

T.C.
BAHÇEŞEHİR ÜNİVERSİTESİ

**VERİ MADENCİLİĞİ YÖNTEMLERİ İLE
BANKA DİJİTAL KANALLARINA EĞİLİMLİ
MÜŞTERİLERİN BELİRLENMESİ**

Yüksek Lisans Tezi

SEMRA KATILMIŞ

İSTANBUL, 2018

**T.C.
BAHÇEŞEHİR ÜNİVERSİTESİ**

**FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİ TEKNOLOJİLERİ**

**VERİ MADENCİLİĞİ YÖNTEMLERİ İLE
BANKA DİJİTAL KANALLARINA EĞİLİMLİ
MÜŞTERİLERİN BELİRLENMESİ**

Yüksek Lisans Tezi

SEMRA KATILMIŞ

Tez Danışmanı: DOÇ. DR. M. ALPER TUNGA

İSTANBUL, 2018

T.C.
BAHÇEŞEHİR ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ BİLGİ TEKNOLOJİLERİ

Tezin Adı: Veri Madenciliği Yöntemleri İle Banka Dijital Kanallarına Eğilimli
Müşterilerin Belirlenmesi

Öğrencinin Adı Soyadı: Semra Katılmış

Tez Savunma Tarihi:

Bu tezin Yüksek Lisans tezi olarak gerekli şartları yerine getirmiş olduğu Fen Bilimleri
Enstitüsü tarafından onaylanmıştır.

Dr. Öğr. Üyesi Yücel Batu Salman

Enstitü Müdürü

İmza

Bu tezin Yüksek Lisans tezi olarak gerekli şartları yerine getirmiş olduğunu onaylarım.

Doç. Dr. Mehmet Alper TUNGA

Program Koordinatörü

İmza

Bu Tez tarafımızca okunmuş, nitelik ve içerik açısından bir Yüksek Lisans tezi olarak
yeterli görülmüş ve kabul edilmiştir.

Jüri Üyeleri

İmzalar

Tez Danışmanı

Doç. Dr. Mehmet Alper TUNGA

.....

Üye

Doç. Dr. Cihangir ÖZEMİR

.....

Üye

Dr. Öğr. Üyesi Cemal Okan ŞAKAR

.....



BU TEZİ, HER ZAMAN YANIMDA OLAN AİLEME İTHAF EDİYORUM.

TEŐEKKÜR

Bu tezin hazırlanmasında ve her aşamasında yardımlarını esirgemeyen danışman hocam Sayın Doç. Dr. Mehmet Alper TUNGA' ya, öğrenim süreci boyunca her zaman yanımda olan Tuğçe Merve DEMİR, Fisun ÖZDEN ve Özgür GÜRKAN' a sonsuz teşekkürler. Çalışmalarımın en sıkıntılı zamanlarında dahi yardımlarını ve desteklerini esirgemeyen aileme sonsuz teşekkürler...

İstanbul, 2018

Semra KATILMIŐ



ÖZET

VERİ MADENCİLİĞİ YÖNTEMLERİ İLE BANKA DİJİTAL KANALLARINA EĞİLİMLİ MÜŞTERİLERİN BELİRLENMESİ

Semra Katılmış

Bilgi Teknolojileri

Tez Danışmanı: Doç. Dr. Mehmet Alper TUNGA

Mayıs 2018, 49 Sayfa

Veri madenciliği, büyük verilerin içerisinde yararlı bilgilerin keşfi olarak ifade edilir. Bankacılık sektöründe dijital kanal kullanan müşterilerin özellikleri bulunarak, benzer özelliklere sahip müşterilerin dijital kanallara yönlendirilmesi bu tezin amacıdır.

Özel bir bankanın 33.503 müşterisinin 100.000 satırlık örtülü verisinin bulunduğu data üzerinden yapılan çalışmada müşterilerin özelliklerinin yoğunluk frekanslarına göre gruplandırılması ve elemeleri yapılmıştır. Datanın madenciliği yapılırken Weka programındaki sınıflandırma algoritmaları olan Navie bayes, Logistic, IBk ve J48 kullanılmıştır. Bu tez çalışmasında kullanılan algoritmaların başarıları değerlendirilmiş ve bu değerlendirmeden öne çıkan sonuçlar doğrultusunda, dijital kanal kullanan müşteriler belirlenip, kullanmayan müşterilerden profili uygun olanlara ilgili ürünün önerilmesi amaçlanmıştır.

Bu uygulamanın sonucunda da dijital kanal kullanmaya yönlendirilen müşteriler sayesinde banka sektörüne farklı bir maliyet avantajı getirilmesi hedeflenmiştir.

Anahtar Kelimeler: Bankacılık, Veri Madenciliği, Weka, Dijital Kanallar

ABSTRACT

DETERMINING BANKING CUSTOMERS WITH A PROPENSITY FOR DIGITAL CHANNELS USING DATA MINING METHODS

Semra Katılmış

Information Technologies

Thesis Consultant: Assoc. Prof. Dr. Mehmet Alper TUNGA

May 2018, 49 Page

Data mining is denoted as the discovery of useful information out of big data. By studying the features of customers who use digital channels, it is the aim of this thesis to divert other customers with similar characteristics to the bank's digital channels.

Grouping and elimination of characteristics of customers using frequency density are performed after a study of 100.000 lines of crypted data of 33.503 customers of a private Turkish bank. While conducting data mining, Naive Bayes, Logistic, IBk and J48, which are the classification algorithms in Weka software, are used. Success of algorithms used in this study is evaluated, and in the light of those outcomes, customers who use digital channels are determined and it is aimed to propose digital channels to those customers who don't use them. Thanks to those customers who will be inverted to use digital channels as a result of this application, it is aimed to bring cost advantage to the banking sector.

Keywords: Banking, Data Mining, Weka, Digital Channels

İÇİNDEKİLER

TABLOLAR	ix
ŞEKİLLER	x
KISALTMALAR	xi
SEMBOLLER	xii
1.GİRİŞ	1
2.LİTERATÜR TARAMASI	4
2.1 VERİLER VE VERİ İŞLEME YÖNTEMLERİ	4
2.2 VERİ MADENCİLİĞİ	4
2.2.1 Veri Madenciliği Kavramı	4
2.2.2 Veri Madenciliğinin Aşamaları	7
2.2.2.1 Sorunun Belirlenmesi	8
2.2.2.2 Verinin Anlaşılması	8
2.2.2.3 Verinin Hazırlanması	9
2.2.2.4 Veri Modelinin Kurulması ve Değerlendirilmesi	10
2.2.2.5 Veri Modelinin Kullanılması	11
2.2.2.6 Veri Modelinin İzlenmesi	12
2.2.3 Veri Madenciliğinin Yöntemleri / Teknikleri	12
2.2.3.1 Sınıflandırma	12
2.2.3.2 Kümeleme	15
2.2.3.3 Birliktelik Kuralları	16
2.2.4 Veri Madenciliğinde Karşılaşılan Sorunlar	17
2.3. BANKACILIKTA DİJİTAL KANALLAR	18
2.3.1 Bankacılığın Tanımı	18
2.3.2 Bankacılıkta Dijital Kanalların Tanımı	20
2.3.3 Bankacılıkta Dijital Kanalların Avantajları	22
2.3.4 Bankacılıkta Dijital Kanalların Tercih Edilmesini Etkileyen Faktörler	23
3. VERİLERİ İŞLEME YÖNTEMLERİ	25
3.1 VERİ KÜMESİ TANIMI	25
3.2 VERİNİN HAZIRLANMASI	26

3.2.1 Özelliklerin Değerlendirilmesi ve Veri Temizliği	26
3.2.2 Verinin Weka Programı İçin Hazırlanması	31
3.3 ALGORİTMA DEĞERLENDİRME KRİTERLERİ	35
3.4 KULLANILAN ALGORİTMALAR	36
4. BULGULAR	39
5. TARTIŞMA VE SONUÇ	48
KAYNAKÇA	50



TABLULAR

Tablo 3.1: Yaş Gruplaması ve Yüzde Değerleri	28
Tablo 3.2: Cinsiyet Yüzde Değerleri	29
Tablo 3.3: Çalışma Şekli Yüzde Dağılımı	29
Tablo 3.4: Maaş Gruplaması Sonrası Yüzde Dağılımı	30
Tablo 4.1: Sınıflandırma Algoritmalarının Karşılaştırılması	43



ŞEKİLLER

Şekil 2.1: Veri Madenciliği Süreci	7
Şekil 2.2: X ve Y Değişkenleri Üzerinden Uygulanan Karar Ağacı	14
Şekil 3.1: Kategori Edilmemiş Ham Data Örneği	25
Şekil 3.2: Kategori Edilmemiş Başlık Eleme Yaptırılmış Data Örneği	27
Şekil 3.3: Yaş Gruplaması Öncesi Dağılım	27
Şekil 3.4: Yaş Gruplaması Sonrası Dağılım	28
Şekil 3.5: Cinsiyet Dağılımı	29
Şekil 3.6: Çalışma Şeklinin Grafikselleştirilmesi	30
Şekil 3.7: Maaş Gruplaması Sonrası Dağılım	31
Şekil 3.8: Data Son Hali	31
Şekil 3.9: Weka Arayüzü Görüntüsü	32
Şekil 3.10: Datanın ARFF Formatı	33
Şekil 3.11: Datanın Weka Görüntüsü	34
Şekil 3.12: Weka Classify Adımı Görüntüsü	35
Şekil 4.1: Naive Bayes Algoritması Sonucu	39
Şekil 4.2: Logistic Algoritması Sonucu	40
Şekil 4.3: J48 Algoritması Sonucu	41
Şekil 4.4: J48 Algoritması Sonucu	42
Şekil 4.5: J48 Algoritması Karar Ağacı Detayları	44

KISALTMALAR

ADK	:	Alternatif Dağıtım Kanalları
ATM	:	Automated Teller Machine
DARPA	:	Defense Advensed Research Project Agency
GNU	:	General Public License
MIT	:	Massachusetts Institute of Tecnology'de
SQL	:	Structured Query Language
SRI	:	Stanford Research Institute
UCLA	:	University of California at Los Angeles
UCSB	:	University of Utah ve son olarak University of California at Santa Barbara
VB	:	Ve Benzeri
WEKA	:	Waikato Environment for Knowledge Analysis
YSA	:	Yapay Sinir Ağları

SEMBOLLER

Yüzde : %



1.GİRİŞ

Veri madenciliği büyük ölçekli verilerden değerli bilginin keşfedilmesidir. Burada amaçlanan birbirleri ile ilişkisi olan verilerin ortaya çıkarılması ile gelecek tahmininin yapılabilmesidir. Veri madenciliğindeki aşamaların işletmelerin sahip oldukları verilerden, nasıl bir bilgi çıkarmak istedikleri, yani amaç ve hedeflerine göre şekillendirilmelidir. Veri madenciliği süreçleri sırası ile uygulanırsa başarı derecesi o oranda artacaktır. Yapılacak madenciliğin hangi sorun için yapılmak istenildiği işletmelerden kesin ve eksiksiz bir şekilde uzmanlar tarafından alınmalıdır. Alınan verilerin anlaşılması ile sorunun doğru bir şekilde anlaşılması birbirini direkt etkiler. Veri anlaşılabilir ve kalitesinin keşfi yapılmalıdır. Büyük ölçekli verilerin sorun çıkarma olasılığı vardır. Bu sebep ile çalışılan verilerin anlaşılması için istatistiksel analizlerden de yararlanılabilir.

Verilerin hazırlanması aşamasında sorunun ne olduğu tanımından yola çıkarak veriler toplanır, bu toplanan veriler içerisinden sonucun farklı yönlerine gitmesine sebep olacak yanlış ya da farklı veriler tespit edilerek temizlenir. Eksik veriler tamamlanır. Birbirlerinden bağımsız kaynaklardan alınan veriler birleştirilir. Doğru olan verilerin kaynaklarının bulunabilmesi çok önemlidir. Verilerde madencilik için kullanılan algoritmaların içerisinde kullanılabilir için model değişikliği yapılabilir. Bu yapılan model değişikliği verilerin genişletilmesi, birleştirilmesi ya da normalize edilmesi şeklinde olabilir. Veriler için bazen de sonucun değişmeyeceğine inanılıyorsa verilerin ve özelliklerinin sayısı azaltılabilir.

Tanımı yapılmış sorun ve hazırlık aşamaları tamamlanmış veriler için en uygun modelin seçilmesi için birden çok model uygulaması yapılabilir. Ayrıca uygun model bulunana kadarda veri hazırlama aşamaları tekrar edebilir. Veriler analiz edilirken en yüksek başarı değeri olan algoritmalar değerlendirilir. Değerlendirme yapılmayan gözden kaçan bir durum olup olmadığına bakılır. Modelin ihtiyaçları karşılaması gerektiği için işlemler gözden geçirilir ve model seçimi tamamlanır. Veri modelinin seçilmesi ve kullanılması bazen yeterli olmayabilir. Ortaya çıkan model hemen kullanılabilir ya da

başka bir uygulamaya temel oluşturabilir. Modeli oluşturan analistler, sorunun sahibine yani müşterilere, modelin kullanılması ile ilgili detaylı bilgi ve rapor vermelidir. Oluşturulan veri modeli zamanla değişen sistemlerin özelliklerine göre düzenlenmelidir. Yoksa model sonuçları etkinliğini kaybedebilir.

Veri madenciliğinde kullanılan yöntem ve teknikler sınıflandırma, kümeleme, birliktelik kuralları şeklindedir. Sınıflandırma algoritmalarının kullanılırken verilerin belirli bir kısmı eğitim amaçlıdır. Bu eğitim amaçlı olan verilerden alınan sonuçlar ile yeni durumlara karşı nasıl karar verilmesi gerektiği bulunur. Kümeleme algoritmalarında veriler benzerlik özelliklerine göre gruplandırılarak ayrıştırılır. Kümeleme, yapılmadan önce karar verilen seçim kriterlerine göre birbirlerine en çok benzeyenlerin aynı grup içinde olmalarını sağlamaktır. Temeli birliktelik kuralı olan algoritmalarda karakteristik olarak birlikte görülen iki ya da ikiden fazla kayıt bulunması amaçlanmaktadır.

Veri madenciliğinde karşılaşılan sorunların birden fazla sebebi olabilir. Bilgilerin sınırlı olması, veriler arasındaki belirsizliğin fazla olması, istenen sonuca götürmeyecek verilerin veri tabanında olması, atık veri şeklinde olması, verilerin ana aktif data da olması ve değişmesi yani dinamik veri olması, veri tabanında olmayan boşluk verilerin olması, veri kümesinin büyüklüğünden kaynaklı eksik verilerin olması, verilerin toplanması ya da verilerin girişinin yapılması sırasında sistem dışı hatalardan kaynaklanan gürültülü ve kayıp değerler olması, farklı tipteki verileri yani kategorik, coğrafi, kesirli verileri değerlendirmeye almak, veri tabanındaki verilerin çok büyük olması, birden çok küçük örneklemin olması karşılaşılan sorunlardır.

Bankalar günümüzde sadece mevduat toplayan ve topladıkları mevduatı kredi olarak kullandıran yapılardan daha geniş işlemler yapmaktadır. Finansal danışmanlık, kredi kartı hizmetleri, otomatik fatura ödemeleri, kiralık kasa hizmetleri, dış ticaret işlemleri, hisse senedi alım satımı gibi işlemler içinde aracılık görevini gerçekleştirir.

Dijitalleşmenin günümüzde çok hızlı bir şekilde gelişmesi bankacılık sektöründe de kendisini göstermektedir. İnternet üzerinden zaman ve mekan kısıtlaması olmadan, kişilerden bağımsız olarak hizmet almalarını sağlamıştır. Nakit para çekimi ve nakit

para yatırılması dışında tüm işlemler dijital kanallar kullanılarak yapılabilmektedir. Ayrıca bankacılık sektöründeki rekabet düşünüldüğünde de bankacılık sektörü operasyonel maliyetlerini minimize etmek için alternatif dağıtım kanallarını tercih etmektedir. Dijitalleşmeye yapılan yatırımların bankalara sağladığı faydaları yeni müşteri kazanımı, eski müşterilerin sadakatinin artırılması, yeni hizmet ve yeni ürünlerin yaratılması, aktif ve hızlı hizmet sunulması, faaliyetlerdeki etkinliğin artırılması, müşteri ile iletişimin kolaylaşması şeklinde söyleyebiliriz. Bankacılık işlemlerindeki dijitalleşme aynı zamanda banka müşterileri içinde avantaj sağlar. Bunları, zaman ve mekandan bağımsız işlem yapmak, banka çalışanlarının dahi görmesine gerek kalmadan gizli ve güvenli işlemler yapabilmek, bilgiye ulaşmak istediğinde detayları kolay bir şekilde bulabilmek, bankanın ürün çeşitliliğinden faydalanabilmek olarak sıralayabiliriz. Bankacılıkta dijital kanalların tercih edilmesini etkileyen en önemli faktör kullanıcıların algılarındaki değişimdir.

Tezin üçüncü bölümünü oluşturan datanın işlenmesi aşamasında özel bir bankanın örtülü verilerinden faydalanılmıştır. Datadaki dijital kanal kullanan müşterilerin özelliklerini belirlemek için verilerden artık ve gürültü yaratan veriler temizlenmiş, datada gruplandırmalar yapılarak da toplanmıştır. Bu sayede data Weka da işlenecek formata getirilmiştir. Weka programı içerisindeki sınıflandırma algoritmaları kullanılarak elde edilen sonuçlar ve algoritmaların içeriği de belirtilerek açıklanmıştır.

Son aşamada sınıflandırma algoritmaları ve sonuçları yorumlanarak yapılan çalışmanın amacından ve daha sonra elde edilmek istenen hedeften bahsedilmiştir.

