I1, I5\}, \{I2, I3\}, \{I2, I4\}, \{I2, I5\} \} = \{$

$\{I1, I2, I3\}, \{I1, I2, I5\}, \{I1, I3, I5\}, \{I2, I3, I4\}, \{I2, I3, I5\}, \{I2, I4, I5\} \}$

–Diğer adımda ise ise sık geçen bir nesne kümesinin tüm alt kümeleri de sık geçmelidir.

– $\{I1, I2, I3\}$ kümesinin 2-elemanlı alt kümeleri $\{I1, I2\}$, $\{I1, I3\}$ ve $\{I2, I3\}$ kümeleridir. Bu kümelerin hepsi L_2 kümesinde olduğundan $\{I1, I2, I3\}$ kümesi C_3 kümesinde tutulur.

– $\{I1, I2, I5\}$ kümesinin 2-elemanlı alt kümeleri $\{I1, I2\}$, $\{I1, I5\}$ ve $\{I2, I5\}$ kümeleridir. Bu kümelerin hepsi L_2 kümesinde olduğundan $\{I1, I2, I5\}$ kümesi C_3 kümesinde tutulur.

– $\{I1, I3, I5\}$ kümesinin 2-elemanlı alt kümeleri $\{I1, I3\}$, $\{I1, I5\}$ ve $\{I3, I5\}$ kümeleridir. $\{I3, I5\}$ kümesi L_2 kümesinde olmadığından sık geçen bir küme değildir. Bu yüzden $\{I1, I3, I5\}$ kümesi C_3 kümesinden atılır.

– $\{I2, I3, I4\}$ kümesinin 2-elemanlı alt kümeleri $\{I2, I3\}$, $\{I2, I4\}$ ve $\{I3, I4\}$ kümeleridir. $\{I3, I4\}$ kümesi L_2 kümesinde olmadığından sık geçen bir küme değildir. Bu yüzden $\{I2, I3, I4\}$ kümesi C_3 kümesinden atılır.

– $\{I2, I3, I5\}$ kümesinin 2-elemanlı alt kümeleri $\{I2, I3\}$, $\{I2, I5\}$ ve $\{I3, I5\}$ kümeleridir. $\{I3, I5\}$ kümesi L_2 kümesinde olmadığından sık geçen bir küme değildir. Bu yüzden $\{I2, I3, I5\}$ kümesi C_3 kümesinden atılır.

– $\{I2, I4, I5\}$ kümesinin 2-elemanlı alt kümeleri $\{I2, I4\}$, $\{I2, I5\}$ ve $\{I4, I5\}$ kümeleridir. $\{I4, I5\}$ kümesi L_2 kümesinde olmadığından sık geçen bir küme değildir. Bu yüzden $\{I2, I4, I5\}$ kümesi C_3 kümesinden atılır.

–Bu yüzden budama sonrasında $C_3 = \{ \{I1, I2, I3\}, \{I1, I2, I5\} \}$ olarak elde edilir.

– C_3 aday kümesindeki kayıtların destek değerleri veri tabanındaki hareketler taranarak minimum destek değeri ile kıyaslandıktan sonra L_3 3-elemanlı sık geçen nesne kümesi elde edilir.

–Algoritma, bir sonraki adımda C_4 aday kümesini belirlemek üzere $L_3 \wedge L_3$

kartezyen çarpımını kullanır. Birleştirme işlemi sonunda {I1,I2,I3,I5} kümesi elde edilmesine rağmen, {I2,I3,I5} L₃ kümesinde yer almadığından bu nesne kümesi budanır ve C₄=[^] olacağından algoritma sona erer. Böylece, birlikte en sık olarak bulunan nesne kümeleri belirlenmiş olur.

Veri tabanındaki sık geçen nesne kümeleri bulunduktan sonra bu kümelerden minimum destek ve güven değerlerini sağlayan güçlü birliktelik kuralları üretmek mümkündür. Bunun için aşağıdaki güven hesaplama formülü kullanılır.

$$\text{Güven}(A \wedge B) = P(B/A) = \text{destek_değeri}(A \cup B) / \text{destek_değeri}(A)$$

Formül 'de belirtilen destek_değeri(AuB) ifadesi A ile B'yi birlikte içeren hareketlerin sayısını, destek_değeri(A) ifadesi ise A'yı içeren hareketlerin sayısını belirtir. Bu eşitliğe göre birliktelik kuralları aşağıdaki gibi oluşturulabilir

Tüm sık geçen nesne kümeleri için (I), I kümesinin boş olmayan tüm alt kümelerini yarat. Boş olmayan her nesne kümesi için, kural çıktısı destek-değeri(I) / destek-değeri(s) > minimum_güven olması durumunda s ^ (I - s) şeklindedir.

Birliktelik kuralları sık geçen nesne kümelerinden üretildikleri için otomatik olarak herbiri minimum destek değerini sağlar.

Örneğin I={I1, I2, I5} bir sık geçen nesne kümesi olsun. I kümesinin alt kümeleri, {I1,I2}, {I1,I5}, {I2,I5}, {I1}, {I2}, {I5}'dir. Sonuç olarak aşağıdaki birliktelik kuralları beraberinde verilen güven değerleri ile çıkartılır.

$$I1=I2_I5 \text{ güven} = 2/4 = \%50$$

$$I1=I5_I2 \text{ güven} = 2/2 = \%100$$

$$I2=I5_I1 \text{ güven} = 2/2 = \%100$$

$$I1=I2_I5 \text{ güven} = 2/6 = \%33$$

$$I2=I1_I5 \text{ güven} = 2/7 = \%29$$

$$I5=I1_I2 \text{ güven} = 2/2 = \%100$$

Minimum güven değeri %70 olarak belirlenmişse, yalnızca 2, 3 ve sonuncu kurallar güçlü birliktelik kuralları olarak elde edilirler.²⁸⁷

Apriori algoritması, aday nesne kümeleri oluştururken minimum destek değerini sağlayamayan nesne kümelerini göz ardı ettiğinden bunun zıttı bir mantıkla çalışan AIS ve SETM algoritmasına göre daha efektif çalışır²⁸⁸

Ayrıca Apriori algoritması, k. geçişte elde edilen sık geçen nesne kümelerinin (L_k-1) ve aday nesne kümelerinin (C_k) belleğe sığmayabileceği ihtimalini göz önüne alarak tampon bellek yönetimini de kullanır. (Agrawal ve Srikant 1994)'te belirli bir aşamada üretilen nesne kümelerinin belleğe sığmaması durumunda, bu sorunu aşmanın yolları anlatılmıştır.

3.2.3 Kümeleme Analizi

Kümeleme analizi, istatistik, biyoloji, tıp, kimya, makine öğrenme gibi birçok alanda uygulanmaktadır. Veritabanlarında toplanan çok fazla sayıdaki verilerden dolayı, kümeleme analizi son yıllarda veri madenciliği araştırmalarında da yaygın olarak kullanılmaktadır. Bilindiği gibi kümeleme analizi en basit şekli ile verileri benzer özelliklerine göre sınıflara veya kümelere ayırma sürecidir.

Kümeleme analizi denetimsiz öğrenme (unsupervised learning) yöntemidir. Denetimli öğrenmede, bir denetçi tarafından ilgili sınıflar önceden belirlenen bir kritere göre ayrılmakta ve her sınıf için çeşitli örnekler verilmektedir. Böylece verilen örneklerden hareketle her sınıfa ilişkin özellikler bulunmakta ve bu özellikler kural cümleleri ile ifade edilmektedir. Süreç tamamlandığında, tanımlanan kural cümleleri verilen yeni örneklerle uygulanmakta ve bu örneklerin hangi sınıfa ait olduğu kurulan model tarafından belirlenmektedir. Denetimsiz öğrenme yönteminde ise, ilgili örneklerin gözlenmesi ve bu örneklerin özellikleri arasındaki benzerliklerden hareketle sınıfların tanımlanması yapılmaktadır²⁸⁹. Kısaca denetimsiz öğrenmede bağımlı değişken olmadığından model kurmak için bağımsız

²⁸⁷ Han ve Kamber 2001

²⁸⁸ Agrawal ve Srikant 1994

²⁸⁹ Akpınar, s.7.

değişkenlerden yararlanılmaktadır. Ayrıca denetimsiz öğrenmede; önceden belirlenmiş sınıflar olmayıp model verilerden üretilmektedir. Yani, nesnelerin kümelere atanması için önceden belirlenmiş kurallar veya bir fonksiyon ya da bir model yoktur.

Kümeleme analizi, pek çok alanda geniş ölçüde uygulanmaktadır. Bu alanlar arasında, örüntü tanıma, veri analizi, pazarlama araştırmaları, metin madenciliği sayılabilir. Ayrıca bölgeleri yoğunluklarına göre tanımlayarak dağılımlarının yapılarını ve özellikleri arasındaki ilişkileri bulmak yine kümeleme analizi yardımı ile mümkün olmaktadır.

3.2.4 Kümeleme Yaklaşımları

Kümeleme, birincil amacı özelliklerine göre nesneleri gruplamak olan çok değişkenli bir analiz tekniğidir. Kümelemede nesnelerin gruplanması önceden belirlenmiş bir kritere göre gerçekleştirilir. Kümeleme sonuçları, kümeler içinde yüksek derecede homojenlik, kümeler arasında ise yüksek derecede heterojenlik göstermelidir. Kümeleme yapısı içerisinde rastlantı değişkeni kavramının önemli bir yeri vardır. Genellikle kümeleme denetimsiz bir biçimde gerçekleşir. Denetimsiz öğrenme yetisinden dolayı, veriler içinde gizli olan örüntüleri ortaya çıkarmayı sağlar. Bu durum kümelemeyi, örüntü tanıma, biyoinformatik ve veri madenciliği gibi çeşitli alanların önemli bir araştırma konusu haline getirmiştir.²⁹⁰

3.2.4.1 Hiyerarşik Kümeleme Yaklaşımı

Hiyerarşik kümeleme, dendogram olarak bilinen kümelerin ağaç yapısını veya diğer bir deyişle bir küme hiyerarşisi ortaya çıkarır. Her küme düğümü, ortak ebeveynlere sahip noktaların çocuk ve kardeş küme bölünmelerini içerir. Böyle bir yaklaşım, farklı seviyelerdeki düğümlerin dikkatle incelenmesine olanak sağlar. Hiyerarşik kümeleme metotları toplanmış (agglomerative) ve bölen (divisive)

²⁹⁰ MA ve CHOW, s. 503.

metotları olarak sınıflandırılabilir²⁹¹. Toplanmış kümeleme tek noktadan (singleton) oluşan kümelerden başlar ve bu kümeleri tekrarlı olarak en uygun kümeleri şekilde, 2 veya daha fazla kümede birleştirir. Bölen kümeleme ise bütün veri noktalarını içeren tek bir kümeden başlar ve bu kümeleri en uygun biçimde tekrarlı olarak parçalara ayırır. Bu süreç bitirme kriteri sağlanana kadar devam eder. Hiyerarşik kümelemenin avantajları, düğümlerin seviyesine ilişkin esneklik özelliğinin olması, benzerlik ve uzaklık biçimlerini ele almanın kolaylığı, çeşitli niteliklere uygulanabilme özelliğinin olmasıdır.

Hiyerarşik kümelemenin dezavantajları ise, bitirme kriterinin belirsizliği, gerçekte çoğu hiyerarşik kümeleme algoritması, geliştirilme amaçlarına göre arada bulunan kümeler (dendogramda) tekrar ulaşılmasına izin vermez.

3.2.4.2 Toplanmış Hiyerarşik Kümeleme Yöntemleri

Başlangıçta tüm birimlerin ayrı küme oluşturduğunu kabul ederek, n birimi hiyerarşik olarak sırasıyla $n, n-1, \dots, n-r, \dots, 3, 2, 1$ kümeye yerleştirmeyi amaçlayan bir yaklaşımdır.²⁹² Toplanmış (agglomerative) hiyerarşik kümeleme yönteminde, kümeleme sürecinin başlangıcında her birey bir kümedir, süreç sonunda ise tüm bireyler tek bir kümede toplanır. İşleyiş daha ayrıntılı bir şekilde aşağıdaki gibi bir algoritma ile ifade edilebilir.

- n tane birey, n tane küme olmak üzere işleme başlanır.
- En yakın (dtj değeri en küçük olan) iki küme birleştirilir.
- Küme sayısı bir azaltılarak yinelenmiş uzaklıklar matrisi bulunur.
- 2 ve 3 nolu adımlar $n-1$ kez tekrarlanır.²⁹³

4. 2 ve 3 nolu adımlar $n-1$ kez tekrarlanır.

²⁹¹ Kaufman, L., ; Rousseuw, P., "Finding Group Data: An Introduction to Cluster Analysis", John Wiley and Sons, New York, 1990, s. 157

²⁹² Özdamar, K., "Paket Programlar ile Veri Analizi-2", Kaan Kitap evi, Eskişehir, 2004, s.294.

²⁹³ Güler, N., "Bulanık Kümeleme Analizi ve Bulanık Modelleme Uygulamaları", Yüksek Lisans Tezi, Muğla Üniversitesi, F.B.E. İstatistik ve Bilgisayar Bilimleri Ana Bilim Dalı, 2006, s.11.

Bu süreçte birden çok gözlemlili kümenin vektör olarak gösterilebilmesi amacıyla değişkenlerin ortalama değerlerinden yeni vektör oluşturulmakta ya da bu kümedeki tüm gözlemler ile başka kümedeki gözlemlerin uzaklık ortalamaları kullanılabilir.

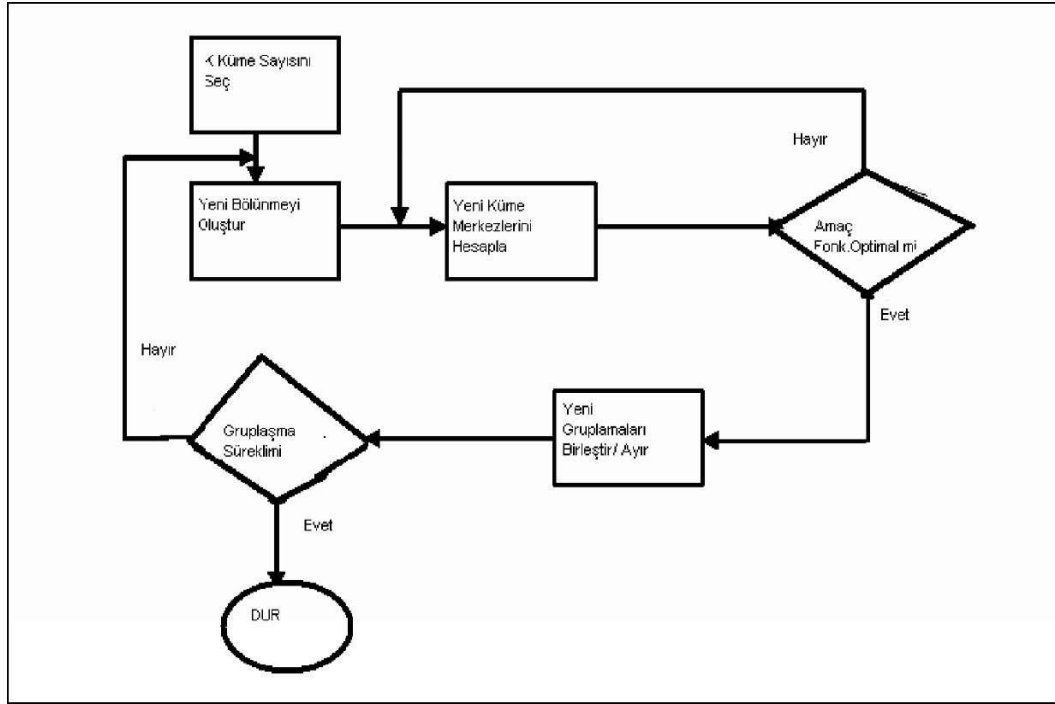
Yukarıda verilen algoritmaya dayalı yedi farklı hiyerarşik teknikten söz edilmektedir. Bu teknikler sırasıyla tek bağlantılı, tam bağlantılı, grup ortalama, merkezi, ortanca, minimum varyans ve Ward teknikleridir. Toplanmış hiyerarşik kümeleme yöntemlerinde, birimlerin birbirleri ile birleştirilmesinde değişik yaklaşımlar uygulanmaktadır. Bu nedenle değişik isimlerle anılan toplanmış hiyerarşik kümeleme yöntemleri bulunmaktadır. Bu yöntemlerden sıklıkla kullanılan ve genel kabul görmüş olanları aşağıdaki gibi sayılabilir.²⁹⁴

3.2.4.3 Bölümlemeli Kümeleme Yaklaşımı

Bölümlemeli kümelemenin temeli, başlangıç bölünmesinden başlayarak, bireyleri yinelemeli olarak kümeleme kriterini azaltacak şekilde kümelere tahsis etmek oluşturur. Yinelemeli bölümlemeli kümelemenin genel algoritması aşağıdaki şekilde görüldüğü gibidir.

²⁹⁴Özdamar, s.296.

²⁹⁴Özdamar, s.293.



Şekil 3. 12 Bölümlemeli Kümelemenin Akış Diyagramı

Kaynak: Y.-M. Cheung, “k star means: A new generalized k-means clustering algorithm,”

2003

Kümeleme algoritmalarının yapılabilmesi için birçok kriter vardır ve bu kriterler şu şekildedir; ²⁹⁵

- Küme Temsil Şeması
- Başlangıç koşulları
- Tahsis etme fonksiyonları
- Birleştir-Ayr koşulları
- En iyilik kriteri

Bölümlemeli kümeleme algoritmaları, yinelemeli yığın yordamı olduğundan dolayı, verilen başlangıç koşulları için global minimumu garanti etmez. Eğer uygulama alanı ile ilgili ön bir bilgi varsa, bu bilgi daha uygun başlangıç koşullarını belirlemek için kullanılabilir. Bununla birlikte özellikleri bilinmeyen bir veri kümesi

²⁹⁵Güler, s.15.

için, bu metot, çok sayıda çalıştırılır ve en sık meydana gelen küme yapısı doğru bölünme olarak kabul edilir.

Mümkün temsil şemaları arasında; kümenin merkezi, kümenin en uçtaki üç bireyi, normal dağılım fonksiyonu, sınıflandırma ağacı, bağlayıcı ifadeler yer alır. Verilen bir küme temsil şeması, küme temsil fonksiyonu, verilen küme için en iyi temsili belirler. Tahsis etme fonksiyonları, temsil etme fonksiyonlarının tersidir: bir küme temsili verildiği zaman her bir kümeye giren bireyleri belirler.²⁹⁶

Küme en iyilik kriteri, kümelerin doğruluğunu belirleyen bir ölçüdür. Bu kriter ayrıca yinelemenin devam edip etmeyeceğini belirtir. Küme en iyilik kriteri bu kümeleme algoritması için oldukça önemlidir. Doğru seçilmediği durumda yanlış kümelerin oluşmasına sebep olacaktır.¹⁹ Bölümlemeli kümeleme algoritmaları, küme temsilcilerinin nasıl oluşturulacağına bağlı olarak, yinelemeli en iyileme bölünme algoritmaları k-medoids ve k-ortalamlar olmak üzere iki gruba ayrılabilir.

3.2.4.3.1 Klasik k-ortalamlar algoritması (KOA)

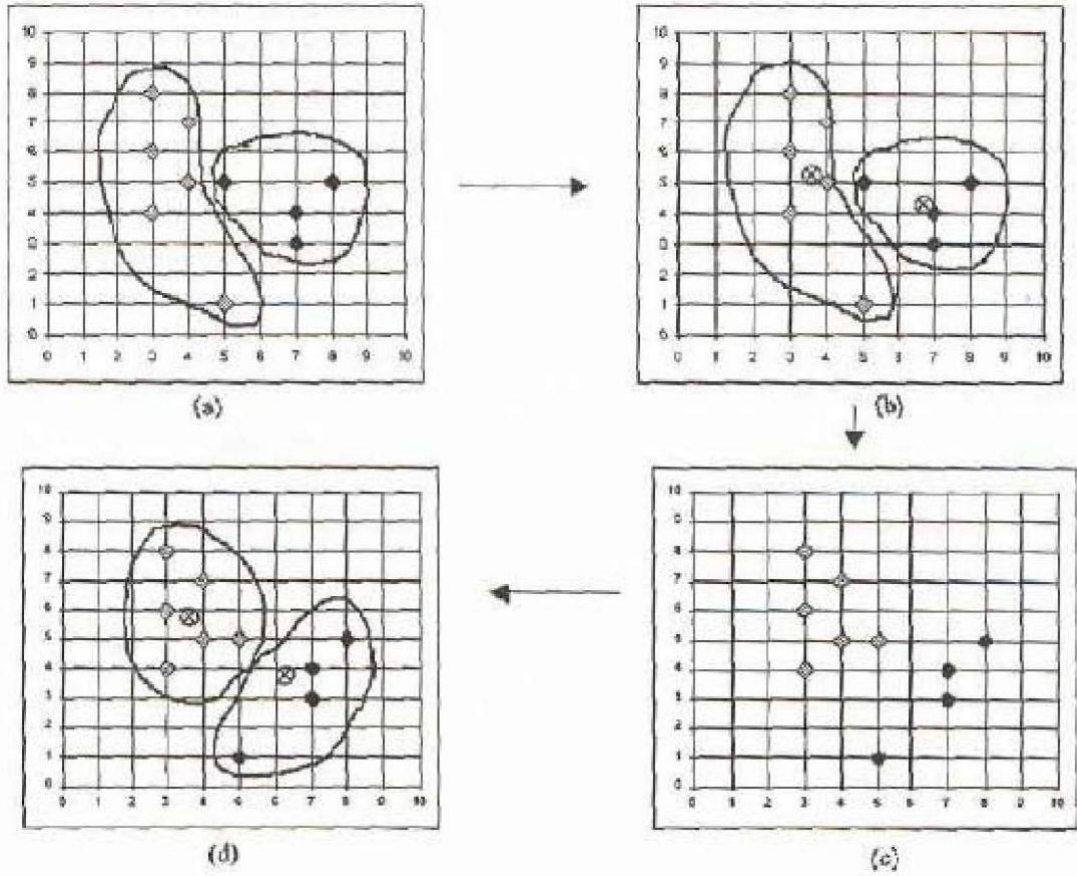
K-ortalamlar algoritması (KOA), 1975 yılında Hartigan tarafından tanımlanan standart kümeleme algoritmasıdır ve hala geniş çaplı kullanılmaktadır. K-ortalamlar algoritması, endüstriyel ve bilimsel anlamda kullanılan en popüler kümeleme araçlarından biridir. Adı, kitle merkezi (centroid) olarak adlandırılan, k tane kümenin her birini temsil eden noktaların ortalamasından gelir. Kesin niteliklerde çok daha iyi çalışmayacağı bellidir. Sayısal nitelikler için iyi bir geometrik ve istatistiksel anlama sahiptir. Nokta ile kitle arasındaki farklılıkların toplamı amaç fonksiyonu olarak kullanılan uygun aşağıdaki uzaklık ölçüsü ile ifade edilir.²⁹⁷

²⁹⁶ Karanfil, S., “Fuzzy Lojik Problemlerinde Üyelik Fonksiyonunun Belirlenmesinde Deneysel Verilere Dayanarak Bir Yöntem Geliştirilmesi”, Doktora Tezi, YTÜ, F.B.E., Matematik Ana Bilim Dalı, 1997

²⁹⁷ Andrew Moore: “K-means and Hierarchical Clustering-Tutorial Slides”<http://www-2.cs.cmu.edu/~awm/tutorials/kmeans.html>, 03, 2009.

$$E(C) = \sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - c_j\|^2$$

Burada x_i : veri noktalarını, c_j : merkezini, k : küme sayısının göstermektedir. K-ortalamalar yöntemi, ilk önce n adet nesneden rasgele k adet nesne seçer ve bu nesnelerin her biri, bir kümenin merkezini veya orta noktasını temsil eder. Geriye kalan nesnelerden her biri kendisine en yakın olan küme merkezine göre kümelere dağılırlar. Yani bir nesne hangi kümenin merkezine daha yakın ise o kümeye yerleşir. Ardından her küme için ortalama hesaplanır ve hesaplanan bu değer o kümenin yeni merkezi olur. Bu işlem tüm nesneler kümelere yerleşinceye kadar devam eder. Bir nesne grubunun, Şekil 15 de görüldüğü gibi uzayda konumlanmış olduğu varsayalım. Kullanıcının bu nesneleri iki kümeye ayırmak istediği varsayılırsa, $k=2$ olur. Şekil 15 'de uzayda ilk önce rasgele iki nesne, iki kümenin merkezi olarak seçilmiş ve diğer nesneler de bu merkezlere olan yakınlıklarına göre iki kümeye ayrılmışlardır. Bu ayırma göre her iki kümenin nesnelerinin yeni ortalaması alınmış ve bu değer kümelerin yeni merkezleri olmuştur. Bu yeni merkezler Şekil 15 (b)'de üstünde çarpı işareti bulunan noktalarla gösterilmektedir. Bu yeni çarpı işaretli merkezlere göre, her iki kümede de birer nesne diğer kümenin merkezine daha yakın duruma gelmişlerdir. Bu durum Şekil 15 (c)'de görülmektedir. (5,1) koordinatındaki nesne ile (5,5) koordinatındaki nesne küme değiştirmişlerdir. Her iki kümedeki bu yeni katılımlar ile kümelerdeki nesnelerin ortalama değerleri ve dolayısıyla merkezleri değişmiştir. Yeni hesaplanan merkezler Şekil 15(d)'de üstünde çarpı işareti bulunan noktalarla gösterilmektedir. Artık açıkta bir nesne kalmadığı ve her nesne içinde bulunduğu kümenin merkezine en yakın durumda bulunduğu için k-ortalamalar yöntemi ile kümelere bölünme işlemi Şekil 15(d)'de görüldüğü gibi sonlanmıştır.



Şekil 3. 13 Bir Nesne Grubunun Uzayda Konumlanmış Hali

k-ortalamlar yöntemi, sadece kümenin ortalaması tanımlanabildiği durumlarda kullanılabilir. Kullanıcıların k değerini, yani oluşacak küme sayısını belirtme gerekliliği bir dezavantaj olarak görülebilir. Esas önemli olan dezavantaj ise dışarıda kalanlar (outliers) olarak adlandırılan nesnelere karşı olan duyarlılıktır. Değeri çok büyük olan bir nesne, dâhil olacağı kümenin ortalamasını ve merkez noktasını büyük bir derecede değiştirebilir. Bu değişiklik kümenin hassasiyetini bozabilir. k-ortalamlar algoritması, varyans analizinin sağlam temeli üzerine kurulmuştur. K- ortalamlar algoritması ile ilgili olağan bazı şüpheler vardır.

- Sonuçlar kitle merkezlerinin başlangıç tahminlerine bağlıdır.
- Küme sayısı k'nın ne olacağı çok açık değildir.

Bu işlem aykırı değerlerin olması durumunda başarılı sonuç vermez.

- Bu algoritma ölçeklenebilirlik özelliğinden yoksundur.
- Sadece sayısal nitelikleri kapsar.
- Meydana gelen kümeler dengeli olmayabilir.²⁹⁸

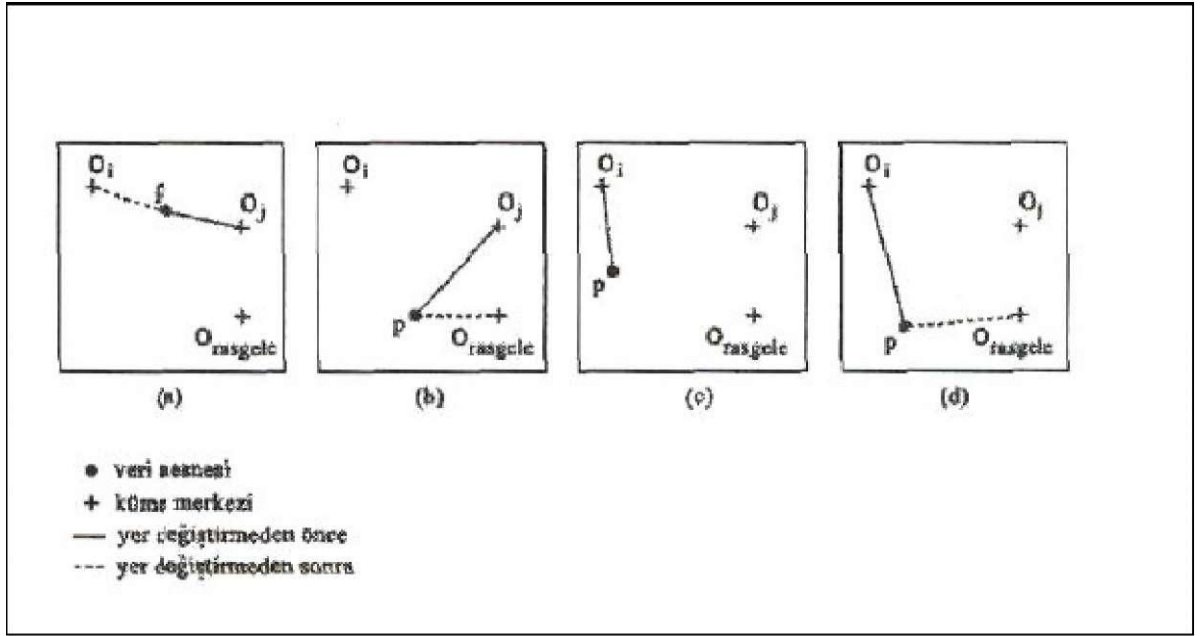
Başlangıç kümelerinin etkilerini azaltmak için basit birkaç yol vardır. İlk olarak, rasgele üretilen birkaç küçük veri örneği üzerinde k-ortalamlar hesaplanır. Bu oluşturulan sistemlerden her biri, tüm örneklerin birleşimi için başlangıç olarak kullanılır. Bu yolla oluşturan en iyi sistemin kitle merkezi, bütün veriler üzerinde k-ortalamlar algoritmasının uygulamak için başlangıç tahmini olarak önerilir.²⁹⁹

3.2.4.3.2 Klasik k-medoid algoritması (KMA)

Değeri çok büyük olan bir nesne, dâhil olacağı kümenin ortalamasını ve merkez noktasını büyük bir ölçüde değiştirebildiğinden kümedeki nesnelerin ortalamasını almak yerine, kümede ortaya en yakın noktada konumlanmış olan nesne anlamındaki medoid kullanılabilir. Bu işlem k-medoids yöntemi ile gerçekleştirilir. k-medoids kümeleme yönteminin temel stratejisi ilk olarak n adet nesnede, merkezi temsili bir medoid olan k adet küme bulmaktır. Geriye kalan nesneler, kendilerine en yakın olan medoide göre k adet kümeye yerleşirler. Bölünmelerin ardından kümenin ortasına en yakın olan nesneyi bulmak için medoid, medoid olmayan her nesne ile yer değiştirir. Bu işlem en verimli medoid bulunana kadar devam eder (URL1). Şekil 2.3'de O_i ve O_j iki ayrı kümenin medoid'lerini, O rasgele seçilen ve medoid adayı olan bir nesneyi, p ise medoid olmayan bir nesneyi temsil etmektedir. Şekil 16'da O rasgele'nin, şu anda medoid olan O_j'nin yerine geçip, yeni medoid olup olamayacağını belirleyen dört durumu göstermektedir.

²⁹⁸ Halkıdı, M., ; Batistakis, Y., ; Cazırgıannis, M., **On Clustering Validation Techniques**, Journal of Intelligent Information Systems, Kluwer Academic Publishers, Manufactred in the Netherlands, 17:2/3, 2001, s.107-145.

²⁹⁹ Güler, s.19.



