

**GSM ŐEBEKELERİNİN PERFORMANSLARININ
VERİ MADENCİLİĐİ İLE ANALİZİ**

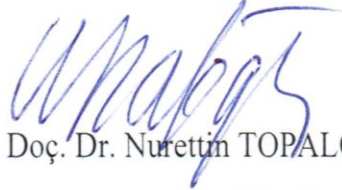
Serkan SAVAŐ

**YÜKSEK LİSANS TEZİ
ELEKTRONİK BİLGİSAYAR EĐİTİMİ**

**GAZİ ÜNİVERSİTESİ
BİLİŐİM ENSTİTÜŐÜ**

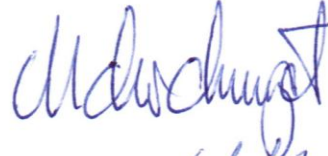
**HAZİRAN 2011
ANKARA**

Serkan SAVAŞ tarafından hazırlanan GSM ŞEBEKELERİNİN PERFORMANSLARININ VERİ MADENCİLİĞİ İLE ANALİZİ adlı bu tezin Yüksek Lisans tezi olarak uygun olduğunu onaylarım.

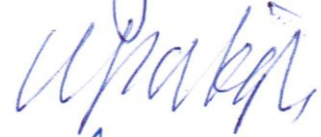

Yrd. Doç. Dr. Nurettin TOPALOĞLU
Tez Yöneticisi

Bu çalışma, jürimiz tarafından oy birliği / ~~oy çokluğu~~ ile Elektronik Bilgisayar Eğitimi Anabilim Dalında Yüksek lisans tezi olarak kabul edilmiştir.

Başkan: : Doç. Dr. M. Ali AKÇAYOL



Üye : Yrd. Doç. Dr. Nurettin TOPALOĞLU



Üye : Yrd. Doç. Dr. Fecir DURAN



Tarih : 02/ 06 /2011

Bu tez, Gazi Üniversitesi Bilişim Enstitüsü tez yazım kurallarına uygundur.

TEZ BİLDİRİMİ

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

Serkan SAVAŞ

**GSM ŐEBEKELERİNİN PERFORMANSLARININ
VERİ MADENCİLİĐİ İLE ANALİZİ
(Yüksek Lisans Tezi)**

Serkan SAVAŐ

**GAZİ ÜNİVERSİTESİ
BİLİŐİM ENSTİTÜSÜ
HAZİRAN 2011**

ÖZET

Bilgi ve iletişim teknolojilerini kullanan tüm kurumlar, bilgisayar sistemleri ve elektronik araçların ürettikleri verileri, kurumlarının veri tabanlarına kaydetmektedirler. Veri madenciliĐi, çok büyük miktarda bilginin depolandığı veri tabanlarından, amacımız doğrultusunda, gelecek ile ilgili tahminler yapmamızı sağlayacak, anlamlı olan veriye ulaşma ve veriyi kullanma işidir. Bilgi ve iletişim teknolojilerinde yaşanan gelişmelere paralel olarak, bu alanlarda kullanılan yazılım teknolojileri de sürekli gelişme halindedir. İnsanların yeni gözdesi, zaman ve mekandan bağımsız olmak ve hareket özgürlüĐü anlamına gelen “mobilité” olmuştur. Günümüzde ülke nüfusu sayısı kadar GSM Şebekesi abonelikleri, her bireyde mobil iletişim araçları, pek çok bireyin çift hatlı kullanıma geçmeleri ve mobil iletişimin sadece sesli görüşme için değil aynı zamanda görüntülü iletişim ve internet gibi başka amaçlarla da kullanılması göstermektedir ki, mobil iletişim hayatımızda çok büyük önem taşımaktadır. Bu kadar çeşitli kullanım alanı ve yoğun kullanım, mobil iletişimde üretilen verilerin boyutunu çok büyük ölçeklere çıkarmıştır. Böylece mobil iletişim alanında veri madenciliĐini kullanmak, yöneticiler ve işletme sahipleri için kaçınılmaz olmaktadır. Bu tez çalışmasında MS Visual Studio C#.NET 2008 programlama dili kullanılarak, üç ayrı GSM şebekesinin sinyal

verilerini cep telefonundan alarak veri tabanı oluşturan bir program geliştirilmiştir. Kaydedilen bu veriler veri madenciliği için uygun hale getirilmiş, ardından Frekans Analizi ve İki Aşamalı Kümeleme yöntemleriyle analiz edilerek sonuçlara ulaşılmıştır. Elde edilen sonuçlar göstermiştir ki, A Şebekesi sinyal gücü baz alındığında en yüksek oranda çekim gücüne ve en istikrarlı çekim seviyesine sahiptir. Daha sonra bu şebekeyi B Şebekesi izlemektedir. Çekim gücü performansı açısından en düşük performans gösteren şebeke C Şebekesi olmuştur. Ayrıca elde edilen diğer farklı bulgular da değerlendirilerek yorumlanmıştır.

Bilim Kodu : 702.1.013
Anahtar Kelime : veri madenciliği, GSM şebekeleri, mobil iletişim, iki aşamalı kümeleme, sinyal gücü, frekans analizi
Sayfa Adedi : 110
Tez Yöneticisi : Yrd. Doç. Dr. Nurettin TOPALOĞLU

**ANALYSIS OF GSM NETWORKS' PERFORMANCE
WITH DATA MINING
(M.Sc. Thesis)**

Serkan SAVAŞ

**GAZI UNIVERSITY
INFORMATICS INSTITUTE
JUNE 2011**

ABSTRACT

All institutions that use information and communication technologies, save the data which is produced with computer systems and electronics devices to databases of their institutions. Data mining is accessing and using the meaningful information job from the very large amount of information stored databases which will allow us to predictions about the future in line with our purposes. Parallel to developments in information and communication technologies, software technologies that used in these areas are in continuous improvement. The new favorite of people has been “mobility” which means to be independent of time and place and freedom of movement. Recently, GSM Networks subscriptions as the number of population of the country, mobile communication devices at everybody, lots of person’s two subscriptions usage, and using mobile communications not only for communication but also for video communication and internet shows that mobile communications are very important for our life. Variety of this usage area and intensive usage made very big amount of data in mobile communication. Thus, using data mining in mobile communication became inevitable for company owners and managers. In this thesis, a program is developed with using MS Visual Stdio C#.NET 2008 programming language which creates database by taking three separate GSM

networks' signal data from mobile phone. The data which is stored from mobile phones is made available for data mining and then is analyzed with Frequency Analysis and Two Step Cluster methods and results are declared. These results showed that A Network has the most powerful range of signal quality and the most stable measurement. Then B Network follows the A Network. C Network has the lowest performance for the signal quality range. Also other findings are declared and commented.

Science Code : 702.1.013

Key Words : data mining, GSM networks, mobile communication, two step cluster, signal power, frequency analysis

Page Number : 110

Adviser : Yrd. Doç. Dr. Nurettin TOPALOĞLU

TEŐEKKÜR

Çalıőmalarım boyunca deęerli yardım ve katkılarıyla beni yönlendiren ve her konuda yol gösteren hocam Yrd. Doç. Dr. Nurettin TOPALOĐLU'na, yine kıymetli fikirlerinden faydalandığım hocam Yrd. Doç. Dr. Fecir DURAN'a, ayrıca çalışmamın her anında bana destek olan deęerli eşim Tuęba SAVAŐ'a ve bana her aşamada yardımcı olan çalıştığım kurumda görevli tüm çalışma arkadaşlarıma teşekkürü bir borç bilirim.

İÇİNDEKİLER

	Sayfa
ÖZET.....	iv
ABSTRACT.....	vi
TEŞEKKÜR.....	viii
İÇİNDEKİLER	ix
ÇİZELGELERİN LİSTESİ.....	xii
ŞEKİLLERİN LİSTESİ	xiii
RESİMLERİN LİSTESİ	xiv
HARİTALARIN LİSTESİ.....	xv
SİMGELER VE KISALTMALAR.....	xvi
1. GİRİŞ	1
2. VERİ MADENCİLİĞİ.....	6
2.1. Veri Madenciliği Tarihi	9
2.2. Veri Madenciliğinin Kullanıldığı Alanlar.....	10
2.3. Daha Önce Gerçekleştirilen Veri Madenciliği Uygulamaları.....	14
2.4. Veri Madenciliğini Etkileyen Etmenler	20
2.5. Veri Madenciliğinde Karşılaşılan Problemler	22
2.5.1. Artık veri	22
2.5.2. Belirsizlik	22
2.5.3. Boş veri	22
2.5.4. Dinamik veri.....	23
2.5.5. Eksik veri.....	23
2.5.6. Farklı tipteki verileri ele alma	24
2.5.7. Gürültülü ve kayıp değerler.....	24
2.5.8. Sınırlı bilgi.....	25
2.5.9. Veri tabanı boyutu	25
2.6. Veri Madenciliği Süreci	25
2.6.1. Problemin tanımlanması.....	27
2.6.2. Verilerin hazırlanması	27

2.6.3. Modelin kurulması ve değerlendirilmesi.....	28
2.6.4. Modelin kullanılması.....	31
2.6.5. Modelin izlenmesi	31
2.7. Veri Madenciliği Yöntemleri.....	32
2.7.1. Sınıflama ve regresyon.....	35
2.7.2. Kümeleme	42
2.7.3. Birliktelik kuralları	45
3. MOBİL İLETİŞİM.....	47
3.1. Mobil İletişim Tarihçesi.....	49
3.1.1. Türkiye’de mobil iletişim süreci	50
3.2. Mobil Cihazlarda Kullanılan İşletim Sistemleri	51
3.2.1. Iphone (iOS).....	53
3.2.2. Android.....	54
3.2.3. Windows mobil	56
3.2.4. BlackBerry	57
3.2.5. Palm OS.....	58
3.2.6. Symbian OS.....	59
3.2.7. Bada.....	60
3.3. Mobil Cihazlar	61
3.4. GSM Sistemi.....	65
3.4.1. Türkiye’de kullanılan GSM şebekeleri	68
3.4.2. GSM şebekeleri ve veri madenciliğini kullanım nedeni	69
4. VERİ MADENCİLİĞİNİN GSM ŞEBEKELERİNE UYGULANMASI.....	70
4.1. Tasarım	70
4.1.1. Program tasarımında kullanılan donanım teknolojileri	70
4.1.2. Program tasarımında kullanılan yazılım teknolojileri	70
4.2. Programın Amaç ve Kapsamı	71
4.3. Program Arayüzü Tanıtımı	72
4.4. Programın Kullanıldığı Güzergah.....	73
4.5. Veri Madenciliği Süreci.....	76
4.5.1. Veri madenciliği yöntemlerinin uygulanması	77

5. SONUÇ VE ÖNERİLER	92
KAYNAKLAR	94
EKLER	99
EK-1. Tasarlanan programın akış diyagramı	100
EK-2. A, B ve C şebekeleri için iki aşamalı kümeleme sonuçları	108
ÖZGEÇMİŞ	110

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 2.1. Veri madenciliğinin tarihsel gelişimi	9
Çizelge 2.2. Risk matrisi	30
Çizelge 2.3. Veri madenciliğinde kullanılan yöntemler	33
Çizelge 3.1. 3G hizmeti kullanıcı verileri	49
Çizelge 3.2. Mobil işletim sistemlerinin özellikleri	52
Çizelge 4.1. Sinyal gücünün sınıflandırılması	76
Çizelge 4.2. A Şebekesi için frekans tablosu	81
Çizelge 4.3. B Şebekesi için frekans tablosu	82
Çizelge 4.4. C Şebekesi için frekans tablosu	83
Çizelge 4.5. A, B ve C Şebekeleri için sınıflandırılmış frekans tablosu	85
Çizelge 4.6. Sinyal gücü için iki aşamalı kümeleme sonuçları	88

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 2.1. Veri madenciliği ve disiplinler.	7
Şekil 2.2. Bilgi keşfi sürecinde veri madenciliği.	26
Şekil 2.3. Veri madenciliği süreci adımları.....	32
Şekil 2.4. Yapay sinir hücresi	38
Şekil 2.5. Şehirleri kümeleme yöntemi örneği	43
Şekil 2.6. Hiyerarşik kümeleme örneği	45
Şekil 3.1. Yıllara göre toplam arama trafiği	48
Şekil 3.2. Dünyadaki işletim sistemleri tercihleri	52
Şekil 3.3. Mobil cihaz üreticileri oranları	63
Şekil 3.4. Dünya genelindeki mobil telefon satışları	64
Şekil 3.5. Firmalar ve yıllara göre satılan cep telefonu markaları ve oranları	64
Şekil 3.6. GSM mimarisi.....	66
Şekil 3.7. İşletmeci bazında toplam abone sayıları, (milyon)	68
Şekil 3.8. Mobil işletmecilerin abone sayılarına göre pazar payları (%)	69
Şekil 4.1. Program arayüzü.....	72
Şekil 4.2. Programın çalışır durumdaki hali.....	73
Şekil 4.3. Elde edilen veri tablosundan bir bölüm	77
Şekil 4.4. A Şebekesi sinyal gücü grafiği	77
Şekil 4.5. B Şebekesi sinyal gücü grafiği.....	78
Şekil 4.6. C Şebekesi sinyal gücü grafiği.....	79
Şekil 4.7. Karşılaştırmalı sinyal gücü grafiği.....	80
Şekil 4.8. A Şebekesi için frekans dağılımı grafiği.....	81
Şekil 4.9. B Şebekesi için frekans dağılımı grafiği	83
Şekil 4.10. C Şebekesi için frekans grafiği	84
Şekil 4.11. A, B ve C Şebekeleri için sınıflandırılmış frekans grafiği.....	86
Şekil 4.12. Sinyal gücü için iki aşamalı kümeleme özellikleri	87
Şekil 4.13. İki aşamalı kümeleme yöntemi, küme elemanları grafikleri	90

RESİMLERİN LİSTESİ

Resim	Sayfa
Resim 3.1. iOS ekranı.	53
Resim 3.2. Android işletim sistemi ekranı.	55
Resim 3.3. Windows mobile işletim sistemi ekranı.	56
Resim 3.4. BlackBerry işletim sistemi ekranı.	58
Resim 3.5. Palm OS ekranı.	59
Resim 3.6. Symbian işletim sistemi ekranı.	60
Resim 3.7. Samsung bada işletim sistemi.	61
Resim 3.8. Örnek bir Tablet PC	62

HARİTALARIN LİSTESİ

Harita	Sayfa
Harita 4.1. Sinyal ölçümü yapılan güzergah (Kızılcahamam – Kazan arası)	74
Harita 4.2. Sinyal ölçümü yapılan güzergah (Kazan – Gazi Osmanpaşa Mah. arası).....	74
Harita 4.3. Sinyal ölçümü yapılan güzergah (Gazi Osmanpaşa Mah. – Anadolu Bulvarı arası).....	75
Harita 4.4. Sinyal ölçümü yapılan güzergah (Anadolu Bulvarı – Gazi Üniv. arası)..	75

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış bazı simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler	Açıklama
GB	GigaByte
TB	TeraByte
%	Yüzde
MHz	MegaHertz
Km	Kilometre
dB	Desibel
Kısaltmalar	Açıklama
VTBK	Veri Tabanında Bilgi Keşfi
2G	Second Generation
3G	Third Generation
WiMAX	Worldwide Interoperability for Microwave Access
GSM	Global System for Mobile Communication
SQL	Structured Query Language
KDD	Knowledge Discovery in Real Databases
İMKB	İstanbul Menkul Kıymetler Borsası
ÖSS	Öğrenci Seçme Sınavı
ÖSYM	Öğrenci Seçme ve Yerleştirme Merkezi
OBP	Ortaöğretim Başarı Puanı
CHAID	Chi-Square Automatic Interaction Detector
E-CHAID	Exhaustive Chi-Square Automatic Interaction Detector
CRT	Classification and Regression Tree
QUEST	Quick, Unbiased, Efficient Statistical Tree
GRI	Generalized Rule Induction
YSA	Yapay Sinir Ağları

Kısaltmalar	Açıklama
AMPS	Advanced Mobile Phone Service
PCS	Personel Communications Services
CDMA	Code Division Multiple Access
TDMA	Time Division Multiple Access
HTML	Hyper Text Markup Language
HDML	Handheld Device Markup Language
WAP	Wireless Application Protocol
WML	Wireless Markup Language
PDA	Personal Digital Assistant
RIM	Research in Motion
RTOS	Real Time Operating System
IrDA	Infrared Data Association
PIM	Personal Information Management
WAN	Wide Area Network
SMS	Short Message Service
MMS	Multimedia Message Service
ETSI	European Telecommunications Standarts Institute
ANSI	American National Standards Institute
NSS	Network Switching Subsystem
BSS	Base Station Subsystem
SIM	Subscriber Identity Module
HSCSD	High Speed Circuit Switched Data
GPRS	General Packet Radio Service
EDGE	Enhanced Data Rate for GSM Evolution
HLR	Home Location Register
VLR	Visitor Location Register
AUC	Authentication Center
EIR	Equipment Identity Register
IMEI	International Mobile Equipment Identity
BIC	Schwarz's Bayesian Information Criterion

1. GİRİŞ

Bilgi ve iletişim teknolojileri her geçen gün gücünü arttırmakta, hızla ilerlemekte ve gelişmektedir. Bu ilerleme ve gelişmelere bağlı olarak bilgi ve iletişim teknolojilerinin kullanıldığı alan sayısı da artmaktadır. Bilim ve mühendislik alanları, sağlık sektörü, uydu ve haberleşme sistemleri, bankacılık ve finans kurumları, askeri sistemler, alışveriş sektörü, devlet ve işletme yönetimleri vb. gibi pek çok alanda bilgi ve iletişim teknolojilerinden yaygın ve etkili şekilde faydalanılmaktadır.

Bilgi ve iletişim teknolojilerini kullanan tüm kurumlar, bilgisayar sistemleri ve elektronik araçların ürettikleri verileri, kurumlarının veri ambarlarına kaydetmektedirler. Zaman içerisinde de kurum idarecileri ve yöneticileri kurumları için kararlar almak, gelecek için planlamalar yapmak ve yeni stratejiler belirlemek için bu verilerden faydalanmak isteyebilmektedirler. Ancak veri ambarlarında kayıtlı tutulan veriler geçen zamanla birlikte çok büyük miktarlara ulaşmaktadır ve çıplak gözle bakıldığında bu verilerin içerisinde anlam ifade eden bilgiye ulaşmak zorlaşmaktadır.

Veri ambarlarında kayıtlı olan veriler belirli bir amaç doğrultusunda işlendiği zaman bir anlam ifade etmeye başlar [1]. Bilgi bir amaca yönelik işlenmiş veridir. Ham veriye dayalı kararlar almak veya geçmişte yaşanan kötü bir tecrübeden kaynaklanan kaybın engellenmesi mümkün değildir. Önemli olan geçmişe ait olaylara dair gizli bilgilerin keşfedilmesi, ileriye yönelik durumsal öngörüler veren modeller ile önceden tedbir alınmasını sağlayacak bir yönetim anlayışına geçmek ve olası kayıpları öngörebilmektir [2]. Bu yüzden büyük miktardaki verileri işleyebilen teknikleri kullanabilmek büyük önem kazanmaktadır.

Ham veriyi bilgiye veya anlamlı hale dönüştürme işlemleri veri madenciliği ile yapılabilmektedir [1]. Veri madenciliği, bu gibi durumlarda kullanılan büyük miktardaki veri kümelerinde saklı durumda bulunan örüntü ve eğilimleri keşfetme işlemidir [3]. Burada örüntü, ilgilenilen verilerdeki gözlenebilir veya ölçülebilir bilgiler anlamına gelmektedir. Diğer bir deyişle örüntü, olay veya nesnelerin düzenli

bir biçimde birbirini takip ederek gelişmesidir. Dolayısıyla veri madenciliğinin amacı, büyük veri yığınları içerisindeki anlam ifade etmeyen verileri, anlamlı olaylara ve sırası düzenlenmiş durumlara dönüştürebilmek ve öngörüler oluşturmak olarak düşünülebilir.

Kurumların veri ambarlarında tutulan verilerin içerisinde anlamlı bilgiler elde edilebilmesi için ilk başlarda bir çözüm yolu olan SQL dilinin, veri boyutlarının artmasıyla birlikte yetersiz kalması, Veri Tabanında Bilgi Keşfi (VTBK) adında yeni bir veri analizi yönteminin gelişmesinin nedenlerinden biri olmuştur. Veri madenciliği, bu analizin içinde bir aşamayı oluşturmaktadır. Ama son zamanlarda VTBK ile veri madenciliği eş anlamlı olarak da kullanılmaktadır [4]. VTBK süreci, çeşitli veri kaynaklarından verinin toplanması ile başlar. Bir sonraki adımda bu veriler analiz için uygun hale getirilir. Bu adımı veri madenciliği takip eder. Genel anlamda, veri madenciliği, ham verinin bilgiye dönüştürülmesidir. Veri madenciliği, çok sayıdaki verinin depolandığı veri tabanlarından elde edilen, modeller ve örüntüler, ilişkiler, sapmalar, anlamlı yapılar gibi daha önce keşfedilmemiş ilginç bilgilerin keşfedilmesi işlemidir. Veri tabanlarından bilgilerin keşfedilmesi süreci farklı birçok yaklaşımı kapsamaktadır. Bu yaklaşımlardan bazıları şunlardır:

- Kümeleme (Clustering),
- Veri özetleme (Data Summarization),
- Sınıflama kurallarının (Classification Rules) öğrenilmesi,
- Birliktelik Kuralları (Dependency Networks) bulunması,
- Değişikliklerin analizi (Analysing Changes) ve
- Anomali tespiti (Detecting Anomaly) 'dir.

Veri madenciliğinin temellerini yapay zeka ve istatistik disiplinlerindeki gelişmeler oluşturmaktadır. Veri madenciliği, çeşitli alanlarda kendine yer bulmuş durumdadır ve her geçen gün bu alanlara yenileri eklemektedir. Bu alanların çeşitliliğini göstermek için bazı örnekler verilebilir [4].

- Makine Öğrenimi
- Uzman Sistemleri

- Veri Tabanları
- Optimizasyon
- Görselleştirme
- Yüksek performanslı paralel işlemciler vb.

Günümüzde veri madenciliği işletmeler için çok önemli hale gelmiştir. Farklı alanlardaki büyük ölçekli veri tabanları, içlerinde değerli verileri bulunduran bir veri madeni gibi düşünülebilir. Bu büyüklükteki verilerin analizi, bu analiz sonucunda daha anlamlı bilgi elde etme ve elde edilen bilgiyi yorumlama işi, insan yeteneği ve ilişkisel veri tabanlarının yapabileceklerini aşmaktadır. Özellikle dijital veri miktarında artış patlaması, buna karşılık bu veriler üzerinde araştırma ve uygulama yapan kişilerin sayısının değişmemesi, çalışmalarını veri madenciliğine doğru zorlamıştır. Veri madenciliği tüm bu durumlara cevap olarak sunulmuş, içerisinde veriyi akıllı ve otomatikleşmiş şekilde analiz ederek işe yarar bilgiye dönüştürebilen yeni nesil teknikler bulunduran bir disiplin olarak ortaya çıkmıştır.

Bilgi ve iletişim teknolojilerinde yaşanan gelişmelere paralel olarak, bu alanlarda kullanılan yazılım teknolojileri de sürekli bir değişim ve gelişim halindedir. Bir zamanlar vazgeçilmez olan masaüstü uygulamaları neredeyse demode olmuş ve yerini çoktan web tabanlı uygulamalara bırakmıştır. Diğer bir taraftan insanların yeni gözdesi, zaman ve mekandan bağımsız olmak ve hareket özgürlüğü anlamına gelen “mobilite” olmuştur.

XG (2G-3G-4G vb.) ve WiMAX (Worldwide Interoperability for Microwave Access) gibi terimlerin hayatımıza girdiği şu günlerde, her alanda farklı bir mobil uygulama ile karşılaşılır olmuştur. Evlere gelen kuryeden, sipariş verilen garsonlara kadar pek çok saha çalışanı mobil cihazlar ve uygulamalar kullanarak işlerini yapar hale gelmişlerdir. Her geçen gün, bir bankanın daha mobil şube veya cep bankacılığı projesi hayata geçmekte veya bir GSM şebekesi yeni bir mobil uygulamanın tanıtımını yapmaktadır. Diğer taraftan web uygulamalarının da artık mobil versiyonları yazılmak zorunlu hale gelmiştir. Örneğin, wap.google.com,

m.yahoo.com, ya da m.facebook.com bunlardan bazılarıdır. Çünkü artık masaüstü bilgisayar teknolojilerinin yerini, cep bilgisayarları, tablet pc'ler ve akıllı telefonlar almaktadır.

Mobilitenin bu kadar yaygınlaştığı, ülkedeki insanların neredeyse yarısının cep telefonu ve mobil iletişim araçlarını kullandığı günümüz hayatında, bu araçların ürettiği dijital veriler de çok büyük boyutlara ulaşmaktadır. Mobilite hizmetlerini sunan firmalar, pazar paylarını arttırmak için sürekli olarak değerlendirmeler yapmalı, yenilikler ortaya koymalı ve gelişim göstermelidir. Bu değerlendirmeler sadece sunucu bazında değil kullanıcı bazında da olmalıdır. Bu noktada veri madenciliğinin, hayatımızda yerini alan bu mobil hizmet alanlarına uygulanması konusu ortaya çıkmaktadır. Kullanıcıların eriştikleri hizmet kalitesi, hizmet durumunun nasıl olduğunun analizi, geleceğe yönelik uygulamalar yapmak ve gelişim sağlamak için elde edilen verilerin değerlendirilmesi ve yorumlanması gerekmektedir. Bu değerlendirme ve yorumlama işlemi, veri madenciliği sayesinde gerçekleştirilebilmektedir.

Bu çalışmanın amacı; günümüzde yaygın bir şekilde kullanılan mobil iletişim alanında oluşturulan verilerin değerlendirilmesi, yorumlanması ve öngörü oluşturulmasını sağlayan veri madenciliğini irdelemektir. Veri madenciliği konusu tüm yönleriyle araştırılmıştır. Veri madenciliği disiplini, mobil iletişim alanına uygulamak için bir program geliştirilmiştir. Geliştirilen program sayesinde mobil iletişim aracı olan cep telefonundan sinyal gücü verileri elde edilmiştir. Bu verilerle GSM şebekelerinin belirlenen bir güzergahtaki çekim gücü performansları, iki aşamalı kümeleme ve frekans analizi yöntemleriyle değerlendirilmiş ve sonuçları yorumlanmıştır.

Bu tezin ikinci bölümünde, veri madenciliğinin tanımı, tarihi, kullanıldığı alanlar, yöntemleri, karşılaşılan problemleri, veri keşfi süreci gibi konulara ve daha önce gerçekleştirilen bazı çalışmalara değinilmiştir.

Üçüncü bölümde, mobil iletişim konusu işlenmiştir. Mobil cihazlar, işletim sistemleri, GSM sistemi, GSM şebekeleri ve mobil iletişimin tarihi konuları bu bölümde anlatılmıştır.

Dördüncü bölümde, mobil iletişimden veri elde etmek için geliştirilen program ve bu program ile GSM Şebekelerinden elde edilen sinyal gücü verilerinin değerlendirilmesi anlatılmıştır. Program sayesinde bilgisayara kaydedilen veriler, veri madenciliği uygulaması için hazır hale getirilmiştir. Verilere frekans analizi yapılmış ve sonuçları değerlendirilmiştir. Daha sonra iki aşamalı kümeleme yöntemi uygulanmış ve elde edilen sonuçlar yorumlanmıştır.

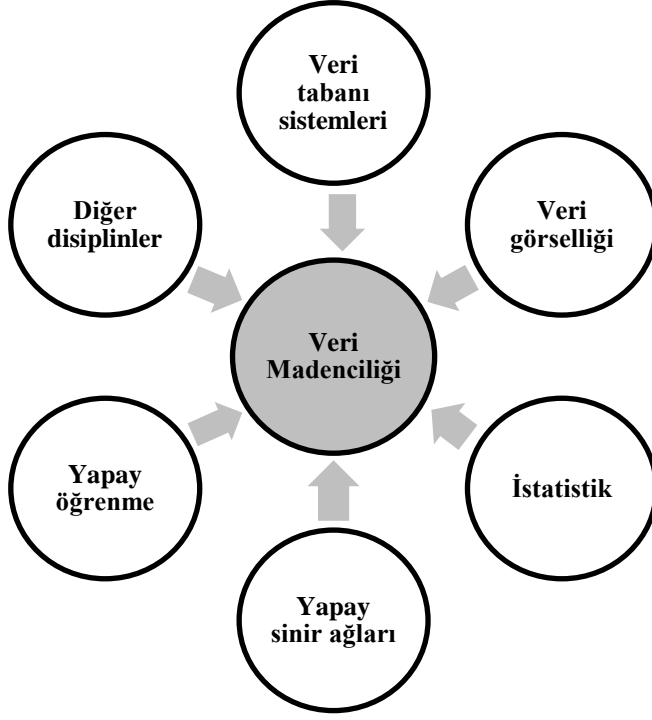
Beşinci bölümde ise, genel olarak tez çalışması içerisinde yapılan çalışmalarda ortaya çıkan bilgiler değerlendirilmiş ve sonuçlar tartışılmıştır. Elde edilen bulgular yorumlanmış, tespit edilen sınırlılıklar ve daha sonra yapılabilecek çalışmalara temel olabilecek durumlar belirtilmiştir.

2. VERİ MADENCİLİĞİ

Çok büyük miktardaki verilerin otomatik veya yarı otomatik bir biçimde çözümlenip kullanılabilir bilgi haline dönüştürülmesi işlemine literatürde pek çok terim karşılık gelmektedir. Bunlardan biri de veri madenciliğidir [2]. Veri madenciliği, veri yığınları arasında kalmış gizli bilgilere ulaşma işidir. Burada madencilik terimi kullanılmasının sebebi, çok büyük veri ortamlarında uygun olan verinin aranarak ortaya çıkarılma işlemine benzetilmesidir.

Günümüzde günlük ve tarihi veriler veri ambarlarında saklanmaktadır. Saklanan bu verilerin boyutları her geçen gün daha da artmaktadır. Bu sebeple gittikçe artan veri tabanlarını kullanarak karar vermekte zorluklar yaşanmaktadır. Ek olarak veri, birçok kaynaktan da gelmekte olabilir. Sonuçta planlamayı ve kurumun başka fonksiyonlarını desteklemek için veriyi analiz etmek gerekmektedir [3]. Veri madenciliği, büyük miktardaki verinin içinden geleceği tahmin edilmesinde yardımcı olacak anlamlı ve yararlı bağlantı ve kuralların bilgisayar programlarının aracılığıyla aranması ve analizidir [4].

Veri madenciliğinin uygulama alanı oldukça geniştir. Bu alanlar içerisinde, Veri Tabanı Sistemleri, Veri Görselliği, Yapay Sinir Ağları, İstatistik, Yapay Öğrenme, vb. gibi disiplinler bulunmaktadır.



Şekil 2.1. Veri madenciliği ve disiplinler

Veri madenciliği yerine pek çok terim kullanılmıştır. Bazıları veri madenciliğinin belli olmayan bilgileri çıkarma anlamına ve bazıları toplanan bilgilerden bir sonuç çıkarmak anlamına gelmekte olduğunu belirtmektedirler. Belirli bir tekniğin veri madenciliği tekniği olmasını fark etmek ve belirlemek zordur. Örneğin, bazıları istatistiğe ait analiz tekniklerini veri madenciliğinin tekniklerine ait olduğunu söyleyerek tartışılar diğerleri buna karşı çıkarlar. Örneğin, veri madenciliğiyle, tıp mallarını satan şirket reklamcılığında belli skorlara baskı yaparak kendi ürünlerinin satışını artırabilir ve bir kredi bürosu uygun ve ödemelerinde kusur olmayan adayları seçerek kendi kayıplarını sınırlayabilir. Veri madenciliği ayrıca anormal davranışı bulmak için kullanılabilir. Örneğin, bir casus ajansı bu teknolojiyi kullanarak onun çalışanları arasında anormal davranışı belirleyebilir [3].

Veri madenciliği araçları kullanılarak, işletmelerin daha etkin kararlar almasına yönelik karar destek sistemlerinde gerekli olan eğilimlerin ve davranış kalıplarının ortaya çıkarılması mümkün olmaktadır. Geçmişteki klasik karar destek sistemlerinin kullanıldığı araçlardan farklı olarak, veri madenciliğinde çok daha kapsamlı ve otomatikleştirilmiş analizler yapmaya yönelik birçok farklı özellik bulunmaktadır.

Veri madenciliğinin işletmelere sunduğu en önemli özellik, veri grupları arasındaki benzer eğilimlerin ve davranış kalıplarının belirlenmesidir. Bu fonksiyon özellikle hedef pazarlara yönelik pazarlama faaliyetlerinde yoğun olarak kullanılmaktadır [2].

Veri madenciliğinin başka bir özelliği ise daha önceden bilinmeyen bilgilerin ortaya çıkarılabilmesidir. Veri madenciliği sayesinde, veri ambarları içerisinde bulunan ancak ilk etapta görülemeyen bilgilerin ortaya çıkarılması mümkündür. Örneğin, bir firma sattığı ürünleri analiz ederek, ilerideki kampanyalarını şekillendirebilir ya da sattığı ürünler arasındaki bağları keşfedebilir. Burada amaç daha önceden fark edilmeyen veri kümelerinin bulunabilmesidir.

Veri tabanlarındaki veri kümelerinin olasılıklarını araştırmada veri madenciliğinde, istatistik alanındaki birçok yöntemin kullanılmasına rağmen nesnelere niteliksel değerlerine bağlı çıkarım yapmada bilinen istatistiksel yöntemlerden ayrılmaktadır [2]. Özellikle veri boyutlarının ulaştığı seviye göz önüne alındığında, istatistiksel yöntemleri kullanmanın veri madenciliği algoritmalarıyla karşılaştırıldığında daha zor olduğu söylenebilir.

Veri madenciliği aslında, bilgi teknolojilerinin doğal gelişim sürecinin sonucu olarak da değerlendirilebilir çünkü bu gelişim sürecinde, bilgisayar ağları ve internet kullanımının artmasıyla birlikte çoğalan veriler, veri tabanlarının da gelişimini hızlandırmış ve toplanan verilerin amaçlı bilgi seviyesine ulaştırılması zorlaşmıştır.

Günümüzün ekonomik koşulları ve yaşanan hızlı değişim ortamlarında, iş deneyimi ve önseziyle dayanarak alınan kararlarda, yanlış karar alma riski çok yüksektir. Riski azaltmanın tek yolu bilgiye dayalı yönetimi öngören karar destek çözümleridir. Veri madenciliği gereçleri gerçek anlamda bir karar destek sistemi oluşturmada olmazsa olmaz araçlardır. Bu noktada bilgi teknolojilerinden yararlanmak kaçınılmaz olmuştur.

Tüm bu bilgilerden yola çıkarak veri madenciliği için şöyle bir tanım yapmak mümkündür: Veri madenciliği, çok büyük miktarda bilginin depolandığı veri

tabanlarından, amacımız doğrultusunda, gelecek ile ilgili tahminler yapmamızı sağlayacak, anlamlı olan veriye ulaşma ve veriyi kullanma işidir.

2.1. Veri Madenciliği Tarihi

Günümüzde bilgi çok hızlı bir şekilde çoğalmaktadır ve bu bilgilerin içerisinde işe yarayacak ve yaramayacak olanları ayırt etmek zorlaşmaktadır. Bu bilgi yığınları arasından, işe yarayacak olanlarını elde etme işi için kullanılan yöntem ise veri madenciliğidir. Veri madenciliğinin tarihsel gelişimi Çizelge 2.1.'de belirtilmiştir.

Çizelge 2.1. Veri madenciliğinin tarihsel gelişimi [5]

Gelişim Adımları	Cevaplanan Karar Problemi	Kullanılabilen Teknolojiler	Ürün Sağlayıcıları	Karakteristikler
Veri Toplama (1960'lar)	“Benim toplam karım geçen 5 yılda ne kadardı?”	Bilgisayarlar, Teypler, Diskler	IBM, CDC	Geriye dönük, statik veri dağıtımı
Veri Erişimi (1980'ler)	“İngiltere’de geçen mart ayında birim satışları ne kadardı?”	İlişkisel Veri tabanları, SQL, OBDC	Oracle, Sybase, Informix, IBM, Microsoft	Kayıt düzeyinde geriye dönük, dinamik veri dağıtımı
Veri Ambarlama ve Karar Destek Sistemleri (1990'lar)	“İngiltere’de geçen mart ayında birim satışları ne kadardı? Boston’a aşağı inebilecek?”	OLAP, Çok Boyutlu Veri tabanı Sistemleri, Veri ambarları	Pilot, Comshare, Arbor, Cognos, Microstrategy	Çoklu düzeylerde, geriye dönük dinamik veri dağıtımı
Veri Madenciliği (Bugün)	“Gelecek ay Boston’daki birim satışlar muhtemelen ne olabilir? Neden?”	İleri düzeyde algoritmalar, çok işlemcili bilgisayarlar, büyük veri tabanları	Pilot, Lockheed, IBM, SGI, SPSS, SAS, Microsoft, vs.	Geleceğe dönük, proaktif enformasyon dağıtımı

Geçmişten günümüze veriler her zaman yorumlanmış, bilgi edinilmek istenmiş ve bunun için de donanımlar oluşturulmuştur. Bu sayede bilgi geçmişten günümüze taşınır hale gelmiştir.