2. LİTERATÜR TARAMASI

2.1 VERİLER VE VERİ İŞLEME YÖNTEMLERİ

2.2 VERİ MADENCİLİĞİ

Veri, eylem, işlem ya da nesnelere niteliklerinin değerleri ile tanımlayabilen bilgi kümesi olarak ifade edilebilir. İşlemler ya da nesnelere niteliklerden oluşmaktadır. Örneğin nesne insan ise, göz rengi, adı, yaşı onun niteliklerini oluşturur. İnsan nesnelere niteliklerinin oluştuğu kümeler ise verilerdir. Para çekme işlemi örnek alınır, çekilen tutar, saat, yer, hesap numarası bu işlemin niteliklerini oluşturur. İçinde birden fazla para çekme işlemine yönelik bilgiler oluşturduğu kümeler ise verilerdir. Nitelikler nominal ya da sayısal olarak değer alabilirler. İdeal olması beklenen bir veri kümesinde tüm verilerin tutarlı, hatasız ve eksiksiz olması gerekmektedir. Fakat bu durumun gerçek hayatta çok geçerli olmadığı söylenebilir (Coşkun 2010, s.12).

Bu bölümde veri madenciliğinin tanımı, aşamaları, yöntemleri ve teknikleri, ayrıca veri madenciliğinde karşılaşılan sorunlar aşağıda detaylı şekilde ele alınmıştır.

2.2.1 Veri Madenciliği Kavramı

Veri madenciliğinin 1960 yıllarda bilgisayarların veri analiz sorunlarını çözebilmek için kullanılmaya başlaması ile ortaya çıktığı söylenebilir. O dönemde veri taraması, veri yakalaması gibi isimler almış ve bilgisayar desteği ile istenilen sorgulamanın yapılarak istenilen bilgiye ulaşılabileceği varsayılmıştır. 1970'li yıllarda İlişkisel Veri Tabanı Yönetim sistemleri kullanılmaya başlanmıştır. 1980'li yıllarda veri tabanı yönetim sistemleri kullanımı daha da yaygınlaşarak bilimsel alanlarda ve mühendislik alanlarında kullanılmıştır. İşletmeler, rakipleri, müşterileri ve ürünleri ile ilgili bilgilerden oluşan veri tabanları çıkarmışlardır. Bu veri tabanlarının içinde büyük miktarlarda veriler bulunmaktadır. Bu verilere SQL veri tabanı sorgulama dili veya benzer diller kullanılarak ulaşılabilmektedir. 1990'lı yıllarda mühendisler, geleneksel

olan istatistiksel yöntemler yerine algoritmik bilgisayar modülleri ile veri analizlerinin gerçekleştirileceğini savunarak veri madenciliği adını ortaya çıkarmışlardır. Veri tabanlarındaki veri miktarlarının sürekli artması için büyük miktarda yer alan verilerden yararlı bilgilere hangi yolla ulaşılabileceği üzerine çalışmaya başlanmıştır. Veri madenciliği için ilk yazılımda bu şekilde gerçekleştirilmiştir. 2000'li yıllar itibari ile veri madenciliğinin geliştiği ve geniş bir alanda uygulanmaya başlandığı söylenebilir (Çalış 2013, s. 6).

Veri madenciliği, basit bir ifade ile büyük ölçekli veriler arasındaki değeri olan bir bilgiyi elde etmektir. Bu şekilde veriler arasında yer alan ilişkileri ortaya çıkarmak ve gerektiği anda ileriye yönelik tahminlerde de bulunulabilir. Bu anlamda, veri madenciliği bir işletmede üretilen bütün verilerin belirlenen yöntemler kullanılarak gelecekte ortaya çıkabilecek gizli olan kıymetli bir bilginin ortaya çıkartılma süreci olarak belirtilebilir. Bu açıdan, veri madenciliğinin işletmelerin karar destek sistemleri için önemli bir yer edindiği söylenebilir (Özkan 2016, s. 12).

Veri madenciliği, ilk başta keşfedilmemiş örneklerin ortaya çıkarılması amacı ile teknik ve bilimsel verileri inceleyen, veri tabanındaki bilgi keşfi aşamalarından biridir (Rokach & Maimon 2005, s. 2).

Veri madenciliğinin büyük veriler ile başa çıkan yüksek performanslı algoritmalar, paralel hesaplama yetenekleri ve kayıtlı olan verilere hızlı şekilde ulaşılabilme gibi kullanımı kolay özellikleri yayıldıkça kullanımı yaygınlaşmaya başlamıştır. Veri madenciliğinin amacı, bir takım bilgi alanları içindeki özellikle denetlenemeyen büyük tutarlardaki veriyi mantıklı duruma getirmektir. Burada anlatılmak istenen mantıklı duruma getirmek ifadesi, veriye kullanıcı tecrübelerine dayanan değişik anlamlar katabilmeyi belirttiği söylenebilir. Anlamlı duruma gelmesi ve kişilerin buldukları bilgileri avantaja çevrilebilmesi için kazanılan yeni bilginin geçerli, faydalı ve anlaşılır olması gibi asıl niteliklere sahip olması gerektiği söylenebilir (Erkuş 2015, s. 6).

Veri madenciliğinin üç önemli ana temeli olduğu belirtilmektedir (Akküçük 2011, s. 23).

- a) **Büyük miktarda veri:** Bir örnek ile belirtmek gerekir ise bir cep telefonu firmasının yaklaşık kırk milyon abonesi hakkında geçmişe yönelik tuttuğu bütün demografik bilgileri, fatura detayları ve kullanım alışkanlıklarını belirten veridir.
- b) **Potansiyel olarak yararlı bilgi:** Yine bir örnek ile belirtirsek, başka telefon firmalarına geçen müşterilerin özelliklerini analiz ederek, şu an başka firmalara geçme eğiliminde bulunan müşterileri öncesinde belirleyerek geçiş yapmalarını engelleyebilmektir.
- c) **Matematiksel ve istatistiksel teknikler:** Yukarıda ikinci verdiğimiz örneğe devam ederek, kaybedilen bu müşterileri tahmin edebilmek için lojistik regresyon gibi bir istatistiksel yöntem kullanma sürecidir.

Veri madenciliği kullanım alanlarını aşağıdaki gibi sıralayabiliriz (Sivri 2015, ss. 3-4);

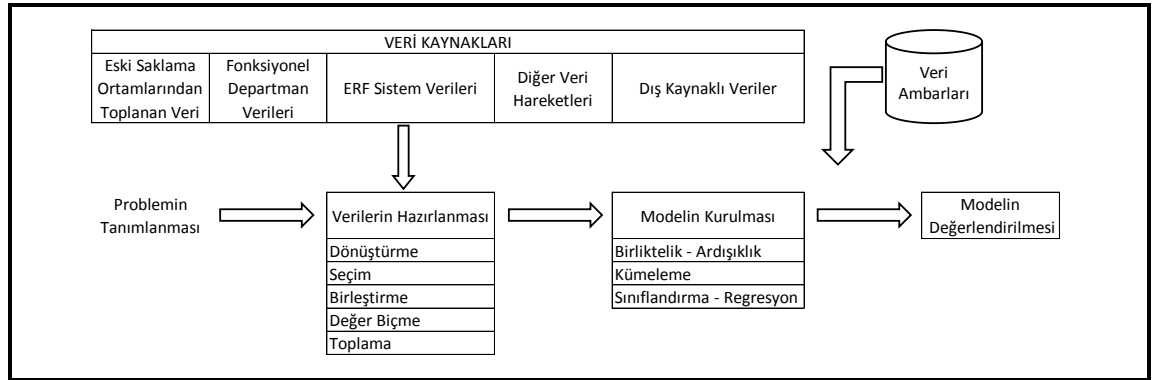
- a) Veri tabanının analizi ve karar verme süreci,
- b) Müşteriler arasındaki benzerliğin bulunması, hedef pazar, sepet analizinin ve çapraz pazarın incelenmesi,
- c) Rekabet analizinin, kalite kontrolünün ve dolandırıcılıkların saptanabilmesi,
- d) Haber kümeleri ve e-posta gibi benzer belgelerin belirlenmesi,
- e) Müşterilerin kredi risk araştırmaları,
- f) Kuruma ait kaynakların en etkin şekilde kullanılması,
- g) Geçmiş ve şuan ki durumu analiz ederek geleceğe yönelik tahminlerde bulunulması,
- h) Müşterilere yönelik satın alma benzerliklerinin çıkarılması,
- i) Müşterilere ait demografik özelliklerin arasındaki bağlantıların belirlenmesi,
- j) Kampanyalarda cevap verme oranının yükseltilmesi,
- k) Var olan müşterilerin elde tutulması ve yeni müşterilerin kazanılması,
- l) Pazar sepeti analizinin gerçekleştirilmesi,
- m) Müşteri ilişkileri yönetimi,
- n) Müşteri değerlendirme süreci,

- o) Satış tahmininin yapılması,
- p) Farklı olan finansal göstergeler arasındaki gizli korelasyonların belirlenmesi,
- r) Kredi kartı dolandırıcılıklarının tespit edilmesi,
- s) Müşteri gruplarının kredi kartı harcamalarına göre gruplandırılması,
- t) Kredi taleplerinin incelenmesi,
- u) Yeni poliçe taleplerinde bunabilecek olan müşterilerin öngörülmesi,
- v) Sigorta dolandırıcılıklarının tespit edilmesi.

2.2.2 Veri Madenciliğinin Aşamaları

İşletmeler, veri madenciliği sürecini belirlerken kendi verilerine, kaynaklarına, iç yapı ve amaçlarına uygun şekilde belirlerler. Veri madenciliğinde başarı sağlamanın ilk şartı uygulamanın, işletmenin hangi amacına yönelik yapılacağını net bir biçimde belirlemesine bağlıdır. Farklı sektörlerde biriken veriler veri tabanlarında toplanır. Bu verilerin içinde bulunan gizli bilgilerin ortaya çıkarılması zordur. Alt işlemlerden oluşan bu süreç 1996 yılında The Cross - Industry Standart Process for Data Mining tarafından çıkarılmıştır. Veri madenciliğinin aşağıda Şekil 3.1’de belirtilen süreçlerden meydana geldiği söylenebilir (Ersöz 2013, s. 26).

Şekil 2.1: Veri Madenciliği Süreci



Kaynak: Ersöz, F., 2013. Veri Madenciliği ve Uygulamaları. 1.Baskı. Ankara: Sage Yayıncılık.

Amaç ve nitelik ne olursa olsun veri madenciliği süreci ile yapılan bütün işlemler yukarıda belirttiğimiz Şekil 2.1’de ki aşamaların tamamından geçmesi gerekir.

2.2.2.1 Sorunun Belirlenmesi

Sorunun belirlenmesi veri analizleri işlemlerinin önemli adımlarından biridir. Uygulamaların hangi kuruluş ya da işletmeye yönelik yapılacağı kesin olarak belirlenmelidir. Soruna odaklanmalı, sorunun ne olduğu doğru şekilde bulunmalı ve sonuçların başarı derecesini ne kadar ne oranla değiştireceği kesin olarak belirtilmelidir. Amaç, sorun ile örtüşmez ise o çalışmadan netice alınmaz. Ayrıca bu hatanın başka bir hatanın doğmasına da sebep olabileceği söylenebilir. Hatalı belirtilen amaç, hatalı kararlar alınmasına neden olarak maliyeti de arttırır. Doğru belirlenen amaç doğru kararlar almayı ve kazanılacak olan faydaları belirler. Amaç ve fayda sorun aşamasında belirlenmelidir (Odabaş 2017, s. 18).

Sorunların araştırılması sürecinde uzmanlık ve iş deneyimi önemlidir. Öncelikle projenin amaç ve gerekliliği belirlenmeli ve iş perspektifinin bulunması gerekmektedir. Bu bilginin veri madenciliği sorun tanımlı olduğu netleştirilmeli ve hedeflere ulaşılma amaçlı planların yapılması gereklidir. Sorunun belirlenmesi amacı, veri madenciliğinin amacını, var olan durumun değerlendirilmesi ile planlama sürecinin belirlenmesinden oluşmaktadır. Sorunun belirlenmesi aşamasında ihtiyaçlar net olarak çıkarılmalıdır. Bu süreçte amaçlar belirlenirken dikkat edilmesi gereken performans ölçütlerinin neler olduğunu ve süreç sonunda doğacak olan sonucun hangi durumlar için kullanılacağına karar verilmelidir. Veri madenciliğindeki başarısızlıklar genellikle sorunlar için kısa ve uzun vadeli amaçların kesin olarak belirlenmemesinden kaynaklandığı söylenebilir. Veri madenciliği uzmanı, hedefleri kesin bir dille, işletmenin hedefini ve veri madenciliğinin nasıl bir etkisinin olacağını belirterek açıklamalıdır (Ersöz 2013, s. 27).

2.2.2.2 Verinin Anlaşılması

Verilerin anlaşılmasında ilk önce veriler toplanır, var olan verilerin uygunluğu değerlendirilir, modelin oluşturulması için gerekli olan verilerin ihtiyaçları tespit edilir ve mevcut olan kayıt sayılarının yeterliliği bu aşamada gerçekleştirilir.

Veriyi anlamak ve işi anlamak birbirleri ile ilişkili iki alt süreçtir. İş anlamaya çalışarak farklı verilere bakmak ve anlamak iş ile ilgili farklı bakış açıları kazandırabilir. Bu

durum kendi içerisinde devam ettikçe veriler netlik kazanmaktadır. Verinin anlaşılmasında aşağıdaki işlemler yapılır.

- a) Başlangıç verilerinin toplanması
- b) Verinin tanımlanması
- c) Verinin keşfedilmesi
- d) Verilerin kalitesinin belirlenmesi

Büyük hacimli verilerde büyük problemler ortaya çıkabilir. Bu sebeple küçük veri gruplarında, simülasyon ortamlarında hazırlanan veri madenciliği sistemleri, gürültülü, eksik, büyük hacimli ya da bilinmeyen veri kümelerinde hatalı çalışabilir. Bu sebeple veri madenciliği sistemleri hazırlama aşamasında bu problemlerin çözülmesi şarttır (Yurdakul 2015, s. 24).

Verilerin anlaşılması verinin toplanması ile başlar. Verileri tanımlayabilmek için bazı süreçlerden geçmesi gerekir. Verilerin kavranması ile işin kavranması arasında yakın bir ilişki bulunmaktadır. Veri madenciliği sorununun tasarımı ve projelerin gereklilikleri fazla olmasa da kullanılacak veriler hakkında bir anlayış gerektirebilir. Bu süreçte istatistiksel analizlere bakabilme gibi verilerin keşfine yönelik bazı işlemler yapılabilir. Ayrıca bu süreçte kümeleme gibi bazı modeller verilerin içindeki örüntüleri anlayabilmek için veri setlerine uygulanabilir (Uyumaz 2017, s. 39).

2.2.2.3 Verinin Hazırlanması

Verilerin hazırlanması aşaması örneklem kümeleri elde edildikten sonra, bu kümelere yer alan yanlış kayıtların bulunduğu ve eksik olan kalite değerlerinin değiştirildiği bölümdür. Bu aşamanın tercih edilen veri madenciliğinin çalışma süresini iyileştirdiği söylenebilir. Modelin kurulması sürecinde ortaya çıkabilecek problemler bu aşamaya sıklıkla geri dönülerek verilerin tekrar iyileştirilmesine sebep olmaktadır. Verilerin hazırlanması aşaması veri madenciliğinin en önemli aşamasıdır. Analistler toplam zamanlarının yüzde elli – yüzde seksen beşini kapsadığı belirtilebilir. Verilerin hazırlanması belirli aşamalardan oluşmaktadır. Bunlar (Çalış 2013, ss. 14-15);

- a) **Verilerin Toplanması:** Oluşan problem için gerekli olan verinin ve bu verinin toplanacağı veri kaynaklarının belirlenmesidir. Bu aşamada analist, kendi veri kaynakları dışında farklı veri kaynaklarından da faydalanabilir.
- b) **Veri Birleştirme ve Temizleme:** Toplanan verilerin arasında farklılık oluşturan, yanlış ya da analizin hatalı yön almasına neden oluşturacak verilerin temizlenmesi bu aşamada gerçekleştirilir. Bu durum da veri madenciliği süreci hızına ve doğruluğuna fayda sağlar. Veri temizleme işlemi sayesinde verilerdeki eksik değerler doldurulur, hatalı değerler yok edilir ve tutarsızlıklar düzeltilir. Veriler birden fazla formdan ve farklı farklı kaynaklar üzerinden gelebilirler. Doğru olan veri kaynaklarını tespit edebilmek ve ihtiyaç duyulanlardan veri oluşturarak bunları bir araya getirmek önemli bir konudur (Berry & Linoff 2004, s. 28).
- c) **Verinin Dönüştürülmesi:** Kullanılacak olan algoritma ve model çerçevesindeki verilerin tanımlama sürecinin değiştirilmesine ihtiyaç duyulabilir. Bu aşamada veriler madencilik için gerekli olan formlara dönüştürülebilir. Veri dönüştürmesi düzleştirme, birleştirme, genelleştirme, normalizasyon ve alan yapılandırma aşamalarından oluşur.
- d) **Verinin İndirgenmesi:** Geniş veri tabanları ile gerçekleştirilen veri madenciliğinde çözümlene işlemleri uzun sürebilir. Özgün verinin bütünlüğünün korunarak, elde edilecek olan sonuçlarda değişiklik olmayacağına inanılıyor ise veri sayısı veya değişkenlerin sayısı minimuma indirilebilir. Bu durumda verilerin indirgenmesi olarak belirtilebilir.

2.2.2.4 Veri Modelinin Kurulması ve Değerlendirilmesi

Bu bölümde seçilmiş olan veri madenciliği sisteminin belirlenen verilere uygulanarak sonucunda da model oluşturmaktır. Uygun olan veri madenciliği sisteminin belirlenmesindeki gibi, sorun ve veri seti net bir şekilde anlaşılabilir farklı algoritmalar denenmeli ve modellerin oluşturulması gerekmektedir. Algoritmalarından elde edilen sonuçlar sonrasında en doğru sonucu çıkaran algoritmanın belirlendiği aşama değerlendirme aşamasıdır. Algoritmalarından en iyi sonucu veren algoritmayı belirlemek

için farklı istatistiksel yöntemler kullanılır. Bu istatistikler sonucu en doğru algoritmayı veren model tercih edilerek oluşturulan modelle yeni durumlar için oranlamalarda bulunulur (Çığışar 2017, ss. 26-27).