Şekil 3. 14 O_i ve O_j İki Ayrı Kümenin Medoidleri

Kaynak: A K-means-like Algorithm for K-medoids Clustering and Its Performance Hae-Sang Park*, Jong-Seok Lee and Chi-Hyuck Jun Department of Industrial and Management Engineering, POSTECH San 31 Hyoja-dong, Pohang 790-784, S. Korea

- p nesnesi su anda O_j medoidine bağlıdır (O_j medoidinin bulunduğu kümededir). Eğer O_j , $O_{rasgele}$ ile yer değiştirir ve p O_i 'ye en yakınsa, p nesnesi O_i 'ye geçer.
- p nesnesi su anda O_j medoidine bağlıdır. Eğer O_j , $O_{rasgele}$ ile yer değiştirir ve p $O_{rasgele}$ 'ye en yakınsa, p nesnesi $O_{rasgele}$ 'ye geçer.
- p nesnesi su anda O_i medoid'ine bağlıdır. Eğer O_j , $O_{rasgele}$ ile yer değiştirir ve p hala $O_{rasgele}$ 'ye en yakınsa, p nesnesi yine O_i 'ye bağlı kalır.
- p nesnesi su anda O_i medoid'ine bağlıdır. Eğer O_j , $O_{rasgele}$ ile yer değiştirir ve P $O_{rasgele}$ 'ye en yakınsa, p nesnesi $O_{rasgele}$ 'ye geçer

3.2.4.4 Yoğunluğa Dayalı Kümeleme Yaklaşımı

Yoğunluğa dayalı kümeleme (YDK), tipik olarak kümelere, veri seti içindeki bireylerin yoğun olduğu bölgeler gözüyle bakar ve bu veri setini düşük yoğunluktaki bölgelere bölmeye çalışır. Bu kümelemeye göre noktaları sonlu sayıda kümeye bölmek için, yoğunluk, bağlanabilirlik ve sınır gibi kavramların bilinmesi gerekir. Bu kavramların noktaların en yakın komşularıyla ilgili olduğu açıktır. Yoğunluğa dayalı kümeleme algoritmaları keyfi şekilli kümeleri ortaya çıkarmada başarılıdır. Bu tür kümeleri bulmak diğer birçok kümeleme algoritması için problem olurken, yoğunluğa dayalı kümeleme algoritmaları bu tür kümeleri rahatlıkla ele alır.³⁰⁰

Bu yöntem, ayrıca aykırı gözlemlere karşı doğal bir koruma sağlar. Ancak bu gibi iyi özellikleri olmasına rağmen, bazı rahatsızlıklardan dolayı ölçülü olarak kullanılmaktadır. Yoğunluğa dayalı kümeleme algoritmalarından yaygın olarak kullanılan algoritma, DBSCAN'dır. Bu algoritmanın ana fikri, küme içindeki her nokta (bu noktalara merkez gözüyle bakıldığında) için; verilen bir yarıçap etrafında en az minimum sayıda nokta olmak zorundadır. DBSCAN Ester tarafından önerilen artan (incremental) kümeleme algoritması gibi kullanılır. Yoğunluğa dayalı doğası yüzünden bir bireyin silinmesi veya eklenmesi, sadece bu bireyin komsuları içindeki geçerli kümelemeyi etkiler ve böylece DBSCAN'a dayanan bu algoritmalar ile var olan kümelemeye ekleme ve silme işlemleri yapılabilir.³⁰¹

Başka bir yoğunluğa dayalı kümeleme algoritması ise DENCLUE'dır. Bu algoritma, büyük çoklu ortam veri tabanlarını kümelemek için kullanılan yeni bir yaklaşım sunar. Bu yaklaşımın temel amacı, veri noktalarının etki fonksiyonunun toplamı şeklinde noktaların yoğunluğunu analitik olarak modellemektir. Etki fonksiyonu, bir veri noktasının komşularının içinde etkisini tanımlayan bir fonksiyon olarak görülebilir. Sonra yoğunluk çekiciler (density attractor) belirlenerek kümeler teşhis edilebilir. Yoğunluk çekiciler, toplam yoğunluk fonksiyonunun yerel

³⁰⁰ Babu, G. P., Murty, M. N. 1994. **Clustering with evolution strategies**. Pattern Recogn. 27, 321–329

³⁰¹ Backer, F. B., Hubert, L. J. 1976. **A graphtheoretic approach to goodness-of-fit in complete-link hierarchical clustering**. J. Am.Stat. Assoc. 71, 870 – 878

maksimumudur. Ek olarak, keyfi şekilli kümeler toplam yoğunluk fonksiyonuna dayanan basit bir eşitlik tarafından kolayca tanımlanabilir. DENCLUE algoritmasının ana avantajı, aykırı değerlerin çok miktarda olduğu veri setlerinde özelliklerinin iyi bir şekilde kümelenmesine ve çok boyutlu veri setlerindeki keyfi şekilli kümelerin matematiksel olarak tanımlanmasına olanak sağlamasıdır. Ancak, DENCLUE kümeleme iki tane parametreye dayanır ve kümeleme sonuçlarının kalitesi bu parametrelerin seçimine bağlıdır. Bu parametreler:

- Bir veri noktasının komsularının içindeki etkisini belirleyen N.
- Yoğunluk çekicilerin sayısını azaltmaya izin veren, performansı geliştirmeye yardımcı olan ve yoğunluk çekicilerin anlamlı olup olmadığını tanımlayanıdır.³⁰²

3.2.4.5 Fuzzy (Bulanık) Kümeleme Yaklaşımları

Fuzzy(Bulanık) kümeleme analizi, desen tanıma, görüntü işleme ve bulanık modelleme gibi uygulamalar için etkin bir şekilde kullanılmaktadır. En iyi bilinen bulanık kümeleme yaklaşımı ilk olarak Dunn(1974) tarafından önerilen ve Bezdek tarafından geliştirilen Bulanık C-Ortalamlar(BCO) algoritmasıdır.³⁰³

Fuzzy(Bulanık) kümeleme yöntemi, kümeler birbirinden belirgin bir şekilde ayrılmıyorsa ya da üyeliklerinde bazı birimler küme üyeliğinde kararsızsa uygun bir yöntem olarak ortaya çıkmaktadır. Bulanık kümeler kümedeki birimin üyeliği olarak tanımlanan 0 ile 1 arasındaki her birimi belirleyen fonksiyonlardır. Birbirine çok benzeyen birimler aynı kümede yüksek üyelik ilişkisine göre yer alırlar. Bundan dolayı Fuzzy(Bulanık) Kümeleme Yöntemi, birimlerin kümeye ya da kümelere ait olabilme katsayılarını hesaplar. Üyelik katsayılarının toplamı daima 1'e eşittir. Böylelikle birim en yüksek üyelik katsayısına sahip olduğu kümeye atanır. Üyelik fonksiyonları, kümedeki elemanlar sürekli veya süreksiz olsun bir bulanık kümedeki

³⁰² Yun-Teng Roan, Pei-Yin Chen, (2000), **A Fuzzy Search Algorithm for the Estimation of Motion Vectors.**

³⁰³ Karanfil, S., "Fuzzy Lojik Problemlerinde Üyelik Fonksiyonunun Belirlenmesinde Deneyisel Verilere Dayanarak Bir Yöntem Geliştirilmesi", Doktora Tezi, YTÜ, F.B.E., Matematik Ana Bilim Dalı, 1997, s. 8.

bulanıklığı karakterize fonksiyonlardır. Klasik kümeleme yöntemlerinde ise her bir birim sıfır olmayan sadece bir üyelik katsayısına sahiptir ve bu değer daima 1 dir. Dolayısıyla klasik kesin kümeleme yöntemleri, bulanık çözümlemenin sınırlı bir durumudur.³⁰⁴

Fuzzy(Bulanık) Kümeleme metotları ikiye ayrılabilir. Bulanık kümeleme için bulanık ilişkileri, diğeri ise amaç fonksiyonunu kullanır. Bulanık ilişkilere dayalı kümeleme adından da anlaşılacağı üzere, orijinal bireyler arasındaki ilişkisel yapıyla ilgilenir. İlişki veya önermeler bireyler arasındaki benzerlik veya benzemezlik kavramlarına göre tanımlanabilir. Diğer bir önerme ise bireyler arasındaki benzemezliği veya farkların derecesini ifade eden bir önerme gibi, tamamlayıcı karakterlerden biri olabilir.³⁰⁵

Amaç fonksiyonuna dayalı algoritmalar ise kümeleme problemini optimizasyon problemi haline dönüştürür. Bu yöntemde küme içindeki benzerliği ölçmek için amaç fonksiyonunu kullanır ve bu amaç fonksiyonu minimize edilerek en iyi bölünme elde edilir. Bu algoritmaların kullanılması için küme sayısının ve küme prototiplerinin nasıl olacağının bilinmesi gerekir. Ayrıca, kümeleme sürecinin başlangıç değerlerine karşı oldukça duyarlıdırlar.³⁰⁶

Fuzzy(Bulanık) kümeleme yöntemleri, diğer kümeleme yöntemlerinin aksine veri uzayındaki her bir bireyin elde bulunan bütün kümelere belirli ölçüde girmesine izin veren üyelik fonksiyonlarını kullanır. Bu üyelik değerleri, veri setinin veya kümelerin doğasını anlama açısından çok önemlidir. Hiyerarşik kümeleme gibi diğer kümeleme yöntemleri sonuç olarak kesin kümeler verir ve bireylerin bu kümelere üyelik dereceleri ya 0'dır yada 1 'dir.Oysa bulanık kümelemede bireylerin üyelik

³⁰⁴ Bezdek, J.C.; Pal, S.K., **Fuzzy Models For Pattern Recognition: Methods that Search for Structures in Data**,IEE Pres, New York, 1992, s. 6.

³⁰⁵ Güler, s.55.

³⁰⁶ Güler, s.55.

dereceleri 0 ile 1 arasındadır ve bireylerin bütün kümelere üyelik dereceleri toplandığından 1 elde edilir.³⁰⁷

Burada küme sayısı olan c kümeleme işlemi boyunca sabittir. Ancak c 'nin farklı değerleri denenip, en iyi sonucu veren küme sayılarından biri seçilebilir.

Diğer kümeleme yöntemlerine benzer olarak bulanık kümeleme de uzaklık ölçümüne dayanır. Bu yöntemin kullandığı uzaklık ölçümlerinin arasında Öklid ve Mahalanobis uzaklık ölçümleri yer alır. Bu uzaklık ölçümlerinden hangisinin seçileceği küme yapısına ve kullanılan algoritmaya bağlıdır. Bulanık kümelemenin kullanışlı bazı özellikleri vardır. Bunlar;

- Yorum açısından kullanışlı olan üyelik değerleri sağlar.
- Uzaklık kullanımı konusunda esnektir
- Üyelik değerlerinin bazıları bilindiğinde sayısal optimizasyon ile birleştirilebilir.³⁰⁸

Fuzzy(Bulanık) kümelemenin geleneksel kümeleme yöntemlerine göre avantajı, veri hakkında daha detaylı bilgi vermesidir. Diğer taraftan dezavantajları da vardır. Çok sayıdaki birey ve küme durumunda çok fazla çıktı olacağından, özetlemek ve bilgiyi tasnif etmek zordur. Ayrıca bulanık kümeleme algoritmaları genellikle karmaşıktır ve daha çok belirsizlik söz konusu olduğunda kullanılır.³⁰⁹

3.2.4.6 Geleneksel Fuzzy(Bulanık) Kümeleme Algoritmaları

Çoğu bulanık kümeleme algoritmaları amaç fonksiyonuna dayanmaktadır. Bu algoritmalar en iyi bölünmeyi amaç fonksiyonunu minimize ederek belirlerler. Amaç fonksiyonuna dayanan bulanık kümeleme algoritmalarında, her kümeyi bu kümenin

³⁰⁷ Özdamar, s.345.

³⁰⁸ Güler, s.56.

³⁰⁹ Şahinli, s.4.

prototipi temsil etmektedir. Bu küme prototipi, küme merkezi ve kümenin şekli, kümenin boyutu gibi bazı bilgileri içerir.³¹⁰

Verilen bir bireyin hangi kümeye gireceğini gösteren üyelik dereceleri, veri noktası ile küme merkezi arasındaki uzaklık hesaplanarak bulunur. Veri noktası hangi küme merkezine daha yakın ise, o kümeye ilişkin üyelik derecesi daha büyük olacaktır. Buradan anlaşılacağı gibi, bir veri setini c tane kümeye bölmek için veri noktaları ile küme merkezleri arasındaki uzaklığın minimize edilmesi ve üyelik derecelerine maksimize edilmesi gerekmektedir. Bu prensiplere dayanan birkaç kümeleme algoritması bulunmaktadır.

3.2.4.6.1 Fuzzy(Bulanık) C-ortalamlar algoritması (BCO)

Fuzzy(Bulanık) C-Ortalamlar algoritması, amaç fonksiyonuna dayanan bütün kümeleme tekniklerinin temelini oluşturmaktadır. İlk olarak Duda ve Hart sert küme bölünmesini (Hard Cluster Partition) hesaplamıştır. Dunn ise, bir bireyin birden fazla kümeye girebilmesini sağlamak için bu algoritmanın bulanık versiyonunu tanıtmıştır. Son olarak, Bezdek bulanıklık indeksini (m) dahil ederek algoritmanın son halini geliştirmiştir. BCO algoritması sonuçlandığında, p-boyutlu uzaydaki noktalar küresel bir şekil alır. Bu kümelerin yaklaşık olarak aynı boyutta olduğu varsayılır. Her bir kümeyi, küme merkezleri (centre) temsil eder ve bunlara prototip denir. Uzaklık ölçüsü olarak, veriler ile küme merkezi arasındaki Öklid uzaklığını kullanır.³¹¹

Bu tekniğin uygulanabilmesi için, küme sayısının ve bireylerin kümeye üyelik derecelerinin önceden bilinmesi gerekmektedir. Bu tür parametrelerin önceden bilinmesi zor olduğundan, bu değerler deneme yanılma yoluyla ya da geliştirilen bazı teknikler ile bulunabilir.

³¹⁰ Güler, s.56.

³¹¹ Eminov, M.; Ballı, S.; Güler, N.; Gökçe, N.; Gökçe, E., "Fuzzy Kümeleme Tabanlı Veri Keşif Sisteminin Geliştirilmesi ve Muğla İlinin Sosyo-Ekonomik Gelişmişlik Gruplarının ve Düzeylerinin Belirlenmesi", Muğla Üniversitesi Yayınları:72, Rektörlük Yayınları:51, 2006, s.4.

BCO algoritmasında, küme prototipleri ile bireyler arasındaki uzaklık sadece küresel şekle sahip kümeler için uygun olan Öklid uzaklık ölçüsü ile hesaplanır. Ancak oval, hat, dörtgen gibi farklı şekillere sahip küme çeşitleri bulunmaktadır ve hatta veri seti içindeki kümeler farklı boyuta ve yoğunluğa sahip olabilirler. BCO algoritması, küme boyutları ve yoğunluklarının farklı olduğu durumlarda da iyi çalışmamaktadır. Ayrıca küçük boyutlara sahip kümeleri de teşhis etmek zor olmaktadır. Bu tip problemleri çözmek için, Gustafson-Kessel, Bulanık C Hatlar gibi bir takım algoritmalar geliştirilmiştir³¹².

3.2.4.6.2 Gustafson-Kessel Algoritması (GKA)

Gustafson-Kessel Algoritması (GKA), küresel kümeler yerine, elipsoidal kümeleri teşhis etmek için geliştirilmiş bir bulanık kümeleme algoritmasıdır. Fuzzy(Bulanık) C- Ortalamalar algoritması bu tür kümeler için uygun değildir. Bu algoritma, BCO algoritmasıyla karşılaştırıldığında, küme merkezlerine ek olarak her bir küme simetrik ve pozitif tanımlı bir A matrisine sahiptir.³¹³

3.2.4.6.3 Gath-Geva Algoritması (GGA)

Gath ve Geva Algoritması, kümelerin yoğunluğunu ve boyutlarını da dikkate alan Gustafson-Kessel Algoritmasının geliştirilmiş bir versiyonudur. Gerçekte bu yaklaşım, amaç fonksiyonu optimize etmeye dayanmaz. Gath-Geva algoritması, Maksimum Olabilirlik Tahmin Edicisinin bulanıklaştırılmasından elde edilen deneye dayalı bir metottur. Temel prensibi, veri noktalarının p-boyutlu normal dağıldığını varsaymaktır

³¹² Zhang, H., **A Note Fuzzy Clustering Department of Computer Science and Engineering University Connecticut**, 2005, s. 144.

³¹³ John Wiley and Sons, New York, 1990

4. MÜŞTERİ KAYIP YÖNETİMİ

Hizmet işletmelerinin müşterileri elde tutmasının bir başka ifade ile müşteri kaybı oranının düşürülmesinin işletmenin karlılığına etkileri, Reichheld ve Sasser (1990: 105-111) tarafından yapılan araştırmalarla ortaya konulmuştur. Bu araştırmalar sonucunda, müşterilerin işletmede ne kadar uzun süre kalırsa işletmenin karlılığının da o derece artacağı ve müşterileri elde tutma oranının %5 arttırılmasının veya müşteri kaybının %5 azaltılmasının işletme karlılığını ortalama olarak şube bankacılığında %85, otomobil tamir hizmetlerinde %30 ve endüstriyel dağıtımda %45 arttırabileceğini ortaya çıkmıştır. Colgate vd., (1996: 28)'in araştırmasında da, müşteri kaybının %17,8'den %15'e düşürülmesinin, işletmenin karlılığını %105 arttırdığı ortaya çıkmıştır. Buna göre, müşteri bağlılığı sağlandığında bağlılık duyan müşterilere yapılan pazarlama yatırımı azalmakta, pazarlama maliyetlerinde önemli azalmalar meydana gelmekte ve işletme karlılığı önemli oranda artmaktadır.³¹⁴

Müşteri bağlılığının rakip işletmelerin rekabet araçlarına ve stratejilerine önemli etkileri vardır. Bunun temel nedeni, bir işletmeye bağlılık duyan müşterilerin, rakip işletmelerin rekabet araçlarına ve pazarlama çabalarına duyarsız kalmalarıdır.³¹⁵

4.1 Müşteri İlişkileri Yönetimi (CRM)

Teknolojik gelişmelerin ve diğer uygulamaların çok kısa süre içerisinde taklit edilebilmesi ve bu durumun rekabet üstünlüğünün sağlanmasında avantaj yaratmaması işletme ve müşteri arasında kurulan olumlu ilişkilerin önemini artırmıştır. İşletmenin, oluşturduğu “müşteri ilişkileri”nin taklit edilmesi zor ve maliyetlidir.³¹⁶

Müşteri ilişkileri yönetimi (CRM), müşterinin söylediklerini ve işletmenin müşteri hakkında bildiklerini temel alan, müşteriye bireysel karşılık veren ilişkisel pazarlamanın

³¹⁴ Griffin, 1995, s. 221; Odabaşı, 2000, s.59.

³¹⁵ Goodwin ve Ball, 1999, s.27.

³¹⁶ Liu D., Shih Y., “Integrating AHP and data mining for product recommendation based on customer lifetime value”, Information & Management, 42 (3): 387-400 (2005).

bir uygulaması olarak kabul edilmektedir. Müşteri ilişkileri yönetimi, uzun dönemli müşteri ilişkileri kurmayı buna bağlı olarak da kârlılığı amaç edinen bir kavramdır. Bu kavram pazarlama stratejileri ile bilgi teknolojileri arasında köprü vazifesi görmektedir. Müşteri ilişkileri yönetimi, kârlı müşterileri cezbetme ve elde tutma yanında, normal müşterileri kârlı müşterilere dönüştürmeyi amaç edinen bir işletme stratejisidir. Andy'ye göre ise müşteri ilişkileri yönetimi, müşteri tatminini geliştirmeyi ve müşteri bağlılığını arttırmayı ve aynı zamanda müşteri ve işletme için değer yaratmayı hedefleyen bir işletme stratejisi olarak tanımlanabilecektir³¹⁷.

Birçok işletmedeki pazarlama faaliyetlerinin birincil odak noktası firmanın müşteri tabanındaki sürekli müşteri sayısını artırmak ve bu sayıyı koruyabilmektir. Marketlerde verilen alışveriş kartları ya da Starbucks'ın "10 tane alana bir tane bedava" kartları buna örnek olarak verilebilecektir. İşletmelerin müşteri tabanını koruyabilmelerinde ürün ya da servis kalite seviyesini belli bir noktada tutmaları yeterli olmamaktadır, aynı zamanda müşteri bağlılığı yaratılmalı ve sürdürülmelidir. Örneğin Starbucks'ta müşteriler sadece orada satılan bir kahveyi almamaktadır, aynı zamanda Starbucks markasının temel düşüncelerini de almaktadır. Müşteri-işletme ilişkilerinin gelişimi ile işletmeler, müşterilerinin bağlılığını kazanmaktadırlar. Sadık müşteriler markalarını daha az sıklıkla değiştirmekteler, sağlanan ilişkiler için prim vermeye istekli olmaktadır ve word-of-mouth ile pozitif etkileşim yaratmaktadırlar³¹⁸.

CRM'de, işletme ile müşteri ve diğer pazarlama tarafları arasındaki ilişkileri geliştirmek üzerine odaklanılmıştır. Dwyer, Schurr ve Oh bu ortak ilişkinin kısa dönemli, aralıklı işlemlere dayalı olmamasını, birbirine bağlı ve uzun dönemli olması gerektiğini savunmuştur. Uzun dönemli olmasının en büyük nedeni, pazardaki tarafların uzun dönemli bir ilişkide fırsatçı davranışlar göstermeyecek olması ve ilişkilerin

³¹⁷ Çetin, E. Ş., s.33-35.

³¹⁸ Clark, J. S., Maher J. K.,2007, **If You Have Their Minds, Will Their Bodies Follow - Factors Effecting Customer Loyalty in a Ski Resort Setting**, Journal of Vacation Marketing, Vol. 13, No. 1, s. 59-71.

karşılıklı kazançlara bağlı olacak olmasıdır³¹⁹. İlişkide, iki tarafın da fayda elde edebilmesi ise bağlılığın daha güçlü olmasını sağlamaktadır.

CRM'nin diğer önemli tarafı müşteri seçiciliğidir. Birçok araştırma bir şirket için her müşterinin aynı kârlılığa sahip olmadığını göstermektedir. Bu nedenle bir şirket pazarlama programını hazırlarken ve pazarlama çabalarını belirlerken doğru müşterileri seçebilmelidir³²⁰. İşletmeler en kârlı müşterilerini tanımlamaya, onların belirgin özelliklerini ortaya koymaya ve bu özelliklere uygun yeni müşterileri kendilerine çekmeye çalışmalıdırlar. Bu da, veri tabanlarının kurulması ile mümkündür. Dolayısıyla bir işletmede müşteri ilişkileri yönetiminin benimsenmesi bölümlene temellerini de etkilemekte ve bölümlene ile müşterilerin homojen gruplara ayrılmasının yanı sıra kâr potansiyelini esas alan yeni bir bölümlendirme de yapılmış olmaktadır³²¹.

Bölümlendirme yapılması, işletmeleri daha az kârlı müşterilere hizmet etmenin maliyetini azaltmaya ilişkin stratejilere doğru yönlendirmektedir. Bu doğrultuda müşteri önceliklerini belirleyen işletmeler, kâr potansiyeli yüksek müşterilerin öncelikleri ile kâr potansiyeli düşük müşterilerin önceliklerini belirleyerek her gruba farklı pazarlama karması tasarlayabilmekte ve kâr potansiyeli düşük grup için daha dar kapsamlı bir tutundurma ve hizmet politikası geliştirebilmektedir. Diğer yandan işletme kâr potansiyeli yüksek grubu elinde tutmak için bu gruba daha fazla önem verebilmekte ve daha fazla kaynak ayırabilmektedir. Böylece işletme, müşteri grubuna yapmış olduğu yatırımın geri dönüşünü daha kısa sürede sağlayabilmektedir.

CRM çerçevesinde uygulanan kampanyalar, yeni müşterinin kampanya süresince söz konusu olduğunu, ancak kampanya tarihinden uzaklaştıkça sadık müşteri sayısının eski seviyelerine yakın bir noktaya geldiğini göstermektedir. Kampanyaların, satış tutundurma etkinliklerinin, reklam giderlerinin kârlılık üzerindeki olumsuz etkileri göz

³¹⁹ Parvatiyar, A., Sheth, J. N., 2001, **Customer Relationship Management: Emerging Practice, Process, and Discipline**, Journal of Economic and Social Research, Vol. 3, 2002 Preliminary Issue, s. 1-34

³²⁰ Ibid, s. 1-34.

³²¹ Çetin, E. Ş., s.34.

önünde bulundurulduğunda, anlamlı olan, sadık müşteri kitlesinin memnuniyetini artırmak ve ilişkiyi bu yönde geliştirmeye çalışmaktır.

Son yıllarda birçok faktör CRM'nin hızlı gelişimine katkıda bulunmaktadır. Bilgisayar ve iletişim teknolojilerinin gelişmesi, aracılık gereksinimlerini azaltmış, işletmelerin son kullanıcılarla doğrudan iletişim kurabilmesini sağlamıştır. Havayolları, bankacılık ve sigortacılık sektörleri ve ev aletleri gibi sektörler pazarlamada müşteri ilişkilerini ön plana çıkarmaya başlamıştır. Veritabanları ve doğrudan pazarlama yöntemleri bu endüstrilere pazarlama çabalarını bireyselleştirecek araçlar sağlamaktadır. On-line bankacılığın kullanılması ve kitap, araba, sigorta vb. gibi ürünlerin internette doğrudan satılmaya başlaması müşterilerin pazarlamacılarla direkt ilişki kurma isteğini artırmaktadır.

CRM'nin yaygınlık kazanmasının önemli sebeplerinden biri de hizmet sektöründe görülen gelişmedir. Hizmetler aynı yerde üretilip dağıtıldığından, aracının rolünü azaltmaktadır. Hizmet veren ile hizmet alan arasındaki duygusal bağ da ilişkilerin sürdürülmesi ve geliştirilmesi ihtiyacını artırmaktadır³²².

Yapılan araştırmalara göre, işletmelerin büyük bir kısmının, müşteri ilişkileri yönetimini benimseme nedeni, müşteriyi bütünüyle tanımayı başararak onlar için vazgeçilmez nitelikte hizmet ya da ürün geliştirmek ve onları bu yolla elde tutmayı başarabilmektir. Birçok araştırma yeni müşteriler elde etmeye göre sürekli müşterilere sahip olmanın rekabet avantajı açısından daha önemli olduğunu ortaya koymaktadır. Çünkü yeni müşteri kazanmanın maliyeti çoğunlukla müşterilerin elde tutulmasından daha fazla olmaktadır³²³. Bu doğrultuda müşteri ilişkileri yönetimine ilişkin geliştirilen işletme stratejilerinin üç önemli hedefi bulunmaktadır³²⁴.

- Müşteri memnuniyetini ve buna bağlı olarak sadakatini arttırmak,
- İşletmenin uzun dönem kârlılığını arttırmak,
- İşletmede karşılaşılan maliyetleri en alt düzeye indirmek.

³²² Parvatiyar, A., Sheth, J. N., 2009, s. 5

³²³ Parvatiyar, A., Sheth, J. N., s. 6.

³²⁴ Çetin, E. Ş., 2006, s.25.

4.1.1 Pazarlama

Pazarlamanın amacı tüketicilerin istek ve ihtiyaçlarını tatmin etmek ve karşılamaktır. Bu amacı gerçekleştirebilmek için yapılması gereken ilk şey tüketici davranışlarını analiz etmektir. Yani mal ve hizmetlerin nasıl, nereden ve niçin satın alındığı, nasıl kullanıldığı ve elden çıkarıldığı araştırılmalıdır. Tüketici davranışı anlaşılmadan ya da tüketiciler tanınmadan istek ve ihtiyaçlar belirlenmesi mümkün olmamaktadır ve onları motive eden değişkenler teşhis edilememektedir. Dolayısıyla, pazarlamacılar hedef tüketicilerin isteklerini, tercihlerini, alışveriş ve satın alma davranışlarını sürekli analiz etmeli ve stratejik kararlarını bu bilgiler üzerine inşa etmelidir.

Tüketici davranışı, kişinin ürün ve hizmetleri seçme, satın alma, kullanma ve elden çıkarmadaki kararları ve bunlarla ilgili faaliyetleri olarak tanımlanabilir.

Esasen tüketici davranışı satın alma öncesi, satın alma anı ve sonrası olmak üzere üç safhalı bir süreci kapsar. Özellikle Pazar bölümlleme, hedef pazar seçimi, konumlandırma ve uygun pazarlama karmasının geliştirilmesi ile ilgili stratejik kararlar bu süreç esas alınarak verilir. Bir taraftan pazarlamacılar, tüketicileri ürün ve hizmetlere yönlendirecek kararlar alırken, diğer taraftan tüketicilerde kendilerine sunulan teklifler içerisinde ihtiyaçlarına en uygun olanı seçmeye çalışırlar.³²⁵

Ömür boyu süren bu aktivite günlük yaşamın bir parçası haline gelir ve tüketici de mal ve hizmet seçiminde etkili olan bir alışveriş tarzı oluşturur. Gelişen teknoloji ve mevcut ekonomik şartlar günümüz tüketicisini ürünlere para vermektense markaları tercih etmeye zorlamaktadır. Çünkü müşteriler markaların benzer ürünlere nazaran sağladığı artı değerleri aramaktadır. Herhangi bir işletme tarafından üretilerek ya da çok aracı kurum tarafından piyasaya arz edilmekte olan, hizmetlere bir kimlik kazandıran,

³²⁵ ATAK, G., **Telekomda Yeni Bir Uygulama Müsteri Memnuniyeti Esas Başarı Kaçınılmaz**, Türk Telekom Dergisi, Kasım-Aralık, 2000.

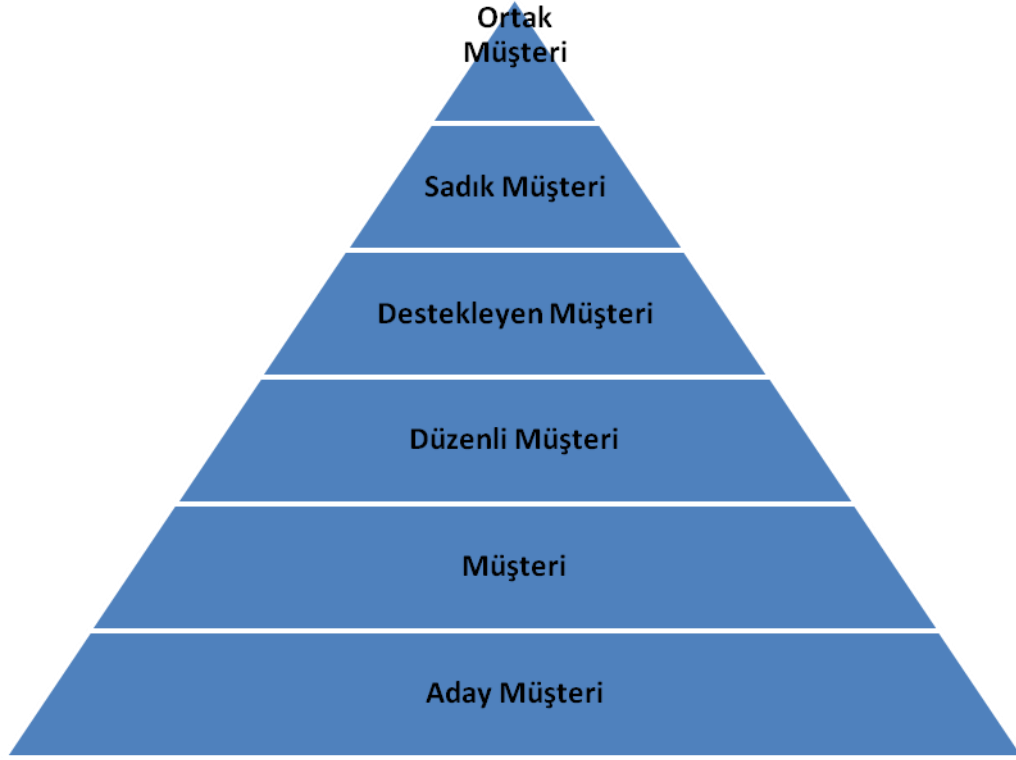
ilgili ürünü rakiplerinden farklı kılan bir terim, sembol, isim ya da bunların kombinasyonudur³²⁶.

Marka tüketiciye; seçim kolaylığı sağlama, kalite açısından bilgi ve güven verme, malın iadesi, onarımı yedek parça ihtiyacının söz konusu olduğu durumlarda isteklerine kolayca ulaşabilmesi dolayısıyla korunmasını sağlama, rekabet yolu ile üreticilere ve aracılara baskı yaratarak sunulan malların kalitelerini artırma ve mamulleri tekrar satın alabilmelerini sağlama açısından yararlar sağlar³²⁷. Ancak pazara aynı malın çok çeşitli markalarının çıkarılması haklı olarak tüketiciyi şaşırtmakta ve rasyonel seçim yapmasını engellemektedir.

Bu bağlamda tercihlerindeki etkin marka olgusu müşterilerle ürün arasında duygusal ve kültürel bir bağ oluşturarak kimlik oluşturma eylemidir.

³²⁶ Kotler, P., 1984, **Marketing Management**, Engelwood Cliffs, N.J.: Prentice Hall, s.482.

³²⁷ Cemalcılar, İ., 1984, **Pazarlama**, A.Ö.F.Yayınları, Yayın No:312, Eskişehir, s. 260-261.



Şekil 3. 15 Müşteri Tipi Hiyerarşisi

Kaynak: Demirel, Y. (2006) Müşteri İlişkileri Yönetimi ve Bilgi Paylaşımı, İstanbul: IQ Kültür Sanat Yayıncılık, Yayın No:19

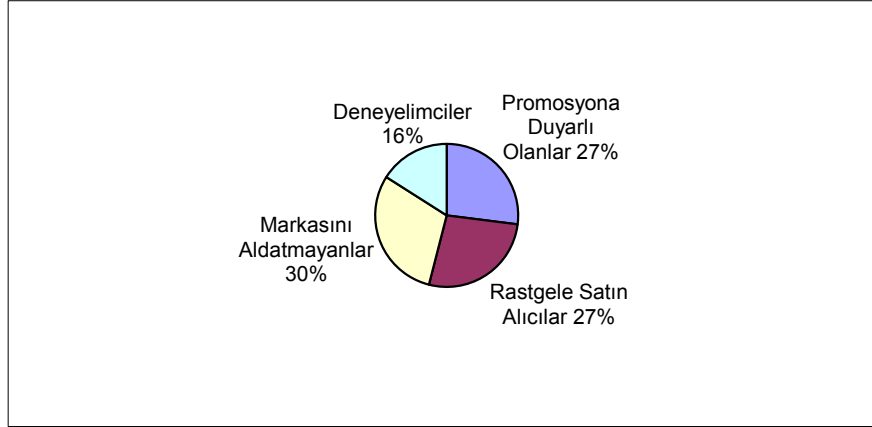
Bir ürün ya da hizmeti satın alma kararını veren müşteriler farklı özellikler gösterebilmektedir.

Yukarıdaki hiyerarşi üçgeninde de görüldüğü üzere altı farklı müşteri tipi tanımlanmaktadır;

Aday (Potansiyel) Müşteri:	Gelecekte işletmenin müşterisi olmaya aday niteliğindeki alıcılardır.
Müşteri:	İşletmeden bir kez ürün ya da hizmet satın almış olan kişi ya da kuruluştur.
Düzenli Müşteri:	İşletme ile sürekli ilişki içinde olan, ancak işletmeye karşı tarafsız tutum içinde bulunan müşteridir.
Destekleyen Müşteri:	Memnun, sadık; ancak işletmeyi diğer alıcılara karşı pasif biçimde destekleyen müşteridir.
Sadık Müşteri:	İşletmeyi diğer alıcılara öneren, sürekli olarak işletme hakkındaki olumlu tutumunu çevresine yansıtan, taraftar müşteridir.
Ortak (Partner) Müşteri:	İşletmenin çıkarlarını kendi çıkarları ile birlikte düşünen müşteridir.

Kaynak: Ersoy, N. F., 2002, **Finansal Hizmetlerde Müşteri İlişkileri Yönetimi (CRM) Kavramı**, Pazarlama Dünyası Dergisi, Mayıs-Haziran, s.7

Eyüp Karagüllü tarafından 2000 yılında yapılan bir araştırma sonucunda ise Türkiye’deki müşteri tiplerinin aşağıda yer alan şekildeki gibi dağıldığı görülmüştür³²⁸.



Şekil 3. 16 Türkiye'deki Müşteri Tiplerinin Oransal Gösterimi

Kaynak: Karagüllü, Eyüp (2000), “Scanner Data Yönetimiyle Türkiye’nin Tüketim Haritası”, Power Dergisi, Haziran.

Yukarıdaki şekilde de görüldüğü üzere Türkiye’de %30 oranıyla en yaygın görülen müşteri tipi markalarını aldatmayan müşterilerdir. Bu gruptaki müşteriler markalarına sadık kalmaktadırlar. Bu nedenle alışverişlerini planlı olarak yapmaktalar, alışveriş öncesinde hangi ürünü ya da hizmeti alacaklarını belirlemektedirler. Bu gruptaki müşterilerin fiyata karşı duyarlılıkları marka bağlılıklarının sınırını çizmektedir. Karagüllü tarafından yapılan araştırma sonucuna göre bu grupta aile yapısı küçük ve ortalama 40 ya da daha yüksek yaşlarda insanlar bulunmaktadır.