1950’li yıllarda ilk bilgisayarlar sayımlar için kullanılmaya başlanmıştır.

1960’larda ise veri tabanı ve verilerin depolanması kavramı teknoloji dünyasında yerini almıştır. Ayrıca 1960’ların sonunda bilim adamları basit öğrenmeli bilgisayarlar geliştirebilmişlerdir.

1970’lerde ilişkisel Veri Tabanı Yönetim Sistemleri uygulamaları kullanılmaya başlanmıştır. Bilgisayar uzmanları bununla beraber basit kurallara dayanan uzman sistemler geliştirmişler ve basit anlamda makine öğrenimini sağlamışlardır.

1980’lerde Veri Tabanı Yönetim Sistemleri yaygınlaşmış ve bilimsel alanlarda, mühendisliklerde vb. alanlarda uygulanmaya başlanmıştır. Bu yıllarda şirketler müşterileri, rakipleri ve ürünleri ile ilgili verilerden oluşan veri tabanları oluşturmuşlardır. Bu veri tabanlarının içerisinde çok büyük miktarlarda veri bulunmakta ve bunlara SQL veri tabanı sorgulama dili ya da benzeri diller kullanılarak ulaşılabilmektedir.

1990’larda içindeki veri miktarı katlanarak artan veri tabanlarından, faydalı bilgilerin nasıl bulunabileceği düşünölmeye başlanmıştır. 1989, KDD (IJCAI)-89 Veri Tabanlarında Bilgi Keşfi Çalışma Grubu toplantısı ve 1991, KDD (IJCAI)-89’un sonuç bildirgesi sayılabilecek “Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Workshop” makalesinin KDD ile ilgili temel tanım ve kavramları ortaya koyması ile süreç daha da hızlanmıştır. 1992 yılında veri madenciliği için ilk yazılım gerçekleştirilmiştir.

2000’li yıllarda veri madenciliği sürekli gelişmiş ve hemen hemen tüm alanlara uygulanmaya başlanmıştır. Alınan sonuçların faydaları göröldükçe bu alana yönelik çalışmalar ve uygulamalar sürekli artmaktadır.

2.2. Veri Madenciliğinin Kullanıldığı Alanlar

Büyük hacimde veri bulunan her yerde veri madenciliğini kullanmak mümkündür. Günümüzde karar verme sürecine ihtiyaç duyulan birçok alanda veri madenciliği uygulamaları yaygın olarak kullanılmaktadır. Örneğin, biyoloji, finans, borsa,

genetik, sađlık, sigorta, endüstri, istihbarat vb. birçok dalda başarılı uygulamaları görölmektedir.

Son 20 yıldır Amerika Birleşik Devletleri'nde çeşitli veri madenciliđi algoritmalarının gizli dinlemelerden, vergi kaçakçılıklarının ortaya çıkartılmasına kadar çeşitli uygulamalarda kullanıldığı bilinmektedir.

Veri madenciliđini kullanan alanlar ve kullandıkları uygulamalar şöyle belirtilebilir:

Pazarlama

- Müşterilerin satın alma örüntülerinin belirlenmesi,
- Müşterilerin demografik özellikleri arasındaki bağlantıların bulunması,
- Posta kampanyalarında cevap verme oranının artırılması,
- Mevcut müşterilerin elde tutulması, yeni müşterilerin kazanılması,
- Pazar sepeti analizi,
- Müşteri ilişkileri yönetimi,
- Müşteri değerlendirme,
- Satış tahmini,
- Müşteri dağılımı,
- Çeşitli pazarlama kampanyaları,
- Pazarlama stratejilerinin oluşturulması,
- Çapraz satış analizleri [2].

Bankacılık, Sigortacılık ve Borsa

- Farklı finansal göstergeler arasında gizli korelasyonların bulunması,
- Kredi kartı dolandırıcılıklarının tespiti,
- Kredi kartı harcamalarına göre müşteri gruplarının belirlenmesi,
- Kredi taleplerinin değerlendirilmesi,
- Müşteri dağılımı,
- Usulsüzlük tespiti,
- Risk analizleri,
- Risk yönetimi,

- Yeni poliçe talep edecek müşterilerin tahmin edilmesi,
- Sigorta dolandırıcılıklarının tespiti,
- Hisse senedi fiyat tahmini,
- Hileli hesapların ve dolandırıcılıkların ortaya çıkarılması [6].

Perakendecilik

- Satış noktası veri analizleri,
- Alışveriş sepeti analizleri,
- Tedarik ve mağaza yerleşim optimizasyonu,
- Genel piyasa analizleri,
- Alım-satım stratejilerinin optimizasyonu.

Telekomünikasyon

- Kalite ve iyileştirme analizleri,
- Hatların yoğunluk tahminleri,
- Web sitesi ziyaretçilerinin profil analizi.

Bilim ve Mühendislik

- Laboratuvar veya bilgisayar ortamında sistemlerin simülasyonu ve analizi sürecinde üretilen yüksek miktarda bilimsel verilerin anlamlandırılması [7].

Sağlık, Biyoloji ve Genetik

- Gen araştırmaları,
- Yeni virüs türlerinin keşfi ve sınıflandırılması,
- Hastalıkların özelliklerinin belirlenerek teşhislerin kolaylaştırılması,
- Birlikte kullanılan ilaçların yan etkilerinin araştırılması,
- Test sonuçlarının tahmini,
- Ürün geliştirme,
- Tedavi sürecinin belirlenmesi,
- Tıbbi teşhis.

Kriminoloji

- Suç eğilimlerinin tanımlanması,
- İstihbarat birimleri.

Veri madenciliğinin kullanıldığı yukarıda belirtilen alanlarda araştırmacılara, firma yöneticilerine ve idarecilerine sağladığı yararlar şu şekilde sıralanabilir [6]:

- Tıp, verilerin en çok tutulduğu alanların başında bulunmaktadır. Özellikle son yıllarda genetiğin inanılmaz hızda ilerlemesi sonucunda oluşan gen haritaları ile hastalıklar sınıflandırılmaya başlanmıştır. Hangi genlere sahip bireylerin hangi hastalıklara yakalanma olasılığı olduğuna dair çalışmaları gerçekleştirmek artık mümkündür.
- Gen bilimi ile ortaklaşa çalışma sayesinde hangi bireylerin suç işlemeye eğilimli olduklarına dair tahminleme çalışmaları yapılarak bu olayların daha ortaya çıkmadan önlenmesinden, kullanıcıların yazım karakterlerinden yola çıkarak birçok olasılığın hesaplanmasına kadar pek çok kriminoloji bilgisi elde edilebilir.
- Simülasyon ortamlarındaki verileri kullanarak mühendislik alanında, üretim veya problem çözme konularında öngörüler oluşturulabilir ve çözümler üretilebilir.
- Veri madenciliği algoritmalarıyla yapılan değerlendirmeler, çok uzun zamanlar alabilecek deney ve test olaylarına gerek kalmadan, güvenilir ve daha kısa zaman alan sonuçlar elde etme imkanı sağlayabilir.
- Bankacılık faaliyetlerinde, küçük işletmelere yönelik olarak makine ve donanım satışı yapan dağıtıcı firmalarla ortak hareket ederek oluşturulacak satış paketleri ile pazarlama stratejileri geliştirilebilir.
- Mevcut müşterilerin yöneticiler tarafından daha iyi tanınmasını sağlayabilir.
- Özellikle finans sektöründe mevcut müşteriler bölümlere ayrılıp, kredi risk davranış modelleri oluşturularak, yeni başvuruda bulunan müşterilere karşı riskin minimize edilmesi sağlanabilir.
- Mevcut müşterilerin ödeme performansları incelenip kötü ödeme performansı gösteren müşterilerin ortak özellikleri belirlenerek, benzer özelliklere sahip tüm müşteriler için yeni risk yönetim politikaları oluşturulabilir.

- En iyi müşteriler veya müşteri bölümlerinin bulunmasında kullanılabilir. Bulunan bu iyi müşteri bölümlerine yönelik yeni pazarlama stratejileri oluşturulabilir.
- Kuruluşlar tarafından düzenlenecek çeşitli kampanyalarda mevcut müşteri kitlesinin seçimi ve bu müşterilerin davranış özelliklerine yönelik kampanya şartlarının oluşturulması sağlanabilir.
- Mevcut müşteriler üzerinde firma ürünlerinin çapraz satış kapasitesinin artırılması sağlanabilir.
- Veri madenciliği ile mevcut müşteriye tanıyarak kuruluşların müşteri ilişkileri yönetimlerinde düzenleme ve geliştirmeler yapılabilir. Bu sayede kuruluşun müşterilerini daha iyi tanıyarak müşteri gibi düşünme kapasitelerinin artırılması sağlanabilir.
- Günümüzde var olan yoğun rekabet ortamında kuruluşların hızlı ve kendisi için en doğru kararı almalarını sağlayabilir.
- Kuruluşlar veri analizi ile müşterilerini kişiselleştirilmiş ürün ve hizmetler hakkında bilgilendirebilirler.
- Veri madenciliği ile kuruluşların müşteriyle bütünleşmiş satış politikaları oluşturması sağlanabilir.
- Telekomünikasyon sektöründe kişilerin kullanım sıklıkları, amaçları ve hat yoğunluk tahminleri yapılarak firmalar altyapı güncellemelerine gidebilir, müşteriye ilişkin, müşteriye özel kampanyalar düzenleyebilirler.

Ülkemizde en çok kullanıldığı alanlar bankacılık, sigortacılık ve borsa olarak sıralanabilir. Kaynaklar incelendiğinde ise veri madenciliğinin en çok kullanıldığı alanlar olarak tıp, biyoloji ve genetik olarak görülmektedir.

2.3. Daha Önce Gerçekleştirilen Veri Madenciliği Uygulamaları

Pek çok alanda yaygın bir şekilde kullanılmaya başlanan veri madenciliği, günümüzün en çok kullanılan uygulama alanlarından birisi olmuştur. Her geçen sene kendisine daha fazla kullanım alanı bulmakla birlikte, kolay uygulanabilirliği ve

etkili sonuçlar ortaya çıkarması sayesinde, firma yöneticileri tarafından en çok başvurulan yöntemlerden bir tanesi olmuştur. Literatür taramasıyla ulaşılan veri madenciliği ile gerçekleştirilmiş bazı uygulamaları şu şekilde özetleyebiliriz.

Veri madenciliğinde kullanılan teknikler ve uygulaması

Elif Özge Özdamar (2002) tarafından gerçekleştirilen bu çalışmada, veri madenciliğinde kullanılan teknikler açıklanmış ve uygulamaları gerçekleştirilmiştir. Ayrıca bu çalışmada IBM intelligent Miner programı ile satışlara yönelik veri madenciliği analizleri gerçekleştirilmiştir [8].

Öğrenci işleri otomasyonu üzerinde veri madenciliği çalışması

Konya Selçuk Üniversitesi'nde Onur İnan (2003) tarafından, hazırlık sınıfı, birinci sınıf ve mezun durumunda olan öğrenciler üzerinde, üniversite veri tabanındaki veriler kullanılarak veri madenciliği uygulaması gerçekleştirilmiştir. Bu uygulama sayesinde, öğrencilerin başarılarını etkileyen etmenler, başarı düzeyleri, üniversiteyi kazanan öğrenci portföyleri ve mezun olamayan öğrencilerin okulu bitirmelerini etkileyen etmenler üzerinde çalışmalar gerçekleştirilmiş ve sonuçları yorumlanmıştır [2].

Mobil sistemlerde veri madenciliği kullanılarak kullanıcı hareketlerinin tahmini

Gökhan Yavaş (2003) tarafından gerçekleştirilen bu çalışmada, mobil kullanıcıların hareket modellerinin veri madenciliği kullanılarak çıkarılması ve bu modeller kullanılarak mobil kullanıcıların daha sonraki hareketlerinin tahmin edilmesi için yeni bir algoritma geliştirilmiştir. Üç aşamadan oluşan bu algoritmanın ilk aşamasında kullanıcı hareket modelleri, kullanıcıların önceden kaydedilmiş mobil yörüngelerinden veri madenciliği kullanılarak çıkarılmaktadır. İkinci aşamada bulunan hareket modellerinden hareket kuralları üretilmekte, son aşamada ise bu hareket kuralları kullanıcının bir sonraki hücreler arası hareketinin tahmini için kullanılmaktadır. Sunulan algoritmanın performansı simülasyonlar yardımıyla iki farklı tahmin yöntemiyle karşılaştırılmıştır. Performans sonuçları algoritmanın diğer yöntemlerden daha doğru tahminler yapabildiğini göstermektedir [9].

Veri madenciliği ve bir e-ticaret uygulaması

Anarberk Kalikov (2006) tarafından, bir yayınevi firmasının internet sitesindeki veriler dikkate alınarak, veri madenciliği birliktelik kuralları tekniği ile sepet ve sipariş tabloları incelenmiştir. Hangi ürünlerin kategorisinin değiştirilmesi gerektiği, kullanıcıların meslek ve ilgi alanı dağılımları, müşteri ilgi alanlarına göre satış grafikleri ve kullanıcıların ödeme seçenekleri ile ilgili bir veri madenciliği uygulaması gerçekleştirilmiştir [1].

Uzaktan eğitimde öğrencilerin ders çalışma etkinliklerinin log verileri analizi ile incelenmesi

Serdar Çiftci (2006) tarafından gerçekleştirilen çalışmada, uzaktan eğitime katılan öğrencilerin ders çalışma etkinliklerinin nasıl değerlendirilebileceğinin ortaya konulması amaçlanmıştır. Ayrıca yapılan anketler ve log dosyaları karşılaştırılarak, sonuçların farklı olup olmadıkları incelenmiştir. Bu çalışmada ayrıca, öğrencilerin öğretim materyali ile olan etkileşimine ait bilgilerinin tutulduğu veri tabanlarındaki log kayıtları analiz edilmiştir [10].

Veri madenciliği teknikleri ile bir kozmetik markanın ayrılan müşteri analizi ve müşteri bölümlenmesi

Sinem Akbulut (2006) tarafından yapılan bir çalışmada, bir kozmetik markasının müşteri grupları ve ayrılma eğilimi gösteren müşteri kesiti belirlenerek; bu müşterilere özel pazarlama stratejileri geliştirilmesi hedeflenmektedir. Bölümlenme için kümeleme teknikleri, ayrılacak müşteri kesitini belirlemek için sınıflama teknikleri kullanılmıştır [11].

Dağıtık zaman-mekan verilerinde gizliliği koruyan veri madenciliği

Ali İnan (2006) tarafından yapılan bu çalışmada zaman-mekan nitelikleri olan veriler için bir gizliliği koruyan veri madenciliği tekniği ve iki ön-işleme tekniği önerilmiştir: (1) Dağıtık kümeleme, (2) Merkezi anonimleştirme ve (3) Dağıtık anonimleştirme. Önerilen tekniklerin güvenlik ve performans analizleri de yapılmış ve sonuçta mantıklı varsayımlar altında minimum mahrem bilgi kaybıyla veri madenciliğinin mümkün olduğu gözlemlenmiştir [12].

Asenkron motorlarda veri madenciliği ile hata tespiti

Kıyas Kayaalp (2007) tarafından yapılan bu çalışmada, veri madenciliği tekniği ile üç fazlı asenkron motordaki sargı spirleri arasında oluşabilecek kısa devre veya yalıtım bozuklukları ve motor milinde oluşabilecek mekanik dengesizlik hatalarının tespiti amaçlanmıştır [13].

Lise türü ve lise mezuniyet başarısının, kazanılan fakülte ile ilişkisinin veri madenciliği ile analizi

Y. Ziya Ayık, Abdülkadir Özdemir ve Uğur Yavuz (2007) tarafından yapılan bu çalışmada, Atatürk Üniversitesi öğrencilerinin mezun oldukları lise türleri ve lise mezuniyet dereceleri ile kazandıkları fakülteler arasındaki ilişki, veri madenciliği teknikleri kullanılarak incelenmiştir. Elde edilen sonuçların, üniversiteyi sonraki yıllarda tercih edecek öğrenci profilinin belirlenmesine yardımcı olması amaçlanmıştır [14].

Veri madenciliği ile deprem verilerinin analizi

N. Duru ve M. Canbay (2007) tarafından gerçekleştirilen bu çalışma, deprem verileri kullanılarak seçilen bir bölgeye ait sismik tehlikenin diğer deyişle gerçekleşme olasılığının veri madenciliği yönünden ele alınarak incelenmesini kapsamaktadır. Çalışma sonuçları Jeofizik sonuçlar ile korele edilerek doğruluk payı da araştırılmıştır. Gelecek her 10 yıl için % sismik tehlike değeri artış göstererek devam etmiştir. Örneğin, 6 magnitüdündeki bir depremin olma olasılığı 10 yıl içinde %27 iken, 30 yıl içinde %60 ve 60 yıl için de %80'leri bulmaktadır [15].

Bu değerler daha önce çalışma bölgesinde yapılmış çalışmalarla uyum göstermektedir. Ancak burada unutulmaması gereken bu çalışmanın deprem tahmini için kullanılan tekniklerden sadece birisi olduğu ve bu çalışmanın konusu itibarıyla çalışma bölgelerinin tektonik özelliklerini hiç irdilemeden dahi olsa olumlu sonuçlara varılabilmesinin mümkün olduğunun gösterilebilmesidir. Ayrıca yapılan çalışmanın sonuçlarının büyük bölgelere göre küçük bölgelerde daha iyi sonuç verdiğinin görülmesidir. Uygulama, dünya ölçeğindeki her noktanın analizini yapacak şekilde geliştirilmiş olup, ihtiyaç halinde programa eklemeler yapmak

suretiyle başka bu tür çalışmalar yapacak şekilde tasarlanmıştır. Sonuç olarak, bu çalışma türünden bazı istatistik yöntemlerin de kullanılabilirliği ve faydalı olabileceği ve veri miktarının önemli bir kriter olduğu, ne kadar çok veri üzerinde çalışılırsa o derece başarılı sonuçlar alınmasının mümkün olacağı söylenebilir [15].

Birliktelik kuralı yöntemi için bir veri madenciliği yazılımı tasarımı ve uygulaması

Feridun Cemal Özçakır ve A. Yılmaz Çamurcu (2007) tarafından gerçekleştirilen bu çalışmada, bir firmanın pastane satış verileri üzerinde veri madenciliği uygulamak için birliktelik kuralları ile bir yazılım tasarlanmıştır. Tasarlanan yazılımda Apriori algoritması kullanılmıştır. Uygulanan Apriori algoritması ile farklı zaman dilimi, farklı satış lokasyonu girdi değerleri doğrultusunda birlikte satın alınan ürünler ile ilgili bağıntılar olduğu gözlemlenmiştir. Genelde aynı ürün grubuna ait ürünlerin, en sık birlikte satın alınan ürünler olduğu görülmüştür [16].

Veri madenciliğinde yaşam çözümlemesi: Kredi kartı sahipleri ile ilgili bir uygulama

Bu çalışmada Nihal Ata, Erengül Özkök ve Uğur Karabey (2007), yaşam çözümlemesi yöntemlerini veri madenciliği konusu çerçevesinde ele aldıktan sonra kredi kartı sahiplerine ait bir veri kümesi için yaşam olasılıkları, hazard olasılıkları ve regresyon modellerini incelemişlerdir. Uygulamada öncelikle yaşam olasılıkları 5'er yıllık 3 dönem (5, 10, 15) için elde edilmiştir. Daha sonra, yaşam ve hazard olasılıklarına ait grafikler verilmiş ve müşteri kaybı açısından yorumlanmıştır. Müşterilerin kredi kartını kullanmayı bırakmasını etkileyen risk faktörleri ise regresyon modelleri ile belirlenmeye çalışılmıştır. Weibull regresyon modelinin veri kümesi için en uygun regresyon modeli olduğu sonucuna ulaşılmıştır. Buna göre çalışmada yaş, gelir ve medeni durumun, müşterilerin kredi kartı kullanmayı bırakmalarını etkileyen önemli risk faktörleri olduğu görülmüştür [17].

KxKNN: k-means ve k en yakın komşu yöntemleri ile ağlarda nüfuz tespiti

Bu çalışmada Sibel Kırmızıgül Çalışkan ve İbrahim Soğukpınar (2008), veri madenciliği yöntemlerinden “K-means” ve “K en yakın komşu” yöntemlerinin iyileştirilmesi amacıyla; nüfuz tespiti için kümelemeyi ve sınıflandırmayı, denetimli ve denetimsiz öğrenimi, k-means ve k en yakın komşu yöntemlerini bir arada

kullanan hibrit bir yapı geliştirmişlerdir. Geliştirilen uygulamada en hızlı sonucu veren K-means uygulaması ile test kümesi daha küçük alt kümelere ayrılarak, k en yakın komşu yönteminin zaman karmaşası ve bellek gereksinimi azaltılmıştır [18].

Türkiye’de bir havayolu işletmesine ait parça söküm raporlarına ilişkin veri madenciliği uygulaması

Feyza Gürbüz, Lale Özbakır ve Hüseyin Yapıcı (2008) tarafından gerçekleştirilen bu çalışmada, Türkiye’de bir hava yolu işletmesinin parça söküm raporları üzerinde veri madenciliği çalışması gerçekleştirilmiştir. Çalışmanın amacı, uçaklarda kullanılan parçaların, herhangi bir arıza oluşmadan önce düzeltici ve önleyici işlemlerin yapılması için ikaz seviyelerinin tespit edilmesine yönelik kural geliştirmektir. Sonuç olarak parçaların ikaz seviyelerini temsil edecek anlamlı bir kural elde edilmiş ve bulunan kurallar doğrulukları ve güvenilirlikleri bakımından test edilmiştir [19].

Karar ağacı algoritmaları ve İMKB verileri üzerine bir uygulama

Ali Sait Albayrak ve Şebnem Koltan Yılmaz (2009) tarafından gerçekleştirilen bu çalışmada, İMKB 100 endeksinde sanayi ve hizmet sektörlerinde faaliyet gösteren 173 işletmenin 2004–2006 yıllarına ait yıllık finansal göstergelerinden yararlanarak veri madenciliği tekniklerinden birisi olan karar ağaçları tekniği uygulanmıştır. Seçilen finansal göstergelere göre sanayi ve hizmet sektörlerinde faaliyet gösteren firmaları ayıran en önemli değişkenler saptanmıştır [20].

Türkiye’de yerli ve yabancı ticaret bankalarının finansal etkinliğe göre sınıflandırılması: Karar ağacı, lojistik regresyon ve diskriminant analizi modellerinin bir karşılaştırması

Bu çalışmada Ali Sait Albayrak (2009), yerli ve yabancı olarak önceden grup üyeliği belirlenmiş bankaların sınıflandırmasında yaygın olarak kullanılan veri madenciliği tekniklerinden diskriminant, lojistik regresyon ve karar ağacı modellerini karşılaştırmıştır. Üç sınıflandırma tekniği, bankalarla ilgili seçilmiş likidite, gelir-gider, karlılık ve faaliyet oranları kullanılarak karşılaştırılmaktadır. Araştırmanın sonuçları, bankaların sınıflandırmasında karar ağacı modelinin geleneksel diskriminant ve lojistik regresyon modellerine üstünlük sağlayarak alternatif etkili bir sınıflandırma tekniği olarak kullanılabileceğini göstermektedir [21].

Öğrenci seçme sınavında (ÖSS) öğrenci başarısını etkileyen faktörlerin veri madenciliği yöntemleriyle tespiti

Ahmet Selman Bozkır, Ebru Sezer ve Bilge Gök (2009) tarafından gerçekleştirilen bu çalışmada, ÖSYM tarafından 2008 ÖSS adayları için resmi internet sitesi üzerinden yapılan anket verileri üzerinde veri madenciliği yöntemleri kullanılarak, öğrencilerin başarılarını etkileyen faktörler araştırılmıştır [22].

Araştırma kapsamında, Ortaöğretim Başarı Puanı (OBP) tüm öğrencilerin ÖSS’de ortak başarı ölçütü olarak ele alınırsa bu puan türüne etkiyen en önemli beş faktörün yaş, okul türü, sanat derslerine ilgi, matematik dersi için harcanan ödev zamanı ve fen bilgisi laboratuvar kullanım oranı olduğu görülmüştür. Diğer tüm puan türlerinde de yaşın önemli bir etken olduğu görülmüştür. Ayrıca okullarda sunulan teknik imkânların ÖSS başarısı üzerinde önemli etkileri tespit edilmiştir. Özellikle kümeleme analizinde görüldüğü üzere öğrencinin sahip olduğu sosyal, kültürel ve ekonomik imkânların ÖSS başarısına büyük katkısı bulunduğu tespit edilmiştir. Sanat derslerine gösterilen ilginin OBP üzerindeki dikkat çekici etkisi ve dershaneye gitme oranının başarı üzerindeki etkisinin beklenenden düşük çıkması araştırmanın sonucunda elde edilen ilginç bulgulardan birkaçı olmuştur. Bu çalışmada, veri madenciliği yöntemlerinden karar ağaçları ve kümeleme kullanılmıştır [22].

Hileli finansal tabloların tespitinde veri madenciliği tekniklerinin kullanımı: İmalat firmaları üzerine bir uygulama

H. Ali Ata ve İbrahim H. Seyrek (2009) tarafından gerçekleştirilen bu çalışmada, denetçiler tarafından yaygın olarak bilinmeyen bazı veri madenciliği teknikleri, finansal tablolardaki hileleri tespit etmeye yardımcı olmak üzere kullanılmıştır. Çalışma İMKB’de işlem gören ve imalat sektöründe faaliyet gösteren 100 firmanın bilgilerine dayalı olarak gerçekleştirilmiştir. Araştırma sonucunda kaldırma oranı ve aktif karlılık oranının finansal tablo hilesini tespit etmede önemli finansal oranlar olduğu belirlenmiştir [23].

2.4. Veri Madenciliğini Etkileyen Etmenler

Veri madenciliği temel olarak 5 ana faktörden etkilenir [4]:

1. Veri: Veri madenciliğinin bu kadar gelişmesindeki en önemli faktördür. Son yirmi yılda sayısal verinin hızla artması, veri madenciliğindeki gelişmeleri hızlandırmıştır. Bir taraftan veri tabanlarında hızla artmakta olan veriler bulunmakta iken diğer taraftan bu verilerle uğraşan bilim adamları, mühendisler ve istatistikçilerin sayısı aynı kalmıştır. Bu yüzden, verileri analiz etme yöntemleri ve teknikleri otomatikleştirilmiş olarak geliştirilme ihtiyacı doğmuştur.

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

3. Bilgisayar ağları: Yeni nesil internet, çok yüksek hızlarda bilgiye erişime izin vermektedir. Böyle bir bilgisayar ağı ortamında, dağıtık verileri analiz etmek ve farklı algoritmaları kullanmak mümkün olmaktadır. Bundan 10 yıl önceki bilgisayar ağları teknolojisinde hayal edilemeyenler artık yapılabilmekte ve buna bağlı olarak, veri madenciliğine uygun ağların tasarımı da oluşturulabilmektedir.

4. Bilimsel hesaplamalar: Bilim adamları ve mühendisler, simülasyonu, bilimin üçüncü yolu olarak görmektedirler. Veri madenciliği ve bilgi keşfi; teori, deney ve simülasyonu birbirine bağlamada önemli bir rol almaktadır.

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

2.5. Veri Madenciliğinde Karşılaşılan Problemler

Büyük hacimli verilerin bulunduğu veri ortamlarında büyük sorunlar ortaya çıkabilir. Bu sebeple küçük veri kümelerinde, benzetim ortamlarında hazırlanmış veri madenciliği sistemleri, büyük hacimli, eksik, gürültülü, boş, atık, aykırı veya belirsiz veri kümelerinin bulunduğu ortamlarda yanlış çalışabilir. Bu nedenle veri madenciliği sistemleri hazırlanırken bu sorunların çözülmesi gerekmektedir. Veri madenciliği uygulamalarında karşılaşılabilecek sorunlar aşağıda belirtilmiştir.

2.5.1. Artık veri

Artık veri, problemde istenilen sonucu elde etmek için kullanılan örneklem kümesindeki gereksiz niteliklerdir. Problemde belirtilen veri kümesi, eldeki probleme uygun olmayan veya artık nitelikler içerebilir. Bu durum pek çok işlem sırasında karşımıza çıkabilir. Örneğin, eldeki problem ile ilgili veriyi elde etmek için iki ilişkiyi birleştirecek elde edilen ilişkide kullanıcının farkında olmadığı artık nitelikler bulunur. Artık nitelikleri elemek için geliştirilmiş algoritmalar özellik seçimi olarak adlandırılır. Özellik seçimi, tümevarıma dayalı öğrenmede budama öncesi yapılan işlem, hedef bağlamı tanımlamak için yeterli ve gerekli olan niteliklerin küçük bir alt kümesinin seçimi problemidir. Özellik seçimi yalnız arama uzayını küçültmekle kalmayıp, sınıflama işleminin kalitesini de artırır [2].

2.5.2. Belirsizlik

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

2.5.3. Boş veri

Bir veri tabanında boş değer, birincil anahtarlar yer almayan herhangi bir niteliğin değeri olabilir. Boş değer, tanımlı gereği kendisi de dahil olmak üzere hiçbir değere eşit olmayan değerdir. Bir çokluda eğer bir nitelik değeri boş ise o nitelik bilinmeyen ve uygulanamaz bir değere sahiptir. Bu durum ilişkisel veri tabanlarında sıkça

karşımıza çıkmaktadır. Bir ilişkide yer alan tüm çoklular, niteliğin değeri boş olsa bile aynı sayıda niteliğe sahip olmalıdır. Örneğin, kişisel bilgisayarların özelliklerini tutan bir ilişkide bazı model bilgisayarlar için ses kartı modeli niteliğinin değeri boş olabilir [24].

2.5.4. Dinamik veri

Kurumsal çevrim içi veri tabanları dinamiktir ve içeriği sürekli olarak değişir. Bu durum bilgi keşfi yöntemleri için önemli sakıncalar doğurmaktadır. İlk olarak sadece okuma yapan ve uzun süre çalışan bilgi keşfi yöntemi, bir veri tabanı uygulaması olarak mevcut veri tabanı ile birlikte çalıştırıldığında mevcut uygulamanın da performansı ciddi ölçüde düşer. Diğer bir sakınca ise, veri tabanında bulunan verilerin kalıcı olduğu varsayıp, çevrim dışı veri üzerinde bilgi keşif yöntemi çalıştırıldığında, değişen verinin elde edilen örüntülere yansımaları gerekmektedir. Bu işlem, bilgi keşfi yönteminin ürettiği örüntüleri zaman içinde değişen veriye göre sadece ilgili örüntüleri yığmalı olarak günleme yeteneğine sahip olmasını gerektirir [24].

2.5.5. Eksik veri

Evrendeki her nesnenin ayrıntılı bir biçimde tanımlandığı ve bu nesnelerin alabileceği değerler kümesinin belirli olduğu varsayalım. Verilen bir bağlamda her bir nesnenin tanımı kesin ve yeterli olsa idi sınıflama işlemi basitçe nesnelerin alt kümelerinden faydalanılarak yapılabilir. Bununla birlikte, veriler kurum ihtiyaçları göz önünde bulundurularak düzenlenip toplandığından, mevcut veri bilgi keşfi açısından uygun olmayabilir [24].

Örneğin, otomobil arızalarının çözümlenmesi için tek bir otomobil markasının verileri kullanılarak hazırlanan bir veri madenciliği uygulaması, diğer marka otomobiller için doğru çözümler üretemezdi. Bu gibi koşullarda bilgi keşfi modeli belirli bir güvenlik (veya doğruluk) derecesinde tahmini kararlar alabilmelidir.

Eksik veriler veri kümesinin büyüklüğünden ya da doğasından kaynaklanmaktadır. Eksik veriler olduğunda yapılması gerekenler şunlardır:

- Eksik veri içeren kayıt veya kayıtlar çıkarılabilir.
- Değişkenin ortalaması eksik verilerin yerine kullanılabilir.
- Var olan verilere dayalı olarak en uygun değer kullanılabilir.

Eksik veriler, yapılacak olan istatistiksel analizlerde önemli problemler yaratmaktadır. Çünkü istatistiksel analizler ve bu analizlerin yapılmasına imkan veren ilgili paket programlar, verilerin tümünün var olduğu durumlar için geliştirilmiştir. Bu analizler, eksik veri içeren veri kümelerine uygulandıklarında istatistiklerin geçerliliğini düşürmektedir [25].

2.5.6. Farklı tipteki verileri ele alma

Gerçek hayattaki uygulamalar makine öğreniminde olduğu gibi yalnızca sembolik veya kategorik veri türleri değil, aynı zamanda tamsayı, kesirli sayılar, çoklu ortam verisi, coğrafi bilgi içeren veri gibi farklı tipteki veriler üzerinde işlem yapılmasını gerektirir. Kullanılan verinin saklandığı ortam, düz bir kütük veya ilişkisel veri tabanında yer alan tablolar olacağı gibi, nesneye yönelik veri tabanları, çoklu ortam veri tabanları, coğrafik veri tabanları vb. olabilir. Saklandığı ortama göre veri, basit tipte olabileceği gibi karmaşık veri tipleri (çoklu ortam verisi, zaman içeren veri, yardımcı metin, coğrafi, vb.) de olabilir. Bununla birlikte veri tipi çeşitliliğinin fazla olması bir veri madenciliği algoritmasının tüm veri tiplerini ele alabilmesini olanaksızlaştırmaktadır. Bu yüzden veri tipine özgü adanmış veri madenciliği algoritmaları geliştirilmektedir [26].

2.5.7. Gürültülü ve kayıp değerler

Veri girişi veya veri toplanması esnasında oluşan sistem dışı hatalara gürültü denir. Büyük veri tabanlarında pek çok niteliğin değeri yanlış olabilir. Veri toplanması esnasında oluşan hatalara ölçümden kaynaklanan hatalar da dâhil olmaktadır. Bu hataların sonucu olarak veri madenciliğinde birçok niteliğin değeri yanlış olabilir ve

bu yanlışlardan dolayı veri madenciliği amacına tam olarak ulaşmayabilir. Bu bilgi yanlışlığı ölçüm hatalarından ya da öznel yaklaşımdan kaynaklanıyor olabilir.

Günümüzde kullanılan ticari ilişkisel veri tabanları, veri girişi sırasında oluşan hataları otomatik biçimde gidermek konusunda az bir destek sağlamaktadır. Eğer veri kümesi gürültülü ise sistem bozuk veriyi tanımalı ve ihmal etmelidir. Deneysel sonuçlar, etiketli öğrenmede makine öğrenim tekniklerinin etiket niteliği üzerindeki gürültülere, diğer koşul niteliklerinde sunulan gürültülere kıyasla, daha duyarlı olduklarını göstermiştir. Buna karşın eğitim kümesindeki nesnelere nitelikleri üzerindeki en çok %10'luk gürültü miktarı ayıklanabilmektedir [24].

2.5.8. Sınırlı bilgi

Veri tabanları genel olarak basit öğrenme işlerini sağlayan özellik veya nitelikleri sunmak gibi veri madenciliği dışındaki amaçlar için hazırlanmışlardır. Bu yüzden, öğrenme görevini kolaylaştıracak bazı özellikler bulunmayabilir. Bir veri tabanı öngörü oluşturmak için yeterli bilgileri barındırmıyorsa veri tabanından bilgi keşfi yapılamaz. Yetersiz veri problemlere sebep olmaktadır çünkü bazı veriler geçerli etki alanında sunulamaz.

2.5.9. Veri tabanı boyutu

Veri tabanı boyutları büyük bir hızla artmaktadır. Veri tabanı algoritması çok sayıda küçük örnekleme ele alabilecek biçimde geliştirilmiştir. Aynı algoritmaların yüzlerce kat büyük örneklerde kullanılabilmesi için çok dikkat gerekmektedir. Örneklemin büyük olması, tahminlerin doğruluğu açısından bir avantaj olsa da dikkatsizlikten kaynaklanacak hatalar göz ardı edilemez.