Tanımlanan sorun ve hazırlanan veri için en doğru modelin kurulması birden fazla modelin kurularak denenmesi ile mümkün olduğu söylenebilir. Model kurmanın yanı sıra veri hazırlamanın en uygun olduğu belirtilen modelin oluşturuluncaya kadar tekrar eden bir süreç olduğu belirtilebilir (Toksöz 2016, s. 13).

Veri analizinde en yüksek kaliteye sahip modeller oluşturulduktan sonra, işletmeler tarafından model değerlendirilir. Model için çalıştırılan adımlar kontrol sürecinden geçer. Buradaki amaç yeteri kadar gözlemlenemeyen önemli iş konularının var olup olmadığının tespit edilmesidir. Değerlendirme sürecinde sonuçların değerlendirilmesi, işlemlerin gözden geçirilmesi ve sonraki adımların belirlenmesi aşamaları yer almaktadır. Bunlar aşağıda kısaca ele alınmıştır (Mocan 2016, s. 7);

- a) **Sonuçların Değerlendirilmesi:** Bu bölümde, modelin doğruluğuna yönelik faktörler ile ilgilenilir. Amaç modelin değerlendirilmesi ve modelin neden tam olmadığı araştırılmasıdır. Ayrıca bütçe ve zaman kapsamında uygulamanın gerçek hayatta denenmesi denilebilir.
- b) **İşlemlerin Gözden Geçirilmesi:** Ortaya çıkacak olan modelin tatmin edici olması ve ihtiyaçları karşılaması gerekmektedir. Bu bölümde veri madenciliği bakımından daha doğru kontrol sağlandığı, kaçan faktörlerin ya da görevlerin olup olmadığı araştırılmaktadır. Bunun ayrıca modellerin doğruluğu konusunda sigortası olduğu belirtilebilir.
- c) **Sonraki Adımların Belirlenmesi:** Bu aşama projenin sonlandırılma aşamasıdır. Bu aşamadan sonra yayılım aşamasına geçilir ya da yeni bir veri madenciliği projesinin başlamasına imkan sağlanır.

2.2.2.5 Veri Modelinin Kullanılması

Modelin oluşturulması projenin sona erdiği anlamına gelmemektedir. Modelin amacı, verinin bilgi değerini arttırmak dışında kazanılmış olan bilgilerin organize edilerek

müşterilere yararlı olacak biçimde sunulması gerekmektedir. İhtiyaçlara göre uygulama süreci bir rapor üretebilmek kadar basit olsa da bir veri madenciliği süreci uygulayabilmek kadar karışık olabilir. Çoğu durumda, uygulama adımlarını yürütecek kişi veri analisti değil müşteridir. Analist gerekli modeli kullansa da müşterilerin yaratacağı modelleri kullanabilmek için yapılması gereken prosedürleri anlamak ayrıca önemli bir noktadır (Karahan Adalı 2017, s. 18).

Kurulmuş olan model direk hazır uygulamada olabilir veya diğer uygulamaların alt bir aşaması olarak da kullanılabilir. Kurulan modeller ayrılan müşterilerin analizi, kaçak olan müşterilerin tespit edilmesi gibi birçok alanda kullanılabilir ya da başka işletmelerin var olan uygulamalarında entegre edilebilir (Odabaş 2017, s. 20).

2.2.2.6 Veri Modelinin İzlenmesi

Zamanla tüm sistemlerin özelliklerinde ve ürettikleri verilerde ortaya çıkabilecek olan değişiklikleri, kurulan modelin sıklıkla izlenmesini ve gerekirse tekrar düzenlenmesini gerektirebilir. Tahmin edilen ve fark edilen değişkenlerdeki farklılığı gösteren çizgelerin model sonuçlarını kullanılabilen faydalı bir yöntem olduğu söylenebilir (Çil 2010, s. 12).

2.2.3 Veri Madenciliğinin Yöntemleri / Teknikleri

Veri madenciliği yöntemleri sınıflandırma, kümeleme ve birliktelik kuralları olarak gruplandırılmaktadır. Bu gruplar aşağıda detaylı şekilde ele alınmıştır.

2.2.3.1 Sınıflandırma

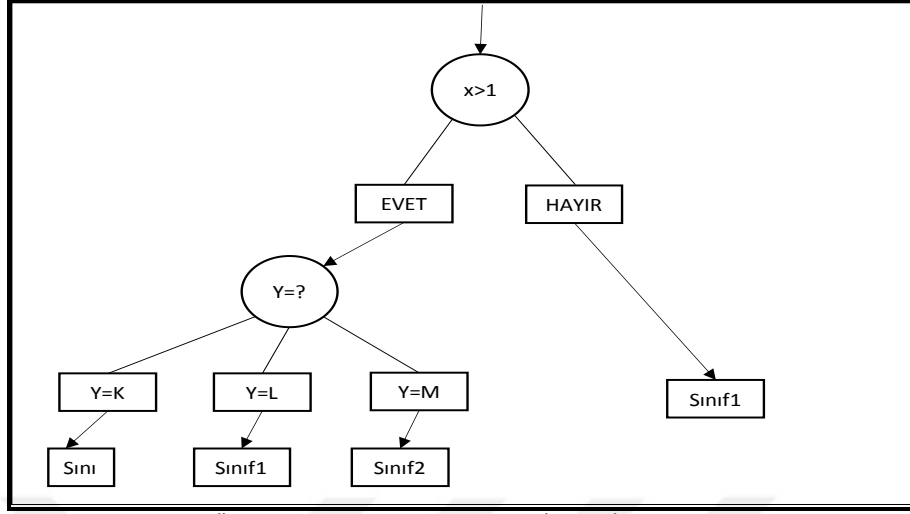
Sınıflandırma veri madenciliğinde sıklıkla kullanılan bir yöntemdir. Veri tabanlarındaki gizli olan verileri ortaya çıkarmakta kullanılır. Verilerin sınıflandırılabilmesi için belli bir süreç izlenmektedir. İlk önce mevcut olan veri tabanının belirli bir bölümü eğitim amacı ile kullanılarak sınıflandırma kurallarının neler olacağı belirlenir, sonrasında

belirlenen bu kurallar yardımı ile çıkacak olan yeni durum hakkında nasıl karar verileceği belirlenmektedir (Özkan 2016, s. 32).

Sınıflandırma veri madenciliğinde bilinen ve en çok kullanılan tekniklerden biridir. Kalite kontrol çalışmaları, dolandırıcılık tespiti, hastalık tanıları ve pazarlama vb. konular sınıflandırma tekniklerinde en çok kullanılan alanlar olduğu söylenebilir. Sınıflandırma tekniği tahminleyici bir modeldir. Örneğin bir kutuda aynı renkten kaç adet top olduğunun tahmin edilmesi bir sınıflandırma işlemidir. Sınıflandırma kendi içerisinde karar ağacı, yapay sinir ağları ve genetik algoritmalar olarak ayrılabilir (Silahtaroglu 2016, s. 67).

a) Karar Ağaçları: Verilerin sınıflandırma işlemlerinin yapılmasında çoğunlukla kullanılan bir algoritmadır. Ağaç görüntüsünde olan tahmin edici bir yöntem olduğu söylenebilir. Veri tabanında yer alan tüm kayıtlar bu ağaç içerisinde dallandırılarak çıkan sonuca göre sınıflandırılır. Karar ağacında ağaçların her bir dalı sonraki süreçte hayır ya da evet şeklinde iki dala ayrılmaktadır. Ağaç üzerindeki tüm dalların bir olasılığı bulunmaktadır. Bütün dalların olasılıkları bulunduğundan dolayı bu tüm dalların birbirleri ile olan olasılıkları hesaplanabilir. Bütün ağaçların hata olasılık oranları ve hesaplama verimliliğinin artırılması için ağaçların bir kısım dallarının kesilerek yani faydası olmayan kurallar çıkarılarak artırılabilir. Verilerin sınıflandırılmasında karar ağacı iki aşamalı bir işlemde gerçekleştirilir. Bunlar öğrenme ve sınıflandırma işlemleridir. Öğrenme aşamasında var olan test verisi, bir model oluşturmak için sınıflandırma algoritmaları kullanılarak çözümlenir. Çözümlenen model sınıflandırma veya model kuralları olarak gösterilebilir. Sınıflandırma evresinde eğitim verisi, karar ağacının veya sınıflandırma kurallarının doğruluğunun belirlenmesi için kullanılmaktadır. Çıkan sonuçların doğruluğu kabul edilebilir derecede ise, oluşan kurallar elde edilen yeni veri kümesini sınıflandırma yapabilmesini sağladığı belirtilebilir (Tekingöz 2016, ss. 16-17).

Şekil 2.2: X ve Y Değişkenleri Üzerinden Uygulanan Karar Ağacı



Kaynak: Tekingöz, Ö. (2016). Dijital Yayıncılıkta İçerik İzleme Oranlarına Göre Müşteri Kümelenmesi.

Yukarıdaki Şekil 2.2’de görüldüğü gibi X ve Y’den oluşan yalın bir karar ağacı bulunmaktadır. Y’nin değerini göz önünde bulundurmadan $X \leq 1$ ve $Y=K$ ve $Y=M$ koşullarını sahip olan örnekler sınıf 2’de $Y=L$ koşuluna sahip olan örnekler sınıf1’de yer aldığı görülmektedir.

- b) Yapay Sinir Ağları:** Günümüzde insan oğlunun doğayı taklit etme ve araştırma çabalarının en yaygın ürünlerinden birinin yapay sinir ağları teknolojisi olduğu söylenebilir. YSA (Yapay Sinir Ağları), sade bir biyolojik sinir sisteminin çalışabilme şekline simüle edebilmek için tasarlanmış programlardır. YSA, simüle edilen nöronları içerir ve bu nöronlar farklı biçimlerde birbirlerine bağlanarak ağ oluştururlar. Bu oluşturulan ağların ise hafızaya alma, öğrenme ve veriler arasındaki ilişkinin ortaya çıkarılma gücüne sahip oldukları söylenebilir. YSA, ağırlıklandırılmış biçimde birbirlerine bağlı olan ve birden fazla nöronlardan oluşan matematiksel sistemleri oluşturmaktadır. Bir işlem elemanının, sıklıkla transfer işlevi olarak hatırlanan bir denklem olduğu belirtilebilir. Bu işlem elemanı, farklı nöronlardan gerekli sinyalleri alarak bunları dönüştürür, birleştirir ve sayısal bir veri ortaya çıkarmaktadır. Genellikle işlem elemanlarının gerçek nöronlara karşılık geldiği söylenebilir ve bir ağ içerisinde birbirlerine bağlanarak bu yapıdaki sinir ağlarını oluştururlar (Karadeniz 2008, s. 93).

- c) **Genetik Algoritmalar:** Genetik Algoritmalar doğal yöntemlerin sistemlerini açıklayabilmek için geliştirilen arama algoritmalarıdır. Bu algoritmalarda çözümün değişkenleri, genlerin bir verisini sunan yapılandırılmış bir sisteme kodlama aşamasıdır. En iyinin hayatta kalabilmesi prensibine göre bütünsel olan en iyi çözümün ne olacağını araştırmaktadırlar. Genetik algoritmalar sorunlara bir tek çözüm bulmak yerine farklı çözümlerden oluşabilecek bir çözüm kümesi üretmeyi amaçlamaktadırlar (Karahana Adalı 2017, s. 21).

2.2.3.2 Kümeleme

Kümeleme, veri madenciliğinin en temel yöntemlerinden biridir. Bu yöntemde amaç, veri tabanında yer alan verileri benzerlik ve özelliklerine göre gruplandırarak ayırtmaktır. Kümeleme analizi, öncesinde belirlenen seçme kriterlerine göre birbirlerine en çok benzeyen nesne ya da bireyleri aynı grup içerisinde gruplandırabilmektir. Analizin sonucunda bir grubu oluşturabilen elemanlar birbirleri ile benzerlik sağlarken, başka kümelerin elemanlarından farklılık gösterebilirler. Oluşturulan bu gruplar çok kapsamlı uzayda gösterilmek istendiğinde eğer kümeleme başarı sağlar ise aynı küme içerisinde yer alan nesne ya da bireylerin birbirlerine yakın çıkmaları, bununla beraber başka kümelerinde birbirinden fark edilebilir derecede uzak olması beklenmektedir. Kümeleme sınıflandırmada olduğu gibi hedef değişkenliği içermez. Hedef değişkenliği içermediği içinde sınıflandırma ya da tahmin yapılamaz. Verilerdeki benzerlikler göz önünde bulundurularak veri setinde meydana gelen bağdaşık alt kümeler oluşturulması amaçlandığı söylenebilir. Kümeleme yöntemleri grupların nasıl ayrıldıklarına yönelik hiyerarşik ve hiyerarşik olmayan kümeler olarak kendi içerisinde ikiye ayrılmaktadır. Bu iki küme yöntemi aşağıda kısaca ele alınmıştır (Boyacı 2017, s. 27).

- a) **Hiyerarşik Kümeleme:** Bu yöntemde objeler arasındaki uzaklık ve yakınlık ilişkilerine göre kümelerden bir ağaç oluşturulur. Hiyerarşik kümelemeye yönelik özellikler aşağıdaki gibi belirtilebilir.
- i. Veri seti birden fazla kümeye ayrılabilir. Bu ayırımı dendogram olarak adlandırılan ağaç sayesinde oluşturulmaktadır.

- ii. Dendogramın kurulumundaki sıralama gövdeden yapraklara doğru ya da yapraklardan gövdeye doğru olabilir. Dendogramlar istenildiği zaman sonlandırılarak kümeler oluşturulabilir.

Yapraktan gövdeye doğru olan hiyerarşik kümeler aşağıdaki gibi sıralanabilir.

- i. Başlangıcı her obje için farklı farklı grupların oluşturulması,
- ii. Birleştirilme işlemleri bir takım kurallara dayanır. Bu kurallara örnek olarakta ortalama, merkezler arası uzaklık vb. durumlar belirtilebilir.
- iii. Birleştirme işleminin bitirilmesi gruplanacak başka bir kümenin olmaması durumunda gerçekleşir. Farklı bir deyişle, elde bulunan tüm objeler sadece bir küme içerisinde kalana kadar ya da talep edilen küme sayısı elde edilene kadar birleştirilme işlemi devam eder.

Gövdeden yaprağa doğru olan hiyerarşik kümeler aşağıdaki gibi sıralanabilir.

- i. Başlanacak olan noktanın, aynı küme içerisinde bulunan objeler ile başlaması,
- ii. Kümelerin daha küçük kümelere ayrılması,
- iii. Sonlandırılma durumuna kadar işlemin sürdürülmesi farklı bir deyişle talep edilen küme sayıları elde edilene kadar ya da her obje ayrı bir küme oluşturabilene kadar işlemin devam edilmesidir (Sivri 2015, s. 16).

b) Hiyerarşik Olmayan Kümeleme: Hiyerarşik olmayan kümelemede birimlerin uygun olan kümelerde toplanarak n birimin k kümeye bölünebilmesinin hedeflendiği belirtilebilir. Küme sayıları hakkında öncesinde bilinen bir bilgi var ise ya da küme sayısına karar verilmişse hiyerarşik olmayan kümeleme tercih edilebilir. Hiyerarşik olmayan kümelemede birden fazla teknik kullanılmaktadır. Bunlardan en çok kullanılanı k -ortalamalar yöntemi olduğu söylenebilir (Boyacı 2017, s. 28).

2.2.3.3 Birliktelik Kuralları

Birliktelik kuralları genelleştirilmiş prediktif yöntemler olarak da adlandırılabilir. Bu yöntemde belirli olan bir sonuç bir dizi kural ile ilişkilendirilmektedir. Karakteristik olarak beraber oluşan iki ya da daha çok kaydın bulunulması çalışılır. Apriori ve GRI olmak üzere iki ilişkiyel kural belirleme tekniği bulunmaktadır. Çoğunlukla prediktif

modellemeden önce sıklıkla olmasa da kümeleme uygulamalarında kullanılan bir veri indirgeme yöntemi olan Temel Bileşenler Analizi Faktör algoritması bulunmaktadır. Apriori ve GRI algoritmaları ilk başta çok basit kuralları oluştururlar. Sonrasında basit kurallar incelenerek karışık kurallar kaydedilir. Birliktelik kuralları, sepet analizi olarak da bilinmektedir. Olayların beraber olma kuralları belli olasılıklar ile ortaya çıkarır. Bu algoritmalar başka kurallara bağımlı olmayan ve herhangi bir çıktı ile sınırlı olmayan kurallar üretirler. Bu algoritmalar bir müşterin bir markette bir seferde yapmış olduğu alışverişteki peynir, zeytin ve ekmek gibi ürünlerden birkaç tanesini aynı anda alma kurallarını keşfederek, bir ya da birden fazla alanın değerlerinin birlikte ortaya çıkmasını açıklamaktadır. Bu durum pozitif ilişkileri ortaya çıkardığı gibi negatif ilişkileri de ortaya çıkarabilir (Ersöz 2013, s. 51).

2.2.4 Veri Madenciliğinde Karşılaşılan Sorunlar

Veri madenciliğinde karşılaşılan sorunların birçok sebepleri bulunmaktadır. Bunlar aşağıdaki şekilde belirtilebilir (Yurdakul 2015, s. 25);

- a) **Sınırlı Bilgi:** Veri tabanları genellikle kolay olan öğrenme işlerini oluşturan özellik ya da nitelikleri kaliteli sunabilmek gibi veri madenciliği dışında olan amaçlar için hazırlanmaktadır. Bu sebeple, öğrenme görevini kolaylaştırabilecek birtakım özellikler olmayabilir.
- b) **Belirsizlik:** Hataların hızı ve verilerdeki gürültünün derecesini ifade etmektedir.
- c) **Atık Veri:** Karşılaşılan problemlerde beklenen sonucu elde edebilmek için kullanılan örneklem kümelerindeki gereksiz olan niteliklerdir.
- d) **Dinamik Veri:** Kurumsal olan çevrimiçi veri tabanları dinamiklerdir. Bunların içeriği sıklıkla değişebilir.
- e) **Boş Veri:** Bir veri tabanındaki boş değer, birincil anahtarda bulunmayan herhangi bir niteliğin değerini ifade edebilir.
- f) **Eksik Veri:** Veri kümelerinin doğasından veya büyüklüğünden kaynaklanmaktadır. Eksik verileri içeren kayıtların olduğu tespit edildiğinde kayıtlar çıkarılabilir ve değişkenlerin ortalamaları eksik verilerin yerine kullanılabilir.