Müşterilerin %27’sini oluşturan promosyona duyarlı olan müşteriler alışverişlerini planlamadan yapmaktadır. Bu grubun ürün ya da hizmeti satın alma kararında satış tutundurma faaliyetleri ağırlıklı bir rol oynamaktadır. Genellikle bilinen markaları tercih etmektedirler ve yeni ürünleri deneme eğilimi taşımaktadırlar. Bu gruptakiler 30-39 yaş ortalamasına sahiptirler.

³²⁸ Çetin, E. Ş., 2005, **Perakendecilik Sektöründe CRM Uygulamaları: Süpermarket Müşterilerinin Mağaza Kartı Kullanımlarıyla İlgili Tercih Nedenlerinin Belirlenmesi**, Çukurova Üniversitesi Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, Adana, s.5.

Deneyilimciler grubunun ana satın alma davranışı, yeni ürünleri, markaları denemektir. Marka bağlılıkları bulunmayan bu grup promosyona da duyarlıdır. “Yenilikçi” de denilen bu grup, aynı zamanda fiyata karşı duyarlılığı en düşük tüketicileri temsil eder; çünkü onlar için fiyattan çok yenilik önem taşımaktadır. Ülkemiz genelinde %16’lık bir oranla tüketiciler arasında en küçük kesimi ifade eden deneyilimcilerin yaş ortalamasının 20-29 arasında olduğu görülmüştür.

Rastgele satın alıcıların ülkemizdeki müşteri tipleri içindeki payı %27 olarak belirlenmiştir. Bu tüketici grubu kararlarını satın alma aşamasında vermektedir. Alacağı ürün ya da markayı önceden planlamamaktadır.

Şirketlerin amaçlarından biri, müşterilerini, markasını aldatmayan, sadık birer müşteri haline getirebilmektir. Müşteri bağlılığı, olumlu yaklaşıma veya davranışsal tepkilere dayanarak, bir müşterinin bir marka, mağaza, üretici veya hizmet verene bağlılığını ifade etmektedir. Acuner’e göre müşteri bağlılığı, müşterilere isteklerini karşılayan hatta aşan şekilde hizmet vererek ilişkinin sürekli tutulması ile gerçekleşecektir³²⁹. Tüketicilerin bir markayı deneyip ürünün performansından tatmin olup tekrar tekrar satın almaları durumunda, o markanın bağımlıları oldukları söylenebilecektir.

4.1.2 Marka Bağımlılığı

Marka bağlılığı kavramına ulaşıncaya kadar birkaç aşamadan geçmek gerekmektedir. Marka bağlılık piramidinin en altında fiyata karşı duyarlı markalara karşı kayıtsız, marka isminin satın alma kararında çok az öneme sahip olduğu ve tüm markaları birbirine benzer olarak algılayan ve markaya sadık olmayan alıcılar yer almaktadır. İkinci sırada ise ürünün özelliklerinden tatmin olan veya en azından tatminsizlik yaşamayan alıcılar yer almaktadır.³³⁰

³²⁹ İbid, s.17-18.

³³⁰ Francisco Jose Mas Ruiz, ‘Strategic Group Analysis in Strategic Marketing: An Application to Spanish Savings Banks’, **Marketing Intelligence and Planning**, Vol.6, No.4, (1998), s.8.

Markadan tatmin olduđu için markayı satın alma davranışı alışkanlık halini almış olan bu gruptaki alıcılar için markayı değiştirmenin sebebi yoktur. Sadakat piramidinin üçüncü sırasında ise markadan tatmin olan ve ek olarak zaman maliyeti, parasal maliyetler ve performans riski gibi markayı değiştirmenin maliyeti olduđu alıcı grubu yer almaktadır. Piramidin dördüncü sırasında ise markaya karşı duygusal olarak bağlı markayı seven ve onu arkadaş gibi gören alıcı grubu yer almaktadır. Piramidin en üst sırasında ise markaya bağlı olan alıcılar yer almaktadır. Markayı kullanmaktan gurur duyan bu grup için marka, duygusal olarak müşterilerin kendilerini ifade etmesinde oldukça önemlidir³³¹.

Marka bağlılığında, tüketiciler daha önceden satın alıp denedikleri bir markayı satın almayı sürdürmektedirler. Marka bağlılığının gerçek mi yoksa sahte marka bağlılığı mı olduğunun ortaya çıkarılması işletme için önem taşımaktadır. Gerçek marka bağımlısı marka özelliğine önem verirken, sahte marka bağımlısı indirim kuponu, mağaza içi raf düzenlenmesi gibi hususlardan etkilenmekte ve bunlara bağlı olarak kullanmakta olduđu markayı kolayca değiştirebilmektedir.

Marka bağlılığı, tüketicinin belirli bir süre içinde bir ürün kategorisi içinde yer alan bir ya da daha çok markaya karşı takındığı olumlu tutum (tercih) ve davranış tepkisi olarak tanımlanmaktadır³³². Bu tanımdan da anlaşılacağı üzere marka bağlılığına iki açıdan incelenebilmektedir.

4.1.2.1 Davranış Olarak Marka Bağlılığı (Behavioral Brand Loyalty)

Müşterilerin devamlı olarak aynı markayı alarak gösterdikleri davranıştır. Tekrarlanan alımlar, belirli bir markanın, diğer markaların toplam alım sayısına göre ya

³³¹ Avcılar, M. Y., Akın, M., **Tüketici Temelli Marka Değeri Kavramı**, Pazarlama Dünyası, Şubat Sayısı, 2007, s.17

³³² Ersoy, A., **Marka Seçimi Davranışı Ve Marka Seçimine Etki Eden Marka Bağlılığının Davranışsal Boyutu**, http://paribus.tr.googlepages.com/a_ersoy.doc, (10 Haziran 2010).

da alım miktarlarına göre alım sıklığını ifade etmektedir³³³. Aynı şirketten hizmet alınmaya devam edilmesi, ilişkilerin her açıdan geliştirilmesi, hatırlama bağlılık davranışının örneklerindendir³³⁴.

4.1.2.2 Tutum Olarak Marka Bağlılığı (Attitudinal Brand Loyalty)

Duygular bir kişinin bir ürüne, hizmete ya da şirkete bağlılığının oluşmasında etkili olmaktadır. Hatta bu duygular ile kişinin bağlılık derecesi belirlenebilmektedir³³⁵. Tutuma dayalı marka bağlılığının sağlanması, devamlı müşteriler için program geliştirme ya da bu müşterilere hediyeler verme gibi stratejilere dayanmamakta, sadakat ve güven gibi unsurlar da önem kazanmaktadır. Araştırmalar marka bağlılığın sadece alım davranışının yinelenmesi olmadığını, markaya olan çok yönlü tutumun da bir sonucu olduğunu göstermektedir.

Bentler ve Speckart 1981 yılındaki araştırmalarında tutumların davranışlardan öncelikli olduğunu bulmuşlardır. Peter ve Olson ise tutumlardaki negatif bir değişimin, müşterilerin çoğunun başka markalara kaymasına neden olduğu, dolayısıyla tutum değiştirmenin marka bağlılığı ve marka değiştirme için iyi bir gösterge olduğu sonucuna ulaşmışlardır³³⁶.

Tüketicilerin farklı markalara karşı özel satın alma davranışı göstermeleri, marka bağımlılığı kavramının önemini ortaya koymaktadır. Tüketicinin belirli markalar karşısında takındığı olumlu tutumun nedenleri arasında tüketicinin demografik, ekonomik ya da psiko-grafik özellikleri, ürünün sahip olduğu özellikler ve çevresel faktörler yer almaktadır³³⁷.

³³³ Back, K., Parks, S. C., 2003, **A Brand Loyalty Model Involving Cognitive, Affective, and Conative Brand Loyalty and Customer Satisfaction**, Journal of Hospitality & Tourism Research, Vol. 27; No. 4, s. 419-435.

³³⁴ Hallowell, R., 1996, **The Relationships of Customer Satisfaction, Customer Loyalty, and Profitability: An Emprical Study**, International Journal of Service Industry Management, Vol. 7, No. 4, s. 27-42.

³³⁵ Hallowell, R., a.g.m., s. 27-42.

³³⁶ Back, K., Parks, S. C., s. 419-435.

³³⁷ Ersoy, A., s. 8.

Bağılılık, gelecekte daha önceden bilinen veya tavsiye edilen mal veya hizmeti yeniden satın alma eğilimine yoğun biçimde yönelmesi ve katılmasıdır. Bu yüzden, çevresel şartlar ve pazarlama çabaları yönelme davranışını potansiyel olarak etkilese bile, markaya veya birden çok sebebe bağlı olmak satın almayı doğuracaktır. Marka bağımlılığı;

- Bazı tüketicilerin belirli bir markayı düzenli ve tutarlı olarak satın alma eğilimidir.
- Tüketicilerin daha önceden satın alıp denedikleri bir markayı satın almayı sürdürme eğilimidir.
- Tüketicinin belirli bir süre içinde bir ürün kategorisi içinde yer alan bir ya da daha çok markaya karşı takındığı olumlu tutum(tercih) ve davranış tepkisidir.

ABD’de “Chicago Tribune Paneli”nce ele alınan 100 hane halkından sağlanan satın alma kayıtlarına göre kahve, portakal suyu, sabun ve margarin gibi sık satın alınan sıralamasına göre dört değişik marka bağımlılığı kategorisi oluşturulmuştur.

Tam Bağımlılık	A,A,A,A,A,.....Sıralamasında
Bölünmüş Bağımlılık	A,B,A,B,A,B,.....Sıralamasında
Kararsız Bağımlılık	A,A,A,B,B,B,.....Sıralamasında
Bağımlılık Yok	A,B,C,D,E,F,G,H,...Sıralamasında

Sınıflandırma kapsamına giren ürün türüne bağlı olarak tüm ürünlerin sık satın alınan ürünler olmasına rağmen herhangi bir ayırmadaki marka bağımlılığı oranı % 54 ile % 95 arasında değiştiği belirlenmiştir. Kısaca sık satın alınan ürünlerde tüketicilerin en az yarısının şu veya bu şekilde bir marka bağımlılığına sahip olduğu sonucuna ulaşılmıştır.³³⁸

Marka bağılılığını belirlemede, özellikle deneysel araştırmalarda en çok kullanılan tanımlamalardan biri, belirli bir ürün türü içindeki toplam satın alımların o tür içindeki en çok satın alınan markaya oranlaması olarak ortaya çıkmaktadır. Satın

³³⁸ Dilek Altas, **Türk Bankacılık Sektörünün _statistik Tekniklerle Analizi (2000 Krizi Öncesi ve Sonrası)**, _stanbul: Der’in Yayınları, 2006, s.85-86.

alımların oranlaması sayısal nitelikte olduğundan, marka bağımlılığını matematiksel modeller olarak ifade edilmesi de söz konusu olmaktadır. Oranlamalar ortaya çıktıktan sonra, ikili ya da üçlü marka bağımlılığı da analiz edilmektedir.

Analizler, belirli bir süre içinde hane halkları tarafından satın alınan markaların dağılımını içeren zaman serileri olarak gerçekleştirilmektedir. Marka bağımlılığının tanımı içinde tüketici tercihlerinin yanı sıra tüketici satın alımlarının da yer alması gerekir. Tüketicinin markayı tekrar satın alması yanı sıra o markaya karşı olumlu tutuma da sahip olması şartı gerçek anlamda bir bağımlılık ifade eder.

Marka bağımlılığının davranışsal ve tutuma dayalı bölümlerinin birlikte yer aldığı bu model, sadece satın alma verilerine dayalı geleneksel modellere üstünlük sağlamaktadır. Ayrıca bu yöntemle marka bağımlılığı genel bir kavram olarak ele alınmamakta, ürün özelliğine dayalı bir husus olarak gözetilmektedir.³³⁹

4.2 Marka Bağımlılığı Yaratma

Marka bağımlılığı basit bir kavram değildir, bu kavramın bir çok tanımları ve ölçütleri vardır. Marka bağımlılığını ayrıntılı ve tam olarak açıklayabilmek için şu tanımı kullanabiliriz. Marka bağımlılığı bir dizi ürün markası arasından bir yada daha çok marka seçeneği ile ilgili, zamana dayalı olarak, bir karar alma birimi tarafından gösterilen davranışsal tepki (satın alma) ve psikolojik süreçlerin (karar alıcı, değerlendirici) bir fonksiyonudur. Marka bağımlılığı, bazı tüketicilerin belirli bir markayı düzenli ve tutarlı olarak satın alma eğilimidir. Tüketicilerin farklı markalara karşı özel satın alma davranışı göstermeleri marka bağımlılığı kavramının önemini ortaya koymaktadır. Marka bağımlılığı ürün özelliklerine dayalı olduğu kadar, tüketici özelliklerine göre de farklılık gösterebilir. Marka bağımlılığı sık sık tekrarlı satın alma davranışıyla karıştırılmaktadır. Bu iki kavram arasında önemli bir farklılık vardır. Markaya bağımlılığı olan tüketici markaya karşı duygusal bir bağlılık duyması nedeniyle aynı markayı tercih

³³⁹ Baurdeau, A., Liebenau, J. and Savin, B. (2005) **A New View of Scale And Scope In The Telecommunications Industry; Implications For Competition And Innovation**. International Journal of Digital Economics. 60(4) pp.85-103

ederken, tekrarlı satın alma davranışını gösteren tüketici ise markayı alış-veriş yaptığı yerde başka bir alternatifin bulunmaması, satın alınan markanın en ucuz olması veya tanıtım ya da promosyon gibi nedenlerden dolayı veya sadece alışkanlık yüzünden aynı markayı satın alabilir.

Bağlılık gelecekte daha önceden bilinen veya tavsiye edilen mal veya hizmeti yeniden satın alma eğilimine yoğun biçimde yönelmesi ve katılmasıdır. Bu yüzden, çevresel şartlar ve pazarlama çabaları yönelme davranışını potansiyel olarak etkilese bile, markaya veya bir den çok sebebe bağlı olmak satın almayı doğuracaktır.³⁴⁰

4.2.1 Marka Müşteri İlişkisi

Pazarlama faaliyetleri sırasında müşteri ziyaretleri önemli bir yer teşkil etmektedir. Her ziyaret, satışla sonuçlanmasa bile müşteri hakkında çok sayıda ayrıntılı bilgi elde edilebilecektir. Bu bilgiler, şirketin pazarlama stratejisinin oluşturulmasında, reklam bütçesinin dağıtılmasında ve müşteri ile ilişkilerin kârlı bir modele dönüştürülmesi sürecinde kullanılmaktadır.

Geleneksel iş modellerinin tersine günümüzde pazarlama teknikleri arasında müşteri ile ilişkiler, iş modelinin başarısında belirleyici rol oynamaktadır. Kitlese pazarlama yerine birebir pazarlama modellerinin kullanılması, ısmarlama çözümler üretilmesi ve birebir müşteri ilişkilerini amaçlayan bir anlayışın gerçekleşmesi için bütün süreçlerin ölçümlenmesi zorunludur. Başarı ölçütü sadece satışlar, ciro ya da kârlılık değildir. Müşterilerin bağlılık oranları, markanın bilinirliği gibi ölçütlerin de başarı açısından izlenmesi gerekmektedir.

Bir marka-müşteri ilişkisi bir değer önerisi üzerine kurulabilir. Örneğin, Saturn markasının müşterilere saygıyla ve arkadaş gibi davranma üzerine kurulmuş olan müşteri-marka ilişkisi, arabanın fonksiyonel, duygusal veya kendini açıklama faydaları ile çevrilmemiştir.

³⁴⁰ Nuri Çalık, **Marka Bağlılığı Ve Marka Bağlılığına Etki Eden Faktörler**, Anadolu Üniversitesi, 2008, s.55.

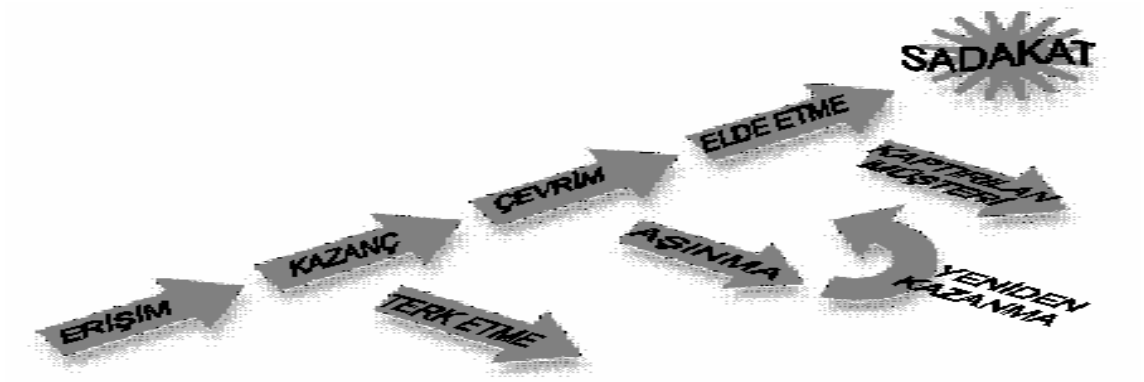
Marka-müşteri ilişkilerinin çoğu, marka, üründen ziyade kurum, kişi veya sembol olarak düşünüldüğünde ortaya çıkmaktadır. Örneğin, kurumsal bağlar bir ilişkinin temelini oluşturan saygıya dönüşebilmektedir.

4.2.2 Müşteri Hayat Seyri Ölçütleri

Rekabet baskısı nedeniyle kitle pazarlamadan giderek uzaklaşan şirketler müşteri merkezli pazarlama programlarına geçmeye başlamışlardır. Bu noktada kazanılan müşteriyle uzun süreli ilişki önem kazanmaktadır. İnternet ortamında müşteriyle uzun süreli ilişki kurulabilmesi için ise ilişkinin süre bakımından ölçülmesi gerekmektedir. İnternet ortamında müşteri ile ilişkide yaşanan sorunlardan en önemlisi müşterinin görünmez olmasıdır. Ziyaretçiler (müşteri veya tüketici) web sitesine gelerek çeşitli izler bırakıp ayrılırlar. Ancak bırakılan bu izlerin kime ait olduğu kesin olarak belli değildir. Kullanılan çeşitli programlarla ziyaretçiler hakkında bazen izinleri olmaksızın da çeşitli bilgiler toplanabilir. Buna rağmen müşterinin kesin olarak kim olduğu, bir kez daha gelip gelmeyeceği, başka bir ifadeyle sadık bir müşteri mi olduğu yoksa rastlantısal olarak mı siteyi ziyaret ettiği tam olarak anlaşılamaz. Bu durumda site üzerinde yürütülen etkinliklerin örneğin promosyonların yararı tam olarak ölçülemez.

Müşteri hayat seyri, pazarlama programının hedef kitleye ulaşmasıyla başlar, sadık müşteri tabanı oluşturuncaya kadar kadar ilerlemektedir. Bu süreç Şekil 3.17’de gösterildiği gibi terketme ve aşınma sonucu kesilebilir.³⁴¹

³⁴¹ Hung S., Yen D., Wang H., “**Applying data mining to telecom churn management**”, Expert Systems with Applications, In Press (2005).



Şekil 3. 17 Müşteri Hayat Seyri

Kaynak: Cutler, Matt; Sterne, Jim (2004): “E-Metrics Customer Centric Solutions”,
www.emetrics.org/articles/whitepaper.html, 14.12.2004.

Eriřim	Geleneksel pazarlamada erişim örneğın reklam mesajının hedef kitlenin dikkatini çekme potansiyelidir. Televizyonda ise iletilen mesajın kaç eve ulařtığı dikkate alınır.
Kazanma	Hedef kitle içerisinde web sitesinin farkındalığını ölçer. Kazanma aşamasında amaç müşterinin katılımıdır. Hedef kitle, promosyonu gördü mü, hatırlıyor mu ve herhangi bir platformda işletme ile ilişki kurdu mu sorularının cevabı aranır. Web sitesinden satılan kitap, oyuncak, yazılım ya da başka bir şey olabilir.
Dönüşüm	Dönüşüm, satışın gerçekleştiğı ve potansiyel alıcının müşteriye dönüřtüğü noktadır.
Elde Tutma	Kazanılan müşterilerin elde tutulması internete dayalı iş modellerinin önemli sorunlarından birisidir. Yeni bir müşteri kazanmanın maliyeti göz önüne alındığında eldeki müşteriye tutmak bütün modellerin tercihidir.
Sadakat	Hayat seyri en uzun müşteriler sadece ürünlere değil, ürünleri satan şirketlere de sadakat gösterirler. Sadık müşterilere yönelik ölçütler müşterilerin elde tutulmasını arttıran, kayıp müşteri oranını düşürmeye yarayan sonuçlar üretirler.
Kayıp Müşterilerin Analizi	İş hayatında karşılaşılan çeşitli sıkıntılar yüzünden sadık müşterilerin bile kaybedilebileceğı muhakkaktır. En doğru iş modellerinde bile süreçlerde yapılan küçük bir hata, basit bir ilgi kaybı "müşterinin hayat seyrinin" yarıda kesilmesine neden olmaktadır. Hayat seyrinin yarıda kesilmesine yol açan etkenlerin belirlenmesi modelin optimizasyonu açısından önem taşımaktadır.
Terk Etme (Abandonment)	Geleneksel alışveriş ortamlarında sepetini doldurup almaktan vazgeçen müşteriye çok az rastlanıyor olursa bile bu "kötü alışveriş davranışı" olarak adlandırılır. Bu nedenle alışverişini terk etme eğiliminde olanların izlenmesi, ölçülmesi ve yönlendirilmesi gerekmektedir.
Kaybetme (Attrition)	Düzenli olarak alışveriş yapan bir müşterinin uzun süre alışveriş yapmaması kaybı işaret eder. Mevcut müşterinin alışverişlerini keserek başka yerlere gitmesi de kayıp olarak değerlendirilir. "Kayıp" bir sadakatsizlik göstergesidir.
Kaptırılan Müşteri (Churn)	Belli bir dönemde müşteri profilinde meydana gelen değişme ne kadar müşterinin başka sitelere kaptırıldığını gösterir. Kaptırılan müşteri ölçütü belli bir zaman periyodunda müşteri temelinin ne kadarının değiştiğini ölçer.

4.2.3 Güven ve Marka Bağlılığı

Olumlu bir müşteri-firma ilişkisinin temelinde güven yatmaktadır. Morgan ve Hunt güven unsurunu; karşı tarafın doğruluğuna ve dürüstlüğüne inanma olarak tanımlamaktadır. Müşteri güveni iki biçimde karşımıza çıkmaktadır. İşletmenin yönetim politika ve uygulamalarına güven ve çalışanların davranışlarına duyulan güven. Bu iki unsur tutuma dayalı bağlılık oluşmasında da önemli rol oynamaktadır³⁴².

Yapılan araştırmalarda müşteri kaybının başlıca nedenleri arasında (%60 oranında) tüketicinin “satış temsilcisinin kendisiyle ilgilenmediğini” düşünmesi olduğu ortaya çıkmıştır³⁴³. İlişki yerine fiyat üzerine kurulu tedbirler almaya çalışan işletmelerin kâr marjları düşmektedir. Dolayısıyla müşterilerin satış elemanları ile geliştirdiği ilişkiler, firma ile ilişkilerin geliştirilmesinde de anahtar rol oynamaktadır. Bir çalışmada, satış elemanına duyulan güvenin, firmaya duyulan güvenin artmasında etkili olmadığı sonucuna ulaşılmıştır. Müşteri her ne kadar satış elemanının kendisi için en iyisinin sağlanması için uğraştığını düşünse de örgütün de müşterilere yarar sağlayacak şekilde çalışacağına ikna olmamaktadır.

Müşterinin satış elemanına güvendiği zaman hizmet ya da ürünlerde ortaya çıkan hatalar firmaya olan güveni de etkilemektedir. Eğer firma için bir güven oluşturulmuşsa satış elemanının daha fazla güven verici ilişki içinde olması gerekmektedir. Çünkü güven satıcı ile müşterilerin birçok karşılıklı etkileşimi ile oluşmaktadır, sadece satış elemanı ile olan ilişkiler vasıtası ile oluşmamaktadır. Çalışmada ayrıca müşteri ile firma arasındaki ilişkinin, firmanın hizmet kalitesi ve satış elemanının sağladığı müşteri tatmini ile sağlandığı anlaşılmaktadır. Satış elemanına güven duyan müşteriler ise daha çok tatmin sağlamaktadır.

³⁴² Clark, J. S., Maher J. K., **If You Have Their Minds, Will Their Bodies Follow? Factors Effecting Customer Loyalty in a Ski Resort Setting**, Journal of Vacation Marketing, Vol. 13, No. 1, 2007, s. 59–71.

³⁴³ Çetin, E. Ş., s.19.

Müşteri ile iletişimin sağlanması için sadece satış elemanlarının kullanılması da istenilen sonuçları vermemektedir. İlk olarak satış elemanları her zaman işverenlerinin istekleri doğrultusunda hareket etmemektedir. Bu ise müşterilerin istenilenden farklı algılara sahip olmasına sebep olabilmektedir. İkinci olarak satış elemanları sürekli değişmektedir. Bu yüzden müşteriler bir satış elemanı ile sıkı bağlar kurduktan sonra o satış elemanı başka bir şirkete geçtiğinde müşteriler de onunla birlikte ya da müşteri aynı markayı kullanmaya devam etse bile firmaya karşı duyduğu güven eskisi gibi olmayacaktır.

Foster ve Cadogan tarafından yapılan bir araştırmada bağlılığın alıcının satıcıya duyduğu güvenle doğrudan ilişkili olduğu belirlenmiştir. Ancak satıcı firmanın ürün ve hizmetlerinin kalitesi ile bağlılık arasında ilişki olduğuna yönelik bir bulguya ulaşılamamıştır. Hizmet kusurlarına ilişkin bir mazisi olan firmaya müşteri tarafından güven duyulmamaktadır ve bu güven eksikliği satıcı firma hakkında olumsuz algıya sahip olunmasına neden olmaktadır. Dolayısıyla hizmet kalitesi doğrudan bağlılığı etkilemese de, gelişmesinde ve azalmasında etkili olmaktadır. Ancak bağlılığın firmaya güvenle ve satış elemanının müşterilerle olan ilişkisi ile doğrudan ilgili olduğu anlaşılmaktadır³⁴⁴.

Türkiye’de telekomünikasyon sektörü üzerinde yapılmış olan araştırmada ise algılanan hizmet kalitesinin, güvenin ve kurumsal imajın müşteri bağlılığı üzerinde pozitif etkisi olduğu görülmüştür. Bu değişkenler içinde en önemli unsurun ise “güven” olduğu bulunmuştur³⁴⁵.

4.2.4 Sadakat ve Marka Bağlılığı

Moorman’a göre sadakat, işletme ile müşteri tarafından değer addedilen bir ilişkinin sürdürülmesi isteğidir. Burada öne çıkan nokta, bağlılığı daha yüksek olan

³⁴⁴ Foster, B. D., Cadogan C. W., **Relationship Selling and Customer Loyalty: An Empirical Investigation**, Marketing Intelligence & Planning, Vol. 18, No. 4, 2000, s. 185-199.

³⁴⁵ Aydın, S., Özer, G., **The Analysis of Antecedents of Customer Loyalty in the Turkish Mobile Telecommunication Market**, European Journal of Marketing; Vol. 39, No. 7/8, 2005, s. 910-925.

müşterinin firma ile ilişkilerini sürdürebilmek için bağlılığı düşük olanlara göre daha fazla çaba sarfedecek olmasıdır³⁴⁶.

Sheth ve Parvatiyar müşterilerin psikolojik ve sosyolojik yarar sağlama amacı ile ilişkisel davranışta bulunduğunu belirlemiştir. Ayrıca müşterilerin ödülleri almak istediği ve yarar sağladıklarında diğer seçenekleri reddettikleri görülmüştür. Hayat seyri en uzun müşteriler sadece ürünlere değil, ürünleri satan şirketlere de sadakat göstermektedirler. Giderek artan rakip iş modelleri nedeniyle sadakat ticari hayatta önemli bir kavramdır. Sadık müşterilere yönelik ölçütler müşterilerin elde tutulmasını arttıran, kayıp müşteri oranını düşürmeye yarayan sonuçlar üretirler. CRM programlarının en önemli tarafı müşteri seçimidir.

Her müşterinin şirket için aynı değerde olmadığı artık herkes tarafından kabul edilmektedir. Pareto, 80/20 kuralı ile müşterilerin %20'sinin gelirin %80'ini sağladığını öne sürmektedir Hatta daha düşük orandaki müşterilerin bile gelirin %80-90'ını kazandırdığı görülmüştür. Bu yüzden bir şirketin her müşteri için aynı oranda kaynak ayırması mantıklı olmamaktadır. Müşteri ayrımı yapma ve programı farklılaştırma verilen hizmetlerle sağlanan gelirin uyumlu olması için önem taşımaktadır. Yüksek gelir sağlama potansiyeli olan kişiler maliyet ve hizmetlerden daha yüksek pay almayı hak etmektedirler. Aksi takdirde rakipler yüksek getirili müşterilere daha fazla kaynak ayırarak ve daha iyi servis sunarak fırsatı elde edecektir. Düşük getirili müşteriler için ise daha düşük maliyetli kaynaklar kullanılarak maliyetlerden tasarruf edilmeye çalışılacaktır.

Bir firmanın sürekli başarı sağlamasındaki en önemli noktalardan biri mevcut müşterilerini koruyabilmesi ve onları markalarına bağlı hale getirebilmesidir. Sadık müşteriler daha fazla miktarlarda satın almakta, primli fiyat vermeyi kabul etmekte ve pozitif WOM etkisi ile firmaya yeni müşteriler kazandırmaktadırlar. Telekomünikasyon sektöründe çalışan firmalar ayda müşterilerinin %2-4'ünü kaybetmektedir. Bu sadık olmayan müşteriler, hasılat ve kârda yüksek tutarlı kayıp anlamına da gelmektedir.

³⁴⁶ Clark, J. S., Maher J. K., s. 59-71.

Örneğin; operatörlerden Orange her yıl müşterilerinin %20'sini kaybetmektedir ve Orange, her yeni müşteriye elde etmek için ortalama 256 pound (1996 yılı için) maliyete katlanmaktadır. Dolayısıyla 1 milyon müşterinin yılda %20'sini değil de %10'u kaybetmesi halinde yıllık tasarrufu 25 milyon pound'un üzerinde olacaktır. Benzer şekilde finansal hizmetler sektöründe müşteri bağlılığının %5 artırılması ile karın %25 ile %75 arasında bir artış göstereceği belirlenmiştir³⁴⁷.

Bağlılığa ilişkin yapılan planlamalarda yüksek artışlar görülmesine rağmen bu planların müşteri ve işletme için değeri konusunda tartışmalar süregelmektedir. Bu nedenle çoğu akademisyen sadakat programlarının başarısız olduklarını düşünmektedir. Sadakat planlarına ilişkin eleştiriler 3 temel grupta incelenmektedir³⁴⁸;

- Sadakat programları ile bağlılık geliştirme arasında çok zayıf bir bağ vardır.
- Örgütler müşterilerine ilişkin verileri doğru kullanmamaktadır.
- Ödüllendirme mekanizmaları bağlılık yaratmak için çok sınırlıdır ve çoğu zaman istenilen sonuçları vermemektedir.

Sadakat programlarındaki ödüllendirmeler tek başına bağlılık yaratamayacaktır. Ancak bu programlarla fiyat, ürün çeşitliliği, kalite ve uygunluk açılarından işletmenin müşterilerine sunduğu hizmetlerin geliştirilmesine yönelik bilgi sağlanmaktadır. Örneğin bağlılık bilgisi stokların planlanmasında kullanılabilecektir. İşletmeler müşterilerin tercihlerini, hangi ürünleri daha fazla aldıklarını görebilecek ve bu ürünlerle birlikte aldıkları başka ürünler olup olmadığını tespit edebileceklerdir. Aynı zamanda bir üretim hattının hangi sıklıkla alındığını, ikame malların ya da benzer malların tercih edilip edilmediği, markalar arası kaymaların olup olmadığı konularında da bilgi sahibi olunabilecektir.

³⁴⁷ Aydın, S., Özer, G., **The Analysis of Antecedents of Customer Loyalty in the Turkish Mobile Telecommunication Market**, European Journal of Marketing; Vol. 39, No. 7/8, 2005, s. 910-925.

³⁴⁸ Rowley, J., Haynes, L., **Customer Relationship Management: The Matalan Way**, The Marketing Review, No:5, 2005, s. 175-187.

4.2.5 Müşteri Değeri ve Marka Bağlılığı

Değer ya da hizmete ilişkin deneyimlere dayalı müşteri algıları daha sonra bağlılığa dönüşebilecek olan hizmet seçiminde önemli rol oynamaktadır. Müşteri değer yargıları ilişkilerin devam etmesinde algılanan faydadan algılanan maliyetlerin çıkarılması ile bulunmaktadır. Fiyat müşterilerin değer belirlemede kullanılan değişkenlerden olduğundan değer olumlu algılanması fiyat yükselmelerine karşı daha az dirençli yapmaktadır. İşletmenin müşteriye ne kadar değer verdiği de müşterinin o işletmeye, dolayısıyla o işletmenin markasına bağlılığını etkilemektedir³⁴⁹.

CRM şirket ve müşteri için yüksek bir değer yaratabilmek amacıyla seçilmiş müşterileri elde etme, tutma ve onlarla ortaklık etme süreci ve stratejisidir. Müşteri değerini artırmak için örgütte yer alan pazarlama, satış, müşteri hizmetleri ve tedarik fonksiyonlarının da bütünleşmesini içermektedir³⁵⁰.

CRM, her şeyden önce çok iyi bir teknolojik altyapı gerektirmektedir. İşletme, müşterilerini ne kadar anlar ve bağlantı kurarsa o kadar kâr elde edebilecektir. Teknolojik gelişmeler, işletmelere çeşitli kanallardan gelebilecek müşteri bilgisini daha çabuk toplama, analiz etme ve kullanma şansını vermektedir. Herhangi bir CRM projesinde teknolojik yatırım kararı verilmeden önce işletme için en uygun CRM stratejisi en baştan belirlenmeli, teknoloji yatırımı ondan sonra yapılmalıdır.