2.6. Veri Madenciliği Süreci

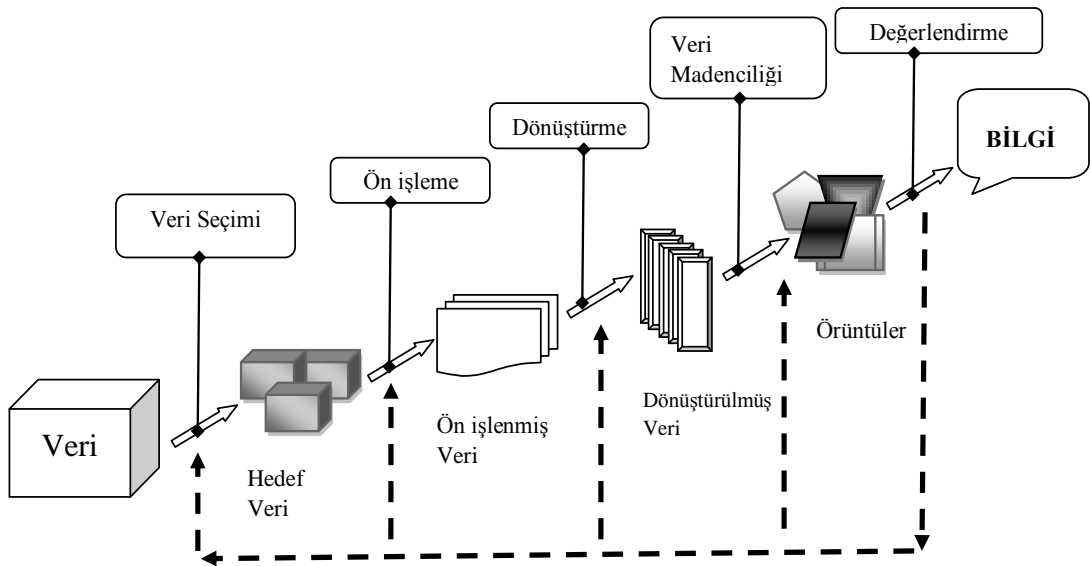
Büyük veri tabanlarında ilginç ve değerli olan bilgiyi algılamak ve erişmek oldukça zordur. Veri Tabanında Bilgi Keşif sürecinin aşamaları (Knowledge Discovery in Databases) bu değerli, önceden bilinmeyen, kullanılabilir olan bilgiye belirli

yöntemler uygulayarak ulaşmada çok büyük rol oynamaktadır. Veri madenciliğinin uygulanabilmesi için kullanılan algoritma bize veri tabanlarında bilgi keşfi sürecinin gerçekleşmesini sağlamaktadır. Bu süreç içerisinde modelin uygulanacağı verilerin özelliklerinin çok iyi bilinmesi gerekmektedir.

Veri madenciliği aynı zamanda bir süreçtir. Veri yığınları arasında soyut kazılar yaparak veriyi ortaya çıkarmanın yanı sıra, bilgi keşfi sürecinde örüntüleri ayrıştırarak süzmek ve bir sonraki adıma hazır hale getirmek de bu sürecin bir parçasıdır. Üzerinde inceleme yapılan işin ve verilerin özelliklerinin bilinmemesi durumunda ne kadar etkin olursa olsun hiç bir veri madenciliği algoritmasının fayda sağlaması mümkün değildir. Bu sebeple, veri madenciliği sürecine girilmeden önce, başarının ilk şartı, iş ve veri özelliklerinin detaylı analiz edilmesidir.

Veri madenciliği sürecinde izlenen adımlar genellikle aşağıdaki şekildedir [27]:

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



Şekil 2.2. Bilgi keşfi sürecinde veri madenciliği

2.6.1. Problemin tanımlanması

Veri madenciliği çalışmalarında başarılı olmanın en önemli şartı, projenin hangi işletme amacı için yapılacağı ve elde edilecek sonuçların başarı düzeylerinin nasıl ölçüleceğinin tanımlanmasıdır. Ayrıca yanlış tahminlerde katlanılacak olan maliyetlere ve doğru tahminlerde kazanılacak faydalara ilişkin tahminlere de bu aşamada yer verilmelidir. Bu aşamada mevcut iş probleminin nasıl bir sonuç üretilmesi durumunda çözüleceğinin, üretilecek olan sonucun fayda-maliyet analizinin, başka bir ifadeyle üretilen bilginin işletme için değerinin doğru analiz edilmesi gerekmektedir.

Bu aşama veri madenciliği sürecinin en önemli aşamasıdır. Araştırma probleminin (konusunun) tanımlanması aşaması araştırmanın amacını, mevcut durumun değerlendirilmesini, veri madenciliğinin amaçlarını ve proje planlama sürecinin belirlenmesini kapsamaktadır.

2.6.2. Verilerin hazırlanması

Modelin kurulması aşamasında ortaya çıkacak sorunlar, bu aşamaya sık sık geri dönülmesine ve verilerin yeniden düzenlenmesine sebep 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 [28].

Bu aşamada firmanın mevcut bilgi sistemleri üzerinde ürettiği sayısal bilginin iyi analiz edilmesi, veriler ile mevcut iş problemi arasında ilişki olması gerektiği unutulmamalıdır. Proje kapsamında kullanılacak sayısal verilerin, hangi iş süreçleri ile yaratıldığı da bu veriler kullanılmadan analiz edilmelidir, bu sayede analist veri kalitesi hakkında fikir sahibi olabilir.

Veri kalitesi, veri madenciliğinde anahtar bir konudur. Veri madenciliğinde güvenilirliğin artırılması için, veri ön işleme yapılmalıdır. Aksi halde hatalı girdi

verileri kullanıcıyı hatalı çıktıya götürecektir. Veri ön işleme, çoğu durumlarda yarı otomatik olan ve yukarıda da belirtildiği gibi zaman isteyen bir veri madenciliği aşamasıdır. Verilerin sayısındaki artış ve buna bağlı olarak çok büyük sayıda verilerin ön işlemeden geçirilmesinin gerekliliği, otomatik veri ön işleme için etkin teknikleri önemli hale getirmiştir [13].

Verilerin hazırlanması şu aşamalarından oluşmaktadır:

- Toplama,
- Değer biçme,
- Birleştirme ve temizleme,
- Örneklem seçimi,
- Dönüştürme.

2.6.3. Modelin kurulması ve değerlendirilmesi

Tanımlanan problem için en uygun modelin bulunabilmesi, olabildiğince çok sayıda modelin kurularak denenmesi ile mümkündür. Bu sebeple veri hazırlama ve model kurma aşamaları, en iyi olduğu düşünülen modele varılıncaya kadar yinelenen bir süreçtir. Model kuruluş süreci, denetimli ve denetimsiz öğrenmenin kullanıldığı modellere göre farklılık göstermektedir.

Örnekten öğrenme olarak da isimlendirilen denetimli öğrenmede, bir denetçi tarafından ilgili sınıflar önceden belirlenen bir kritere göre ayrılarak, her sınıf için çeşitli örnekler verilir. Sistemin amacı verilen örneklerden hareket ederek her bir sınıfa ilişkin özelliklerin bulunması ve bu özelliklerin kural cümleleri ile ifade edilmesidir. Öğrenme süreci tamamlandığında, tanımlanan kural cümleleri verilen yeni örneklerle uygulanır ve yeni örneklerin hangi sınıfa ait olduğu kurulan model tarafından belirlenir [11].

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

Denetimli öğrenmede seçilen algoritmaya uygun olarak ilgili veriler hazırlandıktan sonra, ilk aşamada verinin bir kısmı modelin öğrenilmesi, diğer kısmı ise modelin geçerliliğinin test edilmesi için ayrılır. Modelin öğrenilmesi öğrenim kümesi kullanılarak gerçekleştirildikten sonra, test kümesi ile modelin doğruluk derecesi belirlenir [6].

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.

Sınırlı miktarda veriye sahip olunması durumunda kullanılacak diğer bir yöntem, çapraz geçerlilik testidir. Bu yöntemde veri kümesi rastgele iki eşit parçaya ayrılır. İlk aşamada bir parça üzerinde model eğitimi ve diğer parça üzerinde test işlemi, ikinci aşamada ise ikinci parça üzerinde model eğitimi ve birinci parça üzerinde test işlemi yapılarak elde edilen hata oranlarının ortalaması kullanılır [6].

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

Bootstrapping, küçük veri kümeleri için modelin hata düzeyinin tahmininde kullanılan bir başka tekniktir. Çapraz geçerlilikte olduğu gibi model bütün veri kümesi üzerine kurulur. Daha sonra en az 200, bazen binin üzerinde olmak üzere çok

fazla sayıda öğrenim kümesi tekrarlı örneklemelerle veri kümesinden oluşturularak hata oranı hesaplanır [5].

Model kuruluş çalışmalarına başlamadan önce, hangi tekniğin en uygun olduğuna karar verebilmek güçtür. Bu nedenle farklı modeller kurarak, doğruluk derecelerine göre en uygun modeli bulmak üzere sayısız deneme yapılmasında yarar bulunmaktadır.

Özellikle sınıflama problemleri için kurulan modellerin doğruluk derecelerinin değerlendirilmesinde basit ancak faydalı bir araç olan risk matrisi kullanılmaktadır. Aşağıda bir örneği görülen bu matriste sütunlarda fiili, satırlarda ise tahmini sınıflama değerleri yer almaktadır. Örneğin, fiilen B sınıfına ait olması gereken 46 elemanın, kurulan model tarafından 2'sinin A, 38'inin B, 6'sının ise C olarak sınıflandırıldığı matriste kolayca görülebilmektedir (Çizelge 2.2).

Çizelge 2.2. Risk matrisi [11]

<i>Tahmin Edilmiş Sınıf</i>	<i>Gerçek Sınıf</i>		
	<i>A Sınıfı</i>	<i>B Sınıfı</i>	<i>C Sınıfı</i>
<i>A Sınıfı</i>	45	2	3
<i>B Sınıfı</i>	10	38	2
<i>C Sınıfı</i>	4	6	40

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

Kaldıraç oranı ve grafiği, bir modelin sağladığı faydanın değerlendirilmesinde kullanılan önemli bir yardımcıdır. Örneğin, kredi kartını muhtemelen iade edecek müşterilerin belirlenmesi amacını taşıyan bir uygulamada, kullanılan modelin

belirlediği 100 kişinin 35'i gerçekten bir süre sonra kredi kartını iade ediyorsa ve tesadüfi olarak seçilen 100 müşterinin aynı zaman diliminde sadece 5'i kredi kartını iade ediyorsa kaldıraç oranı 7 olarak bulunacaktır [2].

Kurulan modelin değerinin belirlenmesinde kullanılan diğer bir ölçü, model tarafından önerilen uygulamadan elde edilecek kazancın bu uygulamanın gerçekleştirilmesi için katlanılacak maliyete bölünmesi ile elde edilecek olan yatırımın geri dönüş oranıdır.

Kurulan modelin doğruluk derecesi ne denli yüksek olursa olsun, gerçek dünyayı tam anlamı ile modellediğini garanti edebilmek mümkün değildir. Yapılan testler sonucunda geçerli bir modelin doğru olmamasındaki başlıca sebepler, model kuruluşunda kabul edilen varsayımlar ve modelde kullanılan verilerin doğru olmamasıdır. Örneğin, modelin kurulması sırasında varsayılan enflasyon oranının zaman içerisinde değişmesi, bireyin satın alma davranışını belirgin olarak etkileyecektir [6].

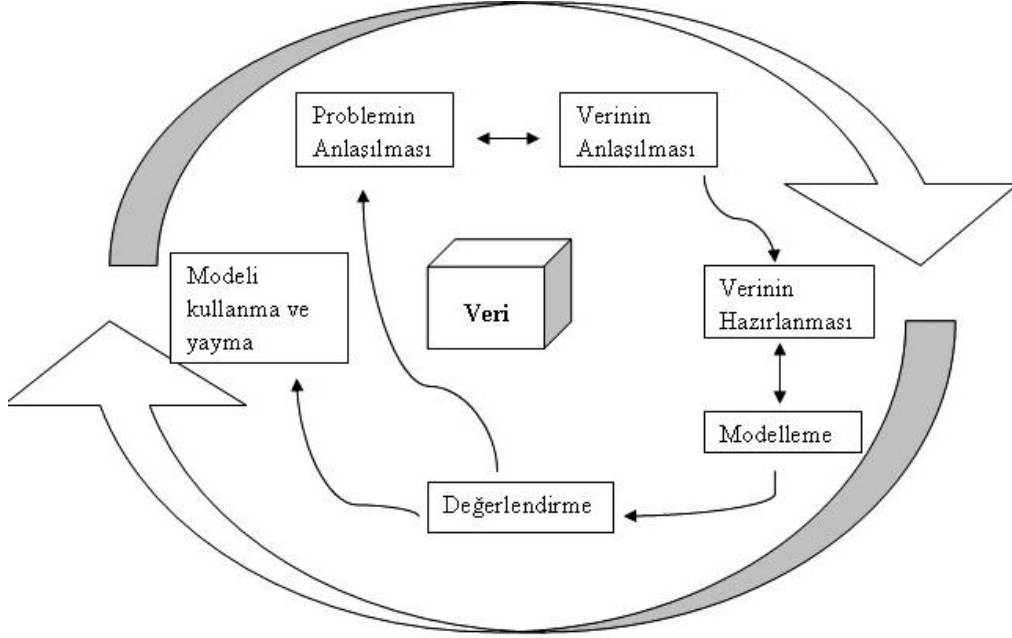
2.6.4. Modelin kullanılması

Kurulan ve geçerliliği kabul edilen model doğrudan bir uygulama olabileceği gibi, bir başka uygulamanın alt parçası olarak kullanılabilir. Kurulan modeller, kullanıldığı alanda amaçlanan problem çözümleri için anlamlı örüntüler ve değerlendirmeler oluşturmaktadır.

2.6.5. Modelin izlenmesi

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

Veri madenciliği süreci adımları Şekil 2.3.'te gösterilmiştir.



Şekil 2.3. Veri madenciliği süreci adımları

2.7. Veri Madenciliği Yöntemleri

Veri madenciliğinde kullanılan modeller, tahmin edici ve tanımlayıcı olmak üzere iki ana başlık altında incelenmektedir [29].

Tahmin edici modellerde, sonuçları bilinen verilerden hareket edilerek bir model geliştirilmesi ve kurulan bu modelden yararlanılarak sonuçlan bilinmeyen veri kümeleri için sonuç değerlerin tahmin edilmesi amaçlanmaktadır. Tanımlayıcı modellerde ise karar vermeye rehberlik etmede kullanılacak mevcut verilerdeki örüntülerin tanımlanması sağlanmaktadır. Veri madenciliği yöntemleri Çizelge 2.3.'te gösterilmektedir.

Çizelge 2.3. Veri madenciliğinde kullanılan yöntemler [30]

VERİ MADENCİLİĞİ YÖNTEMLERİ		
TAHMİN EDİCİ YÖNTEMLER		TANIMLAYICI YÖNTEMLER
Sınıflandırma Yöntemleri	Regresyon	Kümeleme Yöntemleri
Karar Ağaçları		Bağlantı kuralları
Bayes Sınıflandırması		Özetleme
Hatayı Geri Yayma		Tanımsal İstatistik
Karar Destek Makineleri		İstisna Analizi
K-en Yakın Komşu		Diğer Yöntemler
Yapay Sınır Ağları		
Genetik Algoritma		
Zaman Serisi Analizi		
Diğer Yöntemler		

Ayrıca veri madenciliği yöntemlerini denetimli ve denetimsiz olmak üzere iki ana kategoriye ayırırlar da vardır. Veri madenciliğinde iyi tanımlanmış veya kesin bir hedef olduğunda denetimli (supervised) ifadesi kullanılır. Elde edilmesi istenen sonuç için özel bir tanımlama yapılmamışsa veya belirsizlik söz konusu ise denetimsiz (unsupervised) ifadesi kullanılır.

Denetimli ve denetimsiz ifadeleri birbirinin tersidir. Denetimli ve denetimsiz yöntemler sürecin tamamı açısından değerlendirildiğinde;

- Denetimsiz yöntemler daha çok veriyi anlamaya, tanımaya, keşfetmeye yönelik olarak kullanılan ve sonraki uygulanacak yöntemler için fikir vermeyi amaçlamaktadır.
- Denetimli yöntemler ise veriden bilgi ve sonuç çıkarmaya yönelik kullanılmaktadır, denilebilir. Bu nedenle denetimsiz bir yöntemle elde edilen bir bilgi veya sonucu, eğer mümkünse denetimli bir yöntemle teyit etmek, elde edilen bulguların doğruluğu ve geçerliliği açısından önem taşımaktadır.

Denetimsiz yöntemlere örnek olarak Faktör Analizi (Factor Analysis), Temel Bileşenler Analizi (Principle Component Analysis), Hiyerarşik Kümeleme

(Hierarchical Clustering), K-En Yakın Komşuluk (K-Nearest Neighbor), K-Ortalamalar Kümelemesi (K-Means Clustering), İki Aşamalı Kümeleme (Two Step Cluster), Kohonen Ağları (Kohonen Nets) veya Kendi Kendini Düzenleyen Haritalar (Self-Organizing Maps), Aykırılık Saptama (Anomaly Detection) ve Özellik Seçimi (Feature Selection) algoritmaları sayılabilir [31].

Denetimli Veri Madenciliği tekniklerine karar ağacı algoritmaları arasında yer alan Ki-Kare Otomatik Etkileşim Detektörü (Chi-Square Automatic Interaction Detector/CHAID), Ayrıntılı Ki-Kare Otomatik Etkileşim Detektörü (Exhaustive Chi-Square Automatic Interaction Detector/E-CHAID), Sınıflama ve Regresyon Ağacı (Classification and Regression Tree/CRT), Hızlı, Yansız, Etkili İstatistik Ağacı (Quick, Unbiased, Efficient Statistical Tree/QUEST), C5.0 ile Yapay Sinir Ağları, Doğrusal Regresyon Analizi ve Lojistik Regresyon Modelleri ile Birliktelik Kuralları (Association Rules) arasında yer alan Genelleştirilmiş Kural Çıkarsama (Generalized Rule Induction/GRI), Apriori ve CARMA algoritmaları örnek olarak sayılabilir [31].

Veri madenciliği ile ilgili kullanılan pek çok yöntemin yanına hemen her geçen gün yeni yöntem ve algoritmalar eklenmektedir. Bunlardan bir kısmı onlarca yıldır kullanılan klasik teknikler diyebileceğimiz ağırlıklı olarak istatistiksel yöntemlerdir. Diğer yöntemler de genellikle istatistiği temel alan ama daha çok makine öğrenimi ve yapay zekâ destekli yeni nesil yöntemlerdir.

Veri madenciliği modelleri, gördükleri işlemlere göre temel olarak 3 grupta toplanır.

Bunlar:

- 1- Sınıflama (*Classification*) ve Regresyon (*Regression*),
- 2- Kümeleme (*Clustering*),
- 3- Birliktelik Kuralları (*Association Rules*)'dır.

Sınıflama ve regresyon modelleri tahmin edici, kümeleme ve birliktelik kuralları modelleri tanımlayıcı modellerdir.

2.7.1. Sınıflama ve regresyon

Mevcut verilerden hareket ederek geleceğin tahmin edilmesinde faydalanılan ve veri madenciliği teknikleri içerisinde en yaygın kullanıma sahip olan sınıflama ve regresyon modelleri arasındaki temel fark, tahmin edilen bağımlı değişkenin kategorik veya süreklilik gösteren bir değere sahip olmasıdır [25]. Sınıflamada tahmin edilen bağımlı değişken kategorik, Regresyonda ise süreklilik gösteren bir değere sahiptir.

Örneğin, bir sınıflama modeli banka kredi uygulamalarının güvenli veya riskli olmalarını kategorize etmek amacıyla kurulurken, regresyon modeli geliri ve mesleği verilen potansiyel müşterilerin bilgisayar ürünleri alırken yapacakları harcamaları tahmin etmek için kurulabilir. Ancak çok terimli lojistik regresyon (multinomial logistic regression) gibi kategorik değerlerin de tahmin edilmesine imkan sağlayan tekniklerle, her iki model giderek birbirine yaklaşmakta ve bunun bir sonucu olarak aynı tekniklerden yararlanılması mümkün olmaktadır [25].

Sınıflama, verinin önceden belirlenen çıktılara uygun olarak ayrıştırılmasını sağlayan bir tekniktir. Çıktılar, önceden bilindiği için sınıflama, veri kümesini denetimli (supervised) olarak öğrenir [32].

Örneğin, A finans hizmetleri şirketi müşterilerinin yeni bir yatırım fırsatıyla ilgilenip ilgilenmediğini öğrenmek istemektedir. Daha önceden benzer bir ürün satmıştır ve geçmiş veriler hangi müşterilerin önceki teklife cevap verdiğini göstermektedir. Amaç; bu teklife cevap veren müşterilerin özelliklerini belirlemek ve böylece pazarlama ve satış çalışmalarını daha etkin yürütmektir. Müşteri kayıtlarında müşterinin önceki teklife cevap verip vermediğini gösteren “evet”/ “hayır” şeklinde bir alan bulunur. Bu alan “hedef ” ya da “bağımlı” değişken (target variable) olarak adlandırılır. Amaç, müşterilerin diğer niteliklerinin (gelir düzeyi, iş türü, yaş, medeni durum, kaç yıldır müşteri olduğu, satın aldığı diğer ürün ve yatırım türleri) hedef değişken üzerindeki etkilerini analiz etmektir. Analizde yer alan diğer nitelikler “bağımsız” ya da “ tahminci” değişken adını alır [11].

Sınıflama ve regresyon modellerinde kullanılan başlıca teknikler şunlardır [11]:

- 1 - Karar Ağaçları (Decision Trees)
- 2- Yapay Sinir Ağları (Artificial Neural Networks)
- 3- Genetik Algoritmalar (Genetic Algorithms)
- 4- K-En Yakın Komşu (K-Nearest Neighbour)
- 5- Regresyon Analizi
- 6- Naive-Bayes
- 7- Kaba Kümeler

Karar ağaçları

Kuruluşlarının ucuz olması, yorumlanmalarının kolay olması, veri tabanı sistemleri ile kolayca entegre edilebilmeleri ve güvenilirliklerinin iyi olması sebepleri ile en yaygın kullanıma sahip veri madenciliği sınıflama modeli tekniği karar ağaçlarıdır. Karar ağacı, adından da anlaşılacağı gibi bir ağaç görünümünde, tahmin edici bir tekniktir.

Karar ağaçları düğümler ve dallardan oluşan anlaşılması kolay bir tekniktir. Karar ağacında bulunan her bir dalın belirli bir olasılığı mevcuttur. Bu sayede son dallardan köke veya istediğimiz yere ulaşana kadar olasılıkları hesaplamak mümkündür.

SQL sorgusuna kolay dönüştürülebilir yapısıyla en popüler segmentasyon yöntemlerinden birisidir. En yaygın kullanılan karar ağacı yöntemleri şunlardır [33]:

- CHAID (Chi-Squared Automatic Interaction Detector , Kass 1980),
- C&RT (Classification and Regression Trees, Breiman ve Friedman, 1984),
- ID3 (Induction of Decision Trees, Quinlan, 1986),
- C4.5 (Quinlan, 1993).
- C5.0.

Yapay sinir ağıları

Yapay sinir ağıları (YSA), temelde tamamen insan beyni örneklenecek geliştirilmiş bir teknolojidir. Öğrenme, hatırlama, düşünme gibi tüm insan davranışlarının temelinde sinir hücreleri bulunmaktadır. İnsan beyninde tahminen 10^{11} adet sinir hücresi olduğu düşünülmektedir. Bu sinir hücreleri arasında sonsuz denilebilecek sayıda sinaptik birleşme denilen sinirler arası bağ vardır. İnsan beyninin öğrenme yoluyla yeni bilgiler üretebilme, keşfedebilme, düşünme ve gözlemlemeye yönelik yeteneklerini, yardım almadan yapabilen sistemler geliştirmek için tasarlanmışlardır. YSA'nın hesaplama ve bilgi işleme gücünü, paralel dağılmış yapısından, öğrenebilme ve genelleme yeteneğinden aldığı söylenebilir. Genelleme, eğitim ya da öğrenme sürecinde karşılaşılmayan girişler için de YSA'nın uygun tepkileri üretmesi olarak tanımlanır. Bu üstün özellikleri, YSA'nın karmaşık problemleri çözebilme yeteneğini gösterir [14].

Bu yöntem, belirli bir profile uyuması için kalıp düzenlerini kontrol etmekte ve bu süreç içerisinde belli bir öğrenme faaliyeti gerçekleştirerek sistemi geliştirmektedir. Yapay sinir ağlarında kullanılan öğrenme algoritmaları, veriden üniteler arasındaki bağlantı ağırlıklarını hesaplar. Yapay sinir ağıları istatistiksel yöntemler gibi veri hakkında parametrik bir model varsaymaz. Yani uygulama alanı daha geniş ve bellek tabanlı yöntemler kadar yüksek işlem ve bellek gerektirmez [6].

Bir yapay sinir ağı belirli bir amaç için oluşturulur ve insanlar gibi örnekler sayesinde öğrenir. Yapay sinir ağıları tekrarlanan girdiler sayesinde kendi yapısını, ağırlığını değiştirir ve aynen canlıların sinir sistemi gibi adapte olabilen bir yapıya sahiptir. Yapay sinir ağıları, öğrenmenin yanı sıra bilgiler arasında ilişkiler oluşturma yeteneğine de sahiptir.

Yapay sinir ağlarının temel işlevleri şu şekilde belirtilebilir [34]:

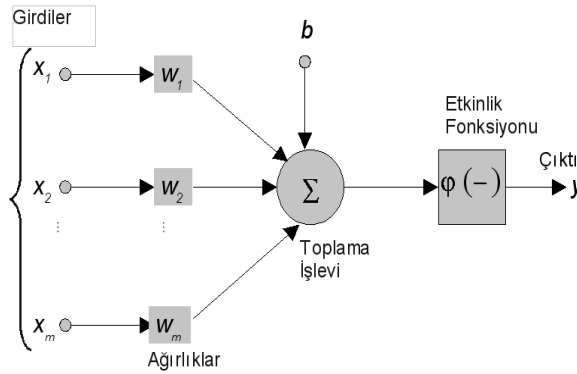
- Öngörü (Prediction) veya tahminleme: İleriki satışlar, hava tahminleri, at yarışları, çevresel risk, ...

- Sınıflandırma (Classification) ve Kümeleme (Clustering): Müşteri profilleri, tıbbi teşhis, ses ve şekil tanıma, hücre tipleri ...
- Kontrol (Control): Erken uyarı için uçaklarda ses ve titreşim düzeyleri, ...

Ayrıca, Veri Birleştirme (Data Association), Kavramsallaştırma (Data Conceptualization) ve Filtreleme (Data Filtering) için de kullanılabilir.

Yapay sinir ağlarının endüstriyel uygulamalar, finans uygulamaları, askeri ve savunma uygulamaları, tıp ve sağlık uygulamaları, mühendislik uygulamaları, robotbilim, görüntü işleme, örüntü tanıma dışında iletişim sanayi, eğlence amaçlı tahmin gibi özel uygulama alanları da bulunmaktadır [34].

Bir yapay sinir ağı hücresi temel olarak girdilerden, ağırlıklardan, toplama işlevinden ve çıktıdan oluşur.



Şekil 2.4. Yapay sinir hücresi [34]

Genelde, verilen bir girdi kümesine karşılık çıktı değerleri verilerek belirtilen öğrenme kuralına göre ağırlık değerleri otomatik olarak değiştirilmektedir. Eğitim verisinin tamamlanmasından sonra eğitilmiş olan ağ, ağırlık değerlerinin son durumuna göre, verilen herhangi bir veri kümesinin sonucunu tahminleyebilmektedir. Yapay sinir ağı bir dizi sinir hücresinin ileri sürümlü ve geri beslemeli bağlantı şekilleri ile birbirine bağlanması ile oluşur [34].

Günümüzde, belirli amaçlarla ve değişik alanlarda kullanılmaya uygun birçok yapay sinir ağı modeli (Perceptron, Adaline, MLP, LVQ, Hopfield, Recurrent, SOM, ART ve PCA gibi) geliştirilmiştir. Öğrenme çeşitlerinden öğreticili öğrenme, öğreticisiz öğrenme, destekleyicili öğrenme ve karma stratejiler kullanılmaktadır [34].

Genetik algoritmalar

Algoritma ilk olarak popülasyon adı verilen bir çözüm kümesi (öğrenme veri kümesi) ile başlatılır. Bir popülasyondan alınan sonuçlar bir öncekinden daha iyi olacağı beklenen yeni bir popülasyon oluşturmak için kullanılır. Evrim süreci tamamlandığında bağımlılık kuralları veya sınıf modelleri ortaya konmuş olur [35].

Yapay sinir ağları gibi biyolojik sistemlerden esinlenerek tasarlanmış bir tür makine öğrenimi sistemidir. Evrim teorisini modelleyen bu algoritmalar, doğal şartlara en fazla uyum sağlayan bireyin hayatta kalabilmesi temeline dayanan doğal seleksiyonu bilgisayar ortamında simüle eder [8].

Genetik algoritmalar, optimizasyon problemlerinde olduğu gibi sınıflandırma için de kullanılabilir. Veri madenciliğinde, diğer algoritmaların uygunluğunu değerlendirmek için kullanılabilirler [36].

Problemlere tek bir çözüm üretmek yerine farklı çözümlerden oluşan bir çözüm kümesi üreten genetik algoritma sayesinde, arama uzayında aynı anda birçok nokta değerlendirilmekte ve sonuçta bütünsel çözüme ulaşma olasılığı yükselmektedir. Çözüm kümesindeki çözümler birbirinden tamamen bağımsızdır. Her biri çok boyutlu uzay üzerinde bir vektördür. Problem için olası pek çok çözümü temsil eden kümeler genetik algoritma terminolojisinde nüfus adını alır. Nüfuslar vektör veya birey adı verilen sayı dizilerinden oluşur. Birey içindeki her bir elemana gen adı verilir. Nüfustaki bireyler evrimsel süreç içinde genetik algoritma işlemcileri tarafından belirlenir [14].

K-En yakın komşu

Veri madenciliğinde sınıflama amacıyla kullanılan bir diğer teknik ise örneksme yoluyla öğrenmeye dayanan k-en yakın komşu algoritmasıdır. Bu teknikte tüm örneklem bir örüntü uzayında saklanır. Algoritma, bilinmeyen bir örneklemin hangi sınıfa dahil olduğunu belirlemek için örüntü uzayını araştırarak bilinmeyen örnekleme en yakın olan k örneklemini bulur. Yakınlık Öklid uzaklığı ile tanımlanır. Daha sonra bilinmeyen örneklem, k en yakın komşu içinden en çok benzediği sınıfa atanır. K-en yakın komşu algoritması, aynı zamanda bilinmeyen örneklem için bir gerçek değer tahmininde de kullanılabilir.

Bu yöntemin tercih edilme sebebi, sayısı bilinen veri kümeleri için hızlı ve verimli olmasıdır [37]. Özellikle büyük veri tabanlarında kullanılan bir sınıflandırma tekniğidir. Sınıflandırılmak istenen nesnenin ait olduğu kümeyi, en yakınında yer alan K birim nesneden en fazla birime ait olanla aynı kümede sınıflandırması mantığına dayanmaktadır [33].

Bu yöntemde eğitim örnekleri n-boyutlu sayısal niteliklerle tanımlanır. Her örnek n-boyutlu uzayda bir noktayı gösterir. Bu yolla, bütün eğitim örnekleri n-boyutlu örnek uzayda depolanır. Bu k-eğitim örnekleri bilinmeyen örneğe k en yakın komşudur. Öklid uzaklığı $X=(x_1,x_2,\dots,x_n)$ ve $Y=(y_1,y_2,\dots,y_n)$ arasındadır. Bilinmeyen örnek k-komşularının arasında en sık olan sınıfa verilir. k=1 olduğunda, bilinmeyen örnek uzayında en yakın eğitim örneğinin sınıfıdır. En yakın komşu sınıflandırıcısı örnek-tabanlı ya da tembel-öğrenenlerdir (lazylearners); bunlar bütün eğitim örneklerini depolarlar ve yeni bir örnek sınıflandırmaya gerek olmadıkça bir sınıflandırıcı yapılandırmazlar. Tembel-öğrenenler, potansiyel komşularının sayısı etiketlenmemiş örneklerin sayısına göre büyükse hesaplama maliyetini arttırabilir. Bu nedenle etkili indeksleme teknikleri gerektirir. Beklendiği gibi lazy-learning yöntemleri eager yöntemlerden daha hızlıdır fakat bütün hesaplamalar geciktirildiğinde sınıflanma yavaşlar. En yakın komşu sınıflandırıcısı kestirim içinde kullanılabilir [36].

Sınıflandırma problemini çözen denetimli öğrenme yöntemleri arasında yer alır. Yöntemde sınıflandırma yapılacak verilerin öğrenme kümesindeki normal davranış verilerine benzerlikleri hesaplanarak, en yakın olduğu düşünülen k verinin ortalamasıyla, belirlenen eşik değere göre sınıflara atamaları yapılır. Önemli olan, her bir sınıfın özelliklerinin önceden net bir şekilde belirlenmiş olmasıdır. Yöntemin performansını k en yakın komşu sayısı, eşik değer, benzerlik ölçümü ve öğrenme kümesindeki normal davranışların yeterli sayıda olması kriterleri etkilemektedir [18].

Regresyon analizi

Regresyon analizi, bir ya da daha fazla bağımsız değişken ile hedef değişken arasındaki ilişkiyi matematiksel olarak modelleyen bir yöntemdir. Veri madenciliğinde yaygın olarak kullanılan regresyon modellerinden doğrusal regresyonda tahmin edilecek olan hedef değişken sürekli değer alırken, lojistik regresyonda hedef değişken kesikli bir değer almaktadır. Doğrusal regresyonda hedef değişkenin değeri, lojistik regresyonda ise hedef değişkenin alabileceği değerlerden birinin gerçekleşme olasılığı tahmin edilmektedir [38].

Doğrusal, doğrusal olmayan ve lojistik modelleme alternatifleri imkanı vardır. Bağımsız değişken olarak anılan tahmin edici değişkenlerin, bağımlı değişken denilen tahmin edilecek değişken değerini belirleyecek ağırlıklandırılmaları içerecek bir bağımsız değişkenler birleşimidir [33].

Naive-Bayes

Naive Bayes algoritması her kriterin sonuca olan etkilerinin olasılık olarak hesaplanması temeline dayanmaktadır. Örneğin, sağlık sektöründe bir kişinin tahlil sonuçlarının değerlendirilerek bir hastalığa yakalanmış olup olmama olasılığının değerlendirilmesinde de sıkça kullanılmaktadır.

Naive Bayes, hedef değişkenle bağımsız değişkenler arasındaki ilişkiyi analiz eden tahminci ve tanımlayıcı bir sınıflama algoritmasıdır. Sürekli veri ile çalışmaz. Bu

sebeple sürekli deęerleri içeren baęımlı ya da baęımsız deęişkenler kategorik hale getirilmelidir. Örneęin; baęımsız deęişkenlerden biri yaş ise, sürekli deęerler “<20” “21-30”, “31-40” gibi yaş aralıklarına dönüştürülmelidir. Naive Bayes modelin öğrenilmesi esnasında, her çıktının öğrenme kümesinde kaç kere meydana geldiğini hesaplar. Bulunan bu deęer öncelikli olasılık olarak adlandırılır. Örneęin, bir banka kredi kartı başvurularını “iyi” ve “kötü” risk sınıflarında gruplandırmak istemektedir. İyi risk çıktısı toplam 5 vaka içinde 2 kere meydana geldiyse iyi risk için öncelikli olasılık 0,4’tür. Bu durum, “Kredi kartı için başvuran biri hakkında hiçbir şey bilinmiyorsa, bu kişi 0,4 olasılıkla iyi risk grubundadır.” olarak yorumlanır. Naive Bayes aynı zamanda her baęımsız deęişken / baęımlı deęişken kombinasyonunun meydana gelme sıklığını bulur. Bu sıklıklar öncelikli olasılıklarla birleştirilmek suretiyle tahminde kullanılır [11].