- g) **Gürültülü Ve Kayıp Değerler:** Verilerin girişi ya da verilerin toplanması sürecinde oluşan sistem dışı hataları ifade etmektedir.
- h) **Farklı Tipteki Verileri Ele Alma:** Günümüzde kullanılan uygulamalar makine öğrenimlerindeki gibi sadece kategorik ya da sembolik veri türleri değildirler. Ayrıca kesirli sayılar, tam sayı, coğrafi bilgi ve çoklu ortam verisi gibi farklı farklı tiplerdeki veriler üzerinde işlem yapılmasını gerektirmektedir.
- i) **Veri Tabanı Boyutu:** Veri tabanı boyutları hızlı bir şekilde artmaktadır. Bu algoritmanın birden fazla küçük örnekleme içerebilecek şekilde geliştirdiği söylenebilir. Çok büyük örneklemlerde kullanılabilmesi için çok dikkat edilmesi gerekmektedir.

2.3 BANKACILIKTA DİJİTAL KANALLAR

2.3.1 Bankacılığın Tanımı

Bankacılığın ilk olarak 15. ve 16. yüzyılda sarraf dükkanlarının ve kasalarının gelişmesi ile ortaya çıktığı söylenebilir. Günümüzde bankacılık 17. yüzyıl başlarında yapılmaya başlanmıştır. 18. yüzyılda ise bankaların ticari senetlerinin iskontoları karşılığında banknot ihraç etme imkanı elde etmişlerdir. Ayrıca hisse senedi ve tahvil ihracına aracı olmaları ve elde ettikleri gelirler ile kazandıkları mevduatları firmaların kullanımına izin vermeleri, bankacılık sektörünün ülkelerin ekonomileri için vazgeçemeyecekleri bir yapı haline getirdiği söylenebilir. 19. yüzyılda ise bankalar ticari faaliyetlerin ve ekonominin yardımcısı hatta bu etkinlikleri geniş ölçüde derleyici kurumlar haline almış ve etkinlik alanlarına göre uzmanlaşmıştır. Modern anlamda bankacılık faaliyetlerini gösteren ilk banka 1609 yılında Amsterdam Bankası olarak kurulmuştur. İkinci banka 1637 yılında Venedik Bankası olarak, 1694 yılında ise ülkeye para sağlamak amacı ile kurulan İngiltere Bankası, emisyon tekelini sahip olmamakla beraber, ülke kontrolünde özel bir banka olarak çağdaş Merkez Bankalarının ilk örneğini oluşturur (Demirel 2017, s 11).

Banka kelimesinin kökeni İtalyanca masa, tezgah anlamına gelen “bonco” kelimesinden gelmektedir. Bankalar, ticari veya bireysel müşterilerden mevduat toplayarak bu

mevduatları en verimli biçimde farklı kredi işlemlerinde kullanmaktadırlar. Fakat günümüzde bankalar mevduat toplama dışında farklı konularda da hizmet vermektedirler. Bankalar mevduat toplarlar, kredi verirler, aracılık yaparlar, para ve kredi politikalarının uygulanmasında destek olurlar, kiralık kasalarda müşterilerin menkul kıymetlerini korurlar, ülkenin kalkınmasına destek olurlar ve müşterilerinin hayatlarını kolaylaştırabilecek kredi kartı, otomatik ödeme vb. birçok ürün sunmaktadırlar. Bankaların, para, sermaye ve kredi konularında her işlemi gerçekleştirebilen özel veya kamusal kişiler ile firmaların bu alandaki bütün ihtiyaçlarını karşılama faaliyetlerini kendine görev edinen ekonomik bir birim olduğu söylenebilir. Bankaların genel faaliyet konularını aşağıdaki gibi sıralayabiliriz (Ulukuş 2011, ss. 13-14).

- a) İşletmelere kredi vererek destek sağlama,
- b) Farklı farklı araçlar yardımı ile para yatırma olanağı sağlama,
- c) Müşterilere tüketici kredileri ile destek olma,
- d) Dış ticaret işlemlerine aracılık etme,
- e) İnternet bankacılığı hizmeti sunma,
- f) Kiralık kasa hizmeti,
- g) Hisse senedi alım satımında aracılık etme,
- h) Kredi kartı hizmeti verme,
- i) Finansal danışmanlık hizmeti sağlamadır.

Bankalar mevduat kabul ederek bu mevduatları en etkin ve verimli biçimde farklı kredi işlemlerinde kullanmak amacı güder. Kısaca, faaliyetlerinin ana konusu düzenli şekilde kredi almak ya da vermek olan bir ekonomik kuruluş olduğu söylenebilir. Ekonomik bakımdan banka tanımı; ekonomiyeye banknot ve kaydi para gibi ödeme seçenekleri yaratan, nakti sermaye ve sermayeyi taşınır değerler ile ticareti alışılmış iş olarak devam ettiren ve nakit kullanmadan yürütülen finansal hizmet ve ödeme işlemlerini gerçekleştiren kamu ya da özel işletmelerdir (İşler 2015, s. 17).

2.3.2 Bankacılıkta Dijital Kanalların Tanımı

İnternet, 1962 yılında Amerika'nın Massachusetts Institute of Technology'de (MIT) üniversitesinde Galaktik Ağ olarak J.C.R. Licklider tarafından bulunmuştur. Licklider, bunun sonucunda küresel şekilde bağlanmış bir sistemden talep edilen herkesin her yerden verilere ve programlara ulaşabilmesini ifade etmiştir. 1965 yılında bilgisayarların birbirleri ile ilk defa konuşabilmesini MIT'de görev yapan Lawrence Roberts ve Thomas Merrill ortaya çıkarmıştır. 1966 yılının sonlarına doğru Roberts, Defense Advanced Research Project Agency (DARPA) da görev almış ve Arpanet adında bir projeyi desteklemiştir. Arpanet ile ilgili ilk bağlantı 1969 yılında dört merkez ile gerçekleştirilmiş ve ana bilgisayarların arasında yer alan bağlantılar ile internete yönelik ilk şekil ortaya çıkarılmıştır. Arpanet'i hayata geçiren dört merkez University of California at Los Angeles (UCLA), Stanford Research Institute (SRI), University of Utah ve son olarak University of California at Santa Barbara (UCSB) olmuştur. Kısa bir sürede çoğu merkezlerde bulunan bilgisayarların Arpanet ağına bağlanarak 1971 yılında ağ kontrol protokolü olan bir protokol ile etkinlik göstermeye başladığı belirtilebilir. Ağ kontrol protokolünden daha çok farklı imkanlar getirebilecek yeni bir protokol, 1983 yılında iletişim kontrol protokolü adı ile Arpanette kullanılmaya başlanmıştır. İletişim kontrol protokolü şu anki günümüzde mevcut olan internet ağının merkez halkası olarak kullanıldığı söylenebilir. Amerika Birleşik Devletleri'nde 1990' lara kadar devletin desteği ile gelişmekte olan internet 1995 yılı itibari ile tamamen özel işletmecilerin yönetimine geçmiştir (Dikkatli 2016, ss 40-41).

Bankalar gelişen teknoloji ile beraber hizmet sunan sektörlerdir. Bankacılık sektörlerinde internetin sağladığı faydaları etkili ve verimli bir biçimde kullanan bankalar dijital kanal hizmetlerinde ve servis alanlarını yaygınlaştırdıkları söylenebilir. Dijital kanallar elektronik bankacılık sektörünün bir gelişim uzantısı olan ve açık ağ sistemlerinin kullanılarak ticari ve bireysel bütün bankacılık işlemlerinin yapıldığı, bankaların sanal ortamda kurdukları şube dışı hizmetlerinin verildiği alternatif bir dağıtım kanalı olarak tanımlanabilir. İnternet bankacılığı ilk olarak 1980 yılında telefon bankacılığı ile ortaya çıkmış ve her yerde kolay kullanımı ile hızla geliştiği söylenebilir. İnternet bankacılığı kişilerin kendileri tarafından, mekan ve zamandan bağımsız şekilde

yapılabilen hizmetleri sağlamaktadır. Günümüzde zaman faktörünün değeri göz önünde bulundurulursa şirketlerin bankacılık faaliyetlerini gerçekleştirebilmeleri için ayrı bir mesai harcamalarına gerek duyulmamaktadır. Bankacılık sektörü, elektronik ticaretin en çok kullanıldığı sektörlerden biri olduğu söylenebilir. Teknolojik gelişmeler sonucunda çıkan ve genellikle şubesiz bankacılık uygulamaları olarak belirtilen şubesiz bankacılık faaliyetleri içerisinde internet şubesi ve telefon bankacılığı örnek olarak gösterilebilir. İnternet bankacılığında para yatırma ve çekme işlemleri dışında birçok işlem yapılabilir. İnternet bankacılığı ile şubelerde yapılabilen işlemlerin hemen hemen tamamı yapılabilmektedir (Özbal 2011, ss. 36-38).

Günümüzde teknolojinin artması ile beraber oluşan rekabet ortamında, bankalar müşterilerine daha fazla tercih sunmaya çalışırlar. Dijital Bankaların olması sebebi ile, bütün işlemler mekan ve zaman tanımaksızın kolay bir hale getirilmiştir. Fiziki olarak paranın olabildiğince az el değiştirmesinin mümkün hale geldiği söylenebilir. Teknolojiyi en çok takip etmesi gereken sektörlerin başında bankacılık sektörü yer almaktadır. Çünkü bankaların teknolojiye en çok yatırım yapan işletmeler olduğu söylenebilir. Teknolojik yenilikler ile bütünleşerek çok fazla hizmeti sağlayabilen dijital kanallar, bankacılık sektörünün müşteri kazanılması ve müşteri sadakatinin sağlanmasında en etkili ADK (Alternatif Dağıtım Kanalları) olduğu söylenebilir. Karlılığın ve verimliliğin arttığı günümüzde bankacılık sektörü, operasyonel maliyetlerinin azalması için müşterilerine internet şubesi, mobil bankacılık ve telefon bankacılığı gibi ADK'yı kullanmaya yönlendirmektedirler. Dijital kanallar ile beraber bankalar geniş bir müşteri kitlesine hizmet sunarlar, müşteriler ise zamanlarını daha etkin kullanarak işlemlerini en kısa sürede sonuçlandırabilirler (Bilgin 2013, ss. 60-61).

2000'li yıllarda kullanımı hızla yayılan internet bankacılığına yönelik hizmet kalitesi için yapılan bilimsel araştırmaların Türkiye'de yapılmaya başlanmıştır. 2010 yılı itibari ile bu araştırmaların hizmet kalitesi yönünde mobil bankacılık için başlangıç aşamasında olduğu söylenebilir (İşler 2015, s. 74).

Dijital kanal işlemleri iki şekilde gerçekleşmektedir. Birincisi, bankaların dağıtım kanalları dışında kurduğu web sitesi üzerinden müşterilere farklı bankacılık hizmetlerini

sağlaması, diğeri ise internet üzerinden sağlanan sanal bankacılık hizmetleridir. Sadece internet üzerinden çalışan bankalar Automated Teller Machine (ATM) ve diğeri dağıtım kanallarını da kullanarak müşterilere para çekme ya da yatırma hizmeti sunabilirler. Günümüzde batı ülkelerde yer alan ticari bankalar üç kategoride internet bankacılığı hizmeti sunmaktadırlar. Bunlar (Ayıgülaalı 2016, ss 10-11);

- a) **Bilgi Hizmetleri:** Bankaların müşterilerine sunabileceği hizmet ve ürünler teşvik edilir, mevduat ve borç verme, döviz kuru sorgulama, faiz oranları gibi konular bu kategoriye dahildir. Bunlar bankaların dijital kanallar ile sağlayabileceği en temel hizmetlerden sayılabilir. Bankalar dijital kanallar için bağımsız bir sunucu kullanırlar ve bu sunucunun banka içi sunucusu ile bağlantısı olmadığından bu kategoride riskin düşük olduğu söylenebilir.
- b) **Müşteri İle İletişim Hizmetleri:** Kredi uygulamaları, e-posta ve hesap sorgulamalarının düzenli olarak güncellenmesi bu kategori içerisinde yer almaktadır. Bu hizmet banka iç ağ sistemi ile müşteriler arasındaki bağlantı olması sebebi ile bankaların gerekli kontrol önlemlerini alarak hackerlerin banka iç ağ sistemlerine girmelerinin önleniminin alınması gerektiği söylenebilir.
- c) **İşlem Hizmetleri:** Kurumsal ve bireysel hizmetleri içine alan bu kategori dijital bankacılığı işlemlerinin temelini oluşturmaktadır. Kurumsal hizmetler; Yatırım bankacılığı, takas ve kredi gibi işlemleri kapsamaktadır. Bireysel hizmetler; havale, ipotek kredileri ve menkul kıymet gibi işlemleri kapsamaktadır. Bankacılık işlem sisteminin sunucusu bankanın iç ağ sistemine bağlı olması sebebi ile işlem hizmetleri ve banka ağ sistemi için büyük ölçüde risk oluşturduğu söylenebilir.

2.3.3 Bankacılıkta Dijital Kanalların Avantajları

Dijital kanalların sağladığı avantajlar müşterilere ve bankalara sağlanan avantajlar olarak ikiye ayrılabilir. Bunlar aşağıda kısaca ele alınmıştır (Ünsal 2013, ss. 10-11).

Müşterilere sağlanan avantajlar;

- a) Şubeye gitmeden ve sıra beklemeden bankacılık işlemleri
- b) Hızlı ve kesintisiz bankacılık işlemleri

- c) Detaylı şekilde bilgi alabilmek
- d) Birden fazla bankacılık ürünlerinden faydalanabilmek
- e) İşlemleri daha ucuz bir ücret karşılığı yapabilmek
- f) İşlemlerin banka çalışanı tarafından görülmemesi sebebi ile gizli ve güvenli bankacılık işlemleri

Bankalara sağlanan avantajlar;

- a) Yeni müşterilerin kazanılması
- b) Mevcut müşterilerin sadakatinin arttırılması
- c) Yeni ürün ve hizmetlerin çıkarılması
- d) Müşteriler ile olan iletişimin kolaylaştırılması
- e) Faaliyet etkinliklerinin arttırılması
- f) Etkin ve hızlı hizmet sunulması
- g) Şubelerdeki işlem yoğunluğunun minimize edilmesi
- h) Hizmet kalitesinin yükseltilmesi
- i) Maliyetlerin düşürülmesidir.

2.3.4 Bankacılıkta Dijital Kanalların Tercih Edilmesini Etkileyen Faktörler

Dijital kanalları hizmetinin tercih edilmesindeki en büyük etkenlerin kişisel algıların ve normların etkisinin olduğu söylenebilir. Özellikle, daha avantajlı olması, yaşam ve çalışma tarzına elverişli olması, güvenlik ve risk algısı, kişilerin dijital kanal tecrübeleri, kullanabilme yeteneği ve dijital kanallar gibi bir dağıtım kanalına ihtiyaç duyulması gibi sebeplerin dijital kanalların kullanılmasında en büyük faktörler olduğu söylenebilir. Banka müşterilerinin şubeler yerine dijital kanalları tercih etmelerindeki en büyük faktör zamandır. Ayrıca dijital kanalların kullanılmasında güven ve hizmet kalitesi de önemli bir faktördür. Dijital kanalların kullanımında ve banka tercihinde genellikle güven faktörü görüldüğü söylenebilir. Ayrıca bankanın tanınması ve büyüklüğünün de tercih edilme sebeplerinden biri olduğu belirtilebilir (Karadirek 2014, ss. 46-47).

Dijital kanallar internet tabanlı bankacılığın ortaya çıkışı ile müşterilerini kabul ederek bununla birlikte bireysel bankaların müşterileri ile iletişim kurma ve bunu koruma sürecinde değişiklikler getirdiği belirtilebilir. Müşteri ihtiyaçlarını doğru anlamak,

dođru ürünü dođru fiyat ile satmak ve bu ürünleri müşterilere tüm dijital kanallar aracılığı ile sağlamaktır. Ayrıca bankaların yeni ürünler ve dijital kanallar ile müşteri samimiyetinin artırılması güçlü rekabet ortamında farklılık sağladığı belirtilebilir (Özođlu 2010, s. 49).



3. VERİLERİ İŞLEME YÖNTEMLERİ

3.1 VERİ KÜMESİ TANIMI

Günümüzde dijitalleşme hayatımızın her alanında karşımıza çıkmaktadır. Özellikle zamandan tasarruf edebilmek için tüm bankacılık işlemleri şube yerine internet bankacılığı ve mobil bankacılık uygulamaları aracılığı ile yapılabilmektedir. Tezin bu bölümü özel bir bankanın verilerinden yararlanarak banka dijital kanallarına eğilimli müşterilerin belirlenmesini içermektedir. 33.503 müşterinin 100.000 satırlık verisine sahip olan data, 75 farklı yaş, kadın ve erkek olarak 2 cinsiyet, 10 farklı çalışma şekli, 52 farklı meslek, 16 farklı ürün grubu, 7.141 maaş değeri içeren bir MS Excel dosyasıdır. Data kişisel verilerin korunması gerekliliği göz önünde bulundurularak örtülü bir şekilde işlenmiştir. Datanın kategori edilmemiş ham halini Şekil 3.1’de detaylı bir şekilde görülmektedir.

