

TC.
İSTANBUL AYDIN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
BİLGİSAYAR MÜHENDİSLİĞİ BİLİM DALI



WEB TABANLI ANKET SİSTEMİ İLE ELDE EDİLEN VERİLERİN VERİ
MADENCİLİĞİ YÖNTEMİ İLE ANALİZİ

YÜKSEK LİSANS TEZİ

Hazırlayan
Zeynep ELABİAD

Tez Danışmanı
Yrd. Doç. Dr. Metin ZONTUL

İSTANBUL - 2013

TC.
İSTANBUL AYDIN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
BİLGİSAYAR MÜHENDİSLİĞİ BİLİM DALI



WEB TABANLI ANKET SİSTEMİ İLE ELDE EDİLEN VERİLERİN VERİ
MADENCİLİĞİ YÖNTEMİ İLE ANALİZİ

YÜKSEK LİSANS TEZİ

Hazırlayan

Zeynep ELABİAD

Tez Danışmanı

Yrd. Doç. Dr. Metin ZONTUL

İSTANBUL - 2013



T.C.
İSTANBUL AYDIN ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ
MÜDÜRLÜĞÜ'NE

Tez Onay Belgesi

Enstitümüz Bilgisayar Mühendisliği Anabilim Dalı, Bilgisayar Mühendisliği (Tezli) Yüksek Lisans Programı Y1013.010009 numaralı öğrencisi **Zeynep ELABİAD'** ın "**WEB TABANLI ANKET SİSTEMİ İLE ELDE EDİLEN VERİLERİN VERİ MADENCİLİĞİ YÖNTEMİ İLE ANALİZİ**" adlı tez çalışması Enstitümüz Yönetim Kurulunun 03.09.2013 tarih ve 2013/18 sayılı kararıyla oluşturulan jüri tarafından **oybirliği /oyçokluğu** ile Yüksek Lisans Tezi olarak **kabul** edilmiştir.

Öğretim Üyesi Adı Soyadı

İmzası

Tez Savunma Tarihi : 17.09.2013

1) Tez Danışmanı : Yrd. Doç. Dr. Metin ZONTUL

2) Jüri Üyesi : Prof. Dr. Ali GÜNEŞ

3) Jüri Üyesi : Yrd. Doç. Dr. Duygu ÇELİK

Metin Zontul
Ali Güneş
Duygu Çelik

Not: Öğrencinin Tez savunmasında **Başarılı** olması halinde bu form **imzalanacaktır**. Aksi halde geçersizdir.

ÖN SÖZ

Veri Madenciliği ile verilerin yeniden şekillenip anlamlı hale geldiği günümüzde, veri madenciliği ve algoritmaları, bilgi keşfi için birçok alanda önemli hale gelmiştir. Tez çalışmamda hazır anket verileri yerine, online anket yöntemiyle ilk elden edinilen veriler, apriori ve karar ağacı algoritmaları yardımıyla işlenip bilgi keşif sürecinden geçmiştir.

Bu çalışmam boyunca yardım ve desteklerinden dolayı değerli hocam Bilgisayar Müh. Bölüm Başkanımız Prof. Dr. Ali GÜNEŞ' e içten teşekkürlerimi sunarım.

Tez sürecimde gösterdiği sabır ve katkılarıyla bilgilerini esirgemeyen danışman hocam Yrd. Doç. Dr. Metin ZONTUL' a teşekkürü bir borç bilirim.

Tez çalışmamda bilgileriyle katkıda bulunan Dr. Vildan GÜLPINAR' a teşekkür ederim.

Ayrıca tez aşamasında evlendiğim sevgili eşim Maged ELABİAD' a gösterdiği sabır ve manevi destek için teşekkür ederim.

Zeynep ELABİAD

İÇİNDEKİLER

ÖN SÖZ	i
İÇİNDEKİLER	ii
SİMGELER / KISALTMALAR DİZİNİ	iv
TABLolar DİZİNİ	v
ŞEKİLLER DİZİNİ	vii
EKLER DİZİNİ	x
1. GİRİŞ	1
2. VERİ TABANLARINDA BİLGİ KEŞİFİ	7
2.1 VERİ TABANLARINDA BİLGİ KEŞİF ADIMLARI	8
2.1.1 Verilerin Elde Edilmesi	8
2.1.2 Veri Seçme ve Dönüştürme	9
2.1.3 Veri Madenciliği ve Örüntü Değerlendirme	9
2.1.4 Bilgi Sunumu	9
3. VERİ MADENCİLİĞİ	10
3.1 VERİ MADENCİLİĞİ VE DİĞER DİSİPLİNLER	11
3.2 MÜŞTERİ İLİŞKİLERİ YÖNETİMİ VE VERİ MADENCİLİĞİ	12
3.3 VERİ MADENCİLİĞİ TEKNİKLERİ	13
3.3.1 Karar Ağaçları (Decision Trees)	15
3.3.2 Apriori Algoritması	18
4. VERİ MADENCİLİĞİ ANALİZ PROGRAMI: WEKA	29
4.1 WEKA NEDİR?	29
4.2 SIMPLE CLI MODÜLÜ	31
4.3 KNOWLEDGEFLOW MODÜLÜ	32
4.4 EXPERIMENTER MODÜLÜ	33
4.4 EXPLORER MODÜLÜ	35
4.5 CLASSIFY	36
4.6 CLUSTER	38
4.7 ASSOCIATE	39
4.8 SELECT ATTRIBUTES	40

4.9 VISUALIZE	41
5. ANKET	42
5.1 ANKET TÜRLERİ	44
5.2 WEB TABANLI ANKET.....	44
5.3 ONLINE ANKET HAZIRLANIŞI	46
5.4 ONLINE ANKET: GSM MÜŞTERİ KAYBI ANALİZİ	51
5.4.1 Online Anket Soruları	52
5.5 ANKET İSTATİSTİKLERİ.....	58
6. WEKA UYGULAMASI	76
6.1 KARAR AĞACI ALGORİTMASININ UYGULANMASI.....	80
6.2 APRIORI ALGORİTMASININ UYGULANMASI	88
6.2.1 Apriori Kuralları.....	92
7. SONUÇ VE ÖNERİLER	97
KAYNAKÇA	99
İNTERNET KAYNAKÇASI	104
EK.....	105
ÖZET	109
ABSTRACT.....	110
ÖZGEÇMİŞ.....	111

SİMGELER / KISALTMALAR DİZİNİ

VM	: Veri Madenciliği
VTBK	: Veri Tabanlarında Bilgi Keşfi
RAM	: Rehberlik Araştırma Merkezi
GSM	: Global System for Mobile Communications
WEKA	: Waikato Environment for Knowledge Analysis
CART	: Classification and Regression Tree
SOM	: Self Organizing Maps
ARFF	: Attribute Relation File Format
CONV	: Conviction
LEV	: Leverage
SUPP(X)	: X'in Destek Oranı
CONF(X)	: X'in Güven Oranı

TABLolar DİZİNİ

Tablo 1. Farklı Müşterinin Aldığı Market Ürünleri Tablosu.....	22
Tablo 2. Destek Değer Tablosu	22
Tablo 3. Eşik Destek Değerine Eşit Ya Da Daha Büyük Desteğe Sahip Ürünler Tablosu.....	23
Tablo 4. İkili Ürün Gruplarının Destek Değerler Tablosu.....	23
Tablo 5. Eşik Destek Sayısı Olan Üç Değerine Eşit Yada Daha Büyük Desteğe Sahip İkili Ürün Gruplar Tablosu.....	24
Tablo 6. Üçlü Ürün Gruplarının Destek Değerleri Tablosu.....	24
Tablo 7. Eşik Destek Sayısına Eşit Ya Da Daha Büyük Destek Sayısına Sahip Üçlü Ürün Grubu Tablosu.....	25
Tablo 8. Birliktelik Kuralları Tablosu.....	27
Tablo 9. Anket Katılımcılarının Cinsiyet Oran Tablosu.....	58
Tablo 10. Anket Katılımcılarının Yaş Aralık Oran Tablosu.....	59
Tablo 11. Anket Katılımcılarının Medeniyet Durum Oran Tablosu.....	60
Tablo 12. Anket Katılımcılarının Mezuniyet Oran Tablosu.....	61
Tablo 13. Anket Katılımcılarının Aylık Gelir Oran Tablosu.....	62
Tablo 14. Anket Katılımcılarının GSM Operatörüne Sahip Olma Süresi Oran Tablosu.....	63
Tablo 15. Anket Katılımcılarının GSM Abonelik Türleri Oran Tablosu.....	64
Tablo 16. Anket Katılımcılarının Numara Taşıma/Taşımama Oran Tablosu.....	65
Tablo 17. Anket Katılımcılarının GSM Operatör Türü Oran Tablosu.....	66

Tablo 18. Turkcell GSM Operatörü Tarife Türü Oran Tablosu.....	67
Tablo 19. Vodafone GSM Operatörü Tarife Türü Oran Tablosu.....	68
Tablo 20. Avea GSM Operatörü Tarife Türü Oran Tablosu.....	69
Tablo 21. Anket Katılımcılarının Cep Telefonu Özelliklerine Sahip Olma Oranı Tablosu.....	70
Tablo 22. Anket Katılımcılarının Operatör Türüne Göre Memnuniyet Durum Tablosu	71
Tablo 23. Anket Katılımcılarının Operatör Türünü Değişirme İstek Tablosu.....	72
Tablo 24. Anket Katılımcılarının GSM Hattı Yorum Oranı Tablosu.....	75
Tablo 25. Nitelik İsimleri ve Açıklamaları.....	78

ŞEKİLLER DİZİNİ

Şekil 1. Veri Tabanlarında Bilgi Keşif Süreci	8
Şekil 2. Veri Madenciliği ve Diğer Disiplinler	11
Şekil 3. Örnek karar ağacı.....	17
Şekil 4. Güven oranı düşük kuralların budanması.....	19
Şekil 5. ARFF dosya yapısı.....	30
Şekil 6. WEKA programı ara yüzü.....	30
Şekil 7. WEKA SimpleCLI paneli.....	32
Şekil 8. WEKA Knowledge Flow Modülü.....	33
Şekil 9. WEKA Experimenter Modülü.....	34
Şekil 10. WEKA Experimenter Viewer Paneli.....	35
Şekil 11. WEKA Explorer Modülü Preprocess Sekmesi.....	36
Şekil 12. WEKA Explorer Modülü Classify Sekmesi.....	37
Şekil 13. WEKA Explorer Modülü Cluster Sekmesi.....	38
Şekil 14. WEKA Explorer Modülü Associate Sekmesi.....	39
Şekil 15. WEKA Explorer Modülü Select Attributes Sekmesi.....	40
Şekil 16. WEKA Explorer Modülü Visualize Sekmesi.....	41
Şekil 17. Anket Hazırlama Panel Örneği a.....	46
Şekil 18. Anket Hazırlama Panel Örneği b.....	47
Şekil 19. Anket Hazırlama Panel Örneği c.....	48
Şekil 20. Anket Hazırlama Panel Örneği d.....	49

Şekil 21. Anket Hazırlama Panel Örneği e.....	50
Şekil 22. Anket Katılımcılarının Cinsiyet Oranı.....	58
Şekil 23. Anket Katılımcılarının Yaş Aralık Oranı.....	59
Şekil 24. Anket Katılımcılarının Medeniyet Durum Oranı.....	60
Şekil 25. Anket Katılımcılarının Mezuniyet Oranı.....	61
Şekil 26. Anket Katılımcılarının Aylık Gelir Oranı.....	62
Şekil 27. Anket Katılımcılarının GSM Operatörüne Sahip Olma Süresi Oranı.....	63
Şekil 28. Anket Katılımcılarının GSM Abonelik Türleri Oranı.....	64
Şekil 29. Anket Katılımcılarının Numara Taşıma/Taşımama Oranı.....	65
Şekil 30. Anket Katılımcılarının GSM Operatör Türü Oranı.....	66
Şekil 31. Turkcell GSM Operatörü Tarife Türü Oranı.....	67
Şekil 32. Vodafone GSM Operatörü Tarife Türü Oranı.....	68
Şekil 33. Avea GSM Operatörü Tarife Türü Oranı.....	69
Şekil 34. Anket Katılımcılarının Cep Telefonu Özelliklerine Sahip Olma Oranı.....	70
Şekil 35. Anket Katılımcılarının Operatör Türüne Göre Memnuniyet Durumu.....	71
Şekil 36. Anket Katılımcılarının Operatör Türünü Değişirme İsteği.....	72
Şekil 37. Anket Katılımcılarının Operatörlerinden Memnun Olma Derecesi.....	73
Şekil 38. Anket Katılımcılarının Operatörlerinden Memnun Olmama Derecesi.....	74

Şekil 39. Anket Katılımcılarının GSM Hattı Yorum Oranı	75
Şekil 40. Öznitelik Değerlerine Göre Memnuniyet Durumu.....	77
Şekil 41. Verilerin J48 Algoritması İle Test Edilmesi.....	81
Şekil 42. J48 Doğruluk Değerleri.....	82
Şekil 43. J48 Algoritması Karar Ağacı.....	84
Şekil 44. J48 Algoritması Görsel Karar Ağacı.....	85
Şekil 45. Apriori Algoritması Destek Ve Güven Parametrelerinin Belirlenmesi	88
Şekil 46. Apriori Destek Ve Güven Parametre Oranı.....	91

EKLER DİZİNİ

EK-1. Bir Anket Katılımcısının Anket Örneđi

1. GİRİŞ

Zaman ilerledikçe teknolojik açıdan birçok alanda ilerlemeler sağlanmıştır. Teknolojinin gelişmesi beraberinde gereksiz verilerin çoğalmasına neden olmakla birlikte, bu verilerin tutulduğu veri tabanlarını birer veri çöplüğü haline getirmektedir. Bu gereksiz veriler veri tabanlarında ciddi boyutlarda yer tutup, istenilen verilere erişmekte zorluk çekilmesine neden olmaktadır. Bir varsayıma göre veriler her 20 ayda bir, bir önceki var olan veri miktarının iki katına çıkmaktadır (Witten ve Frank, 2005).

1995 yılında birincisi düzenlenen Knowledge Discovery in Databases konferansı bildiri kitabı sunuşunda, enformasyon teknolojilerinin oluşturduğu veri dağları aşağıdaki cümleler ile vurgulanmaktadır.

“Dünyadaki enformasyon miktarının her 20 ayda bir ikiye katlandığı tahmin edilmektedir. Bu ham veri seli ile ne yapmamız gerekmektedir. İnsan gözleri bunun ancak çok küçük bir kısmını görebilecektir. Bilgisayarlar bilgelik pınarı olmayı vaat etmekte, ancak veri sellerine neden olmaktadır. “

Veri tabanlarındaki artış, insanları çözüm bulmaya ve işe yarar bilgiyi ortaya çıkarmaya yöneltmiştir. Bu arayış Veri Tabanlarında Bilgi Keşif Süreci (Knowledge Discovery in Databases) olarak adlandırılmaktadır. Bu süreç içerisinde bilgiyi edinmek için veri madenciliğinden yararlanılmaktadır [1] [2].

Literatürde veri madenciliği alanında son yıllarda giderek artan çalışmalara doğru orantılı olarak, bu çalışmaların veri madenciliğinin daha çeşitli alanlarında ortaya konulduğu görülmüştür.

Şengül Doğan 2007 yılında veri madenciliği kullanarak biyokimya verilerinde hastalık teşhisini araştıran yüksek lisans çalışması hazırlamıştır. Bu çalışmasında veri madenciliği algoritmalarından olan karar ağaçlarını kullanarak hastalıkların doktorların yaptığı teşhisle örtüştüğünü analiz etmiştir (Doğan, 2007).

Kerem Üçgün 2009 yılında hazırladığı yüksek lisans çalışmasında ortaöğretim öğrencileri için hazırladığı otomasyon ile öğrenci verileri üzerinde veri madenciliği uygulamıştır. Bilgi edinme sürecinde, öğrenci datalarına apriori algoritmasını uygulayarak başarı analizi yapmıştır (Üçgün, 2009).

F.Sibel Bırtıl 2011 yılında hazırladığı yüksek lisans çalışmasında, kız meslek lisesi öğrencilerinin akademik başarısızlık nedenlerini veri madenciliği tekniği ile analiz etmiştir. RAM (Rehberlik Araştırma Merkezi) yoluyla daha önceden öğrencilere uygulanmış 37 soruluk anket verilerini Clementine programında kümeleme algoritmalarıyla analiz ederek başarısızlık nedenlerini kümelemiştir (Bırtıl, 2011).

Muhsin Özgür Dolgun, 2006 yılında çalıştığı yüksek lisans tezinde büyük alışveriş merkezleri için veri madenciliği uygulamaları üzerine çalışmıştır. Alışveriş merkezinden edinilen fişler yordamıyla topladığı bu verileri, Sas Enterprise Miner programıyla analiz ederek veri üzerinde birliktelik kuralını uygulamıştır (Dolgun, 2011).

Öner ÇELİK, 2009 yılında yüksek lisans tezinde, belirli zaman periyotlarında ATM'lerin para ihtiyaçların belirlenmesinde veri madenciliği algoritmaları üzerine çalışmıştır. ATM'lerin kullanımına yönelik veriler üzerinde birden fazla karar ağaç algoritmalarını uygulayarak, hangi algoritmanın daha verimli olduğunu belirtmiştir (Çelik, 2009).

Buket DOĞAN 2006 yılında çalıştığı doktora tezinde, zeki öğretim sistemlerinde veri madenciliği kullanılması üzerinde çalışmıştır. Veri toplama ve veri madenciliği için oluşturulan yazılımlarda toplanan veriler birliktelik kuralı ve kümeleme algoritmalarıyla incelenmiştir (Doğan, 2006).

Hüseyin ÖZÇINAR tarafından 2006 yılında yazılan yüksek lisans tezinde kpss sonuçları veri madenciliği yöntemiyle analiz edilmiştir. Birden fazla veri madenciliği programı ile analiz edilen veriler, yapay sinir ağları kullanılarak mezun olunan üniversitenin kpss sonuçlarına olan katkısı üzerinde çalışılmıştır (Özçınar, 2006).

Veri Madenciliği ile kredi kartlarında müşteri kaybetme analizi konulu tez çalışmasında, Yapı Kredi Bankası'ndan alınan veriler üzerinde kaybedilmiş müşteri profili üzerinden giderek müşteri kayıp sebepleri ve müşterinin kaybedilme olasılığını belirlemek için karar ağacı algoritması uygulanmıştır (Tosun, 2006).

Emre Güngör, Nesibe Yalçın ve Nilüfer Yurtay'ın 2013 yılında yayınladığı çalışmada apriori algoritması ile teknik seçmeli ders seçim analizi yapılmıştır. Üniversite öğrencileri üzerinde yapılan anket yöntemiyle toplanan veriler, birliktelik kuralı algoritmalarından biri olan apriori ile teknik seçmeli dersi seçmelerindeki nedenleri ve hangi kriterleri göz önünde bulundurdıkları belirlenmiştir (Güngör, Yalçın ve Yurtay, 2013).

Adil Baykasoğlu, 2005 yılında veri madenciliği ve çimento sektöründe uygulama konulu makalesinde çimento basma mukavemeti verileri üzerinde gen denklem programlama, yapay sinir ağları ve regresyon analizi ile tahmini ve hangi yöntemin daha iyi olduğu üzerinde çalışmıştır. Gen denklem programlamanın daha iyi olduğunu çalışmasında belirtmiştir (Baykasoğlu, 2005).

Mustafa Danacı, Mete Çelik ve A. Erhan Akkaya, Irvine California Üniversitesi'nden alınan hasta verileri üzerinde C4.5 karar ağacı algoritmasını uygulayarak meme kanseri hücrelerinin tahmin ve teşhisi üzerinde çalışma gerçekleştirmiştir. Bu algoritmanın meme kanseri gibi bir hastalığın tanı ve teşhisinde % 97.43 lük doğruluk oranı ile çalışması bu hastalığın erken teşhis ve tanısında büyük rol oynayacağı belirtilmiştir (Danacı, Çelik ve Akkaya, 2010).

Müşteri memnuniyetine etki eden faktörleri veri madenciliği algoritmaları ile değerlendirme konulu yayında, 301 kişiden alınan memnuniyet veya memnuniyetsizlik tabanlı bilgileri karar ağacı algoritması olan C5.0 algoritması ve kümeleme analizi algoritması olan K-means ile işlenmiştir (Çınar ve Silahtaroğlu, 2012).

Serkan Savaş ve Nurettin Topaloğlu, veri madenciliği yöntemi ile GSM şebekelerinin performans analizi üzerine çalışma yapmıştır. Geliştirdikleri yazılım ile GSM şebekelerinin sinyal güçlerini bilgilerini toplayarak veri tabanları oluşturulmuştur. Kümeleme analizi kullanılarak GSM firma isimleri verilmeden çıkan sonuçlara göre şebekelerin sinyal güçlerinin yetersiz olduğu belirtilmiştir (Savaş ve Topaloğlu, 2011).

VM teknikleriyle bir kozmetik markanın ayrılan müşteri analizi ve müşteri bölümlenmesini konu alan bir makalede, WEKA programı kullanılarak Naive Bayes, Lojistik Regresyon, ID3, J.48, JRIP, PART, Yapay Sinir Ağları algoritmaları müşteri verilerine uygulanmıştır. Bu çalışma için ayrılan müşteri analizine en uygun algoritmanın J.48 olduğu saptanmıştır (Aydoğan, Gencer ve Akbulut, 2008).

Yrd. Doç. Dr. Mehmet Ali ALAN, 2012 yılındaki yayınında lisansüstü öğrenci verileri üzerinde veri madenciliği uygulaması üzerinde çalışmıştır. WEKA programı üzerinde veri madenciliği algoritmalarını kullanarak başarımlarını derecelerini ölçmüştür. Çalışmada verileri en iyi sınıflandıran algoritmanın SimpleCART olduğunu belirtmiştir (Alan, 2012) .

WEKA programıyla veri madenciliği algoritmalarını uygulayan başka bir tez çalışmasında, asenkron motorlarda veri madenciliği ile hata tespiti yapılmıştır. Veri paketleri WEKA programında bulunan karar ağacı algoritmaları ile test edilmiş olup, en iyi sonucu RepTree ve M5P-M4.0 karar ağaçlarının verdiğini belirtmiştir (Kayaalp, 2007).

Başka bir yayında ise WEKA yazılımı yardımıyla k-ortalama algoritması kullanarak konjestif kalp yetmezliği hastaların teşhisi üzerinde çalışılmıştır. K-ortalama kümeleyicisinin bu hastalık teşhisinde oldukça başarılı olduğu belirtilmiştir (İşler ve Narin, 2012).

Türkiye’de telekomünikasyon alanında olan büyük bir firmanın ayrılma ihtimali olan müşterilerinin profili belirlenerek pazarlama stratejileri geliştirmek ve kaybı önlemek amacıyla veri madenciliği algoritmalarından yararlanılmıştır. Karar ağaçları ve Lojistik Regresyon Analizi kullanılmıştır (Gürsoy, 2010).

Pekko Vehviläinen, Kimmo Hätönen ve Pekka Kumpulainen’in, 2003 yılında, dijital mobil telekomünikasyon ağ kalitesi analizinde veri madenciliği başlıklı yayınlanan makalesinde, sınıflama içerisinde kullanılan karar ağaçlarının temelini oluşturan CART algoritması ve kümeleme algoritması olan SOM kullanılmıştır. VM algoritmalarının bu çalışma içerisinde sorunsuz bir şekilde analiz edildiği belirtilmiştir (Vehviläinen, Hätönen, Kumpulainen, 2003).

Murat Dener, Murat Dörterler ve Abdullah Orman, lisansüstü öğrenci başarısının sebepleri için çalışmalarında veri madenciliğini kullanmıştır. Öncelikle açık kaynak kodlu veri madenciliği programlarını kıyaslayarak, uygulamada kullanacağı programı WEKA olarak belirlemiştir. Daha sonra WEKA içerisinde bulunan algoritmalarından Naive Bayes, Kstar, RBFNetwork, J.48, JRIP, Ridor algoritmaları uygulanmıştır. Bu çalışmada doğruluğu en yüksek veren Naive Bayes algoritması olduğu belirtilmiştir (Dener, Dörterler ve Orman, 2009).

Tez çalışmamın amacı, veri madenciliği analizi için hazır olarak alınan veriler yerine, web tabanlı anket yardımıyla verilerin kişinin kendisinden elde edilen verilerin bilgi keşif sürecinden geçerek WEKA programı içinde var olan apriori, karar ağaçları algoritmaları sayesinde işlenip analiz edilmesidir.

İlk bölümünde, veri tabanlarında bilgi keşfi, veri madenciliği, müşteri ilişkileri ve veri madenciliği ilişkisi, veri madenciliği teknikleri ve tez içinde kullanılan karar ağaçları ve apriori algoritması yer almaktadır.

Tezin ikinci bölümünde veri madenciliğinde yaygın kullanıma sahip WEKA programının işlevi ve kullanımı anlatılmıştır.

Üçüncü bölümde, anket, anket türleri, online anket, online anketin avantajları ve dezavantajları, online anket sorularının hazırlanışı yer almaktadır.

Dördüncü bölümde, tez çalışmamda kullandığım 21 sorudan oluşan web tabanlı anket soruları ve 615 kişinin katıldığı bu anketin istatistikleri ve grafikleri yer almaktadır.

Beşinci bölümde, veri madenciliği için elde edilen verilerin WEKA programına uygulanması, karar ağaçları ve apriori algoritmalarıyla elde edilen kuralların çıkarımı anlatılmaktadır.

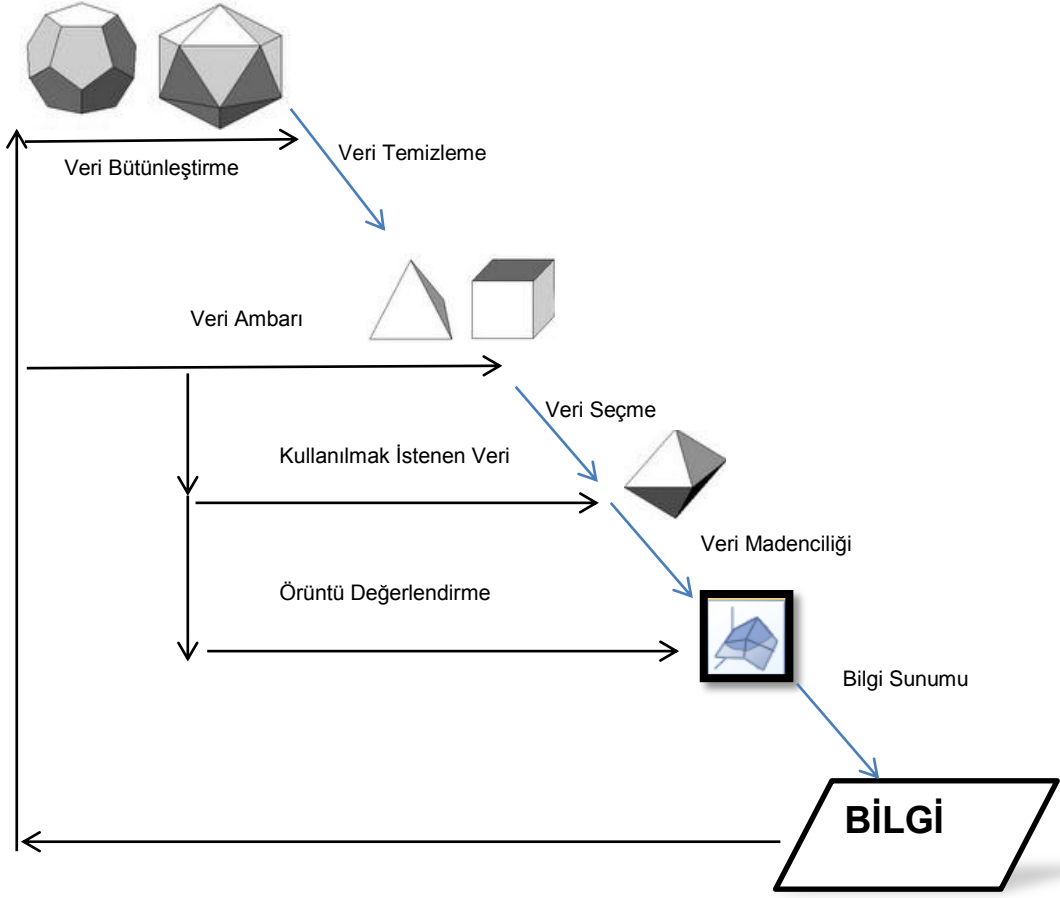
Son bölümde, veri madenciliğinin önemi, elde edilen sonuçlar ve kurallar vurgulanmıştır. İleride yapılabilecek çalışmalar ve uygulama alanları önerilmiştir.

2. VERİ TABANLARINDA BİLGİ KEŞİFİ

Veriler istenilen amaca göre işlenmeden bir değer taşımaz. Bu yüzden veri tabanlarındaki büyük verileri analiz edilerek geleceğe yönelik çıkarım yapmamız gerekmektedir. Veri madenciliği yöntem ve teknikleriyle çıkarılan, aynı zamanda amaca hitap eden anlamlı veriler bize bilgiyi ulaştırır.

Bilgiyi edinme keşfinde aşağıdaki aşamalardan yararlanırız (Han ve Kamber, 2001) :

- Veri Temizleme
- Veri Bütünleştirme
- Veri Dönüşümü
- Veri Seçme
- Veri Madenciliği
- Örüntü Değerlendirme
- Bilgi Sunumu



Şekil 1. Veri Tabanlarında Bilgi Keşif Süreci

2.1 VERİ TABANLARINDA BİLGİ KEŞİF ADIMLARI

2.1.1 Verilerin Elde Edilmesi

Bu aşamada öncelikle gereksiz veriler elemeyi amaçlanmaktadır. Daha sonra eğer varsa başka bir yerden gelebilecek gereksiz verileri önceden süzölmüş olan verilerle birleştirme işlemi yapılarak tek bir data haline getirilmektedir. Ayrıca tutarsız veriler ve gürültü içeren veriler temizlenerek bir sonraki adım için hazır hale getirilmektedir.