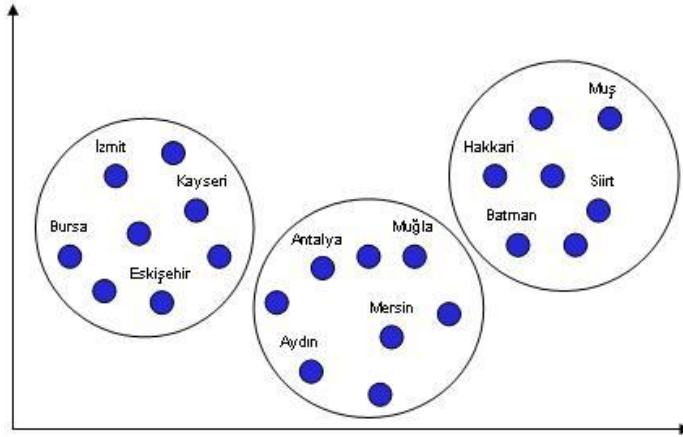
Kaba kümeler

Kaba küme teorisi 1970’li yıllarda Pawlak tarafından geliştirilmiştir. Kaba küme teorisinde bir yaklaştırma uzayı ve bir kümenin alt ve üst yaklaşırmaları vardır. Yaklaştırma uzayı, ilgilenilen alanı ayrı kategorilerde sınıflandırır. Alt yakınlaştırma belirli bir altkümeyle ait olduđu kesin olarak bilinen nesnelerin tanımıdır. Üst yakınlaştırma ise alt kümeyle ait olması olası nesnelerin tanımıdır. Alt ve üst sınırlar arasında tanımlanan herhangi bir nesne ise “kaba küme” olarak adlandırılır [39].

2.7.2. Kümeleme

Nesneleri benzerleriyle gruplama sürecine kümeleme denir. Kümeleme analizi, temel amacı nesneleri (birim) sahip oldukları karakteristik özellikleri baz alarak gruplamak olan çok deęişkenli teknikler grubudur. Eldeki veriler yardımıyla analiz yaparken, verimli ve güvenilir analiz yapma açısından kümeleme analizi oldukça büyük bir önem taşımaktadır. Örneęin, Türkiye’deki şehirlerin profilini çıkarmak üzere bir araştırma yapılacaktır. Tarıma dayalı bir gelir sistemi olan bir şehri, geliri sanayiye dayalı olan şehirlerle kıyaslamak ne kadar güvenilir sonuçlar verir tartışılır. Aynı şekilde nüfusu milyonlarla ölçülen şehirleri nüfusu yüz binler olan şehirlerle

kıyaslamak da yanlıştır. Belirlenen kriterlere göre benzer özellik gösteren şehirler bir gruba toplanır ve kendi aralarında analiz yapılır. Örneğin, Hakkari'yi Ankara ile kıyaslamak yerine profili benzer olabilecek Siirt, Batman, Muş vb. şehirlerle kıyaslamak çok daha güvenilir sonuçlar elde edilmesini sağlayacaktır [6].



Şekil 2.5. Şehirleri kümeleme yöntemi örneği [6]

Kümeler oluşturulurken, kümenin içinde yer alan nesnelere arası benzerlik en büyük, kümeler arası benzerlik ise en küçük olacak şekilde olmalıdır. Benzer bir ifade ile kümeleme analizinin amacı, var olan verileri içsel olarak homojen, kümeler arası heterojen olarak ayırmaktır. Kümeleme analizi genellikle market araştırmaları, gen araştırmaları gibi konularda sıklıkla kullanılır. Bu veriler arasında nitelik bakımından ilginç korelasyonlar elde edilebilir. Örneğin, tıp alanında benzer özelliklere sahip genler aynı küme içine alınabilir.

Kümeleme modellerinde amaç üyelerinin birbirlerine çok benzediği, ancak özellikleri birbirlerinden çok farklı olan kümelerin bulunması ve veri tabanındaki kayıtların bu farklı kümelere bölünmesidir. Veri tabanındaki kayıtların hangi kümelere ayrılacağı veya kümelemenin hangi değişken özelliklerine göre yapılacağı konunun uzmanı olan bir kişi tarafından belirtilebileceği gibi veri tabanındaki kayıtların hangi kümelere ayrılacağını geliştirilen bilgisayar programları da yapabilmektedir.

Marketlerde farklı müşteri gruplarının keşfedilmesi ve bu grupların alışveriş örüntülerinin ortaya konması, biyolojide bitki ve hayvan sınıflandırmaları ve işlevlerine göre benzer genlerin sınıflandırılması, şehir planlanmasında evlerin tiplerine, değerlerine ve coğrafik konumlarına göre gruplara ayrılması gibi uygulamalar tipik kümeleme uygulamalarıdır. Kümeleme aynı zamanda Web üzerinde bilgi keşfi için dokümanların sınıflanması amacıyla da kullanılabilir [40].

Veri kümeleme güçlü bir gelişme göstermektedir. Veri tabanlarında toplanan veri miktarının artmasıyla orantılı olarak, kümeleme analizi son zamanlarda veri madenciliği araştırmalarında aktif bir konu haline gelmiştir. Literatürde pek çok kümeleme algoritması bulunmaktadır. Kullanılacak olan kümeleme algoritmasının seçimi, veri tipine ve amaca bağlıdır.

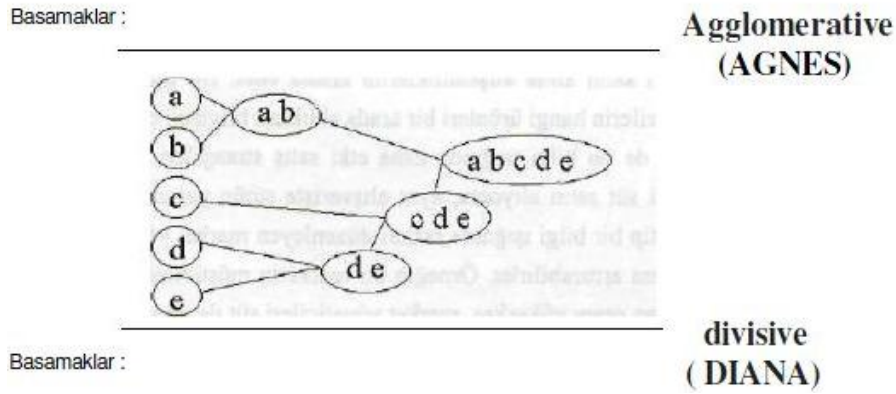
Genel olarak başlıca kümeleme yöntemleri şu şekilde sınıflandırılabilir [36]:

- Bölme yöntemleri (Partitioning methods)
- Hiyerarşik yöntemler (Hierarchical methods)
- Yoğunluk tabanlı yöntemler (Density-based methods)
- Izgara tabanlı yöntemler (Grid-based methods)
- Model tabanlı yöntemler (Model-based methods)

Bölme yöntemlerinde, n veri tabanındaki nesne sayısı ve k oluşturulacak küme sayısı olarak kabul edilir. Bölme algoritması n adet nesneyi, k adet kümeye böler ($k < n$). Kümeler tarafsız bölme kriteri olarak nitelendirilen bir kriterine uygun oluşturulduğu için aynı kümedeki nesnelere birbirlerine benzerken, farklı kümedeki nesnelere farklıdır. Bölme yöntemlerinde içerisinde kullanılan en yaygın yöntem ise k-means yöntemidir [25].

Kümeleme yöntemlerinden biri olan hiyerarşik yöntemler, veri nesnelere kümeler ağacı şeklinde gruplara ayırma esasına dayanır. Hiyerarşik kümeleme yöntemleri, hiyerarşik ayrışmanın aşağıdan yukarıya veya yukarıdan aşağıya doğru olmasına göre agglomerative ve divisive hiyerarşik kümeleme olarak sınıflandırılabilir [41].

Agglomerative hiyerarşik kümelemede, Şekil 2.6.'da görüldüğü gibi hiyerarşik ayrışma aşağıdan yukarıya doğru olur. İlk olarak her nesne kendi kümesini oluşturur ve ardından bu atomik kümeler birleşerek, tüm nesnelere bir kümede toplanıncaya dek daha büyük kümeler oluştururlar. Divise hiyerarşik kümelemede, Şekil 2.6.'da görüldüğü üzere hiyerarşik ayrışma yukarıdan aşağıya doğru olur. İlk olarak tüm nesnelere bir kümededir ve her nesne tek başına bir küme oluşturana dek, kümeler daha küçük parçalara bölünürler [41].



Şekil 2.6. Hiyerarşik kümeleme örneği [41]

2.7.3. Birliktelik kuralları

Birliktelik kuralları, büyük veri kümeleri arasında birliktelik ilişkileri bulurlar. Toplanan ve depolanan verinin her geçen gün artması yüzünden, şirketler veri tabanlarındaki birliktelik kurallarını ortaya çıkarmak istemektedirler. Büyük miktardaki mesleki işlem kayıtlarından ilginç birliktelik ilişkilerini keşfetmek, şirketlerin karar alma işlemlerini daha verimli hale getirmektedir.

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.

Örneğin, bir müşteri süt satın alıyorsa, aynı alışverişte sütün yanında ekmeğin alma olasılığı nedir? Bu tip bir bilgi ışığında rafları düzenleyen market yöneticileri ürünlerindeki satış oranını arttırabilirler. Örneğin, bir marketin müşterilerinin süt ile birlikte ekmeğin satın alınma oranı yüksekse, market yöneticileri süt ile ekmeğin raflarını yan yana koyarak ekmeğin satışlarını arttırabilirler.

Birliktelik kuralının matematiksel modeli Agrawal, Imielinski ve Swami tarafından 1993 yılında sunulmuştur [42]. Bu modelde, $I = \{i_1, i_2, \dots, i_m\}$ kümesine “ürünler” adı verilmektedir. D , veri bütünlüğündeki tüm hareketleri, T ise ürünlerin her bir hareketini simgeler. TID ise, her harekete ait olan tek belirteçtir [16].

Birliktelik kurallarına ait örnekler aşağıda yer almaktadır:

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

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

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

Birliktelik kuralına ilişkin olarak geliştirilen bazı algoritmalar şunlardır; AIS, SETM, Apriori, Partition, RARM, CHARM.

Bu algoritmalar içerisinde, ilk olanı AIS, en bilineni ise Apriori algoritmasıdır [16].

3. MOBİL İLETİŞİM

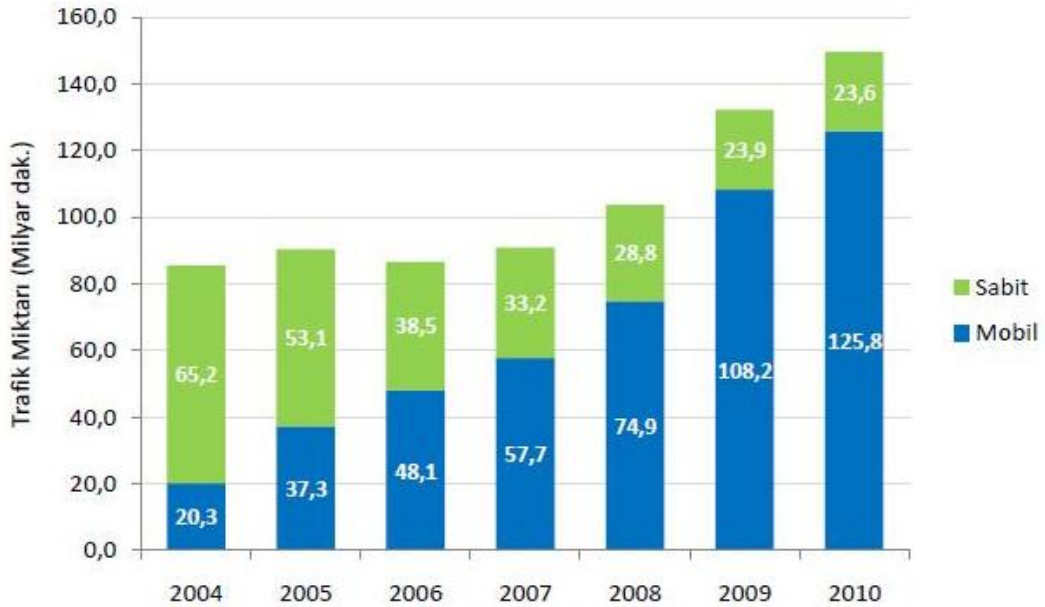
Mobil iletişim alanı bu çalışmada, veri madenciliği uygulaması için kullanılan verilerin sağlandığı alandır. Mobil iletişim aracı olan cep telefonlarından elde edilen GSM şebekelerinin sinyal gücü bilgileri, veri madenciliği yöntemlerinden olan iki aşamalı kümeleme yöntemi ve frekans analizi yöntemiyle değerlendirilmiştir. Elde edilen bulguları açıklamadan önce mobil iletişim alanına değinmenin faydalı olacağı düşünülmüştür.

Mobil iletişim, kullanıcıların iletişim esnasında mekandan bağımsız olabildikleri, hareket özgürlüğüne sahip oldukları bir iletişim şeklidir. Mobil teknolojilere olan ilgi 90'ların ikinci yarısından sonra hızla gelişmektedir. Bu mobilite insanların ihtiyaçları doğrultusunda gelişim göstermektedir. Bugün hayatımızın artık her anında bilgiye, her yerden, cep telefonları, pocket pc'ler, laptoplar aracılığıyla ulaşmak mümkündür.

İletişim ve haberleşme dışında mobil teknolojilerin bir kısmı kamera çekimi ve müzik dinleme gibi eğlence aracı olarak da kullanılabilir. Bunun yanı sıra toplumun önemli bir bölümü mobil web uygulamalarına cep telefonları aracılığıyla erişebilmektedir. Mobil web uygulamaları ile trafik bilgisinden, otobüs, tren ve uçak seferlerine ve hava tahminlerine kadar pek çok konuda bilgi hizmeti sunulabilmektedir. Özellikle son dönemlerde toplumun önemli bir bölümü günlük yaşamda yeri olan aktivitelerin önemli bir bölümünü mobil sosyal ağlar (mobile social networks) üzerinde sürdürme eğilimi içine girmiştir. Mobil sosyal ağlar, günlük ihtiyaçları karşılamak üzere ayrı ayrı webler üzerinde hizmete sunulan uygulamaları aynı ortam içinde bütünleştirmeyi başaran teknolojik bir çözümdür. Mobil sosyal ağlar üzerinde bilgilenme ihtiyacının yanı sıra, eğlence, alışveriş ve haberleşme gibi çok farklı ihtiyaçları karşılayabilmektedir. Bu yönüyle mobil sosyal ağlar, bilgi kaynakları ya da bilgi merkezlerine erişimden bibliyografik verilerin yönetilmesine ya da çeşitli film ve fotoğrafları depolamadan bilet rezervasyonuna kadar farklı tür ve konularda hizmet sağlayan çok boyutlu işlem platformudur [43].

İnsanlık tarihinde her yeni teknolojinin gelişimiyle beraber, kişilerin yaşama şekillerini değiştiren boyutlarda ilerlemeler kaydedilmiştir. Bugüne baktığımız zaman ise, mobil iletişim teknolojilerinin yeni bir dönüm noktası olacağını gözlemlemek olasıdır. Mobil iletişim teknolojileri, geniş perspektiften bakıldığında kişilere sınırsız özgürlük tanıyacak, kurumların ise çok daha etkin çalışmalarını sağlayacaktır [44].

Mobil iletişime olan ilgi geçen zamanla birlikte katlanarak artmaktadır. İnsanlar artık mobil iletişimi sadece görüşme olarak değil, internet sayesinde tüm işleri için de kullanır hale gelmiş bulunmaktalar. Türkiye’de mobil iletişim hatlarını kullananların arama trafiği Şekil 3.1.’de gösterilmektedir. Buna göre, 2004 yılından itibaren yıllık trafik miktarlarına bakıldığında, toplam trafiğin durağan bir trend izlediği, ancak 2008 yılında artışa geçtiği görülmektedir. Diğer taraftan, toplam trafiğin dağılımına bakıldığında yıllar itibariyle mobil arama trafik miktarı artarken sabit arama trafik miktarının düştüğü görülmektedir.



Şekil 3.1. Yıllara göre toplam arama trafiği [45]

Mobil iletişim hizmetlerinin geliştirilmesi ve çeşitlendirilmesiyle birlikte, mobiliteye duyulan ilgi sürekli olarak canlı tutulmaktadır. XG hizmetlerinin de hayatımızda yerini almasıyla birlikte kullanıcılar bu hizmete de giderek artan sayılarda ilgi

göstermişlerdir. Aralık 2010 itibariyle Türkiye’de toplam 61,8 milyon mobil abone bulunmaktadır.

Temmuz 2009’da 3G hizmet sunumunun başlamasıyla Aralık 2010 itibariyle 3G abone sayısı 19,4 milyona ulaşmıştır [45]. 3G hizmetlerinin 2010 yılına ait verilerine bakıldığında, kullanım sayısının gittikçe arttığı ortaya çıkmaktadır. Ayrıca mobil internet hizmetini kullanan kullanıcı sayısı da, 2010 yılının her üç aylık döneminde büyük artışlar göstermiştir.

Çizelge 3.1. 3G hizmeti kullanıcı verileri [45]

	2010-1	2010-2	2010-3	2010-4
3G Abone sayısı	8.717.769	11.433.031	16.615.286	19.407.264
Mobil internet kullanıcı sayısı	640.580	832.321	1.158.866	1.448.020
Mobil internet kullanıcı miktarı (GB)	2.105.643	2.629.253	3.274.139	4.387.315

Çizelge 3.1.’e göre, 2010 birinci çeyrekte 8,7 milyon olan 3G abone sayısı 2010 dördüncü çeyrekte 19,4 milyona ulaşırken, 3G hizmetiyle birlikte mobil internet hizmeti alan kullanıcı sayısı da aynı dönemler için 640.580’den 1.448.020’ye yükselmiştir. Bu dönemde toplam mobil internet kullanım miktarı ise 4.284 TB olarak gerçekleşmiştir.

3.1. Mobil İletişim Tarihçesi

Radyo teknolojisinin geliştirilmesi mobil cihazların başlangıcı olarak kabul edilir. Radyoyu takip eden mobil cihazlar teknolojik olarak uygulamaya geçiş sırasına göre çağrı cihazı, analog araç ve cep telefonu, sayısal cep telefonu, diz üstü bilgisayar, cep bilgisayarı, internet kabiliyetli cep telefonu ile cep bilgisayarıdır [46].

Analog sistemde çalışan ilk taşınabilir telefonların sayıları zaman içerisinde artınca, kullanılan sinyal aralıkları yetersiz kalmıştır. Bunun sonucu olarak da, araştırmacılar tarafından hücre (cell) teknolojisi geliştirilmiştir. Bu teknolojiyi kullanan ilk ticari

servis, 1982 yılında Advanced Mobile Phone Service (AMPS) olmuştur. O tarihten bugüne dek yaşanan çok büyük teknolojik ilerlemeler sonucunda ise artık her bireyde bir (ya da daha fazla) cep telefonu bulunur hale gelmiştir.

Günümüzde kullanımda olan, XG teknolojili kablosuz iletişim cihazları ve uygulamaları sayesinde bilgiye erişme ve onu kullanma yöntemleri çeşitlenmiş, farklılaşmış ve fark edilir şekilde değiştirmiştir. Bu iletişim teknolojileri sayesinde birkaç satırlık metinlerden kurtulup renkli grafiklere, yüksek ses kalitesindeki müzik yayınlara, video akışına ve kolay e-ticaret özelliklerine ulaşmak mümkün olmuştur.

Artık teknolojinin ilerleme hızı, açılı bir eğri olarak büyümeyi aşmış, neredeyse dikey boyutta katlanır hale gelmiştir. Amerika Birleşik Devletleri'nde kullanımda olan 4G kavramından günümüzde artık ülkemizde de bahsedilmektedir. Bu seviye mobil teknolojinin ise, günlük hayata ve iş hayatına daha pek çok yenilikler getireceği kesindir.

3.1.1. Türkiye'de mobil iletişim süreci

Türkiye'de Temmuz 1993 tarihinde, Turkcell ve Telsim, Türk Telekom ile gelir paylaşımı esasına dayalı olarak bir sözleşme imzalamışlardır. 27 Nisan 1998 tarihinde, Ulaştırma Bakanlığı ile ilgili şirketler arasında 25 yıllık GSM lisans imtiyaz sözleşmesi imzalanmış ve ülkemizde ilk defa telekomünikasyon alanında Türk Telekom dışında alternatif işletmeciler ticari olarak faaliyet göstermeye başlamıştır. 1998 yılında telekomünikasyon sektöründe faaliyete başlayan Turkcell ve Telsim, hızlı bir ivme ile tüm Türkiye'de geniş bir coğrafi alanda büyük bir abone potansiyeline ulaşmıştır [47].

2001 yılında, mevcut iki GSM işletmecisine ilaveten, 1800 MHz frekans bandında, Aria Şebekesi de hizmet sunmaya başlamıştır. Daha sonra Türk Telekom'un şirketi Aycell de bu banttan hizmet vermek üzere görev sözleşmesi imzalanmıştır. Böylece GSM pazarında dört adet işletmeci hizmet sunmaya başlamıştır. Devam eden süreçte Aycell ve Aria şebekelerinin birleştirilmesi kararı alınmıştır. Birleşmeden sonra Aria

ve Aycell markaları, 23 Haziran 2004 tarihi itibarıyla Avea markası olarak anılmıştır. Avea, üçüncü mobil işletmecisi olarak 1800 MHz bandında hizmetini sürdürmektedir. İlerleyen süreçte Telsim, Bankalar Kanunu'nun ilgili hükümleri gereğince, 13 Aralık 2005 tarihinde Vodafone firmasına satılmıştır [47].

Üçüncü nesil mobil şebekeler yüksek veri hızı ve kapasitesi gerektiren video, görüntülü görüşme ve eğlence uygulamaları gibi birçok yenilikçi hizmeti desteklemektedir. 30 Temmuz 2009 tarihinden itibaren de Turkcell, Vodafone ve Avea Şebekeleri, abonelerine üçüncü nesil mobil telekomünikasyon hizmetlerini sunmaya başlamışlardır [47].

3.2. Mobil Cihazlarda Kullanılan İşletim Sistemleri

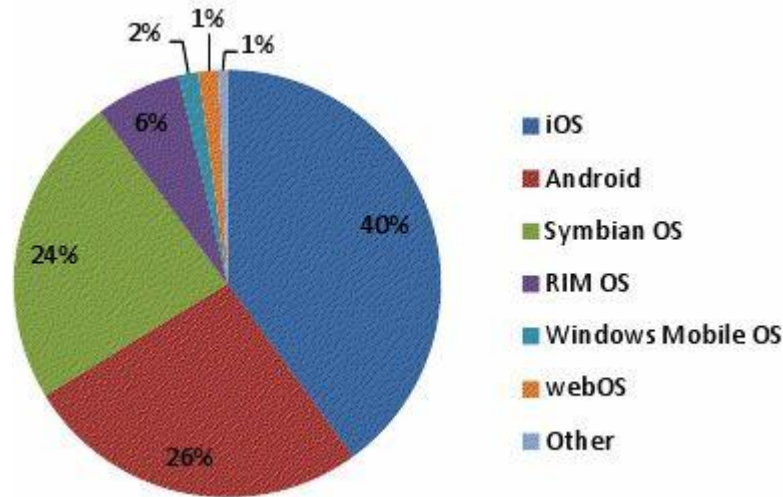
Mobil cihazlar o kadar hızlı geliştirdi ki artık onlar üzerinde de bilgisayarlarda olduğu gibi bir işletim sistemi kurulup çalıştırılabilmektedir. İşletim sistemleri mobil cihazlarla kullanıcılar arasında bağdaştırıcı görevi görmektedir. Bunun yanında birçok görevi yerine getirmektedir. Bunlar arasında kısa film seyretme, mp3 yükleme, dinleme, birçok resim formatını görüntüleyebilme, mikro programlama dilleri ile yazılmış uygulamaları çalıştırma, görsel oyunları yükleme ve oynama gibi sayısız birçok özellik mevcuttur [48].

Bilgisayarlarda bulunan işletim sistemlerinde olduğu gibi, mobil işletim sistemleri de, diğer programların üzerinde çalıştığı platformu ifade eder. Taşınabilir bir cihaz satın alındığında üreticinin o cihaza özel olarak bir mobil işletim sistemi seçmiş olduğu görülür. İşletim sistemi, o cihazdaki mevcut işlevlerin yerine getirilmesini sağlar. Örneğin, klavyenin ve işaretleme cihazının çalışması, WAP bağlantısı, e-mail ve metin mesajlaşmaları gibi. Mobil işletim sistemleri ayrıca hangi tür üçüncü parti yazılımların sisteme kurulabileceğini de kontrol eder. Mobil işletim sistemlerinin genel özellikleri Çizelge 3.2.'de belirtilmektedir.

Çizelge 3.2. Mobil işletim sistemlerinin özellikleri

İşletim Sistemi	Flash Desteği	Kernel	Uygulama Mağazası	Çoklu Dokunmatik	Çoklu İşlem	Firma	Kaynak Kodu
İphone (iOS)	Yok	Hybrid	Var	Var	Kısmen	Apple	Kapalı
Android	Kısmen	Linux	Var	Var	Var	Google	Açık
Windows Mobil	Kısmen	Win. CE	Var	Kısmen	Var	Microsoft	Kapalı
BlackBerry	Yok	Java	Var	Kısmen	Var	RIM	Kapalı
Palm OS	Kısmen	Linux	Var	Var	Var	HP	Kapalı
Symbian OS	Var	Micro kernel	Var	Var	Var	Nokia	Kapalı
Bada	Yok	Linux	Var	Var	Var	Samsung	Kapalı

AdMob imzasını taşıyan Mayıs 2010 tarihli verilere göre dünyadaki işletim sistemi tercihleri Şekil 3.4.'te görülmektedir.



Şekil 3.2. Dünyadaki işletim sistemleri tercihleri [53]

Buna göre küresel bazda iOS' un (Iphone OS) %40 ile epey bir önde götürdüğü yarışta, onun hemen gerisinde %26 ile Google'ın Android'i gelirken, bu ikiliyi Symbian %24 ile takip etmiştir.

Mobil cihazlarda kullanılan işletim sistemlerinin özellikleri aşağıda açıklanmıştır.

3.2.1. Iphone (iOS)

Unix kökenli BSD ve NeXTSTEP işletim sistemlerinin önce Darwin, sonra Mac OS X şeklindeki evriminin bir türevidir. Mac OS X kökenli olduğu için oldukça ileri düzeyde bir altyapıya sahip olmasına rağmen bu konuda isim yaptığı kullanıcı deneyimi kalite standardını korumak için bir takım gelişkin özellikleri kısıtlanmıştır. Kapalı kaynaklı bir işletim sistemidir ve bu sebeple yalnızca Apple'a ait ürünlerde bulunur. Dolayısı ile değişik platformlara uyumluluk zorunluluğu olmadığı için optimizasyon miktarı yüksektir. İşletim sistemi güncellemeleri için bir bilgisayara bağlı olmaya ihtiyaç duyar.



Resim 3.1. iOS ekranı

Sadece parmak etkileşimi ile çalışabilecek şekilde tasarlanmıştır. Bu sebeple rakiplerine kıyasla basit ve sade bir arayüze sahiptir ve kolay bir kullanım sağlar. Ana kullanıcı arabirimi, uygulama kısayollarının bir tablo üzerine yerleşik olduğu bir panodur. Çoklu dokunuş desteği mevcuttur. WebKit tabanlı Safari tarayıcısı ile zengin bir internet, iPod medya oynatıcısı ile de zengin bir çoklu ortam deneyimi sunar. Geliştiriciler için kapsamlı ve gelişmiş bir uygulama geliştirme ortamı (iPhone

SDK) ile dağıtım kanalı (AppStore) sunar. Bundan dolayı uygulama çeşitliliği ile sayısı açısından şu an en zengin platformdur.

Platforma doğal bir dilde uygulama geliştirilmesine izin verir ve bu sebeple uygulama başarımlarının performansı rakiplerine kıyasla yüksektir, kaliteli ve hızlı uygulamalar için yüksek işlemci gücüne ihtiyaç duymaz. Aynı anda birden çok uygulamanın çalıştırılmasını yalnızca Apple'a ait uygulamalarda destekler. 3.0 sürümünden itibaren Apple'a ait olmayan uygulamaların kapalı olduklarında da kullanıcı ile iletişim kurabilmesi için "Push Notification" adı verilen bilgi mesajları gösterimi özelliği eklenmiştir.

3.2.2. Android

Açık kaynak kodlu Android işletim sisteminin arkasında Google firmasının olması bu işletim sistemine güveni artırmaktadır. Fakat programlamada karşılaşılan sorunlar, yeterli dokümanın olmaması, XML editörünün çok temel özelliklere sahip olması karşılaşılan en önemli sorunlardır [49].

Linux çekirdeğine sahip bir işletim sistemidir. Linux kerneli içerdiği için oldukça ileri düzeyde bir altyapıya sahip olmasına rağmen güvenlik ve ticari sebeplerle bir takım özellikleri kısıtlanmıştır veya doğrudan içermemektedir. Alışılmış Linux kernelli işletim sistemlerinde bulunan bazı temel özellikleri ve parçaları içermediği için genel anlamda bir Linux işletim sistemi olarak kabul edilmez.

Açık kaynaklı bir işletim sistemidir ve bu sebeple bir çok üreticinin ürününde bulunabilir, dolayısı değişik platformlara uyumluluk zorunluluğu olduğu için optimizasyon miktarı rakiplerine kıyasla düşüktür. Android işletim sistemi güncellemeleri için bir bilgisayara bağlı olmaya ihtiyaç duymaz. Sadece parmak etkileşimi ile çalışabilecek şekilde tasarlanmıştır, bu nedenle rakiplerine kıyasla basit ve sade bir arayüze sahiptir ve kolay bir kullanım sağlar. Ana kullanıcı arabirimi, üzerinde çeşitli Widget'ların (ufak programcıklar) çalışmasına imkan sağlayan bir bekleme ekranıdır, ayrıca uygulamaların kısayollarının bulunduğu bir menü sistemi

mevcuttur. WebKit tabanlı tarayıcısı ile zengin bir internet, geniş codec desteği ile de zengin bir çoklu ortam deneyimi sunar.

Platforma doğal olmayan, özelleştirilmiş ve benimsenmiş standartların dışında bir dilde uygulama geliştirilmesine izin verir. Android işletim sistemlerinde Java VM değil, kendine özgü Dalvik VM kullanılır. Bu özel dilde yazılan uygulamaların çalışması için bir sanal makineye ihtiyaç duymaları ve bu sanal makinenin düşük bellek yönetim becerisi sebebiyle uygulama başarımlarının performansı rakiplerine kıyasla düşüktür. Kaliteli ve hızlı uygulamalar için yüksek işlemci gücüne ihtiyaç duyar.



Resim 3.2. Android işletim sistemi ekranı

Android'in güzel özelliklerinden bir tanesi de, geliştirilecek programların rahatlıkla Google uygulamaları ile senkronize çalışmasıdır. Android, arkasındaki dev şirketler ve piyasaya yeni çıkan telefonlar ile piyasada hızla yükselmektedir. Bunun temel sebepleri arasında, SDK'sının ücretsiz olması, Android'in üreticiler açısından bir maliyetinin olmaması ve en önemlisi isteyen şirketlerin bu işletim sistemini alıp kendi ürünleri için özelleştirdikten sonra piyasaya sürüyor olması sıralanabilir.

3.2.3. Windows mobil

Windows mobile platformu, birçok cihazda ve birçok operatörle çalışan bir işletim sistemidir. Dell, HP, Palm, I-mate ve son zamanlarda Nokia gibi firmalar, Windows Mobile işletim sistemini kullanan cihazlar üretmektedirler. Windows CE tabanlı bir işletim sistemidir. Windows tabanlı olmasına ve ileri düzeydeki altyapısına rağmen masaüstünde kullanılan Windows uygulamalarını çalıştıramaz. Kapalı kaynak kodlu bir işletim sistemidir fakat bir çok üreticinin ürününde bulunabilir, dolayısı değişik platformlara uyumluluk zorunluluğu olduğu için optimizasyon miktarı rakiplerine kıyasla düşüktür. İşletim sistemi güncellemeleri için bir bilgisayara bağlı olmaya ihtiyaç duyar.



Resim 3.3. Windows mobile işletim sistemi ekranı

Öncelikle stylus ile yönetilen PDA (Personal Digital Assistant-Kişisel Sayısal Yardımcı) kökenli cihazlarda çalışmak üzere geliştirildiğinden, rakiplerine kıyasla daha zor ve karışık bir arayüze sahiptir. Kullanımı rakiplerine kıyasla zordur. Ana kullanıcı arabirimi, üzerinde çeşitli uygulamalara dair kısayolların ve pratik bilgilerin bulunduğu bir bekleme ekranıdır, ayrıca uygulamaların kısayollarının bulunduğu bir menü sistemi mevcuttur. Internet Explorer Mobile tarayıcısı ile internet deneyimi

sunar ve JavaME destekler. Aynı anda birden çok uygulamanın çalıştırılmasını destekler.

Standart ve yaygın bir şekilde benimsenmiş bir uygulama geliştirme ortamı sunar, bu sebeple mevcut uygulama çeşitliliği ile sayısı açısından zengin bir platformdur. Ayrıca kendisine ait bir uygulama dağıtım kanalına (Windows Marketplace for Mobile) sahiptir. Platforma doğal bir dilde uygulama geliştirilmesine izin verir ve bu nedenle uygulama başarımlarının performansı rakiplerine kıyasla yüksektir, fakat bir takım donanım/yazılım platform farklılıkları sebebiyle kaliteli ve hızlı uygulamalar için yüksek işlemci gücüne ihtiyaç duyar.

Windows Mobile kullanan PDA'lara Pocket PC denir. Windows Mobile ile birlikte Word Mobile, Excel Mobile, PowerPoint Mobile gibi programlar yüklü gelir. Bu programlar masaüstü sürümleri kadar özellik içermese de temel görüntüleme ve düzenleme işlemleri için yeterlidir. PowerPoint Mobile ise sadece sunumları görüntülemek için kullanılabilir, düzenlemeye izin vermez. Outlook ve Windows Media Player programları da Windows Mobile ile yüklü gelir [46].

3.2.4. BlackBerry

Research in Motion (RIM) tarafından sadece BlackBerry'ler için geliştirilmiş özel bir işletim sistemidir. Kapalı kaynaklı bir işletim sistemidir ve bu nedenle yalnızca RIM'a ait ürünlerde bulunur. Değişik platformlara uyumluluk zorunluluğu olmadığı halde optimizasyon miktarı düşüktür. İşletim sistemi güncellemeleri için bir bilgisayara bağlı olmaya ihtiyaç duymaz.

Temelde dokunmatik olmayan cihazlar için tasarlanmıştır ve daha sonra dokunmatik ekranları desteklemeye başlamıştır. Basit bir arayüze sahip olsa da rakiplerine kıyasla zor bir kullanım sağlar. Ana kullanıcı arabirimi, üzerinde çeşitli uygulamalara dair kısayolların ve pratik bilgilerin bulunduğu bir bekleme ekranıdır. Ayrıca uygulamaların kısayollarının bulunduğu bir menü sistemi mevcuttur. E-mail deneyimi ve JavaME desteği sunar.



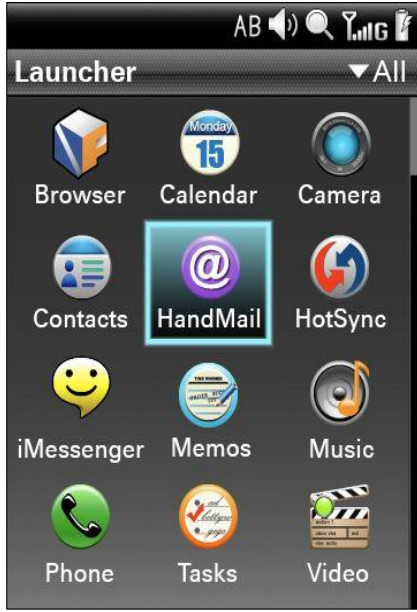
Resim 3.4. BlackBerry işletim sistemi ekranı

Java tabanlı, JavaME ve özel API'ler içeren uygulama geliştirme ortamı (BlackBerry SDK) sunmaktadır. Ayrıca kendisine ait bir uygulama dağıtım kanalına (BlackBerry App World) sahiptir. Platforma doğal bir dilde uygulama geliştirilmesine izin verir ve bu sebele uygulama başarımlarının performansı rakiplerine kıyasla yüksektir. Kaliteli ve hızlı uygulamalar için yüksek işlemci gücüne ihtiyaç duymaz. Aynı anda birden çok uygulamanın çalıştırılmasını destekler.

3.2.5. Palm OS

1996'de piyasaya sürdüğü ilk Palm Pilot'tan sonra Palm OS platformu birçok iş ve eğlence amaçlı taşınabilir cihazda kullanılmıştır.

Şu anda, PDA'lar üzerinde Palm OS ve Windows CE olmak üzere iki ayrı işletim sistemi çalışmaktadır. PDA aslında bir bilgisayardır. Palm mimarisi ile Windows CE mimarisi arasındaki en büyük fark, Palm işletim sisteminin PDA üzerinde yeni yazılımların geliştirilmesi için bedava yazılım geliştirme donatıları ve dokümanları sağlamasıdır. Başka platformlar için de yazılım sağlanmaktadır. Böylece geliştirme Windows ya da Linux ile sınırlandırılmamış, platform bağımsız hale dönüşmüş olmaktadır. [50].