Şekil 3.1: Kategori Edilmemiş Ham Data Örneği

MÜŞTERİ_NO	YAS	CINSİYET	ÇALIŞMA_SEKLI	ÇALIŞMA_SEKLI_KODU	MESLEK_ADI	MESLEK_KODU	MÜŞTERİ_DENEYİM_YIL	ÜRÜN_GRUP_ADI	ÜRÜN_GRUP_KODU	ÜRÜN_AKTIFLIK_AYI	MAAŞ_TUTAR
x	51	E	ÖZEL SEKTÖR ÜCRETLİ	9	DiĞER MESLEKLER	13	24	KANAL KULLANIM ÜRÜNLERİ	7	124	xxxx
x	51	E	ÖZEL SEKTÖR ÜCRETLİ	9	DiĞER MESLEKLER	13	24	NAKİT YÖNETİM ÜRÜNLERİ	11	7	xxxx
x	51	E	ÖZEL SEKTÖR ÜCRETLİ	9	DiĞER MESLEKLER	13	24	NAKİT YÖNETİM ÜRÜNLERİ	11	34	xxxx
x	61	E	EMEKLİ	2	TEKNİSYEN	51	22	NAKİT YÖNETİM ÜRÜNLERİ	11	193	xxxx
x	61	E	EMEKLİ	2	TEKNİSYEN	51	22	NAKİT YÖNETİM ÜRÜNLERİ	11	193	xxxx
x	62	K	EMEKLİ	2	DiĞER MESLEKLER	13	28	NAKİT YÖNETİM ÜRÜNLERİ	11	43	xxxx
x	62	K	EMEKLİ	2	DiĞER MESLEKLER	13	28	NAKİT YÖNETİM ÜRÜNLERİ	11	43	xxxx
x	48	K	KAMU SEKTÖR ÜCRETLİ	7	ÖĞRETİM GÖREVLİSİ/AKADEMİSYEN	36	20	NAKİT YÖNETİM ÜRÜNLERİ	11	192	0
x	48	K	KAMU SEKTÖR ÜCRETLİ	7	ÖĞRETİM GÖREVLİSİ/AKADEMİSYEN	36	20	NAKİT YÖNETİM ÜRÜNLERİ	11	192	0
x	48	K	KAMU SEKTÖR ÜCRETLİ	7	ÖĞRETİM GÖREVLİSİ/AKADEMİSYEN	36	20	NAKİT YÖNETİM ÜRÜNLERİ	11	66	0
x	48	K	KAMU SEKTÖR ÜCRETLİ	7	ÖĞRETİM GÖREVLİSİ/AKADEMİSYEN	36	20	NAKİT YÖNETİM ÜRÜNLERİ	11	192	0
x	48	K	KAMU SEKTÖR ÜCRETLİ	7	ÖĞRETİM GÖREVLİSİ/AKADEMİSYEN	36	20	NAKİT YÖNETİM ÜRÜNLERİ	11	172	0
x	48	K	KAMU SEKTÖR ÜCRETLİ	7	ÖĞRETİM GÖREVLİSİ/AKADEMİSYEN	36	20	NAKİT YÖNETİM ÜRÜNLERİ	11	68	0
x	48	K	KAMU SEKTÖR ÜCRETLİ	7	ÖĞRETİM GÖREVLİSİ/AKADEMİSYEN	36	20	NAKİT YÖNETİM ÜRÜNLERİ	11	5	0
x	71	E	ÖZEL SEKTÖR ÜCRETLİ	9	DOKTÖR/DİŞ HEKİMİ	16	22	NAKİT YÖNETİM ÜRÜNLERİ	11	83	0
x	71	E	ÖZEL SEKTÖR ÜCRETLİ	9	DOKTÖR/DİŞ HEKİMİ	16	22	NAKİT YÖNETİM ÜRÜNLERİ	11	83	0
x	62	E	EMEKLİ	2	ÖĞRETİMEN/EGİTİMEN	37	20	NAKİT YÖNETİM ÜRÜNLERİ	11	102	0
x	62	E	EMEKLİ	2	ÖĞRETİMEN/EGİTİMEN	37	20	NAKİT YÖNETİM ÜRÜNLERİ	11	102	0
x	62	E	EMEKLİ	2	ÖĞRETİMEN/EGİTİMEN	37	20	NAKİT YÖNETİM ÜRÜNLERİ	11	187	0
x	62	E	EMEKLİ	2	ÖĞRETİMEN/EGİTİMEN	37	20	NAKİT YÖNETİM ÜRÜNLERİ	11	187	0
x	56	E	EMEKLİ	2	DiĞER MESLEKLER	13	23	KANAL KULLANIM ÜRÜNLERİ	7	68	0
x	56	E	EMEKLİ	2	DiĞER MESLEKLER	13	23	NAKİT YÖNETİM ÜRÜNLERİ	11	43	0
x	56	E	EMEKLİ	2	DiĞER MESLEKLER	13	23	NAKİT YÖNETİM ÜRÜNLERİ	11	29	0
x	48	K	ÖZEL SEKTÖR ÜCRETLİ	9	DiĞER MESLEKLER	13	24	NAKİT YÖNETİM ÜRÜNLERİ	11	10	0
x	48	K	ÖZEL SEKTÖR ÜCRETLİ	9	DiĞER MESLEKLER	13	24	NAKİT YÖNETİM ÜRÜNLERİ	11	72	0
x	50	E	KAMU SEKTÖR ÜCRETLİ	7	ÖĞRETİM GÖREVLİSİ/AKADEMİSYEN	36	20	NAKİT YÖNETİM ÜRÜNLERİ	11	16	xxxx
x	50	E	KAMU SEKTÖR ÜCRETLİ	7	ÖĞRETİM GÖREVLİSİ/AKADEMİSYEN	36	20	NAKİT YÖNETİM ÜRÜNLERİ	11	16	xxxx
x	50	E	KAMU SEKTÖR ÜCRETLİ	7	ÖĞRETİM GÖREVLİSİ/AKADEMİSYEN	36	20	KANAL KULLANIM ÜRÜNLERİ	7	19	xxxx
x	50	E	KAMU SEKTÖR ÜCRETLİ	7	ÖĞRETİM GÖREVLİSİ/AKADEMİSYEN	36	20	NAKİT YÖNETİM ÜRÜNLERİ	11	15	xxxx
x	50	E	KAMU SEKTÖR ÜCRETLİ	7	ÖĞRETİM GÖREVLİSİ/AKADEMİSYEN	36	20	NAKİT YÖNETİM ÜRÜNLERİ	11	25	xxxx
x	50	E	KAMU SEKTÖR ÜCRETLİ	7	ÖĞRETİM GÖREVLİSİ/AKADEMİSYEN	36	20	NAKİT YÖNETİM ÜRÜNLERİ	11	23	xxxx
x	50	E	KAMU SEKTÖR ÜCRETLİ	7	ÖĞRETİM GÖREVLİSİ/AKADEMİSYEN	36	20	NAKİT YÖNETİM ÜRÜNLERİ	11	31	xxxx
x	50	E	KAMU SEKTÖR ÜCRETLİ	7	ÖĞRETİM GÖREVLİSİ/AKADEMİSYEN	36	20	NAKİT YÖNETİM ÜRÜNLERİ	11	31	xxxx
x	50	E	KAMU SEKTÖR ÜCRETLİ	7	ÖĞRETİM GÖREVLİSİ/AKADEMİSYEN	36	20	KANAL KULLANIM ÜRÜNLERİ	7	19	xxxx
x	50	E	KAMU SEKTÖR ÜCRETLİ	7	ÖĞRETİM GÖREVLİSİ/AKADEMİSYEN	36	20	NAKİT YÖNETİM ÜRÜNLERİ	11	15	xxxx
x	50	E	KAMU SEKTÖR ÜCRETLİ	7	ÖĞRETİM GÖREVLİSİ/AKADEMİSYEN	36	20	NAKİT YÖNETİM ÜRÜNLERİ	11	2	xxxx
x	50	E	KAMU SEKTÖR ÜCRETLİ	7	ÖĞRETİM GÖREVLİSİ/AKADEMİSYEN	36	20	NAKİT YÖNETİM ÜRÜNLERİ	11	2	xxxx
x	54	K	KAMU SEKTÖR ÜCRETLİ	7	ÖĞRETİM GÖREVLİSİ/AKADEMİSYEN	36	20	NAKİT YÖNETİM ÜRÜNLERİ	11	38	xxxx
x	54	K	KAMU SEKTÖR ÜCRETLİ	7	ÖĞRETİM GÖREVLİSİ/AKADEMİSYEN	36	20	NAKİT YÖNETİM ÜRÜNLERİ	11	14	xxxx
x	54	K	KAMU SEKTÖR ÜCRETLİ	7	ÖĞRETİM GÖREVLİSİ/AKADEMİSYEN	36	20	NAKİT YÖNETİM ÜRÜNLERİ	11	73	xxxx

Şekil 3.1'deki tablo 12 ayrı başlıktan oluşan özel bir bankanın müşterilerinin müşteri numarası, yaş, cinsiyet, çalışma şekli, çalışma şekli kodu, meslek adı, meslek adı kodu, müşteri deneyimi yılı, ürün grup adı, ürün grup kodu, ürün aktiflik ayı ve maaş tutarlarını içermektedir.

3.2 VERİNİN HAZIRLANMASI

SQL server aracılığı ile MS Excel aktarılan datadan gereksiz olan veriler çıkarılarak gerekli olan birleştirmeler yapılarak dijital kanal eğilimli kişilerin belirlenmesine yönelik ön hazırlık yapılmıştır.

3.2.1 Özelliklerin Değerlendirilmesi ve Veri Temizliği

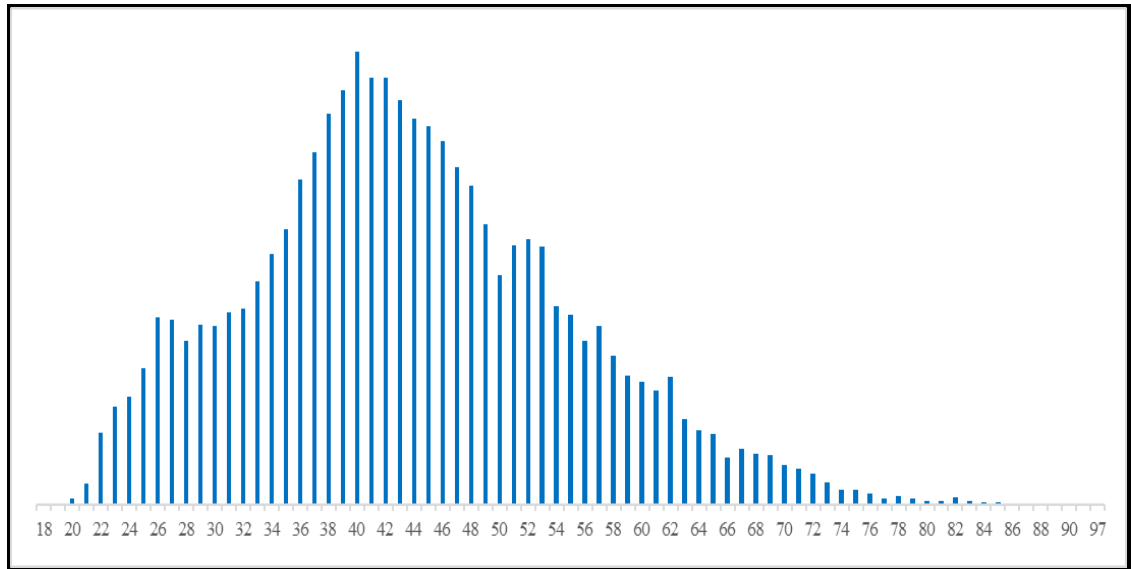
Datada bulunan veriler boş (NULL) değer içermemektedir. Data içerisinde meslek adı, meslek adı kodu, müşteri deneyimi yılı, ürün aktiflik ayı inceleme detayında yer almadığı için silinmiştir. Meslek bilgisi yerine çalışma şeklinin kullanılması uygun bulunmuştur. Ayrıca müşteri deneyimi ve ürün aktifliği bilgisi silinerek sadece müşterinin yaş bilgisi değerlendirme içerisinde tutulmuştur.

Şekil 3.2: Kategori Edilmemiş Başlık Elemesi Yapılmış Data Örneği

MÜŞTERİ_NO	YAS	CINSIYET	CALISMA_SEKLI	CALISMA_SEKLL_KODU	URUN_GRUP_ADI	URUN_GRUP_KODU	MAAŞ_TUTAR
x	51	E	ÖZEL SEKTÖR ÜCRETLİ	9	KANAL KULLANIM ÜRÜNLERİ	7	xxxx
x	51	E	ÖZEL SEKTÖR ÜCRETLİ	9	NAKİT YÖNETİM ÜRÜNLERİ	11	xxxx
x	51	E	ÖZEL SEKTÖR ÜCRETLİ	9	NAKİT YÖNETİM ÜRÜNLERİ	11	xxxx
x	61	E	EMEKLİ	2	NAKİT YÖNETİM ÜRÜNLERİ	11	xxxx
x	61	E	EMEKLİ	2	NAKİT YÖNETİM ÜRÜNLERİ	11	xxxx
x	62	K	EMEKLİ	2	NAKİT YÖNETİM ÜRÜNLERİ	11	xxxx
x	62	K	EMEKLİ	2	NAKİT YÖNETİM ÜRÜNLERİ	11	xxxx
x	48	K	KAMU SEKTÖR ÜCRETLİ	7	NAKİT YÖNETİM ÜRÜNLERİ	11	0
x	48	K	KAMU SEKTÖR ÜCRETLİ	7	NAKİT YÖNETİM ÜRÜNLERİ	11	0
x	48	K	KAMU SEKTÖR ÜCRETLİ	7	NAKİT YÖNETİM ÜRÜNLERİ	11	0
x	48	K	KAMU SEKTÖR ÜCRETLİ	7	NAKİT YÖNETİM ÜRÜNLERİ	11	0
x	48	K	KAMU SEKTÖR ÜCRETLİ	7	NAKİT YÖNETİM ÜRÜNLERİ	11	0
x	48	K	KAMU SEKTÖR ÜCRETLİ	7	NAKİT YÖNETİM ÜRÜNLERİ	11	0
x	48	K	KAMU SEKTÖR ÜCRETLİ	7	NAKİT YÖNETİM ÜRÜNLERİ	11	0
x	71	E	ÖZEL SEKTÖR ÜCRETLİ	9	NAKİT YÖNETİM ÜRÜNLERİ	11	0
x	71	E	ÖZEL SEKTÖR ÜCRETLİ	9	NAKİT YÖNETİM ÜRÜNLERİ	11	0
x	62	E	EMEKLİ	2	NAKİT YÖNETİM ÜRÜNLERİ	11	0
x	62	E	EMEKLİ	2	NAKİT YÖNETİM ÜRÜNLERİ	11	0
x	62	E	EMEKLİ	2	NAKİT YÖNETİM ÜRÜNLERİ	11	0
x	62	E	EMEKLİ	2	NAKİT YÖNETİM ÜRÜNLERİ	11	0
x	56	E	EMEKLİ	2	KANAL KULLANIM ÜRÜNLERİ	7	0
x	56	E	EMEKLİ	2	NAKİT YÖNETİM ÜRÜNLERİ	11	0
x	56	E	EMEKLİ	2	NAKİT YÖNETİM ÜRÜNLERİ	11	0
x	48	K	ÖZEL SEKTÖR ÜCRETLİ	9	NAKİT YÖNETİM ÜRÜNLERİ	11	0
x	48	K	ÖZEL SEKTÖR ÜCRETLİ	9	NAKİT YÖNETİM ÜRÜNLERİ	11	0
x	50	E	KAMU SEKTÖR ÜCRETLİ	7	NAKİT YÖNETİM ÜRÜNLERİ	11	xxxx
x	50	E	KAMU SEKTÖR ÜCRETLİ	7	NAKİT YÖNETİM ÜRÜNLERİ	11	xxxx
x	50	E	KAMU SEKTÖR ÜCRETLİ	7	KANAL KULLANIM ÜRÜNLERİ	7	xxxx
x	50	E	KAMU SEKTÖR ÜCRETLİ	7	NAKİT YÖNETİM ÜRÜNLERİ	11	xxxx
x	50	E	KAMU SEKTÖR ÜCRETLİ	7	NAKİT YÖNETİM ÜRÜNLERİ	11	xxxx
x	50	E	KAMU SEKTÖR ÜCRETLİ	7	NAKİT YÖNETİM ÜRÜNLERİ	11	xxxx
x	50	E	KAMU SEKTÖR ÜCRETLİ	7	NAKİT YÖNETİM ÜRÜNLERİ	11	xxxx
x	50	E	KAMU SEKTÖR ÜCRETLİ	7	NAKİT YÖNETİM ÜRÜNLERİ	11	xxxx
x	50	E	KAMU SEKTÖR ÜCRETLİ	7	KANAL KULLANIM ÜRÜNLERİ	7	xxxx
x	50	E	KAMU SEKTÖR ÜCRETLİ	7	NAKİT YÖNETİM ÜRÜNLERİ	11	xxxx
x	50	E	KAMU SEKTÖR ÜCRETLİ	7	NAKİT YÖNETİM ÜRÜNLERİ	11	xxxx
x	50	E	KAMU SEKTÖR ÜCRETLİ	7	NAKİT YÖNETİM ÜRÜNLERİ	11	xxxx
x	54	K	KAMU SEKTÖR ÜCRETLİ	7	NAKİT YÖNETİM ÜRÜNLERİ	11	xxxx
x	54	K	KAMU SEKTÖR ÜCRETLİ	7	NAKİT YÖNETİM ÜRÜNLERİ	11	xxxx
x	54	K	KAMU SEKTÖR ÜCRETLİ	7	NAKİT YÖNETİM ÜRÜNLERİ	11	xxxx

Şekil 3.2’de kullanılmayacak olan veriler elendikten sonraki durumu görülmektedir.

Şekil 3.3: Yaş Gruplaması Öncesi Dağılım



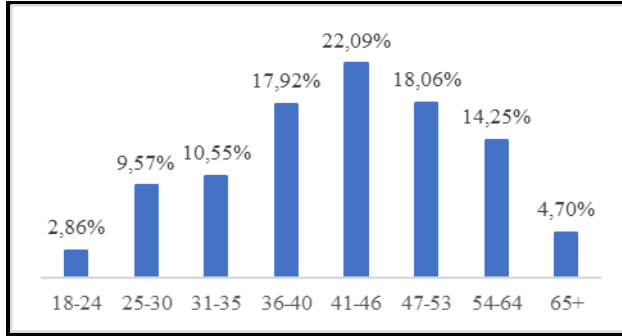
Şekil 3.3 değerlendirildiğinde 75 farklı yaş değeri yoğunluk frekanslarına göre incelenerek 8 gruba bölünmüştür. Bölünen yaş başlığında toplam 7.141 müşteri gruplandırılmıştır. Bu gruplar tablo 3.1 de detaylı olarak belirtilmiştir.

