

VERİ MADENCİLİĞİ YÖNTEMİ İLE GSM ŞEBEKELERİNİN PERFORMANS ANALİZİ

Serkan SAVAŞ*, **Nurettin TOPALOĞLU****

* Kızılcahamam Teknik ve Endüstri Meslek Lisesi, Kızılcahamam, 06890, Ankara

** Gazi Üniversitesi Teknoloji Fakültesi, Beşevler, 06500, Ankara

serkan_savas@hotmail.com, nurettin@gazi.edu.tr

(Geliş/Received: 26.05.2011; Kabul/Accepted: 16.09.2011)

ÖZET

Bu çalışmada, farklı GSM şebekelerinin çekim gücü verilerini cep telefonlarından alarak veri tabanı oluşturan bir program geliştirilmiştir. Programın rastgele seçilen bir güzergâh olan Kızılcahamam ilçesi ile Gazi Üniversitesi Merkez Kampüsü arasında çalıştırılmasıyla elde edilen verilerden bir veri bankası oluşturulmuştur. Bu veriler, veri madenciliği yöntemlerinden iki aşamalı kümeleme tekniği kullanılarak analiz edilmiştir. Bu çalışmanın sonunda; A, B ve C şebekelerinin performansları belirlenmiş olup, sinyal gücü temel alındığında A şebekesinin en yüksek çekim gücüne ve istikrarlı çekim düzeyine, sinyal gücü aralıkları sınıflandırıldığında ise B şebekesinin en yüksek iyi sinyal oranına sahip olduğu gözlenmiştir.

Anahtar Kelimeler: GSM şebekeleri, veri madenciliği, iki aşamalı kümeleme, sinyal gücü analizi

PERFORMANCE ANALYSIS OF GSM NETWORKS WITH DATA MINING METHOD

ABSTRACT

In this study, a program, which creates a database by getting signal strength data of different GSM networks from mobile phones, is developed. The database is obtained by data which were collected by running program on a random route between center campus of Gazi University and Kızılcahamam County. These data are analyzed by using two step clustering method, which is a type of data mining method. The performance of A, B and C GSM networks is determined and network A has the highest signal strength, and the most stable signal level based on the signal strength classification, while network B has the highest signal ratio when the signal strength ranges is classified.

Keywords: GSM networks, data mining, two step cluster, signal strength analysis.

1. GİRİŞ (INTRODUCTION)

Bilgi ve iletişim teknolojilerini kullanan tüm kurumlar, bilgisayar sistemleri ve elektronik araçların ürettikleri verileri, kurumlarının veri ambarlarına kaydetmektedirler. Veri ambarlarında tutulan veriler zamanla çok büyük miktarlara ulaşmakta ve çıplak gözle bakıldığında bu verilerin içerisinde anlamlı bilgiye ulaşmak zorlaşmaktadır. Bu veriler ancak belirli bir amaç için işlendiğinde anlam ifade etmeye başlar [1]. Ham veriye dayalı kararlar almak veya geçmişte yaşanan kötü bir tecrübeden kaynaklanan kayıpların 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. Veri madenciliği, bu gibi durumlarda kullanılan, büyük miktardaki veri kümelerinde saklı durumda bulunan örüntü ve eğilimleri keşfetme işlemidir [3].

Veri madenciliği, çok büyük miktarda bilginin depolandığı veri tabanlarından, amacımız doğrultusunda, gelecek ile ilgili tahminler yapmamızı sağlayabilecek, anlamlı olan veriye ulaşma ve veriyi kullanma işlemidir [4]. Veri madenciliği alanında

daha önce yapılan çalışmaları mühendislik, eğitim, ticaret, mobil, bankacılık ve finans başlıkları altında birleştirmek mümkündür. Veri madenciliği yöntemleriyle deprem verilerinin analizi, asenkron motorlarda hata tespiti, veri madenciliği ile yüz tanıma sistemi, melez bir PSO/ACO algoritması oluşturma ve hiyerarşik kümeleme algoritmalarının ortak noktalarının bulunması gibi çalışmalar mühendislik alanında yapılan çalışmalardan birkaçıdır [5-9]. Eğitim alanında; öğrenci işleri otomasyonu, uzaktan eğitimde “log” verilerinin analizi, lise türü ve mezuniyet başarısının kazanılan fakülteyle ilişkisi ile ÖSS başarı analizi gibi çalışmalar veri madenciliği yöntemleriyle gerçekleştirilmiştir [2,10-12]. Ticaret alanında; sepet analizi, müşteri analizi, satış verileri analizi, parça raporları analizi, online seyahat endüstrisine veri madenciliği uygulaması, havayolu pazarında müşteri ilişkileri yönetimi gibi çalışmalar, birliktelik kuralları, kümeleme yöntemleri gibi veri madenciliği yöntemleri ve teknikleriyle gerçekleştirilmiştir [1,13-17]. Mobil alanda kullanıcıların hareketlerini tahmin etmeye yönelik bir çalışma gerçekleştirilmiştir [18]. Bankacılık ve finans alanlarında ise, kredi kartı kullanım analizleri, İMKB verilerinin değerlendirilmesi ve hileli finansal tabloların tespiti, bankaların faaliyetlerine göre sınıflandırılmaları gibi çalışmalar yine veri madenciliği yöntemleriyle gerçekleştirilmiştir [19-21]. Bu çalışmalardan anlaşılmaktadır ki; veri madenciliği alanında yapılan araştırmalar geçen zamanla birlikte çeşitlendirilmekte, bilgi ve iletişim alanındaki gelişmelere ayak uydurmaktadır.