2.1.2 Veri Seçme ve Dönüştürme

Gereksiz verilerden temizlenen veri içerisinde işe yarayacak verileri seçmek daha kolaylaşmaktadır. Bu aşamada amaçlanan hedef için veriler dönüştürülmek ve daha sonra analiz edilmek için seçilmektedir. Veriler üzerinde öncelikle kurulacak modele göre seçim yapılmaktadır. Daha sonra amaca yönelik belirlenen veri madenciliği algoritması ile dönüştürme işlemi yapılmaktadır. Bu dönüşüm değişkenlerin standartlaştırılması veya normalleştirilmesi için gereklidir.

2.1.3 Veri Madenciliği ve Örüntü Değerlendirme

Bu aşama, ortak özellik barındıran ve aralarında bir ilişki kurulabilen örnekleri veya özellikleri tanımlama ve sınıflama aşamasıdır. Bu aşamada örüntülerin belirlenmesi için veri madenciliği metotları kullanılmaktadır. Yani anlam ifade eden örüntülerin çıkarılması veri madenciliği aşamasında olmaktadır. Bu örüntüler bulunduktan sonra değerlendirilmektedir.

2.1.4 Bilgi Sunumu

Ulaşılan sonuçlar kullanıcılara sunulmaktadır. Sunum aşamasında istenen bilgiler grafiklerle desteklenebilir.

3. VERİ MADENCİLİĞİ

Günümüzde birçok farklı alan sektörleri gerek artan rekabet, gerekse müşteri ilişkileri yönetimini sağlayabilmek için verilerini düzenli olarak tutmaya başlamıştır. Bu veriler teknolojiyle doğru orantılı gelişen veri aygıtlarına uzun süre saklanabilmesi için kaydedilmektedir. Zamanla verilerin çok hızlı artışı insanlara bu verilerden öz ve gelecek açısından yararlı bilgiler için çıkarım yapmalarına yöneltmiştir. Veri tabanlarında biriken veri hedeflenen bilgiye erişebilmek için bilgi keşif sürecine girmektedir. Bir anlam ifade etmeyen verilerden anlamlı bilgiler çıkarabilmek ise veri madenciliğinin işidir. Veri Madenciliği, verilerin işlenmesi için kullanılan teknikler ve analiz edilen örüntüler bütünüdür.

Veri madenciliği, çok büyük veri tabanlarındaki veya veri ambarlarındaki veriler arasında bulunan ilişkiler, örüntüler, değişiklikler, sapma ve eğilimler, belirli yapılar gibi ilginç bilgilerin keşif sürecidir. Veri madenciliğinde kullanılan yöntem ve araçlar, çok kısa zamanlarda işin niteliğine yönelik stratejik soruları cevaplamada yardımcı olurlar. Ham veride gizli kalmış olan örüntüleri ve ilişkileri tahmini bilgilere dönüştürebilirler (Yan vd., 2001)

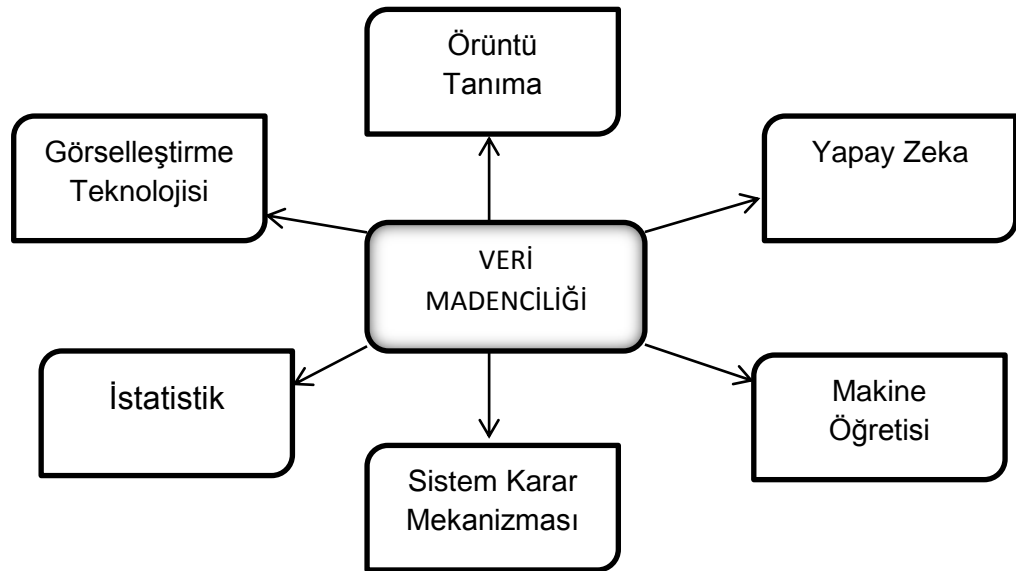
Holsheimer veri madenciliğini, büyük veri kümesi içinde saklı olan genel örüntülerin bulunması olarak tanımlamıştır (Holsheimer ve Siebes,1994).

Jacobs'a göre (1999), işlenmemiş verinin tek başına sunamadığı bilgiyi ortaya çıkaran veri analizi sürecini veri madenciliği olarak tanımlamıştır.

Bransten (1999) ise, veri madenciliğinin insanın asla bulmayı hayal bile edemeyeceği trendlerin keşfedilmesini aracı olduğunu belirtmiştir (Bransten, 1999).

3.1 VERİ MADENCİLİĞİ VE DİĞER DİSİPLİNLER

VM uygulamalarında diğer disiplinlerden yararlanmaktadır. Veri Madenciliği, araştırma ve çözümlere için birden fazla disiplin arasında köprü görevi yapmaktadır. Makine öğrenimi, istatistik, veritabanı teknolojisi, uzman sistemler ve verilerin görüntülenmesi (data visualization) gibi yöntemlerin birlikte kullanıldığı bir yöntemdir (Maindonald, J.).



Şekil 2. Veri Madenciliği ve Diğer Disiplinler (Hinneburg ve Keim,1999)

Veri madenciliği astronomi, biyoloji, finans, pazarlama, sigorta, tıp ve bir çok başka dalda uygulanmaktadır. Son 20 yıldır Amerika Birleşik Devletleri'nde çeşitli veri madenciliği algoritmalarının gizli dinlemeden, vergi kaçakçılıklarının ortaya çıkartılmasına kadar çeşitli uygulamalarda kullanıldığı bilinmektedir (Dilly, 1995).

Günümüzde veri madenciliğinin kullanıldığı sektörler şunlardır:

- İlaç ve Sağlık Sektörü
- Bankacılık, Kredi Analizi
- İletişim ve Telekomünikasyon
- Web Siteleri
- Satış ve Pazarlama
- Finans Sektörü
- Devlet Uygulamaları
- Turizm Alanı
- Kütüphanecilik
- Spor alanı
- Eğitim Alanı

3.2 MÜŞTERİ İLİŞKİLERİ YÖNETİMİ VE VERİ MADENCİLİĞİ

Günümüzde firmaların müşteri ile ilişkileri çok önemli bir yere sahiptir. Müşteri memnuniyetinin ve müşteri sadakatının sağlanabilmesi için artan rekabet piyasasında müşteriye elde tutmak, firma menfaati açısından gerekebilmektedir.

Churchill ve Suprenant (1982), müşteri memnuniyetini, alıcının satın aldığı üründen beklediği performansa karşılık elde ettiği sonuç ile katlandığı maliyeti değerlendirmesinin bir sonucu olduğunu belirtmiştir.

Firmalar müşteri memnuniyetine sadece yeni müşterileri kazanmak için değil, var olan müşterilerinin kaybını engellemek için de müşteri sadakatini kazanmaları açısından gerekli önemi göstermelidir. Firma sayısının artması, firma farkındalıklarının ortaya çıkarılması, daha kaliteli ve daha ekonomik başka bir opsiyonun ortaya çıkması müşteri kaybedilmesini kolaylaştırabilen unsurlardır. Eğer bir müşteri, ilgili firmayla üyelik

anlaşmasını sonlandırır ve başka bir rakip firmanın müşterisi haline gelirse bu müşteri kaybedilmiş müşteridir (Richeldi ve Perrucci, 2002).

Firmalar, hangi müşteri türlerinin kaybedilme riski olduğunu daha önceki verileri analiz ederek çıkarım yapabilmektedir. Firmalar, kaybedebilme riski olan müşterileri veri madenciliği yoluyla tanımlayıp, geri kazanma yollarını bu çıkarımlara göre belirleyebilmektedir. Müşteri kaybı, kaybın çok kolay gerçekleştiği firmalar için önemli bir problemdir. Örnek olarak bankalar, sigorta şirketleri ve telekomünikasyon firmaları verilebilir (Yan vd.,2001). Kim ve arkadaşları (2006) Kore’de Network hizmetleri ile ilgili çalışmalarında müşteri memnuniyetinin müşteri sadakatini pozitif olarak etkilediğini belirtmişlerdir. Venkatesh ve arkadaşları (2002) internet pazarlaması ile ilgili çalışmalarında müşterilerin sadakati ile memnuniyetleri arasında karşılıklı bir etkileşimin olduğunu saptamışlardır. Türkiye’de GSM sektöründe yapılan bir çalışmada da müşteri memnuniyeti ile müşteri sadakati arasında güçlü bir ilişkinin varlığı belirlenmiştir (Arasıl, 2004).

3.3 VERİ MADENCİLİĞİ TEKNİKLERİ

Veri madenciliği teknikleri, tanımlayıcı ve tahmin edici teknikler olmak üzere ikiye ayrılır: Tanımlayıcı teknikler, karar vermeye yardım edecek verilerin tanımlanmasını sağlar. Birliktelik kuralları (association rules) ve kümeleme (clustering) tanımlayıcı tekniklere örnek gösterilebilir. Tahmin edici teknikler ise sonuçları bilinen verileri kullanarak sonuçları bilinmeyen veri kümelerinin sonuçlarının tahmin edilmesini sağlar. Gerileme (regression), sınıflandırma (classification) ve sapma (deviation) tahmin edici tekniklerdendir (Küçüksille, 2009).

Günümüzde kullanılan VM teknikleri 3 grupta toplanır.

- Sınıflama ve Regresyon
- Kümeleme
- Birliktelik Kuralları ve İlişki Analizi

Sınıflama ve Regresyon, tahmin etmede ve regresyon analizinde en çok kullanılan bir yöntemdir. Hastalık teşhisi, dolandırıcılık tespiti, telekomünikasyon ve pazarlama vb. alanlar sınıflama algoritmalarının kullanıldığı yaygın alanlardır. Sınıflamada kullanılan algoritmalar: Karar Ağaçları, Yapay Sinir Ağları, Bayesyen, CART, SLIQ, Sprint, C4.5, C5, ID3 vb.

Birliktelik Kuralları ve İlişki Analizi, belirli bir veri kümesinde yüksek sıklıkta birlikte görülen özellik değerlerine ait ilişki kurallarının keşfidir (Argüden ve Erşahin, 2008). Birliktelik kurallarının en çok kullanıldığı alan pazar sepeti analizidir (Market Basket Analysis). Bu analiz yapılan alışverişlerde alınan ürünlerin birbiriyle olan birliktelik ilişkisini incelemektedir. Ayrıca market sepeti analizinde müşteri ile ilgili veri hareketlerinden gelecekte müşterinin nasıl bir tercih yapacağına dair sonuçlar tahmin edilmektedir (Roiger ve Geatz, 2003). Birliktelik kuralları ve ilişki analizinde kullanılan algoritmalar: AIS Algoritması, SETM Algoritması, Apriori Algoritması, CD, PDM, IDD, PAR vb.

Kümeleme, birbirine yakın veya benzeyen verileri kümelere ayırmaktadır. Böylelikle kümeler kendi içinde bir anlam ifade etmektedir. Kümeleme analizi bir hedef değişken içermediğinden sınıflama analizinden farklı bir yaklaşımdır. Kümeleme analizinde hedef değişkenin değerini belirlemeye yönelik sınıflama, tahmin etme veya kestirim yapılmaya çalışılmaz. Bunun yerine verinin tamamını bölümlere ayırmak için homojen alt gruplar veya kümeler araştırılır. Bu işlem gerçekleştirilirken kümeler içindeki verilerin benzerliği göz önüne alınır. Benzer olmayan veriler kümenin dışında kalacaktır (Larose, 2005).

Kullanılan algoritmalar: SimpleK-Means, Clarans, Birch, Cure, Rock, Chamaleon, Dbscan, Clique vb.

3.3.1 Karar Ağaçları (Decision Trees)

Karar ağaçları, gerek anlaşılması gerekse uygulanıp yorumlanması konusunda kolaylık sağlamasından dolayı sınıflandırma algoritmaları içerisinde yaygın bir şekilde kullanılan algoritmalarından biridir.

Büyük veri tabanlarının kullanıldığı pek çok sınıflama probleminde ve karmaşık ya da hata içeren bilgilerde karar ağaçları yararlı bir çözüm olmaktadır (Türe, Tokatlı ve Kurt, 2008).

Verilere sınıflandırma uygulamak için önce ağaç oluşturulmaktadır. Veriler, bu ağaca uygulanarak sonuçların sınıflandırılması istenmektedir. Karar ağaçları kendi içerisinde düğüm, dallar ve yapraklardan oluşmaktadır. Önce soruları oluşturan düğümler oluşturulur. Daha sonra farklı cevaplara göre dallar oluşturulmaktadır. Yapraklar ise hangi sınıfa ait olduğu sonucunu barındırır. Karar ağacı işlemleri sınıflama işlemini yapabilmek için önce düğümlerden başlamaktadır. Dalların sonucunda yaprak oluşmuyorsa tekrar bir düğüm oluşturularak sonuca gitmeye çalışılır.

Karar ağacı algoritmaları genel olarak aşağıdaki kod bloğunda çalışır: (Silahtaroğlu, 2008).

D: Öğrenme Veritabanı

T: Kurulacak ağaç

T=0 (Ağacın boş küme olması)

Dallara ayırma kriterlerini belirle

T= kök düğümü belirle

T= dallara ayrılma kurallarına göre kök düğümü dallara ayır;

Her bir dal için

do

Bu düğüme gelecek değişkeni belirle

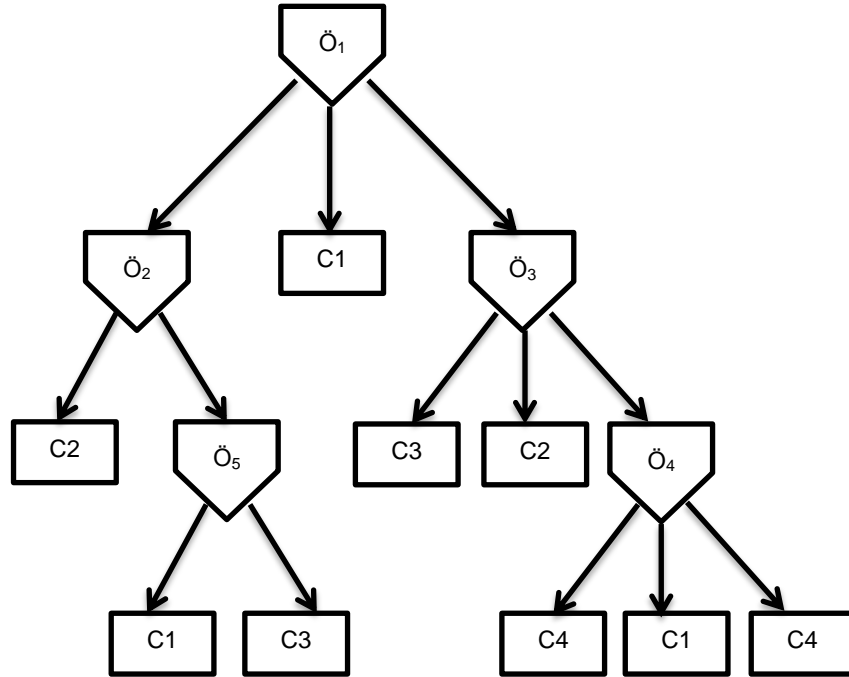
if (durma koşuluna ulaşıldı)

Yaprak ekle ve dur

Else

Loop

Şekil 3'de düğüm, dallar ve yapraklardan oluşan örnek bir karar ağacı gösterilmektedir.



Şekil 3. Örnek Karar Ağacı

Bir karar ağacı oluşturulmadan önce 2 veri kümesi yaratılır. Bunlardan ilki test kümesi, diğeri ise eğitim kümesidir. Test kümesi karar ağaçlarının ya da sınıf kurallarının doğruluğu için kullanılır. Eğer doğruluk kabul edilebilir oranda ise, kurallar yeni verilerin sınıflanması amacıyla kullanılır. Test verisine uygulanan bir modelin doğruluğu, doğru sınıflamanın test verisindeki tüm sınıflara oranıdır. Test örneğinde bilinen sınıf, model tarafından tahmin edilen sınıf ile karşılaştırılır. Eğer modelin doğruluğu kabul edilebilir bir değer ise model, sınıfı bilinmeyen yeni verileri sınıflama amacıyla kullanılabilir (Chaudhuri, 1998).

Karar ağacı oluşturulurken kullanılan veritabanının bir kısmı öğrenme işlemi için kullanılarak ağaç oluşturulacaktır. Veritabanının bir kısmı da

oluşturulan ağacı test etmek için kullanılır. Ağaç oluşturulurken kurulan sistemin çalışıp çalışmadığı belirlenir. Eğer ağaç istenen şekilde çalışıyorsa dallanma durdurulur ve sınıflandırma tamamlanır. Programdaki durdurma kriteri ağacın hassasiyetini de ortaya koymaktadır. Geç durdurulan bir ağaç daha fazla dallanacak ve ağaç daha geniş olup, çalışma süreci daha da uzayacaktır. Fakat daha duyarlı sonuç alınacaktır. Ağacın çalışması erken durdurulmuşsa, daha kısa sürede çalışsa da tam öğrenme gerçekleşmeme olasılığı vardır (Dunham, 2003).

Ağaç oluşturma işleminde yapılan işlemlerden birisi de budama işlemidir. Ağaçta oluşmuş sınıflamaya bir katkısı olmayan veya sonucu etkilemeyen dallar ağaçtan çıkarılmaktadır. Yani gereksiz ayrıntıların sonuçtan alınması işlemidir. Eğer ağaçta birçok düğüm ve dallar olursa, ağacın alt dallarına ve yapraklarına katılan veri sayısı azalır, ağacın hassasiyeti azalacaktır (Cabena, 1998).

3.3.2 Apriori Algoritması

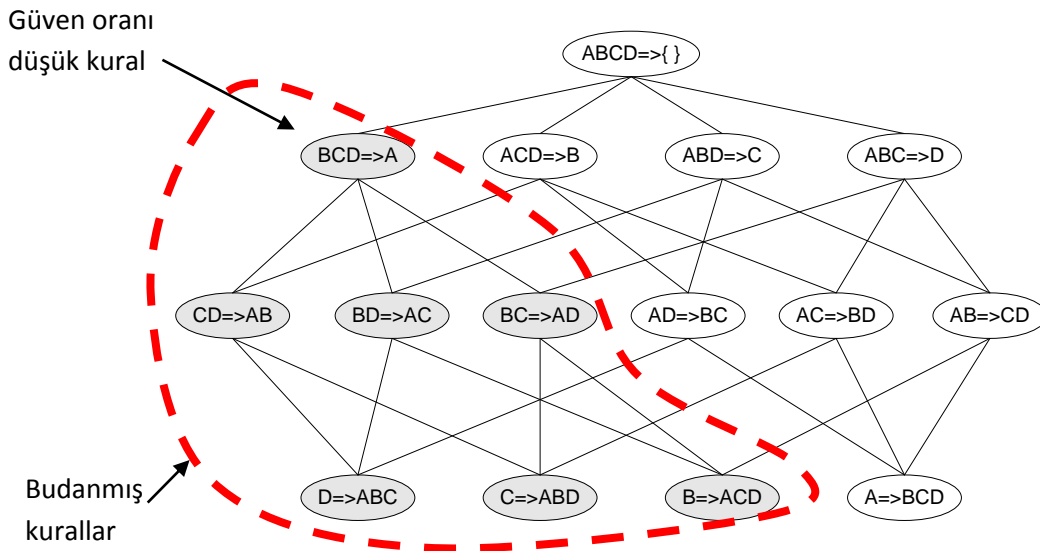
Apriori algoritması, Agrawal ve Srikant tarafından 1994 yılında geliştirilmiştir. Apriori algoritması ismini, sık geçen öge kümelerinin ön bilgisini kullanmasından, yani bilgileri bir önceki adımından almasından dolayı bir önceki (prior) anlamına gelen “apriori” den almaktadır (Chen vd., 2006). Bu algoritma, birliktelik kuralı algoritmaları içinde en fazla bilinen ve uygulanan algoritmadır (Agrawal, Imie ve Swam, 1993).

Apriori algoritması, herhangi bir olayın meydana gelme durumunu, diğer olayların meydana gelme olasılığına bakarak tahmin etmektedir.

Apriori algoritmasına göre , “Eğer k-öge kümesi minimum destek kriterini sağlıyorsa, bu kümenin alt kümeleri de minimum destek kriterini sağlar.” şeklindedir. Öge küme, 1 veya daha fazla elemandan oluşan

kümedir. k -öge küme (k -itemset) ise içinde k adet öge bulunan kümedir. Apriori, $(k+1)$ sık geçen öge kümesini bulmak için en sık geçen öge kümesine ihtiyaç duymaktadır. Sık geçen öge kümelerini bulmak için öncelikle olarak minimum destek kriterini sağlayan sık geçen öge kümesi bulunarak işlem yapılır. Bu süreç, algoritma sık geçen öge kümeleri bulamayınca kadar devam eder (Özseven ve Düğenci, 2011).

Apriori algoritması en çok geçen öge kümelerini bulmak için birçok kez veritabanını taramaktadır. İlk olarak tek elemanlı ve minimum destek sağlayan en sık geçen ögeyi bulur. Tekrarlayan aramalarda bir önceki sık geçen ögeyi aratarak potansiyel olabilecek öge kümelerini bulur. Aday kümelerin destek değerleri tarama esnasında hesaplanır ve aday kümelerinden minimum destek ölçütü sağlayan kümeler o geçişte üretilen sık geçen öge kümeleri olur. Sık geçen öge kümeleri bir sonraki geçiş için aday küme olurlar. Bu süreç yeni bir sık geçen öge kümesi bulunamayana kadar devam eder (Erkan vd., 2005). Şekil 4'de güven desteği düşük olan kuralın veri tabanından o kuralın oluşturduğu diğer kurullarla birlikte silinmesi gösterilmiştir (Tan, Steinbach,Kumar, 2006).



Şekil 4. Güven Oranı Düşük Kuralların Budanması

Apriori algoritmasının en yaygın kullanım alanlarından Pazar sepet çözümlemesinde, satılan ürünler arasındaki ilişkiyi sağlamak için;

- Destek
- Güven

ölçütlerinden yararlanır.

A ve B, birbirinden farklı birer öge küme olarak düşünülürse;

Kural güven ölçütü, A ürün grubunu alan müşterilerin B ürün grubunu da alma olasılığını ortaya koyar.

$$güven(A \rightarrow B) = \frac{sayı(A, B)}{sayı(A)}$$

Kural destek ölçütü bir ilişkinin tüm alışverişler içinde hangi oranda tekrarlandığını gösterir.

$$destek(A \rightarrow B) = \frac{sayı(A, B)}{N}$$

Güven ve destek değerlerini karşılaştırmak için eşik değere gereksinim duyar. Hesaplanan destek veya güven ölçütlerinin destek (eşik) ve güven (eşik) değerlerinden büyük olması istenir.

Hesaplanan destek veya güven ölçütleri ne kadar büyükse birliktelik kurallarının da o derece güçlü olduğuna karar verilir.

Apriori algoritması uygulanması esnasında ekrana çıkan birden fazla birliktelik kuralı ölçü değerleri bulunmaktadır. Bunlar güven (confidence) ve destek (support) ölçümü haricinde lift, leverage, conviction ölçüm değerleridir.

Lift değeri, X ve Y değerinin bağımsız olması durumunda ne sıklıkla birlikte geçtiklerini göstermektedir. Aşağıdaki formül yoluyla lift ya da diğer bir deyişle ilgi değeri hesaplanabilmektedir [3].

$$Lift(X \rightarrow Y) = Lift(Y \rightarrow X) = \frac{conf(X \rightarrow Y)}{supp(Y)} = \frac{conf(Y \rightarrow X)}{supp(X)} = \frac{P(Y \text{ ve } X)}{P(X)P(Y)}$$

Leverage ölçüm değeri, X ve Y'nin birliktelik kuralı uygulanması ile uygulanmaması arasındaki istatistiksel bağlarının ilişkisini hesaplamaktadır. Bu kaldıraç değeri aşağıdaki formül ile hesaplanabilmektedir.

$$Leverage(X \rightarrow Y) = P(X \text{ ve } Y) - (P(X)P(Y))$$

Conviction değeri yani kanaat ise, X elemanlarının Y elemanları olmadan görülme olasılığını hesaplamaktadır [4]. Aşağıdaki formül yoluyla oluşturulan kuralın conviction değeri hesaplanabilmektedir.

$$Conviction(X \rightarrow Y) = \frac{1 - supp(Y)}{1 - Conf(X \rightarrow Y)} = \frac{P(X)P(\neg Y)}{P(X \text{ ve } \neg Y)}$$

Örnek :

N veri tabanına ait işlemler verilmiştir. Bu veri tabanında 5 işlem bulunmaktadır, $|N|=5$ şeklinde gösterilir.

- D veritabanı taranır, her bir nesnenin destek sayısı belirlenir.

Tablo 1. Farklı müşterinin aldığı market ürünleri

Müşteri	Aldığı ürünler
1	Şeker, Çay, Ekmek
2	Ekmek, Peynir, Zeytin, Süt
3	Şeker, Peynir, Deterjan, Ekmek, Süt
4	Ekmek, Peynir, Çay, Süt
5	Peynir, Süt, Şeker, Bira

- Adayların destek değerleri minimum destek değerleriyle karşılaştırılır ve yeni bir küme oluşturulur.

Tablo 2. Destek değerlerin hesaplanması

Urün	Sayı
Şeker	3
Çay	2
Ekmek	4
Süt	4
Peynir	4
Deterjan	1
Bira	1
Zeytin	1

- Adayların destek değerleri minimum destek değerleriyle karşılaştırılır ve yeni bir küme oluşturulur.

Tablo 3. Eşik destek değerine eşit ya da daha büyük desteğe sahip ürünler

Ürün	Sayı
Şeker	3
Ekmek	4
Süt	4
Peynir	4

- D veritabanı taranarak her ikili grup adayının destek değerleri hesaplanır.

Tablo 4. İkili ürün gruplarının destek değerleri

Ürün	Sayı
Şeker, Ekmek	2
Şeker, Süt	2
Şeker, Peynir	2
Ekmek, Süt	3
Ekmek, Peynir	3
Süt, Peynir	4

- Adayların destek değerleri minimum destek değerleriyle karşılaştırılır ve yeni bir küme oluşturulur.

Tablo 5. Eşik destek sayısı olan üç değerine eşit yada daha büyük desteğe sahip ikili ürün grupları

Ürün	Sayı
Ekmek, Süt	3
Ekmek, Peynir	3
Süt, Peynir	4

- D veritabanı taranarak her üçlü grup adayının destek değerleri hesaplanır.

Tablo 6. Üçlü ürün gruplarının destek değerleri

Ürün	Sayı
Ekmek, Süt, Şeker	1
Ekmek, Süt, Çay	1
Ekmek, Süt, Peynir	3
Ekmek, Süt, Deterjan	1
Ekmek, Süt, Bira	0
Ekmek, Süt, Zeytin	1
Ekmek, Süt, Şeker	1
Ekmek, Süt, Çay	1
Ekmek, Süt, Deterjan	1
Ekmek, Süt , Bira	0

Ekmek, Süt, Zeytin	1
Süt, Peynir, Şeker	2
Süt, Peynir, Çay	1
Süt, Peynir, Deterjan	1
Süt, Peynir, Bira	1
Süt, Peynir, Zeytin	1

- Adayların destek değerleri minimum destek değerleriyle karşılaştırılır ve yeni bir küme oluşturulur.

Tablo 7. Eşik destek sayısına eşit ya da daha büyük destek sayısına sahip üçlü ürün grubu

Ürün	Sayı
Ekmek, Süt, Peynir	3

- {ekmek, süt, peynir} kümesi için kural destek sayısı, aşağıdaki gibidir.

$$\text{sayı}(A,B)=\text{sayı}(\text{ekmek, süt, peynir})=3$$

$$\text{destek}(A \rightarrow B) = \frac{\text{sayı}(\text{ekmek, süt, peynir})}{N} = \frac{3}{5} = 0.6$$

Destek ölçütü başlangıç olarak verilen eşik değerden küçük değilse kural destek sayılarına bağlı olarak birliktelik kuralları türeterek bu kurallar için güven ölçütleri elde edilebilmektedir.