CRM, işletmenin tüm kaynaklarını (insan, süreçler, teknoloji, bilgi yönetimi) içine alan ve sürekli tekrar eden bir süreç olmalıdır. Müşteriler hakkındaki bilgiyi çeşitli kanallardan (on-line işlemler, anketler, yüz yüze görüşme, telefon vb..) elde edip, bu bilgiyi daha iyi ürün ve hizmetler dizayn edebilmek ve ortaya çıkan ürün / hizmetlerin sunulabileceği doğru dağıtım kanallarını kullanabilmek üzerinde odaklanmak gerekmektedir.³⁵¹

³⁴⁹ Clark, J. S., Maher J. K., s. 59–71.

³⁵⁰ Parvatiyar, A., Sheth, J. N., s. 1-34.

³⁵¹ Acuner, Ş. A., **Günümüz Pazarlama Anlayışında CRM (Müşteri İlişkileri Yönetimi) Uygulamalarının Yeri ve Önemi, Pazarlama Dünyası**, Mayıs, 2005.

Her sektörde CRM'nin farklı yansımaları görülmektedir. Örneğin havayolları; müşterinin uçtuğu parkurlar, biriktirdiği miller, yılın hangi döneminde ve haftanın hangi gününde daha çok uçtuğu, oturduğu koltuk, yemek tercihleri gibi verileri, bankalar; kredi kartı kullanımı, hizmet kanalı tercihleri, ödemede sorun çıkıp çıkmadığı gibi bilgileri, telefon şirketleri; hangi gün ve hangi saat konuşulduğu, hangi tarifenin tercih edildiği, en çok hangi numaraların arandığı gibi bilgileri, mağazalar; hangi gün ve hangi saatlerde alındığı, hangi ürünlerin hangileriyle birlikte alındığı, ödeme tercihleri gibi verileri toplayarak analiz etmekte ve pazarlama stratejilerini bunun üzerinde kurmaktadır³⁵². Matalan'ın müşteri ilişkileri yönetimini kullanarak müşteri değeri yaratmasındaki başarısı aşağıdaki tabloda özetlenmektedir.

Tablo 3. 24 Müşteri İlişkileri Yönetimi Örneği

Yardım kampanyaları ve sponsorluk anlaşmaları ile halkla ilişkiye girebilmektedir.
Kalite, kalite kontrolü ve maliyetler hakkında tedarikçilerle ilişkiler geliştirmektedir.
Marka bilinirliğini sürekli artırmak, genişleyen ve yenilenen ürünler sunmak, etkili bir hizmet sistemi sağlamak, satış promosyonları yapmak, doğrudan pazarlama yapmak yolu ile müşterilerle iletişim içine girebilmek
Çalışanları ile ilişkilerini geliştirmektedir.

Kaynak: Rowley, J., Haynes, L. , Customer Relationship Management: The Matalan Way, The Marketing Review, No:5, 2005, s. 175-187.

4.2.6 Müşteri Tatmini ve Marka Bağlılığı

Müşteri tatmini, müşterinin hizmet özelliklerine, ürün ya da hizmete, satış elemanına ve diğer değişkenlere ilişkin genel yargısı olarak tanımlanmaktadır³⁵³. Daha genel bir ifadeyle müşteri tatmini, müşterinin ihtiyaç ve isteklerinin karşılanmasıdır. Müşteri tatminin sağlanıp sağlanmadığı satın-alma öncesi beklenti ile satın-alma sonrası performansın karşılaştırılması ile belirlenebilmektedir. Tatmin, müşterinin tutumsal marka bağlılığını artırmaktadır³⁵⁴.

³⁵² Çetin, E. Ş., a.g.e., s.39-40.

³⁵³ Back, K., Parks, S. C., s. 419-435.

³⁵⁴ Clark, J. S., Maher J. K., s. 59-71.

CRM, müşteri odaklı bir anlayış içermektedir. Geleneksel pazarlama anlayışında bir banka bir müşteriye tüm hizmet çeşitlerini sunarken, CRM’de müşterilerin geçmişte aldığı ürün ve hizmetler dikkate alınarak bir hizmet paketi sunulmaktadır. Ya da CRM’de, parfüm alan bir müşteriye 6 ay sonra parfümünün bitmek üzere olacağı düşünülerek ulaşılmakta aynı parfümden almak isteyip istemediği sorulmakta veya yeni parfümler önerilmektedir. Hatta müşteri tarafından beğenileceği düşünülen her yeni koku geldiğinde o müşteriye haber verilebilmektedir. Dolayısıyla satıştan daha önemli olan şey hatırlanabilmektir³⁵⁵. Dolayısıyla CRM, yalnızca kasa sistemleri otomasyonu kurmak ya da puan kazandıran sadakat kartları dağıtmak anlamına gelmemektedir.

CRM, perakende kuruluşlarını kendilerinin etkin bir biçimde tanıtılmalarını, alternatif kanallardan müşteriye ürün satabilmelerini ve müşteri şikayetlerini alıp değerlendirebilmelerini sağlamaktadır. Firmaların satış ve pazarlama gücünü ciddi oranda arttırmakta, müşteri desteği ve hizmetlerini daha etkin ve entegre hale getirerek müşteri ilişkilerinin iyileştirilmesini ve müşteri tatmininin artmasını sağlamaktadır.

Müşteri tatminini sağlamak için müşteriye tanımak, ne istediğini öğrenmek ve pazarlama stratejilerini farklı müşterilere göre farklı bir şekilde yapılandırmak gerekmektedir. İşletme elindeki tüm kaynakları farklı müşterilerin farklı ihtiyaçlarını anlamaya yönlendirmeli ve bu kaynakları farklı müşterilerin farklı ihtiyaçlarına göre ürün ya da hizmet üretim süreç ve teknolojilerine ayırmalıdır³⁵⁶. Marka bağlılığı yaratmak için;

- Yeni müşteri gönderenleri ödüllendirmek
- Teşekkür kartları göndermek,
- Kişisel mektuplar göndermek,
- Telefonla hatırlatmalar yapmak,
- Müşteriye özel promosyonlar hazırlamak,

³⁵⁵ Hamilton, K., **Internet, Branding and CRM**, Westchester County Business Journal, July, 2002, s.11

³⁵⁶ Acuner, Ş. A., **Günümüz Pazarlama Anlayışında CRM (Müşteri İlişkileri Yönetimi) Uygulamalarının Yeri ve Önemi**, Pazarlama Dünyası, Mayıs, 2005, s.58-65

–En önemlisi müşteri şikayetlerini ayrıntılı olarak değerlendirmek ve hızlı cevap vermek gibi yöntemlerin uygulanması yararlı olacaktır.

Günümüzde internet, CRM’yi ve markalamayı destekler nitelikte bulunmaktadır. e-CRM uzun süreli müşteri ilişkilerini sağlamak ve geliştirmek amacıyla internet üzerinden pazarlama faaliyetleri yürütmek, pazarlama araç ve tekniklerini kullanmaktır. Kelley ve Gilbert tarafından yapılan araştırmada, e-CRM’nin internet müşterilerinin bağlılığını doğrudan arttırabildiği görülmüştür. Fiyat duyarlılığı müşterinin satın almasında artık önemli bir faktör olmadığı için e-CRM stratejisinde güvenlik, hizmet kalitesi, hız, uyumluluk gibi diğer faktörler önemli rol oynamaktadır³⁵⁷.

Reichheld ve Sasser’in 2000 yılında yaptıkları araştırmada ise internette giyim pazarında müşteri elde etme maliyetinin geleneksel bir pazara göre %20-40 oranında daha maliyetli olduğu bulunmuştur. Bu yüzden ilk aşamalarda internet aracılığıyla pazarlamada kayıplar büyük olabilmektedir. Ancak 24 ile 30’uncu aylara gelindiğinde internet müşterilerinin ilk 6 aya göre iki kat daha fazla harcama yaptıkları ve fiyata daha az duyarlı oldukları belirlenmiştir.

Oliver’in müşteri tatmini ile marka bağlılığı arasında ilişki olup olmadığını belirlemeye çalışan araştırmasında ilişkinin pozitif yönlü olduğu ortaya çıkmıştır. Araştırma sonucunda müşteri tatmininin tutum geliştirilmesinde önemli rol oynadığı sonucuna ulaşılmıştır. Pozitif tutumlar, ürün ve marka hakkında olan tutumların değiştirilmesini sağlamakta, pozitif düşünceler yeniden alma isteğini arttırmaktadır. Pazarlama literatüründe de müşteri tatmini, müşteri bağlılığı ve karlılık arasında güçlü bir ilişkinin var olduğunu gösteren çalışmalar vardır³⁵⁸. Ancak marka bağlılığının sağlanması için müşteri tatmininin hangi oranda sağlanması gerektiği konusu tartışmalıdır³⁵⁹.

³⁵⁷ Kelley, L. L., Gilbert, D., Mannicom, R., **How e-CRM Can Enhance Customer Loyalty**, Marketing Intelligence & Planning, Vol. 21, No. 4, 2003, s.239-248.

³⁵⁸ Hallowell, R., a.g.m., s. 27-42.

³⁵⁹ Back, K., Parks, S. C., a.g.m., s. 419-435.

Markalar üretici işletmelerle tüketiciler arasında bağlantı noktası oluşturmaktadırlar. Tüketiciler kişisel algılamaları ile markanın kalitesi, yüksek performansı, diğer markalardan farklılığı gibi unsurları değerlendirip ürün veya hizmete fonksiyonel özelliklerin dışında duygusal ve sembolik değerler ekleyerek markalardan faydalanmaktadırlar. Marka değerini oluşturan varlıklar marka bağlılık düzeyini arttırabilmektedir. Algılanan kalite, çağrışımlar ve iyi bilinen marka ismi tüketicilerin satın alma nedenlerini ve tatmin düzeylerini etkileyebilmektedir. Tüketicilerin tatmin düzeyindeki artış markaya karşı sadakat düzeyinin artmasına neden olmaktadır. Marka sadakat düzeyinin artması rakip işletmelerin pazara yeni özelliklere sahip yaratıcı ürünler sunduklarında müşterinin elde tutulmasını sağlamaktadır.

İşletmenin sahip olduğu sadık müşteri düzeyi yüksek olduğu takdirde mevcut müşteriyi elde tutmanın maliyeti yeni müşteri kazanmanın maliyetinden düşük olduğu için işletmenin pazarlama maliyetlerinde azalmaya neden olmaktadır. Mevcut müşterilerin sadakat düzeyi rakiplerin pazara girmesini önleyen doğal engeller oluşturur. Mevcut müşterilerin markadan tatmin olduğu ve markaya sadık olduğu pazarlara girmek ve müşterilerin sadık olduğu markayı değiştirmeye çalışmak rakip işletmeler için oldukça zor ve yüksek maliyetlidir. Marka sadakati, tüketicilerin mağaza seçim kararları üzerinde oldukça etkili bir faktördür.³⁶⁰

Perakendeciler tüketicilerin sadakat düzeyinin yüksek olduğu markaları raflarında bulundurmamak zorundadır. Şayet müşterilerin talep ettikleri markaları raflarında bulundurmazlar ise müşteriler talep ettikleri markaların bulunduğu diğer perakendecileri tercih edeceklerdir. Yine marka sadakati yüksek olan marka isimleri işletmenin yeni geliştirdiği ürünleri markalamada kullanılarak yeni ürünün pazarda başarılı olma ihtimali arttırılabilmektedir. Marka sadakat düzeyi yüksek olan markalar, özellikle riski yüksek satın alma durumlarında muhtemel müşteriler için garanti görevi üstlenirler. Ayrıca yüksek marka sadakat düzeyi rakip işletmelerin yeniliklerine karşı işletmenin ürün ve hizmetlerini rakiplerin düzeyine çekebilmesi için gereken zamanı işletmeye

³⁶⁰ Sohn, S. and Kim, Y. (2008) **Searching customer patterns of mobile service using clustering and quantitative association rule**.Expert Systems with Applications. 34(2) pp.1070-1077

tanınmaktadır. Böylece, marka sadakati tüketicilerin marka değeri algılamalarının en önemli boyutunu oluşturmaktadır³⁶¹.

4.3 İlişkisel Pazarlama

İlişkisel pazarlama en temel anlamda, yeni müşteriler bulmaktan çok mevcut müşterileri elde tutma ve onlarla ilişkileri geliştirme üzerine yoğunlaşan stratejik bir eğilimdir. İlişkisel pazarlamanın amacı belirli bir kârlılıkta müşterilerle ilişki kurulması, sürdürülmesi ve bu ilişkinin geliştirilmesi ve bu şekilde iki tarafın da amaçlarına ulaşmış olmasıdır.

Blomqvist ilişkisel pazarlamayı;

- Her müşterinin ayrı ayrı ele alındığı, her müşterinin ayrı bir birim olarak kabul edildiği,
- Firmanın pazarlama faaliyetlerinin mevcut müşterilere yöneltildiği,
- Firmanın müşteri devir hızını düşürerek ve müşteri ile ilişkilerini güçlendirerek karlılık sağlamaya çalıştığı pazarlama anlayışı olarak tanımlamaktadır.

Peck, başarılı bir ilişkisel pazarlamanın ve yönetiminin uygulanabilmesinin şirket çalışan ve ortakları, tedarikçileri, şirketle anlaşmalı diğer şirketler gibi birçok çıkar grubu ile ilişkilerin kurulmasına bağlamaktadır. İlişkisel pazarlama temel olarak 4 unsura dayandırılmaktadır.

Güven (Trust)

Karşı tarafın sözlerinin itimat edilebilir olduğuna ve yerine getirileceğine ilişkin inançtır. Verilen sözleri yerine getirmek, müşteri tatminini sağlamanın, müşteri tabanını korumanın ve uzun dönemli karlılığı garantilemenin en önemli yollarındandır. İşletme; personel, teknoloji gibi tüm kaynaklarını müşterilerin firmaya güvенеbileceği şekilde kullanmalı ve güçlendirmelidir.

³⁶¹ Avcılar, M. Y., Akın, M. ,**Tüketici Temelli Marka Değeri Kavramı**, Pazarlama Dünyası, Şubat Sayısı, 2007

Bağlılık (Commitment)

Pazarlama ilişkilerinin gücünün belirlenebilmesi ve marka bağlılığının ve gelecekteki satın alma davranışlarının tahmin edilmesinde önemli bir unsurdur. Bağlılık, ilişkilerin devam etmesi yönünde sürekli istek olarak tanımlanabilmektedir. Bu ise ilişkilerin iki taraf için de yararlı ve tatmin edici olması sağlanarak sürdürülebilecektir.

362

İletişim (Communication)

Zamanında ve doğru bilgi sunabilmek olarak tanımlanmaktadır. Günümüzde satış öncesinde, satış esnasında, tüketim esnasında ve tüketim sonrası süreçlerde şirket ile müşteriler arasında karşılıklı diyaloglar da gelişmektedir. Firma için değerli müşterilerle sürekli iletişim halinde olunması ve bu müşterilere hizmetler ve hizmet değişimleri hakkında zamanında ve doğru bilgi sunulması gerekmektedir. Müşterilerin ilgisini çekebilmek, tercihlerini geliştirebilmek (ürünün değerini, performansını ve diğer özelliklerini değiştirerek), onları ikna edebilmek ve alım kararı vermelerini sağlamak için iletişim yolu seçilmelidir. Ayrıca iletişim, tatmin olmayan müşteriler için tatminsizliğe yol açan sebeplerin de düzeltilme fırsatını sağlamaktadır.

Çatışma Yönetimi (Conflict Handling):

İlişkisel pazarlamada çıkabilecek anlaşmazlıklardan kaçınmak, olan anlaşmazlıkları problem yaratmadan önce çözmeye çalışmak ve problemler ortaya çıktığında çözümü için tartışmak gereklidir. Çatışmanın iyi yönetilip yönetilmemesi anlaşmazlığın markaya bağlılıkla mı yoksa markayı almamakla mı sonuçlanacağını belirlemektedir.

³⁶² U. Fayyad, G. Piatetsky-Shapiro ve P. Smyth, Knowledge Discovery and Data Mining: Towards a Unifying Framework, **Proceedings of The Second Conference on Knowledge Discovery and Data Mining**, Portland, Oregon, August, 1996, s.84.

Ndubisi tarafından yapılan arařtırmada güven, baęlılık, iletiřim, çatıřma yönetimi ile müşteri baęlılıęı arasında yüksek bir iliřkinin olduęu görölmüřtür. Baęlı müşterilerin word of mouth ile baęlı oldukları řirket ve ürünleri tanıttıkları görölmektedir. Bu řekilde yeni müşterilerin de kazanılmasını saęlamaktadırlar³⁶³.

İliřkisel pazarlama mevcut müşterilerin bir iřletmenin en deęerli varlıęı olduęunu savunmaktadır. İliřkilerin özünde ise müşteri baęlılıęı yatmaktadır. Müşteri baęlılıęının iřletmeye yararları ařaęıdaki řekilde özetlenebilecektir;³⁶⁴

- Fiyat duyarlılıęını azaltmaktadır.
- Yeni müşteriler kazanılmasında maliyetler azalmaktadır.
- Kârlılık artmaktadır.

³⁶³ Ndubisi, N. O., **Relationship Marketing and Customer Loyalty**, Marketing Intelligence & Planning, Vol. 25, No. 1, 2007, s. 98-106.

³⁶⁴ Wei, C. and Chiu, I. (2002) **Turning telecommunications call details to churn prediction: a data mining approach**. Expert Systems with Applications. 23(2) pp.103-112

5. GSM OPERATÖRLERİ ÜZERİNE BİR ÇALIŞMA

Teknolojik gelişmelerin bir getirisi olan benzer ve hızlı üretim sistemlerinin rekabet şartlarını zorlaması, pazarlama çabalarının iletişim disipliniyle yeniden yorumlanmasını gerektirmiştir. Pazarlama ve iletişim bilimlerinin bir arada yer aldığı pazarlama iletişimi alanı, kurumların ve markaların sunumuna ilişkin yeni yöntem ve kavramlar ortaya çıkarmıştır. İletişim stratejileri, nihai tüketici yerine, geniş bir yelpazede ele alınan hedef kitle üzerine odaklanmıştır.

Bu gelişmelere paralel olarak, hedef kitleyi anlamaya ilişkin tüketici araştırmaları, iletişim çabaları öncesi ve sonrası araştırmalar (pre-test ve post-testlerle geliştirilen ürün ve pazarlama iletişimi çalışmaları) önceki dönemlerde olmadığı kadar önemli hale gelmiştir. Bütün bunlar, hedef kitleye verilen önem kadar, kurum ve markaların imajına ilişkin algı yönetiminin gereklilikleridir.

Benzer ürün ve hizmetlerle rekabet edebilme zorluğu, tutundurma çabalarında, rasyonel vaatler dışında duygusal vaatlere de yer verilmesini gerektirmiştir. Pazarlama iletişimi çabaları, rasyonel önerilerin yanı sıra, duygusal çekiciliklerle de anılır olmuştur. Farklılaşma ve ayrıcalıklı bir konuma yerleşebilme amacıyla geliştirilen tutundurma çabaları kurumsal itibar, marka kişiliği gibi kavramlarla anılır olmuştur.

Günümüzde müşteri odaklı pazarlama için sıklıkla kullanılan veri madenciliği çalışmalarında başarılı olmanın en önemli şartı, projenin hangi işletme amacı için yapılacağını açık bir şekilde tanımlanmasıdır. İlgili işletme amacı, işletme problemi üzerine odaklanmış ve açık bir dille ifade edilmiş olmalı, elde edilecek sonuçların başarı düzeylerinin nasıl ölçüleceği tanımlanmalıdır. 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, üretilen sonucun, diğer bir ifadeyle üretilen bilginin işletme için

değerinin doğru analiz edilmesi gerekmektedir. Analistin işletmede üretilen sayısal verilerin boyutlarını, proje için yeterlilik düzeyinin iyi analiz edilmesi gerekmektedir.

5.1 Araştırmanın Amacı ve Önemi

Günümüzde teknolojik alanda yaşanan gelişmeler, globalleşme, müşterilerin bilinçlenmesi ve artan rekabet gibi nedenlerle birçok işletme stratejisini ve iş yapma biçimini değiştirmektedir. Artık işletmeler müşterilerini elde tutmaya, kaybedilen müşterileri tekrar kazanmaya, onlarla karşılıklı fayda sağlayan uzun süreli bire bir ilişkiler kurmaya çalışmaktadırlar. Müşterilere daha hızlı, daha kaliteli ve daha kişisel ürün ve hizmetler sunabilmek işletmeler için karlılıklarını sürdürebilmek açısından bir zorunluluk haline gelmiştir. Müşteri ilişkileri yönetimi de işletmelerin bu ihtiyaçlarına cevap buldukları yeni bir yönetim anlayışı olarak karşımıza çıkmaktadır.

Yoğunlaşan rekabet ortamında farklı müşterilerin farklı istek ve beklentilerini görebilen, onlara özel kişiselleştirilmiş ürün ve hizmet sunabilen, onları bir değer olarak gören ve iş süreçlerini bu doğrultuda oluşturan, kısaca CRM'i başarılı bir şekilde uygulayan işletmeler müşterilerine ek değerler sunabilmekte ve bu sayede rekabette önemli avantajlar sağlayabilmektedirler. Bununla birlikte, uygulamalara bakıldığında CRM uygulamalarının dünyadaki başarı oranı iyi sonuçlar alınamadığına işaret etmektedir. Amerikan Pazar araştırma kuruluşu olan Meta Group'un 2002 yılında gerçekleştirmiş olduğu araştırma sonucunda tüm CRM uygulamalarının sadece %30'unda istenilen yararların sağlanabildiği ortaya çıkmıştır. Böyle düşük bir oranda geri dönüşlerin alınmasında "işletmelerde gerekli CRM altyapısının oluşturulmadan uygulanması" görüşünün payı büyüktür.

İşletmelerde gerekli altyapı oluşturulmadan CRM'i uygulamak, beklenen yararları sağlamaktan ziyade maliyetleri artırıcı bir faktör olarak büyük sorunlara yol açabilmektedir. Bu nedenledir ki günümüzde CRM'i sadece teknolojik bir çözüm olarak kabul eden işletmelerin girişimleri başarısızlıkla sonuçlanmaktadır. Dolayısıyla, işletmelerin CRM altyapısı ile CRM uygulamalarından elde edilen yararlar arasındaki

ilişkiyi araştırmak hem akademisyenler hem de uygulayıcılar açısından büyük önem taşımaktadır.

Bu çalışmada Müşteri kayıp yönetimi işlemi veri madenciliği araçlarıyla desteklenmiştir. Veri madenciliği karar verme süreçlerinde firmalar için büyük önem taşımaktadır. Yöneticiler açık olmayan veriler içinden veri madenciliği yöntemleri ile birbirleriyle ilişkili verileri ortaya çıkarırlar ve firma için operasyonel, stratejik kararlar verirler. Veri madenciliğinin asıl hedefi saklı bilgiyi geniş veri kümeleri içerisinde çıkarmaktır. Veri madenciliğinde sıklıkla kullanılan bazı teknikler yanında, diğer yöntemlerde sonuçlar kıyaslanmıştır. Çalışmada, Fuzzy(Bulanık) Kümeleme tekniklerinden, amaç fonksiyonlarına dayanan bütün kümeleme tekniklerinin temelini oluşturan Fuzzy(Bulanık) Ortalamalar(BCO) tekniği incelenmiş ve bu teknik kullanılarak Türkiye'deki bir GSM Operatörünün müşteri kaybı analizi yapılmaya çalışılmıştır.

Literatürde CRM'in alt kolu olan müşteri bağımlılığı yönetimi uygulamalarının yararları ve işletmelerin altyapılarını oluşturan boyutlar ayrı ayrı araştırılmıştır. Bu çalışma CRM altyapısını oluşturan boyutların CRM uygulamalarından elde edilen yararları ne denli etkilediğini ortaya koyması nedeniyle önem arz etmektedir.

Tezle ilgili literatür taraması ilgili kitaplar, bilimsel makaleler, konferans ve sempozyum bildirileri, teknik raporlar, ve ilgili tez çalışmalarından yapılmıştır.

Bu amaç ve önem çerçevesinde araştırmanın yapıldığı süreçte, CRM uygulamalarından elde edilen yararlar üzerinde işletmelerin CRM altyapısını oluşturan boyutların etkili olabileceği varsayımı araştırmanın temel dayanak noktasını oluşturmaktadır.

5.2 Araştırmanın Kısıtları

Telekomünikasyon sektörü tüm dünyada ve Türkiye'de büyük bir hızla büyüyor. Geleceğin en parlak sektörü olarak değerlendirilen sektörün dünyadaki büyüklüğü 1.3

trilyon dolar, Türkiye’de ise 20 milyarı aşıyor. Türkiye’de son yıllarda sektörün rekabete açılması büyümede etkili oldu. Araştırmalar, Türkiye mobil iletişim pazarının 2012’ye kadar Avrupa pazarından daha hızlı büyüyerek, yüzde 21’lik artışla 10,5 milyar dolara yükseleceğini gösteriyor. 2009’da küresel krize rağmen Türkiye’nin üç GSM operatörü Turkcell, Vodafone ve Avea’nın yapacağı toplam 3 milyar 300 milyon dolarlık yatırım sektörü daha da hareketlendirecek. Büyük bir hızla gelişen teknoloji hayatımızın her alanını önemli ölçüde etkiliyor. Ses, görüntü ve veri aktarımının gerçekleştirildiği sektör olarak tanımlanabilen “telekomünikasyon” uluslararası alanda teknolojik gelişmelerden en fazla etkilenen sektörler arasında yer alıyor. Telekomünikasyon sektöründeki gelişmeler ve internetin yaygınlaşması dünyadaki sınırları kaldırıyor.

Günümüzde telekomünikasyon hizmetleri sosyal yaşamın ayrılmaz bir parçası olduğu gibi ekonomik kalkınmanın da en temel altyapısıdır. Telekomünikasyonun ekonomik gelişme ve kalkınmadaki önemi, 1980’li yılların başından itibaren dünyayla birlikte ülkemiz ekonomisinde de gözlenen serbestleşme ve küreselleşme eğilimleriyle artmıştır. Bu süreçte, bir yandan bilgiye ve bilgi akışını sağlayan telekomünikasyon hizmetlerine talebin artması ve çeşitlenmesi, aynı doğrultuda telekomünikasyon sektörünü gelişmeye zorlamış, böylelikle hem telekomünikasyon hizmetleri gelişmiş, iç içe yürüyen bu gelişme dünya ekonomisinin küreselleşmesinde önemli rol oynamıştır. Küreselleşme kavramı çerçevesinde, dünya ekonomik sisteminin geçirdiği dönüşümler telekomünikasyon teknolojilerindeki gelişmelerle ilişkilendirilmekte, uluslararası işbölümü telekomünikasyon altyapıları üzerinde yeniden şekillendirilmekte, telekomünikasyona bağımlılığın düzeyinde ve telekomünikasyon aracılığı ile girilen etkinliklerin derecesinde niteliksel bir değişim yaşanmaktadır. Telekomünikasyon hizmetlerinin 1960’lar ve 70’ler boyunca kamu hizmeti olarak sunulması ve evrensel hizmet ilkesi ile yürütülmesi, öte yandan kalkınma düşünceleri, bu doğrultuda sağlık ve eğitim sektörlerinde telekomünikasyonun önemi, ulusal bütünleşme ve güvenlik açısından stratejik bir sektör olarak kabulü bu hizmetlerin kullanımının değişmesine neden olmuş ve telekomünikasyon hizmetlerini yeniden yapılanmakta olan dünya düzeninde önem kazanan hizmet sektörünün en önemli bileşeni haline getirmiştir.

Bu arařtırmada Trkiye'de faaliyet gsteren GSM sektrlerinden, telekomnikasyon hizmetlerinin odağında ki firmalardan veri temini yapılarak saėlanmıřtır. Arařtırmanın kapsamını Trkiye’de faaliyet gsteren GSM operatr řirketleri oluřturmaktadır. Arařtırma gerek mřteri verilerinin temin edilmesinin ok zor olması sebebiyle bir telekomnikasyon firmasına ait gerek 15000 mřterinin 42 farklı zelliėini barındıran veriler zerinde kısıtlanmıřtır.

Bununla birlikte, bazı iřletmeler de kendilerine ulařıldıėı halde arařtırmaya dahil olmayı reddetmiřlerdir. Bu iřletmeler CRM'in yeni bir uygulama olduėunu ve bir rekabet stratejisi olarak grdklerini, mřteri iliřkileri ynetimi hakkında bir tez alıřmasına katkıda bulunmalarının kendilerini gvenlikleri aısından zor durumda bırakabileceėini belirtmiřlerdir.

5.3 Literatr Arařtırması

Mřteri kayıp ynetimi zerine bir ok alıřma mevcuttur. “Churn Management”, “Churn Prediction”, “Data Mining” ve “ANFIS,” anahtar kelimeleri ile science direct veritabanında arama yapıldıėında yapılan yayın alıřmalarının sayısı saptanabilir.

Tablo 5.1’den 5.4’e kadar olan tablolarda 2001 ve 2011 arası ve 2001 ncesi yıllarda yapılan yayın sayıları belirtilmiřtir.

Yapılan aramalar gsteriyor ki veri madenciliėi ve sonrasında churn management en fazla popüler arařtırma yapılan alanlardır.

Mřteri kayıp ynetimi anahtar kelimeleriyle science direct’te arama yapıldıėında toplamda 1665 yayın olduėu grlmektedir.

Tablo 5. 1 Yıllara Göre Yapılan Müşteri Kayıp Yönetimi Yayın Sayısı

Year	Number of journal
2011	179
2010	187
2009	188
2008	188
2007	143
2006	85
2005	65
2004	76
2003	59
2002	63
2001 ve Öncesi	88

Müşteri kayıp tahmini anahtar kelimeleriyle science direct'te arama yapıldığında toplamda 1352 yayın olduğu görülmektedir.

Tablo 5. 2 Yıllara Göre Yapılan Müşteri Kayıp Tahmin Yayınlarının sayısı

Year	Number of journal
2011	127
2010	127
2009	127
2008	104
2007	88
2006	69
2005	65
2004	45
2003	56
2002	44
2001 ve Öncesi	115

Veri madenciliği anahtar kelimeleriyle science direct'te arama yapıldığında toplamda 23161 yayın olduğu görülmektedir.

Tablo 5. 3 Yıllara Göre Yapılan Veri Madenciliği Yayınları Sayısı

Year	Number of journal
2011	3481
2010	3140
2009	3050
2008	2702
2007	2431
2006	2043
2005	1505
2004	1335
2003	1191
2002 ve Öncesi	1730

ANFIS anahtar kelimesiyle science direct'te arama yapıldığında toplamda 1602 yayın olduğu görülmektedir.

Tablo 5. 4 ANFIS Araştırma Yayınları Sayısı

Year	Number of journal
2011	298
2010	225
2009	260
2008	193
2007	133
2006	112
2005	94
2004	61
2003	61
2002	31
2001 ve öncesi	102

Veri madenciliği bilgisayar bilimlerinde çok popüler bir araştırma alanıdır. Veri madenciliği yöntemleri kullanılarak yapılan birçok Müşteri kayıp yönetimi çalışması mevcuttur. Her geçen yıl yeni yayınlar yayınlanmaktadır.

Telekomünikasyon terminolojisinde, müşteri telekomünikasyon firmasından bir sure servis almak üzere sisteme tanımlanır, müşteri kaybı bu müşterinin bağlı olduğu hizmeti iptal etmesiyle gerçekleşir. Aşağıda değişik müşteri kayıp tiplerini istelenmiştir.

Tarife Müşteri Kaybı

Bu müşteri kayıp sınıfında, müşteri aynı telekomünikasyon firmasında farklı bir hizmet paketine değişiklik yapar. Bu bir operator kaybı değildir ve firmaya çok fazla olumsuz etkisi olmaz.³⁶⁵

Hizmet Müşteri Kaybı

Bu sınıfta müşteri aldığı hizmeti tamamen terk eder.³⁶⁶

Operatör Müşteri Kaybı

Müşteri bir başka GSM operatörü için aldığı hizmetten ayrılır. Bu sınıf müşteri kaybında, GSM telekomünikasyon firması müşterilerini kaybetmiş olur. Müşteri kayıp yönetiminin en önemli hedefi operator kayıplarını tespit etmektir, çünkü operator müşteri kaybı müşteri kaybedildiğinin habercisidir.³⁶⁷

Telekomünikasyon sektörü için müşteri kaybetmemek çok ama çok önemli hale gelmiştir. Bu nedenle varolan müşterilerin memnuniyetlerini devam ettirmek yeni müşteri kazanmaktan daha önemli hale gelmiştir.³⁶⁸

Müşteri kayıp yönetiminde, firmalar iki farklı strateji izlerler. Bunlardan bir tanesi Reactive Churn Management iken diğeri Proactive Churn Management'dır. Reactive Churn Management'da GSM operatörleri, hizmetten memnun olmayan ve

³⁶⁵ B.Q. Huang & T.-M. Kechadi, 2010

³⁶⁶ B.Q. Huang & T.-M. Kechadi, 2010

³⁶⁷ B.Q. Huang & T.-M. Kechadi, 2010

³⁶⁸ Reinartz & Kumar, 2003

iptal ettirmek isteyen bayan veya erkek müşterilerini elde tutmak için yeni teklifler sunarlar. Öte yandan proactive Churn management’da müşterinin hizmeti terk etmesi beklenmeden müdahale edilir. Bu sistemde GSM operatörünü terketme olasılığı yüksek olan müşteriler tespit edilmeye çalışılır.³⁶⁹

Müşteri kaybının birçok sebebi vardır. Bunların arasında ücretlendirme, kapsama alanı, görüşme kalitesi müşteri memnuniyeti için en önemli olanlarıdır. Bir veya birden fazla sebep olduğunda müşteri kaybı gerçekleşmiş olacaktır.³⁷⁰ (R.N. Bolton et al., 1998).

Müşteri memnuniyeti müşteri davranışlarının periodik olarak gözlenmesiyle ölçülebilir. Bu gözlem süreci bazı adımlarda oluşmaktadır. Öncelikle müşterilerin demografik özellikleri ve hesap dökümleri very ambarlarında saklanır. Yöneticiler bu veritabanını müşterilerini kaybetme durumlarını elde etmek için kullanırlar. Sonuca ulaşmak için birçok veri madenciliği teknikleri kullanılmaktadır. (Berson, A., Smith, S., & Thearling, K. et al., 2000).

5.4 Materyal

Bu çalışmada aşağıda takip eden programlar kullanılmıştır.