Resim 3.5. Palm OS ekranı

Linux kerneline sahip bir işletim sistemidir. Kapalı kaynaklı bir işletim sistemidir ve bu nedenle yalnızca Palm'a ait ürünlerde bulunur. İşletim sistemi güncellemeleri için bir bilgisayara bağlı olmaya ihtiyaç duymaz. Sadece parmak etkileşimi ile çalışabilecek şekilde tasarlanmıştır. Ana kullanıcı arabirimi, üzerinde çeşitli kısayolların bulunduğu bir bekleme ekranıdır, ayrıca uygulamaların kısayollarının bulunduğu bir menü sistemi mevcuttur. Çoklu dokunuş desteği mevcuttur.

3.2.6. Symbian OS

Symbian Ltd., önceleri Ericsson, Motorola, Nokia, Panasonic, Psion, Samsung, Siemens ve Sony Ericsson şirketlerinin bazıları arasında kurulmuş bir şirket idi. Daha sonra Nokia, diğer hissedarların elindeki hisseleri de satın almış ve Symbian'ın tüm varlığını kendi kurduğu Symbian Vakfı'na bağışlamıştır. Şu an geliştirme işini Symbian Vakfı yapmaktadır. Ancak Nokia firması da artık Symbian işletim sistemini terk etmek üzeredir. Windows mobil işletim sistemini kullanan telefonları piyasalara sunulmuş ve kullanılmaktadır.

Nokia'nın Symbian cihazlarında kullandığı arayüze Series 60 veya S60 ve S80, Sony Ericsson'un Symbian cihazlarında kullandığı arayüze UIQ, Fujitsu, Sony Ericsson

Japonya, Mitsubishi ve Sharp'ın NTT DoCoMo şebekesinde çalışacak Symbian cihazlarda kullandığı arayüze MOAP(S) ismi verilmiştir.



Resim 3.6. Symbian işletim sistemi ekranı

Kendisine ait bir uygulama dağıtım kanalına (Ovi Store) sahiptir. Platforma doğal bir dilde uygulama geliştirilmesine izin verir ve sahip olduğu mikrokernel yapısı sayesinde kaliteli ve hızlı uygulamalar için yüksek işlemci gücüne ihtiyaç duymaz. Aynı anda birden çok uygulamanın çalıştırılmasını destekler.

Symbian işletim sistemi altında çalışan programlar ile SMS, MMS, WAP, kızılötesi ve bluetooth portlar ve hatta telefondaki takvim ve rehber uygulamaları gibi telefonun tüm özelliklerine erişmek mümkündür [46].

3.2.7. Bada

Samsung'un mobil işletim sistemi bada, 3. Nesil kablosuz bağlantı yeteneklerini desteklemesinden çok arayüzü için minimum çoklu dokunmatığe duyarlı 800x400 WVGA ekran gereksinimiyle dikkat çekmektedir.

Samsung Bada, kernel seviyesinde konfigüre edilebilir, C++ API desteği olan gelişmeye uygun bir işletim sistemidir. Kernel-configurable olmasından dolayı Linux ya da herhangi bir RTOS (Real Time Operating System - Gerçek zamanlı işletim

sistemi) üzerinde çalışabilmektedir. Ancak Samsung yeni ürettiği cihazlarda Android işletim sistemini de kullanmaktadır.



Resim 3.7. Samsung bada işletim sistemi

3.3. Mobil Cihazlar

Geleneksel cihazlardan farklı olarak mobil cihazlar iki temel ayırt edici özelliğe sahiptirler. Bunlar; taşınabilirlik ve kapasitedir (bellek, işlemci, hız, ekran boyutu vb.).

Kullanıcılar bir yandan daha küçük ve daha hafif taşınabilir cihazlar tercih ederken, diğer yandan grafiksel ve matematiksel uygulamalar için geniş ekranlar ve güçlü işlemler için de yüksek işlemci kapasitesine ihtiyaç duymaktadırlar [51]. Cep telefonları ve PDA tabanlı cihazları genel olarak mobil cihazlar olarak tanımlanmaktadır. Cep telefonları fiziksel görünümünün yanında yeni teknolojilerle zenginleşirken, PDA tabanlı cihazların yeteneklerini de bünyelerinde barındırarak melez cihazlar olarak ortaya çıkmışlardır. Klasik cep telefonları sınırlı bir işletim sistemine sahip iken, akıllı telefon (smartphone) adı verilen cihaz teknolojisi bu cihazlar için üretilmiş ve kapsamlı işlevleri barındıran işletim sistemlerini beraberinde getirmiştir. Bu tür cihazlar, mp3 dinlemekten, ebook adı verilen kitapları

okumaya dek çok geniş bir platformdaki ihtiyaçları bir arada sunmak şeklinde ciddi çoklu ortam ihtiyaçlarına tek bir cihazla cevap vermektedir [52].

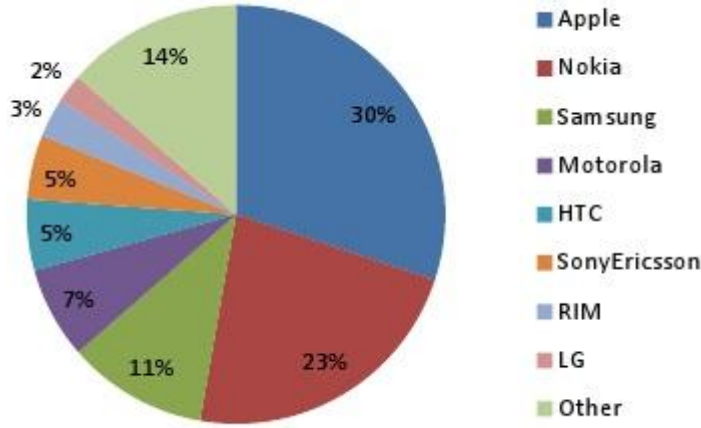
Notebook veya dizüstü sistemler taşınabilir olmalarına rağmen farklı bir teknoloji çerçevesine işaret ettiklerinden dolayı bu bağlamda düşünülmemiştir. Ancak Tablet PC'ler için durum biraz daha farklıdır. Piyasaya çıktığı ilk andan itibaren yoğun ilgi görmeye başlayan Tablet PC'ler hem taşınabilirliği çok ilerletmeleri, hem de notebook ile akıllı cep telefonu karışımı bir yapıda olmalarıyla, mobil cihazlar arasında hızla yerini alan ve gittikçe yaygınlaşan cihazlar haline gelmiştir. Tablet PC'lerin dizüstü bilgisayarlar gibi klavyesi yoktur. Ekranı dokunmatiktir ve dijital bir kalem yardımıyla tüm işlemler gerçekleştirilebilir. Ayrıca kablosuz ağ desteğiyle, internete her yerden bağlanabilme olanağı da sunmaktadır.



Resim 3.8. Örnek bir Tablet PC

Mobil teknolojilere duyulan ilginin çok olması sonucu, mobil iletişim cihazlarını üreten firmalar kendilerini büyük bir rekabet ortamının içerisinde bulmuşlardır. Admob tarafından tutulan istatistiklere göre Mayıs 2010 tarihli verilerde dünyada en

çok tercih edilen üretici firma Apple olarak görülmektedir. Şekil 3.3.'te dünyada tercih edilen mobil cihazların oranları görülmektedir.



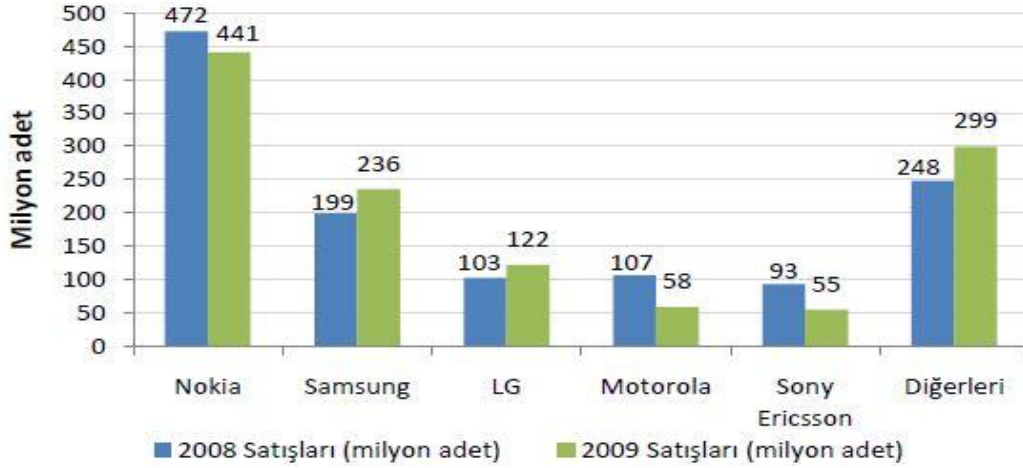
Şekil 3.3. Mobil cihaz üreticileri oranları [53]

Yeni nesil cep telefonları eski teknolojilere göre kullanıcılarına daha fazla yeni ve kullanılabilir teknolojinin yanında daha fazla cihaz hâkimiyeti sunmaktadır. Kullanım kolaylığının artması (arayüz yazılımlarında ve donanımdaki iyileştirmeler) cep telefonlarının kullanım alanlarını ve bu tür cihazlar için üretilen ek donanım-yazılım teknolojilerinin gelişim sürecini ciddi bir dinamizm sürecine sokmuştur [52].

Mobil telefonlar en yaygın ve gelişmiş PIM (Personal Information Management) araçlarını, ağ ve diğer bilgisayar olanaklarını içine alacak şekilde tasarlanmış cihazlar içerir. Bu tür cihazlar, akıllı telefon bazlı tasarımları nedeniyle sınırlı bir ekran genişliğinde olup, yine sınırlı bir bellek kapasitesi ve işlemci hızına sahiptir. Bu tür cihazlar GSM/GPRS gibi WAN (Wide Area Network) ya da 3G hücreli sistemlerle internet erişimi sağlayabilir ve modellerine bağlı olarak Bluetooth protokol, USB ya da kızılötesi bağlantı üzerinden diğer çevresel cihazlar iletişim kurabilirler [51].

Gartner'in 2010 Mayıs ayında yayınlanan araştırmasına göre, 2010 yılının ilk çeyreğindeki mobil telefon satışları 2009 yılının aynı çeyreğine göre %17 artarak 314,7 milyona ulaşmıştır. Aynı araştırmaya göre Smartphone satışları 54,3 milyona ulaşmış ve 2009 yılı aynı dönemine göre %48,7 artış kaydederek mobil telefon

cihazları piyasasının büyümesinde itici güç olmuştur [47]. Dünya genelindeki mobil telefon satışları Şekil 3.4.'te gösterilmiştir.



Şekil 3.4. Dünya genelindeki mobil telefon satışları [47]

Dünya genelinde cep telefonu üreticileri ve bunların yıllara göre satış oranları Şekil 3.5.'te belirtilmiştir.

Mayıs 2008			Mayıs 2009			Mayıs 2010		
Brand	Model	% of Requests	Brand	Model	% of Requests	Brand	Model	% of Requests
Motorola	RAZR V3	5.1%	Apple	iPhone	18.6%	Apple	iPhone	18.8%
Motorola	KRZR K1c	4.1%	Apple	iPod touch	12.8%	Apple	iPod touch	11.0%
Motorola	W385m	2.9%	Samsung	R450	2.2%	Motorola	Droid	3.2%
Nokia	N70	2.7%	Motorola	RAZR V3	1.9%	Samsung	SCH R350	1.6%
RIM	Blackberry 8100	1.6%	Nokia	N70	1.9%	Nokia	5130	1.5%
Motorola	Z6m	1.5%	Nokia	3110c	1.3%	HTC	Magic	1.4%
Kyocera	K24	1.3%	HTC	Dream	1.3%	Nokia	3110c	1.4%
Palm	Centro	1.3%	Nokia	6300	1.2%	Nokia	N70	1.2%
Nokia	N73	1.2%	Motorola	Z6M	1.0%	HTC	Hero	1.1%
Nokia	N95	1.1%	Nokia	N80	1.0%	Nokia	6300	1.1%
Total		22.8%	Total		43.2%	Total		42.2%

Şekil 3.5. Firmalar ve yıllara göre satılan cep telefonu markaları ve oranları [53]

PDA olarak tasarlanmış mobil iletişim özelliklerini barındıran akıllı telefon cihazları, Wireless Markup Language yada Extensible HyperText Markup Language formatındaki hypertext sayfaları göstermek için cep telefonlarına kıyasla uygun olan daha geniş ekranlara sahiptirler. Bunlara ilave olarak; e-posta, SMS ya da MMS mesajlar, hatta daha büyük dokümanlar oluşturmak, el yazısı tanıyabilme,

dokunmaya duyarlı ekran klavye ve hatta katlanabilir klavyeler sayesinde işlemler PDA cihazlar ile daha kolayca gerçekleşir. PDA'lar bellek, işlemci kapasiteleri, ekran boyutları ve geniş ağlara bağlanabilirlik özellikleriyle, mobil telefonlardan daha güçlü cihazlardır [51].

3.4. GSM Sistemi

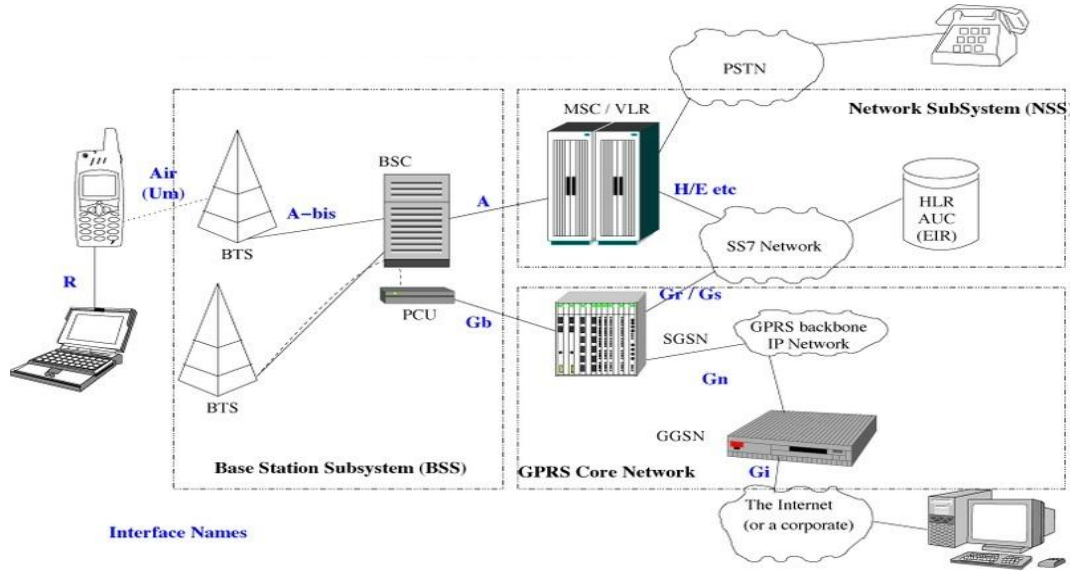
Global System for Mobile Communication (GSM) European Telecommunications Standarts Institute (ETSI)'ye bağlı bir çalışma grubu tarafından geliştirilmiş iletişim sistemidir. GSM dünyada ilk olarak Finlandiya'da kullanılmaya başlanmıştır. Finlandiya'nın coğrafi yapısı, hava şartları, yerleşim alanlarının dağılımı kablosuz haberleşme sisteminin geliştirilmesi için uygun bir zemin hazırlamıştır [54].

Yaklaşık 10 yılda ve milyarlarca dolarlık harcamayla oluşturulabilen GSM, 1987 yılında, 30 Avrupa ülkesi tarafından standart olarak kabul edilmiştir. İlk GSM standartları 1990 yılında yayınlanmıştır. Bir Avrupa standardı olarak başlamasına rağmen GSM kısa sürede dünyaca benimsenerek bir dünya standardı haline gelmiştir. ABD'de bu standartların kullanılabilmesi için ANSI (American National Standards Institute - Amerikan Ulusal Standartlar Enstitüsü) devreye girmiş ve GSM standartları ABD'de de yayınlanarak uygulamalarına bağlanmıştır [55].

GSM gibi hareketli telefon ağının öncelikli hedefi hareketli istasyona sahip olan abonelerin her yerden telefon görüşmesi yapabilmesine imkan sağlamaktır. Bunu başarabilmek için çağrı yönetimi, çağrı işleme, radyo yönetimi, mobilite yönetimi, ücretlendirme ve güvenlik gibi başlıca fonksiyonların yerine getirilmesi gerekmektedir [54].

GSM iki frekans bandı ile çalışır: Konuşmaların cep telefonları üzerinden baz istasyonlarına taşındığı 890–915 MHz'lik alt bant ve 935–960 MHz'lik karşı yön için kullanılan üst banttır. Her frekans bandında 124 kanal vardır. Her kanal aynı zamanda maksimum 8 kanal taşıyabilir. Bir baz istasyonunun eş zamanlı olarak taşıyabildiği maksimum konuşma sayısı yaklaşık 1000'dir [44].

GSM ağı, ağ anahtarlama alt sistemi (NSS; Network Switching Subsystem) ve baz istasyonu alt sisteminden (BSS; Base Station Subsystem) oluşur. NSS çağrı işleme, mobilite yönetimi, güvenlik ve ücretlendirme gibi işlemleri yapar. BSS ise hareketli istasyonlara yönelik olan radyo sistemi ile ilgili fonksiyonları yönetir [54].



Şekil 3.6. GSM mimarisi

Günümüzde kullanılan biçimi ile GSM sisteminin abonelerine sunduğu servis ve avantajlar şu şekilde özetlenebilir [56]:

- Herhangi bir GSM şebekesinin abonesi, artık çoğumuz tarafından kullanılan küçük boyutlara sahip cep telefonu ile GSM sistemine dâhil olan ve kapsama alanı içinde bulunan bütün ülkelerde her yeri arayıp telefon görüşmesi yapabilmekte, nerede olursa olsun aranabilmektedir.
- Kullandığı telefonun kendisine ait olması gerekmemektedir, çünkü abonelik bilgilerini içeren SIM (Subscriber Identity Module - Abone Tanıtım Modülü) kartını yanında taşıması kiralayabileceği ya da ödünç alabileceği herhangi bir GSM el telefonunu kullanması için yeterli olacaktır.
- Ayrıca abone kendisine ait telefonu belirli çağrılara (örneğin, daha pahalı olan uluslararası çağrılara) kapatabilmekte, erişilemediği ya da meşgul olduğu durumlarda çağrı yönlendirme özelliği kazandırılabilir.

Üstelik bütün bunları GSM sistemine özgü güvenlik önlemlerinin sağladığı gizlilik içinde gerçekleştirebilmektedir.

GSM, sayısal olması ve kullandığı radyo erişim teknikleri açısından birçok yenilikler getirmesine karşılık erişilen son nokta olmayıp, daha gelişmiş ve küresel sistemlere geçişi sağlayacak bir çalışma olarak kabul edilmektedir. İkinci nesil GSM sistemlerindeki gelişmeleri, her biri daha hızlı veri iletimine imkan sağlayan HSCSD (High Speed Circuit Switched Data - Yüksek Hızda Devre Anahtarlamalı Veri), GPRS (General Packet Radio Service - Genel Paket Radyo Hizmeti) ve EDGE (Enhanced Data Rate for GSM Evolution - GSM Evrimi için Geliştirilmiş Veri Hızı) sistemleri izlemektedir [52].

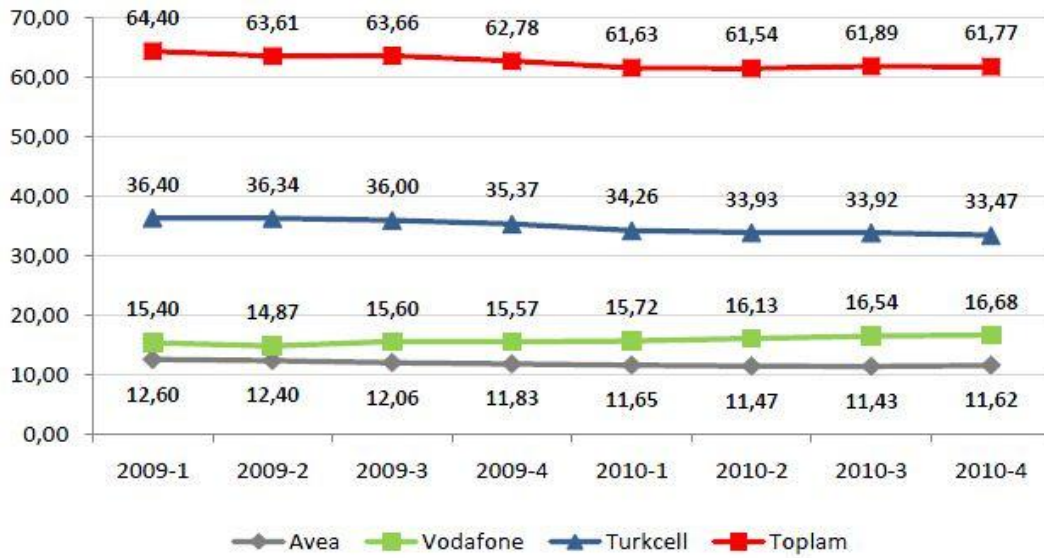
Abonelerle ilgili işlemler ise birçok bölüme ayrılmıştır. Kalıcı konum kütüğü (HLR; Home Location Register), ziyaretçi konum kütüğü (VLR; Visitor Location Register), kimlik denetim merkezi (AUC; Authentication Center), cihaz kimlik kütüğüdür (EIR; Equipment Identity Register). EIR, HLR ve VLR abonelerin bilgilerinin bulunduğu veri tabanlarıdır. HLR, belirli bir operatör için cep telefonunun nerede olduğu, abone kimlik bilgileri gibi her türlü veriyi tutan veri tabanıdır. VLR ise belirli bir MSC sınırları içinde servis verilen abonelerin HLR'deki bilgilerinin kopyasını geçici olarak tutar. Abone bilgilerinin HLR'den VLR'e kopyalamak ağdaki mobilite yönetiminin bir parçasıdır. AUC ise bütün güvenlik operasyonlarını yönetir. EIR, şebeke içerisindeki hareketli istasyonların listesinin tutulduğu veri tabanıdır. Burada tüm hareketli istasyonların uluslararası mobil cihaz kimliği (IMEI; International Mobile Equipment Identity) bulunur [54].

GSM'in en kullanışlı özelliklerinden birisi, kullanıcıların aynı hat ile değişik ülkelerden görüşme (roaming) yapabilmeleridir. Tüm GSM standartları hücrel ağ kullanmakta olup, GSM teknolojisi, hücreler arası geçiş yapabilme kabiliyetine sahiptir.

3.4.1. Türkiye’de kullanılan GSM şebekeleri

Tüm dünyada olduğu gibi ülkemizde de mobil telekomünikasyon sektörü hızlı bir gelişme göstermektedir. Ülkemiz sadece yerli yatırımcıların değil aynı zamanda yabancı sermayenin de ilgisini çekerek önemli bir pazar haline gelmiştir.

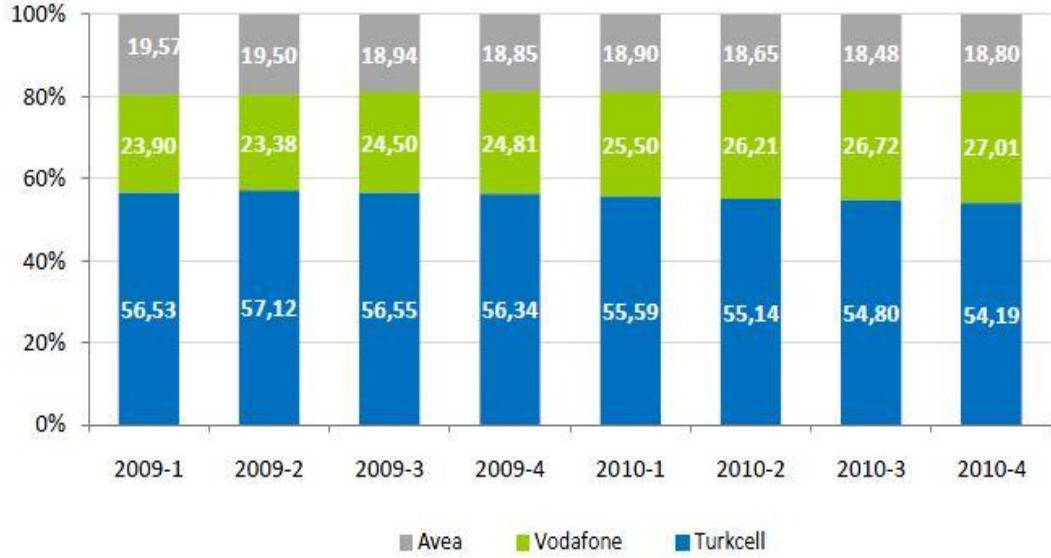
Türkiye’de mobil telekomünikasyon alanında 3 işletmeci faaliyet göstermektedir. Bunlar Turkcell, Vodafone ve Avea’ dır. Bu işletmelerin yıllar bazında abone sayıları Şekil 3.7.’de gösterilmektedir.



Şekil 3.7. İşletmeci bazında toplam abone sayıları, (milyon) [47]

2010 yılının dördüncü üç aylık döneminde, önceki üç aylık döneme göre abone sayıları bakımından Avea’da %1,66, Vodafone’da %0,85 artış görülürken, Turkcell’de %1,33 oranında azalma gerçekleşmiştir (Şekil 3.7.).

Şekil 3.8.’de mobil işletmecilerin abone sayılarına göre pazar paylarına yer verilmektedir. 2010 yılı son üç aylık dönem itibariyle abone sayısına göre Turkcell’in %54,19, Vodafone’un %27,01, Avea’nın ise %18,80’lik paya sahip olduğu görülmektedir.



Şekil 3.8. Mobil işletmecilerin abone sayılarına göre pazar payları (%) [47]

3.4.2. GSM şebekeleri ve veri madenciliğini kullanım nedeni

Ülkemizde hizmet veren GSM Şebekelerinin, mevcut müşterilerini korumak ve yeni müşteriler kazanmak için yaptığı kampanyalar hemen hemen her gün değişmekte ve artmaktadır. Özellikle numara taşınabilirliğine imkan sağlandıktan sonra rekabet piyasası daha da artmıştır. Böyle bir ortamda, GSM Şebekelerinin en çok vurguladığı reklam konusunu; kapsama alanı, çekim gücü ve fiyat avantajı gibi kavramlar oluşturmaktadır.

Bu durum göstermektedir ki insanlar için, kullandığı GSM Şebekesinin çekim gücü yüksek olmalı ve fiyatları da uygun olmalıdır. Tüm Şebekelerin yaptıkları reklamlardan anlaşılmaktadır ki, ülkemizin %100'üne çok yakın bir bölgesi, tüm şebekeler tarafından kapsama alanına alınmıştır. Peki bu kapsama alanındaki bölgelerde insanların elde ettiği "sinyal gücü" hangi şebekede, hangi orandadır? Bu tez çalışmasında cep telefonlarının sinyal verileri, geliştirilen bir program ile telefonda alınarak, veri madenciliği yöntemleriyle değerlendirilmiştir. Çünkü her ne kadar tüm şebekeler kapsama alanlarının yüksek oranlarını duyursalar da bölgesel değişiklikler cep telefonlarının sinyal alım gücüne etki etmektedir.

4. VERİ MADENCİLİĞİNİN GSM ŞEBEKELERİNE UYGULANMASI

Bu bölümde, veri madenciliğinin GSM Şebekelerinin sinyal kalitelerine uygulanmasıyla ilgili veriler aktarılmıştır. Mobil iletişimin hızla yayıldığı günümüz teknolojisinde, ortaya çıkan veriler hem kullanıcı tarafında, hem de sunucu tarafında çok büyük ölçeklere ulaşmaktadır. Bunun sonucu olarak, bu verilerin içerisinde, anlamlı ve yorumlanabilir verilere ulaşmak giderek daha da zorlaşmaktadır.

4.1. Tasarım

Bu çalışmada hazırlanan program sayesinde, mobil iletişim aracı olan cep telefonundan alınan sinyal gücü verileri, veri madenciliği yöntemiyle incelenmiş ve yorumlanmıştır. Programın çalıştırılarak sinyal bilgilerinin elde edildiği güzergah prototip bir bölgesel çalışmadır. İstenildiği takdirde, program başka bölgelerde de kullanılarak sinyal bilgileri alınabilir. Program geliştirilirken kullanılan yazılım ve donanım teknolojileri aşağıda belirtilmiştir.

4.1.1. Program tasarımında kullanılan donanım teknolojileri

Cep telefonundan sinyal bilgisi alabilmek için planlanan güzergahta kullanılmak üzere, mobil bilgisayar sistemi gerekmektedir. Program çalıştırılırken kullanılan bilgisayar sistemi ve diğer cihazlar:

- Bir dizüstü bilgisayar,
- Cep Telefonu,
- USB portu bilgisayara bağlantı kablosu,
- GSM Şebekeleri SIM Kartları'dır.

4.1.2. Program tasarımında kullanılan yazılım teknolojileri

GSM Şebekelerinin sinyal verilerini elde etmek için hazırlanan program tasarımı için, MS Visual Studio C#.NET 2008 programlama dili kullanılmıştır. Bu programın çalışması ve sinyal verilerinin elde edilmesi aşamaları şu şekildedir:

- Bilgisayar ve cep telefonu, bluetooth veya seri bağlantı kablosu yardımıyla eşleştirir.
- Eşleştirilen cep telefonu için bilgisayarda bir port numarası belirlenir.
- Daha sonra hazırlanan uygulama programı çalıştırılır.
- Program içerisinde, cep telefonu için atanan port numarası seçildikten sonra cep telefonundan veri talebi yapılır.
- Bu talep ile beraber, cep telefonları ve modemler için özel olan AT komutları çalıştırılarak, istenilen bilgiler cep telefonundan alınır.
- Alınan bilgiler hem program arayüzünde ekrana, hem de program klasöründe bir dosyaya kaydedilir.

Ayrıca kullanılan cep telefonunun, USB port veya Bluetooth ile bilgisayara bağlantı sırasında sorunsuz çalışması için bilgisayara, USB Modem ve Bluetooth Modem sürücülerini kurulmuştur. Bu sürücüler sayesinde bağlantı sırasında cep telefonu için bir haberleşme portu ayrılmakta ve iletişim bu porttan sağlanmaktadır.

4.2. Programın Amaç ve Kapsamı

İşletmelerde yönetici ve idareciler, gelecek için alacakları kararları daha sağlam temellere dayandırmak istemektedirler. Bu temellerin dayanak noktası ise işletmelerin daha önceki çalışmalarından elde ettikleri verilerdir. Bu verileri kendilerine en iyi şekilde sağlayan disiplin de veri madenciliğidir.

Ülkemizde mobil hat abone sayısının neredeyse ülke nüfusu kadar olduğu düşünüldüğünde, mobil iletişimin insanlar için ne kadar önemli bir konuma geldiği anlaşılacaktır. Böylece, mobil iletişim hizmeti sunan işletmeler için çok çetin bir rekabet ortamı ortaya çıkmıştır. GSM Şebekeleri, mevcut müşterileri korumak ve yeni müşteriler elde etmek için çeşitli stratejiler izlemek, kampanyalar düzenlemek ve insanların istekleri doğrultusunda seçenekler sunmak zorundadırlar.

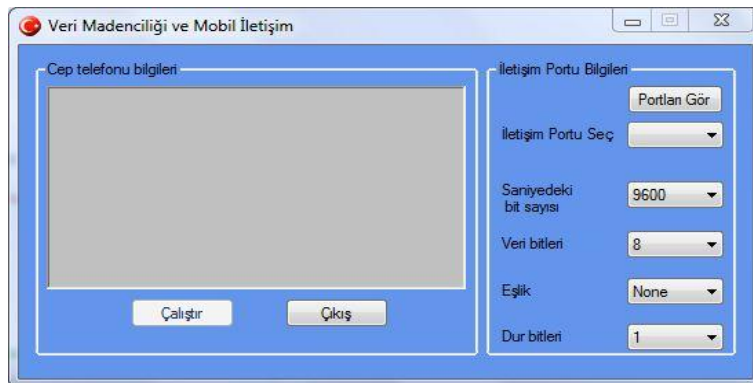
GSM Şebekelerinin mevcut müşterilerini korumak ve bunlara yeni müşteriler eklemek için yaptıkları tanıtım reklamları incelendiğinde, "Kapsama Alanı", "Çekim

Gücü” ve “Sinyal Kalitesi” kavramlarının en çok vurgulanan ve en önemli konuların başında geldiği kolaylıkla görülebilir. Çünkü kaliteli ve kesintisiz iletişim, tüm insanların isteyeceği özelliklerin başında gelmektedir.

Bu tez çalışmasında hazırlanan program, mobil iletişim aracının iletişim sinyallerini toplayarak bir dosyaya kaydedip, bu verileri veri madenciliği yöntemleri ile yorumlayarak; haberleşme sinyalinin kalitesi, gücü ve kapsama alanı hakkında fikir edinme ve öngörüler oluşturma imkanı sunmaktadır. Baz istasyonlarında log kayıtları ve iletişim bilgileri tutulmasına rağmen, işletmelerin kullanıcı tarafındaki sinyal bilgilerini almaları ve bu bilgiler doğrultusunda iyileştirme yapmaları mümkün değildir. Bu sebeple bu çalışmada kullanıcı bazlı bir ölçüm gerçekleştirilmesi ve sonuçlarının değerlendirilmesi büyük önem taşımaktadır. Ayrıca bu çalışmada, yöneticiler tarafından karar destek amaçlı en çok kullanılan disiplinlerden birisi olan veri madenciliği ile hayatımızın ayrılmaz bir parçası haline gelen mobil iletişim konuları bir araya getirilmiş ve GSM işletmelerince göz önüne alınması gereken sonuçlar ortaya çıkarılmıştır.

4.3. Program Arayüzü Tanıtımı

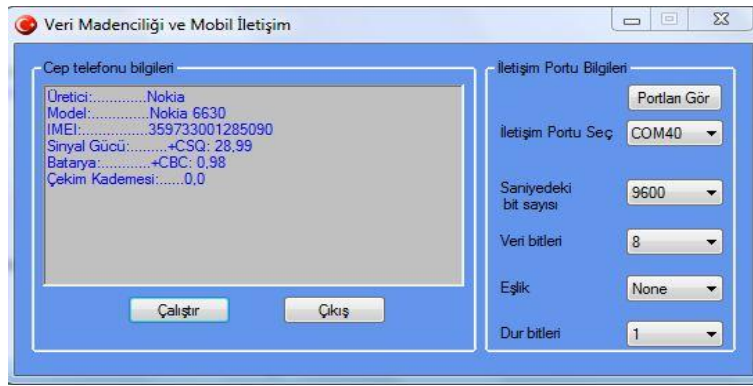
Program, cep telefonu ile bilgisayar arasında bağlantı kurarak, cep telefonundan istenilen bilgilere ulaşılması ve bu bilgilerin dosyaya kaydedilmesi amacıyla tasarlanmıştır. Veri madenciliğinin ana elemanı olan verilere bu şekilde ulaşılması amaçlanmıştır. Programın arayüzü Şekil 4.1.’de gösterilmiştir.



Şekil 4.1. Program arayüzü

Program penceresi iki bölüme ayrılmıştır. Bir bölümde cep telefonu ile bilgisayarın iletişimi için kullanılacak port ile ilgili bilgiler ayarlanmaktadır. Diğer bölümde ise cep telefonundan gelen bilgilerin gösterildiği alan yer almaktadır.

Program çalıştırıldığı ilk anda, çalıştır düğmesi aktif değildir. Öncelikle iletişim portu bilgileri ayarlanmalıdır. Cep telefonuna bağlanacak port seçildikten sonra, çalıştır düğmesi aktif duruma geçmektedir. Program çalıştırılırken ortaya çıkan sonuç Şekil 4.2.'de gösterilmiştir.