Sonuçlar

1)

$$\text{güven}(\text{Ekmek, süt} \rightarrow \text{peynir}) = \frac{\text{sayı}(\text{Ekmek, süt, peynir})}{\text{sayı}(\text{Ekmek, süt})} = \frac{3}{3} = \%100$$

2)

$$\text{güven}(\text{Ekmek} \rightarrow \text{peynir, süt}) = \frac{\text{sayı}(\text{Ekmek, süt, peynir})}{\text{sayı}(\text{Ekmek})} = \frac{3}{4} = \%75$$

3)

$$\text{güven}(\text{peynir} \rightarrow \text{ekmek, süt}) = \frac{\text{sayı}(\text{Ekmek, süt, peynir})}{\text{sayı}(\text{peynir})} = \frac{3}{4} = \%75$$

4)

$$\text{güven}(\text{süt} \rightarrow \text{ekmek, peynir}) = \frac{\text{sayı}(\text{Ekmek, süt, peynir})}{\text{sayı}(\text{süt})} = \frac{3}{4} = \%75$$

Tablo 8. Birliktelik kuralları

Birliktelik kuralı	Anlamı	Güven
Ekmek&Süt→Peynir	Ekmek ve Sütün bulunduğu ürün kümesinde Peynirin olma olasılığı	%100
Ekmek→Peynir&Süt	Ekmeğin yer aldığı bir ürün kümesinde peynir ve sütün olma olasılığı	%75
Peynir→Ekmek&Süt	Peynirin yer aldığı bir ürün kümesinde ekmek ve sütün olma olasılığı	%75
Süt→Ekmek&Peynir	Sütün yer aldığı bir ürün kümesinde ekmek ve peynirin olma olasılığı	%75

Yukarıda bulunan Tablo 8’de oluşturulan birliktelik kuralı, anlamları ve bu kurallara ait güven oranları belirtilmektedir.

3.3.2.1 Apriori Algoritmasının Yapısı

1994 yılında “20th Very Large Database Endowment” konferansında apriori algoritması sunulmuştur. Bu konferansta, Agrawal ve Srikan algoritmanın ayrıntılarını ve pseudo kodunu aşağıdaki şekilde sunmuştur (Agrawal ve Srikant, 1994):

- Verilerin ilk tarandığı esnada, geniş nesne kümelerinin bulunması için, tüm nesnelere sayılır.
- Bir sonraki tarama, k'inci tarama olup iki aşamadan oluşur.
- Apriori-gen fonksiyonu kullanılarak, k-1'inci taramada elde edilen (L_{k-1}) nesne kümeleriyle, C_k aday nesne kümeleri oluşturulur.
- Daha sonra veritabanı taranarak, C_k daki adayların desteği sayılır.
- Hızlı bir sayım için, verilen bir / işleminde C_k yı oluşturan adayların çok iyi belirlenmesi gerekmektedir.

- 1) $L_1 = \{ \text{sık geçen 1 - öge kümesi} \};$
- 2) for ($k=2; L_{k-1} \neq \emptyset; k++$) do begin
- 3) $C_k = \text{apriori - gen} (L_{k-1});$ // Yeni adaylar
- 4) for all transactions - hareketler $t \in D$ do begin
- 5) $C_t = \text{subset} (C_k, t);$ // Adaylar t içindedir
- 6) for all candidates – adaylar $c \in C_t$ do
- 7) $c.\text{count}++;$
- 8) end
- 9) $L_k = \{ c \in C_k \mid c.\text{count} \geq \text{minsup} \}$
- 10) end
- 11) Answer = $\bigcup_k L_k;$

4. VERİ MADENCİLİĞİ ANALİZ PROGRAMI: WEKA

4.1 WEKA NEDİR?

WEKA programı veri madenciliği uygulaması olup yapay zeka tekniğiyle kullanılan asıl bilgiyi edinmeye yardımcı bir programdır. Bu program sayesinde istatistiksel tahminler çıkarılabilir, kesin olmamak kaydıyla ihtimaller oluşturulabilir. 1993 yılında Yeni Zelanda Waikato Üniversitesi'nde java diliyle geliştirilmiştir. WEKA, açık kaynak kodlu olmakla birlikte GNU lisansı ile ücretsiz olarak dağıtılmaktadır. Veri Madenciliği denildiğinde akla gelen yazılımların başında olan bu program Waikato Environment for Knowledge Analysis kelimelerinin baş harflerini alarak WEKA adını almıştır.

WEKA programı kendine özgü bir uzantı olan .arff uzantısını kullanmaktadır. Attribute Relationship File Format olarak geçen ASCII tabanlı bu uzantı .csv uzantısının çevrilmesiyle de elde edilebilmektedir.

ARFF dosyasında dosyanın ilişik olduğu adı, değişkenleri ve aldığı değerleri yazmaktadır. Değişkenlerin alabileceği değerler kendi içinde virgül (,) ile ayrılmaktadır.

@relation <ilişik adı>

@attribute <değişken adı> <alabileceği değer>

@data <değişkenlerin sıralamasına göre aldığı değerler>

Aşağıda bulunan Şekil 5'te ARFF dosya yapısı örneği bulunmaktadır.

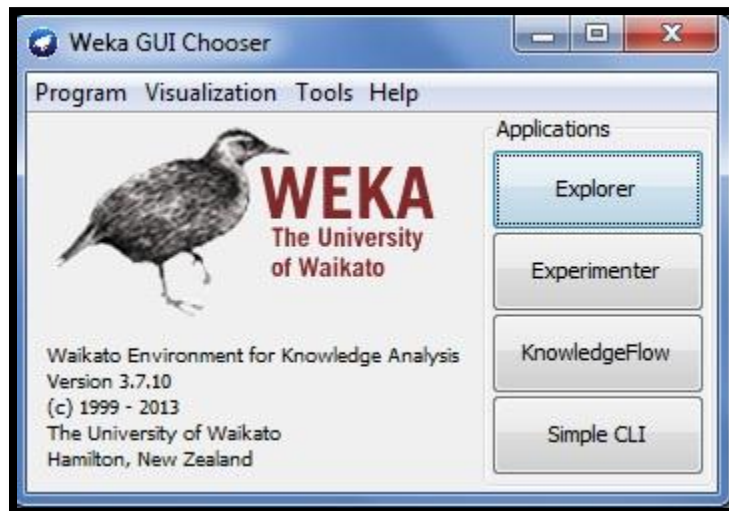
```

1 @relation csv
2
3 @attribute cins {erkek,kadin}
4 @attribute yas {15-24,25-34,35-44,45-54,'55 ustu'}
5 @attribute medeni {bekar,evli,dul,bosanmis}
6 @attribute egitim {onlisans,lisans,lise,y.lisans,doktora,ilkogretim}
7
8 @data
9 erkek,15-24,bekar,onlisans
10 erkek,15-24,bekar,onlisans
11 kadin,15-24,bekar,onlisans
12 erkek,15-24,bekar,onlisans
13 kadin,15-24,bekar,onlisans
14

```

Şekil 5. ARFF Dosya Yapısı

WEKA programı görsel arayüzü ve içerisinde bulundurduğu hazır algoritmaları ile veri madenciliği uygulamalarını kullanıcıları için kolaylaştırmaktadır. Ücretsiz olarak son sürümü internet üzerinden indirilebilmektedir. Kurulum tamamlanmasından sonra program çalıştırıldığında aşağıda bulunan Şekil 6'daki ekran, kullanıcı karşısına çıkmaktadır.



Şekil 6. WEKA Programı Ara Yüzü

Applications panelinde Explorer, Experimenter, KnowledgeFlow ve Simple CLI modülleri bulunmaktadır. Bu modüller sayesinde verileri işleme, regresyon analizi, sınıflandırma, kümeleme ve ilişkilendirme yapılabilmektedir.

4.2 SIMPLE CLI MODÜLÜ

“ Basit Komut Satırı” adı altında temel komutları kullanabileceğimiz bir alan gelmektedir. Veri madenciliği çalışmaları bu komutlarla yapılabilmektedir.

Bu komutlar :

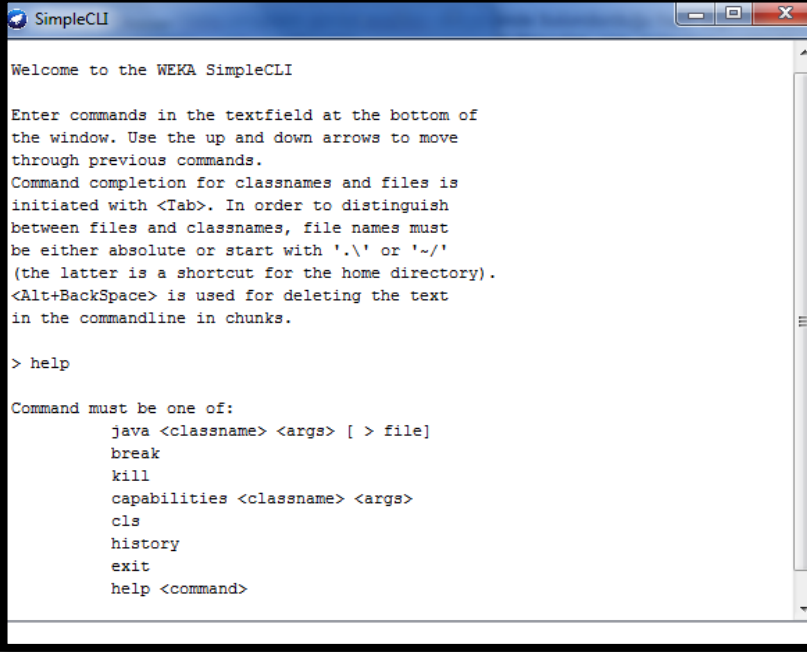
- java <classname> <args> [> file]

Argümanları ile java sınıflarını çalıştırır.

- Break
Çalışan uygulamayı durdurur.
- Kill
Çalışan uygulamayı durdurarak sonlandırır.
- capabilities <classname> <args>
Argümanları ile birlikte belirtilen sınıfın özelliklerini listeler.
- cls
Ekranı temizler.
- History
Komut geçmişini listeler.
- Exit
Konsoldan çıkış yapar.
- help <command>

Yardım komutlarını ekrana getirir.

Şekil 7’de WEKA programının SimpleCLI paneli gösterilmektedir.



```

SimpleCLI

Welcome to the WEKA SimpleCLI

Enter commands in the textfield at the bottom of
the window. Use the up and down arrows to move
through previous commands.
Command completion for classnames and files is
initiated with <Tab>. In order to distinguish
between files and classnames, file names must
be either absolute or start with './' or '~/ '
(the latter is a shortcut for the home directory).
<Alt+BackSpace> is used for deleting the text
in the commandline in chunks.

> help

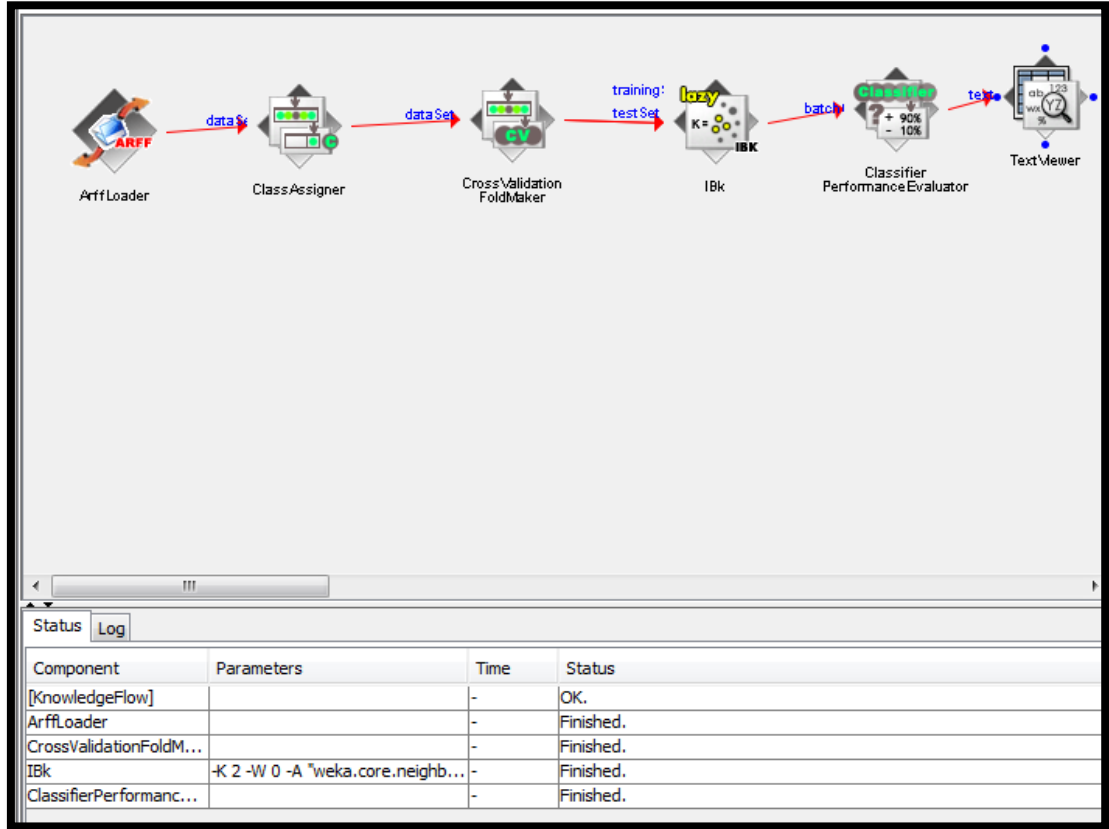
Command must be one of:
    java <classname> <args> [ > file]
    break
    kill
    capabilities <classname> <args>
    cls
    history
    exit
    help <command>
  
```

Şekil 7. WEKA SimpleCLI Paneli

4.3 KNOWLEDGEFLOW MODÜLÜ

Knowledge Flow modülü görsel bir ortam yardımıyla WEKA içindeki kütüphanelere erişmeyi sağlamaktadır. Explorer modülünde yapılabilecek işlevleri akış yöntemleriyle gösteren bir arayüzdür. Sürükle bırak yöntemiyle “Associations”, “Classifiers”, “DataSources”, “DataSinks”, “Filters”, “Clusterers”, “Evaluation”, “Visualization” sekmelerinden görsel nesnelere çekerek veri madenciliği uygulanabilmektedir.

Şekil 8'de WEKA programına ait KnowledgeFlow modülü gösterilmektedir.



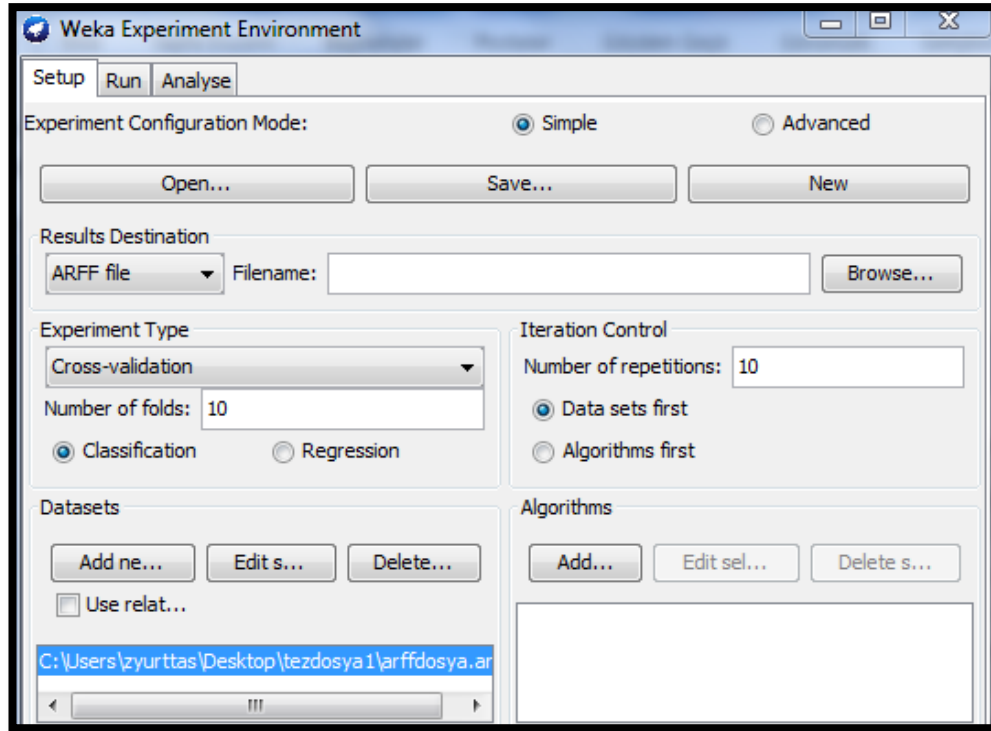
Şekil 8. WEKA KnowledgeFlow Modülü

4.4 EXPERIMENTER MODÜLÜ

Bu modül içerisinde birden fazla kaynak dosyayı seçme imkanı sağlayarak, kullanıcılara aynı anda uygulanabilen algoritmalarla karşılaştırma imkanı sağlamaktadır. Böylelikle aynı anda kullanıcı hızlı, kolay kıyaslayabilme olanağına sahip olmaktadır.

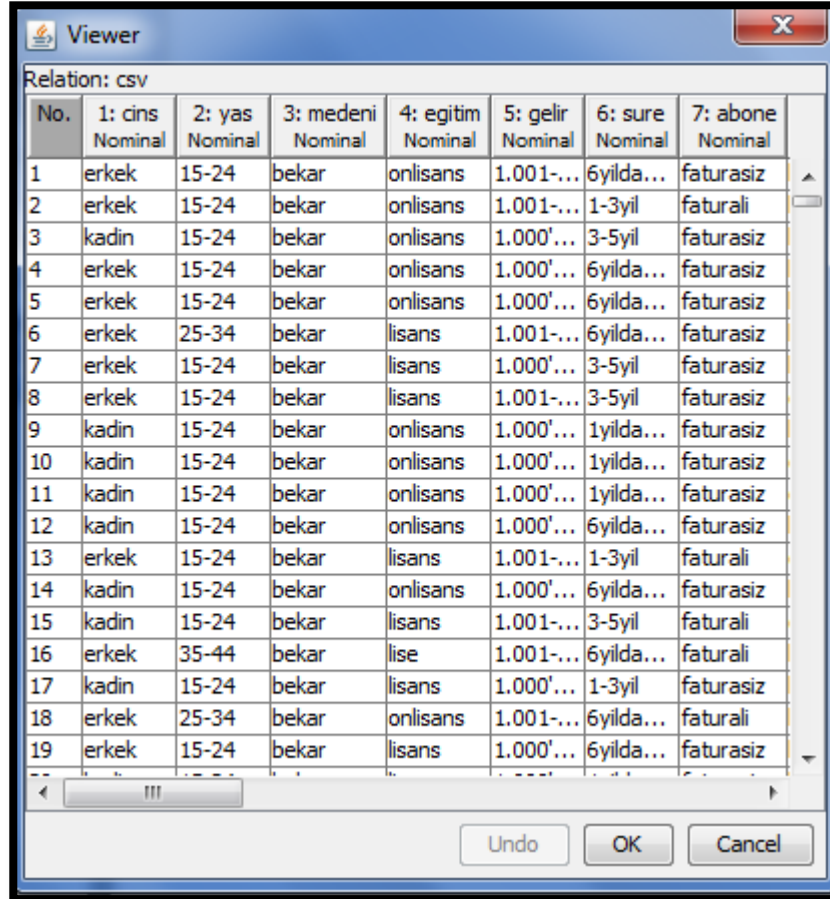
Bu modülde birden fazla dosya eklenebilir, istenilen algoritma uygulanabilir ve görüntülenebilir. Basit ve gelişmiş olmak üzere arayüzü seçenekleri bulunmaktadır. Yeni bir csv ya da arff uzantılı kaynak dosya

çağırma imkanı olmaktadır. Ayrıca yapılan uygulama kaydedilebilir ve daha sonra tekrar çağırılabilir. Şekil 9'da WEKA programının Experimenter modülü bulunmaktadır.



Şekil 9. WEKA Experimenter Modülü

Aşağıda bulunan Şekil 10'da WEKA Experimenter Viewer paneli gösterilmiştir.

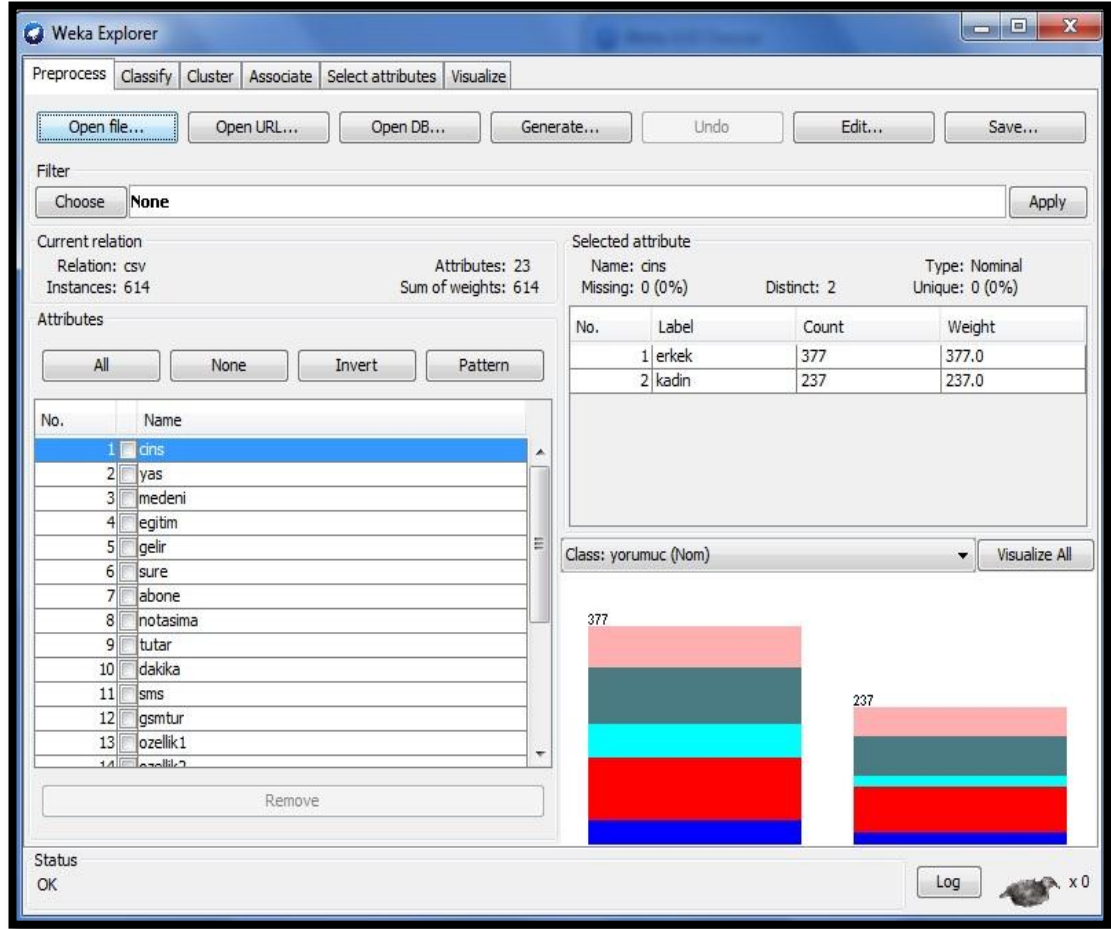


No.	1: cins Nominal	2: yas Nominal	3: medeni Nominal	4: egitim Nominal	5: gelir Nominal	6: sure Nominal	7: abone Nominal
1	erkek	15-24	bekar	onlisans	1.001-...	6yilda...	faturasiz
2	erkek	15-24	bekar	onlisans	1.001-...	1-3yil	faturali
3	kadin	15-24	bekar	onlisans	1.000'...	3-5yil	faturasiz
4	erkek	15-24	bekar	onlisans	1.000'...	6yilda...	faturasiz
5	erkek	15-24	bekar	onlisans	1.000'...	6yilda...	faturasiz
6	erkek	25-34	bekar	lisans	1.001-...	6yilda...	faturasiz
7	erkek	15-24	bekar	lisans	1.000'...	3-5yil	faturasiz
8	erkek	15-24	bekar	lisans	1.001-...	3-5yil	faturasiz
9	kadin	15-24	bekar	onlisans	1.000'...	1yilda...	faturasiz
10	kadin	15-24	bekar	onlisans	1.000'...	1yilda...	faturasiz
11	kadin	15-24	bekar	onlisans	1.000'...	1yilda...	faturasiz
12	kadin	15-24	bekar	onlisans	1.000'...	6yilda...	faturasiz
13	erkek	15-24	bekar	lisans	1.001-...	1-3yil	faturali
14	kadin	15-24	bekar	onlisans	1.000'...	6yilda...	faturasiz
15	kadin	15-24	bekar	lisans	1.001-...	3-5yil	faturali
16	erkek	35-44	bekar	lise	1.001-...	6yilda...	faturali
17	kadin	15-24	bekar	lisans	1.000'...	1-3yil	faturasiz
18	erkek	25-34	bekar	onlisans	1.001-...	6yilda...	faturali
19	erkek	15-24	bekar	lisans	1.000'...	6yilda...	faturasiz

Şekil 10. WEKA Experimenter Viewer Paneli

4.4 EXPLORER MODÜLÜ

Explorer Modülü kullanıcıya sınıflandırma, kümeleme, ilişki analizi yapma, nitelik seçme ve görselleştirme imkanı sağlayan araştırmaya yönelik ara yüzüdür. Önışlem panelinde bulunan alanda, kaynak dosya istenen diskten seçilebilir, URL'den ya da kayıtlı database' den de çekilebilmektedir. Uygulanmak istenen filtre bu modülde seçilebilmektedir. Şekil 11'de WEKA Explorer Modülü Preprocess sekmesi gösterilmiştir.



Şekil 11. WEKA Explorer Modülü Preprocess Sekmesi

4.5 CLASSIFY

Modülde bulunan Classify sekmesi sınıflandırma algoritmalarının bulunduğu ve uygulandığı bölümdür. “J48”, “BayesNet”, “SMO”, “LibSVM”, “IBk”, “AdaBoostM1”, “ZeroR”, “OneR”, “JRip” vb. algoritmalar kullanılabilir.

Test Options:

Verinin nasıl parçalanacağını ve kaçının test kaçının training olarak kullanılacağını belirlediğini bölümdür.

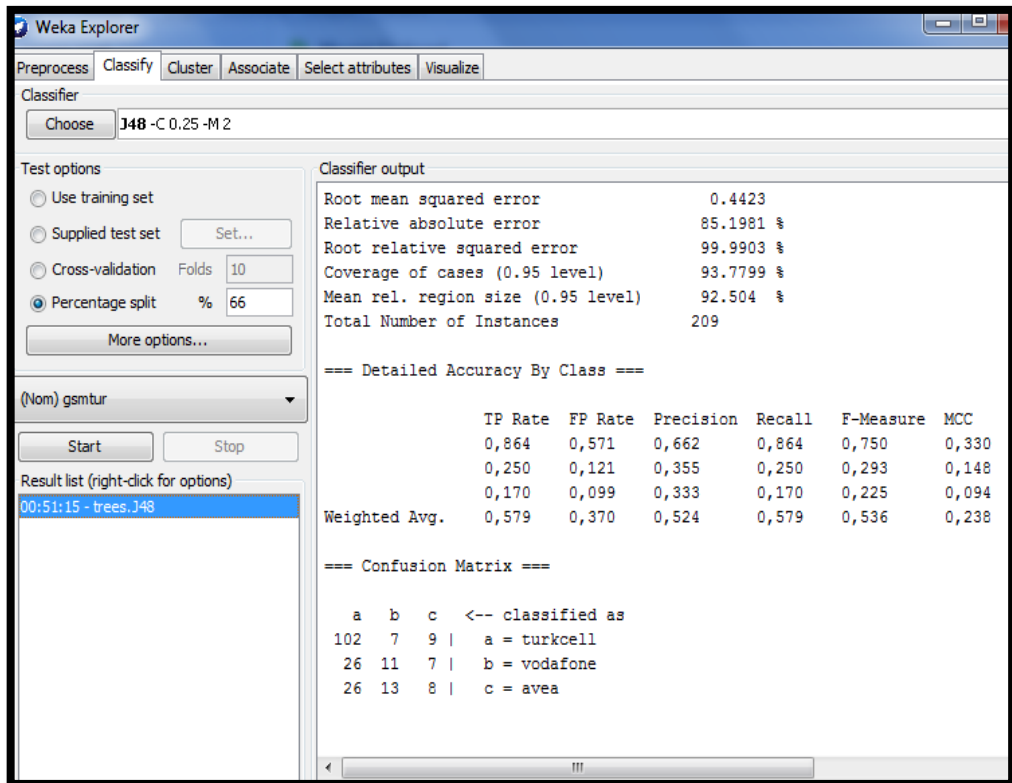
Use Training set: Önışlem sekmesinde yüklenen veri seti kullanarak sınıflandırma yapılmaktadır.

Supplied test set: Dış ortamdaki alınan veri seti test, var olan veri seti ise training olarak kullanılabilir.

Cross Validation: K-Cross Validation yöntemiyle veri seçme işlemi yapmaktadır.

Percentage Split: Mevcut verilerin training olarak yüzde kaçının kullanılacağına karar verilmektedir.

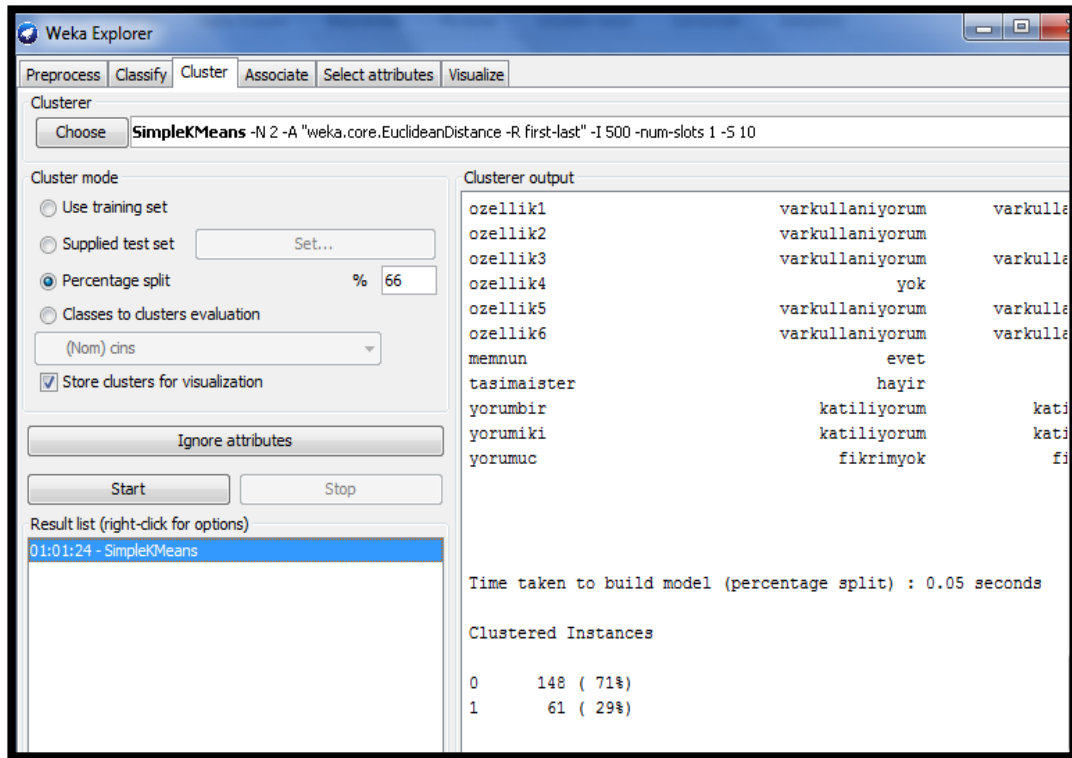
Şekil 12'de WEKA Explorer Modülü Classify sekmesi gösterilmiştir.