- Weka Version 3.6: WEKA veri madenciliği görevleri için, makine öğrenme algoritmalarının bulunduğu Javada yazılmış bir programdır. WEKA veri ön işleme, sınıflandırma, regresyon, dernek kuralları, görselleştirme birliktelik kuralları ve kümeleme için araçlarda içerir.
- KNIME (Konstanz Information Miner) Version 2.3.0: KNIME veri entegrasyonu ve analizi için kullanışlı bir programdır. KNIME açık kaynak kodlu olup, very madenciliği araçlarında içerir.
- MATLAB Versiyon 7.1: MATLAB temel olarak nümerik hesaplama, grafiksel

³⁶⁹ Burez and Van den Poel , 2007

³⁷⁰

veri gösterimi ve programlamayı içeren teknik ve bilimsel hesaplamalar için yazılmış yüksek seviyeli bir dildir.

5.4.1 Verilerin Hazırlanması ve Analizi

Burada kullanılacak verinin kalitesi sonuçları da etkileyeceğinden kullanılacak verilerin öncelikle ön işlemden geçirilmesi büyük bir önem taşımaktadır. Sonuçta kaliteli verilerden ancak kaliteli çıktılar elde edilebilecektir. Bu nedenle verilerin kalitesini artırmanın yolu, verilerin ön işlemden geçirilmesidir.

Modelin kurulması aşamasında ortaya çıkacak sorunlar, bu aşamaya sık sık geri dönülmesine ve verilerin yeniden düzenlenmesine neden olacaktır. Bu durum verilerin hazırlanması ve modelin kurulması aşamaları için, bir analistin veri keşfi sürecinin toplamı içerisinde enerji ve zamanının %50 - %85'ini harcamasına neden olmaktadır. Bu aşamada firmanın mevcut bilgi sistemleri üzerinde ürettiği sayısal bilginin iyi analiz edilmesi, veriler ile mevcut iş problemi arasında ilişki olması gerektiği de unutulmamalıdır. Proje kapsamında kullanılacak sayısal verilerin, hangi iş süreçleri ile elde edildiği de bu veriler kullanılmadan analiz edilmelidir. Bu sayede analist veri kalitesi hakkında fikir sahibi olabilir. Verilerin hazırlanması aşaması kendi içerisinde toplama, değer biçme, birleştirme ve temizleme, seçme ve dönüştürme adımlarından oluşmaktadır.

Bu çalışmada kullanılan veriler Türkiye’de bir telekomünikasyon firmasından temin edilmiştir. Veri kümesi içerisinde sadece demografik, görüşme ve ücretlendirme verileri mevcuttur. 15000 müşteri için iki veri kümesi vardır. Demografik veri kümesi ve Görüşme veri kümesidir.

Demografik veri kümesinde müşterinin ailesi ve finansal durumu hakkında bilgiler bulunur. Öte yandan görüşme veri kümesinde, müşteri görüşme alışkanlıkları bilgileri mevcuttur.

Veri hazırlanma sürecinde boş ve eksik değerleri anlamlı değerlerle değiştirmeyi denedik. Bu sorunu çözmek için KNIME kullanıldı. Knime’nin eksik

değer algoritması, değişecek eksik değeri seçmemizi sağlamıştır. Her sütundaki eksik değerler yerine minimum değerler verildi.

Aşağıdaki veritabanında her bir farklı müşteri için veriler aittir. Her özniteliğin değeri bir müşteri için değişken anlamındadır.

Veri kümesinde 42 farklı alan vardır, ve bu alanlar Weka analiz programı tarafından işleme sokulur. Veri Weka program tarafından analiz edilir ve veriye ait özet istatistikler(Maksimum, minimum, ortalama, standart sapma, varyans ve Overall Sum) olarak belirtilir.

Bu çalışmada kullanılan veriler Türkiye’de bir telekomünikasyon firmasından temin edilmiştir. Veri kümesi içerisinde sadece demografik, görüşme ve ücretlendirme mevcuttur. 15000 müşteri için iki veri kümesi vardır.

–Demografik veri kümesi

–Görüşme veri kümesi

Demografik veri kümesinde müşterinin ailesi ve finansal durumu hakkında bilgiler bulunur.Öte yandan görüşme veri kümesinde, müşteri arama alışkanlıkları bilgi mevcuttur.

Veri hazırlanma sürecinde boş ve eksik değerleri anlamlı değerlerle değiştirmeyi denedik. Bu sorunu çözmek için KNIME kullanıldı. Knime’nin eksik değer algoritması, değişecek eksik değeri seçmemizi sağlamıştır. Her sütundaki eksik değerler yerine minimum değerler vermeye karar verdik.

Aşağıdaki veritabanında her bir farklı müşteri için veriler aittir. Her özniteliğin değeri bir müşteri için değişken anlamındadır.

Tablo 5. 5 Veri Kümesindeki Değişkenler

Veri İsmi	Veri Tipi	Detay
customerid	numerik	ID – Müşteri Kimlik Numarası
region	numerik	Geographic Indicator - Bölge
tenure	numerik	Months with service – Kullanım Süresi
age	numerik	Age in years – Yaş Durumu
marital	numerik	Marital Status – Medeni Durum
address	numerik	Years at current address – Güncel Adreste bulunulan süre
income	numerik	Household income in thousand – Ev Halkı gelir bilgisi (1000 ile çarpılacak)
ed	numerik	Level of education – Eğitim Seviyesi
employ	numerik	Years with current employer – Güncel iş
retire	numerik	Retired – Emeklilik Durumu
gender	numerik	Gender - Cinsiyet
reside	numerik	Number of people in household – Ev halkı sayısı
tollfree	numerik	Toll free service - Ücretsiz Görüşme Servisi
equip	numerik	Equipment rental – Cihaz kiralama
Callcard	numerik	Calling card service – Arama Kartı Hizmeti
wireless	numerik	Wireless service – Kablosuz hizmeti
longmon	numerik	Long distance last month – Uzun mesafe geçen ay arama
tollmon	numerik	Toll free last month – geçen ay ücretsiz görüşme
equipmon	numerik	Equipment last month – geçen ayki cihaz
Cardmon	numerik	Calling card last month – geçen ayki arama kartı
wiremon	numerik	Wireless last month – Geçen ay kablosuz hizmeti
longten	numerik	Long distance over tenure – üyelik üzerinde uzun mesafe
Veri İsmi	Veri Tipi	Detay
tollten	numerik	Toll free over tenure – ücretsiz görüşme üyelik üzerinde
equipten	numerik	Equipment over tenure – üyelik üzerinde cihaz
cardten	numerik	Calling card over tenure – üyelik üzerinde arama kartı
wireten	numerik	Wireless over tenure – üyelik üzerinde kablosuz
multline	numerik	Multiple lines- çoklu bağlantı
voice	numerik	Voice mail – sesli posta
pager	numerik	Paging Service – çağrı cihazı
internet	numerik	Internet - İnternet
callid	numerik	Caller ID - arayan kimliği
callwait	numerik	Call waiting – arama bekletme
forward	numerik	Call forwarding – arama yönlendirme
confer	numerik	3- way calling - 3 yönlü arama
ebill	numerik	Electronic billing – elektronik fatura
loglong	numerik	Log-long distance – uzun mesafe oturum
logtoll	numerik	Log-toll free – ücretsiz görüşme oturum
logequi	numerik	Log-equipment – oturum ekipmanı
logcard	numerik	Log-calling card – arama kartı oturumu
logwire	numerik	Log-wireless – kablosuz oturumu
lninc	numerik	Log income – gelir oturumu
custcat	numerik	Customer category – müşteri tipi

5.4.2 Veri Anlaşılabilirliği

Veri kaynağında 15000 müşteri mevcuttur. Veri kaynağı iki ayrı veri kümesine bölünmeliydi. Bunlar demografik ve görüşme detayı bilgileridir.

Demografik Veri Girişi

- Müşteri bölgesi,
- Yaş,
- Medeni durum,
- Adres,
- Gelir,
- Eğitim Durumu,
- Cinsiyet,
- Müşteri kayıt tarihi,
- Emekliye ayrılma durumu.
- Aynı adreste bulunma süresi
- İş - Uzmanlık
- Ev halkı sayısı

Görüşme Detayı Veri Girişi

- Ücretsiz konuşma hizmeti
- Cihaz kiralama
- Kablosuz hizmeti
- Kartvizit hizmeti
- Uzun mesafe kullanım süresi
- Internet
- Sesli posta
- Arama bekletme
- Arama yönlendirme

- 3 kişilik görüşme
- E-fatura
- Müşteri kategorisi
- Çağrı hizmeti

5.5 Araştırma Modelinin 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 yenilenen bir süreçtir.

Bir modelin doğruluğunun test edilmesinde kullanılan en basit yöntem basit geçerlilik 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. (Doğruluk Oranı = 1 - Hata Oranı)

Sınırlı miktarda veriye sahip olunması durumunda, kullanılabilecek diğer bir yöntem ise çapraz geçerlilik (Cross Validation) testidir. Bu yöntemde veri kümesi tesadüfi olarak iki eşit parçaya ayrılır.

Bir diğer önemli değerlendirme kriteri ise modelin anlaşılabilir olmasıdır. Bazı uygulamalarda doğruluk oranlarındaki küçük artışlar çok önemli olsa da, bir çok işletme uygulamasında ilgili kararın niçin verildiğinin yorumlanabilmesi çok daha büyük önem taşıyabilir.

Bu çalışmada kümeleme metotlarından bölümlmeli yöntem kullanılmıştır. Bu yöntemde n tane nesnenin olduğu veri tabanında, nesneler mantıksal gruplara ayrılarak

analiz edilir. Merkez tabanlı kümeleri tespit etmede başarılıdır. Gruplandırma yapılırken değişik kriterler değerlendirilebilir. Yapılan gruplandırma analizin kalitesine etki eder.

Bu metotta çok boyutlu uzaydaki nesnelerin birbirlerine geometrik uzaklıklarının ve benzerliklerin belirlenmesi Öklid uzaklık ölçütü ile sağlanmıştır.

5.5.1 Modelin İzlenmesi

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.

Ne kadar etkin olursa olsun hiçbir veri madenciliği algoritmasının üzerinde inceleme yapılan işin ve verilerin özelliklerinin bilinmemesi durumunda fayda sağlanması mümkün değildir. Bu nedenle aşağıda tanımlanan tüm aşamalardan önce, iş ve veri özelliklerinin öğrenilmesi başarının ilk ve temel şartı olacaktır.

Başarılı bir veri madenciliği projelerinde sırasıyla; Problemin Tanımlanması, Verilerin Hazırlanması, Modelin Kurulması ve Değerlendirilmesi, Modelin Kullanılması, Modelin izlenmesi adımları yer almaktadır.

5.5.2 Modelin Kullanılması

Kurulan ve geçerliliği kabul edilen model doğrudan bir uygulama olabileceği gibi, bir başka uygulamanın alt parçası olarak da kullanılabilir. Kurulan modeller risk analizi, kredi değerlendirme, dolandırıcılık tespiti gibi işletme uygulamalarında doğrudan kullanılabileceği gibi, promosyon planlaması simülasyonuna entegre edilebilir veya tahmin edilen envanter düzeyleri yeniden sipariş noktasının altına

düştüğünde, otomatik olarak sipariş verilmesini sağlayacak bir uygulamanın içine de gömülebilir.

5.5.2.1 C Ortalamaları Algoritması

Kümelemenin sonucunda, N tane farklı örüntü c kümesi için bulunur. Bir küme bulunurken, veri içerisinde en önemli nokta benzerliktir.(yüksek veya düşük) Bu sebeple Veriler arasındaki uzaklığı bulmak gereklidir ve böylece benzerlik ölçülebilir. Her bir kümenin bir geometrik merkezi vardır;bu sebeple, merkezler ve kümelerin sayısı eşittir. Kümeler merkezleriyle tanımlanırlar.

C ortalamaları algoritması N tane nesneyi(X_k) kısımlara bölerek “ c boş olmayan küme” üretir. Bu algorithmadan önce her bir küme bir geometric merkez sahiptir ve merkez nokta aşağıda gibi hesaplanır.

$$V_i = \frac{\sum_{X_k \in U_i} X_k}{|C_i|}$$

Bu formülde;

$|C_i|$ U_i ’deki toplam nesne sayısını temsil eder.

The process of computing centroids and finding new membership is repeated until no new membership can’t be found.

Geometrik merkez hesaplama ve yeni üyelikleri bulma işlemi yeni bir üyelik bulunmayana kadar devam edecektir.

5.5.2.2 Bulanık C Ortalamaları Algoritması

Bulanık C ortalaması (FCM) bulanık bölünmede ki kümeleme teknikleri içerisinde en iyi bilinen ve en geniş kullanılanıdır. FCM metod nesneleri, iki veya daha fazla kümeye sahip olabilirler. FCM mantığında her bir nesne $[0,1]$ aralığındaki değerlerle bir kümeye aittir. Her bir nesnenin üyelik değerlerinin toplamı “1” olmak zorundadır. Bir veri noktası merkeze ne kadar yakınsa o kümeye ait olduğu anlaşılır.

FCM kümeleme algoritması, Dunn tarafından 1973 yılında yaratılmıştır ve Bezdek tarafından 1981 yılında geliştirilmiştir. Bulanık C ortalaması (FCM) iki ana basamakta oluşur. İlk basamakta algoritma her bir kümenin merkez noktasını bulacaktır. Sonraki basamakta her bir nesne merkeze uzaklığına bakılarak ait olduğu kümeye atanacaktır.

Algoritma hedef fonksiyonu minimize etmek için yürütülür. Bu en küçük kareler metodunun genellemesidir.

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2, \quad 1 \leq m < \infty,$$

Burada m bulanıklaştırıcıdır. Algoritma rastgele U üyelik matrisleri üretmekle başlar. Bir sonraki basamak ise merkezlerin hesaplanmasıdır. Merkez aşağıdaki denklem ile bulunur:

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m x_i}{\sum_{i=1}^N u_{ij}^m}$$

Kümeye göre, U matris merkezleri ikinci denklem ile hesaplanır. Bir önceki ve güncel U matrisleri kıyaslanırlar ve bu basamak ikisi arasındaki fark ε ’den az olana kadar devam eder.

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{\|x_i - c_i\|}{\|x_i - c_k\|} \right)^{2/(m-1)}}$$

5.5.2.3 ANFIS - Adaptif Ağ Yapısına Dayalı Bulanık Çıkarım Sistemi

ANFIS öğrenme algoritması iki metod içeren hybrid bir algoritmadır. “Adaptive Neuro-Fuzzy Inference System” (ANFIS) algoritmaları öncelikle yapay öğrenmeye dayalı elektro-motor sistemlerinde kullanılmıştı. Daha sonraları algoritmanın adaptif olması ve bulanık mantığı da içermesi bulanık veri madenciliği işlemlerinde kullanılabileceği sonucunu doğurdu. Telekomünikasyon dolandırıcılığı, gibi bazı alanlarda veri madenciliği anlamında kullanılan sistem çalışmamıza konu olan müşteri kayıp yönetimi uyum göstermektedir.

Adaptif ağ biçimindeki bulanık çıkarım sistemi yada kısaca ANFIS yapısı, modellemenin yanı sıra denetim uygulamalarında da kullanılmakta olup, bulanık ve sinirsel yaklaşımların birlikte kullanımına oldukça güzel bir örnek teşkil etmektedir.

ANFIS ağ yapısına ait parametreler, giriş ve çıkış parametreleri olarak iki kısımda ele alınır. Toplam parametre kümesi; $S = S1 + S2$ şeklinde ifade edilecek olursa, S1 giriş parametrelerine, S2 de çıkış parametrelerine karşılık gelir. Hibrid algoritmanın ilk aşaması yada ileri yön geçişi ile ANFIS ağına ait giriş parametreleri, ikinci aşamada yada geri yön geçişinde ise ağın sonuç parametreleri güncellenir. Hibrid öğrenme algoritmasının ileri yön geçişini oluşturan ve en küçük kareler tahmini (LSE) yöntemi kullanılmak suretiyle gerçekleştirilen kısımda, ağın girişindeki üyelik fonksiyonlarına ait parametreler yada kısaca S1 ile belirtilen giriş parametreleri sabit tutulur. Böylece ağın çıkışı, S2 parametre kümesi içinde bulunan çıkış parametrelerinin, lineer bir kombinasyonu haline gelir. Modellenecek sisteme ait P tane giriş-çıkış verisi yada eğitim datası ile S1 giriş parametrelerinin verilen değerleri kullanılarak, ağın çıkışı, matrisel formda, $A\theta = B$ şeklinde ifade edilebilir. Matrisel eşitlikte yer alan θ

vektörü, S2 çıkış parametrelerinin elemanlarından oluşan, bilinmeyen vektördür. Bu denklem, standart lineer en küçük-kareler (least-squares) problemini gösterir ve θ için en iyi çözüm, $\|A\theta - B\|^2$ 'nin minimum değeri olan, en küçük kareler tahmini (LSE) θ^* dır.

ANFIS modeli uygulanırken bazı önemli özelliklere dikkat edilmesi gerekir. Bunlar üyelik fonksiyonu, bulanık kurallar sayısı ve eğitim devirleridir. Bu özelliklerinin değerleri çok önemlidir. Eğer özelliklere yanlış değerler verilirse, algoritma tabiki yanlış sonuçlar üretecektir.

Eğitim işleminin hedefi ANFIS çıktısı ve gerçek hedef arasındaki hata oranını azaltmaktır.

5.6 Bulgular

Bulgular üç ana aşamada gözden geçirilecektir. İlk aşamada, veri analzi edilir ve sonuçlar ikinci aşamaya girdi olarak verilir. İkinci aşamada kümeler yaratılır ve veri alt gruplara toplanır. İkinci aşama olan “kümeleme aşaması” son aşama için çok önemlidir. Üçüncü aşama “tahmin aşaması”dır. Bu aşamada bir müşteri kaybı olup olmadığı tahmin edilir ve kurallar oluşturulur.

5.6.1 Veri Analiz Aşaması

Veri kümesinde 42 farklı alan vardır, ve bu alanlar Weka analiz programı tarafından işleme sokulur.

Veri Weka program tarafından analiz edilir ve veriye ait özet istatistikler (Maksimum, minimum, ortalama, standart sapma, varyans ve Overall Sum) aşağıda gösterilmiştir.

Tablo 5. 6 İlk 7 Özniteliğin Analizi

Row ID	D region	D tenure	D age	D marital	D address	D income	D ed
Minimum	1	1	18	0	0	9	1
Maximum	3	72	77	1	55	1,668	5
Mean	2.022	35.526	41.684	0.495	11.551	77.535	2.671
Std. deviation	0.816	21.35	12.553	0.5	10.082	106.994	1.222
Variance	0.666	455.816	157.577	0.25	101.646	11,447.758	1.493
Overall sum	30,330	532,890	625,260	7,425	173,265	1,163,025	40,065

Satır No	Bölge	Kullanım Süresi	Yaş	Medeni Hal	Adres	Gelir	Eğitim
Minimum	1	1	18	0	0	9	1
Maksimum	3	72	77	1	55	1,668	5
Ortalama	2.022	35.536	41.684	0.495	11.551	77.535	2.671
Standart Sapma	0.816	21.35	12.553	0.5	10.082	106.994	1.222
Varyans	0.666	455.816	157.577	0.25	101.646	11,447.758	1.493
Genel Toplam	30,330	532.890	625,260	7,425	173,265	1,163,025	40,065

Tablo 5. 7 İkinci 7 Özniteliğin Analizi

Row ID	D employ	D retire	D gender	D reside	D tollfree	D equip	D callcard
Minimum	0	0	0	1	0	0	0
Maximum	47	1	1	8	1	1	1
Mean	10.987	0.047	0.517	2.331	0.474	0.386	0.678
Std. deviation	10.077	0.212	0.5	1.435	0.499	0.487	0.467
Variance	101.554	0.045	0.25	2.06	0.249	0.237	0.218
Overall sum	164,805	705	7,755	34,965	7,110	5,790	10,170

Satır No	istihdam	çekilme	Cinsiyet	ikamet	Ücretsiz görüşme	ekipman	Arama kartı
Minimum	0	0	0	1	0	0	0
Maksimum	47	1	1	8	1	1	1
Ortalama	10.987	0.047	0,517	2.331	0.474	0.386	0.678
Standart Sapma	10.077	0.212	0.5	1.435	0.499	0.487	0.467
Varyans	101.554	0.045	0.25	2.06	0.249	0.237	0.218
Genel Toplam	164,805	705	7,755	34,965	7,110	5,7900	10,170

Tablo 5. 8 Üçüncü 7 Özniteliğin Analizi

Row ID	D wireless	D longmon	D tollmon	D equipmon	D cardmon	D wiremon	D longten
Minimum	0	0.9	0	0	0	0	0.9
Maximum	1	99.95	173	77.7	109.25	111.95	7,257.6
Mean	0.296	11.723	13.274	14.22	13.781	11.584	574.05
Std. deviation	0.457	10.359	16.894	19.06	14.078	19.71	789.606
Variance	0.208	107.302	285.415	363.27	198.188	388.493	623,476.973
Overall sum	4,440	175,846.5	199,110	213,297	206,715	173,758.5	8,610,750.75

Satır No	Kablosuz	Geçen ay Uzun mesafe	Geçen ay Ücretsiz görüşme	Geçen ay ekipman	Geçen ay arama kartı	Geçen ay kablosuz hizmeti	Üyelik üzerinde uzun mesafe
Minimum	0	0.9	0	0	0	0	0.9
Maksimum	1	99.95	173	77.7	109.25	111.95	7,257.6
Ortalama	0.296	11.723	13.274	14.22	13.781	11.584	574.05
Standart Sapma	0.457	10.359	16.894	19.06	14.078	19.71	789.606
Varyans	0.208	107.302	285.415	363.27	198.188	388.493	623,747.973
Genel Toplam	4,440	175,846.5	199,110	213,297	206,715	173,758.5	8,610,750.75

Tablo 5. 9 Dördüncü 7 Özniteliğin Analizi

Row ID	D tollten	D equipten	D cardten	D wireten	D multiline	D voice	D pager
Minimum	0	0	0	0	0	0	0
Maximum	5,916	5,028.65	7,515	7,856.85	1	1	1
Mean	551.259	465.633	605.774	442.737	0.475	0.304	0.261
Std. deviation	915.319	856.873	829.739	971.018	0.499	0.46	0.439
Variance	837,809.341	734,231.15	688,466.08	942,875.519	0.249	0.212	0.193
Overall sum	8,268,877.5	6,984,492.75	9,086,606.25	6,641,053.5	7,125	4,560	3,915

Satır No	Ücretsiz görüşme üyelik üzerinde	Üyelik üzerinde cihaz	Üyelik üzerinde arama kartı	Üyelik üzerinde kablosuz	Çoklu bağlantı	Sesli posta	Çağrı cihazı
Minimum	0	0	0	0	0	0	0
Maksimum	5,916	5,028.65	7,515	7,856.85	1	1	1
Ortalama	551.259	465.633	605.774	442.737	0.475	0.304	0.261
Standart Sapma	915.319	856.873	829.739	971.018	0.499	0.46	0.439

Varyans	837,809.341	734,231.15	688,466.08	942,875.519	0.249	0.212	0.193
Genel Toplam	8,268,877.5	6,984,492.75	9,086,606.25	6,641,053.5	7,125	4,560	3,915

Tablo 5. 10 Beşinci 7 Özniteliğin Analizi

Row ID	D internet	D callid	D callwait	D forward	D confer	D ebill	D loglong
Minimum	0	0	0	0	0	0	-0.105
Maximum	1	1	1	1	1	1	4.605
Mean	0.368	0.481	0.485	0.493	0.502	0.371	2.182
Std. deviation	0.482	0.5	0.5	0.5	0.5	0.483	0.734
Variance	0.233	0.25	0.25	0.25	0.25	0.233	0.539
Overall sum	5,520	7,215	7,275	7,395	7,530	5,565	32,731.65

Satır No	İnternet	Arama kimliği	Arama bekletme	Arama yönlendirme	Çok yönlü arama	Elektronik fatura	Uzun mesafe oturum
Minimum	0	0	0	0	0	0	-0.105
Maksimum	1	1	1	1	1	1	4.605
Ortalama	0.368	0.481	0.485	0.493	0.502	0.371	2.182
Standart Sapma	0.482	0.5	0.5	0.5	0.5	0.483	0.734
Varyans	0.233	0.25	0.25	0.25	0.25	0.233	0.539
Genel Toplam	5,520	7,215	7,275	7,395	7,530	5,565	32,731.65

Tablo 5. 11 Son Altı Özniteliğin Analizi

Row ID	D logtoll	D logequi	D logcard	D logwire	D lninc	D custcat
Minimum	1.749	2.734	1.012	2.701	2.197	1
Maximum	5.153	4.353	4.694	4.718	7.419	4
Mean	3.24	3.568	2.854	3.598	3.957	2.487
Std. deviation	0.413	0.277	0.557	0.367	0.803	1.12
Variance	0.171	0.077	0.31	0.134	0.645	1.254
Overall sum	23,082.901	20,659.251	29,027.303	15,976.435	59,358.05	37,305

Satır No	Ücretsiz görüşme oturumu	Oturum ekipmanı	Arama kartı oturumu	Kablosuz oturumu	Gelir oturumu	Müşteri tipi
Minimum	1.749	2.734	1.012	2.701	2.197	1
Maksimum	5.153	4.353	4.694	4.718	7.419	4
Ortalama	3.24	3.568	2.854	3.598	3.957	2.487
Standart Sapma	0.413	0.277	0.557	0.367	0.803	1.12
Varyans	0.171	0.077	0.31	0.134	0.645	1.254
Genel Toplam	23,082.901	20,659.251	29,027.303	15,976.435	59,358.05	37,305

Veride 42 farklı öznitelik mevcuttur. Takip eden tablolarda kategorize edilmiş değişkenlerin analiz bilgileri vardır. Değişkenler ve değerleri aşağıda gösterilmektedir.

Tablo 5.12’de Medeni Durum değerleri gösterilmektedir.

Tablo 5. 12 Müşteri Medeni Durum Değerleri

Değerler	Anlam
1	Evli
0	Bekar

Tablo 5.13’de Cinsiyet değerleri gösterilmektedir.

Tablo 5. 13 Müşteri Cinsiyet Değerleri

Değerler	Anlam
1	Erkek
0	Kadın

Tablo 5.14’da Eğitim Durumu değerleri gösterilmektedir.

Tablo 5. 14 Müşteri Eğitim Değerleri

Değerler	Anlam
1	İlköğretim Okulu
2	Lise
3	Üniversite
4	Yüksek Lisans
5	Doktora

Tablo 5.15’de Adres değerleri gösterilmektedir.

Tablo 5. 15 Adres Değerleri

Adana	Bilecik	Erzurum	Karaman	Mersin	Tekirdağ
Adıyaman	Bingöl	Eskişehir	Kars	Muğla	Tokat
Afyon	Bitlis	Gaziantep	Kastamonu	Muş	Trabzon
Ağrı	Bolu	Gazimağusa	Kayseri	Nevşehir	Tunceli
Aksaray	Burdur	Giresun	Kilis	Niğde	Uşak
Amasya	Bursa	Girne	Kırıkkale	Ordu	Van
Ankara	Çanakkale	Gümüşhane	Kırklareli	Osmaniye	Yalova
Antalya	Çankırı	Hakkâri	Kırşehir	Rize	Yozgat
Ardahan	Çorum	Hatay	Kocaeli	Sakarya	Zonguldak.
Artvin	Denizli	Iğdır	Konya	Samsun	
Aydın	Diyarbakır	Isparta	Kütahya	Siirt	
Balıkesir	Düzce	İstanbul	Lefkoşa	Sinop	
Bartın	Edirne	İzmir	Malatya	Sivas	
Batman	Elazığ	Kahramanmaraş	Manisa	Şanlıurfa	
Bayburt	Erzincan	Karabük	Mardin	Şırnak	

Tablo 5.16’de Uzmanlık Alanı değerleri gösterilmektedir.

Tablo 5. 16 Uzmanlık Alanı Değerleri

Akademisyen / Üniversite Öğretim Görevlisi	Hatalı Girilmiş	Pilastik Sanatlar (Ressam Heykeltıraş ... vb.)
Analist / Programcı	Hemşire / Ebe	Polis / Güvenlik Görevlisi
Astsubay	Hizmet / Ticaret	Reklam / Halkla ilişkiler
Aşçı / Garson / Barmen	Hostes	Sağlık Personeli
Avukat	İmalat / Sanayi	Sahne Sanatları (Bale Tiyatro ... vb.)
Büyük Sanayici	İnsan Kaynakları	Sekreter / Yönetici Asistanı
Çalışmayan	İşçi	Serbest Meslek / Esnaf (Bakkal Nalbur ... vb.)
Çiftçi / Balıkçı	Küçük Sanayici	Serbest Meslek / Zanaatkâr (Terzi ... vb.)
Dealer / Broker	Manken / Fotomodel	Sosyal ve İdari Bilimci
Diğer	Memur	Sporcu / Antrenör
Diplomat	Mevsimlik İşçi	Subay (Teğmen Yüzbaşı Binbaşı Albay ... vb.)
Diş Hekimi	Mimar / İçMimar / Dekoratör	Şöför
Eczacı	Muhasebeci / Mali Müşavir	Teknisyen
Emekli	Mühendis	Temel Bilimci (Fizikçi Kimyager ... vb.)
Ev Hanımı	Noter	Tercüman / Çevirmen
Finans	Öğrenci	Tıp Doktoru

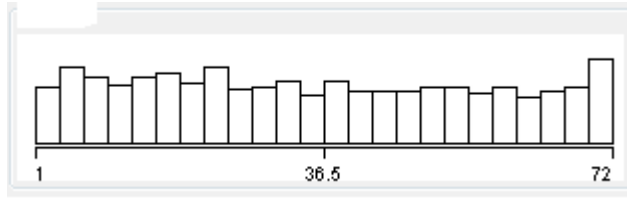
Takip eden grafiklerde, veri özniteliklerinin dağılımları gösterilmiştir. Her bir veri özniteliğinin dağılımı öznitelikleri ve verileri analiz etmede önemli rol oynarlar.

–Müşterilerin bölge değerleri aşağıda gösterilmektedir.



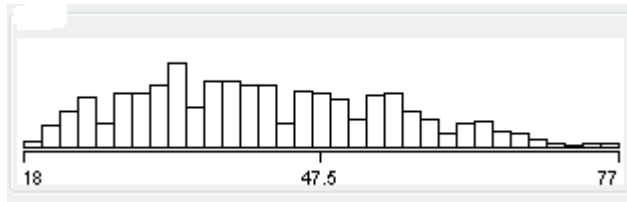
Grafik 5. 1 Bölge Gösterimi

–Müşterilerin kullanım süresi değerleri aşağıda gösterilmektedir.



Grafik 5. 2 Kullanım Süresi Değerleri

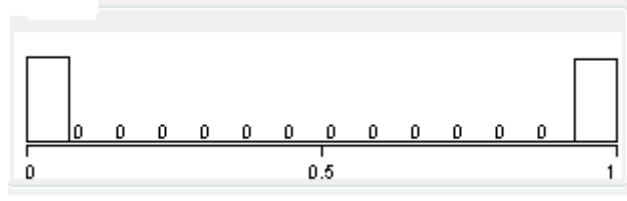
–Müşterilerin yaş değerleri aşağıda gösterilmektedir.



Grafik 5. 3 Yaş Değerleri

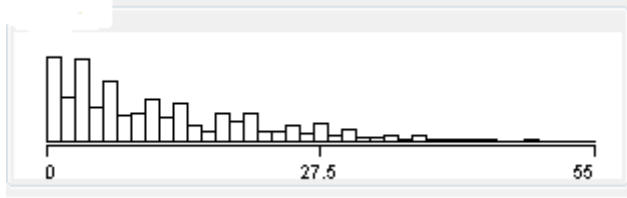
–Müşterilerin medeni durum değerleri aşağıda gösterilmektedir.

- “0” bekarı temsil eder,
- “1” evliyi temsil eder,



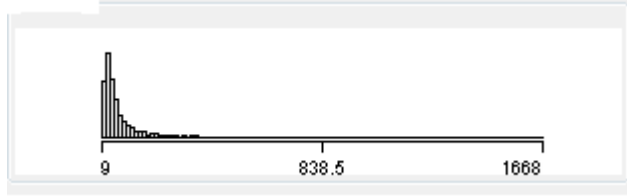
Grafik 5. 4 Medeni Durum Değerleri

–Müşterilerin adres değerleri aşağıda gösterilmektedir.



Grafik 5. 5 Adres Değerleri

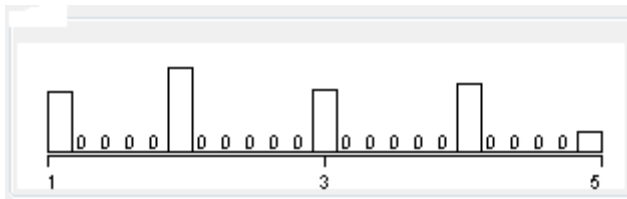
–Müşterilerin gelir değerleri aşağıda gösterilmektedir.



Grafik 5. 6 Gelir Değerleri

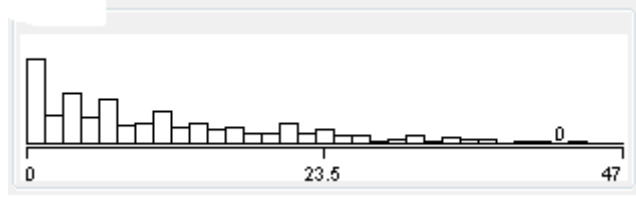
Aşağıda müşterilerin eğitim seviyeleri karşılık değerleri belirtilmektedir.

- “1” ilköğretimi temsil eder,
- “2” liseyi temsil eder,
- “3” üniversiteyi temsil eder,
- “4” yüksek lisansı temsil eder,
- “5” doktorayı temsil eder,



Grafik 5. 7 Eğitim Durumu Değerleri

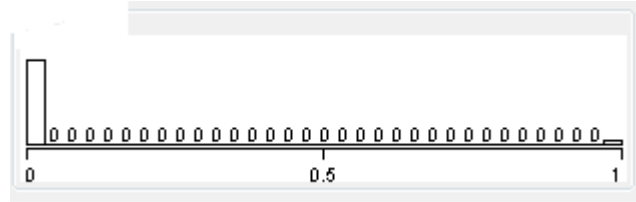
–Müşterilerin iş değerleri aşağıda gösterilmektedir.



Grafik 5. 8 İş Değerleri

–Müşterilerin emeklilik değerleri aşağıda gösterilmektedir.