Bilgi ve iletişim teknolojilerinde yaşanan gelişmelere paralel olarak, bu alanlarda kullanılan yazılım teknolojileri de sürekli bir değişim ve gelişim halindedir. İnsanların yeni gözdesi, zaman ve mekândan bağımsız olmak ile hareketlilik ve hareket yeteneği anlamına gelen “mobilité” olmuştur [22]. Artık sabit bilgisayar sistemlerinin yerini cep bilgisayarları, tablet pc’ler ve akıllı telefonlar gibi mobil cihazlar almaktadır.

Ülkemizde mobil hat abone sayısının neredeyse ülke nüfusu kadar olduğu düşünüldüğünde, mobil iletişimin insanlar için ne kadar önemli bir konuma geldiği anlaşılmaktadır. Günümüzde mobil iletişim hizmeti sunan işletmeler için çok çetin bir rekabet ortamı oluşmuştur. Özellikle numara taşınabilirliğine imkân sağlandıktan sonra piyasadaki rekabet daha da artmıştır. GSM 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? Çünkü her ne kadar tüm şebekeler kapsama alanlarının yüksek oranlarını duyursalar da bölgesel değişiklikler cep telefonlarının sinyal alım gücüne etki etmektedir.

GSM şebekelerinin mevcut müşterilerini korumak ve yeni müşteriler kazanmak için yaptıkları tanıtım reklamları incelendiğinde, “Uygun Fiyat”, “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 durum göstermektedir ki insanlar için, kullandığı GSM şebekesinin çekim gücü yüksek ve fiyatları da uygun olmalıdır. Türkiye’de mobil iletişim hizmetlerini sunan üç adet GSM şebekesi bulunmaktadır. Bunlar; Turkcell, Vodafone ve Avea’dır. Bu işletmelerin 2010 yılı 4. çeyreğindeki abone sayıları; Turkcell 33,47 milyon, Vodafone 16,68 milyon ve Avea 11,62 milyondur. Türkiye’deki abone sayısı böylece toplam 61,77 milyondur [23].

Mobil iletişim alanı bu çalışmada, veri madenciliği uygulaması için kullanılan verilerin sağlandığı alandır. Bu çalışmanın amacı; GSM şebekelerinin çekim gücü verilerini, veri madenciliği yöntemleriyle analiz ederek bulgular elde etmek ve bunları yorumlayarak öngörüler oluşturmaktır.

2. VERİ MADENCİLİĞİNİN GSM ŞEBEKELERİNE UYGULANMASI (DATA MINING APPLICATION TO GSM NETWORKS)

Veri madenciliği disiplinini, mobil iletişim alanına uygulamak için bir program geliştirilmiştir. Bu program vasıtasıyla GSM şebekelerinin sinyal gücü verileri cep telefonundan alınarak bir veri tabanı oluşturulmuştur. Veri tabanına kaydedilen bu verilere iki aşamalı kümeleme yöntemi uygulanmış, böylece GSM şebekelerinin belirlenen bir güzergâhtaki çekim gücü performansları değerlendirilmiş ve sonuçları yorumlanmıştır.

2.1. Program Tasarımında Kullanılan Donanım ve Yazılım Sistemleri (Hardware and Software Systems Used in Program Design)

Cep telefonundan sinyal bilgisi alabilmek için planlanan güzergâhta kullanılmak üzere, mobil bir bilgisayar sistemi gerekmektedir. Veri tabanı oluşturulurken kullanılan bilgisayar sistemi ve diğer cihazlar:

- Bir dizüstü bilgisayar,
- Cep Telefonu,
- USB portundan bilgisayara bağlantı kablosu,
- GSM şebekeleri SIM kartlarıdır.

GSM şebekelerinin sinyal verilerini elde etmek için hazırlanan program tasarımı için, MS Visual Studio C#.NET programlama dili kullanılmıştır. Ayrıca kullanılan cep telefonunun USB port ve Bluetooth ile bilgisayara bağlantı sırasında sorunsuz çalışması için, bilgisayara USB Modem ve Bluetooth Modem sürücülere kurulmuştur.