Şekil 12. WEKA Explorer Modülü Classify Sekmesi

4.6 CLUSTER

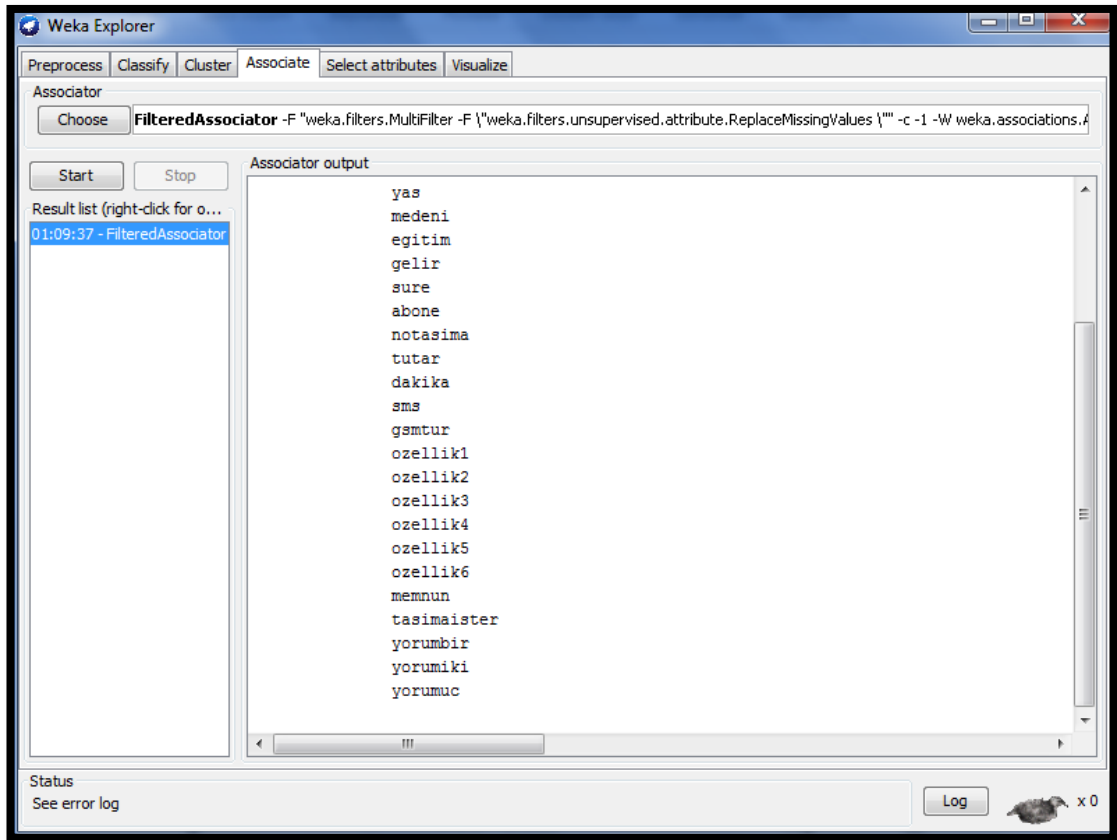
Cluster sekmesinde kümeleme algoritmaları kullanılmaktadır. “SimpleKMeans”, “FilteredClusterer”, “Cobweb”, “EM”, “HierarchicalClusterer”, “FathestFirst”, “MakeDensityBasedClusterer” adlı kümeleme algoritmalarını uygulayabilme imkanı vermektedir. Şekil 13’de WEKA Explorer Modülü Cluster sekmesi gösterilmiştir.



Şekil 13. WEKA Explorer Modülü Cluster Sekmesi

4.7 ASSOCIATE

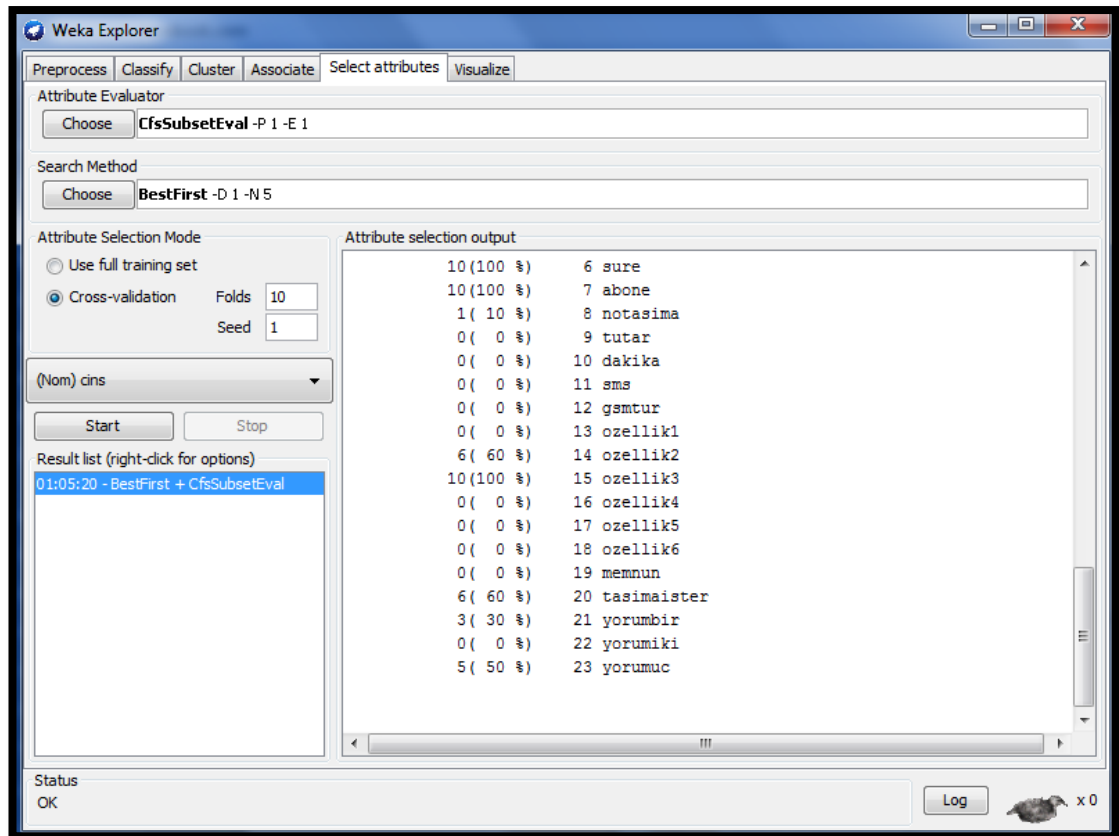
Associate sekmesi ilişki kurallarının uygulandığı bölümdür. Apriori algoritmasının uygulandığı bölüm bu sekmede yer almaktadır. Şekil 14'de WEKA Explorer Modülü Associate sekmesi gösterilmiştir.



Şekil 14. WEKA Explorer Modülü Associate Sekmesi

4.8 SELECT ATTRIBUTES

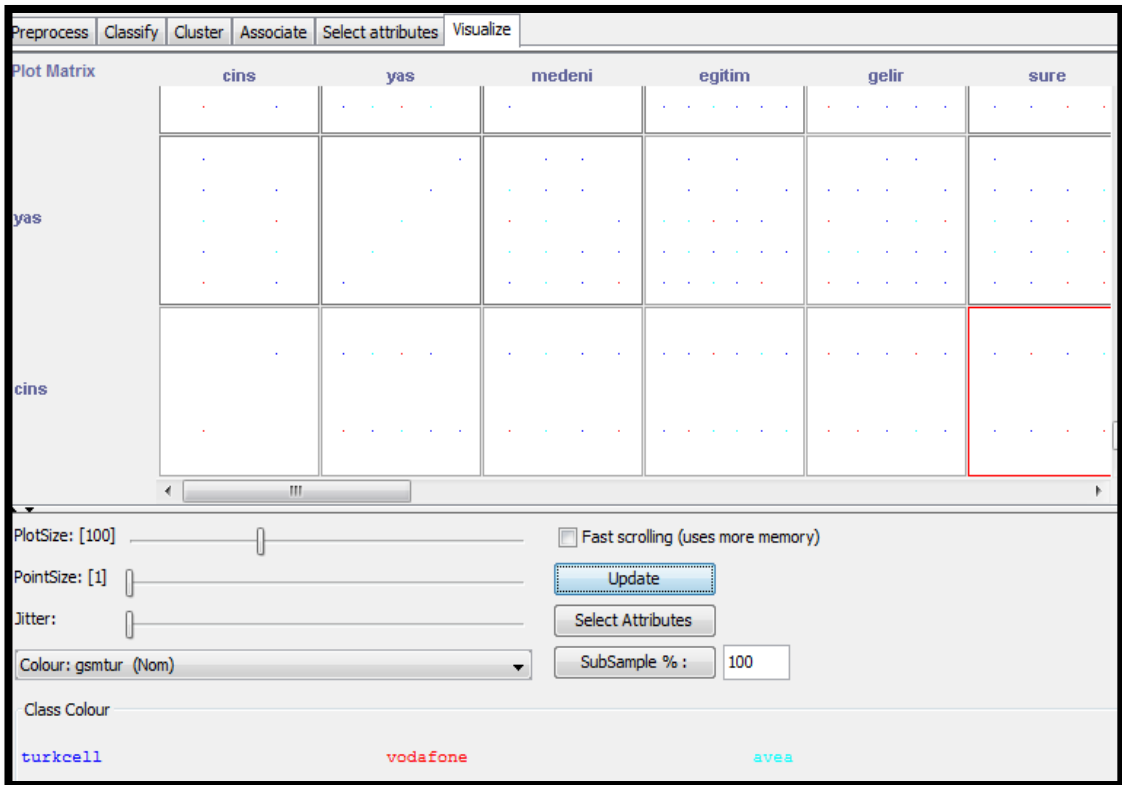
Select attributes, öznitelik seçiminin yapıldığı sekmedir. Gereksiz özniteliklerin işlemi uzatmasını engellemek için çıkarılması gerekmektedir. İlişkisi olmayan öznitelikler bu sekmede görülmektedir. Şekil 15'de WEKA Explorer Modülü Select Attributes sekmesi gösterilmiştir.



Şekil 15. WEKA Explorer Modülü Select Attributes Sekmesi

4.9 VISUALIZE

Visualize, verilerin özniteliklere göre dağılımının yapıldığı görsel bölümdür. Farklı değişkenlerin birbiriyle olan ilişkisi grafik olarak incelenebilmektedir. Her bir görsel kareye tıklandığında ayrı bir panelde açılarak değişkenlerin dağılımı detaylı bir şekilde görülmektedir. Şekil 16'da WEKA Explorer Modülü Visualize sekmesi gösterilmiştir.



Şekil 16. WEKA Explorer Modülü Visualize Sekmesi

5. ANKET

Anket, daha önceden belirlenmiş bir konu üzerine kişilerin fikirlerini, duygularını, deneyimlerini öğrenebilmek için hazırlanmış sözlü veya yazılı olarak sıralı sorulardan oluşan bir bilgi edinme ve araştırma yöntemidir. Anket yöntemiyle, kişilerin sağlık durumları, ihtiyaçlarının belirlenmesi, eğitsel planları, aile ve ev yaşamları, okul başarısı, çalışma yöntemleri, sosyal alışkanlıkları, memnuniyet vb. gibi farklı türlerde bilgi sahibi olunarak geleceğe yönelik karar ve çıkarımlarda bulunulabilmektedir. Anketler yapılan araştırmalara katkı sağlayarak asıl verilere kişinin kendisinden ulaşılmasını sağlamaktadır. Araştırmalarda anket kullanımı hem daha hızlı hem de daha az maliyetli olması nedeniyle tercih sebebi olabilmektedir.

Anket kullanımının avantajları:

- 1) Anket yapılırken kişilerin şahsi bilgilerinin gizli tutulması güven doğurduğundan dolayı daha doğru ve daha tutarlı veriler doğrudan elde edilebilmektedir.
- 2) Planlanması ve uygulanması daha kolay ve sorunsuzdur.
- 3) Anket sonuçlarının analiz edilmesi veri madenciliği programları yardımıyla daha kolay olmaktadır.
- 4) Maliyet açısından diğer veri toplama yöntemlerine göre daha az maliyetli olmaktadır.
- 5) Çok kişiye ulaşılabildiğinden dolayı sonuç alabilmek daha hızlı olmaktadır.
- 6) Anket yöntemine göre gerek sözlü, gerek yazılı veya online olarak daha çok ilgili kitlelere erişilebilmektedir.
- 7) Aynı anda farklı türlerde sorular sorularak çok sayıda bilgi edinilmek mümkündür.

Anket kullanımının dezavantajları:

- 1) Yazılı anketlerin cevaplanabilmesi için anket kitlesinin en az okuryazarlık derecesine sahip olması gereklidir.
- 2) Soruların kişi bazlı anlaşılabilirliğinin farklı olması anket hızını ve ölçümlerini negatif yönde etkileyebilmektedir.
- 3) Anketin hedef kitlesinin yanlış bir şekilde anket sorularına cevap vermesi veya gerçek dışı cevaplar vermesi objektif sonuç elde edilmesini etkileyebilmektedir.
- 4) Anket sonuçları bir araştırma için kesin bilimsel sonuçları elde etmek için değil, sadece o konuda yorum yapılabilmesine ve fikir geliştirmesine olanak sağlamaktadır.
- 5) Ankete cevap veren kişi sayısının azalması, sonuç edinebilme sürecini uzatabilmektedir.
- 6) Anketin sonuçlarının yapılan araştırmaya katkısı olabilmesi için öncelikle anket sorularının dikkatli bir şekilde hazırlanması gereklidir. Aksi takdirde anketin tutarlılığını etkilemektedir.
- 7) Cevaplanması gereken soru miktarının fazlalığı kişinin anketi bitirme isteğini etkilemektedir. Ayrıca soru miktarının fazla olması, anket cevaplama süresinin artmasına neden olarak tutarlı cevaplar alınmasını etkilemektedir.

Bir anketi araştırmanın hedef kitlesine sunmadan önce aşağıdaki aşamalardan geçmesi gerekmektedir:

- 1) Anketin amacının belirlenmesi ve anketin planlanması
- 2) Anket sorularının hazırlanması
- 3) Anket sorularının tutarlılığı amacıyla bir uzmana danışılması
- 4) Küçük bir grup üzerinde ön uygulama yapılması
- 5) Ankete son halini vererek hedef kitleyle paylaşılması

5.1 ANKET TÜRLERİ

- 1) Yüz yüze Anket
- 2) Telefon ile Anket
- 3) Posta ile Anket
- 4) Web Tabanlı Anket
- 5) Karma Anket

5.2 WEB TABANLI ANKET

Anketlerin bilgi kaynağı insandır. Araştırmaya katılan katılımcıları bulmak her geçen gün daha zor ve daha maddiyatı yüksek olmaktadır. Online anket, internet üzerinde bir web server sayesinde yayınlanan ve kişilerin internet bağlantısı üzerinden her yerden erişim yapabileceği bir anket türüdür. Bu anket türü, diğer anket türlerine nazaran daha çok katılımcı erişebilmekte, daha hızlı ve daha güvenli sonuç elde edilmesini sağlamaktadır. Web tabanlı anket ya da online anket olarak da isimlendirilmektedir.

Online anket hazırlanırken dikkat edilmesi gereken hususlar şunlardır:

- 1) Online anket yöntemi veri toplama yöntemleri içerisinde en yaygın olanlarından biridir. Fakat veri toplamak için online anketin mi yoksa diğer anket türlerinin mi daha yararlı olduğuna en başında karar verilmesi gerekmektedir.
- 2) Online anket hazırlanırken dikkat edilmesi gereken en önemli hususlardan bir tanesi yapılan anketin tutarlı olmasıdır. Tutarlılığı olmayan bir anket vakit kaybına neden olacaktır.

- 3)** Online anket, bir amaca hizmet etmelidir. Konu önceden belirlenmeli ve sorular istenen hedefe yönelik hazırlanmalıdır.
- 4)** Online anket içerisinde kişinin adı, soyadı, telefon numarası vb. gibi kişisel bilgilerin sorulduğu sorulardan kaçınılması gerekmektedir. Kişisel bilgi içeren sorular, anketin tutarlılığını etkileyebilmektedir.
- 5)** Online anket yapılırken, soruların internet üzerinden cevaplanması istenmektedir. Fakat uzun online anket soruları veya soru sayısının fazla olması kişinin anketi yarım bırakmasına veya ankete verilen ciddiyetin azalmasına neden olabilmektedir. Bu yüzden anket soruları karşı tarafın anlayabileceği en kısa şekilde hazırlanmalıdır. Ayrıca soru sayısının karşı tarafı sıkmayacağı şekilde belirlenmesi gerekmektedir.
- 6)** Online anket soruları, kişilerin algılayış farklılığını gözetererek, anketi cevaplayacak kitlenin anlayabileceği şekilde anlaşılır ve sade bir şekilde sunulmalıdır.

5.3 ONLINE ANKET HAZIRLANIŞI

Online anket hazırlanırken, anket sonuçlarının sağlığı, aynı ip girişinin engellenmesi ve veri istatistikleri sağlanması sebebiyle <http://surveymonkey.com> adresinden yararlanılmıştır.

Aşağıda bulunan Şekil 17, 18, 19, 20, 21’de online ankette sorulmuş bazı soruların hazırlanış yöntemleri bulunmaktadır.

Örnek 1.

The image shows a screenshot of the SurveyMonkey question creation interface. It is titled "Question" and contains the following elements:

- Question Text:** A text input field containing "Cinsiyetiniz?".
- Insert response from... ?** A button with a question mark icon.
- Question Type:** A dropdown menu set to "Multiple Choice (Only One Answer)".
- Pick a display format:** A dropdown menu set to "Display Choices as Buttons (1 column)".
- Answer Options:** A section titled "Answer Choices: Enter each choice on a separate line." with a text input field containing "Kadın" and "Erkek" on separate lines.

Şekil 17. Anket Hazırlama Panel Örneği a

Anketin 1. sorusu olan cinsiyet sorusunun cevabı olarak Answer Choices alanında "Kadın, Erkek" olarak iki seçenek sunulmuştur. Question Type kısmında ise bu sorunun birden fazla cevabı olabileceği fakat sadece bir tanesinin seçilebileceğini belirtmektedir. Pick a display format kısmında, cevapların 1 sütun üzerinde gözükeceğini belirtmektedir.

Örnek 2.

Question

Question Text:
Ortalama Aylık GSM Ödeme Tutarınız? (TL) (Faturalı/ Faturasız) Sadece rakamsal değer yazınız.

Insert response from... ?

Question Type:
Single Textbox

Require an answer to this question (Optional)
When the question is not answered, display this error message:
Bu soruya cevap vermeniz gerekmektedir.

Validate answer format (Optional)
Format: whole number between 0 and 10000
When the text is an invalid format, display this error message:
0-10000 arasında sayı giriniz.

Şekil 18. Anket Hazırlama Panel Örneği b

Anketin 9. Sorusu olan “Ortalama Aylık GSM Ödeme Tutarı” sorusu cevabının textbox kutucuğu içerisinde yazılması gerekmektedir. Cevap formatı rakam seçilip, bu rakam aralığının ise 0 – 10.000 TL aralığında olduğu belirtilmektedir. “Require an answer to this question” bölümü işaretlendiği zaman, anketi cevaplayan kişinin bu soruya muhakkak cevap yazması gerektiğini belirtilmektedir. Herhangi bir cevap yazılmaması halinde anket bitiş onayı istendiğinde “Bu soruya cevap vermeniz gerekmektedir” şeklinde hata mesajı vermektedir.

Örnek 3.

Question

Question Text:
Cep Telefonu cihazınız aşağıdaki belirtilen özelliklerden hangisine sahiptir?

Insert response from... ?

Question Type:
Matrix of Choices (Only One Answer Per Row) ▾

Answer Options

Row Choices: Enter each choice on a separate line.

1 İnternet
2 Görüntülü Arama
3 GPS
4 Karasal TV Alıcı
5 Bluetooth
6 Oyun

Column Choices: Enter the text for each column on a separate line.

Var kullanıyorum
Yok
Var ama kullanmıyorum

Şekil 19. Anket Hazırlama Panel Örneği c

Anketin 16. Sorusu olan “Cep telefonu cihazınız belirtilen özelliklerden hangisine sahiptir” sorusu için öncelikle Row Choices kısmında cep telefonunda var olabilecek özellikler belirtilmiştir. Bu özellikler, satırlar alt alta gelecek şekilde sıralanmıştır. Ayrıca her özelliğin yan tarafında mevcut olarak “Var kullanıyorum - Yok - Var ama kullanmıyorum“ seçeneklerinin işaretlenmesi gerekmektedir.

Örnek 4.

Question

Question Text:

"On sekizinci soruya "cevabınız " hayır" ise aşağıda belirtilen sebep/sebepleri kendi önceliğinize göre (1: en önemsiz - 10: en önemli'yi göstermek üzere) 1-10 arasında derecelendiriniz.

Insert response from... ?

Question Type:

Rating Scale

Answer Options

Add Row Labels (Optional)

Enter row labels for your question, each choice on a separate line:

1. Fiyat
2. Kapsama Ağı Kalitesi
3. Ses Kalitesi
4. Gizlilik İlkesine Bağlılık
5. Şeffaflık İlkesine Bağlılık
6. Teknik Destek
7. Tarife Avantajı

Rating Scale: 10 ratings Add N/A column

Şekil 20. Anket Hazırlama Panel Örneği d

Anketin 18. Sorusu olan “ Kullandığınız Operatörü Değiştirmeyi Düşünüyor musunuz?” sorusuna hayır cevabı veren kişilerin cevaplayabileceği bu soru anketin 19. Sorusudur. Bu anket sorusu derecelendirme ölçeği türünde bir soru olup 1 ile 10 arasında derecelendirme imkanı vermektedir. Operatörü değiştirmeme nedenleri ise alt alta gelecek şekilde sıralanarak, derecelendirme ölçeği ile her birinin anketi cevaplayan kişi için ne kadar önemli olup olmadığını belirtilmesi istenmektedir.

Örnek 5.

Question Text:
GSM hattınızla ilgili aşağıda yer alan yorumları ' 1: kesinlikle katılmıyorum' ve ' 5: kesinlikle katılıyorum' arasında değerlendiriniz.

[Insert response from...](#) ?

Question Type:
Matrix of Choices (Only One Answer Per Row) ▼

Answer Options

Row Choices: Enter each choice on a separate line.
Kullandığım GSM Operatörünü Tavsiye ediyorum
Gelecekte de aynı GSM Operatörüyle devam etmek istiyorum.
Kullandığım GSM Operatörü Sadakatime Layıktır.

Column Choices: Enter the text for each column on a separate line.
Kesinlikle katılmıyorum
Katılmıyorum
Ne katılıyorum ne katılmıyorum(fikrim yok)
Katılıyorum
Kesinlikle katılıyorum

Şekil 21. Anket Hazırlama Panel Örneği e

Anketin 21. ve son sorusu olan “GSM hattınızla ilgili yer alan yorumları kesinlikle katılmıyorum – kesinlikle katılıyorum arasında değerlendiriniz.” sorusu içinde 3 farklı GSM cümlesinin değerlendirilmesi istenmektedir. Bunlar : “ Kullandığım GSM Operatörünü Tavsiye Ediyorum- Gelecekte de aynı GSM Operatörüyle devam etmek istiyorum. – Kullandığım GSM Operatörü Sadakatime Layıktır.” cümlelerinin her birini “ kesinlikle katılıyorum – kesinlikle katılmıyorum“ aralığında değerlendirilmesi gerekmektedir. Bu sorunun türü tercih matrisidir ve sadece her bir satır için bir cevap seçilmesine izin verilmektedir.

5.4 ONLINE ANKET: GSM MÜŞTERİ KAYBI ANALİZİ

Anketin amacı, farklı yaş ve cinsiyetlerdeki insanların GSM alışkanlıklarının yanı sıra müşteri memnuniyetini sağlayan veya sağlayamayan unsurları belirlemektir. Anket sonucunda elde edilen veriler, veri madenciliği amacıyla günümüzde yaygın olarak kullanılmakta olan WEKA programı ve algoritmaları ile analiz edilerek müşteri kaybının nelere bağlı olduğu ve nelerle ilişkili olduğu saptanmıştır.

Anket online ortamda hazırlanarak kişilerin istediği zaman aralığında kendilerini herhangi bir zorunda olma duygusu hissettirilmeden ankete tutarlı cevap vermeleri sağlanmaktadır. Kullanılan anket sisteminde kişilerin aynı IP' den girilmesi engellenerek anket tutarlılığı artırılmıştır.

Anket soruları çoktan seçmeli, açık uçlu ve likert ölçek tipli sorulardan oluşmaktadır. Uygulanan anket 21 adet sorudan oluşmakta ve kişilerden 20 adet soruya internet üzerinden cevap verilmeleri istenmektedir.

Anket verilerinin toplanması 6 ay sürüp, gereksiz veriler temizlendikten sonra 615 kişinin verisi istatistiğe alınmıştır.

Anketin hedef kitlesi olarak üniversite öğrencileri ve genel olarak öğrenci çevresinde bulunan aile, arkadaş grubu alınmıştır. Anket sorularının içerisinde kimlik bilgilerini taşıyan sorular sorulmamıştır.

Anket soruları GÜLPINAR ve ALTAŞ'ın (2013) çalışmasından faydalanarak oluşturulmuştur.

5.4.1 Online Anket Soruları

WEB-TABANLI ANKET SİSTEMİ İLE ELDE EDİLEN VERİLERİN VERİ MADENCİLİĞİ YÖNTEMİ İLE ANALİZİ

Bu araştırma; Web-tabanlı anket sistemi ile elde edilen verilerin veri madenciliği yöntemi ile analizini amaçlamaktadır. Anket 21 adet çoktan seçmeli sorudan oluşmaktadır. Araştırmada hiçbir kişisel bilgi alınmamaktadır. Ankete vereceğiniz içten ve doğru yanıtlar araştırmanın doğruluğu açısından önemlidir. Katkılarınız için şimdiden teşekkür ederiz.

Sevgi ve Saygılarımızla

Yrd.Doç.Dr. Metin ZONTUL
İstanbul Aydın Üniversitesi

Öğr. Gör. Zeynep YURTTAŞ ELABİAD
İstanbul Gelişim Üniversitesi

* 1. Cinsiyetiniz?

Kadın Erkek

*2. Yaşınız?