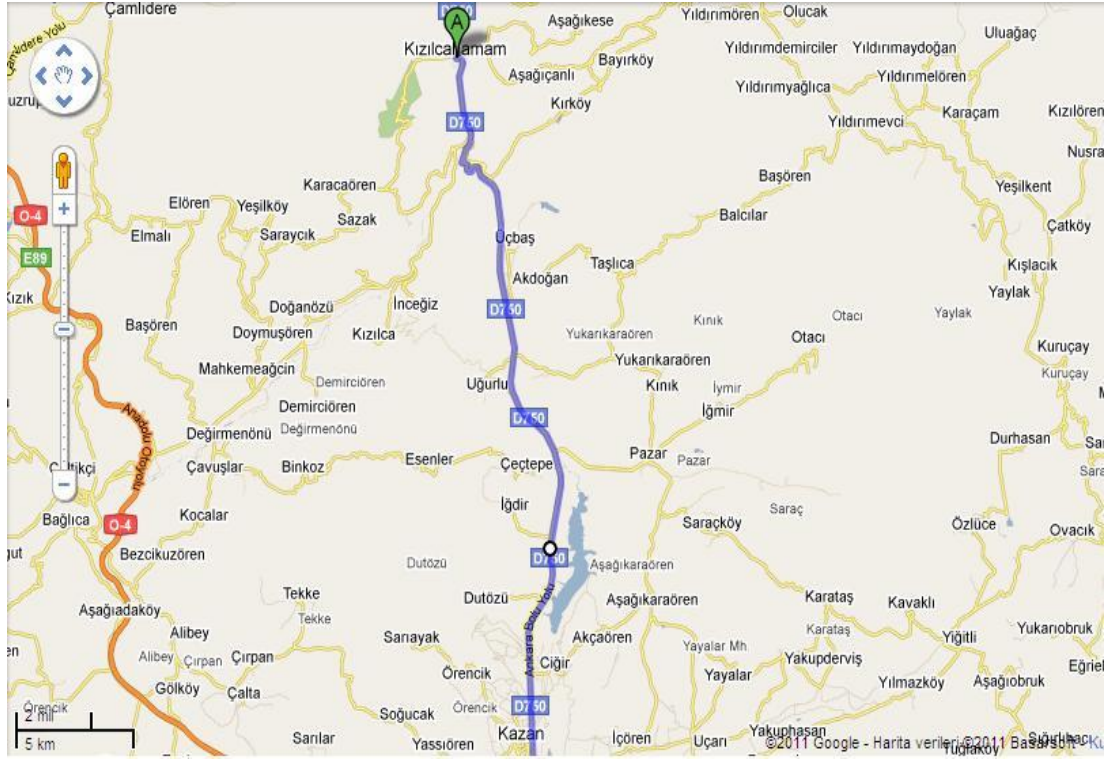


Şekil 4.2. Programın çalışır durumdaki hali

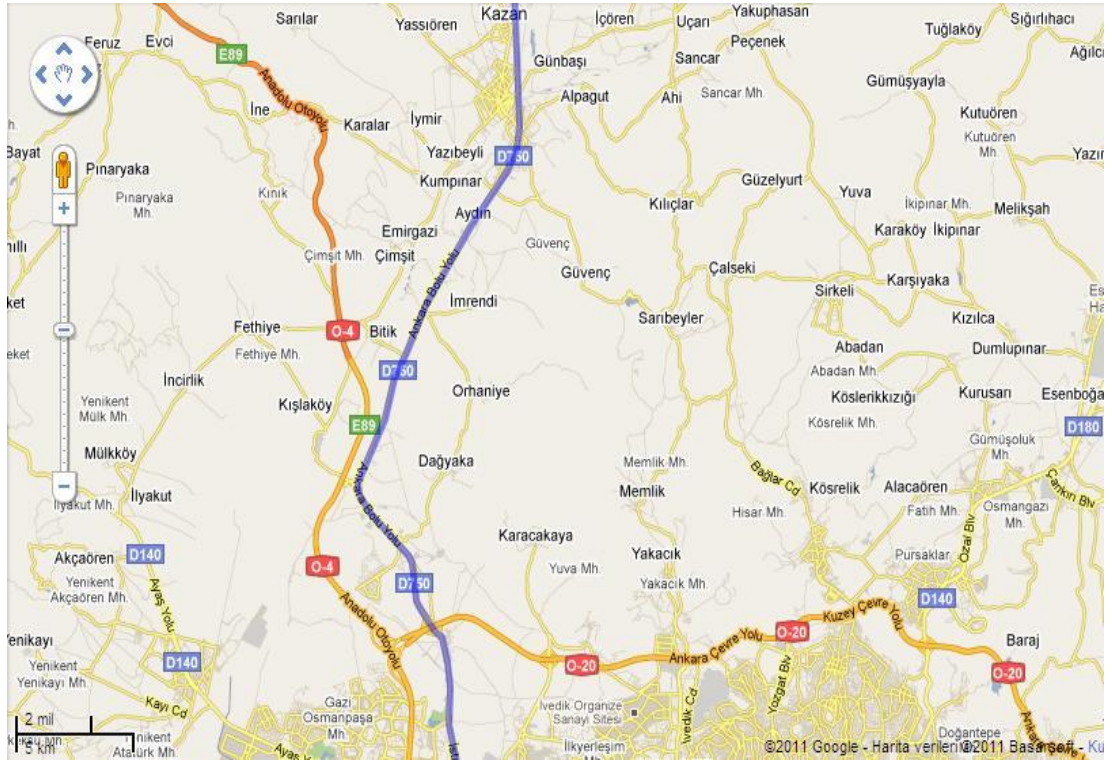
Telefondan istenilen bilgilere AT komutları ile erişildikten sonra bu bilgiler hem program ekranına yazılmakta hem de veriler isimli bir metin dosyasına kaydedilmektedir. Böylece veri madenciliği için gerekli olan veriler elde edilmektedir. Programın akış diyagramı EK-1'de verilmiştir.

4.4. Programın Kullanıldığı Güzergah

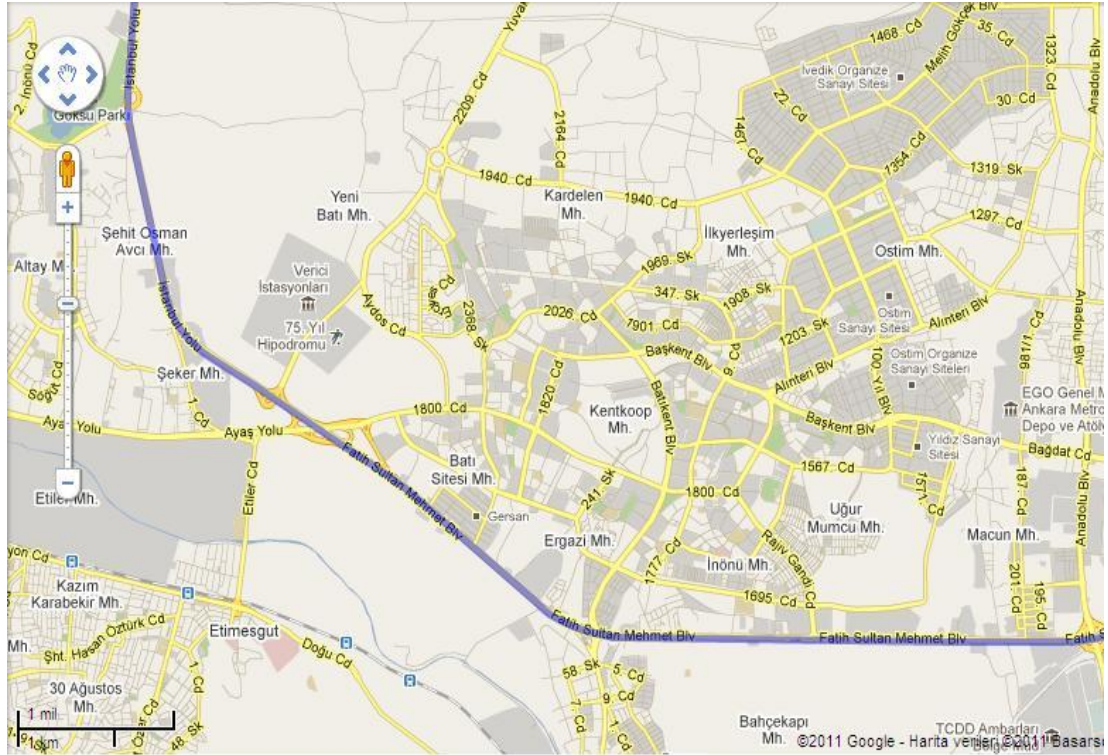
Tasarlanan program, sinyal gücü verilerini elde etmek için önceden gelişi-güzel belirlenen bir güzergahta kullanılmıştır. Seçilen güzergah Kızılcahamam – Gazi Üniversitesi Merkez Kampüsü arasındaki yaklaşık 80 km'lik bir yolu kapsamaktadır. Bu güzergahta bazı bölgeler dağlık arazi, bazı bölgeler de ovalık arazidir. Ayrıca güzergahta, şehir merkezi sayılabilecek bir bölgeden de geçilmektedir. Tüm şebekelerin sinyal bilgilerini elde etmek için aynı cep telefonu kullanılmıştır. Program güzergah boyunca bir araç içerisinde, yasal hız sınırlarında seyir halinde çalıştırılmıştır. Bu güzergahın harita bilgisi aşağıda belirtilmiştir.



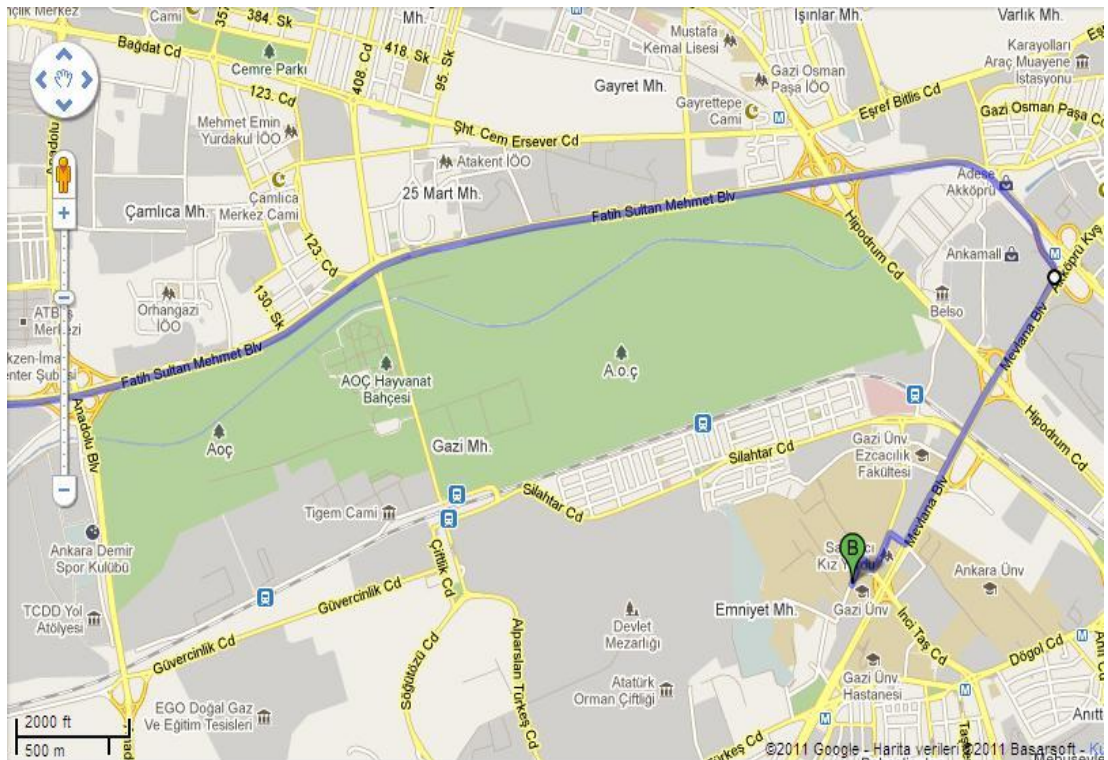
Harita 4.1. Sinyal ölçümü yapılan güzergah (Kızılcahamam – Kazan arası)



Harita 4.2. Sinyal ölçümü yapılan güzergah (Kazan – Gazi Osmanpaşa Mah. arası)



Harita 4.3. Sinyal ölçümü yapılan güzergah (Gazi Osmanpaşa Mah. – Anadolu Bulvarı arası)



Harita 4.4. Sinyal ölçümü yapılan güzergah (Anadolu Bulvarı – Gazi Üniv. arası)

4.5. Veri Madenciliği Süreci

Tasarlanan program, belirtilen güzergah boyunca üç ayrı GSM şebekesi için çalıştırılmıştır. Bu şebekelerin isimleri, yasal sorumluluktan dolayı belirtilmemiştir. GSM Şebekelerinin isimleri bu çalışmanın ileri bölümlerinde “A, B ve C Şebekeleri” olarak anılacaktır. Program çalıştırdıktan sonra elde edilen veriler bir dosyaya kaydedilmiştir.

Bu aşamadan sonra oluşturulan veriler üzerinde veri madenciliği süreci başlatılmıştır. Öncelikle elde edilen veriler içerisinde, gereksiz veriler temizlenmiş ve gerekli olan sinyal verileri düzene sokulmuştur. Böylece verilerin temizlenmesi ve hazırlanması aşaması tamamlanmış ve sinyal tablosu oluşturulmuştur. Daha sonra bu tablo, sinyal gücü ve sinyal kalitesine göre sınıflandırılmıştır. Tabloya değinmeden önce belirtilmesi gereken bir konu da sinyal gücünün sınıflandırılmasıdır. Sinyal gücünün sınıflandırılması Çizelge 4.1.’de gösterilmiştir.

Çizelge 4.1. Sinyal gücünün sınıflandırılması

Sinyal Gücü Değeri	dB Karşılığı (desibel)	Sınıflandırması (kapsama)
0...6,99	-101 veya daha düşük	Yetersiz
7...11,99	-100 ...-91	Zayıf
12...16,99	-90 ... -81	Orta
17...31,99	-80 veya daha yüksek	İyi
,99	-	Bilinmiyor veya algılanamıyor

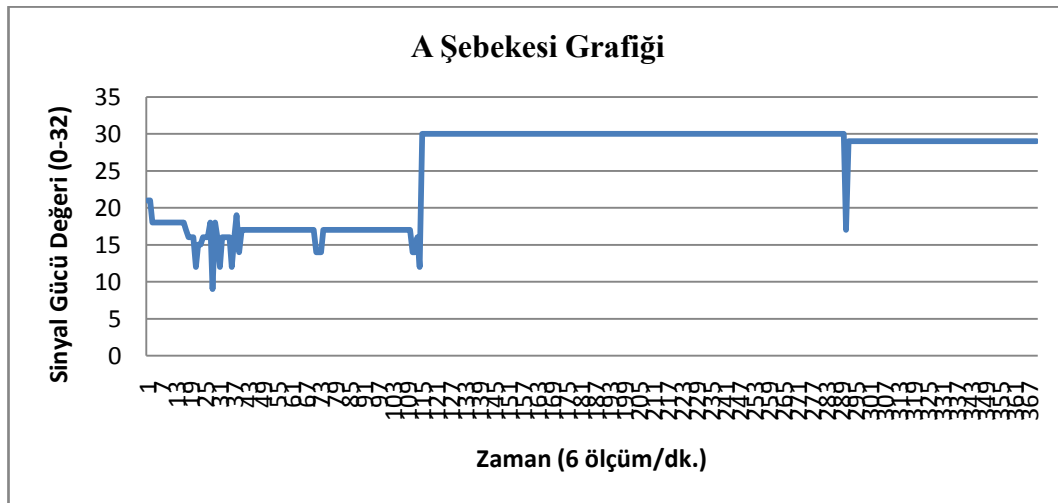
Verilerin temizlenmesi ve hazırlanması aşamasından sonra her bir şebeke için 370 adet sinyal verisi bulunan bir tablo oluşturulmuştur. Bu tablodan her onuncu verinin seçilmesiyle oluşturulan bir bölüm Şekil 4.3.’te gösterilmiştir. Her bir sinyal verisinin karşılık geldiği sınıflar da belirtilmiştir.

S.No	A ŞEBEKESİ		B ŞEBEKESİ		C ŞEBEKESİ	
	SİNYAL KALİTESİ	DEĞERİ	SİNYAL KALİTESİ	DEĞERİ	SİNYAL KALİTESİ	DEĞERİ
1	20,99	İYİ SİNYAL	15,99	ORTA SİNYAL	15,99	ORTA SİNYAL
11	17,99	İYİ SİNYAL	15,99	ORTA SİNYAL	12,99	ORTA SİNYAL
21	15,99	ORTA SİNYAL	14,99	ORTA SİNYAL	8,99	ZAYIF SİNYAL
31	17,99	İYİ SİNYAL	18,99	İYİ SİNYAL	15,99	ORTA SİNYAL
41	13,99	ORTA SİNYAL	21,99	İYİ SİNYAL	16,99	ORTA SİNYAL
51	16,99	ORTA SİNYAL	21,99	İYİ SİNYAL	15,99	ORTA SİNYAL
61	16,99	ORTA SİNYAL	15,99	ORTA SİNYAL	15,99	ORTA SİNYAL
71	16,99	ORTA SİNYAL	6,99	YETERSİZ SİNYAL	17,99	İYİ SİNYAL
81	16,99	ORTA SİNYAL	16,99	ORTA SİNYAL	21,99	İYİ SİNYAL
91	16,99	ORTA SİNYAL	16,99	ORTA SİNYAL	13,99	ORTA SİNYAL
101	16,99	ORTA SİNYAL	10,99	ZAYIF SİNYAL	13,99	ORTA SİNYAL
111	16,99	ORTA SİNYAL	24,99	İYİ SİNYAL	22,99	İYİ SİNYAL
121	29,99	İYİ SİNYAL	21,99	İYİ SİNYAL	22,99	İYİ SİNYAL
131	29,99	İYİ SİNYAL	21,99	İYİ SİNYAL	22,99	İYİ SİNYAL
141	29,99	İYİ SİNYAL	21,99	İYİ SİNYAL	22,99	İYİ SİNYAL
151	29,99	İYİ SİNYAL	21,99	İYİ SİNYAL	22,99	İYİ SİNYAL
161	29,99	İYİ SİNYAL	21,99	İYİ SİNYAL	22,99	İYİ SİNYAL
171	29,99	İYİ SİNYAL	21,99	İYİ SİNYAL	22,99	İYİ SİNYAL
181	29,99	İYİ SİNYAL	21,99	İYİ SİNYAL	22,99	İYİ SİNYAL
191	29,99	İYİ SİNYAL	21,99	İYİ SİNYAL	18,99	İYİ SİNYAL
201	29,99	İYİ SİNYAL	21,99	İYİ SİNYAL	15,99	ORTA SİNYAL
211	29,99	İYİ SİNYAL	21,99	İYİ SİNYAL	17,99	İYİ SİNYAL
221	29,99	İYİ SİNYAL	21,99	İYİ SİNYAL	17,99	İYİ SİNYAL
231	29,99	İYİ SİNYAL	21,99	İYİ SİNYAL	17,99	İYİ SİNYAL
241	29,99	İYİ SİNYAL	21,99	İYİ SİNYAL	17,99	İYİ SİNYAL
251	29,99	İYİ SİNYAL	21,99	İYİ SİNYAL	17,99	İYİ SİNYAL
261	29,99	İYİ SİNYAL	21,99	İYİ SİNYAL	17,99	İYİ SİNYAL
271	29,99	İYİ SİNYAL	19,99	İYİ SİNYAL	17,99	İYİ SİNYAL
281	29,99	İYİ SİNYAL	19,99	İYİ SİNYAL	17,99	İYİ SİNYAL
291	29,99	İYİ SİNYAL	19,99	İYİ SİNYAL	17,99	İYİ SİNYAL
301	28,99	İYİ SİNYAL	17,99	İYİ SİNYAL	17,99	İYİ SİNYAL
311	28,99	İYİ SİNYAL	17,99	İYİ SİNYAL	17,99	İYİ SİNYAL
321	28,99	İYİ SİNYAL	22,99	İYİ SİNYAL	21,99	İYİ SİNYAL
331	28,99	İYİ SİNYAL	22,99	İYİ SİNYAL	16,99	ORTA SİNYAL
341	28,99	İYİ SİNYAL	22,99	İYİ SİNYAL	16,99	ORTA SİNYAL
351	28,99	İYİ SİNYAL	22,99	İYİ SİNYAL	16,99	ORTA SİNYAL
361	28,99	İYİ SİNYAL	22,99	İYİ SİNYAL	17,99	İYİ SİNYAL
370	28,99	İYİ SİNYAL	22,99	İYİ SİNYAL	17,99	İYİ SİNYAL

Şekil 4.3. Elde edilen veri tablosundan bir bölüm

4.5.1. Veri madenciliği yöntemlerinin uygulanması

Her üç GSM Şebekesinden elde edilen veriler düzenlenerek ve çeşitli yöntemlerle analiz edilerek yorumlanması mümkün hale getirilmiştir. Elde edilen bu veriler sonucunda her bir şebeke için zaman-sinyal gücü grafiği oluşturulmuştur. Bu grafikler aşağıda yorumlanmıştır.

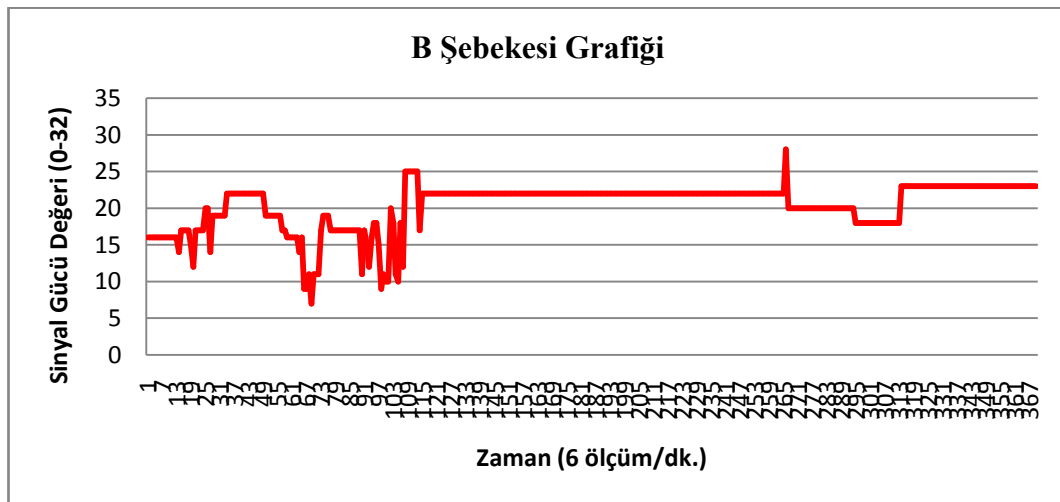


Şekil 4.4. A Şebekesi sinyal gücü grafiği

A Şebekesi hareket bölgesinin ilk kilometrelerindeki tabiat şartlarına bağlı olarak, ilk sinyal gücü ölçümlerinde dalgalanmalar gösterse de daha sonra sinyal gücünü, en yüksek seviyeye yakın bir noktaya çekmiş (29,99) ve istikrarlı bir şekilde sinyal gücünü bu seviyede tutmayı başarmıştır. En yüksek sinyal gücü olan 31,99 değerine yakın değerler olan 29,99 ve 28,99 değerlerinde, uzun süreli ve mesafeli ölçümleri sağlanabilmiştir.

A Şebekesi tüm verilerin ortalamasında 25,99 değerine yaklaşarak, aritmetik ortalamalar baz alındığında en yüksek ortalamayı sağlayan şebeke olmayı başarmıştır. Daha kapsamlı incelemelere bu tezin ilerleyen bölümlerinde yer verilecektir.

B Şebekesinde de sinyal ölçümünün ilk kilometrelerinde benzer şekilde sinyal gücünde dalgalanmalar olmuştur. Ancak ilerleyen zamanlarda B Şebekesi de sinyal gücünü bir istikrara kavuşturmuştur. Dikkat çeken bir nokta, sinyal ölçümünün ilerleyen zamanlarında B Şebekesinin sinyal gücü, A Şebekesiyle sınıflandırma anlamında aynı sınıfta olsa da, sinyal gücü anlamında daha düşük seviyede olmuştur. B Şebekesinin sinyal grafiği Şekil 4.5.'te gösterilmiştir.

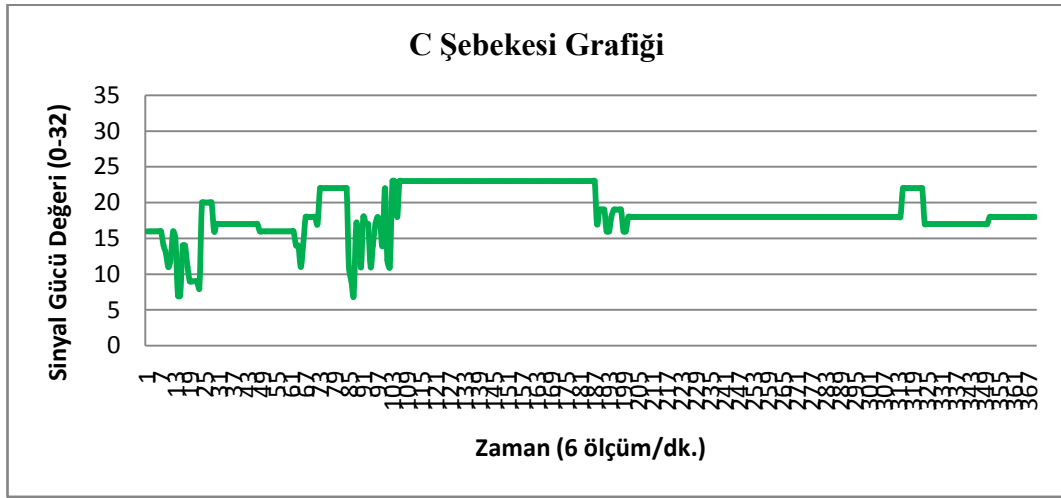


Şekil 4.5. B Şebekesi sinyal gücü grafiği

B Şebekesi sadece bir noktada, sinyal gücü anlamında A Şebekesine yaklaşmış ancak daha sonra yine daha düşük seviyede seyretmiştir. Ayrıca B şebekesi bir noktada

yetersiz sinyal seviyesine düşmüştür. B Şebekesinin ortalama sinyal gücü, aritmetiksel olarak alındığında 20,99 seviyesine yakın bir seviyede olmuş ve bu üç şebeke içerisinde aritmetik olarak ortalama alındığında, orta sırada yer almıştır.

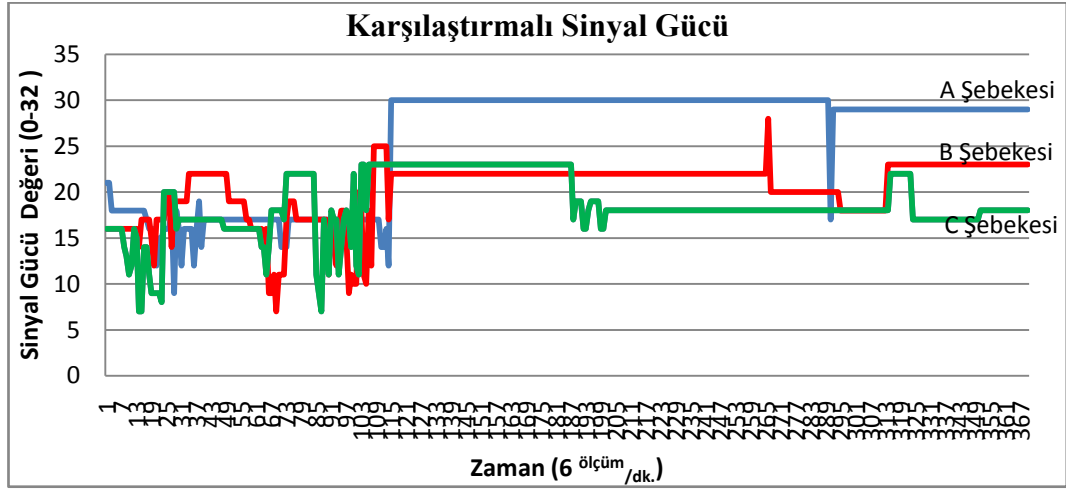
C Şebekesinde sinyal gücü ölçümlerinde diğer şebekelere oranla fark edilir şekilde dalgalanmalar gerçekleşmiştir. Sinyal ölçümlerinin ilerleyen zamanlarında sınıflandırma anlamında iyi sinyal seviyesinde bulunsa da, sinyal gücü ölçümlerinde diğer şebekelerden daha düşük seviyede ölçümler elde edilmiştir. C Şebekesine ait ölçüm grafiği Şekil 4.6.'da gösterilmiştir.



Şekil 4.6. C Şebekesi sinyal gücü grafiği

C Şebekesinin sinyal gücü, ölçüm yapılan üç noktada yetersiz sinyal seviyesine düşmüştür. Bu anlamda diğer şebekelerle kıyaslandığında en çok yetersiz sayıda ölçüm yapılan şebeke olmuştur. Bunun sebebi büyük ölçüde, altyapı yetersizliği olarak düşünülebilir. Ayrıca aritmetik ortalama alındığında 18,99 ortalama ile en düşük sinyal gücü ortalaması elde edilen şebeke C Şebekesi olmuştur.

Ölçüm yapılan üç şebekenin karşılaştırmalı grafiği Şekil 4.7.'de gösterilmiştir.



Şekil 4.7. Karşılaştırmalı sinyal gücü grafiği

Frekans analizi

Elde edilen sinyal gücü ölçüm verilerine frekans analizi yöntemi uygulanmıştır. Böylece her bir sinyal seviyesi için elde edilen ölçümler, toplam ölçüm sayısı içerisinde yüzde olarak hesaplanmış ve hangi şebekenin hangi sınıfta ve seviyede çekim seviyesine sahip olduğu ortaya çıkarılmıştır.

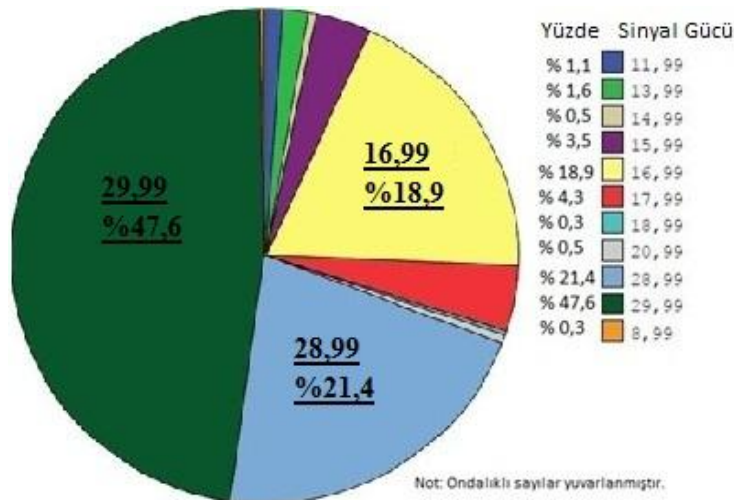
A şebekesi için elde edilen frekans tablosu Çizelge 4.2.'de gösterilmiştir. Buna göre A şebekesinde %47,6 ile en yüksek yüzdeye sahip olan sinyal gücü seviyesi 29,99 olmuştur. Aynı zamanda bu ölçüm değeri, tüm şebekeler için ölçülen değerler içerisinde en yüksek değerdir. Buradan anlaşılmaktadır ki, A Şebekesi yüksek bir yüzde ile yüksek sinyal gücü seviyesine ulaşmıştır. A Şebekesi için yetersiz sinyal seviyesinde ölçüm hiç gerçekleştirilmemiş ve sadece bir noktada bu seviyenin iki üst noktası olan 8,99'luk sinyal gücü ölçümü gerçekleşmiştir. Bu ise 0,3 lük çok düşük bir orana denk gelmektedir.

Çizelge 4.2. A Şebekesi için frekans tablosu

A Şebekesi Değerleri	Frekans	Yüzde	Geçerli Yüzde	Birikimli Yüzde
11,99	4	1,1	1,1	1,1
13,99	6	1,6	1,6	2,7
14,99	2	0,5	0,5	3,2
15,99	13	3,5	3,5	6,8
16,99	70	18,9	18,9	25,7
17,99	16	4,3	4,3	30
18,99	1	0,3	0,3	30,3
20,99	2	0,5	0,5	30,8
28,99	79	21,4	21,4	52,2
29,99	176	47,6	47,6	99,7
8,99	1	0,3	0,3	100
Toplam	370	100	100	

Not: Kullanılan programda, ondalıklı kesirler yuvarlanmıştır.

A Şebekesinde %21,4 oran ile ikinci yüksek paya sahip oran, 28,99 sinyal gücü ölçümü olmuştur. Bu oran da yine tüm şebekeler içinde ölçülen en yüksek sinyal gücü seviyelerinden birisi olmuştur. Buradan anlaşılmaktadır ki, toplamda %69 gibi yüksek bir oranda sinyal gücü seviyesi A Şebekesinde, en üst sınır olan 31,99'a yakın seviye olan 28,99 ve 29,99 değerlerinden oluşmaktadır. Bu da A Şebekesi için çekim gücü performansında büyük bir artı olmuştur.



Şekil 4.8. A Şebekesi için frekans dağılımı grafiği

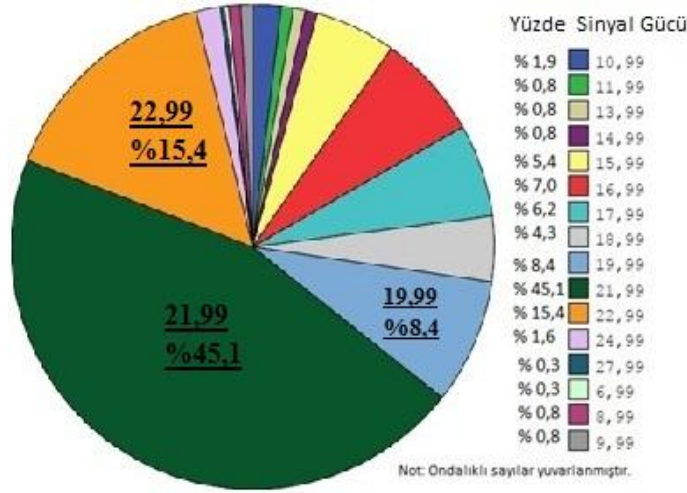
B Şebekesinin frekans tablosu incelendiğinde, %45,1 ile en çok orana sahip olan sinyal gücü değeri 21,99 olmuştur. Geriye kalan ölçüm değerleri içerisinde diğerlerinden ayrılan bir değer ise %15,4 ile 22,99 olmuştur. Diğer ölçüm verilerinin oranları, birbirlerine yakın değerlerde dağılım göstermiştir. Yetersiz sinyal seviyesi olan 6,99 değeri 1 defa ölçülmüş ve bu da %0,3'lük bir orana denk gelmiştir. Bu seviyeye yakın diğer ölçümler ise üçer defa tespit edilen 8,99 ve 9,99 değerleri olmuştur. Toplamda 6 noktada tespit edilen bu ölçümler %1,6 oranına karşılık gelmektedir. B şebekesine ait frekans tablosu Çizelge 4.3.'te gösterilmiştir.

Çizelge 4.3. B Şebekesi için frekans tablosu

B Şebekesi Değerleri	Frekans	Yüzde	Geçerli Yüzde	Birikimli Yüzde
10,99	7	1,9	1,9	1,9
11,99	3	0,8	0,8	2,7
13,99	3	0,8	0,8	3,5
14,99	3	0,8	0,8	4,3
15,99	20	5,4	5,4	9,7
16,99	26	7	7	16,8
17,99	23	6,2	6,2	23
18,99	16	4,3	4,3	27,3
19,99	31	8,4	8,4	35,7
21,99	167	45,1	45,1	80,8
22,99	57	15,4	15,4	96,2
24,99	6	1,6	1,6	97,8
27,99	1	0,3	0,3	98,1
6,99	1	0,3	0,3	98,4
8,99	3	0,8	0,8	99,2
9,99	3	0,8	0,8	100
Toplam	370	100	100	

Not: Kullanılan programda, ondalıklı kesirler yuvarlanmıştır.

B Şebekesi için dikkat çeken bir diğer nokta ise, ölçülen değer aralıklarının A Şebekesine göre daha fazla dağılım gösterdiğidir. Buradan da B Şebekesinin sinyal gücü dalgalanmasının A Şebekesinden daha fazla olduğu anlamı çıkmaktadır. Bu çıkarım B Şebekesinin frekans grafiğinde de açıkça görülmektedir (Şekil 4.9.).