A ŞEBEKESİ		B ŞEBEKESİ		C ŞEBEKESİ	
SİNYAL DEĞERİ	SINIFLANDIRMA	SİNYAL DEĞERİ	SINIFLANDIRMA	SİNYAL DEĞERİ	SINIFLANDIRMA
20,99	İYİ SİNYAL	15,99	ORTA SİNYAL	15,99	ORTA SİNYAL
20,99	İYİ SİNYAL	15,99	ORTA SİNYAL	15,99	ORTA SİNYAL
17,99	İYİ SİNYAL	15,99	ORTA SİNYAL	15,99	ORTA SİNYAL
17,99	İYİ SİNYAL	15,99	ORTA SİNYAL	15,99	ORTA SİNYAL
17,99	İYİ SİNYAL	15,99	ORTA SİNYAL	15,99	ORTA SİNYAL
17,99	İYİ SİNYAL	15,99	ORTA SİNYAL	15,99	ORTA SİNYAL
17,99	İYİ SİNYAL	15,99	ORTA SİNYAL	13,99	ORTA SİNYAL
17,99	İYİ SİNYAL	15,99	ORTA SİNYAL	12,99	ORTA SİNYAL
17,99	İYİ SİNYAL	15,99	ORTA SİNYAL	10,99	ZAYIF SİNYAL
17,99	İYİ SİNYAL	15,99	ORTA SİNYAL	11,99	ZAYIF SİNYAL
17,99	İYİ SİNYAL	15,99	ORTA SİNYAL	15,99	ORTA SİNYAL
17,99	İYİ SİNYAL	15,99	ORTA SİNYAL	14,99	ORTA SİNYAL
17,99	İYİ SİNYAL	13,99	ORTA SİNYAL	6,99	YETERSİZ SİNYAL
17,99	İYİ SİNYAL	16,99	ORTA SİNYAL	6,99	YETERSİZ SİNYAL
17,99	İYİ SİNYAL	16,99	ORTA SİNYAL	13,99	ORTA SİNYAL
16,99	ORTA SİNYAL	16,99	ORTA SİNYAL	13,99	ORTA SİNYAL
15,99	ORTA SİNYAL	16,99	ORTA SİNYAL	10,99	ZAYIF SİNYAL
15,99	ORTA SİNYAL	14,99	ORTA SİNYAL	8,99	ZAYIF SİNYAL
15,99	ORTA SİNYAL	11,99	ZAYIF SİNYAL	8,99	ZAYIF SİNYAL
11,99	ZAYIF SİNYAL	16,99	ORTA SİNYAL	8,99	ZAYIF SİNYAL
14,99	ORTA SİNYAL	16,99	ORTA SİNYAL	8,99	ZAYIF SİNYAL
14,99	ORTA SİNYAL	16,99	ORTA SİNYAL	7,99	ZAYIF SİNYAL
15,99	ORTA SİNYAL	16,99	ORTA SİNYAL	19,99	İYİ SİNYAL
15,99	ORTA SİNYAL	19,99	İYİ SİNYAL	19,99	İYİ SİNYAL
15,99	ORTA SİNYAL	19,99	İYİ SİNYAL	19,99	İYİ SİNYAL
17,99	İYİ SİNYAL	13,99	ORTA SİNYAL	19,99	İYİ SİNYAL
8,99	ZAYIF SİNYAL	18,99	İYİ SİNYAL	19,99	İYİ SİNYAL

Şekil 3. Elde edilen veri tablosundan bir bölüm (A part of the data table)

2.3. Veri Madenciliği Süreci (Data Mining Process)

Tasarlanan programın kullanıldığı şebekelerin isimleri, yasal sorumluluktan dolayı belirtilmemiştir. GSM şebekelerinin isimleri bu çalışmanın ileri bölümlerinde “A”, “B” ve “C” Şebekeleri olarak anılmıştır. Program çalıştırdıktan sonra elde edilen veriler bir dosyaya kaydedilmiştir. Bu aşamadan sonra oluşturulan veri tabanı üzerinde veri madenciliği süreci başlatılmıştır. Oluşturulan tabloya değinmeden önce belirtilmesi gereken bir konu da sinyal gücünün sınıflandırılmasıdır. Sinyal gücünün sınıflandırılması Tablo 1’de gösterilmiştir.

Tablo 1. Sinyal gücünün sınıflandırılması (Signal strength classification) [24]

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

Oluşturulan sinyal gücü tablosunun bir bölümü Şekil 3’te gösterilmiştir. Bu tabloda her bir şebeke için 370 adet sinyal verisi bulunmaktadır.