15 Altı 15-24 25-34 35-44 45-54 55 üstü

***3. Medeni Durumunuz?**

- Evli Bekar Dul Boşanmış

***4. Eğitim Durumunuz?**

- İlköğretim Lise Önlisans Lisans Y.Lisans Doktora

***5. Ortalama Aylık Gelirini? (TL)**

- 1.000'nin altında 1.001-2.000 2.001-3.000 3.001-4.000 4.000 üstünde

***6. Kullanmakta olduğunuz GSM hattına(operatörünüze) sahip olma süreniz?**

- 1 yıldan az 1-3 yıl 3-5 yıl 6 yıldan fazla

***7. GSM Abonelik Türünüz nedir?**

- Faturalı Faturasız

***8. Farklı bir operatörden ilgili operatöre numaranızı taşıdınız mı?**

- Evet Hayır

***9. Ortalama Aylık GSM Ödeme Tutarınız? (TL) (Faturalı/ Faturasız)
Sadece rakamsal değer yazınız.*****10. GSM üzerinden Ortalama Aylık Konuşma Süreniz? (dakika). Sadece rakamsal değer yazınız.****11. GSM üzerinden Ortalama Aylık SMS/MMS Sayınız?**

***12. Hangi GSM Operatörünü kullanıyorsunuz?**

- Avea Vodafone Turkcell

13. Turkcell GSM operatörü kullanıyorsanız tarifenizi seçiniz

- Platinum Gencturkcell Turkcell Kamu Profesyoneller
 Diğer

14. Vodafone GSM operatörü kullanıcısı iseniz tarifenizi seçiniz.

- Freezone Cepavantaj Cepkamu Cepogrenci
 Diğer

15. Avea GSM operatörü kullanıyorsanız tarifenizi seçiniz.

- Fırsatbizim Heryonbizim Bizimofis Prestijbizde
 Diğer

***16. Cep Telefonu cihazınız aşağıdaki belirtilen özelliklerden hangisine sahiptir?**

	Var kullanıyorum	Yok	Var ama kullanmıyorum
1 İnternet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2 Görüntülü Arama	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3 GPS	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4 Karasal TV Alıcı	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5 Bluetooth	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6 Oyun	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

20. " On sekizinci soruya cevabınız "evet" ise aşağıda belirtilen sebep/sebepleri kendi önceliğinize göre 1-10 arasında derecelendiriniz

	(1)(2)	(3)(4)	(5)(6)	(7)(8)	(9)	(10)
1. Fiyat	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Kapsama Ağı Kalitesi	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Ses Kalitesi	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Gizlilik İlkesine Bağlılık	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Şeffaflık İlkesine Bağlılık	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Teknik Destek	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Extra Maliyetler	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Marka İmajı	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Teknolojik Yetersizlik	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10.Güven	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11.Hizmet Çeşitlerinin Yetersizliği	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. Hatalı Faturaların/İşlemlerin Sıklığı	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. Yetersiz Promosyonlar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14. Aile ve Arkadaş Çevresinin Farklı GSM op. Kullanması	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15. Diğer GSM operatörlerinden daha cazip olması	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

***21. GSM hattınızla ilgili aşağıda yer alan yorumları ' 1: kesinlikle katılmıyorum' ve ' 5: kesinlikle katılıyorum' arasında değerlendiriniz.**

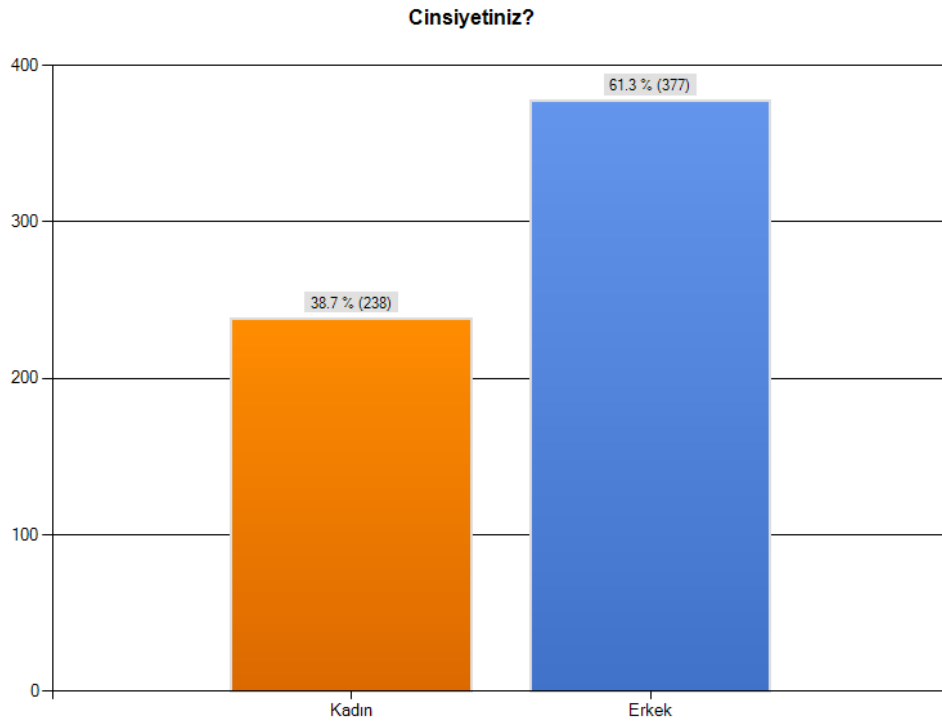
		Kesinlikle katılmıyorum	Katılmıyorum	Fikrim yok	Katılıyorum	Kesinlikle katılıyorum
Kullandığım Operatörünü tavsiye ediyorum	GSM Tavsiye	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Gelecekte de aynı GSM Operatörüyle devam etmek istiyorum.	GSM devam	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Kullandığım Operatörü Sadakatime Layıkır.	GSM Sadakatime	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

5.5 ANKET İSTATİSTİKLERİ

1) Anket katılımcılarının cinsiyetleri, cevaplama yüzdesi ve cevaplayan sayısı Tablo 9'da gösterilmiştir. Şekil 22'de anket katılımcılarının cinsiyet oranı, istatistiksel grafik halinde gösterilmiştir.

Tablo 9. Anket Katılımcılarının Cinsiyet Oran Tablosu

Cinsiyetiniz?		
Cevap Seçenekleri	Cevaplama Yüzdesi	Cevaplayan Sayısı
Kadın	38,7%	238
Erkek	61,3%	377
Toplam cevaplayan sayısı		615

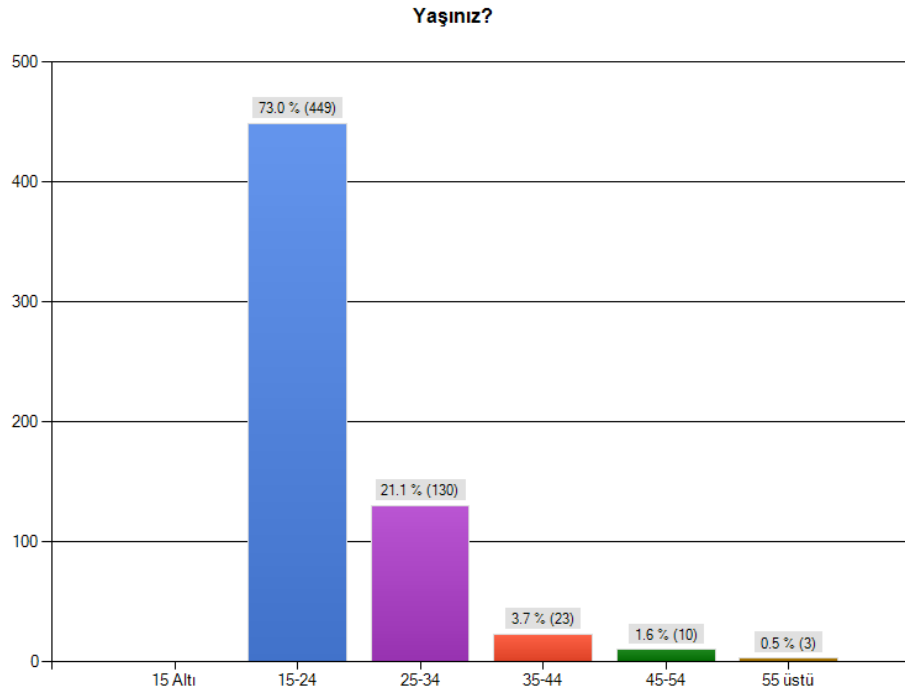


Şekil 22. Anket Katılımcılarının Cinsiyet Oranı

2) Anket katılımcılarının yaş aralık oran tablosu Tablo 10'da gösterilmiştir. Şekil 23'de anket katılımcılarının yaş aralık oranı, istatistiksel grafik halinde gösterilmiştir.

Tablo 10. Anket Katılımcılarının Yaş Aralık Oran Tablosu

Yaşınız?		
Cevap Seçenekleri	Cevaplama Yüzdesi	Cevaplayan Sayısı
15 Altı	0,0%	0
15-24	73,0%	449
25-34	21,1%	130
35-44	3,7%	23
45-54	1,6%	10
55 üstü	0,5%	3
Toplam Cevaplayan Sayısı		615

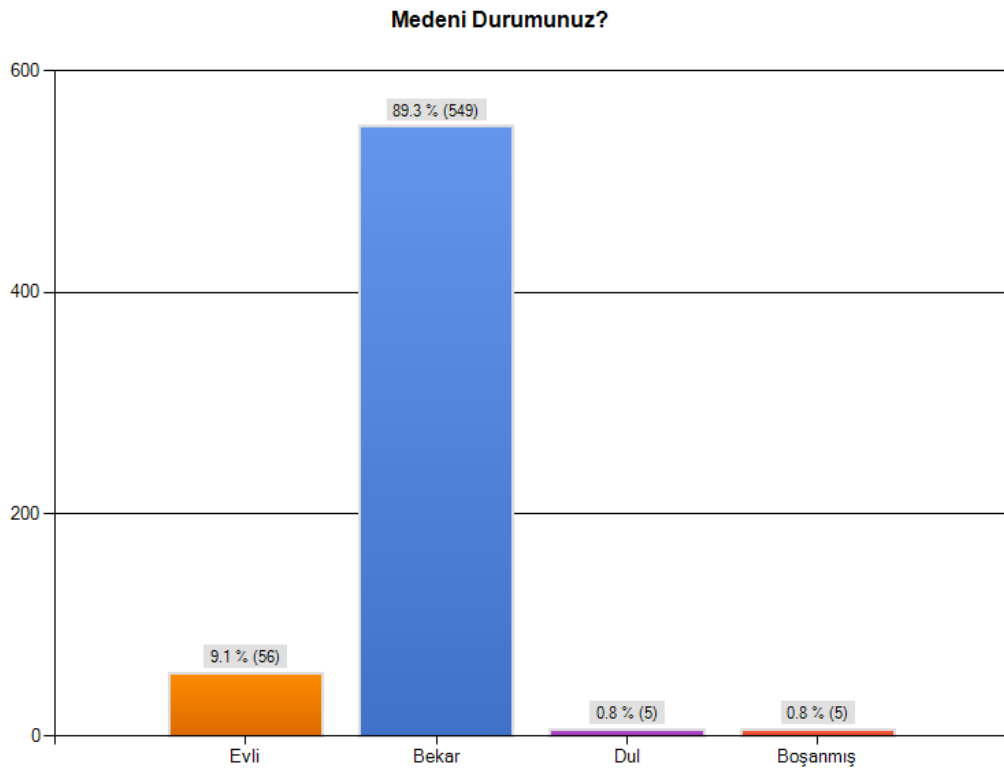


Şekil 23. Anket Katılımcılarının Yaş Aralık Oranı

3) Anket katılımcılarının medeni durumunun oran verileri Tablo 11’de gösterilmiştir. Şekil 24’de anket katılımcılarının medeni durum oranı, istatistiksel grafik halinde gösterilmiştir.

Tablo 11. Anket Katılımcılarının Medeniyet Durum Oran Tablosu

Medeni Durumunuz?		
Cevap Seçenekleri	Cevaplama Yüzdesi	Cevaplayan Sayısı
Evli	9,1%	56
Bekar	89,3%	549
Dul	0,8%	5
Boşanmış	0,8%	5
Toplam Cevaplayan Sayısı		615

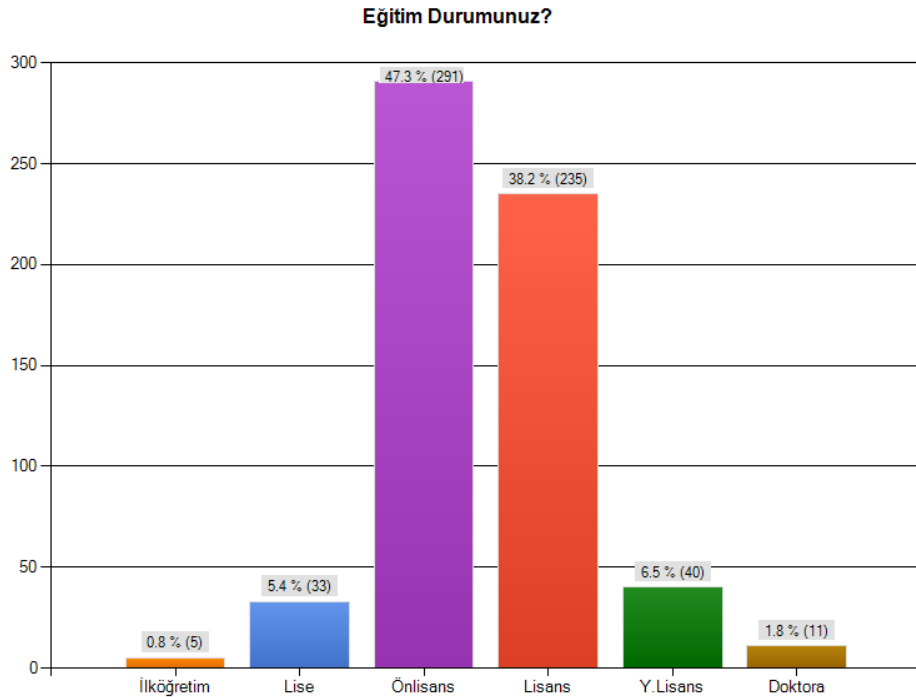


Şekil 24. Anket Katılımcılarının Medeniyet Durum Oranı

4) Anket katılımcılarının mezuniyet oran tablosu Tablo 12'de gösterilmiştir. Şekil 25'de anket katılımcılarının mezuniyet oran verileri, istatistiksel grafik halinde gösterilmiştir.

Tablo 12. Anket Katılımcılarının Mezuniyet Oran Tablosu

Eğitim Durumunuz?		
Cevap Seçenekleri	Cevaplama Yüzdesi	Cevaplayan Sayısı
İlköğretim	0,8%	5
Lise	5,4%	33
Önlisans	47,3%	291
Lisans	38,2%	235
Y.Lisans	6,5%	40
Doktora	1,8%	11
Toplam Cevaplayan Sayısı		615

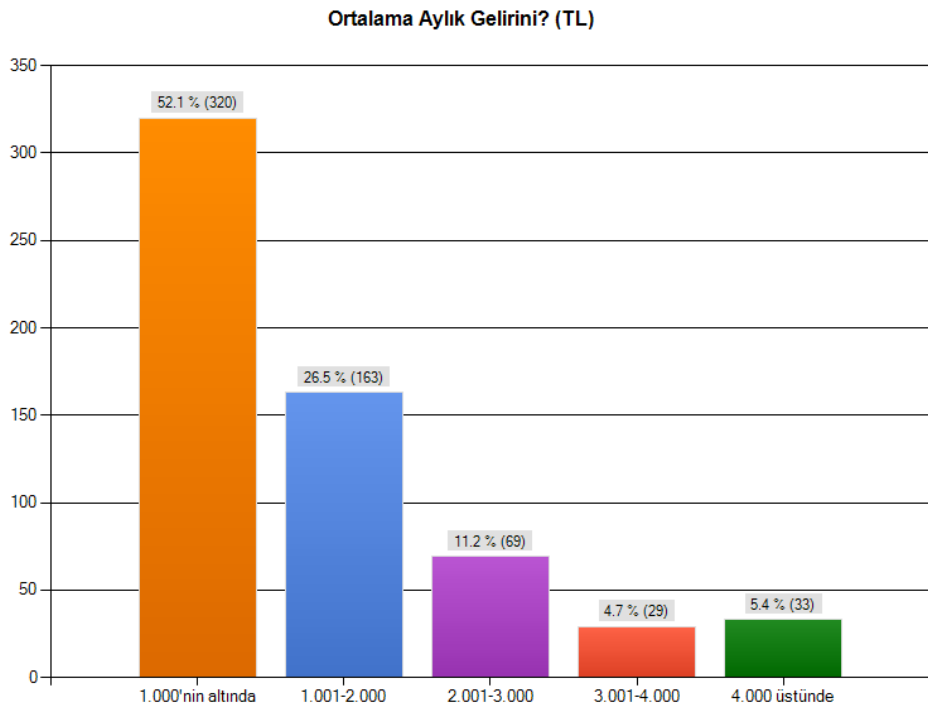


Şekil 25. Anket Katılımcılarının Mezuniyet Oranı

5) Anket katılımcılarının aylık gelir oranı Tablo 13'de gösterilmiştir. Şekil 26'da anket katılımcılarının aylık gelir oranı, istatistiksel grafik halinde gösterilmiştir.

Tablo 13. Anket Katılımcılarının Aylık Gelir Oran Tablosu

Ortalama Aylık Gelirini? (TL)		
Cevap Seçenekleri	Cevaplama Yüzdesi	Cevaplayan Sayısı
1.000'nin altında	52,1%	321
1.001-2.000	26,5%	163
2.001-3.000	11,2%	69
3.001-4.000	4,7%	29
4.000 üstünde	5,4%	33
Toplam Cevaplayan Sayısı		615

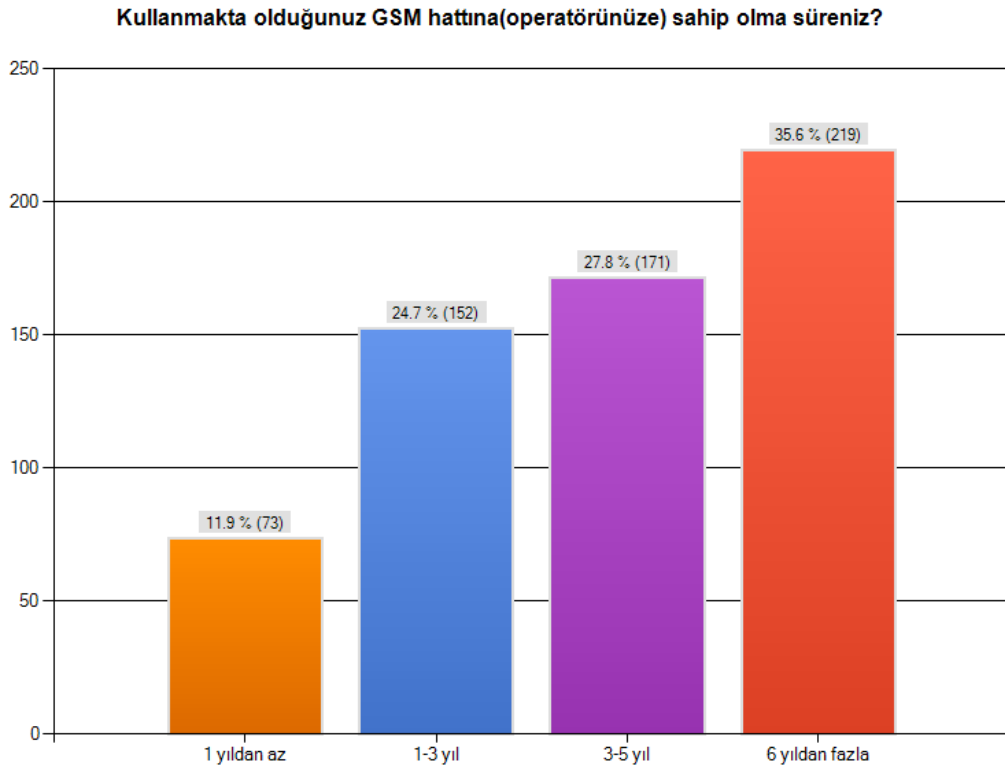


Şekil 26. Anket Katılımcılarının Aylık Gelir Oranı

6) Anket katılımcılarının GSM hat operatörüne sahip olma süresi Tablo 14’de gösterilmiştir. Şekil 27’de anket katılımcılarının GSM hat operatörüne sahip olma süresi, istatistiksel grafik halinde gösterilmiştir.

Tablo 14. Anket Katılımcılarının GSM Operatörüne Sahip Olma Süresi Oran Tablosu

Kullanmakta olduğunuz GSM hattına(operatörünüze) sahip olma süreniz?		
Cevap Seçenekleri	Cevaplama Yüzdesi	Cevaplayan Sayısı
1 yıldan az	11,9%	73
1-3 yıl	24,7%	152
3-5 yıl	27,8%	171
6 yıldan fazla	35,6%	219
Toplam Cevaplayan Sayısı		615

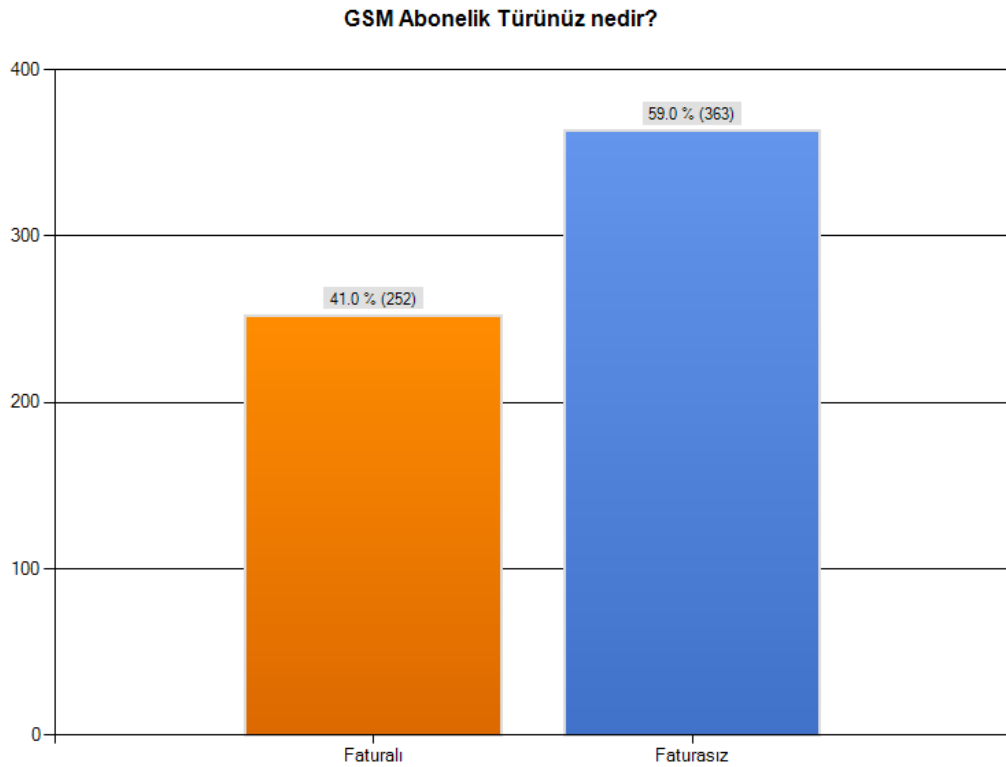


Şekil 27. Anket Katılımcılarının GSM Operatörüne Sahip Olma Süresi Oranı

7) Anket katılımcılarının GSM abonelik türü oran tablosu Tablo 15’de gösterilmiştir. Şekil 28’de anket katılımcılarının GSM abonelik tür verileri istatistiksel grafik halinde gösterilmiştir.

Tablo 15. Anket Katılımcılarının GSM Abonelik Türleri Oran Tablosu

GSM Abonelik Türünüz nedir?		
Cevaplama Seçenekleri	Cevaplama Yüzdesi	Cevaplayan Sayısı
Faturalı	41,0%	252
Faturasız	59,0%	363
Toplam Cevaplayan Sayısı		615

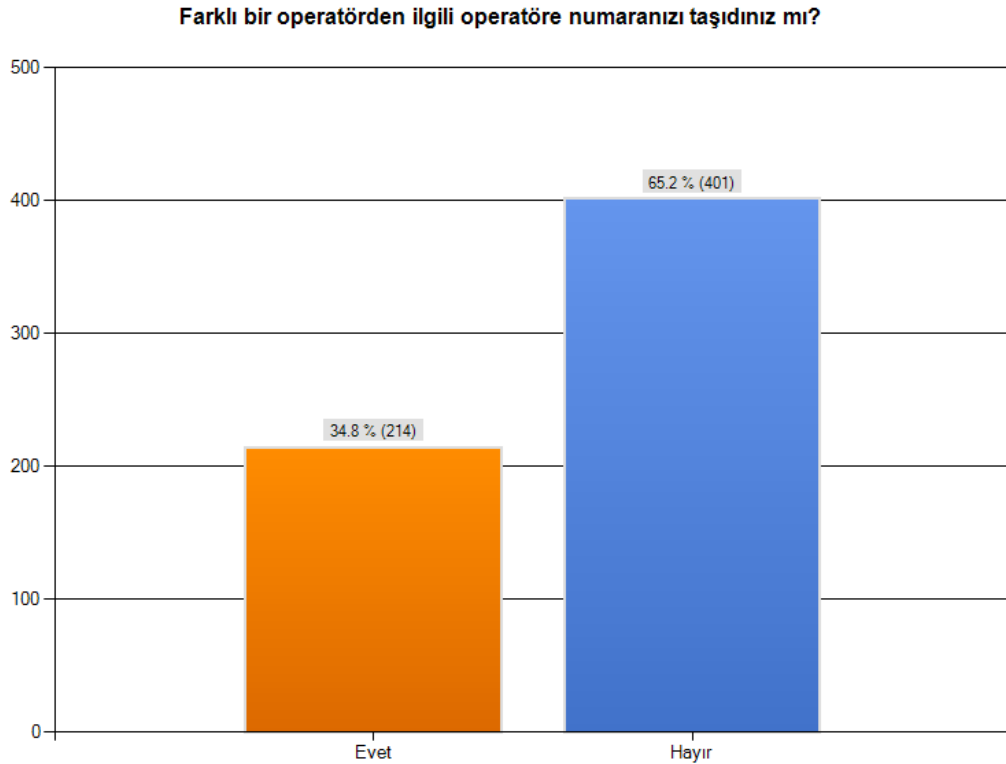


Şekil 28. Anket Katılımcılarının GSM Abonelik Türleri Oranı

8) Anket katılımcılarının GSM numaralarını başka bir operatöre taşıma oranı Tablo 16'da gösterilmiştir. Şekil 29'da anket katılımcılarının GSM numaralarını başka bir operatöre taşıma/taşıyamama oranı istatistiksel grafik halinde gösterilmiştir.

Tablo 16. Anket Katılımcılarının Numara Taşıma/Taşıyamama Oran Tablosu

Farklı bir operatörden ilgili operatöre numaranızı taşıdınız mı?		
Cevaplama Seçenekleri	Cevaplama Yüzdesi	Cevaplayan Sayısı
Evet	34,8%	214
Hayır	65,2%	401
Toplam Cevaplayan Sayısı		615

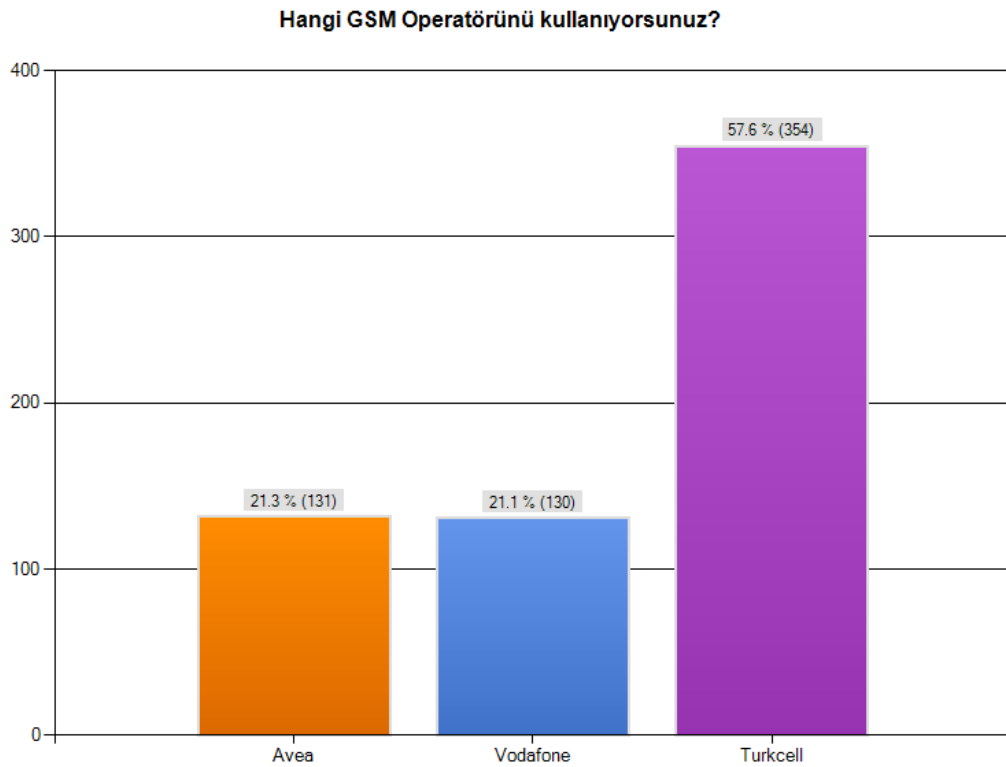


Şekil 29. Anket Katılımcılarının Numara Taşıma/Taşıyamama Oranı

12) Anket katılımcılarının GSM operatör türü oran tablosu Tablo 17’de gösterilmiştir. Şekil 30’da anket katılımcılarının GSM operatör türü verileri istatistiksel grafik halinde gösterilmiştir.

Tablo 17. Anket Katılımcılarının GSM Operatör Türü Oran Tablosu

Hangi GSM Operatörünü kullanıyorsunuz?		
Cevaplama Seçenekleri	Cevaplama Yüzdesi	Cevaplayan Sayısı
Avea	21,3%	131
Vodafone	21,1%	130
Turkcell	57,6%	354
Toplayan Cevaplama Sayısı		615

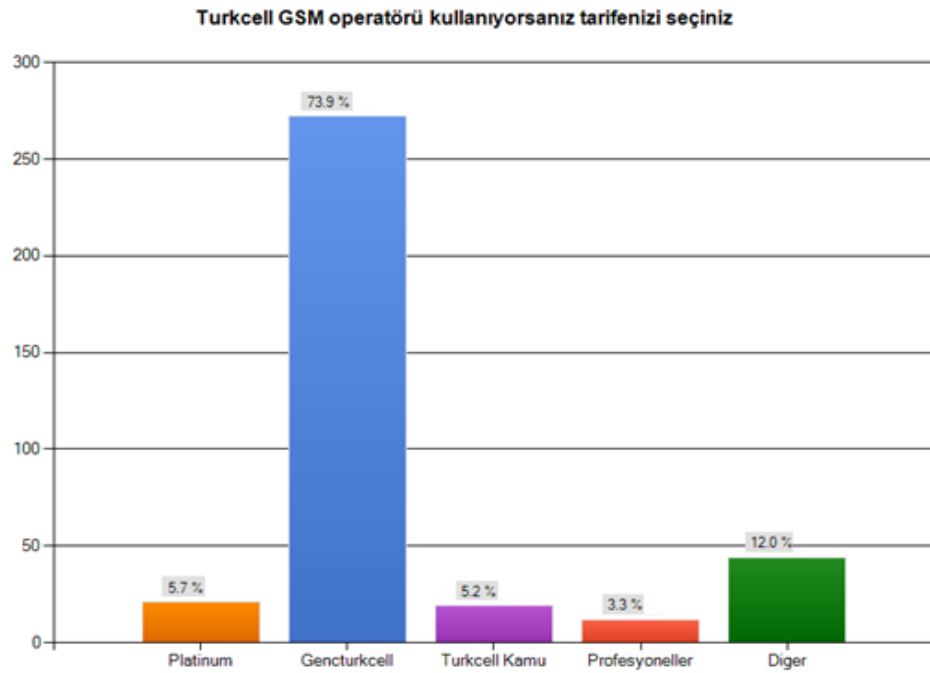


Şekil 30. Anket Katılımcılarının GSM Operatör Türü Oranı

13) Turkcell GSM operatörü kullanan anket katılımcılarının tarife türleri ve oranlarına ait tablo Tablo 18'de gösterilmiştir. Şekil 31'de Turkcell GSM operatörü kullanan anket katılımcılarının tarife türü verileri, istatistiksel grafik halinde gösterilmiştir.