Tablo 3.1: Yaş Gruplaması ve Yüzde Değerleri

YAŞ ARALIĞI	ADET	YÜZDE ORANI
18-24	958	2,9%
25-30	3.206	9,6%
31-35	3.534	10,5%
36-40	6.004	17,9%
41-46	7.401	22,1%
47-53	6.051	18,1%
54-64	4.773	14,2%
65+	1.576	4,7%
TOPLAM	33.503	

Tablo 3.1’de 8 gruba ayrıldıktan sonra yaş aralıkları adetsel ve yüzdeler oranları görülmektedir.

Şekil 3.4: Yaş Gruplaması Sonrası Dağılım



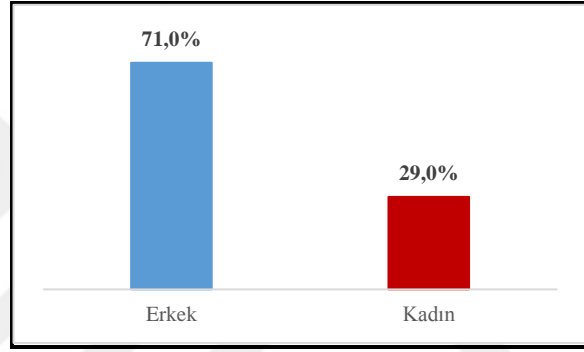
Şekil 3.4’te yaş gruplaması sonrası dağılımının grafiksel boyutu görülmektedir.

Tablo 3.2: Cinsiyet Yüzde Değerleri

CİNSİYET DAĞILIMI KODU	CİNSİYET DAĞILIMI	ADET	YÜZDE ORANI
0	Erkek	23.801	71,0%
1	Kadın	9.702	29,0%
	TOPLAM	33.503	

Tablo 3.2’de cinsiyet kodu, cinsiyet yüzde değerleri incelendiğinde de 33.503 müşterinin yüzde 71’i erkek, yüzde 29 u kadından oluştuğu görülmektedir.

Şekil 3.5: Cinsiyet Dağılımı



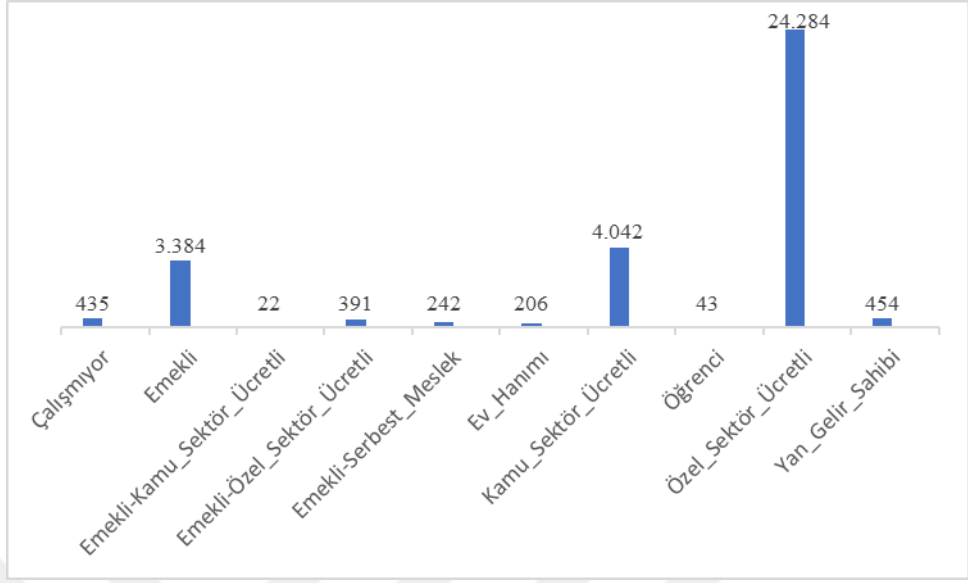
Şekil 3.5’de cinsiyet dağılımının grafiksel durumu görülmektedir.

Tablo 3.3: Çalışma Şekli Yüzde Dağılımı

ÇALIŞMA ŞEKLİ	ADET	YÜZDE
Çalışmıyor	435	1,3%
Emekli	3.384	10,1%
Emekli-Kamu_Sektör_Ücretli	22	0,1%
Emekli-Özel_Sektör_Ücretli	391	1,2%
Emekli-Serbest_Meslek	242	0,7%
Ev_Hanımı	206	0,6%
Kamu_Sektör_Ücretli	4.042	12,1%
Öğrenci	43	0,1%
Özel_Sektör_Ücretli	24.284	72,5%
Yan_Gelir_Sahibi	454	1,4%
	33.503	

Tablo 3.3 değerlendirildiğinde bulunan çalışma şekli 10 farklı içerikten oluşmaktadır. Bu çalışma şekilleri aynen korunmuştur.

Şekil 3.6: Çalışma Şeklinin Grafiksnel Dağılımı



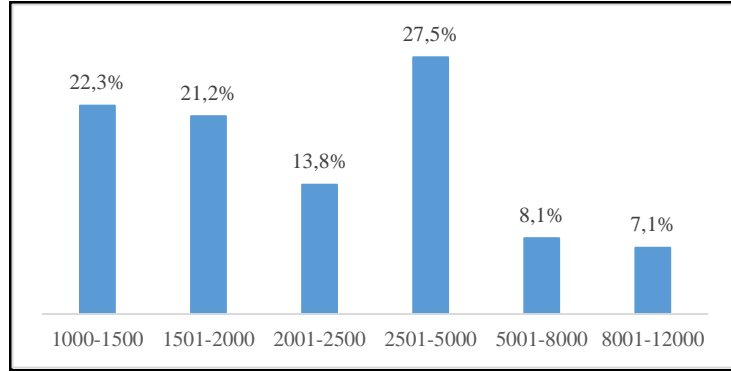
Şekil 3.6. da çalışma şeklinin grafiksnel durumu belirtilmiştir.

Tablo 3.4: Maaş Gruplaması Sonrası Yüzde Dağılımı

MAAŞ ARALIKLARI	ADET	YÜZDE ORANI
1000-1500	1.869	22,3%
1501-2000	1.772	21,2%
2001-2500	1.159	13,8%
2501-5000	2.300	27,5%
5001-8000	678	8,1%
8001-12000	594	7,1%
	8.372	

Tablo 3.4'te maaş gruplaması sonrası yüzde dağılımı değerlendirdiğimizde maaş aralık tutarları 6 kategoride ele alınmıştır. Adetsel ve yüzdelik oranları tabloda görülmektedir.

Şekil 3.7: Maaş Grublama Sonrası Dağılım



Şekil 3.7’de maaş grublama sonrası dağılımın grafiksel durumu gösterilmiştir.

Şekil 3.8: Data Son Hali

MUSTERI	YAS_ARAL	CINSIYET_KOD	CALISMA_SEKLI	MAAS_ARA	MEVDUA	NAKIT_YONETI	KREDI_KARTLI	HAYAT_SIGOR	KREDI	class
		U				M_URUNL	ARI	TALARI		
4	47-53	1	KAMU_SEKTOR_UCRETLI	0	0	1	0	0	0	0
5	65+	0	OZEL_SEKTOR_UCRETLI	0	0	1	0	0	0	0
6	54-64	0	EMEKLİ	0	0	1	0	0	0	0
14	41-46	1	OZEL_SEKTOR_UCRETLI	0	0	1	0	0	0	0
15	47-53	0	KAMU_SEKTOR_UCRETLI	0	0	1	0	0	0	0
20	47-53	0	OZEL_SEKTOR_UCRETLI	0	0	1	1	0	0	1
23	54-64	1	OZEL_SEKTOR_UCRETLI	0	0	0	1	0	0	1
27	47-53	0	EMEKLİ	0	0	1	0	0	0	0
38	54-64	1	OZEL_SEKTOR_UCRETLI	0	0	0	1	0	0	0
40	54-64	0	EMEKLİ	0	0	0	0	0	1	0
41	41-46	1	OZEL_SEKTOR_UCRETLI	0	0	0	1	0	0	0
60	41-46	1	OZEL_SEKTOR_UCRETLI	0	0	0	1	0	0	0
62	47-53	0	OZEL_SEKTOR_UCRETLI	0	0	0	1	0	0	0
71	41-46	0	OZEL_SEKTOR_UCRETLI	0	0	1	0	0	0	0
76	47-53	0	OZEL_SEKTOR_UCRETLI	0	0	0	1	0	0	0
77	47-53	0	KAMU_SEKTOR_UCRETLI	0	0	0	1	0	0	0
82	65+	0	OZEL_SEKTOR_UCRETLI	0	0	1	1	0	0	1
84	36-40	1	OZEL_SEKTOR_UCRETLI	0	0	0	0	0	0	1
87	41-46	0	OZEL_SEKTOR_UCRETLI	0	1	1	1	1	0	1
91	47-53	0	OZEL_SEKTOR_UCRETLI	0	1	0	0	0	0	0

Yukarıdaki Şekil 3.8’de 33.503 müşteri bilgisinden kullanımı en yüksek 5 ürün seçilerek data 11.478 müşteri bilgisine indirilmiştir. Bu müşterilerin banka dijital kanallarını kullanıp kullanmama durumuna göre data classı yapılarak son haline getirilmiştir.

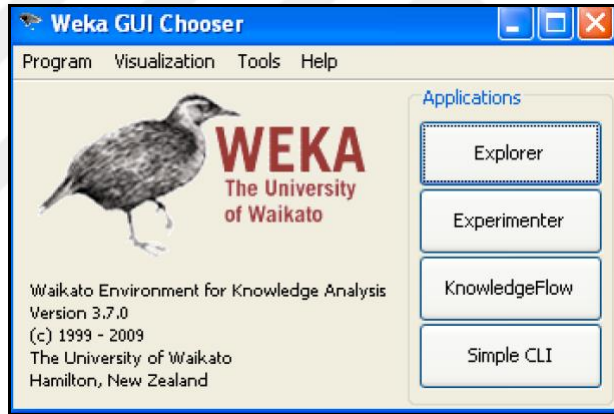
3.2.2 Verinin Weka Programı İçin Hazırlanması

Waikato Environment for Knowledge Analysis (Weka), Yeni Zelanda’daki Waikato Üniversitesinin geliştirmiş olduğu bir programdır. Weka kelimesi “Waikato Environment for Knowledge Analysis” ın kısaltmasıdır. Programın yaratılma sebebi, ilk zamanlar Yeni Zelanda’da tarım ve çiftçilik üretiminden elde edilen verilerin analiz

edilmesi olarak görülmektedir. Programın içeriğindeki makine öğrenme algoritmaları ve kullanım kolaylığının üst seviyede olmasından dolayı hızlı bir biçimde gelişim göstermiş ve şu anda veri madenciliği uygulama alanlarının birçoğunda kullanılabilir hale gelmiştir.

Program içerisinde makine öğrenmesi algoritmalarının birçoğunu bulundurması ile beraber sıfırdan algoritma yapılabilmesi yönünde de tasarlandığı söylenebilir. Java dilinde yazılan, General Public License (GNU) lisansı ile yayınlanmış açık kaynaklı bir yazılım olarak hayata geçirildiği görülmektedir. Program Windows, Macintosh, Linux vb. farklı işletim sistemlerinde çalışabilen bir yapıya sahiptir. Program yüklendikten sonra çalıştırmak için Weka'nın grafiksel görselinin bulunduğu kullanıcı ara yüzünden faydalanılmaktadır. Grafiksel görselinde beş adet seçenek vardır (Bilen 2009, s. 51).

Şekil 3.9: Weka Arayüzü Görseli



Şekil 3.9'da Weka'nın arayüzü kullanım görseli görülmektedir. Kullanım görselinde yer alan Explorer, Experimenter, Knowledge Flow, Simple CLI seçenekleri aşağıda kısaca ele alınmıştır.

Explorer: Datanın ana işleme adımlarının yürütüldüğü bölümdür. Burada veri madenciliği sınıflandırmaları, kümelemeleri ve birliktelik kuralları uygulanabilir. Ayrıca datada istenilen filtrelemeler, silmeler ve değiştirmeler gerçekleştirilebildiği söylenebilir (Çığışar 2017, s.38).

Experimenter: Deneylerin, öğrenim planlamalarının ve istatistik testlerinin yapıldığı bölümü oluşturur. Kullanılacak olan dataya birbirinden değişik tekniklerin

uygulanmasını ya da farklı parametreler ile aynı tekniğin uygulanmasını sağladığı ve bir defada birden fazla deneyin yapılabildiği alan olduğu söylenebilir (Bilen 2009, s. 53).

Knowledge Flow: Explorer seçeneği ile aynı işlevi gören, sürükle bırak yönteminin kullanılarak Weka paketlerinin oluşturduğu ve kullanım kolaylığının da sağlandığı alternatif bölümdür. Ayrıca Experimenter seçeneğinde bulunmayan ya da eksiklikleri olan bazı özelliklerin gelişimin sürdüğü bölüm olduğu da söylenebilir (Bilen 2009, s. 53).

Simple CLI: Programdaki seçeneklerin kurlsız ve basit bir şekilde işlenebilmesine yarayan ara yüzü olduğu belirtilebilir.

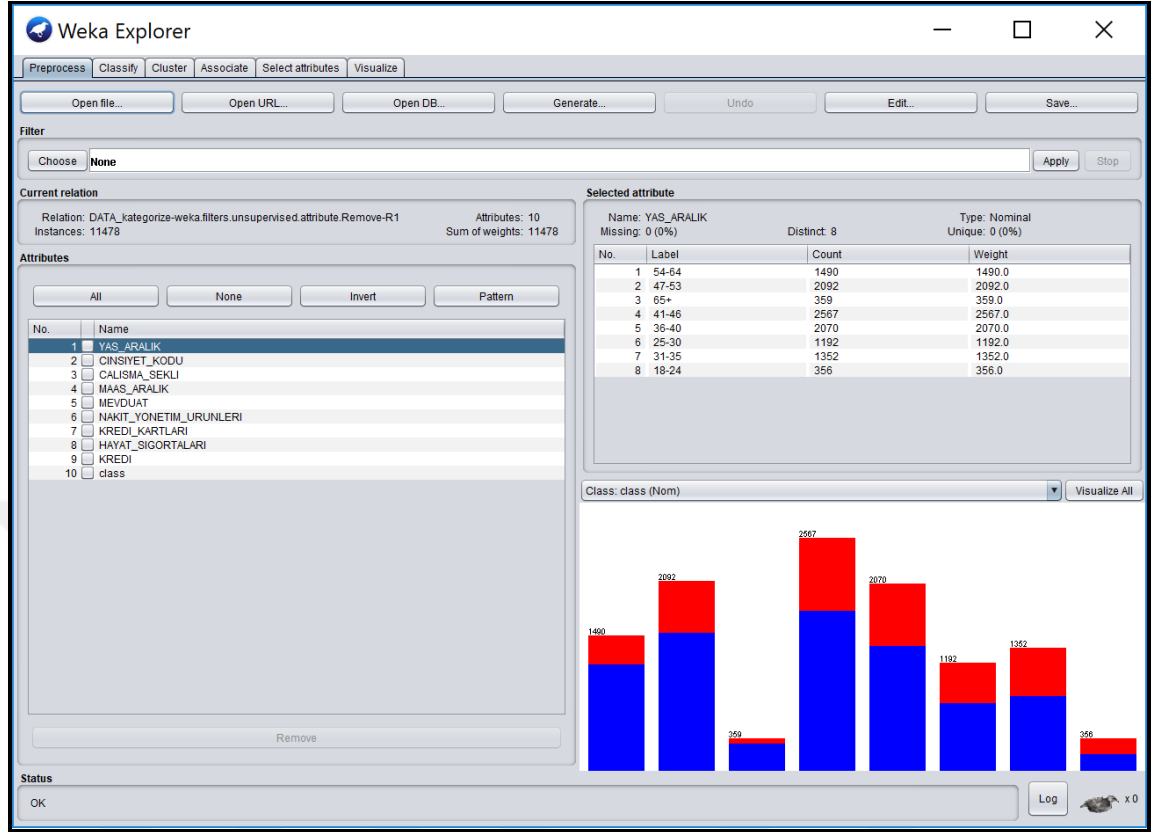
Şekil 3.10: Datanın ARFF Formatı



```
File Edit Format View Help
@relation DATA_kategorize-weka.filters.unsupervised.attribute.Remove-R1
@attribute YAS_ARALIK {54-64,47-53,65+,41-46,36-40,25-30,31-35,18-24}
@attribute CINSIYET_KODU numeric
@attribute CALISMA_SEKLI {EHEKLI,KAMU_SEKTOR_UCRETLI,OZEL_SEKTOR_UCRETLI,EHEKLI-OZEL_SEKTOR_UCRETLI,VAN_GELIR_SAHIBI,EHEKLI-SERBEST_MESLEK,CALISMIYOR_EV_HANIMI,EHEKLI-KAMU_SEKTOR_UCRETLI,OZURENCI}
@attribute MIA5_ARALIK {2001-2500,0,5001-8000,8001-12000,2501-5000,1501-2000,1000-1500}
@attribute MEVDUAT numeric
@attribute NAKIT_YONETIM_URUNLERI numeric
@attribute KREDI_KARTLARI numeric
@attribute HAYAT_SIGORTALARI numeric
@attribute KREDI numeric
@attribute class {0,1}
@data
54-64,0,EHEKLI,2001-2500,0,1,0,0,0,0
47-53,1,KAMU_SEKTOR_UCRETLI,0,0,1,0,0,0,0
65+,0,OZEL_SEKTOR_UCRETLI,0,0,1,0,0,0,0
54-64,0,EHEKLI,0,0,1,0,0,0,0
47-53,0,KAMU_SEKTOR_UCRETLI,5001-8000,0,1,0,0,0,1
54-64,1,KAMU_SEKTOR_UCRETLI,8001-12000,0,1,0,0,0,0
41-46,1,OZEL_SEKTOR_UCRETLI,0,0,1,0,0,0,0
47-53,0,KAMU_SEKTOR_UCRETLI,0,0,1,0,0,0,0
47-53,0,OZEL_SEKTOR_UCRETLI,0,0,1,1,0,0,0,1
54-64,1,OZEL_SEKTOR_UCRETLI,0,0,0,1,0,0,0,1
47-53,0,EHEKLI-OZEL_SEKTOR_UCRETLI,0,0,1,0,0,0,0
54-64,1,OZEL_SEKTOR_UCRETLI,0,0,0,1,0,0,0,0
54-64,0,EHEKLI,0,0,0,0,0,1,0
41-46,1,OZEL_SEKTOR_UCRETLI,0,0,0,1,0,0,0,0
47-53,0,OZEL_SEKTOR_UCRETLI,2501-5000,0,0,0,0,1,0
41-46,1,KAMU_SEKTOR_UCRETLI,1501-2000,0,1,1,0,0,0
65+,0,KAMU_SEKTOR_UCRETLI,5001-8000,0,1,1,0,0,0
47-53,1,EHEKLI,2501-5000,0,1,1,0,0,0
41-46,1,OZEL_SEKTOR_UCRETLI,0,0,0,1,0,0,0,0
47-53,0,OZEL_SEKTOR_UCRETLI,0,0,0,1,0,0,0,0
47-53,1,OZEL_SEKTOR_UCRETLI,2501-5000,0,1,0,0,0,0
65+,0,KAMU_SEKTOR_UCRETLI,5001-8000,0,0,1,0,0,0
41-46,0,OZEL_SEKTOR_UCRETLI,0,0,1,0,0,0,0,0
65+,0,EHEKLI,2001-2500,0,0,1,0,0,0,0
47-53,0,OZEL_SEKTOR_UCRETLI,0,0,0,1,0,0,0,0
54-64,0,OZEL_SEKTOR_UCRETLI,2501-5000,0,1,1,0,0,0
65+,0,OZEL_SEKTOR_UCRETLI,0,0,1,1,0,0,1
36-40,1,OZEL_SEKTOR_UCRETLI,0,0,0,0,0,0,0,1
41-46,0,OZEL_SEKTOR_UCRETLI,0,1,1,1,0,0,0,1
47-53,1,OZEL_SEKTOR_UCRETLI,1000-1500,1,1,0,0,0,1
47-53,0,OZEL_SEKTOR_UCRETLI,2001-2500,1,1,0,0,0,1
41-46,0,OZEL_SEKTOR_UCRETLI,0,1,1,1,0,0,0,0
41-46,0,OZEL_SEKTOR_UCRETLI,0,1,1,1,0,0,1,1
54-64,0,OZEL_SEKTOR_UCRETLI,1501-2000,1,1,1,0,0,0
65+,0,OZEL_SEKTOR_UCRETLI,0,1,0,1,0,0,1
54-64,1,EHEKLI,1000-1500,1,1,0,0,1,1
47-53,0,OZEL_SEKTOR_UCRETLI,0,1,0,0,0,0,1
```