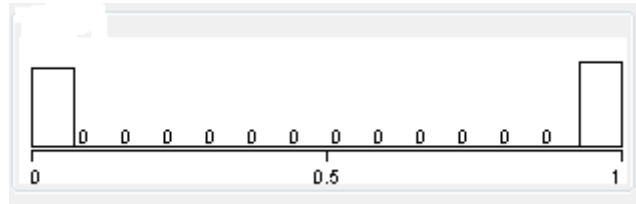
- “0” işten atılmama durumunu temsil eder,
- “1” işten atılmayı temsil eder,



Grafik 5. 9 Emeklilik Değerleri

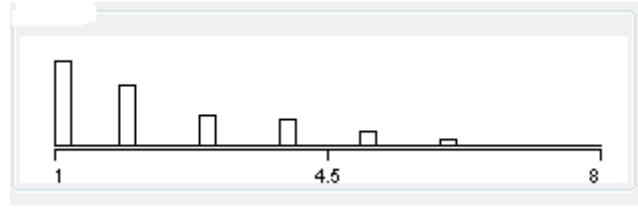
–Müşterilerin cinsiyet değerleri aşağıda gösterilmektedir.

- “0” kadını temsil eder,
- “1” erkeği temsil eder,



Grafik 5. 10 Cinsiyet Durumu Değerleri

–Müşterilerin ikamet değerleri aşağıda gösterilmektedir.



Grafik 5. 11 İkamet Değerleri

- Ranked³⁷¹ öznitelikler ve değerleri Tablo 5.17 'de gösterilmektedir.

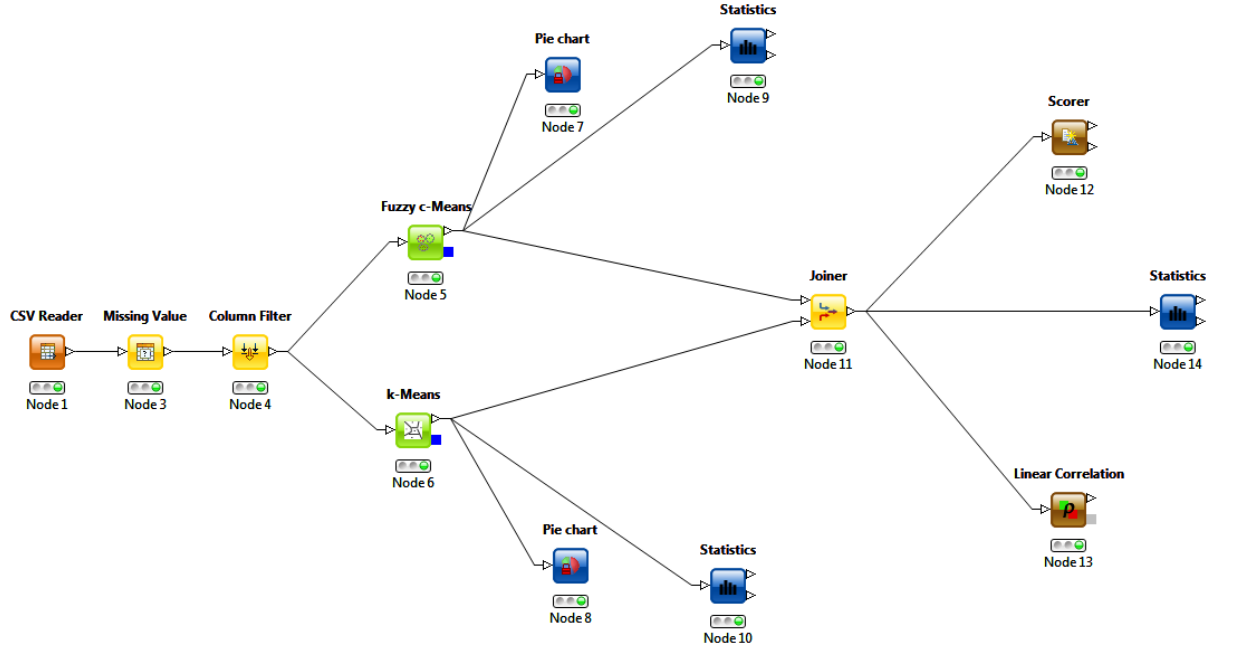
Tablo 5. 17 Öznitelikler ve Rank(Sıralama) Değerleri

Öznitelik İsmi	Kazanan Rank Değerleri
Longten	0.7898796
Equipten	0.3987685
Tollten	0.3678763
Equipmon	0.3345324
Logequi	0.3167987

5.6.2 Kümeleme Aşaması

Bu çalışmada, kümeleme aşaması için, müşteri öznitelikleri ve iki farklı kümeleme algoritması uygun müşteri gruplarının kararında denenmiştir. Bu kümeleme algoritmaları k ve bulanık c ortalamalarıdır. İki farklı algortmada aynı verileri kümeledikten sonra, bu algoritmalar bu aşama için en iyi sonucu alabilmek için kıyaslanmıştır.

³⁷¹ Rank: Bir veri kümesindeki herhangi bir değerin kaçınıcı sırada olduğunu gösterir.



Şekil 5. 1 Bulanık C ve K Ortalamaları Karşılaştırılması İşlemi

Takip eden grafik bize k ve bulanık c ortalamaları sonuçlarını göstermektedir. “Winner Cluster” yani kazanan küme bulanık c ortalamaları sonuçlarını göstermektedir. “Winner Cluster Count” yani kazanan küme sayısı, her bir kümedeki toplam nesne sayısını göstermektedir. “Cluster” yani küme k ortalama algoritmasının sonuçlarını göstermektedir. “Cluster Count” yani küme sayıcı, k ortalama sonuçlarından sonra, her bir kümedeki toplam nesne sayısını göstermektedir.

Tablo 5. 18 Bulanık C Ortalama ve K Ortalama Sınıflandırma Algoritmalarının Dağılımları

Row ID	Winner Cluster	Winner Cluster_Count	Cluster	Cluster_Count
Row0	cluster_1	10397	cluster_1	10427
Row1	cluster_2	2089	cluster_2	2057
Row2	cluster_3	1661	cluster_3	1676
Row3	cluster_0	590	cluster_0	577

Satır No	Bulanık C Kazanan Küme	Kazanan Küme Sayısı	K Ortalama Kazanan Küme	Kazanan Küme Sayısı
Satır 0	Küme1	10397	Küme1	10427
Satır 1	Küme2	2089	Küme2	2057
Satır 2	Küme3	1661	Küme3	1676
Satır 3	Küme0	590	Küme0	577

Takip eden grafik aynı veri kümesi için iki farklı algoritmanın doğruluk değerlerini göstermektedir.

Tablo 5. 19 İki Farklı Doğruluk İstatistiği: K Ortalama ve Bulanık C Ortalama

Winner Cluster \ Cluster	cluster_2	cluster_1	cluster_3	cluster_0
cluster_2	2044	45	0	0
cluster_1	0	10382	15	0
cluster_3	0	0	1661	0
cluster_0	13	0	0	577
Correct classified: 14,664		Wrong classified: 73		
Accuracy: 99.505 %		Error: 0.495 %		

Kazanan Küme	Küme2	Küme1	Küme3	Küme0
Küme2	2044	45	0	0
Küme1	0	10382	15	0
Küme3	0	0	1661	0
Küme0	13	0	0	577
Doğru Sınıflandırılmış:14,664 Yanlış Sınıflandırılmış:73				
Doğruluk:99.505% Hata:0.495				

Bir sonraki grafik yukarıdaki iki farklı algoritmanın sonuçlarının True Positives, False Positives, True Negatives and False Negatives değerleridir.

True Positive yani doğru pozitif, kümeleme ile doğru tahmin edilen pozitif örnek sayısı anlamına gelmektedir.

False negative yani yanlış negatif, kümeleme ile yanlış tahmin edilen pozitif örnek sayısı anlamına gelmektedir.

False positive yani yanlış pozitif, kümeleme ile yanlış tahmin edilen negatif örnek sayısı anlamına gelmektedir.

True negative yani doğru negatif, kümeleme ile doğru tahmin edilen negatif örnek sayısı anlamına gelmektedir.

Tablo 5. 20 K Ortalama ve Bulanık C Ortalama Algoritmalarının Doğru Pozitif Yanlış Pozitif Doğru Negatif Yanlış Negatif Değerleri

Row ID	TruePositives	FalsePositives	TrueNegatives	FalseNegatives
cluster_2	2044	13	12635	45
cluster_1	10382	45	4295	15
cluster_3	1661	15	13061	0
cluster_0	577	0	14147	13

Satır No	Doğru Pozitif	Yanlış Pozitif	Doğru Negatif	Yanlış Negatif
Küme2	2044	13	12635	45
Küme1	10382	45	4295	15
Küme3	1661	15	13061	0
Küme0	577	0	14147	13

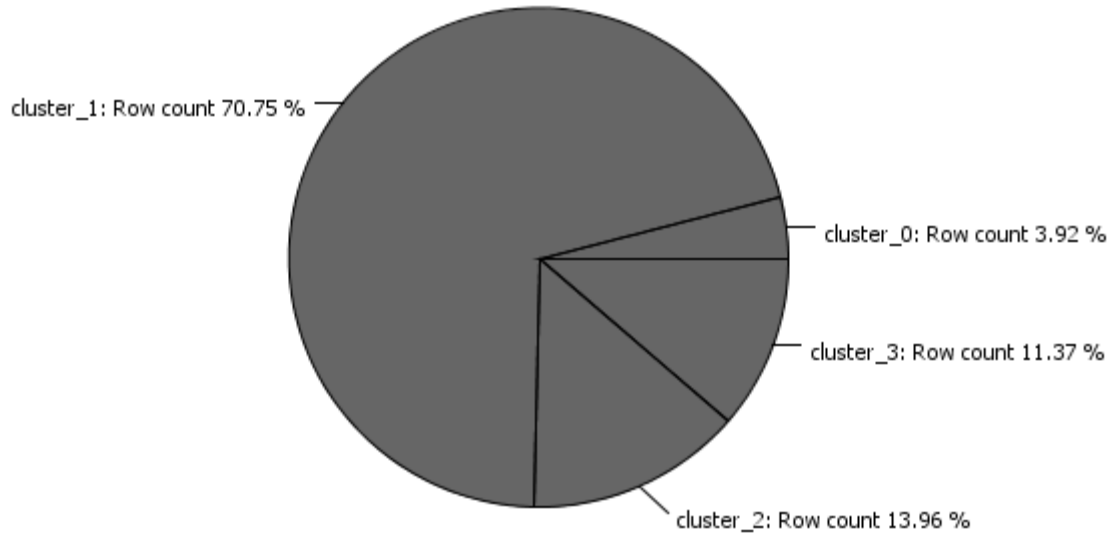
Bir sonraki şekil K ve bulanık C ortalamaları sonuçları arasındaki lineer korelasyonu göstermektedir.

Tablo 5. 21 K ve Bulanık C Ortalamaları Sonuçları Arasındaki Lineer Korelasyon

Row ID	D Winner Cluster	D Cluster
Winner Cluster	1	0.989
Cluster	0.989	1

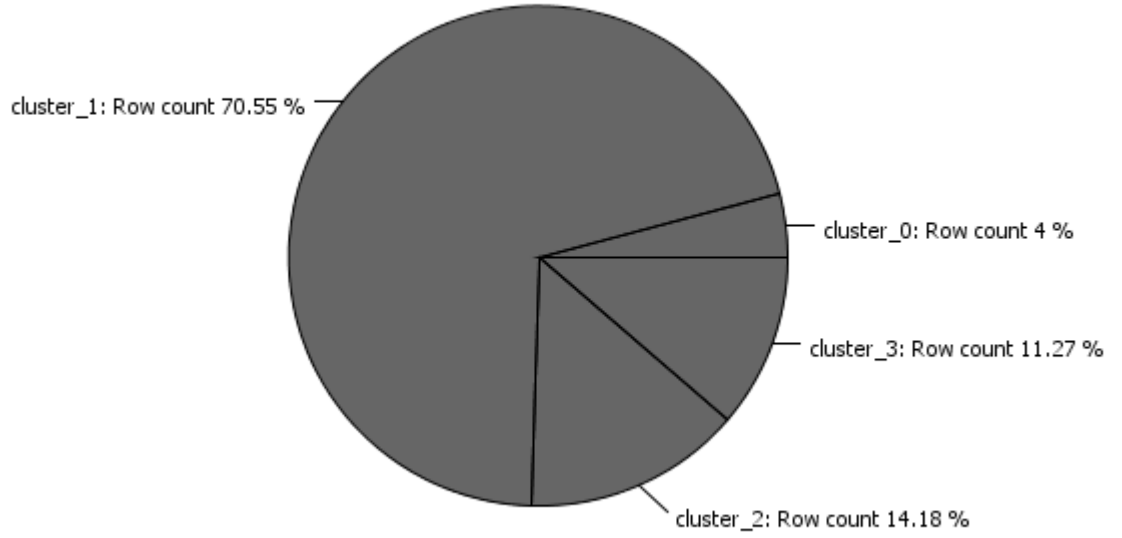
Satır No	Kazanan Küme	Küme
Kazanan Küme	1	0.989
Küme	0.989	1

Şekil 36'da K ortalamalar yöntemi uygulandıktan sonra kümelerin dağılımını göstermektedir.



Şekil 5. 2 K Ortalama Sınıflarının Dağılımı

Bir sonraki şekil bulanık c ortalamalar yöntemi uygulandıktan sonra kümelerin dağılımını göstermektedir.



Şekil 5. 3 Bulanık C Ortalama Sınıflarının Dağılımı

Her bir küme için Tüm toplam istatistik(minimum,maksimum,ortalama ve standart sapma) şekil 38 'de gösterilmektedir. Bu istatistiksel değerler bulanık c ortalamalar algoritması uygulandıktan sonra veri kümesine alınmıştır.

Tablo 5. 22 Bulanık C Ortalaması İstatistik Bilgileri

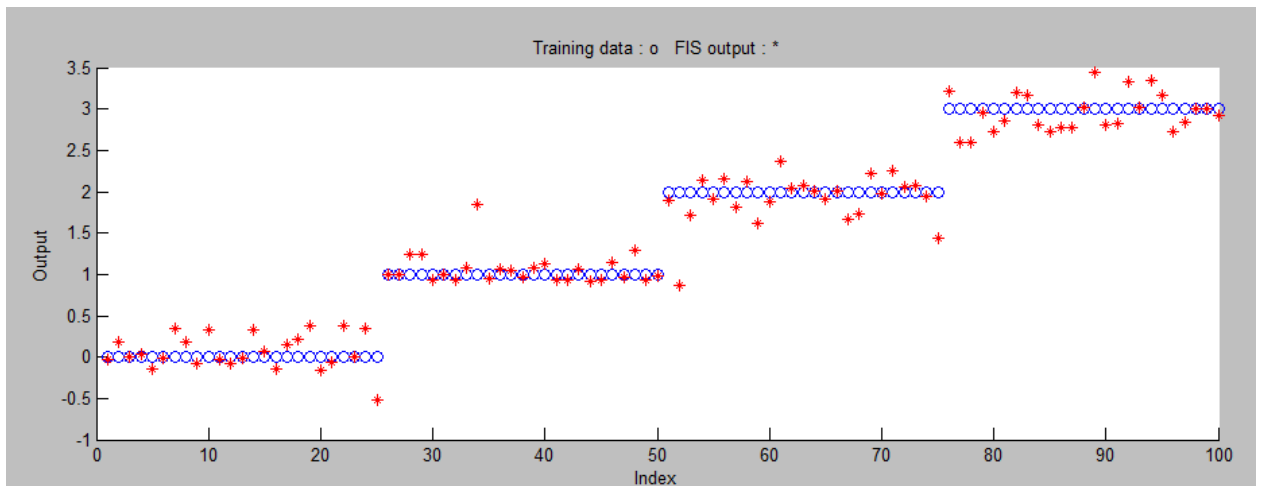
Row ID	D cluster_0	D cluster_1	D cluster_2	D cluster_3
Minimum	0	0	0	0
Maximum	1	1	1	1
Mean	0.041	0.704	0.142	0.114
Std. deviation	0.194	0.448	0.341	0.311

Satır No	Küme0	Küme1	Küme2	Küme3
Minimum	0	0	0	0
Maksimum	1	1	1	1
Ortalama	0.041	0.704	0.142	0.114
Standart Sapma	0.194	0.448	0.341	0.311

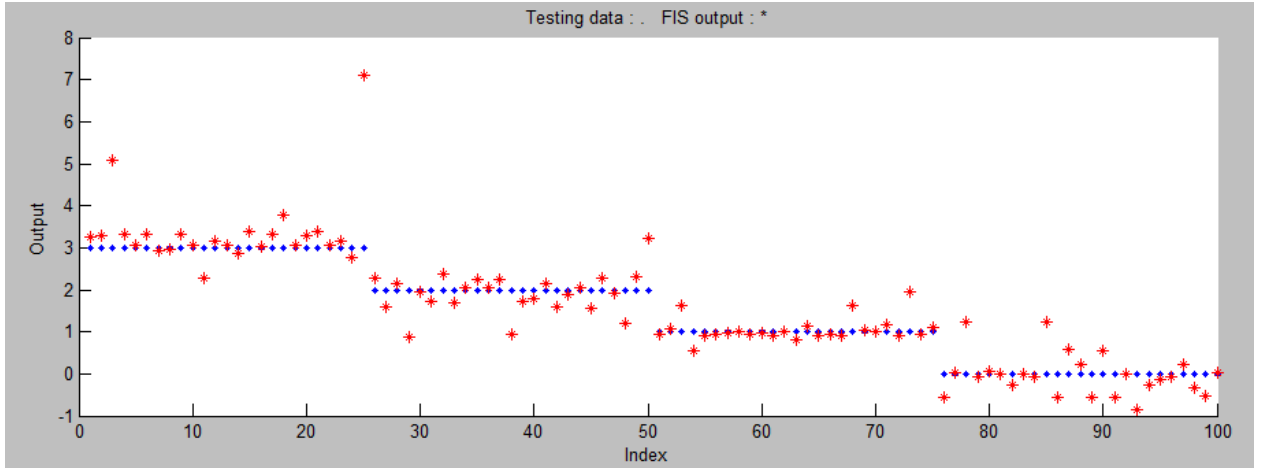
5.6.3 Tahmin Aşaması

Tahmin aşaması için, Matlab'te ANFIS kullandık. Bu çalışmada sonuca ulaşmak için alt kümeleme kullanıldı. Modeli eğitmek için parametre değerleri bunlardır; Influence aralığı 0.5, squash factor 1.25 ,accept oranı 0.5 ve rejection oranı 0.15 tir.

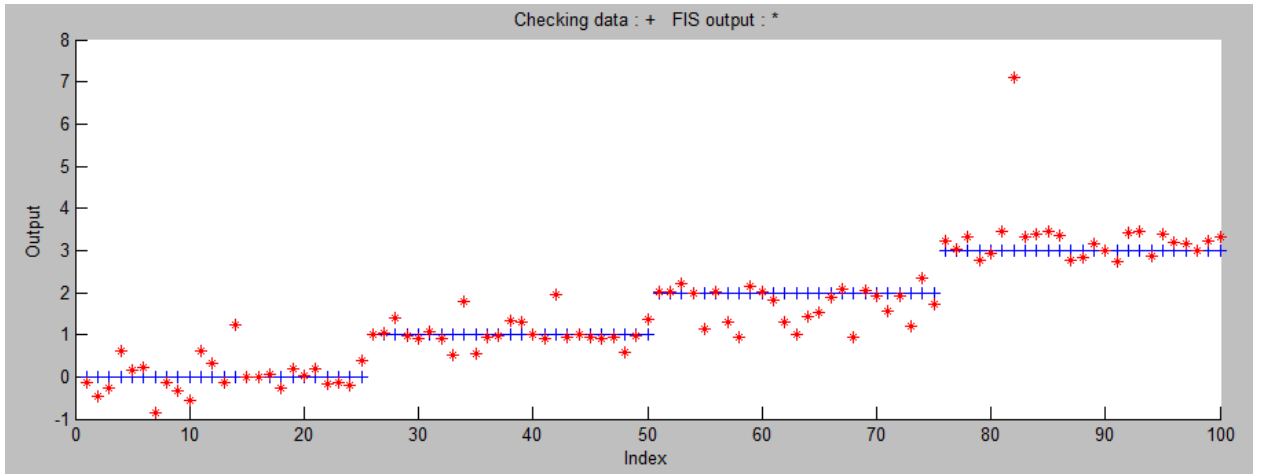
ANFIS'e belirtilen parametreleri verdiğinizde sonuç aşağıdaki gibidir.



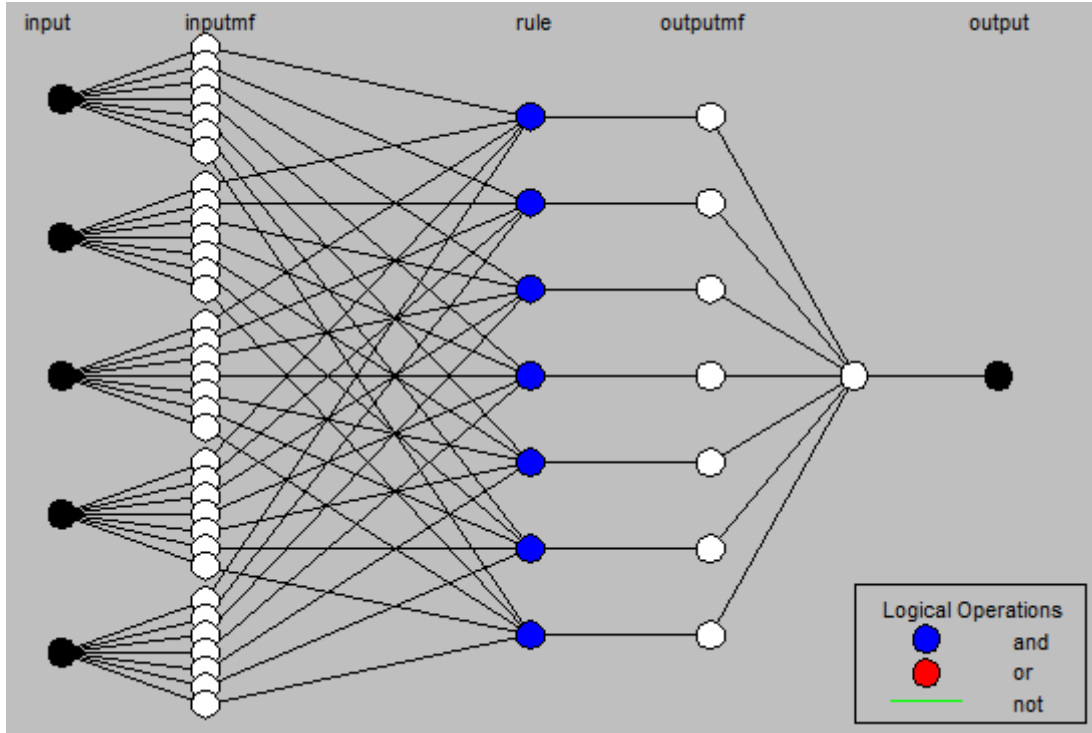
Şekil 5. 4 Eğitilmiş Verilerin ANFIS Sınıflandırması I



Şekil 5. 5 Eğitilmiş Verilerin ANFIS Sınıflandırması II

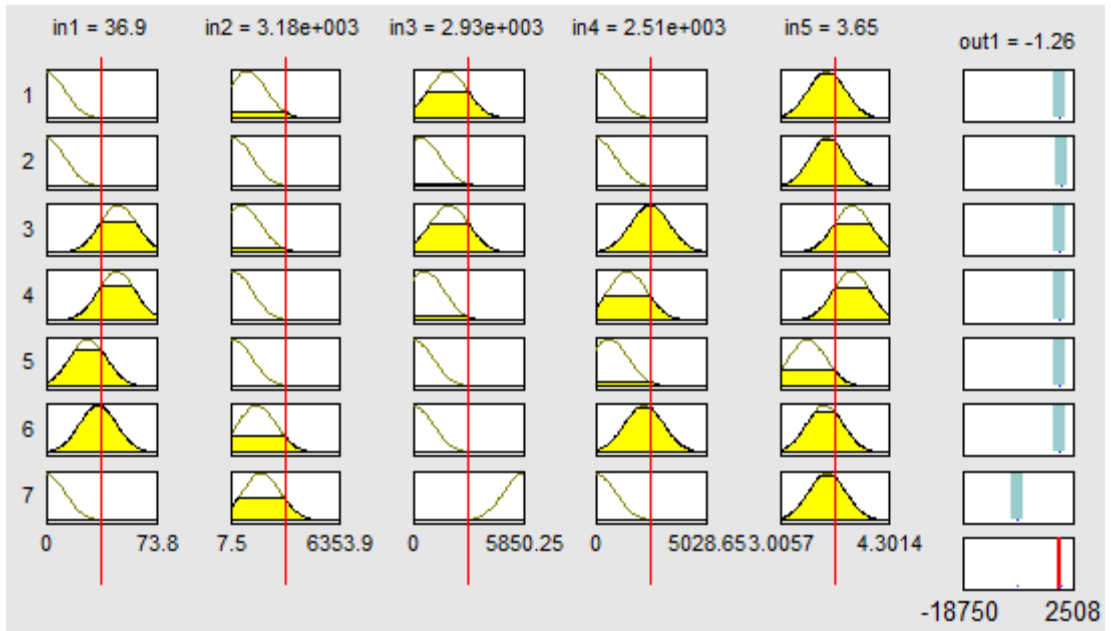


Şekil 5. 6 Kontrol Verilerinin ANFIS Sınıflandırması



Şekil 5. 7 Bulanık Çıkarım Sisteminin ANFIS Modeli

Şekil 42 bulanık çıkarım için giriş faktörlerini ve çıkış sonuçlarını çizmiştir. Bulanık çıkarım diyagramı tüm factor diyagramlarının bileşimidir. Bu diyagram bulanık çıkarım işleminin tüm parçalarını içermektedir.

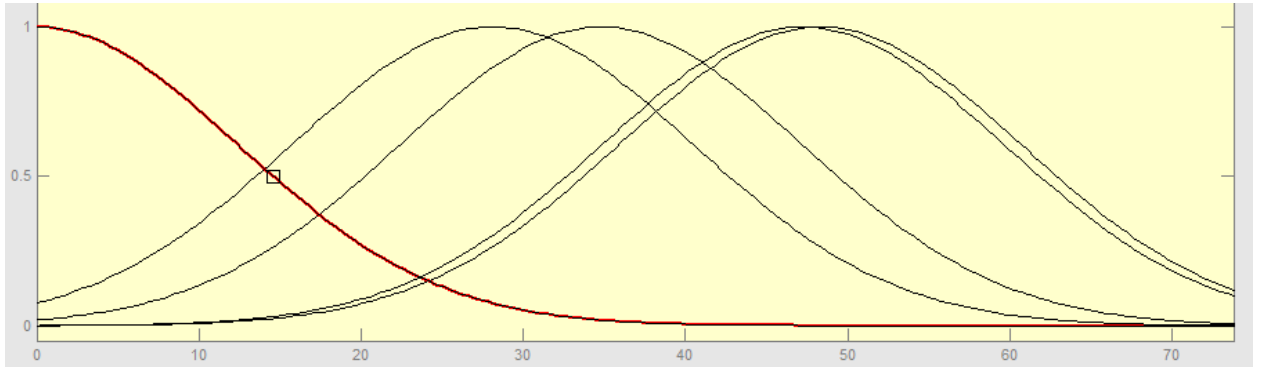


Şekil 5. 8 Bulanık Çıkarım Sisteminin Şeması

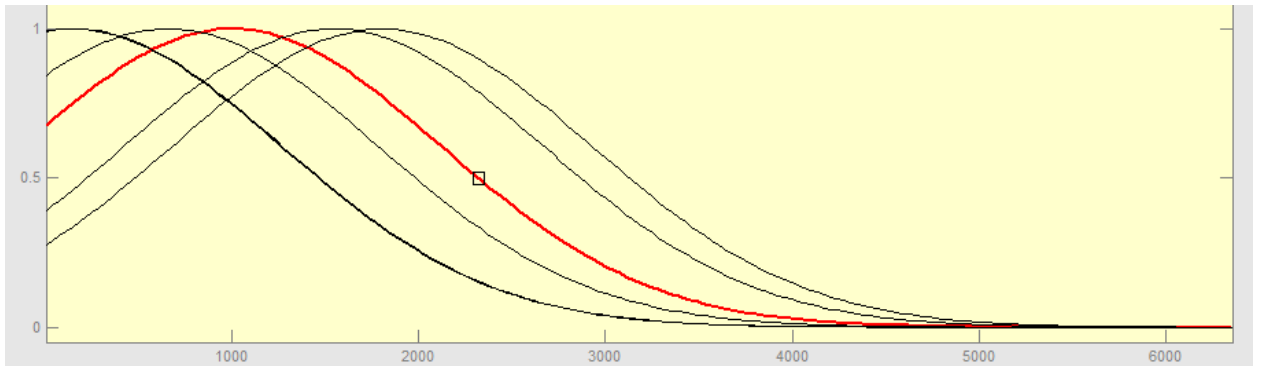
(ANFIS) Adaptif Ağ Yapısına Dayalı Bulanık Çıkarım Sistemini yöntemini kullanarak, ANFIS her bir giriş için üyelik fonksiyonları oluşturmuştur. Takip eden grafiklerde her bir giriş için üyelik fonksiyonu gösterilmektedir. Bu girişler Long distance over tenure (longten), Equipment over tenure (equipten), Toll free over tenure (tollten), Equipment last month (equipmon) and Log equipment (logequi) değişkenleridir. Her bir girişin üyelik fonksiyonu bu grafikleri kullanarak eğitimden önce ve sonra gözden geçirilmiştir.

Eğer bir girişin ortalama üzerinde etkisi olursa, orjinel eğriden düşündürücü bir sapma yapacaktır.

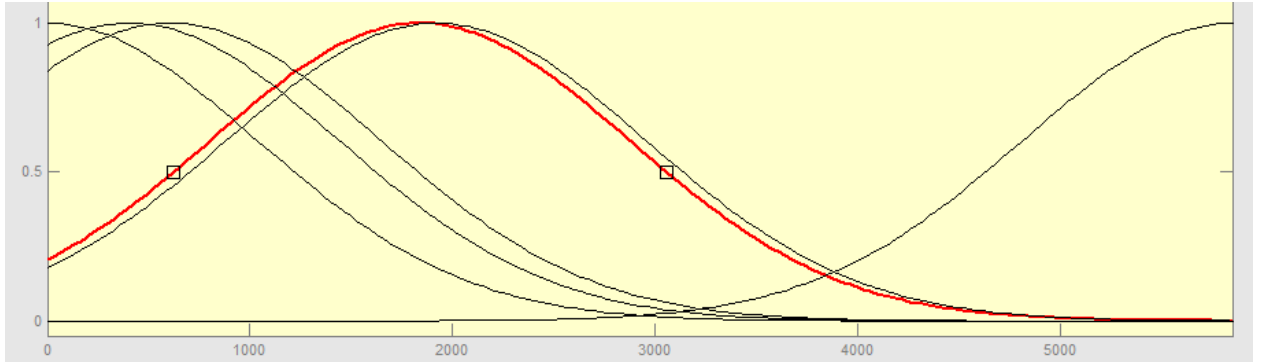
5.9 – 5.13 grafiklerinde, dikey eksen üyelik fonksiyonunun değeridir, yatay eksen ise giriş faktörünü gösterir.



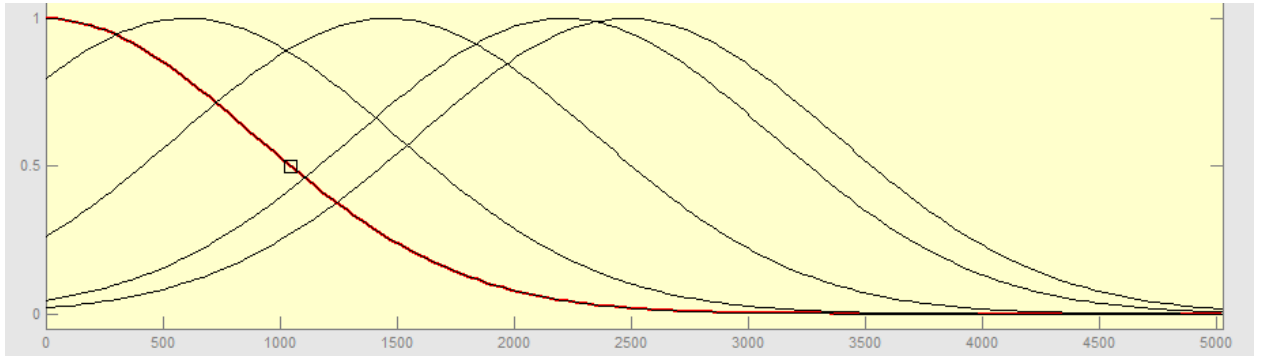
Şekil 5. 9 Equipment Üyelik Fonksiyonu (equipmon)



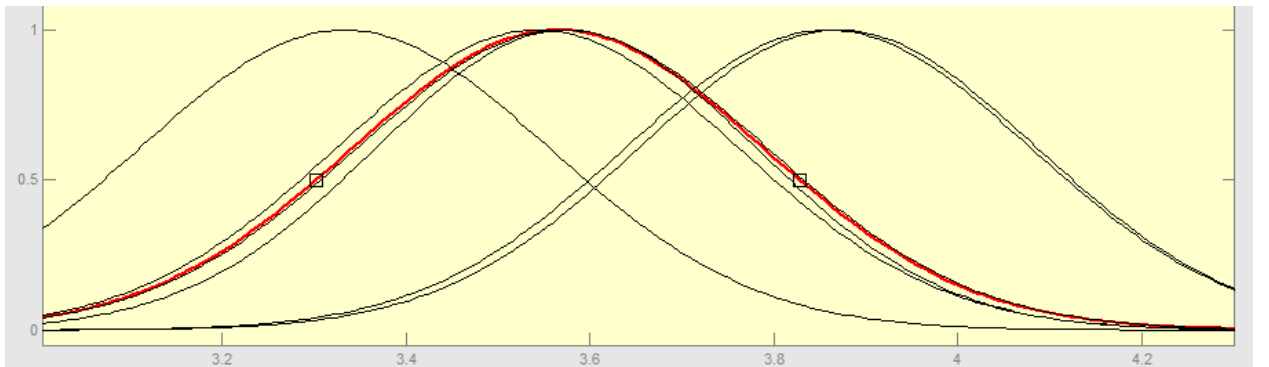
Şekil 5. 10 Long Distance Over Tenure Üyelik Fonksiyonu(longten)



Şekil 5. 11 Toll Free Over Tenure Üyelik Fonksiyonu(toll ten)



Şekil 5. 12 Equipment Over Tenure Üyelik Fonksiyonu (equipten)



Şekil 5. 13 Log Equipment Üyelik Fonksiyonu(losequi)

Bu tabloda değişik yöntemlerin sonuçlarını görebilirsiniz. Bu sonuçlar eğitilmiş veriler sonucu alınmıştır.