Tablo 18. Turkcell GSM Operatörü Tarife Türü Oran Tablosu

Turkcell GSM operatörü kullanıyorsanız tarifenizi seçiniz	
Cevap Seçenekleri	Cevaplayan Yüzdesi
Platinum	5,7%
Gencturkcell	73,9%
Turkcell Kamu	5,2%
Profesyoneller	3,3%
Diger	12,0%

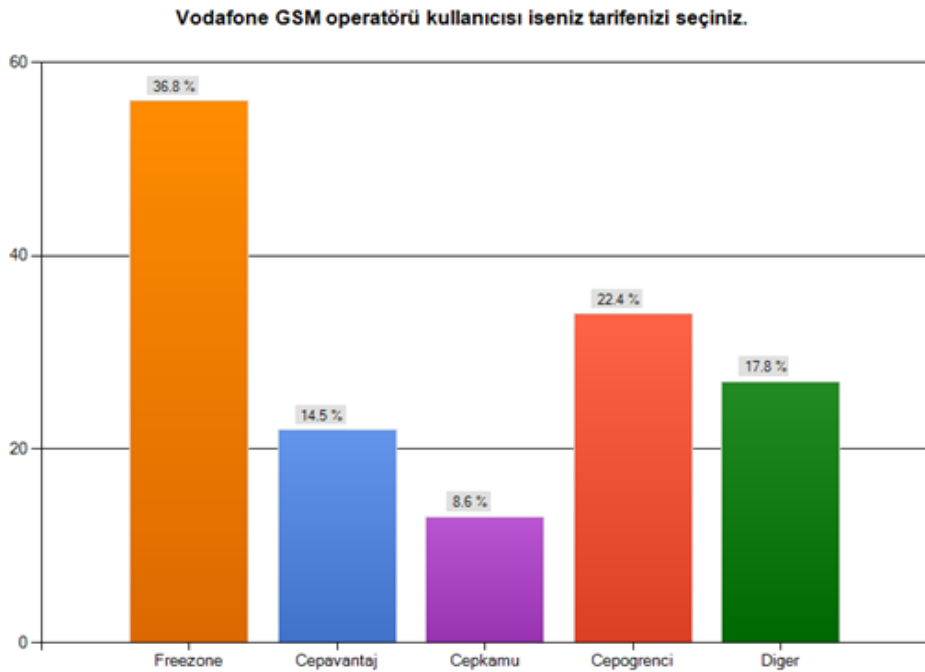


Şekil 31. Turkcell GSM Operatörü Tarife Türü Oranı

14) Vodafone GSM operatörü kullanan anket katılımcılarının tarife türleri ve oranlarına ait tablo Tablo 19'da gösterilmiştir. Şekil 32'de Vodafone GSM operatörü kullanan anket katılımcılarının tarife türü verileri, istatistiksel grafik halinde gösterilmiştir.

Tablo 19. Vodafone GSM Operatörü Tarife Türü Oran Tablosu

Vodafone GSM operatörü kullanıcısı iseniz tarifenizi seçiniz.	
Cevap Seçenekleri	Cevaplayan Yüzdesi
Freezone	36,8%
Cepavantaj	14,5%
Cepkamu	8,6%
Cepogrenci	22,4%
Diger	17,8%

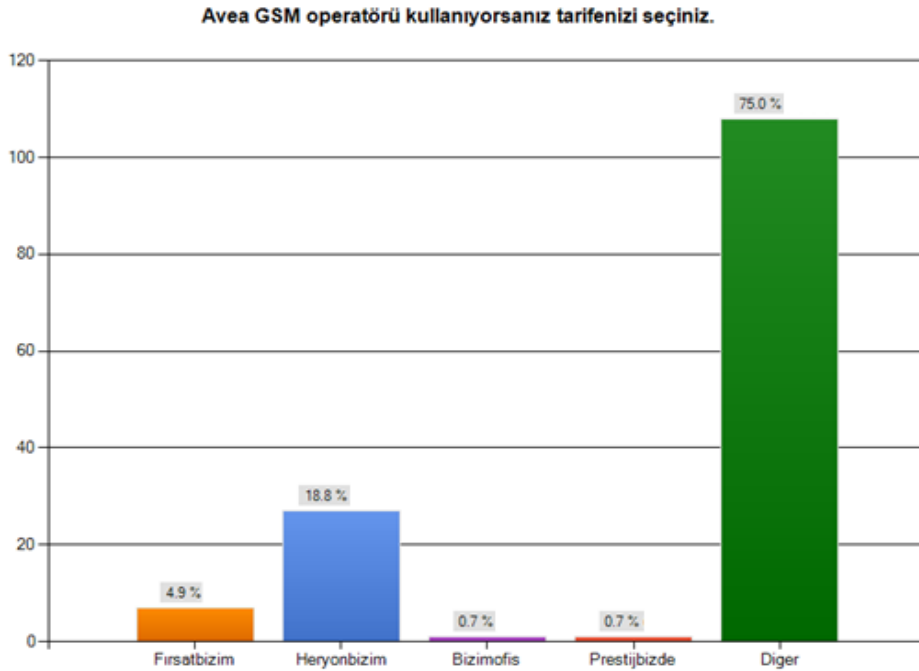


Şekil 32. Vodafone GSM Operatörü Tarife Türü Oranı

15) Avea GSM operatörü kullanan anket katılımcılarının tarife türleri ve oranlarına ait tablo Tablo 20'de gösterilmiştir. Şekil 33'de Avea GSM operatörü kullanan anket katılımcılarının tarife türü verileri, istatistiksel grafik halinde gösterilmiştir.

Tablo 20. Avea GSM Operatörü Tarife Türü Oran Tablosu

Avea GSM operatörü kullanıyorsanız tarifenizi seçiniz.	
Cevap Seçenekleri	Cevaplayan Yüzdesi
Fırsatbizim	4,9%
Heryonbizim	18,8%
Bizimofis	0,7%
Prestijbizde	0,7%
Diger	75,0%



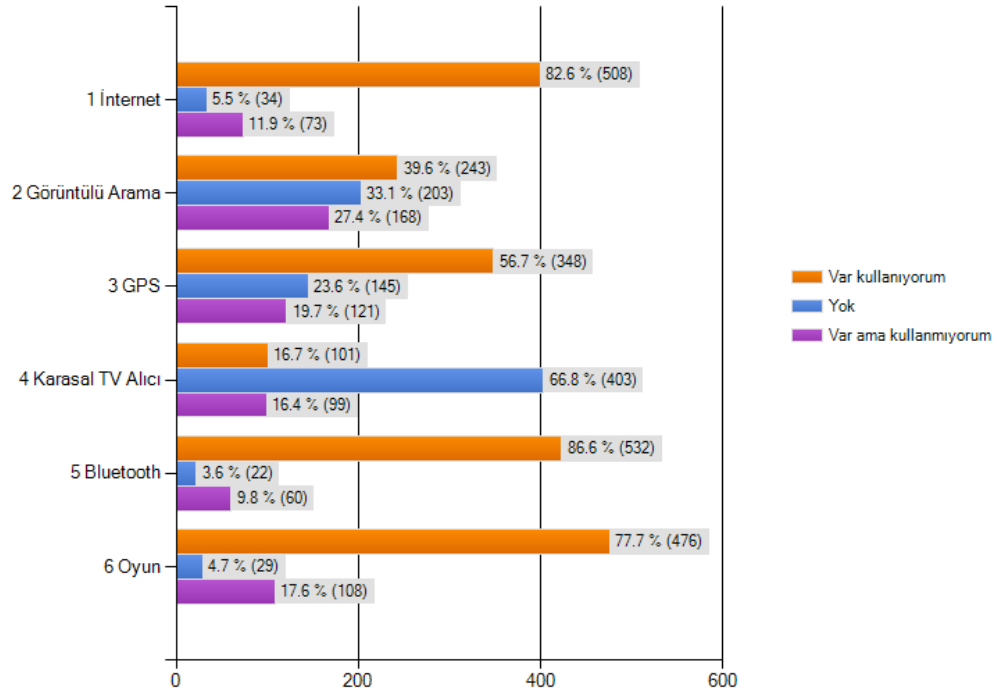
Şekil 33. Avea GSM Operatörü Tarife Türü Oranı

16) Anket katılımcılarının cep telefonu cihazlarında belirtilen özelliklere sahip olma oranı Tablo 21’de gösterilmiştir. Şekil 34’de anket katılımcılarının cep telefonu cihazlarında belirtilen özelliklerine sahip olma oran verileri, istatistiksel grafik halinde gösterilmiştir.

Tablo 21. Anket Katılımcılarının Cep Telefonu Özelliklerine Sahip Olma Oranı Tablosu

Cep Telefonu cihazınız aşağıdaki belirtilen özelliklerden hangisine sahiptir?				
Cevap Seçenekleri	Var kullanıyorum	Yok	Var ama kullanmıyorum	Cevaplayan Sayısı
1 İnternet	508	34	73	615
2 Görüntülü Arama	243	203	168	614
3 GPS	348	145	121	614
4 Karasal TV Alıcı	101	403	99	603
5 Bluetooth	532	22	60	614
6 Oyun	476	29	108	613
Toplam Cevap Sayısı				615

Cep Telefonu cihazınız aşağıdaki belirtilen özelliklerden hangisine sahiptir?

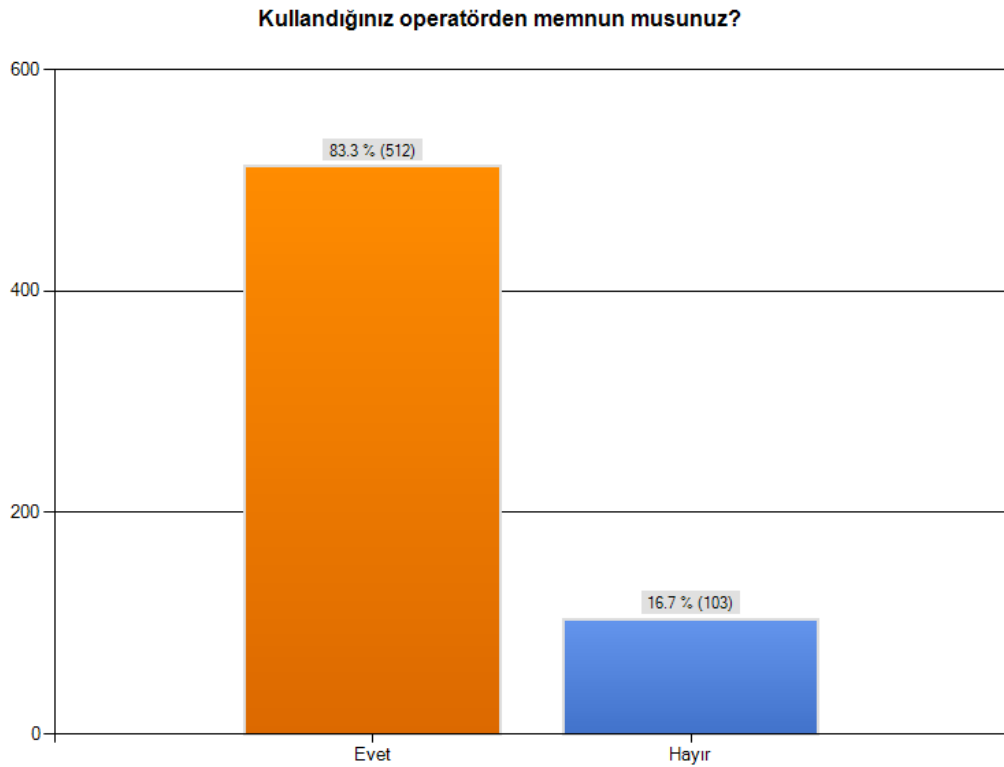


Şekil 34. Anket Katılımcılarının Cep Telefonu Özelliklerine Sahip Olma Oranı

17) Anket katılımcılarının kullandıkları operatörlerinden memnun olma/olmama durum verileri Tablo 22'de gösterilmiştir. Şekil 35'de anket katılımcılarının kullandıkları operatörlerinden memnun olma/olmama durum verileri istatistiksel grafik olarak gösterilmiştir.

Tablo 22. Anket Katılımcılarının Operatör Türüne Göre Memnuniyet Durum Tablosu

Kullandığınız operatörden memnun musunuz?		
Cevap Seçenekleri	Cevaplama Yüzdesi	Cevaplayan Sayısı
Evet	83,3%	512
Hayır	16,7%	103
Toplam Cevaplayan Sayısı		615

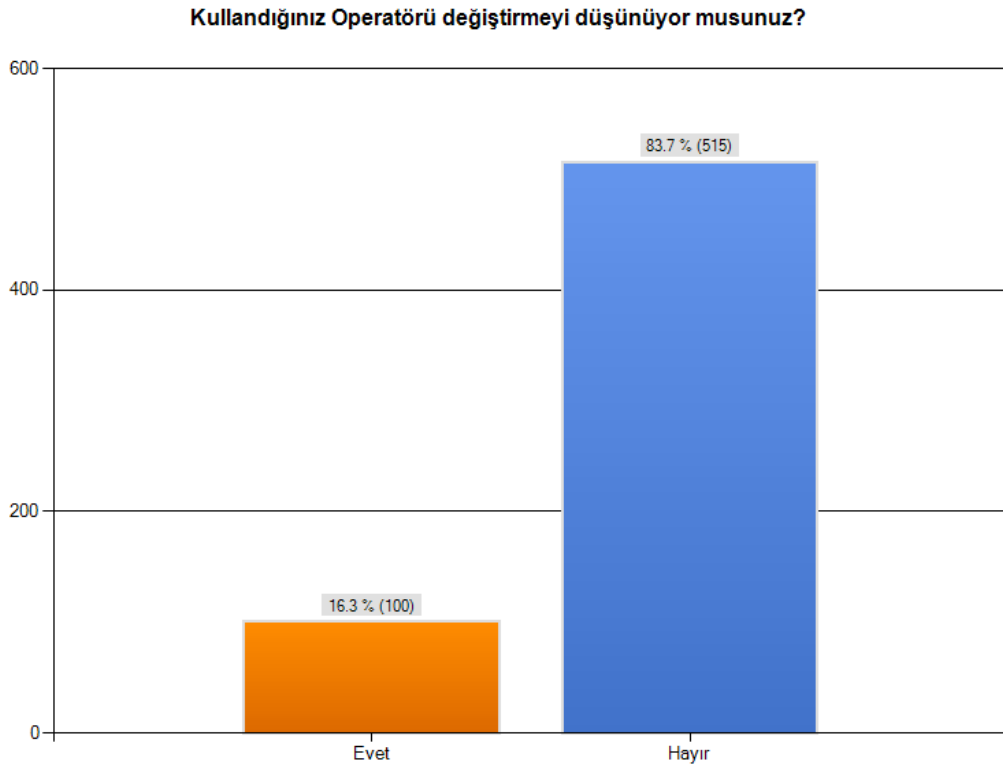


Şekil 35. Anket Katılımcılarının Operatör Türüne Göre Memnuniyet Durumu

18) Anket katılımcılarının GSM operatörünü deęiřtirme istek tablosu Tablo 23'de gösterilmiřtir. Őekil 36'da anket katılımcılarının GSM operatörünü deęiřtirme isteęi istatistiksel grafik olarak gösterilmiřtir.

Tablo 23. Anket Katılımcılarının Operatör Türünü Deęiřtirme İstek Tablosu

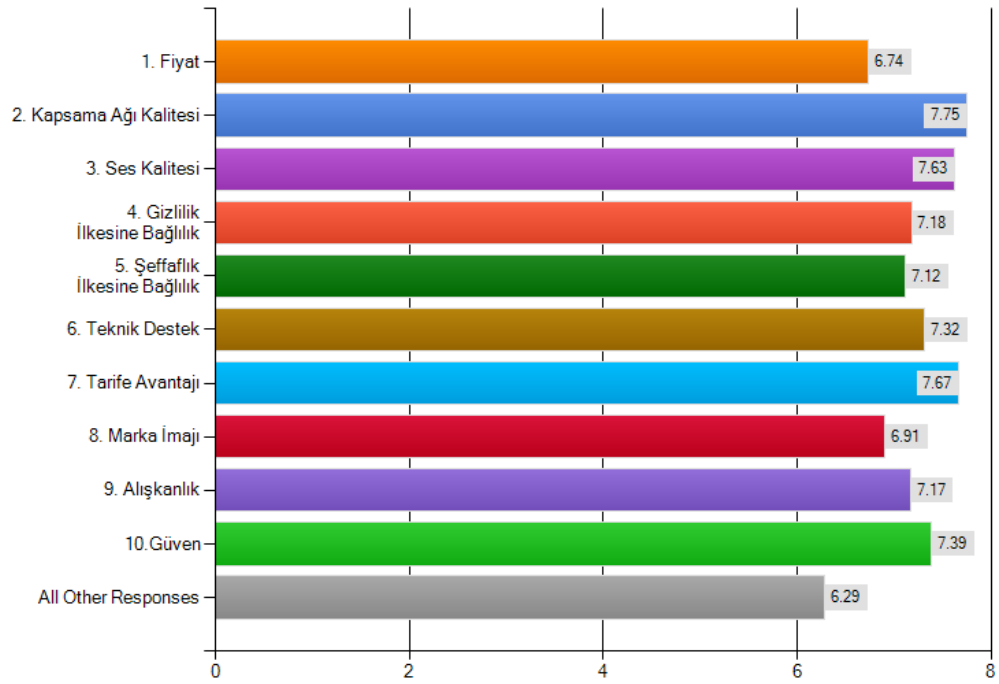
Kullandığınız Operatörü deęiřtirmeyi düşünüyor musunuz?		
Cevap Seçenekleri	Cevaplama Yüzdesi	Cevaplayan Sayısı
Evet	16,3%	100
Hayır	83,7%	515
Toplam Cevaplayan Sayısı		615



Őekil 36. Anket Katılımcılarının Operatör Türünü Deęiřtirme İsteęi

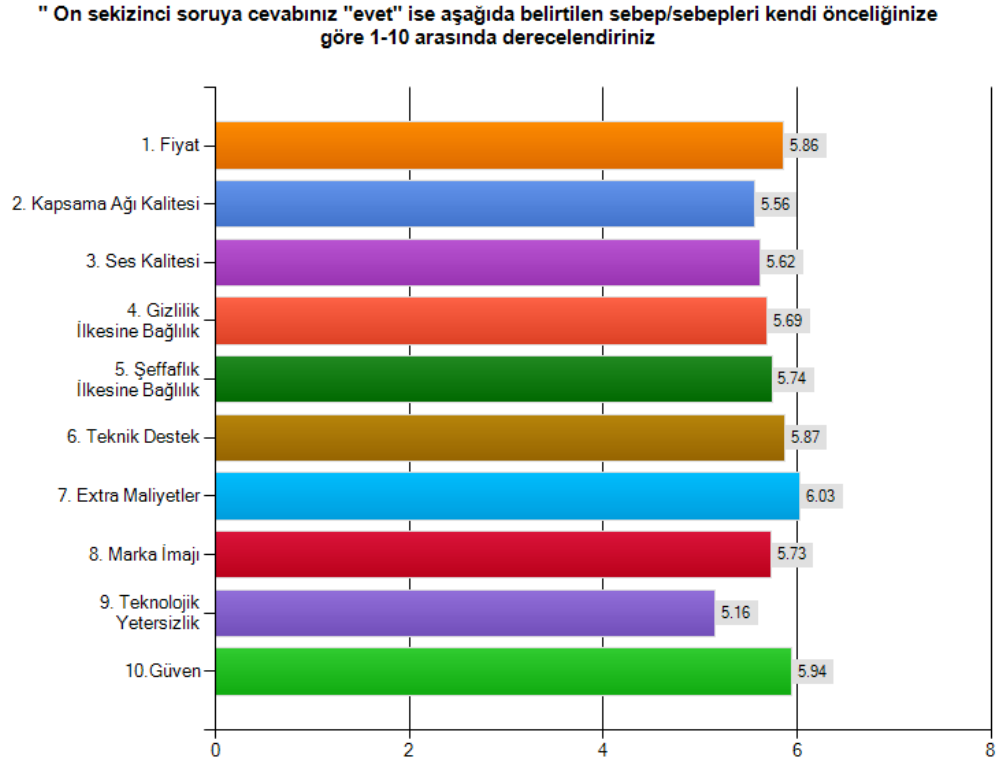
19) GSM operatörünü değiştirmek istemeyen anket katılımcılarının değiştirmeme isteği nedenleri ve oranları Şekil 37’de gösterilmiştir.

"On sekizinci soruya "cevabınız " hayır" ise aşağıda belirtilen sebep/sebepleri kendi önceliğinize göre (1: en önemsiz - 10: en önemli'yi göstermek üzere) 1-10 arasında derecelendirisiniz.



Şekil 37. Anket Katılımcılarının Operatörlerinden Memnun Olma Derecesi

20) GSM operatörünü değiştirmek isteyen anket katılımcılarının değiştirme isteği nedenleri ve oranları Şekil 38'de gösterilmiştir.



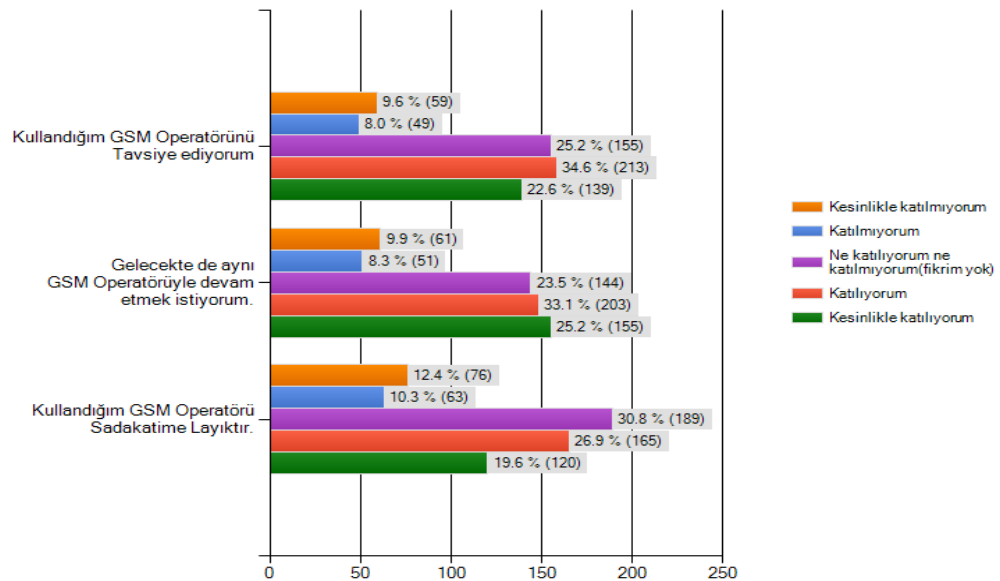
Şekil 38. Anket Katılımcılarının Operatörlerinden Memnun Olmama Derecesi

21) Anket katılımcılarının kullandıkları GSM hattı ile ilgili yorumları ve ilgili oranları Tablo 24'de gösterilmiştir. Şekil 39'da anket katılımcılarının kullandıkları GSM hattı ile ilgili yorumları ve ilgili oranları istatistiksel grafik olarak gösterilmiştir.

Tablo 24. Anket Katılımcılarının GSM Hattı Yorum Oranı Tablosu

GSM hattınızla ilgili aşağıda yer alan yorumları ' 1: kesinlikle katılmıyorum' ve ' 5: kesinlikle katılıyorum' arasında değerlendiriniz.						
Cevap Seçenekleri	Kesinlikle katılmıyorum	Katılmıyorum	Ne katılıyorum ne katılmıyorum (fikrim yok)	Katılıyorum	Kesinlikle katılıyorum	Cevaplayan Sayısı
Kullandığım GSM Operatörünü Tavsiye ediyorum	59	49	155	213	139	615
Gelecekte de aynı GSM Operatörüyle devam etmek istiyorum.	61	51	144	203	155	614
Kullandığım GSM Operatörü Sadakatime Layıktır.	76	63	189	165	120	613
Toplam Cevaplayan Sayısı						615

GSM hattınızla ilgili aşağıda yer alan yorumları ' 1: kesinlikle katılmıyorum' ve ' 5: kesinlikle katılıyorum' arasında değerlendiriniz.



Şekil 39. Anket Katılımcılarının GSM Hattı Yorum Oranı

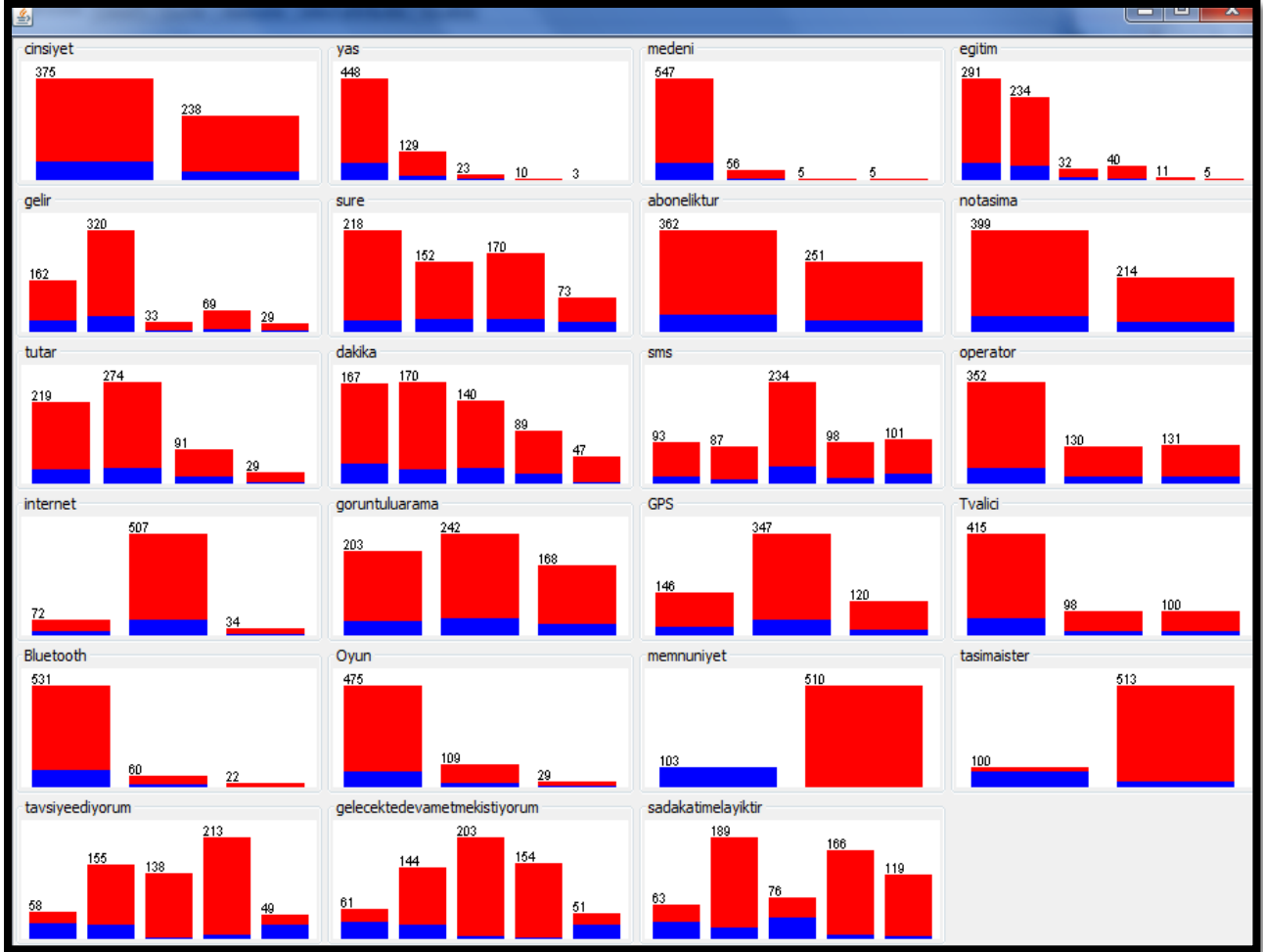
6. WEKA UYGULAMASI

23 öznitelikten oluşan ve 613 online anket katılımcısından edinilen veriler, veri madenciliği programı olan WEKA ile analiz edilmiştir. Analiz sırasında karar ağacı algoritması olan j48 ve birliktelik kuralları çıkarımı için apriori algoritması kullanılmıştır.

Web tabanlı anket sistemiyle elde edilen ve excel dosyasında tutulan veriler, veri madenciliği ile analiz edilmeden önce eksik ve hatalı verilerden temizlenmiştir. Özellikle anket soruları içerisinde yer alan GSM türüne göre paket isimleri sorusu açısından paketlerin sürekli güncellenmesi, çok sayıda paket olması, aynı anda farklı paketlere sahip olabilmeleri ve gsm hat sahiplerinin kullandıkları paket isimlerini bilmemeleri ya da eksik bilmelerinden dolayı WEKA programı ile uygulama esnasında GSM paket türü verileri anketin tutarlılığını korumak açısından çıkarılmıştır.