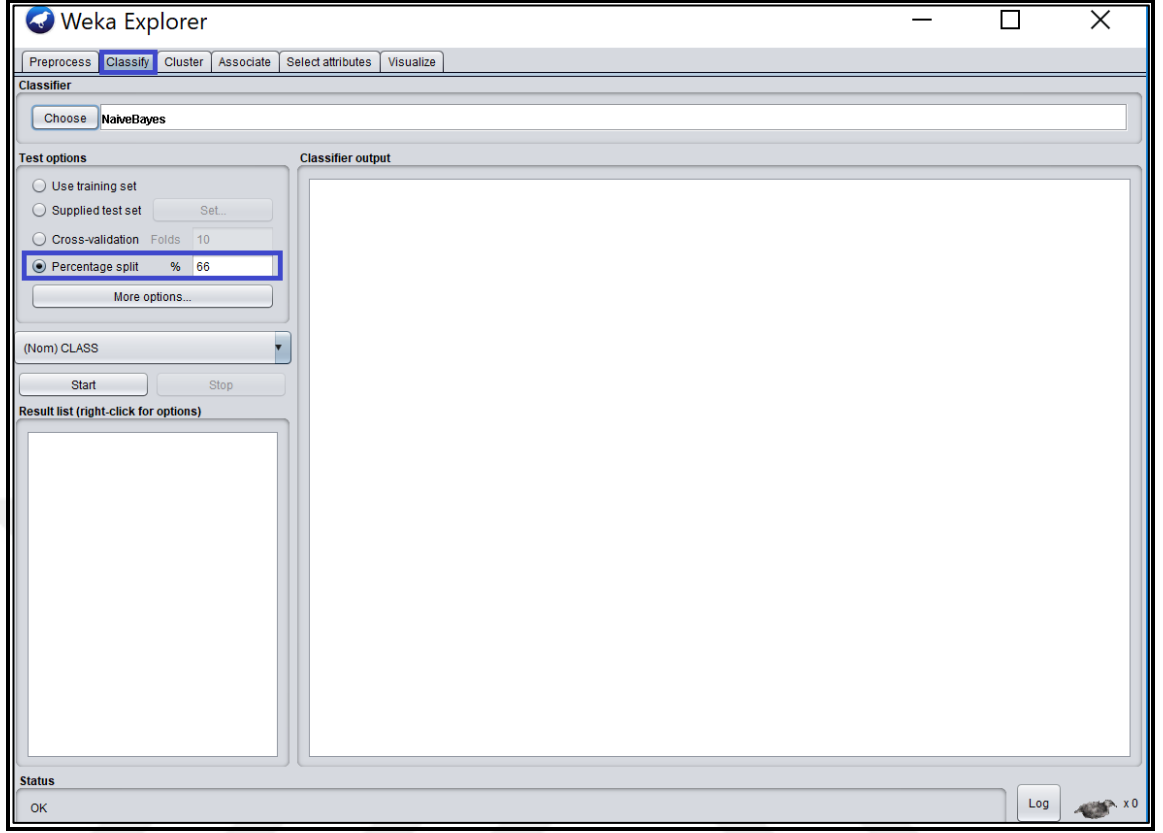
Şekil 3.10’da görüldüğü gibi MS Excel olarak hazırlanan data Weka programında kullanılmak üzere arff formatına çevrilmiştir.

Şekil 3.11: Datanın Weka Görüntüsü



Şekil 3.11’de Weka programı çalıştırıldıktan sonra datanın son durumu görülmektedir. Dijital kanal kullanıcısı olanların 1, dijital kanal kullanıcısı olmayanların 0 olması koşulu datada “class” olarak belirlenmiştir.

Şekil 3.12: Weka Classify Adımı Görüntüsü



Şekil 3.12 Weka classify adımı görüntüsü sınıflama algoritmaları uygulanmadan önce percentage split seçimi yüzde 66 olacak şekilde tercih edilmiştir.

3.3 Algoritma Değerlendirme Kriterleri

Veri madenciliği sınıflandırması yapılırken algoritma sonuçlarının karşılaştırılması ve en başarılı sınıflandırma algoritmasının belirlenmesi en fazla dikkat edilmesi gereken konulardan biridir. (Witten & Frank & Pal 2011, s.154).

Algoritma sonuçlarının karşılaştırılması sınıflandırılan değerlerin yüzdesel ölçütlerine göre alınmıştır. Bunları kısaca tanımladığımızda;

ACC: Gözlem sonrası doğru sınıflandırılmaların yüzdeselini gösterir.

TPR: Gözlem pozitif ve pozitif olarak tahmin edilir.

FPR: Gözlem olumsuzdur, ancak pozitif olarak tahmin edilir.

PRECISION: Gözlem çıktısında pozitif olup pozitif değerlendirilenlerin, gözlem çıktısındaki pozitif olup pozitif değerlendirilen ve gözlem çıktısı olumsuz ancak pozitif olarak tahmin edilenlerin bölümünü ifade eder.

$$Precision = \frac{TPR}{TPR + FPR}$$

RMSE: Kök kareler karesi olarak ifade edilir. Hata ortalamalarının ne kadar büyüklükte olduğunu ölçer.

F-MEASURE: Aritmetik ortalama yerine Harmonik ortalama kullanılan bir ölçüm hesaplanmaktadır.

ROC AREA: ROC alanı eğrisinde birbirinden farklı olan sınıflandırma algoritmalarının tahmini performansları belirlenir. Eğri altında kalan alan, en doğru sınıflandırma algoritmasının tercih edilebilmesi için kullanılan değerlendirme ölçütlerinden birini oluşturur. Eğrinin altında kalan alanın 1 e yaklaşıyor olması sınıflandırma sürecinin doğru gerçekleştiğini göstermektedir.

3.4 Kullanılan Algoritmalar

Hazırlanan arff formatındaki datanın Weka da algoritmaları çalıştırıldıktan sonra ortaya çıkan algoritma sonuçları aşağıda detaylı şekilde anlatılmıştır.

Navie Bayes (Sade Bayes) Algoritması: Bayes sınıflama tekniklerinin istatistiksel sınıflama teknikleri arasında yer aldığı söylenebilir. Navia bayes algoritması şu şekilde açıklanabilmektedir; X sınıflandırma yapılacak üyeliği bilinmeyen veri örneği olabilir. Örnek $X = \{ x_1, x_2, \dots, X_n \}$ özelliklerini içeriyor olsun. Örnek sınıfta m kadar sınıf olduğunu düşünelim. C_1, C_2, \dots, C_n 'inde sınıfların öznitelik değerleri olduğunu düşünürsek sınıf bilgisi bulunacak örneğin olasılık hesaplarını $P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$

formülü ile bulunabiliriz. Yapılacak karşılaştırma sonucunda değerlerin arasından en yüksek olan seçilerek X'in sınıfının bulunduğu söylenebilir (Özkan 2016, ss. 157-158).

Logistic Algoritması: Lojistik regresyon analizi iki kategorisi olan ve birbirleri ile bağımlılığı bulunan değişkenler ile yapılabilir. Bağımlılığı bulunan değişken 0 ya da 1 değerinden birisi olabilir. Lojistik regresyonun bu açıklamadan da anlaşılacağı gibi olasılıklar üzerinde çalıştığı söylenebilir. Algoritma olayın gerçekleşmesi olasılığının ve gerçekleşmemesi olasılığının birbirlerine bölünmesinden oluşan değerlerin logaritması

alınarak uygulanmaktadır. Logaritmik dağılımın kullanılmasındaki amaç dağılımı normalize etmektir. Basitçe söylemek gerekirse kategorik değişken 1 ya da 0 olduğunda artı sonsuz ile eksi sonsuz arasında değer alabilir (Academia (online), Lojistik Regresyon, 2013, http://www.academia.edu/11479607/Lojistik_Regresyon (erişim tarihi 23.04.2018)).

IBk Algoritması: K en yakın komşu algoritmasının Weka programındaki adı Ibk algoritmasıdır. K en yakın komşu sınıflandırması yapılırken verilerin bulunduğu datadaki kayıtların her biri için öteki kayıtlar ile olan uzaklığı hesaplanmaktadır. Her bir veri için öteki kayıtlardaki k adedi değerlendirmeye alınmaktadır. Adından anlaşılacağı gibi belirtilen k adet kayıt, diğer bir ifade ile veri tabanında yer alan nokta, uzaklık hesaplanan noktaya diğer kayıtlardan en yakın kayıt olduğu söylenebilir. K en yakın komşu algoritması coğrafi bilgilerin değerlendirildiği sistemler içinde sıklıkla kullanılmaktadır. Örneğin; belirlenmiş bir yerin en yakınındaki il, istasyon vb. noktaların bulunmasında bu algoritmadan yararlandığı söylenebilir.

Algoritmanın çalışması için k değeri öncelikli seçilmelidir; seçilen değer yüksek tutulması benzerliği fazla olan verilerin birbirine yaklaşarak toplanmalarına, seçilen değer düşük tutulması ise verinin benzerliği fazla olmasına rağmen bir araya gelmemesine sebep olabilir. Kısaca belirtmek gerekirse aynı sınıf içinde değerlendirilmesi gereken veriler ya bir araya gelemez ya da farklı sınıflara kayabilirler. K değerlerinin çoğunlukla 3,5 ve 7 olarak seçilmekte olduğu söylenebilir.

K en yakın komşu algoritmanın çalışması kısaca aşağıdaki gibi belirtilebilir (Silahtaroglu 2016, ss. 118-119);

- a) Uygulanabilir mesafede ölçümleme alanını belirlemeli (Euclid mesafe ölçümlemesi kullanılan arasında ilk sıradadır).
- b) Birbirlerine en yakın mesafedeki k adet nokta belirlenmeli.
- c) Belirlemeden çıkan grubun en sık karşılaştığı sınıf seçilmeli.
- d) Buradaki yer alan gruba sınıf adı verilmeli

J48 Algoritması: C4.5 algoritmasının Weka programındaki adı J48 algoritmasıdır. C4.5 algoritmasında hataların oransal olarak tahmin edilmesi sağlanmaktadır. Bu hata

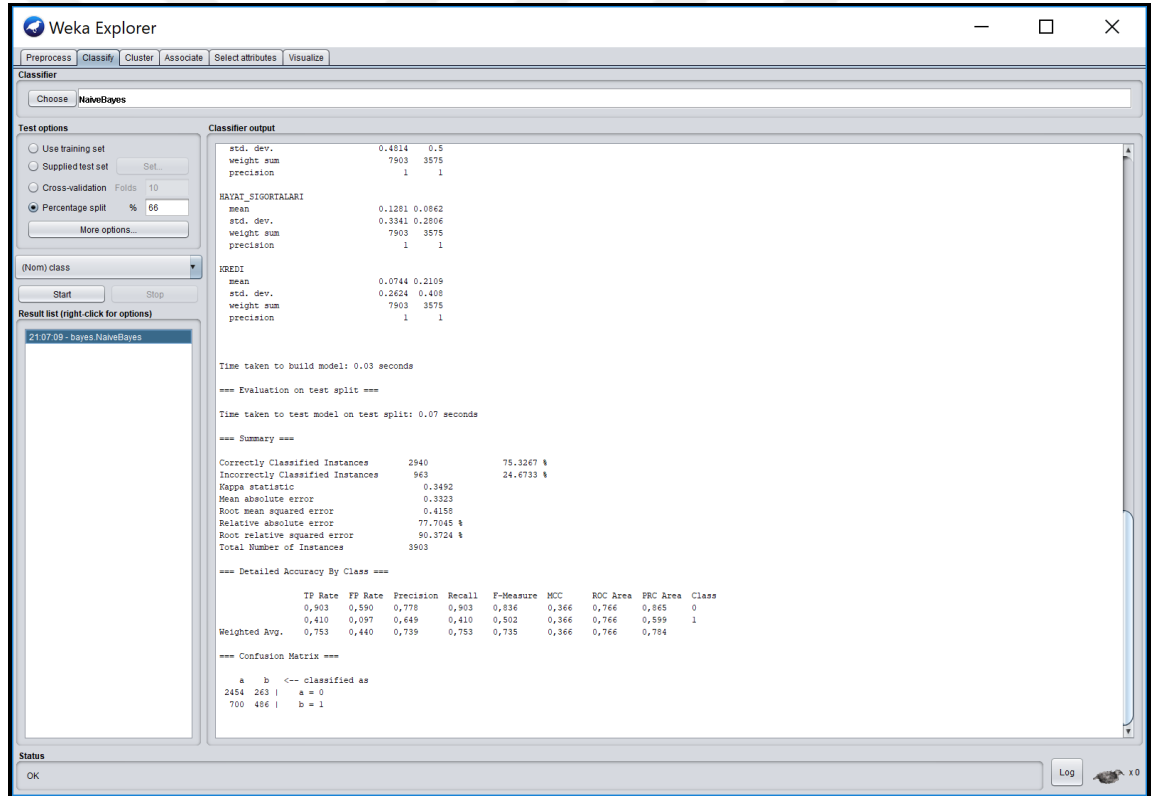
oranlarının hesaplanması için de belirli bir yöntemden faydalandığı görülmektedir. “Kötümser budama” ismi verilen bu yöntemde ağaç üzerinde bulunan her düğüm için üst güven sınırındaki iki terimli dağılım istatistiksel tablolarından yararlanıldığı söylenebilir. Verilmekte olan düğümler için C4.5 algoritmasında yüzde 25 güven sınırlaması kullanılmaktadır. Ağaçta yeniden yaprak oluşturabilmek için alttaki ağacın kaldırılıp yerine kök düğüm konulması gerekmektedir. Buradaki amaç istenen budamanın yapılıp yapılamayacağını tespit edilmesidir. Mevcutta olan ağacın beklenen hata değerinin yerine geçeceği ağacın beklenen hata değerinden daha yüksek olması durumunda karar ağacında budama ve alttaki ağaçta yeni yaprak düğüm konması gerekir (Özkan 2016, ss. 79-80).



4. BULGULAR

Oluşturulan veri seti ile incelendiğinde ulaşılmak istenen sonuç banka dijital kanal kullanan müşterilerin eğilimlerinin tespit edilmesidir. Datada weka programındaki algoritmalar kullanılarak müşteri eğilimlerini en iyi tespit eden algoritmanın seçilmesi hedeflenmiştir. Seçilen veri madenciliği sınıflandırma yöntemleri algoritmaları Naive Bayes, Logistic, Ibk ve J48 şeklindedir. Bu bölümde algoritmalar çalıştırılma sırasına ve ortaya çıkan bulgulara göre analiz edilmiştir.