2.3.1. Veri madenciliği yönteminin uygulanması (Application of data mining method)

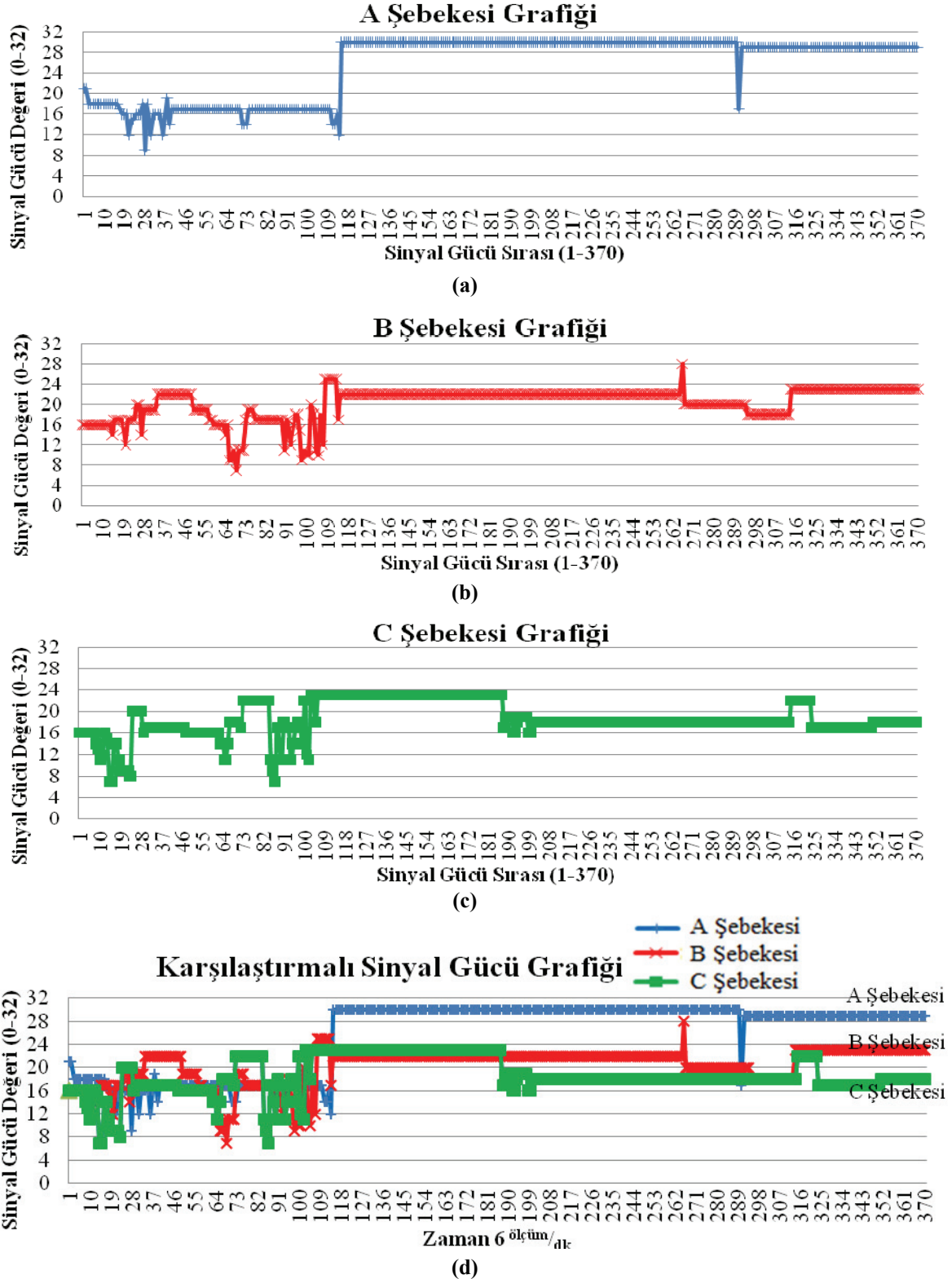
Her bir şebeke için ayrı ayrı elde edilen sinyal gücü verilerinden oluşturulan zaman-sinyal gücü grafikleri Şekil 4’te belirtilmiştir.

Şekil 4 ve Tablo 2 incelendiğinde şu sonuçlara ulaşılmıştır: 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 olan 29,99 ve 28,99 değerlerinde, uzun süreli ölçümler sağlanabilmiştir. A Şebekesi tüm verilerin ortalamasında 25,99 değerine yaklaşarak, aritmetik ortalamalar baz alındığında en yüksek ortalamayı sağlayan şebeke olmayı başarmıştır.

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 ölçülmüştür. B Şebekesi bir noktada yetersiz sinyal seviyesine düşmüştür. B Şebekesinin sinyal gücü, aritmetiksel ortalama ile incelendiğinde 20,99 seviyesine yakın bir seviyede olmuş ve bu üç şebeke içerisinde aritmetik ortalama, 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 Şebekesinin sinyal gücü, ölçüm yapılan

üç noktada yetersiz sinyal seviyesine düşmüştür. Bu anlamda diğer şebekelerle kıyaslandığında en çok yetersiz sayıda ölçüm yapılan şebeke olmuştur. Ayrıca aritmetik ortalama alındığında 18,99 ortalama ile en düşük sinyal gücü ortalaması elde edilen şebeke, C Şebekesi olmuştur.



Şekil 4. A (a), B (b), C (c) Şebekeleri ve Karşılaştırmalı sinyal gücü grafikleri (d) (Networks A, B, C and comparative signal strength graphics)

Tablo 2. A (a), B (b) ve C (c) Şebekeleri için sinyal gücü tabloları (Signal strength tables of networks A, B and C)

A Şebekesi: Değer	Frekans	Yüzde	Değer Yüzdesi	Kümülatif 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	

(a)

B Şebekesi: Değer	Frekans	Yüzde	Değer Yüzdesi	Kümülatif 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	

(b)

C Şebekesi: Değer	Frekans	Yüzde	Değer Yüzdesi	Kümülatif 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	

(c)

Not: Kullanılan programda, ondalıklı kesirler yuvarlanmıştır.