Şekil 4.9. B Şebekesi için frekans dağılımı grafiği

Çizelge 4.4. C Şebekesi için frekans tablosu

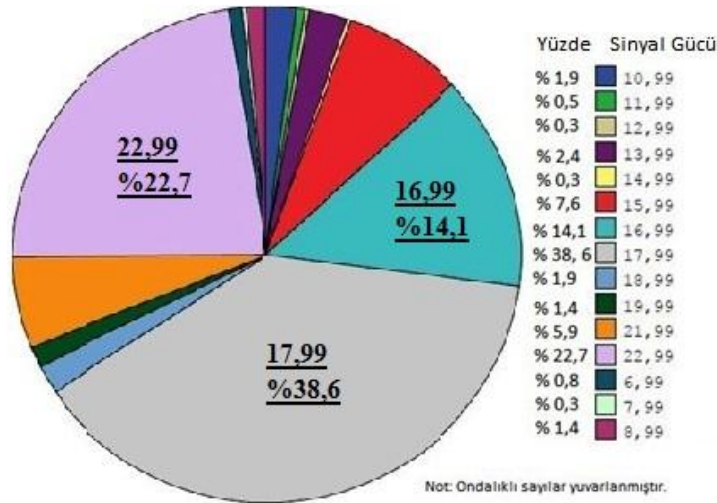
C Şebekesi Değerleri	Frekans	Yüzde	Geçerli Yüzde	Birikimli Yüzde
10,99	7	1,9	1,9	1,9
11,99	2	0,5	0,5	2,4
12,99	1	0,3	0,3	2,7
13,99	9	2,4	2,4	5,1
14,99	1	0,3	0,3	5,4
15,99	28	7,6	7,6	13
16,99	52	14,1	14,1	27
17,99	143	38,6	38,6	65,7
18,99	7	1,9	1,9	67,6
19,99	5	1,4	1,4	68,9
21,99	22	5,9	5,9	74,9
22,99	84	22,7	22,7	97,6
6,99	3	0,8	0,8	98,4
7,99	1	0,3	0,3	98,6
8,99	5	1,4	1,4	100
Toplam	370	100	100	

Not: Kullanılan programda, ondalıklı kesirler yuvarlanmıştır.

C Şebekesinin frekans tablosu incelendiğinde en yüksek orana %38,6 ile 17,99 sinyal gücü ölçümü karşılık geldiği görülmektedir. A ve B şebekelerine oranla en düşük çekim gücü performansı gösteren C Şebekesi, sinyal dalgalanması açısından B

Şebekesinden bir değer daha düşük olmuştur. B Şebekesinde 16 ayrı sinyal değeri tespit edilirken C Şebekesinde 15 ayrı sinyal değeri ölçülmüştür. Bu anlamda yine en istikrarlı şebeke A Şebekesi olmuştur.

C Şebekesinin en yüksek oranını izleyen ikinci oran %22,7 ile 22,99 sinyal gücü değeri olmuştur. Bu değer aynı zamanda C Şebekesinde tespit edilen en yüksek sinyal gücü değeri olmuş ve toplamda 84 defa tespit edilmiştir. C Şebekesi 3 noktada yetersiz sinyal seviyesi ölçümü ile bu konuda en başarısız şebeke olarak görülmüştür. Bu ölçümlere karşılık gelen oran ise %0,8 olmuştur. Bu düşük seviyeyi izleyen 7,99 ve 8,99 değerleri ise toplamda 6 defa ölçülmüş ve toplamda %1,7 oranına karşılık gelmiştir. Sinyal gücü performansları frekans tabloları ve grafikleri ile incelendiğinde, üç şebeke arasında performansı en düşük tespit edilen şebeke C Şebekesi olmuştur.



Şekil 4.10. C Şebekesi için frekans grafiği

Her bir şebekenin frekans tablosunu ve grafiğini sinyal gücü değerlerine göre inceledikten sonra aynı tablo ve grafikleri bir de sınıflandırılmış aralıklara göre incelemekte fayda vardır. Çünkü bu tablolarda, aslında veri madenciliğinin asıl amacını oluşturan, ilk bakışta kestirilemeyen ancak daha sonra veri madenciliğiyle ortaya çıkan ilginç bilgiye de ulaşılmıştır. Sinyal gücü ölçümlerinde, en yüksek değere ulaşılan ve en yüksek oranda yüksek çekim gücü tespit edilen şebeke A Şebekesi olarak görülürken, sınıflandırılmış aralıklara göre frekans dağılımı

uygulandığında, B Şebekesinin iyi sinyal oranının A Şebekesinden daha çok olduğu ortaya çıkmıştır. Yine farklı ve ilginç bir bilgi olarak B Şebekesinin, zayıf sinyal sınıfı oranının diğer iki şebekeden de daha fazla olduğu ortaya çıkmıştır.

Çizelge 4.5. A, B ve C Şebekeleri için sınıflandırılmış frekans tablosu

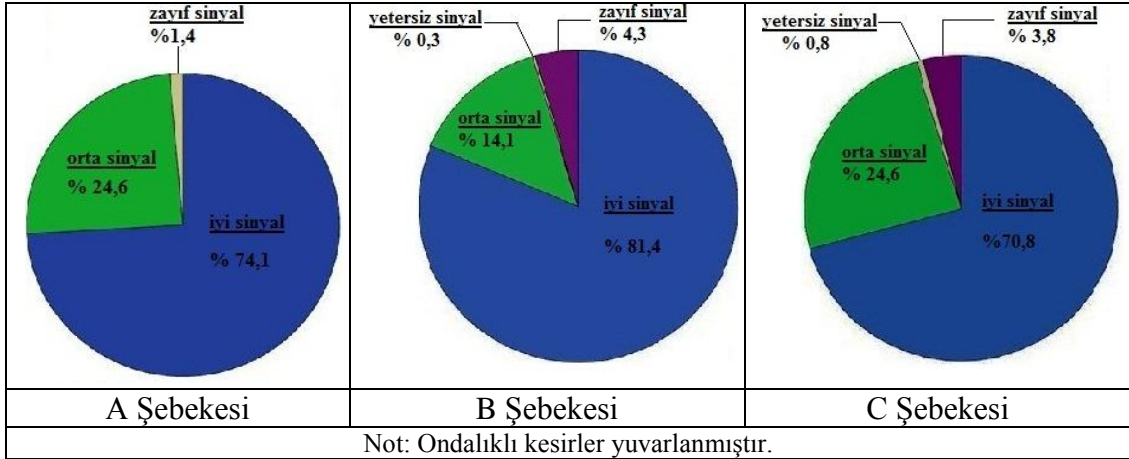
A Şebekesi Değerleri	Frekans	Yüzde	Geçerli Yüzde	Birikimli Yüzde
iyi sinyal	274	74,1	74,1	74,1
orta sinyal	91	24,6	24,6	98,6
zayıf sinyal	5	1,4	1,4	100
Toplam	370	100	100	
B Şebekesi Değerleri	Frekans	Yüzde	Geçerli Yüzde	Birikimli Yüzde
iyi sinyal	301	81,4	81,4	81,4
orta sinyal	52	14,1	14,1	95,4
yetersiz sinyal	1	0,3	0,3	95,7
zayıf sinyal	16	4,3	4,3	100
Toplam	370	100	100	
C Şebekesi Değerleri	Frekans	Yüzde	Geçerli Yüzde	Birikimli Yüzde
iyi sinyal	262	70,8	70,8	70,8
orta sinyal	91	24,6	24,6	95,4
yetersiz sinyal	3	0,8	0,8	96,2
zayıf sinyal	14	3,8	3,8	100
Toplam	370	100	100	

Not: Kullanılan programda, ondalıklı kesirler yuvarlanmıştır.

Sinyal gücü frekans oranlarına bakıldığında en düşük performanslı şebeke olarak görülen C Şebekesi, sınıflandırılmış aralıklara göre de yine en düşük orana sahip olmasına rağmen, orta sinyal sınıfında A Şebekesi ile aynı orana sahip olmuştur. Ayrıca zayıf sinyal sınıfı ölçümlerinde de B Şebekesinden daha düşük oran

yakalamıştır. Ancak yetersiz sinyal sınıfında %0,8 ile en çok yetersiz sinyal tespit edilen şebeke olmuştur.

Tüm şebekeler için frekans grafikleri aşağıda verilmiştir.



Şekil 4.11. A, B ve C Şebekeleri için sınıflandırılmış frekans grafiği

İki Aşamalı Kümeleme (Two Step Cluster)

İki aşamalı kümeleme algoritmalarında, ilk akla gelen Punj ve Steward (1983) tarafından önerilen klasik iki aşamalı kümeleme algoritmasıdır. Bu algoritma, Ward'ın minimum varyans yöntemi ile “Kmeans” yönteminden oluşan bir hibrid yaklaşımdır. Böyle bir karma yaklaşımın avantajı, Ward'ın minimum varyans yönteminin, “K-means” yönteminin gerektirdiği küme sayısını hesaplamasından ileri gelmektedir [57].

İki aşamalı kümeleme algoritması, öncelikli amaç olarak büyük veri tabanlarını analiz etmek için tasarlanmıştır. Bu algoritma, kümelerdeki gözlemleri ayırıcı özellik yaklaşımını kullanarak gruplandırır. Klasik kümeleme algoritmalarına kıyasla iki aşamalı kümeleme analizi, hem daha devamlı hem de daha öznelikli kategoriler sağlar. Ayrıca bu yöntem otomatik olarak optimum küme sayısını kendisi belirler.

İki aşamalı kümeleme algoritması, ön-kümeleme, tipik veri türlerini çözümüleme ve kümeleme aşamalarını gerçekleştirir. Ön kümeleme sırasında verilerin her biri

incelenerek, her bir veri bir önceki kümeye dahil edilebilir mi yoksa başka bir küme başlatılmalı mı diye karar verilir. Bu karar verilerin birbirlerine uzaklığına göre alınır. Öklid uzaklığı ve log-olasılık uzaklığı olmak üzere iki uzaklık ölçütü vardır. Tipik veri çözümleme aşamasında, herhangi bir kümeye dahil edilemeyen veriler değerlendirilir. Ekleme işlemi için tüm denemeler gerçekleştirildikten sonra dahiliyet yine sağlanamaz ise, bu veriler dış veriler olarak ayrılır. Kümeleme aşamasında ise bir ağaç yapısı oluşturulur. Tüm veriler kökten yapraklara doğru dağıtılmaya başlanır. Her bir veri kendisine yakın bir dala tutturulur, eğer dahil edilecek grup sayısı optimum grup üyesi sayısına ulaşmışsa, uzaklık ölçütüne göre en uygun diğer bir dalda diğer bir kümeye tutturulur.

En uygun küme sayısını otomatik olarak belirlemek için, BIC (Schwarz's Bayesian Information Criterion) veya AIC (Akaike's Information Criterion) yöntemleri kullanılır.

Elde edilen sinyal verilerine, iki aşamalı kümeleme yöntemi uygulanmıştır. Uygulama sonrasında ortaya çıkan algoritma özeti Şekil 4.12.'de gösterilmiştir.

Output Created		
Comments		
Input	Active Dataset Filter Weight Split File N of Rows in Working Data File	DataSet1 <none> <none> <none> 370
Missing Value Handling	Definition of Missing Cases Used	User-defined missing values are treated as missing. Statistics are based on all cases with valid data for all variables in the analysis.
Syntax		TWOSTEP CLUSTER /CATEGORICAL VARIABLES = A B C /DISTANCE LIKELIHOOD /NUMCLUSTERS AUTO 15 BIC /HANDLENOISE 0 /MEMALLOCATE 64 /CRITERIA INITHRESHOLD (0) MXBRANCH (8) MXLEVEL (3) /PRINT COUNT SUMMARY .
Resources	Elapsed Time Processor Time	0:00:00,09 0:00:00,16

Şekil 4.12. Sinyal gücü için iki aşamalı kümeleme özellikleri

Belirtilen özelliklerden de anlaşılacağı üzere, otomatik olarak BIC metoduna göre kümeleme işlemi gerçekleştirilmiştir. Daha sonra bu kümeler olasılık uzaklığı

ölçütüne göre 3 kümede kombine edilmiştir. Bunun sonucunda da ortaya çıkan küme dağılımı Çizelge 4.6.'daki gibi olmuştur.

Çizelge 4.6. Sinyal gücü için iki aşamalı kümeleme sonuçları

Küme	Eleman Sayısı	Kombinasyon Yüzdesi	Toplam Yüzdesi
1	115	31,10%	31,10%
2	104	28,10%	28,10%
3	151	40,80%	40,80%
Kombine	370	100%	100%
Toplam	370		100%

Bu tablodan çıkan sonuç şöyledir: Her üç şebeke için 370'er sinyal ölçümü sonrasında oluşan veri grubundan birinci kümeye 115, ikinci kümeye 104 ve üçüncü kümeye 151 adet veri eklenmiştir. Her bir kümeye hangi sinyal derecesinden kaç tane ekleneceği bilgisi, sinyal gücü dağılımı için uygulanan iki aşamalı kümeleme yöntemi sonucunda ortaya çıkmış ve bu tablolar EK-2'de olduğu gibi birleştirilerek sunulmuştur.

EK-2'deki çizelge incelendiğinde her bir küme grubu için, hangi şebekeden, hangi sinyal gücünden kaç eleman seçildiği ilk bakışta fark edilmeyebilir. Bu sebeple şöyle bir düzenlemeye gidilebilir.

$$\text{Şebeke}_{(\text{küme no})} = \{ [\text{Sinyal gücü}](\text{seçilen eleman sayısı}) \} = \text{toplam eleman sayısı} \quad (4.1)$$

$$\text{Ş}_{kn} = \{ [\text{Sg}](\text{SES}) \} = \text{TES}$$

Eş. 4.1.'den yola çıkarak her bir şebeke için küme elemanları gösterildiğinde ortaya çıkan sonuç şu şekilde olmuştur:

$$A_1 = \{ [8,99](1), [11,99](4), [13,99](6), [14,99](2), [15,99](13), [16,99](70), [17,99](16), [18,99](1), [20,99](2) \} = 115 \text{ eleman.}$$

$B_1 = \{[6,99](1), [8,99](3), [9,99](3), [10,99](7), [11,99](3), [13,99](3), [14,99](3), [15,99](20), [16,99](26), [17,99](4), [18,99](16), [19,99](4), [21,99](16), [24,99](6)\} = 115$ eleman.

$C_1 = \{[6,99](3), [7,99](1), [8,99](5), [10,99](7), [11,99](2), [12,99](1), [13,99](9), [14,99](1), [15,99](24), [16,99](24), [17,99](9), [19,99](5), [21,99](13), [22,99](11)\} = 115$ eleman.

$A_2 = \{[28,99](79), [29,99](25)\} = 104$ eleman.

$B_2 = \{[17,99](19), [19,99](27), [22,99](57), [27,99](1)\} = 104$ eleman.

$C_2 = \{[16,99](27), [17,99](68), [21,99](9)\} = 104$ eleman.

$A_3 = \{[28,99](151)\} = 151$ eleman.

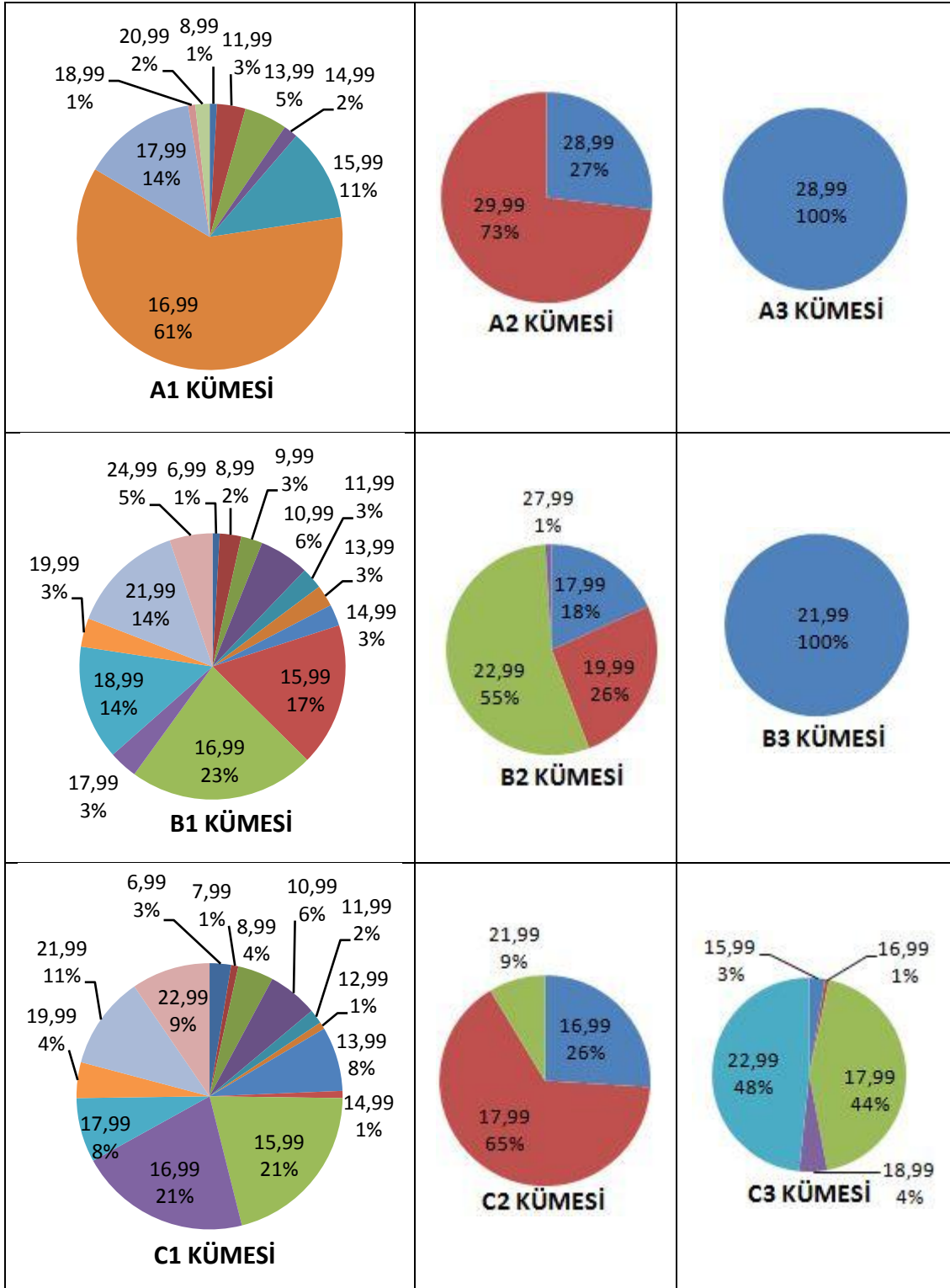
$B_3 = \{[21,99](151)\} = 151$ eleman.

$C_3 = \{[15,99](4), [16,99](1), [17,99](66), [18,99](7), [22,99](73)\} = 151$ eleman.

Buradaki kümeler incelendiğinde, A, B ve C Şebekelerinin tüm küçük oranlı sinyal gücü ölçümlerinin tamamının 1. kümede toplandığı görülmüştür. A Şebekesinin en yüksek değerleri 2. ve 3. kümelerde toplanırken, 3. küme için tek bir veri grubu seçilmiştir. B Şebekesinin de benzer şekilde yüksek değerleri 2. ve 3. kümelerde toplanmıştır ve 3. küme için tek bir veri grubu seçilmiştir. C Şebekesi için tüm kümelere sinyal gücü dağılımı birden fazla veri grubu ile olmuştur.

İki aşamalı kümeleme yöntemi uygulandıktan sonra ortaya çıkan bu tablo göstermektedir ki, A ve B Şebekeleri, sinyal gücü ölçümlerinde C Şebekesine göre daha istikrarlı sonuçlar göstermiştir. A ve B Şebekeleri için elde edilen ölçüm sonuçları, belirli sürelerle sabit kalarak devam etmiştir. Ancak C Şebekesinin ölçüm değerleri, farklı zaman aralıklarından değişerek devam etmiştir. İki aşamalı

kümeleme yöntemiyle elde edilen küme elemanları, her bir küme için Şekil 4.13.'te belirtilen oranlara sahip olmuştur.



Şekil 4.13. İki aşamalı kümeleme yöntemi, küme elemanları grafikleri

Bu yöntemden çıkarılabilecek bir başka sonuç ise her bir şebeke kümesinin, diğer şebekenin aynı kümesindeki sinyal gücü oranına karşılık gelebilecek olası değeri göstermesidir. Örneğin, A_3 kümesindeki ölçüm periyodu üyesi 28,99 değeridir ve 100% orana sahiptir. Aynı şekilde 100% oranla B_3 kümesindeki ölçüm değeri, 21,99 sinyal gücü değeridir. Buradan, A Şebekesinde 28,99 değeri ölçülen bölgelerde B Şebekesi için 21,99 değeri ölçülecektir olasılığı çıkmaktadır. Bu örnek benzer şekilde diğer kümelerle de eşleştirilebilir.

Kullanılan yöntemlerden frekans analizi yöntemi sonucunda, sinyal gücü verilerinin oranları belirlenmiş ve buradan elde edilen bilgiyle, en yüksek sinyal gücü seviyesine A Şebekesinin ulaştığı görülmüştür. Ayrıca yüksek seviyede istikrarlı sinyal gücü ölçümü de A Şebekesinde tespit edilmiştir. Ancak sinyal gücü aralıkları sınıflandırılarak yapılan frekans analizinde çıkan sonuç bundan farklı olmuştur. B Şebekesi, iyi sinyal sınıfında A Şebekesinden daha fazla oran elde etmiştir. Her iki frekans analizinde de sinyal gücü performansı en düşük şebeke C Şebekesi olmuştur. İki aşamalı kümeleme yönteminde A Şebekesi, B Şebekesine oranla daha istikrarlı ölçüm sonuçları vermiştir. Bu yöntemde de yine C Şebekesi sinyal gücü ölçümünde performansı en düşük şebeke olmuştur.

Tez çalışmasının bu bölümünde elde edilen verilerin temizlenmesi ve hazırlanması aşamasından sonra veri madenciliği modeli oluşturulmuş ve uygulama sonucunda ortaya çıkan veriler izlenmiş ve sonuçlar yorumlanmıştır. Böylece veri madenciliği süreci tamamlanmıştır. Bu bilgiler ışığında A, B ve C Şebekelerinin, belirlenen güzergahta çekim gücü performansları değerlendirilmiştir. Sinyal gücü ölçümü için geliştirilen program ve gelişi-güzel seçilen güzergahta yapılan sinyal gücü ölçümlerinin yorumlanması, gelecekte gerçekleştirilebilecek uygulamalar için örnek teşkil etmektedir.

5. SONUÇ VE ÖNERİLER

Tez kapsamında, GSM şebekelerinin sinyal gücü verilerini elde edebilmek için, cep telefonlarının çekim gücünü alarak veri tabanı oluşturmak üzere bir program geliştirilmiştir. Bu program ihtiyaçlar doğrultusunda hazırlanabilecek bu türden programlara örnek teşkil edebilecek bir programdır. Özellikle de bu programı mobil olarak uygulamak, sürekli olarak veri elde edebilmek adına GSM firmalarının yararına olacaktır.

Geliştirilen programın amacı, cep telefonun anlık sinyal verisini alarak belirlenen bir konuma kaydetmesidir. Bu veriler daha sonraki aşamada değersiz bilgilerden temizlenmiş, analiz edilmek üzere düzenlenmiş ve veri madenciliği için uygun hale getirilmiştir. Belirlenen güzergahta çalıştırılan program, üç ayrı şebeke için de kullanılmıştır. Bunun sonucunda, her bir şebeke için 370'er adet sinyal gücü verisi elde edilmiştir. Bu sinyal verilerini elde etmenin amacı ise, bu veriler üzerinde veri madenciliği çalışması yaparak, çekim gücü performanslarına yönelik çıkarımlar yapmak ve gelecekteki iyileştirmeler ve gelişmeler için öngörüler oluşturabilmektir.

Program tasarımı için MS Visual Studio 2008 C#.NET programlama dili kullanılmıştır. Sinyal gücü elde etmek için bir cep telefonundan veriler alınmıştır. Elde edilen bu veriler Frekans Analizi ve İki Aşamalı Kümeleme yöntemleriyle analiz edilerek sonuçlara ulaşılmıştır. Elde edilen sonuçlar göstermiştir ki, A Şebekesi sinyal gücü baz alındığında en yüksek oranda çekim gücüne ve en istikrarlı çekim seviyesine sahiptir. Daha sonra bu şebekeyi B Şebekesi izlemektedir. Çekim gücü performansı açısından en düşük performans gösteren şebeke C Şebekesi olmuştur.

Sinyal güçleri belirli aralıklara göre sınıflandırıldığında, elde edilen sonuçlarda yukarıda belirtilenden farklı bir durum ortaya çıkmıştır. Veri madenciliğinin de asıl amacı farklı durumları ortaya koyabilmektir. Sinyal gücü aritmetik oranı olarak ve sinyal gücü ölçümünde en yüksek noktalar ve bu noktalarda istikrarlı ölçüm bazında hep en önde olan A Şebekesi olmuştur. Buna rağmen, sınıflandırma sonrası

değerlendirmede B Şebekesi iyi sinyal seviyesi oranında, A Şebekesinden daha çok oran elde etmiştir. Ortaya çıkan başka bir durum ise, orta sinyal seviyesinde C Şebekesi ve A Şebekesinin oranlarının aynı olmasıdır. Ayrıca C Şebekesi zayıf sinyal sınıfında, B Şebekesinden daha düşük bir oran ortaya çıkarmıştır. Bu bilgiler ışığında tüm şebekeler için iyileştirme ve geliştirme bilgileri ortaya çıkmıştır.

Bilgi ve iletişim çağında bulunduğumuz günümüz yaşantısında, bu ikiliyi bir araya getiren bu tez çalışması, mobil iletişim araçları ve bu araçların ürettikleri verileri yorumlayarak gelişmeler ve iyileştirmeler göstermek adına uygulamalar sunmuştur. Günümüzde ülke nüfusu sayısı kadar GSM Şebekesi abonelikleri, her bireyde mobil iletişim araçları, pek çok bireyin çift hatlı kullanıma geçmeleri ve mobil iletişimi sadece sesli görüşme için değil, görüntülü iletişim ve internet gibi başka amaçlarla da kullanmak göstermektedir ki, mobil iletişim hayatımızda çok büyük önem taşımaktadır.

Veri madenciliği genel anlamda, çok büyük veri yığınları arasından, tahmin, yorum, iyileştirme ve geliştirme amaçlı bilgileri keşfetme işlemidir. Hayatın olmazsa olmazı haline gelen mobil iletişim alanında veri madenciliğini kullanmak, yöneticiler ve işletme sahipleri için kaçınılmaz olmuştur. Ancak işletmelerin kullanıcı tarafındaki veriye ulaşmalarında bazı zorluklar vardır. Çünkü işletmeler mobil iletişim hizmetleriyle ilgili verilere, ancak baz istasyonları log kayıtları veya uygulanabilecek anketler vasıtasıyla ulaşabilirler. Kullanıcı tarafındaki verilerin ise, baz istasyonu ile aynı oranda olması beklenmemelidir. Çünkü coğrafi özellikler çok büyük farklar yaratmaktadır. Bu sebeple işletmeler kullanıcı tarafındaki veriye de, uygulamalar geliştirerek ulaşmak durumunda kalmışlardır. Bu tez çalışmasında gerçekleştirilen uygulama programı, gelecekte yapılacak çalışmalar ve tasarlanacak programlar için örnek teşkil etmektedir.

Veri madenciliğinin uygulama alanı gittikçe yaygınlaşmaktadır. Bu çalışmada, veri madenciliği yöntemleri, hızla yayılmakta ve gelişmekte olan mobil iletişim alanına uygulanmıştır. Uygulama sonucundaki veriler yorumlanarak, elde edilen bulgular değerlendirilmiştir. Sonuçlar göstermiştir ki; veri madenciliği, mobil iletişim alanına uygulandığında, değerlendirilmesi mümkün veriler ve ilginç bilgiler ortaya koyabilecek sonuçlar elde edilebilir.

KAYNAKLAR

1. Kalikov, A., “Veri madenciliği ve bir e-ticaret uygulaması”, Yüksek Lisans Tezi, **Gazi Üniversitesi Fen Bilimleri Enstitüsü**, Ankara, 1-41 (2006).
2. İnan, O., “Veri madenciliği”, Yüksek Lisans Tezi, **Selçuk Üniversitesi Fen Bilimleri Enstitüsü**, Konya, 1-50 (2003).
3. Thuarisingham, B.M., “Web Data Mining and Applications in Business Intelligence and Counter Terrorism”, **CRC Press LLC**, Boca Raton FL USA, 35 (2003).
4. Akpınar, H., “Veri tabanlarında bilgi keşfi ve veri madenciliği”. **İ.Ü. İşletme Fakültesi Dergisi**, 29(1): 1-22 (2000).
5. Aldana, W.A., “Data mining industry: emerging trends and new opportunities”, Yüksek Lisans Tezi, **Massachusetts Institute of Technology**, Massachusetts, 11 (2000).
6. Albayrak, M., “EEG sinyallerindeki epileptiform aktivitenin veri madenciliği süreci ile tespiti”, Doktora Tezi, **Sakarya Üniversitesi Fen Bilimleri Enstitüsü**, Sakarya, 56-70 (2008).
7. Akgöbek, Ö. ve Çakır, F., “Veri madenciliğinde bir uzman sistem tasarımı”, **Akademik Bilişim 09**, Harran Üniversitesi, Şanlıurfa, 801-806 (2009).
8. Özdamar, E.Ö., “Veri madenciliğinde kullanılan teknikler ve bir uygulama”, Yüksek Lisans Tezi, **Mimar Sinan Üniversitesi Fen Bilimleri Enstitüsü**, İstanbul, 50-65 (2002).
9. Yavaş, G., “Using a data mining approach for the prediction of user movements in mobile environments”, Yüksek Lisans Tezi, **Bilkent University Institute of Engineering and Science**, Ankara, 1-5 (2003).
10. Çiftçi, S., “Uzaktan eğitimde öğrencilerin ders çalışma etkinliklerinin log verilerinin analiz edilerek incelenmesi”, Yüksek Lisans Tezi, **Gazi Üniversitesi Eğitim Bilimleri Enstitüsü**, Ankara, 1-5 (2006).
11. Akbulut, S., “Veri madenciliği teknikleri ile bir kozmetik markanın ayrılan müşteri analizi ve müşteri segmentasyonu”, Yüksek Lisans Tezi, **Gazi Üniversitesi Fen Bilimleri Enstitüsü**, Ankara, 1-25 (2006).
12. İnan, A., “Privacy preserving distributed spatio-temporal data mining”, Yüksek Lisans Tezi, **Sabancı University Engineering and Natural Sciences**, İstanbul, 1-5 (2006).

13. Kayaalp, K., "Asenkron motorlarda veri madenciliği ile hata tespiti", Yüksek Lisans Tezi, *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü*, Isparta, 1-45 (2007).
14. Ayık, Y.Z., Özdemir, A. ve Yavuz, U., "Lise türü ve lise mezuniyet başarısının kazanılan fakülte ile ilişkisinin veri madenciliği tekniği ile analizi", *Sosyal Bilimler Enstitüsü Dergisi*, 10(2): 441-454 (2007).
15. Duru, N. ve Canbay, M., "Veri madenciliği ile deprem verilerinin analizi", *International Earthquake Symposium*, Kocaeli, 556-560, (2007).
16. Özçakır, F.C. ve Çamurcu, A. Y., "Birliktelik kuralı yöntemi için bir veri madenciliği yöntemi tasarımı ve uygulaması", *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 6(12): 21-37 (2007).
17. Ata, N., Özkök, E. ve Karabey, U., "Survival data mining: an application to credit card holders", *Sigma Mühendislik ve Fen Bilimleri Dergisi*, 26(1): 33-42 (2008).
18. Çalışkan, S.K. ve Soğukpınar, İ., "KxKNN: K-Means ve k en yakın komşu yöntemleri ile ağlarda nüfuz tespiti", **2. Ağ ve Bilgi Güvenliği Sempozyumu**, Girne, 120-124 (2008).
19. Gürbüz, F., Özbakır, L. ve Yapıcı, H., "Türkiye'de bir havayolu işletmesine ait parça söküm raporlarına ilişkin veri madenciliği uygulaması", *Gazi Üniversitesi Mimarlık Mühendislik Fakültesi Dergisi*, 24(1): 73-78 (2009).
20. Albayrak, A.S. ve Yılmaz, Ş.K., "Veri madenciliği: karar ağacı algoritmaları ve İMKB verileri üzerine bir uygulama", *S.D.Ü. İktisadi ve İdari Bilimler Fakültesi Dergisi*, 14(1): 31-52 (2009).
21. Albayrak, A.S., "Classification of domestic and foreign commercial banks in turkey based on financial efficiency: a comparison of decision tree, logistic regression and discriminant analysis models", *S.D.Ü. İktisadi ve İdari Bilimler Fakültesi Dergisi*, 14(2): 113-139 (2009).
22. Bozkır, A.S., Sezer, E. ve Gök, B., "Öğrenci seçme sınavında (ÖSS) öğrenci başarılarını etkileyen faktörlerin veri madenciliği yöntemleriyle tespiti", **5. Uluslararası İleri Teknolojiler Sempozyumu (IATS'09)**, Karabük, 37-43 (2009).
23. Ata, A.H. ve Seyrek, İ.H., "The use of data mining techniques in detecting fraudulent financial statements: an application on manufacturing firms", *S.D.Ü. İktisadi ve İdari Bilimler Fakültesi Dergisi*, 14(2): 157-170 (2009).
24. Sever, H. ve Oğuz, B., "Veri tabanlarında bilgi keşfine formel bir yaklaşım", *Bilgi Dünyası*, 3(2): 173-204 (2002).

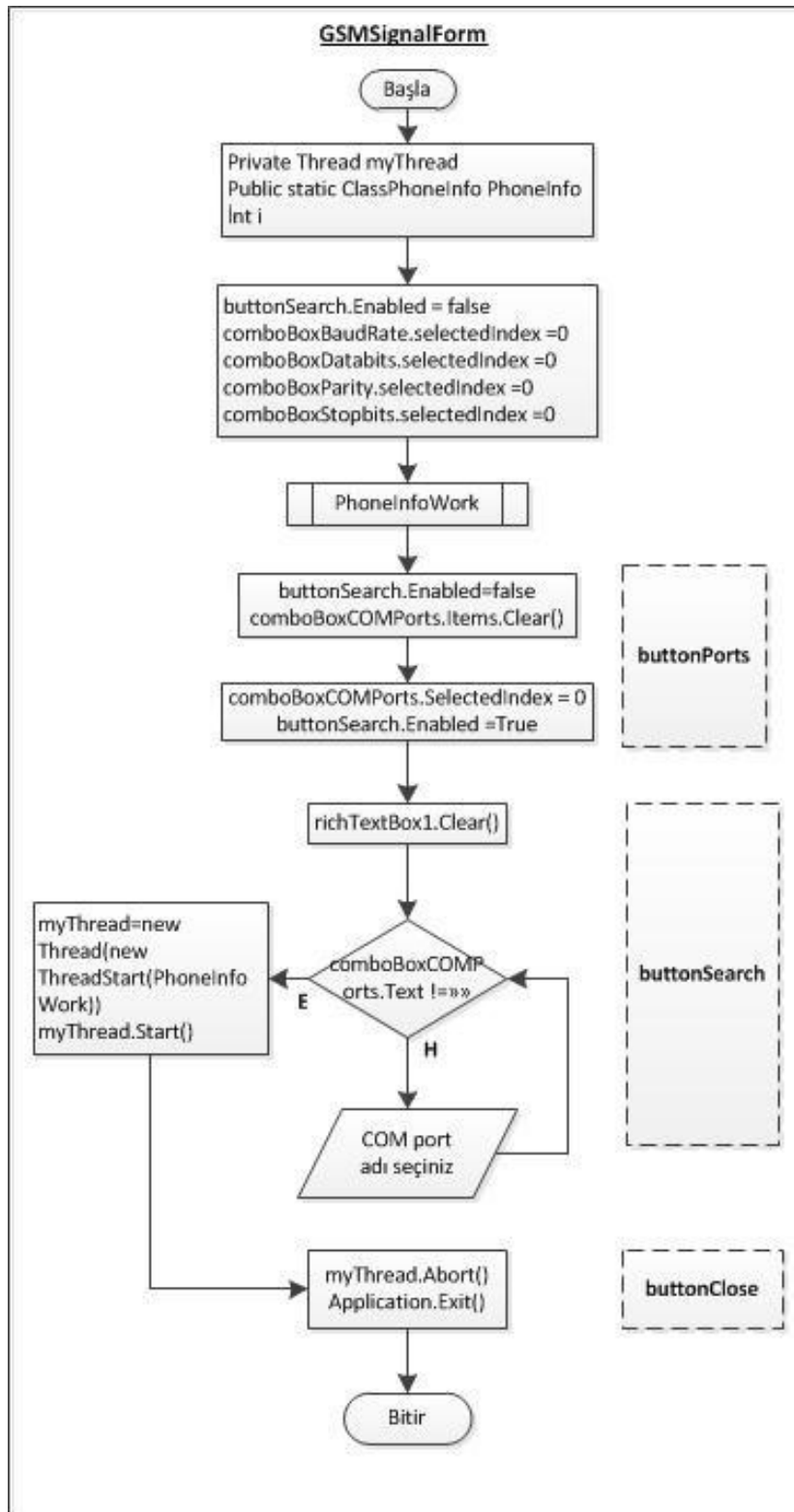
25. İnternet: Hacettepe Üniversitesi “Veri Madenciliğine Giriş” <http://yunus.hacettepe.edu.tr/~hcingi/ist376a/6Bolum.doc> (2011).
26. Ching, J.Y., Wong, A.K.C. ve Chan, K.C.C., “Class-dependent discretization for inductive learning from continuous and mixed mode data”, *IEEE Transactions on Knowledge and Data Engineering*, 17(7): 641-651(1995).
27. Shearer, C., “The crisp-dm model: the new blueprint for data mining” *Journal of Data Warehousing*, 5(4), 13-23 (2000).
28. Piramuthu, S., “Evaluating feature selection methods for learning in data mining applications”, *Thirty-First Annual Hawaii International Conference on System Sciences*, Hawaii, 5: 294 (1998).
29. Zhong, N. - Zhou, L., “Methodologies for knowledge discovery and data mining”, *Third pacific-asia conference Pakdd-99*, Beijing China, (1999).
30. Kaya, H. ve Köymen, K., “Veri madenciliği kavramı ve uygulama alanları”, *Doğu Anadolu Bölgesi Araştırmaları*, 159-164 (2008).
31. Albayrak, A.S. ve Akbulut, R., “Sermaye yapısını belirleyen faktörler: İMKB sanayi ve hizmet sektörlerinde işlem gören işletmeler üzerine bir inceleme”, *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 22: 22 (2008).
32. Giudici, P., “Applied Data Mining: Statistical Methods for Business and Industry”, *John Wiley & Sons Inc.*, Chichester, 85-100 (2003).
33. Koyuncugil, A.S. ve Özgülbaş, N., “Veri madenciliği: tıp ve sağlık hizmetlerinde kullanımı ve uygulamaları”, *Bilişim Teknolojileri Dergisi*, 2(2): 21-32 (2009).
34. Uğur, A. ve Kınacı, A.C., “Yapay zeka teknikleri ve yapay sinir ağları kullanılarak web sayfalarının sınıflandırılması”, *Inet-tr 2006, XI. Türkiye’de İnternet Konferansı*, TOBB Ekonomi ve Teknoloji Üniversitesi, Ankara, (2006).
35. Shah. S. ve Kursak. A., “Data mining and genetic algorithms based gene / SNP selection”, *Artificial Intelligence in Medicine*, 31, 183 -196 (2004).
36. Han, J. ve Kamber, M., “Data Mining: Concepts and Techniques”, *Morgan Kaufmann*, San Francisco, USA, 45-53 (2001).
37. Larose, D.T., “Discovering Knowledge in Data: An Introduction to Data Mining”, *John Wiley & Sons Inc.*, 42-70 (2005).
38. Hui, S. ve Jha, G., “Application data mining for customer service support”, *Information and Management*, 38: 1-13 (2000).