Bu formüller duyarlılık(sensitivity), özgüllük(specificity), tahmin(precision) ve doğruluk(accuracy) hesaplanırken kullanılan yöntemlerdir.

TP: doğru pozitif sayısı

FN: yanlış negatif sayısı

FP: yanlış pozitif sayısı

TN: doğru negatif sayısı

$$sensitivity = \frac{TP}{TP + FN}$$

$$specificity = \frac{TN}{FP + TN}$$

$$precision = \frac{TP}{TP + FP}$$

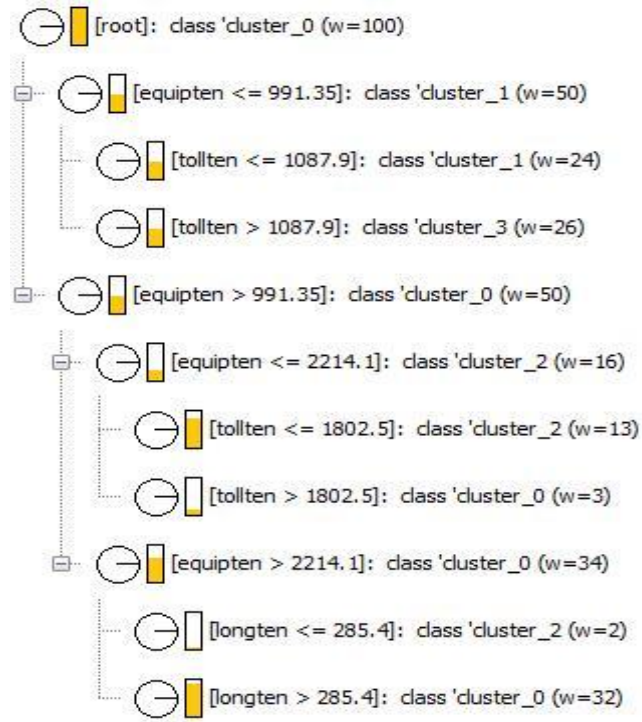
$$accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

Tablo 5. 23 Kullanılan Yöntem ve Eğitim Sonuçları

Method	Duyarlılık	Özgüllük	Tahmin	Doğruluk
Karar ağaçları	0.81	0.94	0.82	0.88
Sade Bayes Sınıflandırıcısı	0.61	0.86	0.63	0.82
SVM	0.52	0.83	0.51	0.75
Olasılıksal Yapay Sinir Ağları	0.84	0.91	0.84	0.92
ANFIS	0.87	0.98	0.87	0.94

Dört farklı yöntemin test edilmesi sonucu, ANFIS'in veri kümeleri için en iyi yöntem olduğunu söyleyebiliriz.

Karar ağacı tahmini sonucu aşağıdaki şekilde gösterilmektedir.



Şekil 5. 14 Karar Ağacı Tahmin Sonucu

Olasılıksal yapay sinir ağıları uygulandıktan sonra kurallar:

Tablo 5. 24 Olasılıksal Yapay Sinir Ağıları Sonrası Oluşan Kurallar

Row ID	D equipmon	D longten	D tollten	D equipten	D logequi	S Winner Cluster
Rule_1	46.55	6,353.9	0	3,411.2	3.841	cluster_0
Rule_6	51.2	447.15	1,653.9	2,986.65	3.936	cluster_0
Rule_7	46.7	825.35	2,624.25	2,590.95	3.844	cluster_0
Rule_8	34.3	2,368.9	2,522.75	2,170.85	3.535	cluster_0
Rule_9	51.35	3,983.6	4,905.85	3,761.8	3.939	cluster_0
Rule_21	42	1,292	1,492.1	2,394.75	3.738	cluster_0
Rule_2	28.25	7.5	0	48	3.341	cluster_1
Rule_10	0	495.85	1,087.9	0	3.568	cluster_1
Rule_11	29.55	229.5	0	991.35	3.386	cluster_1
Rule_12	0	893.3	0	0	3.568	cluster_1
Rule_22	36.25	45.6	609.2	654.75	3.59	cluster_1
Rule_3	54.3	492.8	910.25	2,059.7	3.995	cluster_2
Rule_14	33.35	646.7	0	1,108.05	3.507	cluster_2
Rule_16	46.9	141.95	0	1,262.7	3.848	cluster_2
Rule_17	31.3	946.9	1,767.6	1,788.95	3.444	cluster_2
Rule_18	34.8	1,546.55	0	2,214.1	3.55	cluster_2
Rule_23	40.9	350.05	593.1	1,347.7	3.711	cluster_2
Rule_24	49.6	253.35	0	2,499.8	3.904	cluster_2
Rule_25	55.8	279.7	1,665	2,197.6	4.022	cluster_2
Rule_4	0	350.85	2,423.3	0	3.568	cluster_3
Rule_5	0	1,519.2	4,064.3	0	3.568	cluster_3
Rule_19	0	918	1,447.4	0	3.568	cluster_3
Rule_20	0	2,201	1,981.65	0	3.568	cluster_3

Satır No	Geçen Ay Cihaz	Geçen Ay Uzun Mesafe	Geçen Ay Ücretsiz Görüşme	Üyelik Üzerinde Cihaz	Oturum Ekipmanı	Kazanan Küme
Kural1	46.55	6,353.9	0	3,411.2	3.841	Küme0
Kural6	51.2	447.15	1,653.9	2,986.65	3.936	Küme0
Kural7	46.7	825.35	2,624.25	2,590.95	3.844	Küme0
Kural8	34.3	2,368.9	2.522.75	2,170.85	3.535	Küme0
Kural9	51.35	3,983.6	4,905.85	3,761.8	3.939	Küme0
Kural21	42	1,292	1,492.1	2,394.75	3.738	Küme0
Kural2	28.25	7.5	0	48	3.341	Küme1
Kural10	0	495.85	1,087.9	0	3.568	Küme1

Kural11	29.55	229.5	0	991.35	3.386	Küme1
Kural12	0	893.3	0	0	3.568	Küme1
Kural22	36.25	45.6	609.2	654.75	3.59	Küme1
Kural3	54.3	492.8	910.25	2,059.7	3.995	Küme2
Kural14	33.35	646.7	0	1,108.05	3.507	Küme2
Kural16	46.9	141.95	0	1,262.7	3.848	Küme2
Kural17	31.3	946.9	1,767.6	1,788.95	3.444	Küme2
Kural18	34.8	1,546.55	0	2,214.1	3.55	Küme2
Kural23	40.9	350.05	593.1	1,347.7	3.711	Küme2
Kural24	49.6	253.35	0	2,499.8	3.904	Küme2
Kural25	55.8	279.7	1,665	2,197.6	4.022	Küme2
Kural4	0	350.85	2,423.3	0	3.658	Küme3
Kural5	0	1,519.2	4,064.3	0	3.658	Küme3
Kural19	0	918	1,447.4	0	3.658	Küme3
Kural20	0	2,201	1,981.65	0	3.658	Küme3

Olasılıksal yapay sinir ağıları algoritması 25 farklı kural çıkarmıştır ve her bir veri kümesinde 100 farklı müşteri vardır.(eğitim ve test). ANFIS bulanık c ortalamaları algoritması yedi kural çıkarmıştır ve bu kurallar aşağıdaki gibi açıklanır.

Kural 1: [33 660 0 5028.65 3.4965] [Küme 0]

Eğer geçen ayki equipment 33 ,long distance over tenure 660 ,toll free over tenure 0, equipment over tenure 5028.65 ve log equipment 3.4965 ise çıktı küme 0'dır.

Kural 2: [51.2 447.15 1238.2 2170.85 3.5351] [Küme 0]

Eğer geçen ayki equipment 51.2 ,long distance over tenure 447.15 ,toll free over tenure 1238.2, equipment over tenure 2170.85 ve log equipment 3,5351 ise çıktı küme 0'dır.

Kural 3: [0 7.5 0 0 3.005] [Küme 1]

Eğer geçen ayki equipment 0 ,long distance over tenure 7.5 ,toll free over tenure 0, equipment over tenure 0 ve log equipment 33.001 ise çıktı küme 1'dir.

Kural 4: [48.4 279.7 1802.22 1481.35 4.02] [Küme 2]

Eğer geçen ayki equipment 48.4 ,long distance over tenure 279.7 ,toll free over tenure 1802.22 , equipment over tenure 1481.35 ve log equipment 4.02 ise çıktı küme 2 ‘dir.

Kural 5: [49.2 447.15 1238.2 2498.35 3.895] [Küme 2]

Eğer geçen ayki equipment 49.2 ,long distance over tenure 447.15,toll free over tenure 1238.2, equipment over tenure 2498.35 ve log equipment 3.895 ise çıktı küme 2 ‘dir.

Kural 6: [31.3 1037.65 4.176 1395.1 3.171] [Küme 2]

Eğer geçen ayki equipment 31.3,long distance over tenure 1037.65 ,toll free over tenure 4.176, equipment over tenure 1395.1 ve log equipment 3.171 ise çıktı küme 2 ‘dir.

Kural 7: [0 2291.1 1447.4 0 3.568] [Küme 3]

Eğer geçen ayki equipment 0,long distance over tenure 2291.1 ,toll free over tenure 1447.4, equipment over tenure 0 ve log equipment 3.568 ise çıktı küme 3 ‘tür.

5.7 Araştırmanın Sonucu

Bu çalışmada müşteri davranışları hakkında analiz ve bazı sonuçlara ulaşmaya çalışıldı. Telekomünikasyon endüstrisinden alınan örnek bir veri elimizde mevcuttu. Bu veri kümesinde 42 farklı öznitelik ve 15000 kayıt vardı. Bazı boş ve eksik olan değerleri doldurmak için bazı yöntemler kullanıldı. Çalışmanın ilk aşaması analizdi. Her bir öz niteliğin değerleri gözden geçirildi. Daha sonra veri kümesinden anlamlı bilgiler elde etmek için veri madenciliği yöntemleri uygulandı. Bir sonraki aşamada veri kümesinde hangi özniteliklerin daha etkili ve verimli olduğu ortaya çıkartıldı. Daha sonra veri kümesi içerisinde model oluşturmak için 5 öznitelik seçildi. Bulanık c ve k ortalamaları sonuçlarını karşılaştırıldı. Sınıflandırma işlemi için giriş olarak bulanık c ve k ortalamaları sonuçları kullanıldı ve bulanık c ortalamaları algoritmasının bu çalışmada en iyisi olduğu anlaşıldı.

Bu alıřmada en iyi sonucu ANFIS bulanık c ortalamaları yöntemi verdi. Tahmin aşamasında ANFIS bulanık c ortalamaları yöntemi kullanıldı ve ANFIS sınıflandırma şekilleri, kullanılan her bir özniteliğın üyelik fksiyonları ve bulanık ıkarım diyagramı alıřmaya eklendi. Aynı veri giriřleriyle üç farklı yöntem daha kullanıldı ve denendi. Bu alıřmayla dört farklı sınıflandırma yöntemi karşılaştırılmıř oldu. Bu sınıflandırma yöntemleri karar ağaları, Sade bayes sınıflandırıcısı, Destek vector makinası ve ANFIS oldu. Her yöntem için duyarlılık, özğüllük, tahmin ve doğruluk değeri alıřmada belirtildi. ANFIS 7 kural genelledi ve duyarlılık, özüllük, tahmin ve doğruluk değeri diğeri yöntemler arasında en iyisi olduğı anlaşıldı. Olasılıksal yapay sinir ağları ANFIS'ten sonra en iyi sonucu 25 kural ortaya ıkararak vermiř oldu.

6. SONUÇ

Veri Madenciliği Çözümleri, kurumlar için artık olmazsa olmaz bir kaynak. Çünkü bununla kurumlar, hedef kitlelerinin özelliklerini, müşterilerinin neyi neden satın aldıklarını öğrenerek, yeni stratejiler oluşturuyor. Veri Madenciliği'ni kullanan şirketlerin öncülüğünü ise telekomünikasyon ve finans şirketlerinin yaptığını söyleyen Dünyaca ünlü guru Gordon Linoff, artık Türkiye gibi zenginleşen ülkelerde de Veri Madenciliği'ne ilginin arttığını söylüyor. Kurumlar için hedef kitlesinin nasıl bir özelliğe sahip olduğu, müşteri portföyünün neyi, neden satın aldığını bilmesi çok önemlidir. Dünya'daki tüm kurumsal şirketler, bu bilgilere ulaşmak ve bu bilgileri kullanarak, atacakları adımda doğru karar almak ister. Kurumsal veri kaynaklarının kısa sürede analiz edilmesi ve şirketle ilgili stratejik kararlar alınmasını olanaklı kılan 'Veri Madenciliği Çözümleri' de işte bu noktada şirketlere kılavuzluk ediyor. Veri Madenciliği Yazılımı da; kısaca büyük şirketlerin stratejik adımlar atarken çok büyük veri yığınları arasından, kendilerine yol gösterecek kritik verileri ayıklayarak, analiz etmelerine dayanıyor.

Her geçen gün bilgisayar sistemlerinin ucuzlaması ve güçlerinin artması, daha fazla verinin bilgisayar ortamında saklanabilmesini sağladı. Böylece, fazla miktardaki bu verilerin işlenebilmesi fırsatı ortaya çıktı ve büyük önem kazandı. Bu veriler içinde bulunan örüntü ve eğilimleri keşfetme işlemi olan veri madenciliği, gelişen teknolojinin kullanılmasıyla birlikte daha büyük keşiflerin ortaya çıkmasını sağlıyor. Gelişmiş algoritmaların, çok işlemcili bilgisayarların ve büyük veri tabanlarının kullanılması, daha kısa sürede daha fazla verinin analiz edilmesini mümkün kılıyor. Günümüzün sıkı rekabet ortamında ilk elde elindeki müşteriye kaybetmeme güdüsü ile hareket ederek büyük bir pazar haline gelen Müşteri İlişkileri Yönetimi yaklaşımını geliştiren firmalar da, bu yaklaşım dahilinde ellerindeki müşteriye ait çok sayıda veriyi değerlendirmek için veri madenciliği tekniklerini kullanmaya başladı. Verimadenciliğinin CRM kapsamındaki kullanımı, müşteri verilerinin analiz edilerek doğru müşteriler hakkında doğru tahminlerde bulunulmasını sağladı. Diğer alanlarla birlikte özellikle CRM konusunda kullanımı gitgide yaygınlaşan ve başarılı sonuçlar doğuran veri madenciliği

teknikleri, firmalara mevcut müşterileriyle iletişimlerinde, onlara erişiminde büyük faydalar sağlamış, geleceğe yönelik doğru tahminlere müşteri sadakatını iyi ölçerek ellerinde tuttıkları müşterilerle kâr elde etmeyi başaramışlardır.

Bu çalışmada müşteri bağlılığının, sürdürülebilir rekabet avantajı oluşturulmasında önemli bir unsur olduğunu, işletmenin faaliyetlerine olumlu katkıda bulunduğunu ve müşteri bağlılığı yönetimi görevinin yönetsel çabaların merkezinde bulunması gerektiğini vurgulanmıştır. Gelecekte pazarlamanın gideceği yönlerden birinin, müşterilerine hayat boyu mal ve hizmet sunulması stratejilerin uygulanması olacağı açıkça görülmektedir.

Müşteri bağlılığını etkileyen değişkenlerin belirlenmesine yönelik farklı araştırmalar yapılmıştır. Farklı çalışmalarda kalite, müşteri memnuniyeti ve marka imajının müşteri bağlılığı üzerindeki etkisini, müşterilerin elde tutulmasında güvenin etkisini, imajın kalite ve müşteri memnuniyeti ile birlikte müşteri bağlılığı üzerindeki etkisini, bankalara olan bağlılıkta imaj, hizmet kalitesi ve müşteri memnuniyetinin etkisini, müşteri memnuniyeti, kalite, değer, imaj ve müşteri bağlılığı arasında kavramsal bir yapının oluşturulmasıyla ilgili etkisini, değiştirme maliyetinin müşteri bağlılığı üzerindeki etkisini ve güven, işletme imajı, değiştirme maliyeti ve algılanan hizmet kalitesinin müşteri bağlılığının sağlanmasındaki etkileri analiz edilmiştir.. Bu araştırmalar sonucunda, müşteri memnuniyetinin sağlanması, kaliteli ürünlerin sunulması, olumlu işletme ve marka imajı, işletmenin müşteriye sunduğu değer, müşterilerin alışkanlıkları, sürekli müşterilerin ödüllendirilmeleri, güven ve satış sonrası hizmetlerin etki düzeyleri farklı olmakla beraber, müşteri bağlılığının oluşturulmasında önemli rolleri olduğu ortaya çıkmıştır.

Müşteri bağlılığının oluşturulmasında bu değişkenlerin yanında kulüp pazarlama ve ilişki pazarlaması programlarının da önemli bir yeri vardır. İşletmenin kurduğu kulüplere üyelik müşterilere yeterince özel yararlar getiriyorsa, üyelerin bu ayrıcalıkları kaybetmeyi göze almak istemeyecekleri için işletmeye bağlılık duyacaklarını diğer bir düşünce ise kulüp pazarlama programı çerçevesinde kurulan

bu kulüplerin, müşteri bağlılığının kurulması, korunması ve geliştirilmesinde en etkili yollarından biri olduğunu ifade etmişlerdir. Örneğin, ABD'nin en büyük telefon servisi sağlayıcılarından biri olan MCI işletmesi, 1991 yılında başlattığı "Dostlar ve Aile" programı çerçevesinde müşterilerine bir form göndererek, bu formlara indirimli arama yapmak istedikleri 20 aile üyesi veya arkadaşlarının telefon numaralarını yazmasını istemiştir. Bu programa üye olan kişiler, başka kişileri de bu programa girmeye teşvik ederek, taraftar müşteri haline gelmişler ve grubun üyeleri diğer üye/üyelerini bu kulüpten ayrılmaması yönünde baskı yapmışlardır. Bu sayede işletme iki yıl içinde 5 Milyon yeni abone de kazanmıştır. Kulüp pazarlama programına bir başka örnek de dünyaca ünlü motosiklet üreticisi Harley-Davidson tarafından 1983 yılında kurulan "Harley-Davidson Kulübü"dür. Müşterileri üyelik grupları veya oluşturulan kulüplere üye yaparak bütünleştirmek yoluyla müşterilerin tekrar tekrar satın alması güçlendirilmiş ve müşteri bağlılığı sağlanmıştır. GSM sektöründe çok sayıda kulüp pazarlama programı çerçevesinde belirli meslek grupları arasında düşük ücretlerle ve/veya ücretsiz konuşma imkanı verilerek hem yeni müşteriler kazanılmış hem de mevcut müşterilerini koruma yönlü programlar uygulanmıştır.

Colgate ve Danaher 2000 yılında yapılan araştırmada, ilişki pazarlaması stratejisinin etkin bir şekilde uygulanmasının müşteri bağlılığı ve memnuniyeti üzerindeki etkisi araştırılmıştır. Bu araştırmada ilişki pazarlaması stratejinin mükemmel bir şekilde uygulanmasının müşteri bağlılığı ve memnuniyetini arttıracığı, buna karşın uygulamanın zayıf olması durumunda ise işletmenin hedefine ulaşamadığı gibi müşterilerin işletmeye bakış açılarının da olumsuz yönde etkilenebileceği ortaya çıkmıştır. Raval ve Grönroos ise ilişki pazarlamasının temelini işletme ile işletmenin mikro çevresi müşteriler, tedarikçiler, aracılar vb. arasındaki karşılıklı ilişki olduğunu belirterek, ilişki pazarlamasının en önemli amacının müşterilerle sürekli, karlı ve uzun dönemli ilişki kurarak müşteri bağlılığı sağlamak olduğunu açıklamışlardır. Başka bir araştırmada ise banka müşterileri üzerinde yaptığı

araştırma sonucunda, ilişki pazarlaması uygulamalarının müşteri bağlılığı oluşturulmasında önemli bir rolü olduğunu ortaya çıkarmışlardır.

Bir veri madenciliği çalışması için öncelikli olarak çok miktarda kaliteli veriye ihtiyaç vardır. Amaç bu veri içinde saklı, gelecekle ilgili tahmin yapmakta kullanılabilecek kural ve bağıntıların çıkarılmasıdır. Böyle bir çalışmanın başarılı olması için uygulama konusundaki uzmanların veri tabanları ve veri madenciliği konusundaki uzmanlarla beraber çalışması gerekir. Çalışma uzun sürebilir; zaman ve sabır gerekir.

Sonuç olarak tüm telekomünikasyon müşterileri benzerliklerine bakılarak değişik yöntemler ışığında gruplandı. Son olarak veri kümesi 4 farklı yöntem için giriş oldu. Bu çalışmada farklı yöntemler ile müşterileri gruplara ayırarak belirledik, müşteri kayıp yönetimi için hangi yöntemin en iyisi olabileceğini bulmaya gayret ettik. Bu çalışma sonraki yayınlar için bir kaynak niteliğinde olacaktır.

Tüketici pazarlama faaliyetlerinin esas noktası olmalıdır. Tüketicinin güdümleyici, sosyo- kültürel ve demografik faktörler itibariyle, pazarı bölümlere ayırmak ve buna uygun yeni pazarlama programları geliştirmek hem firmanın hem de etkin bir pazarlama faaliyeti için önemlidir. İşletmelerin pazarlama faaliyetlerinde tüketici görüş açısını benimsemeleri ve pazarlama programların tüketiciye ve topluma dönük hale getirmeleri gerekmektedir. Büyük bir hızla değişen ve artan rekabet koşulları içinde faaliyet gösteren işletmeler özellikle ürünlerin fiziksel olarak farklılaştırmayacağı durumlarda rekabet üstünlüğü elde etmek ve pazarda yaşamını sürdürebilmek için ürünlerine bir takım kişilikler kazandırmak yoluna gitmektedir.

Pazarlamacılar sadece ürün kalitesini iyileştirerek satışlarını arttıramayacaklarını anlamışlardır. Günümüzde sadece ürün değil marka oluşturmanın, diğer bir ifadeyle kaliteli bir ürüne markalama yoluyla kişilik kazandırmanın daha önemli olduğu görülmektedir. Marka oluşturduktan sonra da müşterilerinin markaya bağlılığı sağlanmalıdır. Ancak müşteri bağlılığını arttırmak için müşterinin markayı sürekli tercih

etmesinde etkili olan özelliklerin ve bu özelliklerin tercih kararını ne derecede etkilediğinin bilinmesi marka ile ilgili pazarlama kararlarının oluşturulmasında büyük önem taşımaktadır. İşte bu aşamada da müşteri ilişkileri öne çıkmakta ve marka bağlılığının sağlanmasında, müşterinin markayı sürekli hatırlaması ve alışkanlık haline getirmesinde büyük rol oynamaktadır.

EKLER

EK 1: TELEKOMÜNİKASYON VERİSİ

Çalışmada müşteri kayıp yönetimi için kullanılan 15000 müşteriye ait 42 özellik içeren excel listesinin ilk 15 müşterisine ait kayıtları aşağıda örnek teşkil etmesi açısından gösterilmektedir.

	A	B	C	D	E	F	G	H	I	J
1	customerid	region	tenure	age	marital	address	income	ed	employ	retire
2	ID1	2.000.000	41.000.000	46.000.000	1.000.000	14.000.000	47.000.000	2.000.000	11.000.000	0.000000
3	ID2	2.000.000	18.000.000	69.000.000	0.000000	28.000.000	11.000.000	1.000.000	17.000.000	1.000.000
4	ID3	2.000.000	45.000.000	22.000.000	1.000.000	2.000.000	36.000.000	4.000.000	0.000000	0.000000
5	ID4	1.000.000	28.000.000	33.000.000	1.000.000	3.000.000	14.000.000	4.000.000	1.000.000	0.000000
6	ID5	2.000.000	15.000.000	35.000.000	0.000000	10.000.000	25.000.000	2.000.000	3.000.000	0.000000
7	ID6	3.000.000	55.000.000	37.000.000	1.000.000	15.000.000	23.000.000	2.000.000	0.000000	0.000000
8	ID7	3.000.000	72.000.000	53.000.000	0.000000	34.000.000	108.000.000	1.000.000	23.000.000	0.000000
9	ID8	2.000.000	13.000.000	38.000.000	1.000.000	8.000.000	104.000.000	4.000.000	6.000.000	0.000000
10	ID9	2.000.000	2.000.000	28.000.000	0.000000	7.000.000	46.000.000	3.000.000	0.000000	0.000000
11	ID10	3.000.000	48.000.000	35.000.000	0.000000	10.000.000	24.000.000	2.000.000	6.000.000	0.000000
12	ID11	2.000.000	38.000.000	30.000.000	1.000.000	10.000.000	25.000.000	1.000.000	5.000.000	0.000000
13	ID12	3.000.000	59.000.000	43.000.000	1.000.000	4.000.000	101.000.000	2.000.000	22.000.000	0.000000
14	ID13	1.000.000	3.000.000	25.000.000	0.000000	0.000000	65.000.000	4.000.000	0.000000	0.000000
15	ID14	2.000.000	25.000.000	55.000.000	0.000000	15.000.000	37.000.000	1.000.000	7.000.000	0.000000

K	L	M	N	O	P	Q	R	S	T
gender	reside	tollfree	equip	callcard	wireless	longmon	tollmon	equipmon	cardmon
0.000000	5.000.000	1.000.000	1.000.000	1.000.000	1.000.000	6.300.000	41.250.000	55.800.000	20.000.000
0.000000	1.000.000	0.000000	0.000000	0.000000	0.000000	3.850.000	0.000000	0.000000	0.000000
1.000.000	5.000.000	1.000.000	1.000.000	0.000000	0.000000	9.950.000	14.750.000	26.150.000	0.000000
0.000000	3.000.000	1.000.000	1.000.000	0.000000	1.000.000	4.550.000	15.750.000	33.000.000	0.000000
0.000000	1.000.000	0.000000	1.000.000	0.000000	0.000000	4.750.000	0.000000	21.400.000	0.000000
1.000.000	5.000.000	1.000.000	0.000000	1.000.000	0.000000	12.950.000	32.500.000	0.000000	26.250.000
0.000000	1.000.000	1.000.000	0.000000	1.000.000	1.000.000	32.550.000	35.500.000	0.000000	14.750.000
0.000000	7.000.000	1.000.000	0.000000	1.000.000	1.000.000	10.250.000	21.500.000	0.000000	25.750.000
1.000.000	1.000.000	0.000000	0.000000	0.000000	1.000.000	3.450.000	0.000000	0.000000	0.000000
0.000000	1.000.000	1.000.000	0.000000	1.000.000	0.000000	16.950.000	43.000.000	0.000000	28.500.000
1.000.000	4.000.000	0.000000	0.000000	1.000.000	0.000000	3.950.000	0.000000	0.000000	12.000.000
1.000.000	5.000.000	0.000000	0.000000	1.000.000	0.000000	13.650.000	0.000000	0.000000	20.750.000
1.000.000	1.000.000	0.000000	1.000.000	1.000.000	0.000000	3.300.000	0.000000	34.250.000	11.250.000
1.000.000	1.000.000	0.000000	0.000000	0.000000	0.000000	10.950.000	0.000000	0.000000	0.000000

U	V	W	X	Y	Z	AA	AB	AC	AD
wiremon	longten	tollten	equipten	cardten	wireten	multline	voice	pager	internet
50.950.000	279.700.000	1.665.000.000	2.197.600.000	830.000.000	1.956.700.000	1.000.000	1.000.000	1.000.000	0.000000
0.000000	62.750.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.000000	412.100.000	663.100.000	1.180.500.000	0.000000	0.000000	1.000.000	0.000000	0.000000	0.000000
23.950.000	145.950.000	504.900.000	869.850.000	0.000000	618.650.000	1.000.000	0.000000	1.000.000	1.000.000
0.000000	41.400.000	0.000000	291.400.000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000.000
0.000000	716.250.000	1.783.550.000	0.000000	1.355.000.000	0.000000	1.000.000	0.000000	0.000000	0.000000
40.750.000	2.361.350.000	2.535.400.000	0.000000	1.060.000.000	2.886.050.000	1.000.000	0.000000	0.000000	0.000000
58.850.000	160.400.000	274.000.000	0.000000	325.000.000	723.350.000	0.000000	1.000.000	1.000.000	1.000.000
15.100.000	6.550.000	0.000000	0.000000	0.000000	10.450.000	1.000.000	0.000000	1.000.000	0.000000
0.000000	874.450.000	2.007.400.000	0.000000	1.270.000.000	0.000000	0.000000	0.000000	0.000000	0.000000
0.000000	179.750.000	0.000000	0.000000	430.000.000	0.000000	0.000000	0.000000	0.000000	0.000000
0.000000	817.650.000	0.000000	0.000000	1.195.000.000	0.000000	1.000.000	0.000000	0.000000	0.000000
0.000000	5.900.000	0.000000	47.650.000	5.000.000	0.000000	0.000000	1.000.000	0.000000	1.000.000
0.000000	302.800.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
41.000.000	138.500.000	282.300.000	634.100.000	265.000.000	595.300.000	0.000000	1.000.000	0.000000	0.000000

AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ
callid	callwait	forward	confer	ebill	loglong	logtoll	logequi	logcard	logwire	lninc	custcat	churn
1.000.000	1.000.000	1.000.000	1.000.000	0.000000	1.840.550	3.719.651	4.021.774	2.995.732	3.930.845	3.850.148	4.000.000	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	1.348.073					2.397.895	1.000.000	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	2.297.573	2.691.243	3.263.849			3.583.519	2.000.000	1.000.000
0.000000	1.000.000	1.000.000	0.000000	1.000.000	1.515.127	2.756.840	3.496.508		3.175.968	2.639.057	4.000.000	1.000.000
0.000000	0.000000	0.000000	0.000000	1.000.000	1.558.145		3.063.391			3.218.876	1.000.000	1.000.000
1.000.000	1.000.000	1.000.000	1.000.000	0.000000	2.561.096	3.481.240		3.267.666		3.135.494	3.000.000	0.000000
1.000.000	1.000.000	1.000.000	1.000.000	0.000000	3.482.777	3.569.533		2.691.243	3.707.456	4.682.131	3.000.000	0.000000
1.000.000	1.000.000	1.000.000	1.000.000	1.000.000	2.327.278	3.068.053		3.248.435	4.074.992	4.644.391	4.000.000	1.000.000
1.000.000	1.000.000	0.000000	1.000.000	1.000.000	1.238.374				2.714.695	3.828.641	4.000.000	0.000000
1.000.000	0.000000	1.000.000	1.000.000	0.000000	2.830.268	3.761.200		3.349.904		3.178.054	3.000.000	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	1.373.716			2.484.907		3.218.876	1.000.000	0.000000
0.000000	0.000000	0.000000	1.000.000	0.000000	2.613.740			3.032.546		4.615.121	2.000.000	0.000000
1.000.000	1.000.000	0.000000	1.000.000	1.000.000	1.193.922		3.533.687	2.420.368		4.174.387	4.000.000	1.000.000
0.000000	0.000000	0.000000	0.000000	0.000000	2.393.339					3.610.918	1.000.000	1.000.000
1.000.000	1.000.000	1.000.000	1.000.000	0.000000	2.224.624	3.044.522	3.723.281	2.876.386	3.713.572	3.401.197	3.000.000	1.000.000

EK 2: RANK (SIRALAMA) TANIMI

Elimizde aşağıdaki gibi bir veri seti olduğu düşünülürse, Bu veri setinde 10 değerinin rankını bulmak için izleyeceğimiz yol;

7
12
4
10
3
21
9

Rankın sıralama değeri olduğunu söylemiştik. 10 değerinin veya diğer herhangi bir değerin rankını öğrenmek için elimizdeki veri seti küçükten büyüğe doğru sıralanır. Bu sıralamadaki sıra değeri, o verinin rankıdır. Bu açıklama yukarıdaki örneğe uygulandığında;

Veri	Rank
3	1
4	2
7	3
9	4
10	5
12	6
21	7

Bu verilerde 10 değerinin rankı 5'tir. Aynı şekilde 21 değerinin rankı 7, 3 değerinin rankı ise 1'dir. Buraya kadar her şey kolay görünüyor. Peki veri setinde aynı değerler var ise ne olacak? Örneğin yukarıdaki verilere 10 değerinde bir veri daha ekleyelim ve 10'un rankını soralım. Verilere bir 10 değeri daha eklenmiş ve küçükten büyüğe doğru sıralanmış hali aşağıda yer almaktadır.

Veri	Sıralama
3	1
4	2
7	3
9	4
10	5
10	6
12	7
21	8

Yukarıda görüldüğü gibi normal şekilde sıralama yapılır. Ancak 10 değerini taşıyan veri iki tanedir ve dolayısıyla da rankları aynı olmalıdır. Peki bu rank kaçtır? Bunun için aynı değere sahip verilerin kaçınıcı sıralarda olduklarına bakılır. Yukarıda ki sette 10 değeri 5 ve 6. sıradadır. Öyle ise 5 ve 6. sırayı 10 değerlerine eşit şekilde paylaşmalıyız. Yani paylaştıkları sıralama değerlerinin ortalamasını bu verilere atamalıyız. Şöyle ki $5+6=11$ 'dir. $11/2=5,5$. Demek ki veri setinde ki 10 değerlerinin rankları 5,5'tir. Rank sıralaması aşağıdaki gibidir.