Her bir şebekenin sinyal gücü tablosunu inceledikten sonra bir de sınıflandırılmış aralıklara göre inceleme gerçekleştirilmiştir. Burada, aslında veri madenciliğinin asıl amacını oluşturan, ilk bakışta kestirilemeyen ancak daha sonra veri madenciliğiyle ortaya çıkan ilginç bir bilgiye ulaşılmıştır (Tablo 3).

Tablo3. A, B ve C Şebekeleri için sınıflandırılmış frekans tablosu (Classified frequency tables of Networks A, B and C)

A Şebekesi: Değer	Frekans	Yüzde	Değer Yüzdesi	Kümülatif 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	

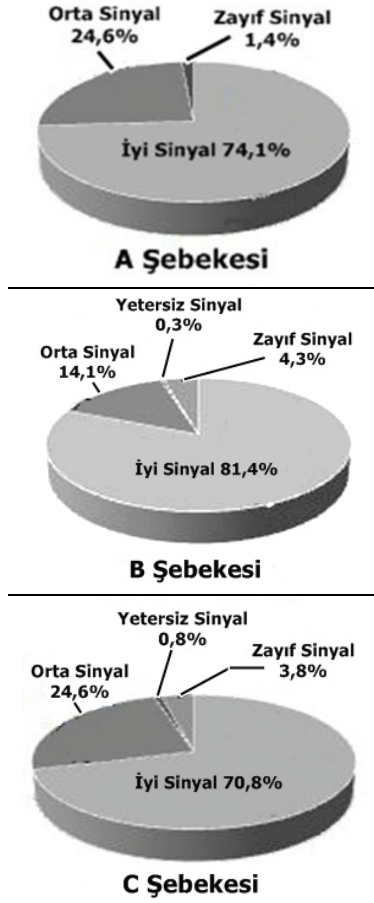
B Şebekesi: Değer	Frekans	Yüzde	Değer Yüzdesi	Kümülatif Yüzde
iyi sinyal	301	81,4	81,4	81,4
orta sinyal	52	14,1	14,1	95,4
yetersiz sinyal	1	0,3	0,3	95,7
zayıf sinyal	16	4,3	4,3	100
Toplam	370	100	100	

C Şebekesi: Değer	Frekans	Yüzde	Değer Yüzdesi	Kümülatif 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 ölçümlerinde, en yüksek değere ulaşılan ve en yüksek oranda yüksek çekim gücü tespit edilen şebeke A Şebekesi olarak görülürken, sınıflandırılmış aralıklara göre değerlendirme yapıldığında, 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. Sinyal gücü 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. Buradan anlaşıldığı üzere, ilk bakışta sinyal gücü performansı en iyi olan şebeke olarak A Şebekesi düşünülmesine rağmen, sınıflandırma sonucunda B Şebekesinin sinyal gücü performansının en iyi olduğu ortaya çıkmıştır. Yüksek sinyal gücü oranı her ne kadar A Şebekesi için olumlu bir etken olsa da, seçilen güzergâhta B Şebekesi iyi sinyal sınıfında daha çok oran yakalayarak performansı en iyi şebeke olmuştur. C Şebekesi her iki değerlendirmede de, sinyal gücü performansı en

düşük şebeke olmuştur. Tüm şebekeler için sınıflandırılmış sinyal gücü grafikleri Şekil 5'te gösterilmiştir.



Şekil 5. A, B ve C Şebekeleri için sınıflandırılmış sinyal gücü grafiği (Classified signal strength graphics of Networks A, B and C)

Tasarlanan program ile elde edilen sinyal gücü ölçüm verilerine “İki Aşamalı Kümeleme” yöntemleri uygulanmıştır. Sinyal gücü verilerine, veri madenciliği yöntemleri içerisinde uygulanabilecek en uygun yöntemler tanımlayıcı yöntemlerdir. Tanımlayıcı yöntemler içerisinde yer alan kümeleme yöntemleri ve kümeleme yöntemleri içerisinde de, bu verilere uygulanması sonucunda en iyi sonuçları veren iki aşamalı kümeleme yöntemi, bu çalışmada kullanılan veri madenciliği yöntemi olmuştur.

2.4.1.1. İki Aşamalı Kümeleme (Two step cluster)

İki aşamalı kümeleme algoritmalarında, ilk akla gelen Punjve Steward (1983) tarafından önerilen klasik iki aşamalı kümeleme algoritmasıdır. Bu algoritma, Ward'ın minimum varyans yöntemi ile “Kmeans” yönteminden oluşan bir hibrid yaklaşımdır [25]. Büyük veri tabanlarını analiz etmek için tasarlanmıştır. 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 çözümü ve kümeleme aşamalarını gerçekleştirmektedir. Ön kümeleme sırasında, verilerin her biri incelenerek 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 çözümü aşamasında, herhangi bir kümeye dahil edilemeyen veriler değerlendirilir. Ekleme işlemi için tüm denemeler gerçekleştirildikten sonra dahiliet yine sağlanamaz ise, bu veriler dış veriler olarak ayrılır. Kümeleme aşamasında ise bir ağaç yapısı oluşturulur. Tüm veriler kökten yapraklara doğru dağıtmaya 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 [4].