39. Pawlak, Z., "Rough sets, decision algorithms and bayes theorem", *European Journal of Operation Research*, 136: 181-189 (2002).
40. Seidman, C., "Data Mining With Microsoft SQL Server 2000", *Microsoft Press*, Washington, USA, (2001).
41. Özekes, S., "Veri madenciliği modelleri ve uygulama alanları", *İstanbul Ticaret Üniversitesi Dergisi*, 3: 65-82 (2003).
42. Agrawal, R., Imielinski, T. ve Swami, A., "Mining association rules between sets of items in large databases", *In Proceedings of the ACM SIGMOD International Conference on Management of Data (ACMSIGMOD '93)*, Washington, USA, 207-216 (1993).
43. Kroski, E., "On the move with the mobile web: libraries and mobile technologies", *American Library Association*, 44(5): 1-48 (2008).
44. Alagöz, F., "Mobil ağlar ve veri erişim stratejileri", Yüksek Lisans Tezi, *Kahramanmaraş Sütçü İmam Üniversitesi Fen Bilimleri Enstitüsü*, Kahramanmaraş, 1-6, 22-23 (2005).
45. Sektörel Araştırma ve Stratejiler Dairesi Başkanlığı, "Üç aylık pazar verileri raporu", *Bilgi Teknolojileri ve İletişim Kurumu 2010 Yılı 4. Çeyrek S:12*, Ankara, 36-59 (2011).
46. Boztaş, A., "Mobil cihazlar üzerinde güncel programlama teknikleri ve karşılaştırılması", Yüksek Lisans Projesi, *Beykent Üniversitesi Fen Bilimleri Enstitüsü*, İstanbul, 3-7, 15-25 (2006).
47. Güngör, M., Kibar, Y. Ş., Ulaşanoğlu, M. E., Güçlü, T. ve Yılmaz, R., "İstatistikî veriler ışığında dünyada ve Türkiye'de elektronik haberleşme sektörü", *Bilgi Teknolojileri ve İletişim Kurumu*, Ankara, 33-48 (2010).
48. Andrade, R., Aldo von Wangenheim ve Bortoluzzi, M., K., "Wireless and PDA: a novel strategy to access DICOM-compliant medical data on mobile devices", *International Journal of Medical Informatics*, 71: 157-163 (2003).
49. Işık, A.H., Özkaraca, O. ve Güler, İ., "Mobil öğrenme ve podcast", *Akademik Bilişim 2011*, Malatya, 147 (2011).
50. Albayrak, M., Aydoğan, T., Uzun, E. ve Görgülü, Y. E., "PDA'larda kullanılmak üzere geliştirilmiş bir öğretmen danışmanlık sistemi", *International Twelfth Turkish Symposium on Artificial Intelligence & Neural Networks*, Çanakkale, 1-9 (2003).

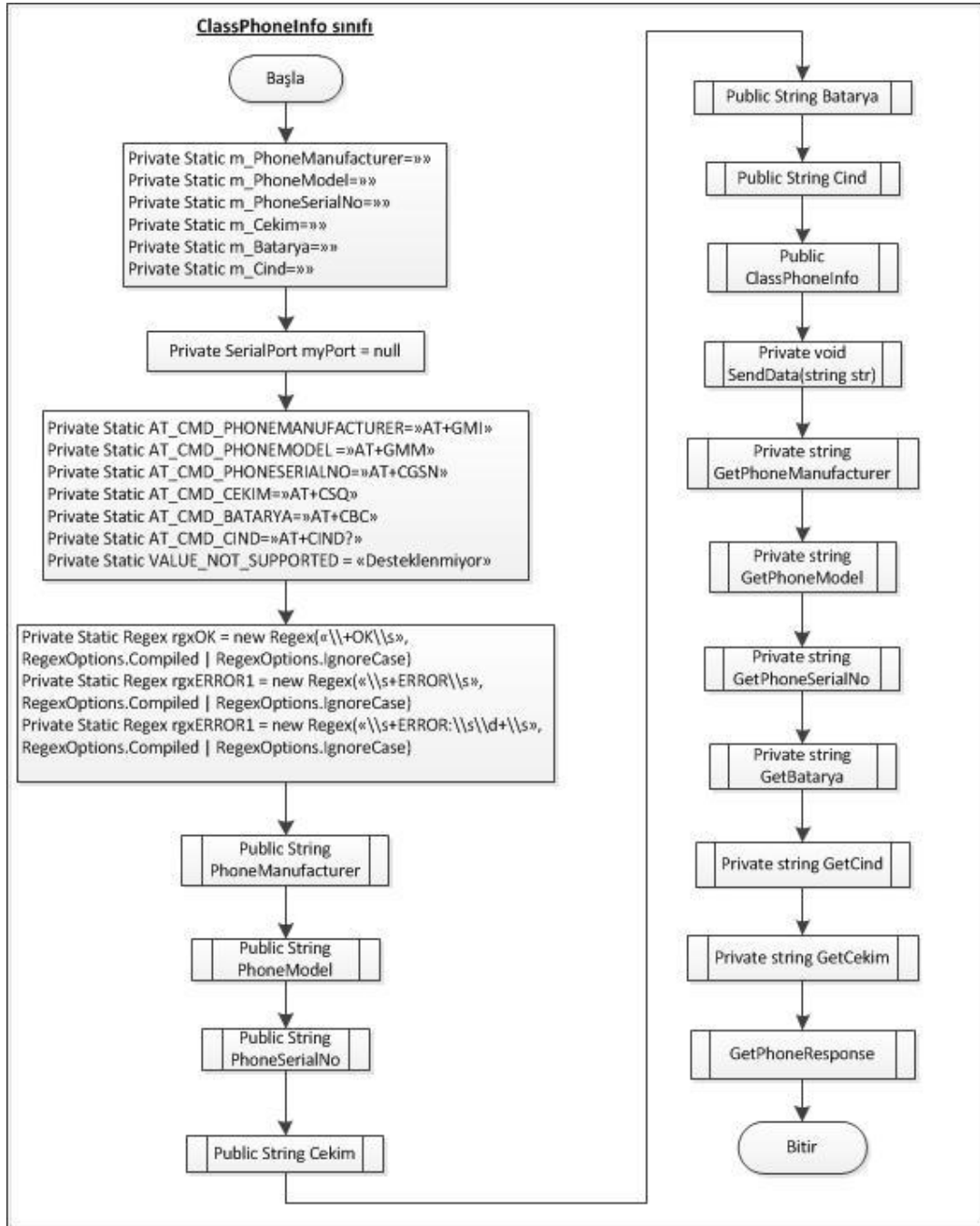
51. Yomralıođlu, T. ve Döner, F., “Mobil GIS: gezici cođrafi bilgi sistemleri ve uygulamaları”, *Hkm - Jeodezi, Jeoinformasyon ve Arazi Yönetimi Dergisi*, 93: 1-9 (2005).
52. Dayak, M.E., “3. nesil gezgin telefonlar üzerinde multimedya destekli güvenlik artırımı tekniklerinin araştırılması”, *Kahramanmaraş Sütçü İmam Üniversitesi Fen Bilimleri Enstitüsü*, Kahramanmaraş, 4 (2007).
53. İnternet: Admob Mobile Metrics “Metrics Highlights” <http://metrics.admob.com> (2010).
54. Birođul, S., “GSM şebekelerinde frekans planlamasının veri füzyonu ile gerçekleştirilmesi”, Doktora Tezi, *Gazi Üniversitesi Fen Bilimleri Enstitüsü*, Ankara, 39-40 (2008).
55. Arslan, S., “GSM 900 baz istasyon servis sayılarının simülasyon ve amaç programlama tekniklerinin kullanılması” Yüksek Lisans Tezi, *Gazi Üniversitesi Fen Bilimleri Enstitüsü*, Ankara, (2000).
56. Candan, M.M., “Üçüncü nesil mobil haberleşme sistemleri için Türkiye’de uygulanacak frekans bandı, lisans, servisler, uygulamalar ve ülkemizdeki durumu”, Uzmanlık Tezi, *Telekomünikasyon Kurumu*, Ankara, (2002).
57. Taşkın, Ç. ve Emel, G.G., “Veri madenciliğinde kümeleme yaklaşımları ve kohonen ađları ile perakendecilik sektöründe bir uygulama”, *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 15(3): 395-409 (2010).

EKLER

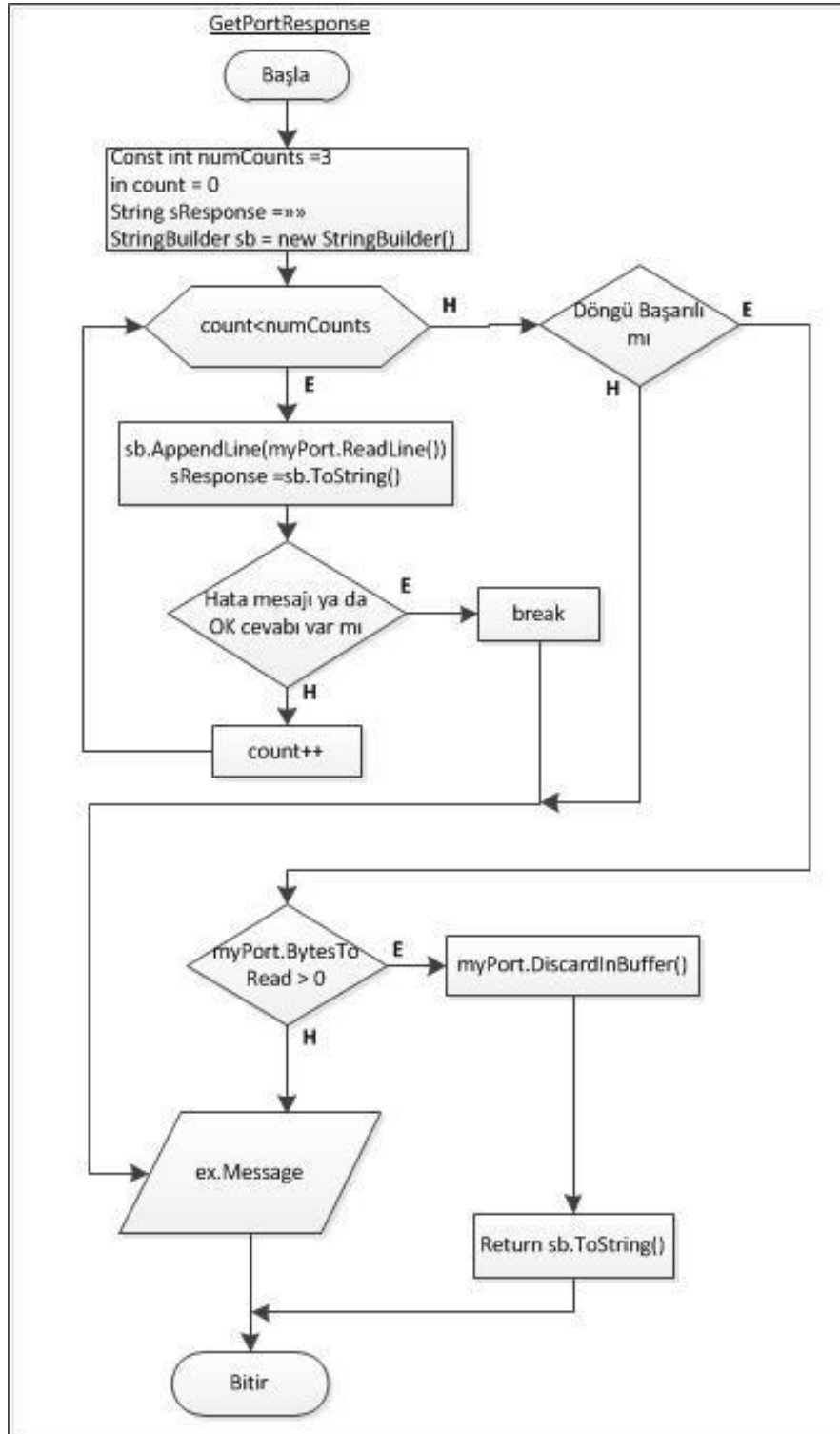
EK-1. Tasarlanan programın akış diyagramı



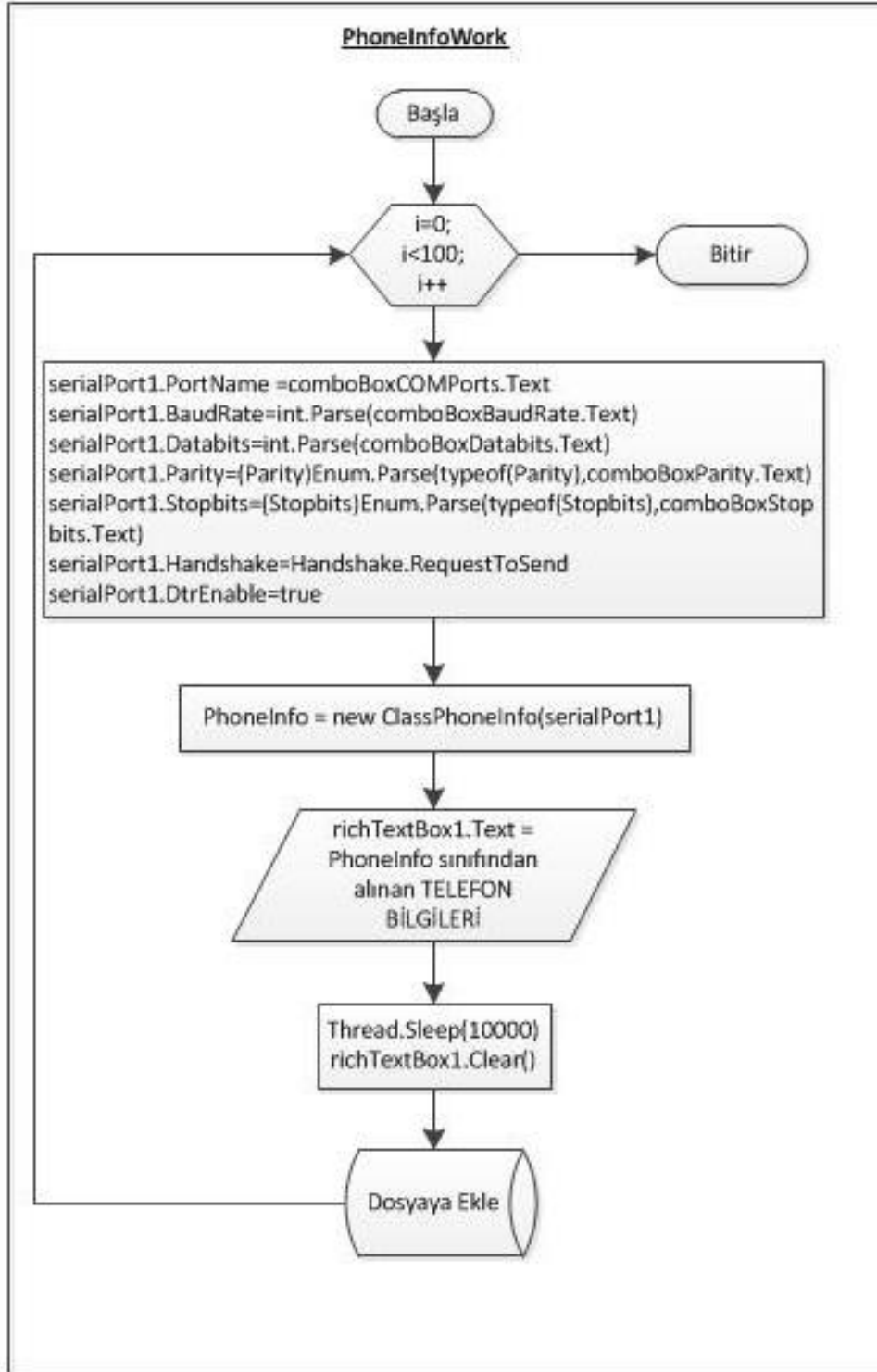
EK-1 (Devam). Tasarlanan programın akış diyagramı



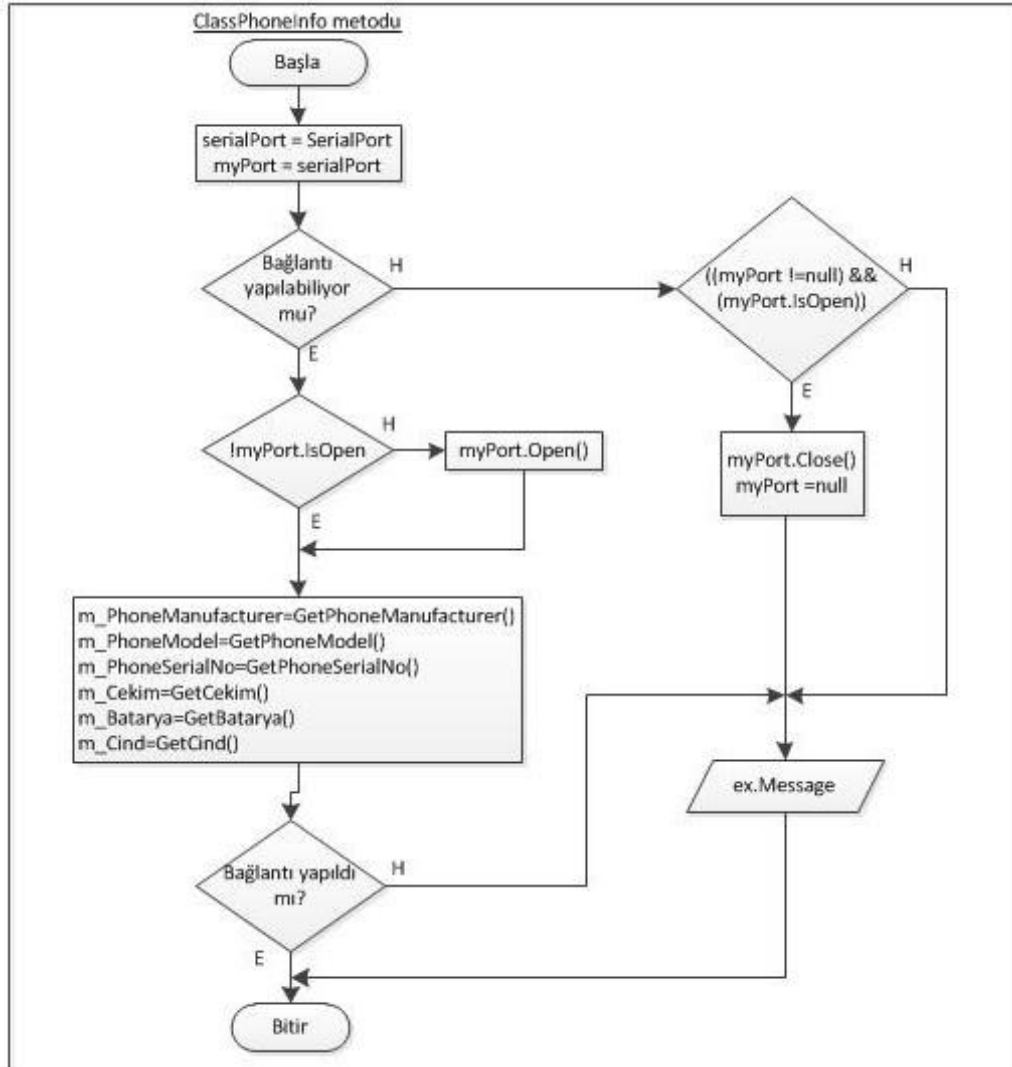
EK-1 (Devam). Tasarlanan programın akış diyagramı



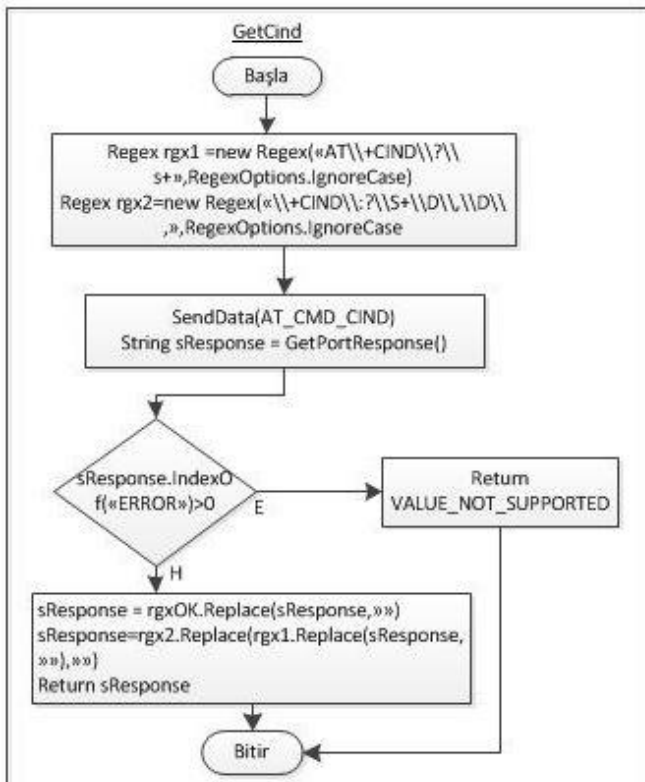
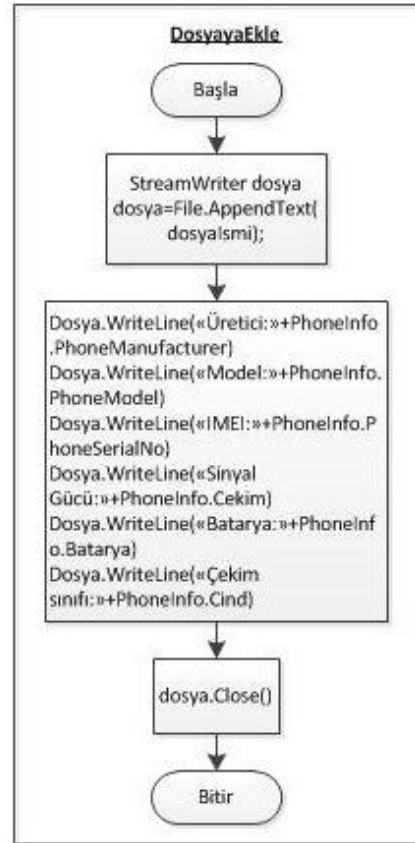
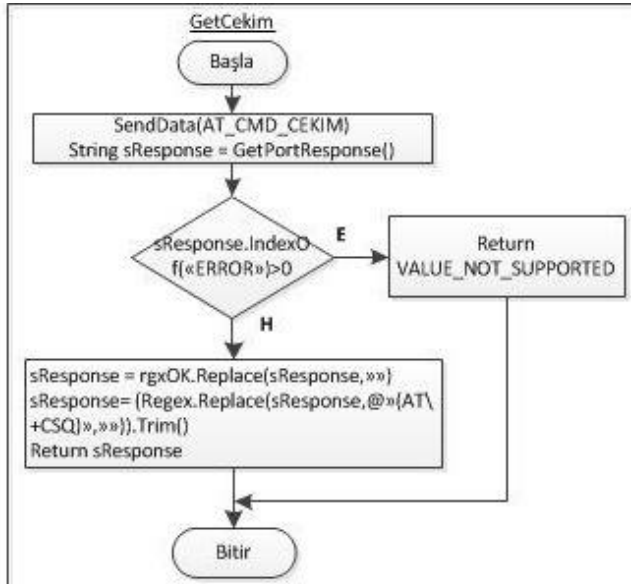
EK-1 (Devam). Tasarlanan programın akış diyagramı



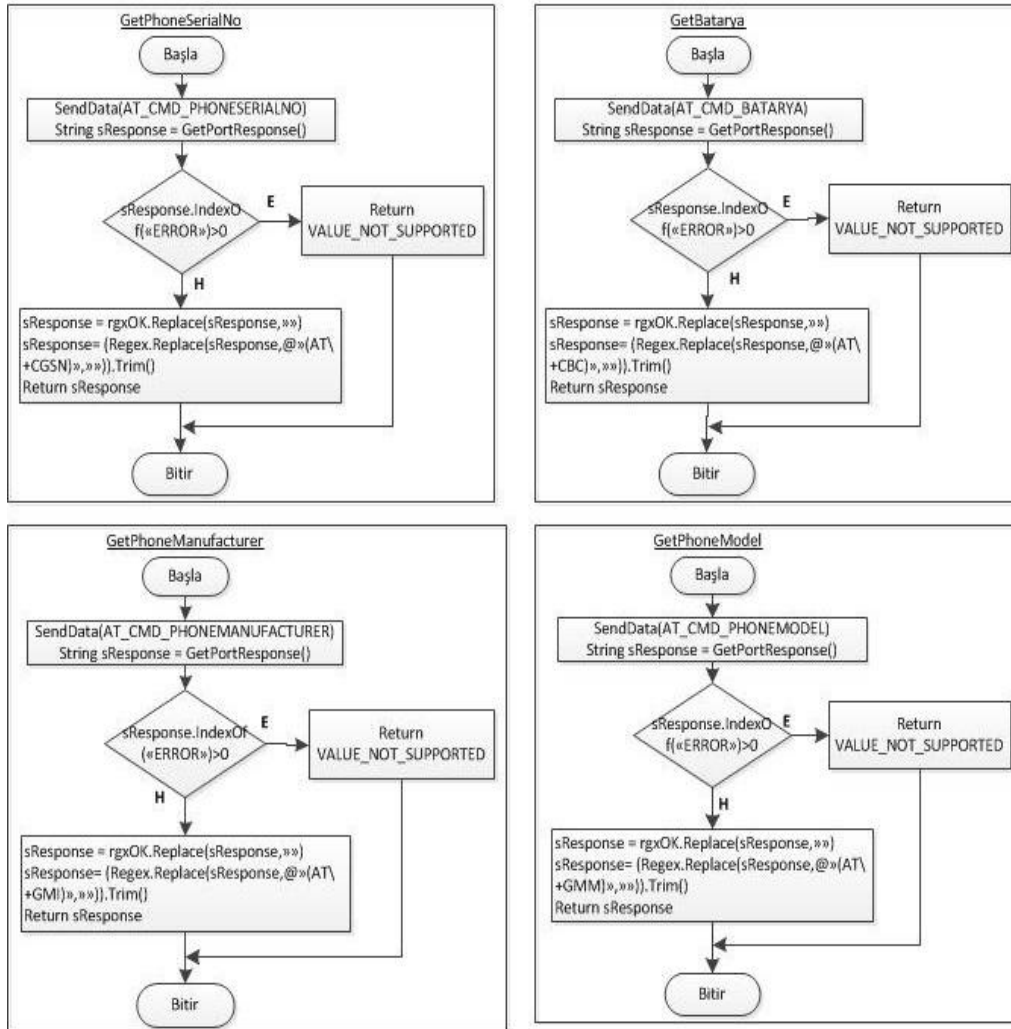
EK-1 (Devam). Tasarlanan programın akış diyagramı



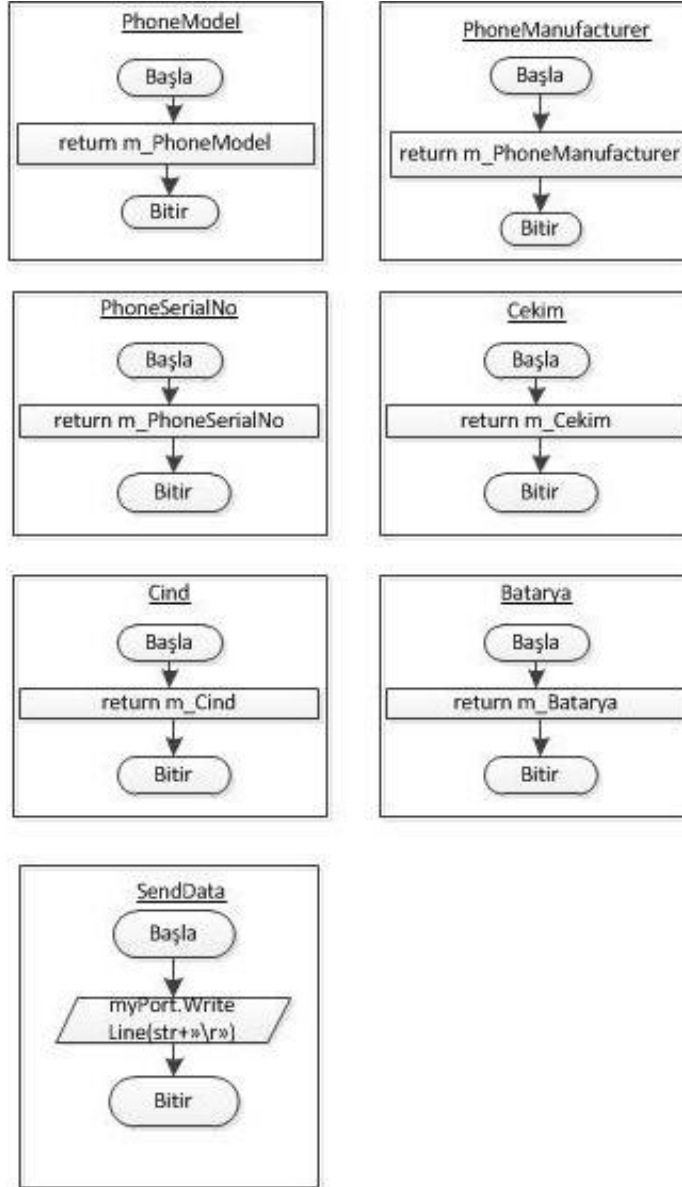
EK-1 (Devam). Tasarlanan programın akış diyagramı



EK-1 (Devam). Tasarlanan programın akış diyagramı



EK-1 (Devam). Tasarlanan programın akış diyagramı



EK-2. A, B ve C şebekeleri için iki aşamalı kümeleme sonuçları

A ŞEBEKESİ												
Kümelere	11,99		13,99		14,99		15,99		16,99		17,99	
	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde
cluster 1	4	100%	6	100%	2	100%	13	100%	70	100%	16	100%
cluster 2	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
cluster 3	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
combined	4	100%	6	100%	2	100%	13	100%	70	100%	16	100%

A ŞEBEKESİ												
Kümelere	18,99		20,99		28,99		29,99		8,99			
	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde		
cluster 1	1	100%	2	100%	0	0%	0	0%	1	100%		
cluster 2	0	0%	0	0%	79	100%	25	14,20%	0	0%		
cluster 3	0	0%	0	0%	0	0%	151	85,80%	0	0%		
combined	1	100%	2	100%	79	100%	176	100%	1	100%		

B ŞEBEKESİ												
Kümelere	10,99		11,99		13,99		14,99		15,99		16,99	
	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde
cluster 1	7	100%	3	100%	3	100%	3	100%	20	100%	26	100%
cluster 2	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
cluster 3	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
combined	7	100%	3	100%	3	100%	3	100%	20	100%	26	100%

B ŞEBEKESİ												
Kümelere	17,99		18,99		19,99		21,99		22,99		24,99	
	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde
cluster 1	4	17,40%	16	100%	4	12,90%	16	9,60%	0	0%	6	100%
cluster 2	19	82,60%	0	0%	27	87,10%	0	0%	57	100%	0	0%
cluster 3	0	0,00%	0	0%	0	0%	151	90,40%	0	0%	0	0%
combined	23	100%	16	100%	31	100%	167	100%	57	100%	6	100%

B ŞEBEKESİ												
Kümelere	6,99		8,99		9,99		27,99					
	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde				
cluster 1	1	100%	3	100%	3	100%	0	0%				
cluster 2	0	0%	0	0%	0	0%	1	100%				
cluster 3	0	0%	0	0%	0	0%	0	0%				
combined	1	100%	3	100%	3	100%	1	100%				

EK-2 (Devam). A, B ve C şebekeleri için iki aşamalı kümeleme sonuçları

C ŞEBEKESİ												
Kümelere	10,99		11,99		12,99		13,99		14,99		15,99	
	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde
cluster 1	7	100%	2	100%	1	100%	9	100%	1	100%	24	85,70%
cluster 2	0	0%	0	0%	0	0%	0	0%	0	0%	0	0
cluster 3	0	0%	0	0%	0	0%	0	0%	0	0%	4	14,30%
combined	7	100%	2	100%	1	100%	9	100%	1	100%	28	100%
C ŞEBEKESİ												
Kümelere	16,99		17,99		18,99		19,99		21,99		22,99	
	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde
cluster 1	24	46,20%	9	6,30%	0	0%	5	100%	13	59,10%	11	13,10%
cluster 2	27	51,90%	68	47,60%	0	0%	0	0%	9	40,90%	0	0%
cluster 3	1	1,90%	66	46,20%	7	100%	0	0%	0	0%	73	86,90%
combined	52	100%	143	100%	7	100%	5	100%	22	100%	84	100%
C ŞEBEKESİ												
Kümelere	6,99		7,99		8,99							
	frekans	yüzde	frekans	yüzde	frekans	yüzde						
cluster 1	3	100%	1	100%	5	100%						
cluster 2	0	0%	0	0%	0	0%						
cluster 3	0	0%	0	0%	0	0%						
combined	3	100%	1	100%	5	100%						

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : SAVAŞ, Serkan
Uyruğu : T.C.
Doğum tarihi ve yeri : 14.10.1982 Ankara
Medeni hali : Evli
e-mail : serkan_savas@hotmail.com

Eğitim

Derece	Eğitim Birimi	Mezuniyet tarihi
Lisans	Gazi Üniversitesi/ Elkt.-Bilg. Bölümü	2006
Lise	Anafartalar And. Tic. Mes. Lisesi	2000

İş Deneyimi

Yıl	Yer	Görev
2006- --	Kızılcahamam Tek. ve End. Mes. L.	Bilişim Tek. Öğretmeni

Yabancı Dil

İngilizce

Hobiler

Bilgisayar teknolojileri, Seyahat, Spor.