Veri	Sıralama	Rank
3	1	1
4	2	2
7	3	3
9	4	4
10	5	5,5
10	6	5,5
12	7	7
21	8	8

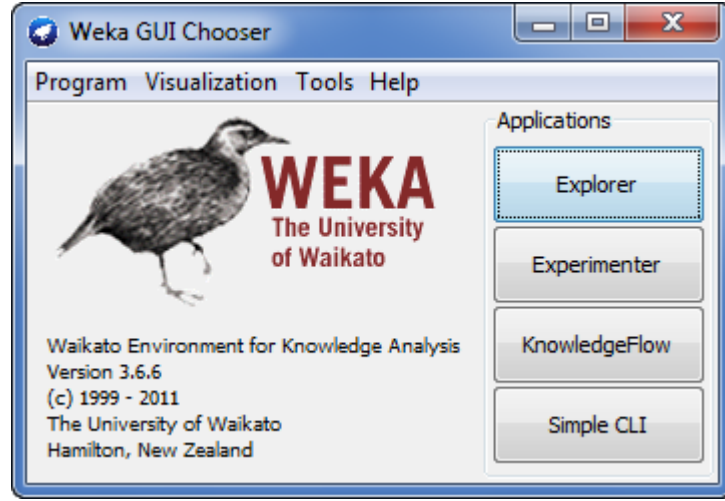
EK 3: WEKA PROGRAMI HAKKINDA GENEL BİLGİLER

WEKA(Waikato Environment for Knowledge Analysis), Yeni Zelanda Waikato Üniversitesi' nde geliştirilen bir veri madenciliği ve makine öğrenmesi (machine learning) yazılımıdır. WEKA yazılımı nesneye yönelik programlama dillerinden olan Java ile geliştirilmiştir. Java birçok değişik öğrenme algoritmaları için düzenli bir platform sağlamaktadır. WEKA' nın en güçlü özelliği birçok sınıflandırma tekniklerini içermesidir. Diğer bir özelliği de uygulamaların komut girilerek gerçekleştirilmesine imkân tanınmasıdır.³⁷²

WEKA dört temel modülden oluşmaktadır. Bu modüllere GUI(General User Interface) genel kullanıcı arayüzünden ulaşılabilir.

Simple CLI	Basit komut satırı arayüzü olarak adlandırılan bu modül, WEKA komutlarının direkt olarak çalıştırılmasını sağlar.
Explorer	WEKA ile veri keşfi yapılmasına imkan sağlayan bir platformdur. Veri madenciliği tekniklerini kullanmayı ve görselleştirme yapmayı sağlayan modüldür. Bu modül veri madenciliği açısından detaylı olarak incelenecektir.
Experimenter	Öğrenme setlerinin denemelerinin ve aralarındaki istatistiksel testlerin yapılmasını sağlayan modüldür(Deneme Modülü).
Knowledge Flow	Explorer modülünün grafik olarak temsil edilmesidir. Bilgi akışının modellenmesini sağlayan bir modüldür.

³⁷² WEKA, Machine Learning Software İn Java, <http://www.Cs.Waikato.Ac.Nz /Ml/Weka>, 3,2011



Şekil 1 WEKA Genel Kullanıcı Arayüzü

Explorer ilk çalıştırıldığında preprocess menüsü dışındakiler aktif değildir. Bunun sebebi öncelikle bir veri seti seçilmesini sağlamaktır. Veri seti bir dosyadan, veritabanından yada URL adresi girilerek seçilebilir. WEKA, veri setleri için *.arff dosya formatını kabul etmektedir. Veri setini bu formata dönüştürmek için şu yol izlenir.

- 1)Excel gibi bir tablo düzenleyicide veriler girilir. (a)
- 2)Düzenlenen tablo farklı kaydet menüsüyle *.csv (virgülle ayrılmış) formatında kaydedilir. (b)

	A	B	C	D	E
	outlook	temperature	humidity	windy	play
1					
2					
3	sunny	85	85	FALSE	no
4	sunny	80	90	TRUE	no
5	overcast	83	86	FALSE	yes
6	rainy	70	96	FALSE	yes
7	rainy	68	80	FALSE	yes
8	rainy	65	70	TRUE	no
9	overcast	64	65	TRUE	yes
10	sunny	72	95	FALSE	no
11	sunny	69	70	FALSE	yes
12	rainy	75	80	FALSE	yes
13	sunny	75	70	TRUE	yes
14	overcast	72	90	TRUE	yes
15	overcast	81	75	FALSE	yes
16	rainy	71	91	TRUE	no
17					
18					
19					
20					

(a)

```

outlook,temperature,humidity,windy,play
sunny,85,85,FALSE,no
sunny,80,90,TRUE,no
overcast,83,86,FALSE,yes
rainy,70,96,FALSE,yes
rainy,68,80,FALSE,yes
rainy,65,70,TRUE,no
overcast,64,65,TRUE,yes
sunny,72,95,FALSE,no
sunny,69,70,FALSE,yes
rainy,75,80,FALSE,yes
sunny,75,70,TRUE,yes
overcast,72,90,TRUE,yes
overcast,81,75,FALSE,yes
rainy,71,91,TRUE,no

```

(b)

Şekil 2 Örnek Bir Veri Seti Düzeni

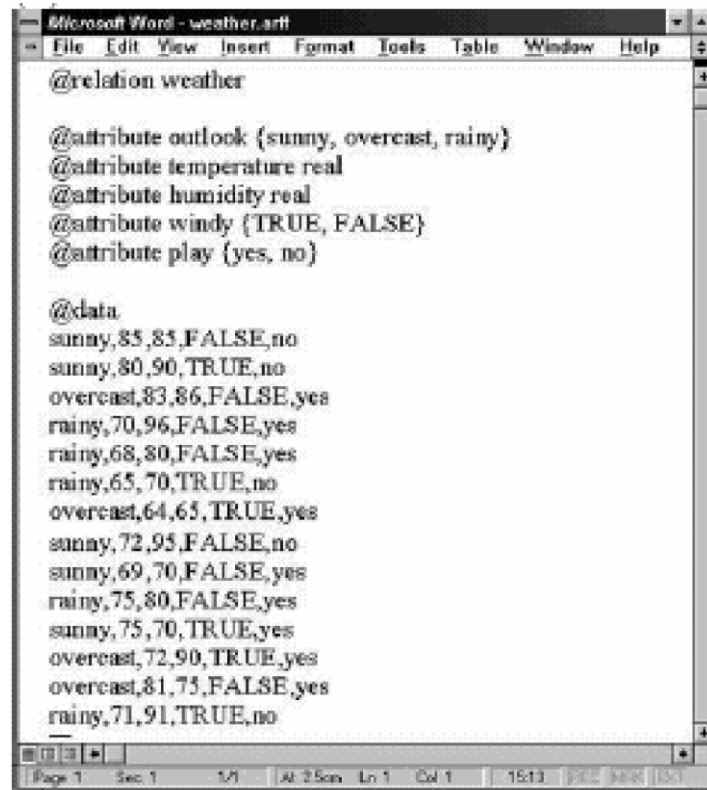
3. *.csv formatında kaydedilen bu belge herhangi bir metin düzenleyicide açılarak çeşitli düzenlemeler yapılır. Girilen veri setindeki sütun isimleri bir nitelik (attribute) olarak temsil edilir. Her nitelik için bir tanımlama yapılır. Nitelik değerleri sayısal (real yada integer), nominal, string yada tarih(date) olabilir(c).

@relation weather

@attribute outlook {sunny, overcast, rainy} @attribute temperature

real @attribute humidity real @attribute windy {TRUE, FALSE}

@attribute play {yes, no} Nominal değerler aralarına virgül konularak {} aralığında yazılır.



Şekil 3 Arff Veri Formatı

Bu şekilde düzenlendikten sonra metin düzenleyicide *.txt olarak kaydedilip kapatılarak uzantısı *.arff olarak değiştirilir. Eğer WEKA yazılımı bilgisayarda kurulu ise bu işlemten sonra belge WEKA sembolünü alacaktır.

WEKA içerisinde var olan weather.arff örnek veri setini açtığımızda diğer tüm alt modüller (Classify, Cluster, Associate, Select Attribute, Visualize) aktif hale gelir.³⁷³

³⁷³ WEKA, Machine Learning Software In Java, <http://www.Cs.Waikato.Ac.Nz /Ml/Weka>, 3,2011

KAYNAKÇA

Adrian M. Kentsch, Walter A. Kusters, Peter van der Putten and Frank W. Takes. **“Exploratory Recommendations Using Wikipedia’s Linking Structure.”** Benelearn 2011, The Hague, ss 61-68, 2011.

Alabay, M.N, **Müşteri İlişkileri Yönetimi MİY: Turizm Sektörü Üzerinde Bir Uygulama**, Doktora Tezi, Gazi Üniversitesi SBE, Ankara, 2007.

Alptürk, H., **Yeni Yasa Yeni Yapılanma ve Anlayış**, Türk Telekom Dergisi, Kozan Ofset Baskı, Sayı:5, Eylül Ekim, 2000.

Alptürk, H.,**Türk Telekom Radyo TV İletişim Fuarında**, Türk Telekom Dergisi, Eylül-Ekim 2000.

Agrawal R., **“Database Mining : A Performance Perspective”**, IEEE Transactions on Knowledge and Data Engineering, ss 914-925,1993.

Agrawal, R. ve Srikant, R., **“Fast Algorithms for Mining Association Rules in Large Databases”**, Proceedings of the Twentieth International Conference on Very Large Databases, 1994.

Allahverdi N. , **Uzman Sistemler, Bir Yapay Zeka Uygulaması**, Atlas Yayın Dağıtım, Mart 2001.

Alpaydın, E., **“Zeki Veri Madenciliği: Ham Veriden Altın Bilgiye Ulasma Yöntemleri”**, Bilisim 2000 Eğitim Semineri, İstanbul, 2000.

Al-Radaideh, Qasem A., AI-Shawakfa, Emad M. ve AI-Najjar, Mustafa. , **“Mining Student Data using Decision Trees.”** International Arab Conference on Information Technology.ss 5-15 ,2006

ATAK, G., **Telekomda Yeni Bir Uygulama Müsteri Memnuniyeti Esas Başarı Kaçınılmaz**, Türk Telekom Dergisi, Kasım-Aralık, 2000.

Ayers, E., Nugent, R., ve Dean, N. **“A comparison of student skill knowledge estimates.”** 2nd International Conference on Educational Data mining, Cordoba, Spain, ss 1–10, 2009.

Babu, G. P., Murty, M. N. **Clustering with evolution strategies**. Pattern Recogn. 27, 321–329, 1994.

Bach, F., Harchaoui, Z. **DIFFRAC: “A discriminative and exible framework for clustering.”**, In Advances in Neural Information Processing Systems 20, MIT Press, ss 49-56, 2008.

Backer, F. B., Hubert, L. J., **A graphtheoretic approach to goodness-of-fit in complete-link hierarchical clustering**. J. Am. Stat. Assoc. 71, 870 – 878, 1976.

Baker, M., **“Marketing Management and Strategy”**, 3rd edition, Macmillan Business, 2000.

Baker B. M. ve Ayeche M. A. , **“A Genetic Algorithm For The Vehicle Routing Problem”**, Computers & Operations Research, s. 1-14, 2002.

Bentayeb, F. ve Darmont, J., **“Decision Tree Modeling with Relational Views”**, 13th International Symposium, ISMIS 2002, Lyon ,22-27 Temmuz 2002.

Berry, M.J.A. ve Linoff, G.S., **“Data Mining Techniques For Marketing, Sales, and Customer Relationship Management”**, First Edition, Wiley Publishing, ss 187-216, 1997.

Berson, A., Smith, S. and Thearling, K., **“Building Data Mining Applications For CRM”**, Second Edition, McGraw-Hill Professional Publishing, ss 55 115, 2001.

Berry, M. Survey of Text Mining : Clustering, Classification, and Retrieval (Hardcover)Springer; 1edition, 2003.

Bishop C. , **Neural Networks for Pattern Recognition**, Oxford Univ Press, 1996.

Borgelt, C. ve Kruse, R., **“Induction of Association Rules: Apriori Implementation”**, 15th Conference on Computational Statistics, Osaka, Japan, 31Agustos, 2002.

Bozdogan, Hamparsum , **“Statistical Data Mining and Knowledge Discovery”** , Washington: A CRC Press Company, 2004.

Bhattacharyya S., **“Direct Marketing Performance Modeling Using Genetic Algorithms”**, Journal on Computing, Volume: 11, Issue: 3, s. 248-267, 1999.

Blake, C. L., & Merz, C. J. **UCI repository of machine learning databases**. <http://www.ics.uci.edu/~mllearn/MLRepository.html>,1998.

Blaschko, M.ve Gretton, A., **“Learning taxonomies by dependence maximization.”** , In Advances in Neural Information Processing Systems 21, ss 153-160. MIT Press, 2009.

Brassington, F and Pettitt, S, **“Principles of Marketing”**, Second Edition, Prentice Hall, Harlow, 2000.

Breiman, L., Friedman, J. H., Olshen, R.A. and Stone, C. J.,**“Classification and Regression Trees”**,Second Edition, CRS Press LLC, ss 1-87,1998.

Breese, J. S., Heckerman, D., & Kadie, C., **Empirical analysis of predictive algorithms for collaborative filtering**. Proc. UAI-98 ss. 43–52, 1998.

Brown C. E. ve Sumichrast R.T., “**CF-GGA: A Grouping Genetic Algorithm For The Cell Formation Problem**”, **International Journal of Production Research**, Volume: 39, Number: 16, s. 3651-3669,2001.

Shaw B. ve Jebara T. , “**Structure preserving embedding.**”, In Proceedings of the Twenty-Sixth International Conference on Machine Learning, Montreal, Quebec, Canada, June 2009.

Caswell, B., M. Roesch , “**Snort: The open source network intrusion detection system**” , <http://www.snort.org/.ddd>, 2004.

CERT/CC Statistics, 1988-2005, Mellon Software Engineering Institute, CERT Coordination Center, http://www.cert.org/stats/cert_stats.html, [Ziyaret Tarihi: 10 Ocak 2010]

Civanlar, S., **Yeni Servis Altyapısına Hazırmıyız?**, Telekom Dünyası Dergisi, İstanbul, Mayıs, 2005.

Chebrolu, S., Abraham, A. ve Thomas, J. P. , **Feature Deduction and Ensemble Design of Intrusion Detection Systems**, Computers & Security, Elsevier, 24, 295-307, 2005

Chickering, D. M. The WinMine Toolkit (Technical Report MSR-TR-2002-103). Microsoft, Redmond, WA, 2002.

Coleman T., Saunderson J. ve Wirth A., “**Spectral Clustering with Inconsistent Advice**”, Proc. Intl. Conference on Machine Learning (ICML 2008), Helsinki, Finland, ,ss 152–159, 2008.

Cover TM, Hart PE. "**Nearest neighbor pattern classification**". IEEE Transactions on Information Theory 13 (1): 21–27. doi:10.1109/TIT.1967.105396,1967.

Cho, Hyunbo; Lee, Yeonhee; and Lee, Hyeseon, **"Toward Optimal Churn Management: A Partial Least Square (PLS) Model"**, AMCIS 2010 Proceedings. s.78, 2010.

Dan L., Kefei W., Jitender S. D., **"FADS: Fuzzy Anomaly Detection System"**, LNAI 4062, ss 792-798, 2006.

Dasgupta D., **Artificial Neural Networks and Artificial Immune Systems: Similarities and Differences**, In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC '97), Orlando, October ss 12-15 1997.

Davidson I. ve Ravi S. S., **"Identifying and Generating Easy Sets of Constraints For Clustering"**, Proceedings of the Twenty-First National Conference on Artificial Intelligence (AAAI 2006), Boston, MA, 2006.

Davidson I. ve Ravi S. S., **"The Complexity of NonHierarchical Clustering with Constraints"**, J. Knowledge Discovery and Data Mining (DMKD), Vol. 14, No. 1, ss 25–61, 2007.

Davidson I. ve Ravi S. S., **"Intractability and Clustering with Constraints"**, Proc. Intl. Conference on Machine Learning (ICML 2007), Corvallis, OR, ss 201–208, 2007.

Davidson I. ve Ravi S. S., **"Using Instance-Level Constraints in Hierarchical Agglomerative Clustering: Theoretical and Empirical Results"**, Data Mining and Knowledge Discovery, Vol. 18, No. 2, ss 257–282, 2009.

Davidson I., Wagstaff K. and Basu S., **"Measuring Constraint-Set Utility for Partitional Clustering Algorithms"**, Proc. Intl. Conf. Principles and Practice of Knowledge Discovery in Databases (PKDD 2006), Berlin, Germany, ss 115–126, 2006.

Davis L. ve Coombs S. (1987), **"Genetic Algorithms And Communication Link Speed Design: Theoretical Considerations"**, Grefenstette, s. 252-256

Dejan R., Peter van der Putten and K. K. L. “**Customer Satisfaction and Network Experience in Mobile Telecommunications**” , Benelearn 2011, The Hague, ss 91-92, 2011.

Dejan R., Peter van der Putten and Kim K. L.. “**The Impact of Experimental Setup in Prepaid Churn Prediction for Mobile Telecommunications: What to Predict, for Whom and Does the Customer Experience Matter?**” Transactions on Machine Learning and Data Mining, Volume 3, Number 2, , ss 80-99., 2010

Dejan R., Peter van der Putten ve Kim K. L. “**The Impact of Experimental Setup on Prepaid Churn Modeling: Data, Population and Outcome Definition.**” Workshop on Data Mining in Marketing (DMM 2010). Berlin, Germany, 2010.

Dejan R., “**Prepaid Churn Modeling Using Customer Experience Management Key Performance Indicators**”, second supervisor with Hans Borgman and Kim K. Larsen. 2009.

Demirel, Y, **Müşteri İlişkileri Yönetimi ve Bilgi Paylaşımı**, İstanbul: IQ Kültür Sanat Yayıncılık, 2006.

Denning, D. E. , **An intrusion-detection model**. IEEE Transactions on Software Engineering 13 (2), ss 222–232, 1987.

Dengiz, Berna, Altıparmak, Fulya, “**Genetik Algoritmalar Genel Bir Giriş**”, Endüstri mühendisliği Dergisi, Cilt:9, Sayı:3, ss 3-14, 1998.

Dunham M.H., “**Data Mining Introductory and Advanced Topics**”, Prentice Hall, Pearson Education Inc., 2003.

Elibol, Semih, **Telekomünikasyon Sektöründe Değişim ve Türk Telekom’da Yeniden Yapılanma Süreci**, İstanbul: Marmara Üniversitesi, 2002.

Engin O., **Akış Tipi Çizelgeleme Problemlerinin Genetik Algoritma ile Çözüm Performansının Arttırılmasında Parametre Optimizasyonu**, İTÜ, Arttırılmasında Parametre Optimizasyonu , ITU,Fen Bilimleri Enstitüsü, Yayınlanmamış Doktora Tezi, 2001.

Freitas, A. Alex. **Data Mining and Knowledge Discovery with Evolutionary Algorithms**, New York: Springer-Verlag Berlin Heidelberg, 2002.

Friedman, N. **The Bayesian structural EM algorithm**. Proc. UAI-98 ss 129–138,1998.

Gel, O. , **MİY Yolculuğu**, Ankara: Sistem Yayıncılık, 2004.

Geray, Haluk, “**Yeni İletişim Teknolojileri: Toplumsal Bir Yaklaşım**”, Kılıçaslan Matbaacılık, Ankara, 1994

Giudici, Paolo. “**Applied Data Mining Statistical Methods for Business and Industry**”, New York, Chichester: Wiley, 2003.

Goharian, Nazli ve David Grossman, **Data Preprocessing Tutorial**. Chicago: Illinois Institute of Technology, 2003.

Greg Hamerly, Erez Perelman, Jeremy Lau, Brad Calder, ve Timothy Sherwood. “**Using machine learning to guide architecture simulation.**”, Journal of Machine Learning Research, 7:343–378, 2010.

Gülhan Örekici Temel, Sınıflama **Ağaçları Yardımıyla Restless Legs Syndrome (RLS) Hastalarına Tanı Koyma**, İnönü Üniversitesi Tıp Fakültesi Dergisi, 2005, s.111-117.

Han, Jiawei ve Kamber Micheline. **Data Mining: Concept and Techniques**. USA: Morgan Kaufmann Publishers, 2001.

Hand, D., Mannila, H. and Smyth, P., **Principles of Data Mining**, Mit Press, 2001.

Hand, David, Mannila Heikki ve Smyth Padhraic. **Principles of Data Mining**, Cambridge, CA: MIT Press, 2001.

Hassan, M. R., Nath, B., **Stock Market Forecasting Using Hidden Markov Model: A New Approach**, In Proceedings of 5th international conference on intelligent system design and application, 192–196, 2005.

Haykin, S., **Neural Networks: A Comprehensive Foundation**, MacMillan College Publishing Company, New York, 1994.

Heckerman, D., Geiger, D., & Chickering, D. M. **Learning Bayesian networks: The combination of knowledge and statist. data.** Machine Learning, 20, 197–243, 1995.

Hoyer, W.D. and MacInnis, D.J. , **“Consumer Behaviour”**, 2nd Edition, USA: Houghton Mifflin Company, 2001.

James, V.H., Ray, D. N. **Neural Networks and Traditional Time Series Methods: A Synergistic Combination in State Economics Forecasts**, IEEE Trans. Neural Networks , 8(4), 863-873, 1997

Jain, A. ve R. Dubes. **Algorithms for Clustering Data**. Englewood Cliffs, New Jersey: Prentice-Hall, 1988.

Jiang, S., Song, X., Wang, H., Han, J. and Li, O., **A Clustering-Based Method for Unsupervised Intrusion Detections**, Elsevier, 27.7, 802-10, 2006.

Jiawei Han ve Micheline Kamber, **Data Mining: Concept and Techniques**, USA: Morgan Kaufmann Publishers, s.39-40, 2001.

Jiu-Ling Z., Jiu-Fen Z., Jian-Jun L., **“Intrusion detection based on clustering genetic algorithm”**, Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on Volume: 6, ss 3911- 3914 Vol. 6, 2005.

Kantardzic, Mehmed. **Data Mining: Concepts, Models, Methods and Algorithms**. USA: A John Wiley and Sons, Inc., Publication, IEEE Press, 2003.

Kaufman, Leonard ve Peter J. Rousseeuw. **Finding Groups in Data: An Introduction to Cluster Analysis**, USA: John Wiley and Sons. 1990.

Keskin, Nuray, **Telekomünikasyon Sektöründe Yeniden Yapılanma**, Ankara: Ankara Üniversitesi, 2001

Kerin, Hartley, and Rudelius, **“Marketing”**, s.299, 2011.

Kulis, B., S. Basu, I. Dhillon, and R. Mooney. **“Semi-supervised graph clustering: A kernel approach.”** Machine Learning, 74(1):ss 1-22, 2009.

Kulis, B., A. Surendran, and J. Platt. , **“Fast low-rank semidefinite programming for embedding and clustering.”** In Proceedings of the Eleventh International Workshop on Artificial Intelligence and Statistics, ss 235-242, San Juan, Puerto Rico, 2007.

Kurt, Mustafa.-Cumali, Sametay. **“Genetik Algoritma ve Uygulama Alanları”**, Mühendis ve Makina, Sayı:501, Ekim 2001.

Latin, M. James, Carroll J. Douglas ve Paul E. Green. **Analyzing Multivariate Data**. United States: Thomson Brooks-Cole, 2003.

Lee, W. and Stolfo, S. J., **A Framework for Constructing Features and Models for Intrusion Detection Systems**, ACM, 3.4, ss 227 – 261, 2000.

Leung, K. and Leckie, C., **Unsupervised Anomaly Detection in Network Intrusion Detection Using Clusters**, Proceedings of the Twenty-eighth Australasian conference on Computer Science, ACM, 38, 333-342, 2005.

Li, Y. and Guo, L., **An active learning based TCM-KNN algorithm for supervised network intrusion detection**, Computers & Security, Elsevier, s 26, 2007.

Lovelock, “**Services Marketing, people, technology, strategy**”, Prentice Hall. , 2001.

Marina T., Chuanyi J., “**Anomaly Detection in IP Networks**”, IEEE Transaction on signal processing, Vol. 51, No. 8, 2003.

Mitchell T. , **Machine Learning**, McGraw-Hill, 1997.

Mızrak. E., **Türk Telekom Kalite Yolculugunda İlk Adımı Attı**, Telekomünikasyon Mühendisleri Dernegi Dergisi, Mart 2000.

Murthy, S.K., "Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey," Data Mining and Knowledge Discovery, 1998, s.345-389.

Naresh, K. Malhotra. **Marketing Research An Applied Orientation**. Second Edition, New Jersey: Prentice Hall Inc, 1996.

Nikulin, V., **Threshold-Based Clustering with Merging and Regularization in Application to Network Intrusion Detection**, Computational Statistics & Data Analysis, Elsevier, 51.2, 1184-1196, 2005.

Quadrianto N., Song L., ve Smola A.. “**Kernelized sorting.**” In Advances in Neural Information Processing Systems 21, ss 1289-1296. MIT Press, 2009.

Oğuzlar, Ayse. **Veri Madenciligine Giriş**, Bursa: Ekin Kitabevi, 2004.

Oh, S. H. and Lee, W. S., **An Anomaly Intrusion Detection Method by Clustering Normal User Behavior**, Computers & Security, Elsevier, 22.7., 596-612, 2003.

Qin, Z., **ROC Analysis for Predictions Made by Probabilistic Classifiers**, Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, IEEE, 5, 3119 – 3124, 2005.

Paul S. Bradley ve Usama M. Fayyad. **“Refining initial points for k-means clustering.”** In Jude W. Shavlik, editor, ICML, ss 91–99. Morgan Kaufmann, 1998.

Peter van der Putten, Cor Veenman, Joaquin Vanschoren, Menno Israel ve Hendrik Blockeel. ,**“Benelearn 2011”**: Proceedings of the 20th Annual Belgian Dutch Conference on Machine Learning. The Hague, 2011.

Peter van der Putten. **“On Data Mining in Context: Cases, Fusion and Evaluation.”**, PhD Thesis, Leiden Institute of Advanced Computer Science (LIACS), Leiden University. 2010.

Peter van der Putten, Joost N. K. ve Ling J. M., **“Profiling Novel Classification Algorithms: Artificial Immune Systems.”** , In Proceedings 7th IEEE International Conference on Cybernetic Intelligent Systems (CIS2008), London, United Kingdom, 2008.

Pietraszek, T. and Tanner, **A Data Mining and Machine Learning Towards Reducing False Positives in Intrusion Detection**, Information Security Technical Report, Elsevier, 10,3, 169-83, 2005.

Pride, W. and Ferrell O. **“Marketing”**, s.36, 2010.

Proctor, T. , **“Essentials of Marketing research”**, UK: Financial Times-Prentice Hall, 2000.

Rakesh Agrawal, **Mining association rules between sets of items in large databases**, In Proceedings of the ACM SIGMOD Conference on Management of Data, Washington, D. C., 1993, s.207-216.

Ratchford, Brian T. , “**Online Pricing: Review and Directions for Research**”, Journal of Interactive Marketing, 23 (1), 82-90, 2009.

Randall, G. , “**Principles of Marketing**”, 2nd edition, Thomson Learning, 2001.

Ravinda, Khattree ve Naik N. Dayanand. **Multivariate Data Reduction and Discrimination**. First Edition. North Caroline: Cary, 2002.

Rencher A.C. **Methods of Multivariate Analysis**, Wiley, 1995.

Roiger, Richard ve Michael Geatz. **Data Mining A Tutorial-Based Primer**. NewYork, USA: Addison Wesley, 2003.

Safel, R., **Telekomünikasyon Sektörü**, Türkiye Vakıflar Bankası T.A.O Planlama ve İktisadi Araştırmalar Grup Yönetmenliği Sektör Araştırmaları, Seri No: 25, Ekim 2001.

Sankar, K. Pal ve Pabitra Mitra. **Pattern Recognition Algorithms for Data Mining**. Washington: A CRC Press Company, 2004.

Saygı, Nur, **Telekomünikasyon ve Bilgi teknolojileri Pazarı**, Ankara: Telekomünikasyon Kurumu, 2002.

Sarısu, A., **Dünya’da ve Türkiye’de Özelleştirme**, Genel Bir Değerlendirme, Kasım,2003.

Sheen, Shina Rajesh, R. , “**Network intrusion detection using feature selection and Decision tree classifier**” TENCON 2008 - 2008, TENCON 2008. IEEE Region 10 Conference, pp:1-4, 2008.

Shai Ben-David, D'avid P'al, ve Hans Ulrich Simon. **“Stability of k-means clustering”** Lecture Notes in Computer Science, 4539, ss 20–34, 2007

Shun, J. Malki, H.A., **“Network Intrusion Detection System Using Neural Networks”**, ICNC '08. Fourth International Conference on, Volume: 5, pp: 242-246, 2008.

Sönmez, D. **Müşteri İlişkileri Yönetiminde Bilişim Teknolojilerinin Müşteri Odaklılığını Sağlamadaki Rolü**. Yüksek Lisans Tezi, Gazi Üniversitesi, Ankara, 2007.

Stein, G., Chen, B., Wu, A. S. and Hua, K. A., **Decision Tree Classifier For Network Intrusion Detection With GA-based Feature Selection**, ACM, 2, 136 – 141, 2005.

S. Khot and A. Naor. ,**“Approximate kernel clustering.”**, In Proceedings of the Forty-Ninth Symposium on Foundations of Computer Science, pages 561-570, Philadelphia, PA, USA, 2008.

Stephen, M. S., **“Thomas bayes' bayesian inference”**, Journal of the Royal Statistical Society, Series A, ss 250–258 ,1982.

Şen, Z., **“Bulanık Mantık ve Modelleme İlkeleri”**, Bilge Kültür Sanat, İstanbul, 2001.

Süel, H., **Türk Telekomünikasyon Sektörü ve Özelleştirme**, İktisat İşletme ve Finans Dergisi, Nisan,1998.

Süel, H., **Türk Telekomünikasyon Sektörü ve Özelleştirme**, İktisat İşletme ve Finans Dergisi, Nisan,1998.

Tan, P., Steinbach, M. ve Kumar V., **Introduction to Data Mining**, Boston, MA: Addison-Wesley is an Imprint of Pearson Education Inc., 2005.

Tanaka, K.: **“An Introduction to Fuzzy Logic for Practical Applications”**,Rassel Inc., Japan, 1997.

Tarek A., Adel B., Michael R., **“Protocol analysis in intrusion detection using decision tree”** **Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'04) Volume 2**, 2004.

Tek, Ö.B., **Pazarlamada Değer Yaratmak**. İstanbul: Hayat Yayınları, 2006.

Tek, Ö. B. ve Özgül E. , **Modern Pazarlama İlkeleri**. İzmir: Birleşik Matbaacılık, 2005.

Tuğ E., Şakiroğlu M., Bulun M., **Tıbbi Veri Tabanlarında Gizli Bilgilerin Keşfedilmesi**, 20. Bilişim Kurultayı, İstanbul, 2003.

Tuncer, T., Tatar, Y., **"Detection SYN Flooding Attacks Using Fuzzy Logic"**, International Conference on Information Security and Assurance, ss 321-325, 2008.

Utgoff, P.E, **Decision trees**, 1998, s. 222-224.

Weiss S. M., N. Indurkha , **Predictive Data Mining: A Practical Guide**, Morgan Kaufmann, 1998.

Vasilios A. Siris, Fotini P., **“Application of Anomaly Detection Algorithms for Detecting SYN Flooding Attacks”**, Computer Communications, ss 1433-1442, 2006.

Yılmaz, K., **Türk Telekomünikasyon Sektöründe Reform:Özelleştirme, Düzenleme ve Serbestleşme**, İstanbul: Koç Üniversitesi,1999.

Zaki, M. ve Sobh, T.S., **NCDS: Data Mining for Discovering Interesting Network Characteristics**, Information and Software Technology, Elsevier, 47.3, 18998, 2005.

Zaki, M.J., Parthasarathy, S., Ogihara, M. ve LI. W., “**New Parallel Algorithms for Fast Discovery of Association Rules**”, Data Mining and Knowledge Discovery , 4, 343-373, 1997.

ÖZGEÇMİŞ

1981 tarihinde İstanbul'da doğdu. İlk, Orta ve Lise öğrenimini babasının işi dolayısıyla Kdz.Ereğli'de tamamladı. Daha sonra Üniversite eğitimi için 2000 senesinde İstanbul'a geldi ve 2007 senesinde evlendi.

Bahçeşehir Üniversitesi Bilgisayar Mühendisliğinden 2004'de lisans, Haliç Üniversitesi Bilgisayar Mühendisliğinden 2007'de yüksek lisansını aldı. 11/2005-11/2009 tarihleri arasında Haliç Üniversitesinde Rektörlüğe bağlı olarak Enformatik Okutmanlığı, görevlendirme ile Bilgisayar Mühendisliği bölümünde öğretim görevliliği ünvanlarıyla çalıştı. 11/2009 tarihinden bu yana İstanbul Arel Üniversitesi Mühendislik Mimarlık Fakültesi Bilgisayar Mühendisliği bölümünde Öğretim Görevlisi ünvanıyla görevine halen devam etmektedir.

Çalışma süresi içerisinde bir çok Fakülte, Yüksekokul'da Türkçe ve İngilizce olarak Lisans dersleri verdi. Ayrıca yüksek lisans derslerinde araştırma asistanlığı yaptı. Haliç Üniversitesinde İşletme, Fen Edebiyat, Mühendislik Fakülteleri, Konservatuar, Beden Eğitimi Spor Yüksekokullarında Temel bilgi teknolojileri, programlama, yapay sinir ağları konularında dersler verdi. 2009 kasım itibariyle İstanbul Arel Üniversitesinde Lisans düzeyinde Muhasebe Bilgi Sistemlerinde Yönetim Bilgi Sistemleri, Bilgisayar, Elektrik Elektronik, Endüstri Mühendislikleri ve Fen Edebiyat Fakültelerinde ingilizce olarak Computer I,II, Algorithms & Programming, Introduction to Information Technologies derslerini verdi. Mimarlık ve diğer tüm fakültelerde Matematik I, Bilgisayar I, II derslerinin yanısıra, tüm fakülte ve yüksekokullarda ortak zorunlu ders kabul edilen Bilgisayar I dersini uzaktan eğitim sistemi üzerinden ekim 2011 tarihinden bu yana vermeye başlamıştır.

Ekim 2011'de İstanbul Arel Üniversitesinde, bilimsel etkinliklere katkı plaketi ile ödüllendirildi. Bir çok farklı alanda 25 akademik yayın yaptı. Uluslararası bilimsel toplantılarda sunulan ve bildiri kitabında basılan bir tanesi SSCI indeksli olmak üzere 18 yayını, ulusal bilimsel toplantılarda sunulan ve bildiri kitabında basılan 5 yayını,

ulusal toplantılarda sunulan ve bildiri kitabında basılan 1 yayını, Uluslararası hakemli dergilerde yayınlanan 1 yayını vardır. Çok iyi derecede İngilizce, ispanyolca bilmektedir.

Şimdiki araştırma alanları, yapay zeka, uzaktan eğitimidir.