Verilere iki aşamalı kümeleme uygulaması sırasında, otomatik olarak BIC yöntemine göre kümeleme işlemi gerçekleştirilmiştir. Daha sonra bu kümeler olasılık uzaklığı ölçütüne göre 3 kümede kombine edilmiştir. Bunun sonucunda da ortaya çıkan küme dağılımı Tablo 4'teki gibi olmuştur.

Tablo 4. Sinyal gücü için iki aşamalı kümeleme sonuçları (Two step cluster results for signal strength)

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

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 Tablo 5'teki gibi birleştirilerek sunulmuştur.

Tablo 5 incelendiğinde her bir küme grubu için hangi şebekeden, hangi sinyal gücünden, kaç eleman seçildiği ilk bakışta fark edilmeyebilir. Bu sebeple şöyle bir düzenlemeye gidilebilir.

Tablo 5. İki aşamalı kümelemede şebekeler için sinyal gücü dağılımı (The distribution of signal strength for networks with two step cluster)

A ŞEBEKESİ												
Küme	11,99		13,99		14,99		15,99		16,99		17,99	
	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde
küme 1	4	100%	6	100%	2	100%	13	100%	70	100%	16	100%
küme 2	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
küme 3	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
kombine	4	100%	6	100%	2	100%	13	100%	70	100%	16	100%
A ŞEBEKESİ												
Küme	18,99		20,99		28,99		29,99		8,99			
	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde		
küme 1	1	100%	2	100%	0	0%	0	0%	1	100%		
küme 2	0	0%	0	0%	79	100%	25	14,20%	0	0%		
küme 3	0	0%	0	0%	0	0%	151	85,80%	0	0%		
kombine	1	100%	2	100%	79	100%	176	100%	1	100%		
B ŞEBEKESİ												
Küme	10,99		11,99		13,99		14,99		15,99		16,99	
	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde
küme 1	7	100%	3	100%	3	100%	3	100%	20	100%	26	100%
küme 2	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
küme 3	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%
kombine	7	100%	3	100%	3	100%	3	100%	20	100%	26	100%
B ŞEBEKESİ												
Küme	17,99		18,99		19,99		21,99		22,99		24,99	
	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde
küme 1	4	17,40%	16	100%	4	12,90%	16	9,60%	0	0%	6	100%
küme 2	19	82,60%	0	0%	27	87,10%	0	0%	57	100%	0	0%
küme 3	0	0,00%	0	0%	0	0%	151	90,40%	0	0%	0	0%
kombine	23	100%	16	100%	31	100%	167	100%	57	100%	6	100%
B ŞEBEKESİ												
Küme	6,99		8,99		9,99		27,99					
	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde				
küme 1	1	100%	3	100%	3	100%	0	0%				
küme 2	0	0%	0	0%	0	0%	1	100%				
küme 3	0	0%	0	0%	0	0%	0	0%				
kombine	1	100%	3	100%	3	100%	1	100%				
C ŞEBEKESİ												
Küme	10,99		11,99		12,99		13,99		14,99		15,99	
	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde
küme 1	7	100%	2	100%	1	100%	9	100%	1	100%	24	85,70%
küme 2	0	0%	0	0%	0	0%	0	0%	0	0%	0	0
küme 3	0	0%	0	0%	0	0%	0	0%	0	0%	4	14,30%
kombine	7	100%	2	100%	1	100%	9	100%	1	100%	28	100%
C ŞEBEKESİ												
Küme	16,99		17,99		18,99		19,99		21,99		22,99	
	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde	frekans	yüzde
küme 1	24	46,20%	9	6,30%	0	0%	5	100%	13	59,10%	11	13,10%
küme 2	27	51,90%	68	47,60%	0	0%	0	0%	9	40,90%	0	0%
küme 3	1	1,90%	66	46,20%	7	100%	0	0%	0	0%	73	86,90%
kombine	52	100%	143	100%	7	100%	5	100%	22	100%	84	100%
C ŞEBEKESİ												
Küme	6,99		7,99		8,99							
	frekans	yüzde	frekans	yüzde	frekans	yüzde						
küme 1	3	100%	1	100%	5	100%						
küme 2	0	0%	0	0%	0	0%						
küme 3	0	0%	0	0%	0	0%						
kombine	3	100%	1	100%	5	100%						

$\text{Şebeke}_{(\text{küme no})} = \{ [\text{Sinyal gücü}](\text{seçilen eleman sayısı}) \} = \text{toplam eleman sayısı}$ (1)

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

Eşitlik 1'den yola çıkarak her bir şebeke için küme elemanları gösterildiğinde ortaya çıkan sonuç şu şekilde olmuştur:

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

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

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

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

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

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

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

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

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

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 göstermiştir 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 aralıklarında değişerek devam etmiştir. Bu sonuca varılmasının en önemli sebebi, log-olasılık uzaklığı ölçümünün dikkate alınarak kümeleme oluşturulmasıdır. Çünkü elde edilen her bir kayıt,

kendisinden önceki kayıtlarla karşılaştırılarak, uzaklık ve yakınlık bağlarına göre en uygun sayıda elemana sahip, optimum sayıda küme oluşturulmuştur. Böylece farklı değerlerin birbirine farklı uzaklıklarda ölçülmesi sonucu, küme gruplarının daha fazla sinyal gücü veri grubuna sahip olması durumu gözlemlenmiştir.

3. BULGULAR VE TARTIŞMA (RESULTS AND DISCUSSION)

Elde edilen veriler iki aşamalı kümeleme yöntemiyle değerlendirilmiş 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. A Şebekesi, B ve C Şebekesine oranla daha istikrarlı ölçüm sonuçları vermiştir. C Şebekesi sinyal gücü ölçümünde performansı en düşük şebeke olmuştur.

Sinyal güçleri belirli aralıklara göre sınıflandırıldığında, elde edilen sonuçlarda belirtilenden farklı bir durum ortaya çıkmıştır. 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.

Çalışmanın bu bölümünde elde edilen verilerin temizlenmesi ve hazırlanması aşamasından sonra veri madenciliği modeli oluşturulmuş, uygulama sonucunda ortaya çıkan veriler izlenmiş ve sonuçlar yorumlanmıştır. Böylece veri madenciliği süreci tamamlanmıştır. Bu bilgiler ışığında A, B ve C Şebekelerinin, belirlenen güzergâhta çekim gücü performansları değerlendirilmiş ve tüm şebekeler için iyileştirme ve geliştirme bilgileri ortaya çıkmıştır.

4. SONUÇLARIN DEĞERLENDİRİLMESİ (CONCLUSIONS)

Çalışma 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.

Program tasarımı için MS Visual Studio C#.NET programlama dili kullanılmıştır. Sinyal gücü elde etmek için bir cep telefonundan veriler alınmıştır. Elde edilen bu veriler iki aşamalı kümeleme yöntemiyle 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. Sınıflandırılmış sinyal gücü oranları incelendiğinde ise B Şebekesinin iyi sinyal oranı, A Şebekesinden daha yüksek olmuştur. Bu değerlendirmede yine C Şebekesi en düşük orana sahip şebeke olmuştur. “İyi sinyal” sınıfı, mobil iletişim için en uygun ortam olduğu için, sınıflandırılmış sinyal gücü oranlarına göre yapılan değerlendirmenin daha doğru olduğu düşünülmüştür. Böylece bu üç şebeke içinde, B Şebekesi performansı en iyi şebeke, A Şebekesi orta seviyede ve C Şebekesi ise performansı en düşük şebeke olarak ortaya çıkmıştır.

Seçilen güzergâhta her üç şebeke için de iyileştirme çalışmaları gerekmektedir. A Şebekesi, yüksek sinyal gücüne paralel olarak orta sinyal gücü bölgelerinde sinyal iyileştirmeleri yapmalı ve iyi sinyal oranını yükseltmelidir. B Şebekesi, yetersiz sinyal bölgesi ve özellikle de zayıf sinyal bölgelerini göz önüne almalı ve buralarda sinyal gücü iyileştirme çalışmaları yapılmalıdır. C Şebekesi performansı ise diğer şebekelere oranla gözle görülür seviyede düşüktür. C Şebekesi işletmecileri, diğer şebekelerle rekabet edebilmek için iyi sinyal oranını yükseltmelidir. Geliştirilen program başka bölgelerde farklı GSM şebekeleri için de kullanılabilir ve sonuçlar değerlendirilebilir.

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. Bu çalışmada geliştirilen uygulama programı, gelecekte yapılacak çalışmalar ve tasarlanacak programlar için örnek teşkil etmektedir. Özellikle bu programın mobil versiyonlarının yazılması ve GSM şebekelerinin bu programları kullanıcılarına kendi uygulama marketlerinde sunması, işletmelere kullanıcı bazlı veri elde etme imkânı sunacaktır. Bu programları mobil iletişim araçlarına yükleyen kullanıcılardan elde edilen verilerin, işletmelerce analiz edilerek yorumlanması, gelecek için öngörüler oluşturma, iyileştirme ve geliştirmeler yapmak için işletmelere imkanlar sunacaktır.