Temizlenen veriler .csv uzantısına çevirilerek WEKA programının anlayacağı hale getirilmiştir. WEKA programı karşımıza geldiğinde veri madenciliği algoritmalarının kullanılabilmesi Explorer modülü seçilmelidir. Preprocess paneli, veriler üzerinde filtre uygulanabilen, niteliklerin seçilebildiği, verilerin düzenlenebildiği ve nitelik istatistiklerinin görülebildiği Explorer modülü sekmesidir. Open file butonuyla açılan .csv ve diğer uzantılı dosyalar, istenildiğinde Save butonu ile .arff, csv, c4.5 vb. uzantılı dosya türlerine dönüştürülebilir.

Verilerin olduğu dosya .arff dosyası uzantısına çevrildikten sonra, her özneliğin cevapları ve ağırlıkları görülebilmektedir. Ayrıca istenilen özneliğin istatistikleri veya tüm özneliklerin nitelikleri tek bir pencerede gözükebilmektedir. Veri madenciliği analiz sırasında istenilmeyen öznelikler silinebilir ve düzenlenebilmektedir. Aşağıda bulunan Şekil 40'da memnuniyet durumunun, veri içindeki tüm özneliklere göre istatistiksel oranı verilmektedir.



Şekil 40. Öznitelik Değerlerine Göre Memnuniyet Durumu

Tablo 25. Nitelik İsimleri ve Açıklamaları

Öznitelikler	Seçenekler	Açıklama
Cinsiyet	Erkek Kadın	Anketi cevaplayan kişinin cinsiyeti sorulmaktadır
Yas	15-25 25-34 35-44 45-54 55 üstü	Anketi cevaplayan kişinin yaş aralığı sorulmaktadır.
medeni	Bekar Evli Boşanmış Dul	Anketi cevaplayan kişinin medeni durumu sorulmaktadır.
egitim	İlköğretim Lise Önlisans Lisans Y.Lisans Doktora	Anketi cevaplayan kişinin eğitim durumu sorulmaktadır.
gelir	1000'in altında 1.001-2.000 2.001-3.000 3.001-4.000 4.000'ün üstünde	Anketi cevaplayan kişinin gelir aralığı sorulmaktadır.
sure	1 yıldan az 1-3 yıl 3-5 yıl 6 yıl ve fazlası	Anketi cevaplayan kişinin kullandığı GSM hattını ne zamandır kullandığı sorulmaktadır.
abonelik türü	Faturasız Faturalı	Anketi cevaplayan kişinin GSM hattının abonelik türü sorulmaktadır.
notasına	Evet Hayır	Anketi cevaplayan kişinin GSM hattından önce başka bir operatörden numara taşıyıp taşımadığı sorulmaktadır.
tutar	1-25 tl aylık 26-50 tl aylık 51-100 tl aylık 101 tl üstü	Anketi cevaplayan kişinin GSM hattı için aylık ne kadar yükleme yaptığı sorulmaktadır.
dakika	0-100 dakika 301-700 dakika 101-300 dakika 701-1000 dakika 1000 dakika üstü	Anketi cevaplayan kişinin GSM hattıyla aylık kaç dakika görüşme yaptığı sorulmaktadır.
sms	0-30 sms 31-150 sms 150-500 sms 501-1000 sms 1001 sms üstü	Anketi cevaplayan kişinin GSM hattıyla aylık kaç sms yolladığı sorulmaktadır.
operator	Turkcell Vodafone Avea	Anketi cevaplayan kişinin GSM operatörünün ismi sorulmaktadır.

Öznitelikler	Seçenekler	Açıklama
internet	Yok var ama kullanmıyorum var kullanıyorum	Anketi cevaplayan kişinin internet kullanım durumu sorulmaktadır
goruntuluarama	Yok var ama kullanmıyorum var kullanıyorum	Anketi cevaplayan kişinin görüntülü arama kullanma durumu sorulmaktadır
GPS	Yok var ama kullanmıyorum var kullanıyorum	Anketi cevaplayan kişinin GPS kullanma durumu sorulmaktadır
Tvalici	Yok var ama kullanmıyorum var kullanıyorum	Anketi cevaplayan kişinin Tv alıcı kullanma durumu sorulmaktadır
Bluetooth	Yok var ama kullanmıyorum var kullanıyorum	Anketi cevaplayan kişinin Bluetooth kullanma durumu sorulmaktadır
Oyun	Yok var ama kullanmıyorum var kullanıyorum	Anketi cevaplayan kişinin internet kullanma durumu sorulmaktadır
memnuniyet	Evet Hayır	Anketi cevaplayan kişinin GSM hattından memnuniyet durumu sorulmaktadır.
tasimaister	Evet Hayır	Anketi cevaplayan kişinin GSM hattını başka bir operatöre ya da başka bir numaraya taşıyıp taşımama isteği sorulmaktadır.
tavsiyeediyorum	Kesinlikle katılmıyorum Katılmıyorum fikrim yok Katiliyorum Kesinlikle Katiliyorum	Anketi cevaplayan kişinin kullandığı GSM hattını tavsiye etme durumu sorulmaktadır.
gelecekte devam etmek istiyorum	Kesinlikle katılmıyorum Katılmıyorum fikrim yok Katiliyorum Kesinlikle Katiliyorum	Anketi cevaplayan kişinin kullandığı GSM hattıyla gelecekte devam etme isteği sorulmaktadır.
sadakatimelayiktir	Kesinlikle katılmıyorum Katılmıyorum fikrim yok Katiliyorum Kesinlikle Katiliyorum	Anketi cevaplayan kişinin kullandığı GSM hattı kullanıcının sadakatine layık olup olmama durumu sorulmaktadır.

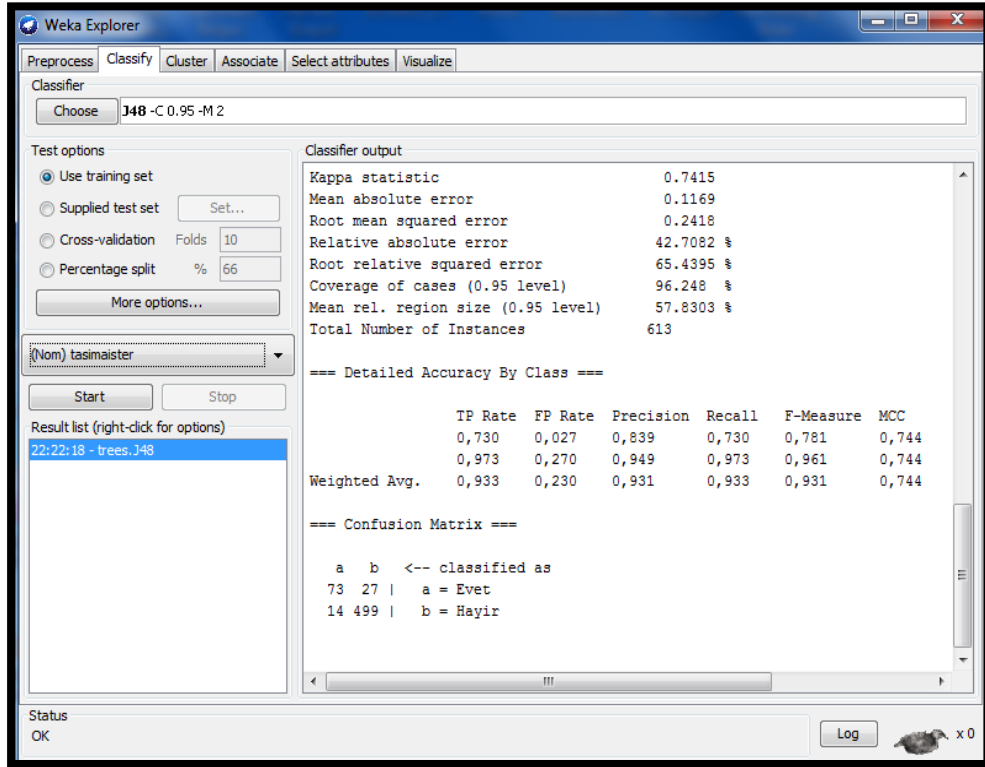
Yukarıda bulunan Tablo 25'de WEKA programında kullanılan nitelik isimleri ve açıklamaları gösterilmiştir.

6.1 KARAR AĞACI ALGORİTMASININ UYGULANMASI

Preprocess panelinden “sure, abonelik, notasima, operatör, memnuniyet, tasimaister” öznitelikleri haricinde geri kalan öznitelikler temizlenerek bir sonraki adım olan sınıflandırma paneli için hazır hale getirilmiştir. Classiffy paneli sınıflandırma algoritmalarının uygulandığı kısımdır. Sınıflayıcı algoritmalarından karar ağaçları üretilmesi için trees kısmının içinde yer alan “j48” algoritması seçilmektedir. Bu panelde çapraz geçерleme(cross validation) veya yüzde ayırma (percentage split) seçilebilmektedir. Bu seçenekler ile kuralın doğruluğu test edilmektedir.

Çapraz geçерleme veri kümesinin eşit boyutta parçalara bölmektedir. Her defasında 1 parçayı modelin testi için, n-1 parçayı ise modelin öğrenilmesi için kullanmaktadır. Programda çapraz geçерleme değeri 10’dur. Büyük veritabanlarında isteğe bağlı olarak bu değeri değıştirilebilir. Yüzde ayırma yöntemi, veri kümesinin isteğe bağlı yüzdesini modelin öğrenmesine, geri kalan yüzde değerini modelin testine ayırarak analiz yapmaktadır. Use training set ile verinin tamamını test kümesi için kullanabiliriz. Supplied test set ise, dışarıdan tedarik edilen bir test kümesini modelin öğrenilmesi için kullanılmaktadır. Bu çalışma içerisinde “tasimaister” özniteliđi için “use training set” kullanılmıştır.

Şekil 41'de j48 algoritmasının uygulanması gösterilmektedir.



Şekil 41. Verilerin j48 algoritması ile test edilmesi

J48 algoritması %93,3116 doğruluk oranı ile verileri sınıflandırma yapmıştır. Yani 613 verinin 572 adeti doğru sınıflandırılmıştır. Şekil 42'de analiz sonrasında ekrana çıkan j48 doğruluk algoritması oranları gösterilmektedir.

```

Correctly Classified Instances      572          93.3116 %
Kappa statistic                    0.7415
Mean absolute error                0.1169
Root mean squared error            0.2418
Relative absolute error            42.7082 %
Root relative squared error        65.4395 %
Coverage of cases (0.95 level)    96.248 %
Mean rel. region size (0.95 level) 57.8303 %
Total Number of Instances          613

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0,730   0,027   0,839     0,730   0,781     0,744   0,872    0,709    Evet
                0,973   0,270   0,949     0,973   0,961     0,744   0,872    0,954    Hayir
Weighted Avg.   0,933   0,230   0,931     0,933   0,931     0,744   0,872    0,914

=== Confusion Matrix ===

 a  b  <-- classified as
73 27 |  a = Evet
14 499 |  b = Hayir

```

Şekil 42. J48 Algoritması Doğruluk Değerleri

Doğruluk değer oranları için kullanılan terimler şunlardır (Akbulut, 2006):

Confusion Matrix: Yakınsaklık matrisidir. Doğru olarak sınıflandırılan örneklerin sayısı bu matrisin diyagonal elemanlarının toplamına eşittir.

True Positive (TP): Sınıflandırma algoritması tarafından a sınıfına atanan birimlerden gerçekte a sınıfına ait olanların oranını gösterir. Örneğin; Şekil 42’de Evet sınıfı için TP oranı, 0.730 olup, bu değer yakınsaklık matrisinde gösterilen ve gerçekte Evet sınıfına ait olan 73 birimin; E sınıfına atanan 100 birime oranlanması ile bulunmuştur.

False Positive (FP): Sınıflandırma algoritması tarafından a sınıfına atandığı halde gerçekte a sınıfına ait olmayan birimlerin oranını gösterir. Örneğin; Şekil 42’de Hayır sınıfı için FP oranı, 0.027 olup, bu değer yakınsaklık

matrisinde gösterilen ve gerçekte Hayır sınıfına ait olmayan 14 birimin; H sınıfına atanan 513 birime oranlanması ile bulunur.

Precision: Gerçekte a sınıfına ait olan birimlerin hangi oranda sınıflandırma algoritması tarafından a sınıfına atandığı gösterir. Şekil 42'de Evet sınıfı için Precision oranı 0.839 olup bu değer yakınsaklık matrisinde gösterilen ve E sınıfına atanan 73 birimin gerçekte E sınıfında yer alan 87 birime oranlanması ile bulunur.

Kappa istatistiği: Tahmin doğruluğunun ölçüsüdür.

J48 algoritmasının verilere uygulanmasından sonra ekrana analiz ile ilgili bilgiler gelmektedir. 613 kayıtlı veri olup, özniteliklerinin 6 adet olduğu ve özniteliklerin adı yazmaktadır. Ayrıca hangi test kümesinin kullanıldığı bilgisi ve budanmış ağacın gösterimi bulunmaktadır. Şekil 43'de J48 algoritmasının uygulanması sonrasında oluşan karar ağacı ve uygulama bilgileri yer almaktadır.

```

Instances:    613
Attributes:   6
              sure
              abonelikdur
              notasima
              operator
              memnuniyet
              tasimaister
Test mode:    evaluate on training data

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

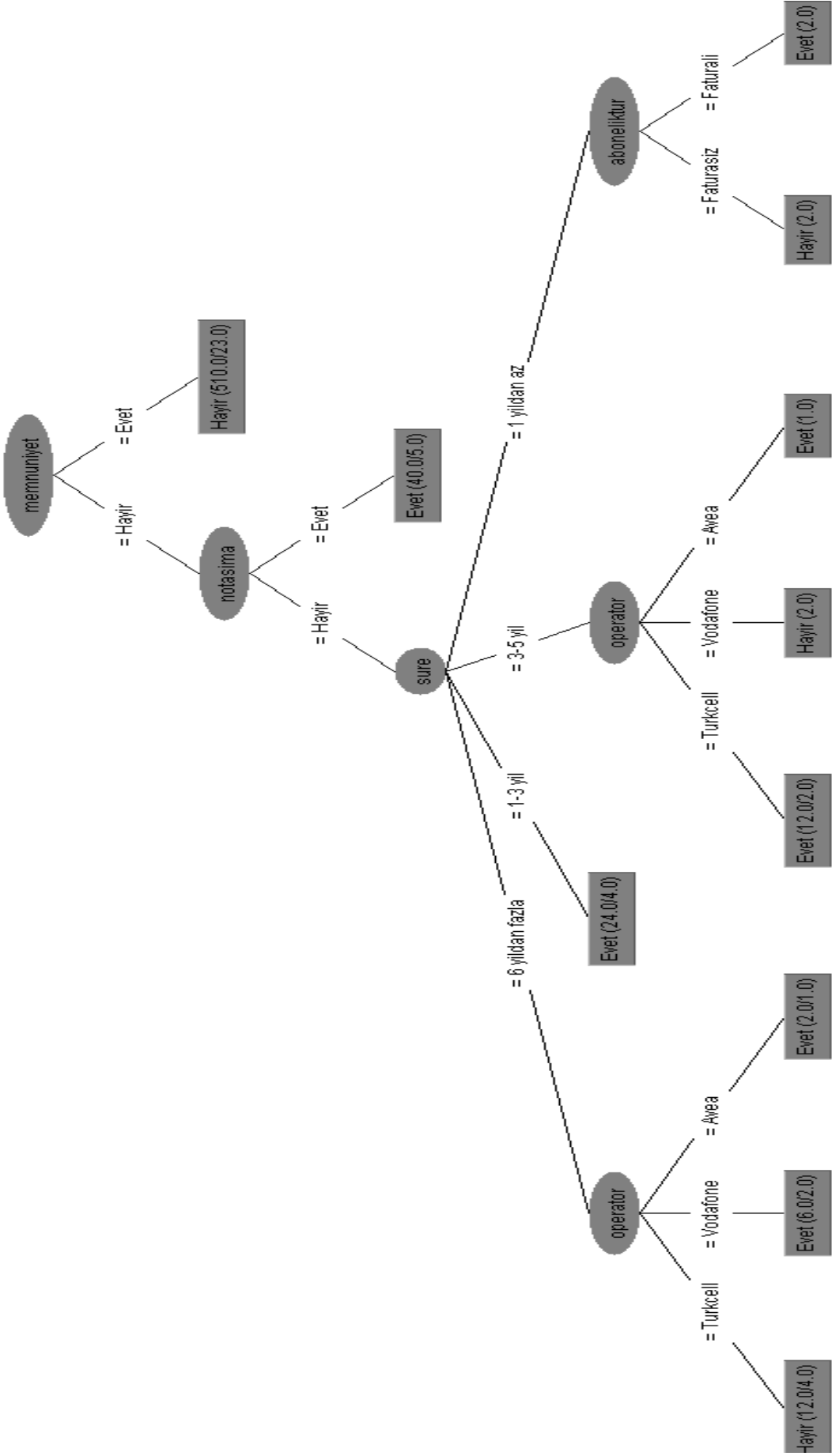
J48 pruned tree
-----

memnuniyet = Hayir
|  notasima = Hayir
|  |  sure = 6 yildan fazla
|  |  |  operator = Turkcell: Hayir (12.0/4.0)
|  |  |  operator = Vodafone: Evet (6.0/2.0)
|  |  |  operator = Avea: Evet (2.0/1.0)
|  |  sure = 1-3 yil: Evet (24.0/4.0)
|  |  sure = 3-5 yil
|  |  |  operator = Turkcell: Evet (12.0/2.0)
|  |  |  operator = Vodafone: Hayir (2.0)
|  |  |  operator = Avea: Evet (1.0)
|  |  sure = 1 yildan az
|  |  |  abonelikdur = Faturasiz: Hayir (2.0)
|  |  |  abonelikdur = Faturali: Evet (2.0)
|  notasima = Evet: Evet (40.0/5.0)
memnuniyet = Evet: Hayir (510.0/23.0)

```

Şekil 43. J48 Algoritması Karar Ağacı

J48 algoritmasının sınıflayıcı panelinde bulunan karar ağacı, görsel bir şekilde de gösterilebilmektedir. Görsel bir şekilde gösterim, daha anlaşılır ve açıklayıcıdır. Aşağıda bulunan Şekil 44'de görsel karar ağacı görülmektedir.



Şekil 44. J48 Algoritması Görsel Karar Ağacı

Şekil 44'de ilk dallanma anket katılımcılarının GSM hatlarından memnuniyet durumunu göstermektedir. İkinci dallanmadan itibaren GSM hatlarından memnun olmayanların durumunu göstermektedir. İkinci dallanma da memnun olmayan GSM hat kullanıcıların daha önceden numara taşıma işlemi gerçekleştirip gerçekleştirmediği durumunu; üçüncü dallanmada GSM hatlarına sahip olma sürelerini; dördüncü dallanmada ise operatör türünü ve abonelik türünü göstermektedir.

İlk dallanmada GSM hat sahibi 613 kişinin 510'u GSM hattından memnun olduğu, bunların sadece 23'ünün taşıma istediğini, geriye kalan 487 kişinin taşıma istemediği görülmektedir.

İkinci dallanmada GSM hatlarından memnun olmayanların daha önceden numara taşıma durumu gözükmemektedir. Daha önceden numara taşıma işlemi yapmış ve hatlarından memnun olmayan 40 kişinin 5'i numarasını taşımak istemediğini, geriye kalan 35 kişinin numarasını taşımak istediği görülmektedir.

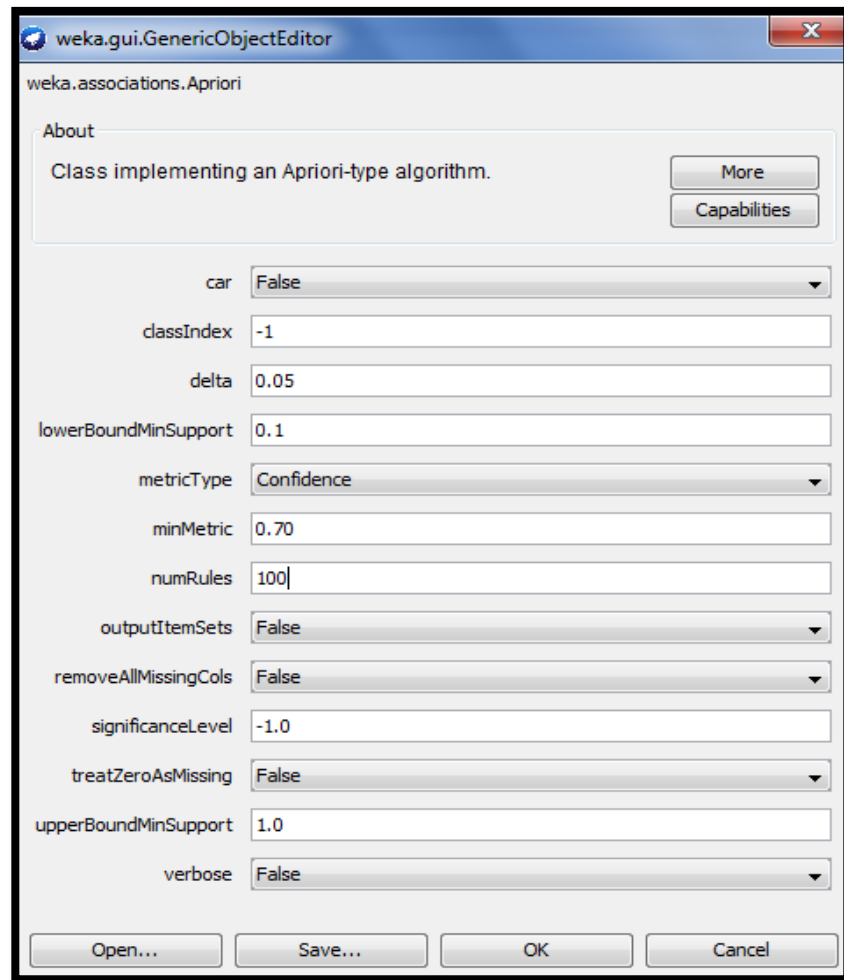
Üçüncü dallanmada GSM hattından memnun olmayan ve daha önceden numara taşıma işlemi yapmamış kişilerin hatlarına sahip olma süreleri gözükmemektedir. GSM hatlarına 1 yıldan daha az sürede sahip kişiler, abonelik türüne göre yapıllara ayrılmıştır. Faturalı ve 1 yıldan daha az sürede GSM hattına sahip olan kişiler memnun olmasalar da taşıma işlemi yapmak istedikleri gözükmemektedir. Fakat faturasız GSM hat kullanıcıları numara taşıma işlemi yapmak istemedikleri gözükmemektedir. Memnun olmayan ve daha önceden numara taşıma işlemi yapmamış, 1-3 yıl arasında GSM hatlarına sahip kişiler, numara taşıma işlemi yapmak istedikleri 24/4 oranda belirtilmiştir. Yani aynı GSM hattına 1-3 yıl arasında sahip olan 24 kişinin 4'ü hariç geriye kalan 20 kişi numara taşıma işlemi yapmak istemektedir. 3-5 yıl arasında GSM hattına sahip, memnun olmayan ve daha önceden numara taşıma işlemi yapmamış kişilerden Turkcell ve Avea hat kullanıcıları numara taşıma işlemi istediğini, Vodafone kullanıcıları ise taşımak istemediğini belirtmiştir.

6 yıldan fazla Turkcell GSM hattına sahip kişiler, numara taşıma işlemi yapmak istemedikleri, fakat 6 yıldan fazla Vodafone ve Avea kullanıcılarının numara taşıma işlemi yapmak istedikleri gözükmektedir.

Sonuç olarak; GSM hat kullanıcıları memnuniyet durumunda müşteri sadakati göstermektedir. Memnun olmama durumunda ise Faturalı GSM hattına sahip kişiler, faturasız hat sahiplerine göre numara taşıma işlem eğilimi daha fazla olduğu görülmektedir. Daha önceden numara taşıma işlemi yapmış kişiler, daha önceden numara taşıma işlemi yapmamış kişilere nazaran GSM numara taşıma işlemi eğilimi daha fazladır. Memnun olmayan ve numara taşıma işlemi yapmamış kişiler kendi aralarında alışkanlık, çevresel faktörler vb. gibi durumlardan dolayı numara taşıma işlemi isteği farklılık gösterse de, numarasını daha önceden taşımış kişiler, memnuniyetsizlik durumunda GSM hatlarını net bir şekilde taşıma isteğini belirtmiştir.

6.2 APRIORI ALGORİTMASININ UYGULANMASI

Karar ağacı uygulanan aynı veri ve özniteliklere, birliktelik kuralı algoritması uygulanmıştır. Apriori uygulanmadan önce minimum güven değerini ifade eden minMetric kısmı 0.70'e ayarlanmıştır. Ayrıca bulacağı en fazla kural sayısını da numRules kısmından istenilen değere getirilebilmektedir. Bu uygulamada en fazla 100 adet birliktelik kuralı bulunması istenmiştir. Şekil 45'de apriori algoritması yazısına çift tıkladığında çıkacak pencerede minimum destek ve güvenleri belirlenebilmektedir.



Şekil 45. Apriori Algoritması Destek Ve Güven Parametrelerinin Belirlenmesi

Apriori kurallarının özelliklerinin belirlendiği Şekil 45'deki parametreler şöyledir:

Car: Apriori algoritmasının genel bir şekilde birliktelik kurallarının uygulanmasına true / false değerleriyle izin veren bir parametredir. True değeri seçildiğinde çıkan kuralların sadece güven(confidence) değeri görülmektedir. Aksi taktirde güven ile birlikte lift, lev, cov. istatistik değerleri hesaplanmaktadır. Bu uygulamada false değeri seçilerek diğer ölçü değerleri de gösterilmiştir.

Classindex: Sınıfın öznitelik indeksidir. Yani birliktelik kuralının hangi öznitelik değerine uygulanacağı seçilebilmektedir. -1 değeri girildiğinde kurala dahil olan özniteliklerden en sonuncusuna birliktelik kuralı uygulanır. Özniteliklerin sayısını aşmamak kaydıyla classindex parametresine hangi özniteliği birliktelik kuralı uygulanacaksa onun sıralamadaki sayısı yazılmaktadır. Örnek olarak classindex' e 3 rakamı girildiğinde, 3 farklı özniteliğin, öznitelikler arasında 3. Sırada olan özniteliğine birliktelik kuralı uygulanmaktadır. Bu uygulamada -1 değeri seçilmiştir.

Delta: Destek sayısı azaltılarak tekrarlayan kural sayısının ortaya çıkmamasını sağlamaktadır. Bu uygulamada 0.05 delta değeri verilmiştir.

LowerBoundMinSupport: Minimum destek oranının en düşük sınır değerini göstermektedir. Yani değer olarak en az bulunması istenen destek değeri bu bölüme yazılmaktadır. Bu uygulamadaki en az minimum destek değeri 0.1 olarak belirlenmiştir.

MetricType: Kuralları sıralarken hangi ölçüm değerine göre sıralaması gerektiği bu bölümden ayarlanmaktadır. Seçenek olarak Confidence, Lift, Leverage ve Conviction bulunmaktadır. Bu değerlerden hangisi seçildiyse, o değeri, belirlenen parametreler dahilinde en yüksek değerden en düşük değere göre sıralamaktadır. Bu uygulamada MetricType, confidence yani güven değerine göre sıralanması istenmiştir.

MinMetric: Apriori algoritması kurallarında, seçilen MetricType' ına göre en düşük gösterilmesi istenen değer yazılmaktadır. Bu uygulamada gösterilmesi gereken en düşük confidence (güven) değeri 0.70 olarak belirlenmiştir.

NumRules: Apriori algoritması uygulanırken belirtilen özelliklere göre bulunması gereken kural sayısının yazıldığı bölümdür. Bu uygulamada en fazla 100 kural gösterilmesi istenmiştir.

OutputItemSets: Apriori algoritmasının uygulandığı tekli, ikili, üçlü vb. itemsetlerini çıkış ekranında true/false seçimleriyle gösterilip gösterilmemesini sağlayan parametredir. Bu uygulamada false değeri seçilmiştir.

RemoveAllMissingCols: Eksik değer barındıran sütunlar silinip silinmemesini true/false değerleriyle belirleyen parametredir. True değeri seçildiğinde, eksik değer taşıyan sütunlar silinerek, algoritmaya dahil edilmemektedir. Bu uygulamada false değeri seçilmiştir.

SignificanceLevel: Kuralın sağ ve sol tarafı için anlamlılık testi uygulanmaktadır. Bu uygulamada -1.0 olarak belirlenmiştir.

UpperBoundMinSupport: Minimum destek oranının en yüksek sınır değerini göstermektedir. Yani en fazla bulunması istenen destek değeri bu bölüme yazılmaktadır. Bu uygulamada en yüksek 1.0 destek oranı belirlenmiştir.

Verbose: Etkin olduğu zaman algoritma ayrıntılı modda çalışmaktadır. Bu uygulamada false değeri seçilmiştir.

Apriori algoritmasının uygulanmasından sonra ekrana öznitelik sayısı, hangi özniteliklerin kullanıldığı, minimum güven ve minimum destek oranı, kaç adet öge setinin incelendiği gösterilmektedir. Apriori algoritmasının uygulaması aşamasında ekranda gösterilen minimum destek, güven oranları Şekil 46'da gösterilmektedir.

```
Attributes: 6
            sure
            aboneliktur
            notasima
            operator
            memnuniyet
            tasimaister
=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.1 (61 instances)
Minimum metric <confidence>: 0.7
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 15

Size of set of large itemsets L(2): 49

Size of set of large itemsets L(3): 56

Size of set of large itemsets L(4): 35

Size of set of large itemsets L(5): 10

Size of set of large itemsets L(6): 1
```

Şekil 46. Apriori Algoritması Destek Ve Güven Parametre Oranı

6.2.1 Apriori Kuralları

Apriori Algoritması uygulamasının sonucunda çıkan bazı birliktelik kuralları ve istatistiksel değerlerine ait güven oranı çoktan aza doğru olacak şekilde aşağıda sıralanmıştır.

1. aboneliktur=Faturali tasimaister=Hayir 207 ==> memnuniyet=Evet 199
<conf:(0.96)> lift:(1.16) lev:(0.04) [26] conv:(3.86)
2. memnuniyet=Evet 510 ==> tasimaister=Hayir 487 <conf:(0.95)>
lift:(1.14) lev:(0.1) [60] conv:(3.47)
3. sure=6 yildan fazla notasima=Hayir 171 ==> tasimaister=Hayir 161
<conf:(0.94)> lift:(1.13) lev:(0.03) [17] conv:(2.54)
4. aboneliktur=Faturasiz tasimaister=Hayir 306 ==> memnuniyet=Evet
288 <conf:(0.94)> lift:(1.13) lev:(0.05) [33] conv:(2.71)
5. notasima=Evet memnuniyet=Evet 174 ==> tasimaister=Hayir 160
<conf:(0.92)> lift:(1.1) lev:(0.02) [14] conv:(1.89)
6. sure=6 yildan fazla operator=Turkcell 137 ==> notasima=Hayir 124
<conf:(0.91)> lift:(1.39) lev:(0.06) [34] conv:(3.42)
7. operator=Turkcell 352 ==> memnuniyet=Evet 299 <conf:(0.85)>
lift:(1.02) lev:(0.01) [6] conv:(1.1)
8. aboneliktur=Faturasiz 362 ==> tasimaister=Hayir 306 <conf:(0.85)>
lift:(1.01) lev:(0) [3] conv:(1.04)
9. operator=Turkcell 352 ==> memnuniyet=Evet tasimaister=Hayir 287
<conf:(0.82)> lift:(1.03) lev:(0.01) [7] conv:(1.1)
10. operator=Avea 131 ==> memnuniyet=Evet tasimaister=Hayir 103
<conf:(0.79)> lift:(0.99) lev:(0) [-1] conv:(0.93)
11. operator=Vodafone 130 ==> memnuniyet=Evet tasimaister=Hayir 97
<conf:(0.75)> lift:(0.94) lev:(-0.01) [-6] conv:(0.79)
12. aboneliktur=Faturasiz tasimaister=Hayir 306 ==> notasima=Hayir 220
<conf:(0.72)> lift:(1.1) lev:(0.03) [20] conv:(1.23) notasima=Hayir
memnuniyet=Evet 336 ==> aboneliktur=Faturasiz 213 <conf:(0.63)>
lift:(1.07) lev:(0.02) [14] conv:(1.11)
13. aboneliktur=Faturasiz 362 ==> operator=Turkcell memnuniyet=Evet
202 <conf:(0.56)> lift:(1.14) lev:(0.04) [25] conv:(1.15)
14. tasimaister=Hayir 513 ==> notasima=Evet memnuniyet=Evet 160
<conf:(0.31)> lift:(1.1) lev:(0.02) [14] conv:(1.04)
15. aboneliktur=Faturali 251 ==> operator=Avea 64 <conf:(0.25)>
lift:(1.19) lev:(0.02) [10] conv:(1.05)

1. kuralda GSM hattından memnun 510 kullanıcıdan 487'si, 0.95 güven oranı ile hattını taşımak istememektedir. Yani müşteri memnuniyeti ile müşteri sadakati doğru orantılı olduğu anlaşılmaktadır.

2. kuralda 6 yıldan fazla GSM hattına sahip ve daha önce numarasını taşımamış olan 171 müşterinin 161'i, 0.94 güven oranı ile hattını başka bir GSM operatörüne taşımak istemediği görülmüştür.

3. kuralda GSM hattını başka operatöre taşımak istemeyen 207 faturalı hat kullanıcının 199'u, 0.96 güven oranı ile hattından memnunken, 4. Kuralda başka operatöre hattını taşımak istemeyen 306 faturasız kullanıcının 288'i 0.94 güven oranı ile GSM hattından memnun olduğu görülmüştür.

9. kuralda Turkcell GSM hat müşterisi olan 352 kişiden 287'si, 0.82 güven oranı ile aynı anda GSM hattından memnun olduğu gözükmekte ve numarasını başka operatöre taşımak istememektedir.

10. kuralda Avea GSM hat müşterisi olan 131 kişiden 103'ü, 0.79 güven oranı ile aynı anda GSM hattından memnun olduğu gözükmekte ve numarasını başka bir GSM operatörüne taşımak istememektedir.

11. kuralda Vodafone GSM hat müşterisi olan 130 kişiden 97'si, 0.75 güven oranı ile aynı anda kullandığı hattından memnun olduğu gözükmekte ve numarasını başka operatöre taşımak istemediği görülmüştür.

12. kuralda hattını başka operatöre taşımak istemeyen Faturasız GSM hat kullanıcısı olan 306 kişinin 220'sinin 0.72 güven oranı ile daha önceden numarasını taşıma işlemi yapmadığı görülmüştür.

14. kuralda GSM hattını taşıma işlemi yapmamak istemeyen 513 kişiden 160'ı, daha önceden numara taşıma işlemi yapmış ve hattından memnun olduğu 0.31 düşük güven oranı ile görülmüştür.

Apriori algoritmasının kuralları, öznitelikler başta olmak üzere destek ve güven ölçüt oranı değiştirilerek 10000'lerce kural elde edilebilmektedir ya da istenilen oranlar dahilinde kural sayısı kısıtlanabilmektedir.

WEKA programı sayesinde birliktelik kuralı algoritması olan Apriori algoritmasının uygulanması sonucunda elde edilen ilk kuralın ölçü değerlerinin hesaplanması aşağıdaki şekildedir:

- memnuniyet=Evet 510 ==> tasimaister=Hayir 487 <conf:(0.95)>
lift:(1.14) lev:(0.1) [60] conv:(3.47)

Toplam katılımcı sayısı 613 olan hat kullanıcısı bilgilerine göre, GSM hattından memnun olan 510 hat sahibinin, 487'si numarasını başka bir GSM operatörüne taşımak istememektedir. Bu verilere göre;

Güven (Confidence) değeri;

$$Güven(X \rightarrow Y) = \frac{sayı(X, Y)}{sayı(X)} \longrightarrow \text{Conf: } 487/510 = 0,954$$

şeklinde hesaplanmaktadır.

İlgi (Lift) değeri;

$$Lift(X \rightarrow Y) = \frac{conf(X \rightarrow Y)}{Supp(Y)}$$

Bu formülün uygulanması için öncelikle X'in destek oranı bulunmalıdır.

$$Supp(Y) = \frac{Y \text{ Küme Üye Sayısı}}{\text{Toplam Öğe Küme Sayısı}}$$

Formüldeki Y Küme Üye sayısı, tasimaister özneliğine Toplam Öğe Küme Sayısı olan 613 kişi arasında hayır diyenleri temsil etmektedir. Anket verisinde 399 GSM hat kullanıcısı numara taşıma isteğine hayır cevabı vermiştir. Yani Y'nin destek oranı aşağıdaki şekilde hesaplanmaktadır.

$$P(Y) = supp(Y)$$

$$P(Y) = \frac{513}{613} = 0,836$$

Formüle göre X' in Y' ye göre güven oranı ile Y' nin destek oranının birbirine oranlanması ile Lift değerine ulaşılmaktadır.

$$\text{İlgi} = \text{Lift} = \frac{0,954}{0,836} = 1.14$$

şeklinde hesaplanmaktadır.

Kaldıraç (Leverage) değeri;

$$\text{Kaldıraç} = \text{Leverage}(X \rightarrow Y) = P(X \text{ ve } Y) - (P(X)P(Y))$$

Bu formülün uygulanması için öncelikle P(X), P(Y) ve P(X,Y) değerlerinin hesaplanması gerekmektedir.

$$\text{Supp}(X) = \frac{X \text{ Küme Üye Sayısı}}{\text{Toplam Öğe Küme Sayısı}} = \frac{510}{613} = 0,8319$$

$$\text{Supp}(Y) = \frac{Y \text{ Küme Üye Sayısı}}{\text{Toplam Öğe Küme Sayısı}} = \frac{513}{613} = 0,836$$

$$\text{Supp}(X \text{ ve } Y) = \frac{\text{Kural İçinde } Y \text{ Geçme Sayısı}}{\text{Toplam Öğe Küme Sayısı}} = \frac{487}{613} = 0,794$$

Formüllere göre;

$$P(X) = 0,8319 \quad P(Y) = 0,836$$

$P(X \text{ ve } Y) = 0,794$ olarak hesaplanmıştır.

Bu değerler Leverage formülüne yerleştirildiğinde;

$$Leverage(X \rightarrow Y) = P(0,794) - (P(0,8319)P(0,836))$$

$$Leverage(X \rightarrow Y) = 0,1 \text{ (0,098 yuvarlanarak 0.1' tamamlanmıştır.)}$$

(Bu oran 613 katılımcıya oranlandığında $613 \times 0.098 = 60$ olarak kuralda gösterilen değer elde edilir.)

Kanaat (Conviction) değeri;

$$Kanaat = Conviction(X \rightarrow Y) = \frac{1 - P(Y)}{1 - Conf(X \rightarrow Y)}$$

$$Conviction(X \rightarrow Y) = \frac{1 - 0,836}{1 - 0,954} = \frac{0,16}{0,046} = 3,47$$

şeklinde hesaplanmaktadır.

7. SONUÇ VE ÖNERİLER

Web tabanlı anket yöntemiyle, birden fazla öznitelik değeri verisi toplanabilmektedir. Toplanan öznitelikler arasında sadece istenilen amaca yönelik öznitelik seçimi yapılarak birbirleri arasındaki ilişki, güven ve destek bulunabilmektedir. Kullanılan algoritmalar sayesinde 1000'lerce kural bulunabilmekte ve geleceğe yönelik çıkarım yapılmasına yardımcı olmaktadır.

Bu çalışmada hazır veri yerine, 613 GSM hattı sahibinin online bir şekilde katıldığı anketten 21 sorudan 20'sine cevap verilmeleri istenerek gerekli datalar veri madenciliği analizi için toplanmıştır. Verilerin analizi için WEKA programı kullanılmıştır. En çok kullanılan algoritmalarından j48 karar ağacı algoritması ve birliktelik kuralı algoritması olan apriori kullanılmıştır. İsteğe bağlı olarak VM algoritmalarından kümeleme analizi kullanılarak müşteri grupları kümelenebilmektedir. Yapılan çalışmadaki sonuçlara göre, faturalı hatta sahip kullanıcılar, memnuniyet konusunda, faturasız hat sahiplerine göre daha hassas durumda oldukları, memnuniyetsiz olma durumunda, kullanılan GSM hattını başka bir operatöre taşıma isteği daha fazla olmaktadır. GSM firmalarından Turkcell, en eski GSM firması olması nedeniyle 6 yıldan fazla hat sahiplerinin memnun olmama durumunda bile numaralarını taşımadıkları görülmektedir. Fakat GSM hattına 3-5 yıl arasında sahip olan kişiler, memnun olmama durumunda operatöre ve alışkanlığa bakmaksızın numara taşıma yapabilecekleri görülmektedir. Ayrıca daha önce hat taşımış kullanıcılar, GSM hatları için risk grubu olarak gözükmemektedir. Çünkü anket verilerine göre, daha önce numarasını taşımış kullanıcılar, memnun olmama durumunda numarasını taşımamış olan kullanıcılara oranla, numara taşıma ihtimali daha fazladır. GSM firmaları için müşteri kaybı yaşamamaları adına, müşteri sadakati önemli bir durumdur. Bu yüzden faturalı hat sahiplerine belirli zamanlar periyodlarında faturalarında indirim yaparak müşteri sadakati sağlanabilir, numarasını daha önceden taşımış

kullanıcılara ise internet, sms, konuşma dakikası gibi hediyeler ile bu müşteri grubunun numara taşıma riski azaltılabilir.

Web tabanlı anket sistemiyle sadece GSM sektöründe değil pazar araştırmaları, telekomünikasyon, hastaneler, oteller, ticari firmalar, eğitim sektörü gibi sahalarda hızlı bir şekilde veriler toplanarak bilgiye erişmek mümkün olabilmektedir. Uzaktan eğitim sistemine sahip eğitim kurumlarında, web tabanlı anket sistemi yöntemiyle öğrenci dataları toplanarak herhangi bir veri madenciliği programında analiz edilip, gerek öğrenci başarısını etkileyen faktör tahmininde gerekse memnuniyet durumuna yönelik çıkarımlarda bulunulabilir. İlköğretim öğrencilerine ise belirli periyotlarda online anketlere tabi tutarak, veri madenciliği analizi yardımıyla öğrencinin yeteneğine göre meslek seçimine yönlenebilir. Firmaların iş alımlarında kullanabileceği anket verisi üzerinde veri madenciliği analizi yaparak, işe alım kararı veya iş başvurusu yapan kişinin geleceğe yönelik performans analizi yapılabilir. Ayrıca günümüzde hastanelerde hasta muayenesi, randevulu sistem üzerinden internet aracılığı ile yapılmaktadır. Fakat kimi hastalar hangi bölümden randevu alması gerektiğini bilmemektedir ya da tahmin yöntemiyle bölüm randevusu almaktadır. Web tabanlı anket yöntemiyle hasta semptomlarına göre hastanın gitmesi gereken hastane bölümüne yönlendirmesi veri madenciliği algoritmaları ile mümkün olabilir. Hastaların vakit kaybetmeden doğru yönlendirilmesi, kimi hastalar için hayati önem taşıdığından dolayı, hastanın kendisinden elde edilen veriler üzerinde veri madenciliği analizi ile yüksek güven oranı çıkan bölümlere yönlendirilerek zaman kaybı veya geç kalma olasılığı azaltılabilir.

KAYNAKÇA

AGRAWAL, R., IMIE linski, T., SWAM i, (1993), "Mining association rules between sets of items in large databases", In ACM SIG MOD Conf. Management of Data.

AGRAWAL, R. ve SRIKANT, R., (1994), Fast Algorithms for Mining Association Rules in Large Databases, Proceedings of the 20th International Conference on Very Large.

AKBULUT, Sinem, (2006), Veri Madenciliği Teknikleri İle Bir Kozmetik Markanın Ayrılan Müşteri Analizi Ve Müşteri Segmentasyonu, Gazi Üniversitesi Fen Bilimleri Enstitüsü, (Yayınlanmamış Master Tezi).

ALAN, M. A., (2012), Veri Madenciliği Ve Lisansüstü Öğrenci Verileri Üzerine Bir Uygulama, **Dumlupınar Üniversitesi Sosyal Bilimler Dergisi**, Sayı: 33.

ALTINTOP, Ümmühan,(2006), İnternet Tabanlı Öğretimde Veri Madenciliği Tekniklerinin Uygulanması, Kocaeli Üniversitesi Fen Bilimleri Enstitüsü, (Yayınlanmamış Master Tezi).

ARASIL Ö.,KARAÇUHA E.,AYDIN S.,ÖZER G.,(2004), Türk Gsm Sektöründe Müşteri Sadakati, Memnuniyeti, Güven Değişirme Maliyeti Arasındaki Dinamik İlişkiler: Yapısal Denklem Modelleme Tekniği, **İktisat, İşletme ve Finans Dergisi**, cilt:19, sayı:219, s. 46-61.

BAYKASOĞLU, A. (2005), Veri madenciliği ve çimento sektörüne bir uygulama, Akademik Bilişim Konferansı, Gaziantep, 82-83.

BIRTIL, F. S.,(2011) Kız Meslek Lisesi Öğrencilerinin Akademik Başarısızlık Nedenlerinin Veri Madenciliği Tekniği ile Analizi, Afyon Kocatepe Üniversitesi Fen Bilimleri Enstitüsü, (Yayınlanmamış Master Tezi).

BRANSTEN,L.,(1999), "Technology – power tools – looking for patterns: data mining enables companies to better manage the reamofstatistics theycollect; the goal: spot the unexpected", Wall Street Journal, 27 (12): 16- 20.

CHEN, Y.L., CHEN, J.M. ve TUNG, C.W., (2006), A Data Mining Approach For Retail Knowledge Discovery With Consideration of the Effect of Shelf-Space Adjacency on Sales, Decisions Support Systems, cilt: 42, sayı: 3, 1503-1520.

ÇELİK, Öner, (2009), Belirli Periyotlarda Atmlerdeki Para Talebinin Belirlenmesinde Veri Madenciliğinin Kullanılması, Sakarya Üniversitesi Fen Bilimleri Enstitüsü, (Yayınlanmamış Master Tezi).

CHAUDHURI, (1998), S.: “Data Mining and Database Systems : Where is the Intersection?”, IEEE Bulletin of the Technical Committee on Data Engineering , Vol.21 No.1, 48.

ÇINAR, A., SİLAHTAROĞLU G., (2012), Veri Madenciliği Teknikleri ile Müşteri Memnuniyetine Etki Eden Gizli Nedenlerin Keşfi, **Marmara Üniversitesi İ.İ.B. Dergisi**, cilt: XXXIII, sayı: II, S. 309-330.

DANACI M., ÇELİK M., AKKAYA A. E.,(2010), Veri Madenciliği Yöntemleri Kullanılarak Meme Kanseri Hücreleri Tahmin ve Teşhisi, Danacı, M., Çelik, M., Akıllı sistemlerde Yenilikler ve Uygulamaları Sempozyumu(ASYU' 2010), 21-24 Haziran, Kayseri, 9–12.

DILLY, Ruth. (12/1995). Data Mining: An Introduction.

DOĞAN, Buket,(2006), Zeki Öğretim Sistemlerinde Veri Madenciliği Kullanılması, Marmara Üniversitesi Fen Bilimleri Enstitüsü, (Yayınlanmamış Doktora Tezi).

DOĞAN, Şengül, (2007), Veri Madenciliği Kullanarak Biyokimya Verilerinden Hastalık Teşhisi, Fırat Üniversitesi Fen Bilimleri Enstitüsü, (Yayınlanmamış Master Tezi).

DOLGUN, M. R., (2006), Büyük Alışveriş Merkezleri İçin Veri Madenciliği Uygulamaları, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, (Yayınlanmamış Master Tezi).

ERKAN, Ü., EYÜP, S., (2005), "Web Basın Verilerine Apriori Algoritması Uygulanarak Düzenli Birliktelik Kurallarının Bulunması", 4rd International Advanced Technologies Symposium..

GÜLPINAR, Vildan, ALTAŞ, Dilek, (2013), Customer Churn Analysis through Artificial Neural Networks in Turkish Telecommunications Market International Journal of Economic Perspectives.

HAN, J, KAMBER, S.F.,(2001), "**Data Mining: Concepts and Techniques**", Morgan Kaufmann Publishers.

HİNNEBURG, A., KEİM, D.A.,(1999), "Clustering Techniques for Large Data Sets from Past to Future".

HOLSHEİMER, M; SİEBES, A.P.J.M.:(1994)," Data Mining: the search for knowloedge in databases", Technical Report CS-R9406, Amsterdam, The Natherlands.

İŞLER Yalçın, NARİN, Ali,(2012), WEKA Yazılımında k-Ortalama Algoritması Kullanılarak Konjestif Kalp Yetmezliği Hastalarının Teşhisi, **SDU Teknik Bilimler Dergisi**, cilt:2, sayı:4, s. 21-29..

KAYAALP Kıyas, (2007), Asenkron Motorlarda Veri Madenciliği ile Hata Tespiti, Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü, (Yayınlanmamış Master Tezi).

KÜÇÜKSİLLE, E., (2009), Veri Madenciliği Süreci Kullanılarak Portföy Performansının Değerlendirilmesi ve IMKB Hisse Senetleri Piyasasında Bir Uygulama, Süleyman Demirel Üniversitesi Sosyal Bilimler Enstitüsü, Doktora Tezi, Isparta.

LAROSE, Daniel T., Discovering Knowledge in Data(USA: John Wiley&Sons, 2005), s.16.

ÖZÇINAR, Hüseyin, (2006), Kpss Sonuçlarının Veri Madenciliği Yöntemleriyle Tahmin Edilmesi, Pamukkale Üniversitesi Fen Bilimleri Enstitüsü, ,(Yayınlanmamış Master Tezi).

ÖZSEVEN, T. ve DÜĞENCİ M., (2011), LOG Analiz: Erişim Kayıt Dosyaları Analiz Yazılımı ve GOP Üniversitesi Uygulaması, Bilişim Teknolojileri Dergisi, cilt: 4, sayı: 2, 55- 66 .

RICHELDİ, M. and Perrucci , (2002), A Churn Analysis Case Study, Telecom Italia Lab. Torino. 3-6.

ROIGER, R. J. ve M. W. Geatz, (2003), “**Data Mining A Tutorial-Based Primer**”, Addison Wesley,USA..

SİLAHTAROĞLU, Gökhan,(2008), **Kavram ve Algoritmalarıyla Temel Veri Madenciliği**, İstanbul, Papatya Yayıncılık.

TAN, Pang-Ning, STEINBACH, Michael, KUMAR, Vipin, (2006), **Introduction to Data Mining**, Pearson Addison Wesley, Boston.

TOSUN, Tuğba, (2006), Veri Madenciliği Teknikleriyle Kredi Kartlarında Müşteri Kaybetme Analizi, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, (Yayınlanmamış Yüksek Lisans Tezi).

TÜRE, Mevlut, TOKATLI, Füsun ve KURT, İmran, (2008), “Using Kaplan-Meier Analysis Together With Decision Tree Methods(C&RT, CHAID, QUEST, C4.5 and ID3) In Determining Recurrence-Free Survival of Breast Cancer Patients”, **Expert Systems With Applications,Article in Pres.**

ÜÇGÜN, Kerem, (2009), Ortaöğretim Okulları için Öğrenci Otomasyonu Tasarımı ve Öğrenci Verileri Üzerine Veri Madenciliği Uygulamaları, Marmara Üniversitesi Fen Bilimleri Enstitüsü, (Yayınlanmamış Master Tezi).

WITTEN, I. H. ; Frank, E. ; (2005) Data Mining, Practical Machine Learning Tools and Techniques ; Morgan Kaufmann , USA.

YAN, L., MILLER, J., MOZER, M., WOLNIEWICZ, R.,(2001), Improving Prediction of Customer Behaviour in Nonstationary Environments.

İNTERNET KAYNAKÇASI

- [1] <http://www.cs.xu.edu/csci390/07s/dmintro.pdf>
- [2] www.spss.com/cool/papers/clem_healthcare1.htm
- [3] <http://www.bilgisayarkavramlari.com/2011/09/09/birliktelik-kurallarinin-pay-olcumleri-interest-measures-for-association-rules/>
- [4] http://michael.hahsler.net/research/association_rules/measures.html

EK

ÖRNEK BİR ANKET SONUCU**1. Cinsiyetiniz?**

Erkek

2. Yaşınız?

15-24

3. Medeni Durumunuz?

Bekar

4. Eğitim Durumunuz?

Önlisans

5. Ortalama Aylık Geliriniz? (TL)

1.001-2.000

6. Kullanmakta olduğunuz GSM hattına(operatörünüze) sahip olma süreniz?

6 yıldan fazla

7. GSM Abonelik Türünüz nedir?

Faturasız

8. Farklı bir operatörden ilgili operatöre numaranızı taşıdınız mı?

Hayır

**9. Ortalama Aylık GSM Ödeme Tutarınız? (TL) (Faturalı/ Faturasız)
Sadece rakamsal değer yazınız.**

20

10. GSM üzerinden Ortalama Aylık Konuşma Süreniz? (dakika). Sadece rakamsal değer yazınız.

60

11. GSM üzerinden Ortalama Aylık SMS/MMS Sayınız?

250

12. Hangi GSM Operatörünü kullanıyorsunuz?

Turkcell

13. Turkcell GSM operatörü kullanıyorsanız tarifinizi seçiniz

Gencturkcell

14. Vodafone GSM operatörü kullanıcısı iseniz tarifinizi seçiniz.

No Response

15. Avea GSM operatörü kullanıyorsanız tarifinizi seçiniz.

No Response

16. Cep Telefonu cihazınız aşağıdaki belirtilen özelliklerden hangisine sahiptir?

	Var kullanıyorum	Yok	Var ama kullanmıyorum
1 İnternet			X
2 Görüntülü Arama		X	
3 GPS		X	
4 Karasal TV Alıcı		X	
5 Bluetooth	X		
6 Oyun	X		

17. Kullandığınız operatörden memnun musunuz?

Hayır

18. Kullandığınız Operatörü değiştirmeyi düşünüyor musunuz?

Evet

21. GSM hattınızla ilgili aşağıda yer alan yorumları ' 1: kesinlikle katılmıyorum' ve ' 5: kesinlikle katılıyorum' arasında değerlendiriniz.

	Kesinlikle katılmıyorum	Katılmıyorum	Ne katılıyorum ne katılmıyorum(fikrim yok)	Katılıyorum	Kesinlikle katılıyorum
Kullandığım GSM Operatörünü Tavsiye ediyorum					
Gelecekte de aynı GSM Operatörüyle devam etmek istiyorum.					
Kullandığım GSM Operatörü Sadakatime Layıktır.		X			

ÖZET

Teknolojinin çok hızla geliştiđi çağımıza paralel olarak bilgi paylaşımı yaygınlaşmış ve bilgiler dijital ortama aktarılmıştır. Teknolojinin gelişmesiyle hafıza birimleri ve hafıza kapasiteleri de artmıştır. Böylelikle veriler dijital ortamda daha fazla ve daha kalıcı şekilde hafıza birimlerinde saklanmaktadır. Veriler arttıkça ulaşmak istediğimiz bilgiye ulaşmak daha da zorlaşmaktadır. Bu noktada gerekli gereksiz veri yığınlarından, işe yarar bilgi çıkarılması için Veri Madenciliğinden yararlanmaktayız. Veri Madenciliđi işlenmemiş verilerden anlamlı ve amaca yönelik bilgilerin çıkarılması işlevidir.

Bu tez çalışmasında daha önceden var olan kayıtlı veriler yerine web tabanlı anket yöntemiyle farklı cinsiyet ve yaş aralıklarında bulunan kişilerin kendisinden edinilen verilerin analizi için veri madenciliđi yöntemleri kullanılarak gerekli bilgiye ulaşılması amaçlanmıştır. Bu bilgiler sonucunda GSM şirketlerinin müşteri kaybetme potansiyeli olan müşteri grupları ve alışkanlıkları veri madenciliđi ile tespit edilebilmektedir. Veri Madenciliđi analizi için çeşitli dillerde yazılmış bilgisayar yazılımlarına ihtiyaç vardır. Bu yazılımlar farklı teknik ve algoritmaları kullanarak verilerin işleyerek analiz sonuçlarını sunmaktadır. Web tabanlı anket sistemi ile elde edilen veriler WEKA programıyla işlenip analiz edilmiştir.

Anahtar Kelimeler: Veri Madenciliđi, J48, Apriori, WEKA, GSM, Online Anket

ABSTRACT

DATA ANALYSIS OF A WEB-BASED SURVEY USING DATA MINING METHODS

The rapid development of computing technology has led to a large increase in knowledge and data sharing. Additionally, the capacity of memory modules to hold data has increased dramatically making it more difficult to find the desired data and information. Due to the abundance of information and difficulty in locating it, the use of data mining is needed for the extraction of meaningful and purposeful information.

This thesis examines a web-based survey of GSM customers utilizing data mining analysis. The purpose of the data mining analysis is to learn why GSM customers switch from one provider of services to the other. For the purpose of this study, the data mining software "WEKA" was used to identify habits and trends amongst different customer groups. Having knowledge of this information can be of critical importance for GSM providers to retain their customers. In this thesis, the results of the analysis done by the WEKA program will be presented.

Keywords: Data Mining, J48, Apriori, WEKA, GSM, Online Survey

ÖZGEÇMİŞ

Zeynep YURTTAŞ ELABİAD

Cihangir Mah. Duygu sokak no:2 Avcılar/ İSTANBUL/TURKIYE

GSM: +90 539 847 48 05

: +90 212 422 06 70 (152)



Genel Bilgiler

İsim, Soyisim	Zeynep YURTTAŞ ELABİAD
Doğum Tarihi	08/05/1985
Doğum Yeri	İskenderun
Cinsiyeti	Kadın
Uyruğu	T. C.
Mesleği	Öğretim Görevlisi
Medeni Durumu	Evli

EĞİTİM

	Okul	Dönem
1	İskenderun Anadolu Meslek Lisesi (Bilgisayar)(High School)	1999-2003
	Süleyman Demirel Üniversitesi (Isparta) (Bachelor)	
2	The faculty of Technical Education Computer Systems Education	2004-2009
3	İst. Aydın Üniversitesi Computer Engineering Yüksek Lisans Programı (Master)	2010- 2013

Bilgisayar Becerileri

İşletim Sistemleri	DOS, Vista, Windows XP, Windows NT/2003, Linux,Windows7, Server 2003
Yazılımlar	Ms Office,Flash, Adobe Captivate
Programlama Dilleri	Turbo C, Turbo Pascal, C ++, C#, Assembly,Java
Grafik Tasarımı	Photoshop, Illustrator, Coreldraw, Indesign, Gimp,Fireworks
Veritabanı	Ms SQL Server, Access
Simulasyon Prog.	Routing-Cisco Packet Tracer, MatLABMultiSIM, Proteus,MS Robotics Studio
WebTasarımı	Html, Css ,Ms FrontPage, Dreamweaver
Mikrodenetleyiciler	Atmel 8051-8251, PIC 16F877,Keil

Çalışma Deneyimi

	Görevi/Kurumu	Dates
1	Öğr. Gör, İst. Gelişim Üniversitesi Gelişim MYO /Vocational School	8.2009 - Devam
2	Stajyer, IMKB Teknik Lise, Anadolu Meslek ve Endüstri Meslek Lisesi İSPARTA	20.02 – 08.05.2009
8	Stajyer, İskenderun Kaymakamlığı-Sosyal Yard. Day.Vakfi	01.07.2002 – 15.06.2003

Dil Becerisi

Orta Seviye İngilizce

İletişim Bilgileri

Address	Cihangir Mah. Duygu sok. no:2 Avcılar / İSTANBUL
GSM	+212 422 0670
e-mail	zyurttas@gelisim.edu.tr