5. KAYNAKLAR (REFERENCES)

1. Kalikov, A., **Data Mining and an e-Commerce Application**, M.Sc. Thesis, Gazi University, Institute of Science and Technology, 2006.
2. İnan, O., **Data Mining**, M.Sc. Thesis, Selçuk University, Institute of Science and Technology, 2003.
3. Thuarisingham, B.M., **Web Data Mining and Applications in Business Intelligence and Counter Terrorism**, CRC Press LLC, 35, Boca Raton FL, USA, 2003.
4. Savaş, S., **Analysis of GSM Networks' Performance With Data Mining**, M.Sc. Thesis, Gazi University, Institute of Informatics, 2011.
5. Kayaalp, K., **Fault Detection in Induction Motors Using Data Mining**, M.Sc. Thesis, Süleyman Demirel University, Graduate School of Applied and Natural Sciences, 2007.
6. Duru, N. ve Canbay, M., “Analysis of Earthquake Data With Data Mining”, **International Earthquake Symposium**, Kocaeli, 556-560, 22-26 Ekim 2007.
7. Tsao, W.K., Lee, A.J.T., Liu Y.H., Chang, T.W., Lin, H.H., “A Data Mining Approach to Face Detection”, **Pattern Recognition**, PR, Cilt 43, No 3, 1039-1049, 2010.
8. Holden, N., Frietas, A.A., “A Hybrid PSO/ACO Algorithm for Classification”, **Journal of Artificial Evolution and Applications**, 2745-2750, 2007.
9. Chen, Y., Sprague, A., “Exploration on the Commonality of Hierarchical Clustering Algorithms”, **ACM-SE 42 Proceedings of the 42nd Annual Southeast Regional Conference**, The University of Alabama in Huntsville, Huntsville, AL, 2004.
10. Çiftçi, S., **Analysing Student Learning Behaviour from Log Data in Distance Education**, M.Sc. Thesis, Gazi University, Faculty of Education, 2006.
11. Ayık, Y.Z., Özdemir, A. ve Yavuz, U., “Analysis relationship of Type of School and High School Graduation with Acquired Faculty with Data Mining Techniques”, **Journal of Graduate School of Social Sciences**, Cilt 10, No 2, 441-454, 2007.
12. Bozkır, A.S., Sezer, E. ve Gök, B., “Determination Of The Factors Influencing Student's Success In Student Selection Examination (OSS) Via Data Mining Techniques”, **5. International Advanced Technologies Sempodium (IATS'09)**, Karabük University, Karabük, 37-43, 13-15 Mayıs 2009.
13. Akbulut, S., **Churn Analysis and Customer Segmentation of a Cosmetics Brand Using Data Mining Techniques**, M.Sc. Thesis, Gazi University, Institute of Science and Technology, 2006.

14. Özçakır, F.C. ve Çamurcu, A. Y., “A Data Mining Software Design and Application for Association Rule”, **Journal of Science and Technology of Istanbul Commerce University**, Cilt 6, No 12, 21-37, 2007.
15. Gürbüz, F., Özbakır, L. ve Yapıcı, H., “Data Mining Application on Component Reports of an Airline Company in Turkey”, **Journal of the Faculty of Engineering and Architecture of Gazi University**, Cilt 24, No 1, 73-78, 2009.
16. Hoontrakul, P., Sahadev, S., “Application of data mining techniques in the on-line travel industry: A case study from Thailand”, **Marketing Intelligence & Planning**, Cilt 26, No 1, 60-76, 2008.
17. Liou, J.J.H., “A Novel Decision Rules Approach For Customer Relationship Management ff The Airline Market”, **Expert Systems with Applications**, Cilt 36, No 3, 4374-4381, 2009.
18. Yavaş, G., **Using a Data Mining Approach for The Prediction of User Movements in Mobile Environments**, M.Sc. Thesis, Bilkent University, Institute of Engineering and Science, 2003.
19. Albayrak, A.S. ve Yılmaz, Ş.K., “Data Mining: Decision Tree Algorithms and an Application on ISE Data”, **Suleyman Demirel University The Journal of Faculty of Economics and Administrative Sciences**, Cilt 14, No 1, 31-52, 2009.
20. Ata, A.H. ve Seyrek, İ.H., “The Use of Data Mining Techniques in Detecting Fraudulent Financial Statements: an Application on Manufacturing Firms”, **Suleyman Demirel University The Journal of Faculty of Economics and Administrative Sciences**, Cilt 14, No 2, 157-170, 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”, **Suleyman Demirel University The Journal of Faculty of Economics and Administrative Sciences**, Cilt 14, No 2, 113-139, 2009.
22. Türk Dil Kurumu, Büyük Türkçe Sözlük “Mobilite”, <http://tdkterim.gov.tr/bts/>, Mart 2011.
23. Sectoral Research and Strategies Department, “Three-month Market Data Report”, **Information and Communication Authority**, 2010 Yılı 4. Çeyrek, S 12, 36-59, 2011.
24. GSM / GPRS modems, Telemetry and data accessories “AT+CSQ in dbm”, <http://www.gprsmodems.co.uk/images/csq1.pdf>, Mart 2011.
25. Taşkın, Ç. ve Emel, G.G., “Clustering Approaches in Data Mining and an Application with Kohonen Networks in Retailing Sector”, **Suleyman Demirel University The Journal of Faculty of Economics and Administrative Sciences**, Cilt 15, No 3, 395-409, 2010.