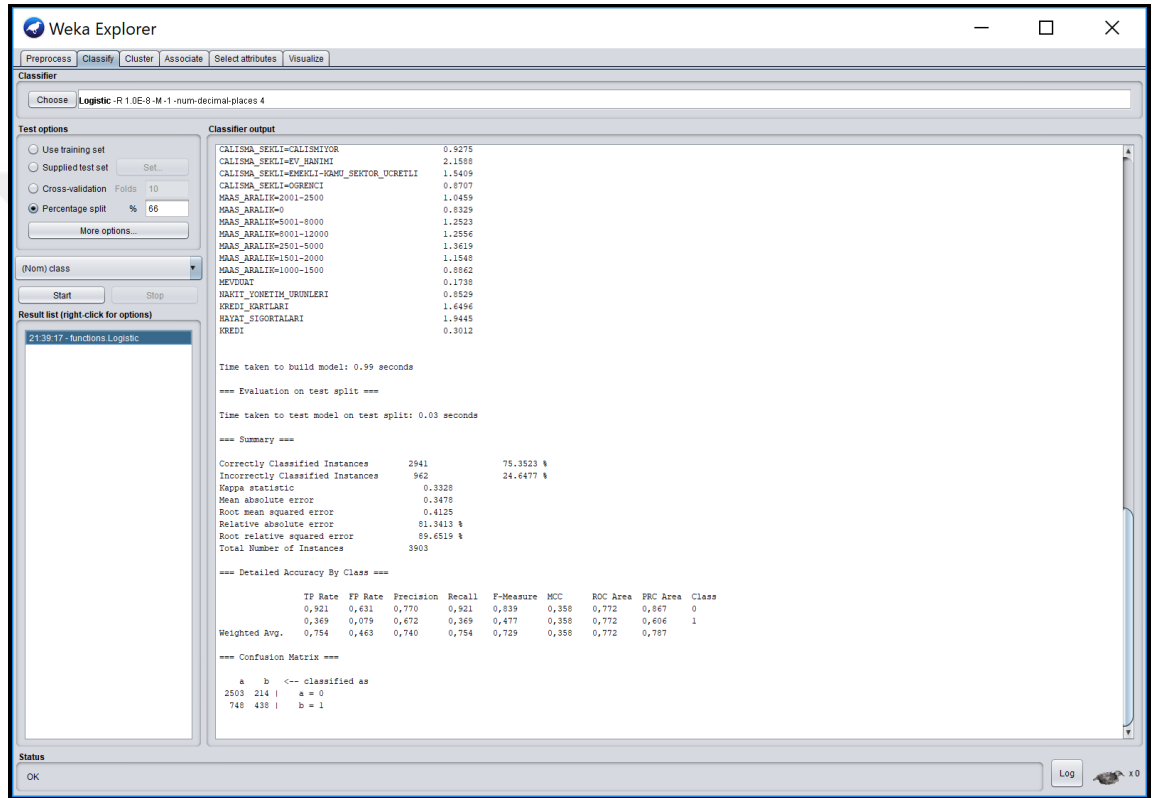
Şekil 4.1: Naive Bayes Algoritması Sonucu



Şekil 4.1'de Naive Bayes algoritmasının sınıflandırma yüzdesinin doğruluğuna baktığımızda oranın yüzde 75,33 olduğu görülmektedir. Dijital kanal kullananlar 1 numaralı sınıfta gösterilirken, dijital kanalı olmayanlar 0 olarak gösterilmiştir. Hata matrisi incelendiğinde de bu durumu destekleyen sonuçlar görülmektedir. Doğru olarak sınıflandırma yapılan 1.sınıfta gözlem sayısı 2.454 ve 1. sınıfta yer alması gerekirken 0. sınıfta yer alan gözlemler 263; 0. sınıfta doğru olarak sınıflandırılan gözlem sayısı 486

ve 0. sınıfta yer alması gerekirken 1. sınıfta olan gözlem sayısı 700'ü göstermektedir. Bununla birlikte Precision incelediğimizde yüzde 73,9 ile doğru duyarlılıkta olan algoritmanın yüzdesini destekleyici sonuçlar çıkardığı söylenebilir. Kök ortalama hata kareler RMSE düşük olması gerekir. Burada yer alan kök ortalama hata karelerinin değeri de 0,42'yi gösterir ve bu durum sonuç olarak kabul edilebilir.

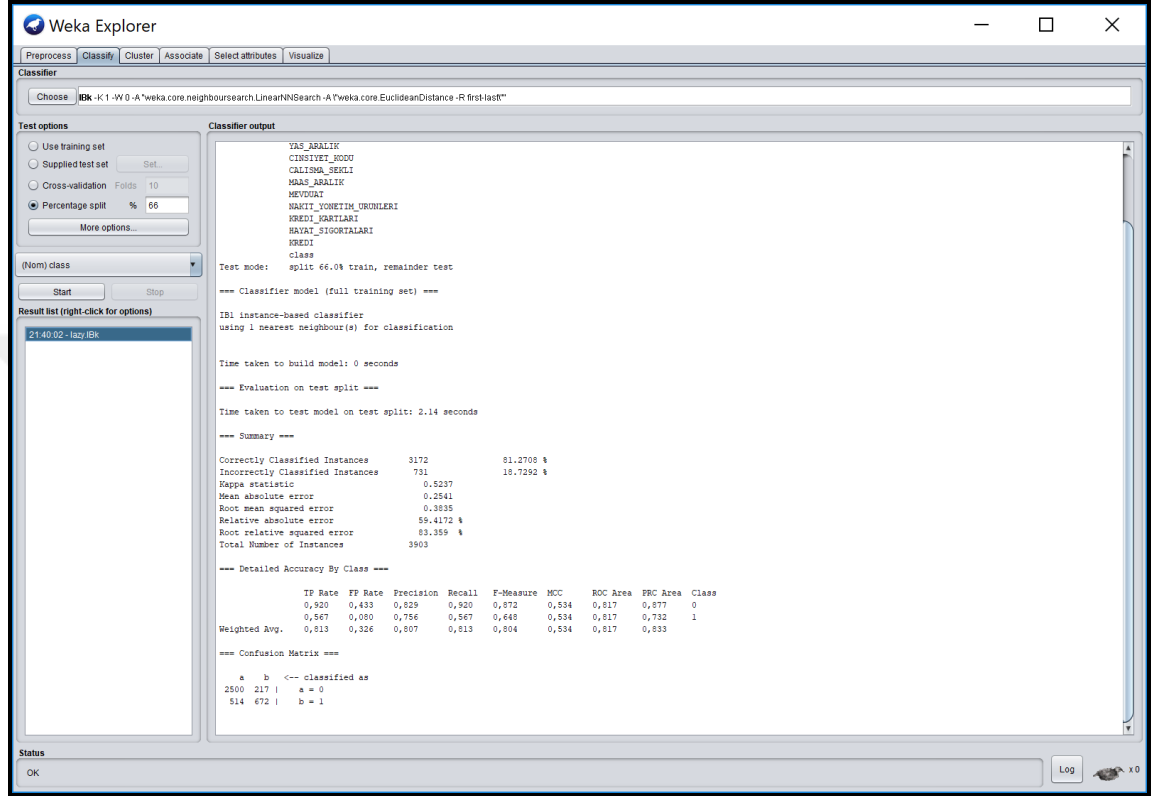
Şekil 4.2: Logistic Algoritması Sonucu



Şekil 4.2'de Logistic algoritmasının sınıflandırma yüzdesinin doğruluğuna baktığımızda oranın yüzde 75,35 olduğu görülmektedir. Dijital kanal kullananlar 1 numaralı sınıfta gösterilirken, dijital kanalı olmayanlar 0 olarak gösterilmiştir. Hata matrisi incelendiğinde de bu durumu destekleyen sonuçlar görülmektedir. Doğru olarak sınıflandırma yapılan 1.sınıfta gözlem sayısı 2.503 ve 1. sınıfta yer alması gerekirken 0. sınıfta yer alan gözlemler 214; 0. sınıfta doğru olarak sınıflandırılan gözlem sayısı 438 ve 0. sınıfta yer alması gerekirken 1. sınıfta olan gözlem sayısı 748'i göstermektedir. Bununla birlikte Precision incelediğimizde yüzde 74 ile doğru duyarlılıkta olan algoritmanın yüzdesini destekleyici sonuçlar çıkardığı söylenebilir. Kök ortalama hata

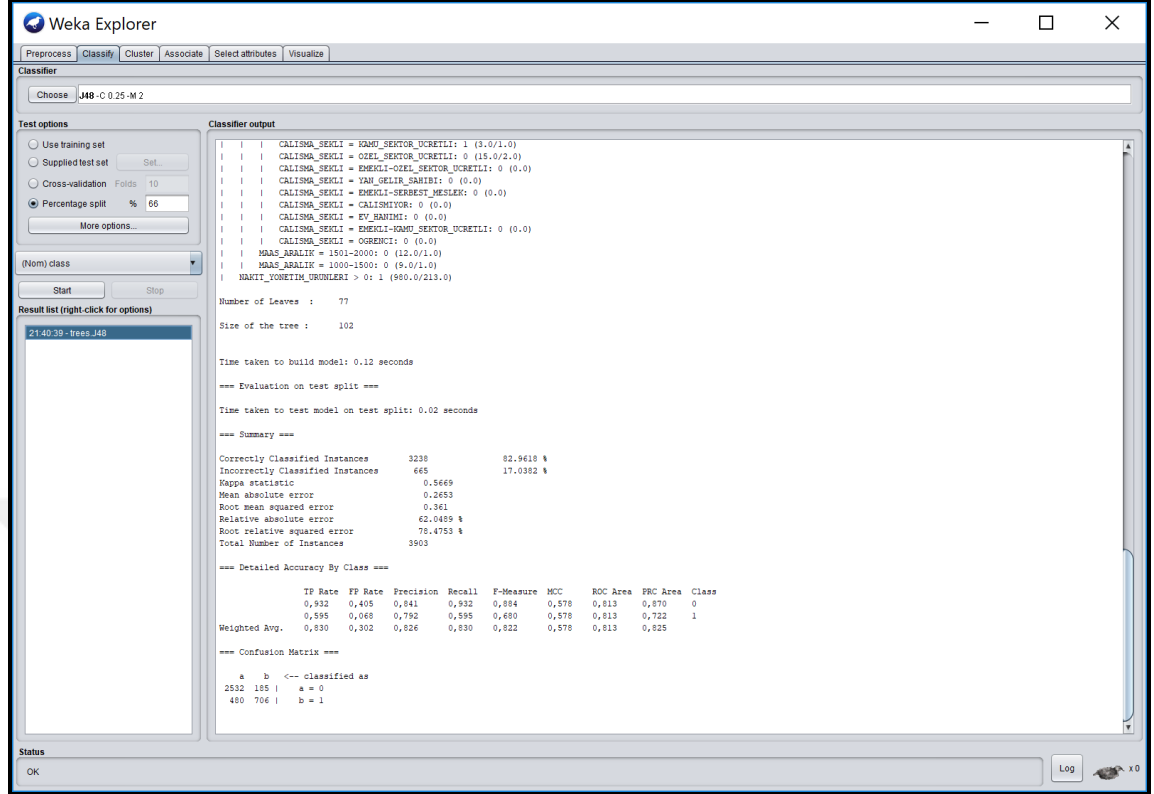
kareler RMSE düşük olması gerekir. Burada yer alan kök ortalama hata karelerinin değeri de 0,41'i gösterir ve bu durum sonuç olarak kabul edilebilir.

Şekil 4.3: Ibk Algoritması Sonucu



Şekil 4.3. Ibk algoritmasının sınıflandırma yüzdesinin doğruluğuna baktığımızda oranın yüzde 81,27 olduğu görülmektedir. Dijital kanal kullananlar 1 numaralı sınıfta gösterilirken, dijital kanalı olmayanlar 0 olarak gösterilmiştir. Hata matrisi incelendiğinde de bu durumu destekleyen sonuçlar görülmektedir. Doğru olarak sınıflandırma yapılan 1.sınıfta gözlem sayısı 2.500 ve 1. sınıfta yer alması gerekirken 0. sınıfta yer alan gözlemler 217; 0. sınıfta doğru olarak sınıflandırılan gözlem sayısı 672 ve 0. sınıfta yer alması gerekirken 1. sınıfta olan gözlem sayısı 514'u göstermektedir. Bununla birlikte Precision incelediğimizde yüzde 80,7 ile doğru duyarlılıkta olan algoritmanın yüzdesini destekleyici sonuçlar çıkardığı söylenebilir. Kök ortalama hata kareler RMSE düşük olması gerekir. Burada yer alan kök ortalama hata karelerinin değeri de 0,38'i gösterir ve bu durum sonuç olarak kabul edilebilir.

Şekil 4.4: J48 Algoritması Sonucu



Şekil 4.4'te J48 algoritmasının sınıflandırma yüzdesinin doğruluğuna baktığımızda oranın yüzde 82,96 olduğu görülmektedir. Dijital kanal kullananlar 1 numaralı sınıfta gösterilirken, dijital kanalı olmayanlar 0 olarak gösterilmiştir. Hata matrisi incelendiğinde de bu durumu destekleyen sonuçlar görülmektedir. Doğru olarak sınıflandırma yapılan 1.sınıfta gözlem sayısı 2.532 ve 1. sınıfta yer alması gerekirken 0. sınıfta yer alan gözlemler 185; 0. sınıfta doğru olarak sınıflandırılan gözlem sayısı 706 ve 0. sınıfta yer alması gerekirken 1. sınıfta olan gözlem sayısı 480'i göstermektedir. Bununla birlikte Precision incelediğimizde yüzde 82,60 ile doğru duyarlılıkta olan algoritmanın yüzdesini destekleyici sonuçlar çıkardığı söylenebilir. Kök ortalama hata kareler RMSE düşük olması gerekir. Burada yer alan kök ortalama hata karelerinin değeri de 0,36'yı gösterir ve bu durum sonuç olarak kabul edilebilir.

Tablo 4.1: Sınıflandırma Algoritmalarının Karşılaştırılması

Algoritma	ACC	TPR	FP Rate	Precision	RMSE	F-Measure	ROC Area
Navie_Bayes	75,33%	75,30%	44,00%	73,90%	0,42	73,50%	76,60%
Logistic	75,35%	75,40%	46,30%	74%	41,25%	72,90%	77,20%
IBk	81,27%	81,30%	32,60%	80,70%	38,35%	80,40%	81,70%
J48	82,96%	83%	30,20%	82,60%	36,10%	82,20%	81,30%

Tablo 4.1 de dijital kanal kullanan ve kullanmayan 11.478 müşterileri için Wekadaki sınıflandırma algoritmaları değerlendirilmiştir.

J48 algoritması, gözlem sonrası doğru sınıflandırılmaların yüzdeleri gösteren ACC'si yüzde 82,96 ve algoritmanın duyarlılığını gösteren precision yüzdesi 82,60 ile değerlendirmeler arasında en yüksek başarıya sahiptir. Bu değerler göz önünde bulundurulduğunda J48 algoritma sonucunun banka dijital kanal kullanan müşterilerin eğilimlerinin tespiti konusunda en başarılı olduğu söylenebilir.

Şekil 4.5: J48 Algoritması Karar Ağacı Detayları

```

J48 pruned tree
-----
MEVDUAT <= 0
| KREDI <= 0
| | KREDI_KARTLARI <= 0
| | | NAKIT_YONETIM_URUNLERI <= 0
| | | | HAYAT_SIGORTALARI <= 0: 1 (751.0)
| | | | HAYAT_SIGORTALARI > 0: 0 (670.0/26.0)
| | | | NAKIT_YONETIM_URUNLERI > 0: 0 (1915.0/288.0)
| | | KREDI_KARTLARI > 0: 0 (5209.0/820.0)
| | KREDI > 0
| | | NAKIT_YONETIM_URUNLERI <= 0: 0 (330.0/105.0)
| | | NAKIT_YONETIM_URUNLERI > 0
| | | | CALISMA_SEKLI = EMEKLI: 0 (24.0/7.0)
| | | | CALISMA_SEKLI = KGMU_SEKTOR_UCRETLI: 0 (140.0/65.0)
| | | | CALISMA_SEKLI = OZEL_SEKTOR_UCRETLI
| | | | | YAS_ARALIK = 54-64
| | | | | | MAAS_ARALIK = 2001-2500: 1 (0.0)
| | | | | | MAAS_ARALIK = 0: 1 (10.0/4.0)
| | | | | | MAAS_ARALIK = 5001-8000: 0 (1.0)
| | | | | | MAAS_ARALIK = 8001-12000: 1 (2.0)
| | | | | | MAAS_ARALIK = 2501-5000
| | | | | | | CINSIYET_KODU <= 0: 0 (2.0)
| | | | | | | CINSIYET_KODU > 0: 1 (2.0)
| | | | | | | MAAS_ARALIK = 1501-2000: 0 (1.0)
| | | | | | | MAAS_ARALIK = 1000-1500: 0 (1.0)
| | | | | | | YAS_ARALIK = 47-53: 1 (77.0/33.0)
| | | | | | | YAS_ARALIK = 65+: 0 (9.0/2.0)
| | | | | | | YAS_ARALIK = 41-46
| | | | | | | KREDI_KARTLARI <= 0: 0 (26.0/8.0)
| | | | | | | KREDI_KARTLARI > 0: 1 (72.0/33.0)
| | | | | | | YAS_ARALIK = 36-40
| | | | | | | CINSIYET_KODU <= 0: 1 (58.0/22.0)
| | | | | | | CINSIYET_KODU > 0: 0 (17.0/7.0)
| | | | | | | YAS_ARALIK = 25-30
| | | | | | | | MAAS_ARALIK = 2001-2500: 1 (0.0)
| | | | | | | | MAAS_ARALIK = 0: 1 (14.0/4.0)
| | | | | | | | MAAS_ARALIK = 5001-8000: 1 (0.0)
| | | | | | | | MAAS_ARALIK = 8001-12000: 1 (0.0)
| | | | | | | | MAAS_ARALIK = 2501-5000: 0 (6.0/1.0)
| | | | | | | | MAAS_ARALIK = 1501-2000: 0 (3.0/1.0)
| | | | | | | | MAAS_ARALIK = 1000-1500: 1 (7.0/3.0)
| | | | | | | | YAS_ARALIK = 31-35
| | | | | | | | KREDI_KARTLARI <= 0
| | | | | | | | | HAYAT_SIGORTALARI <= 0: 0 (5.0/1.0)
| | | | | | | | | HAYAT_SIGORTALARI > 0: 1 (2.0)
| | | | | | | | | KREDI_KARTLARI > 0: 1 (28.0/8.0)
| | | | | | | | | YAS_ARALIK = 18-24: 1 (10.0/2.0)
| | | | | | | | | CALISMA_SEKLI = EMEKLI-OZEL_SEKTOR_UCRETLI: 0 (8.0/2.0)
| | | | | | | | | CALISMA_SEKLI = YAN_GELIR_SAHIBI: 0 (1.0)
| | | | | | | | | CALISMA_SEKLI = EMEKLI-SERBEST_MESLEK: 0 (11.0/2.0)
| | | | | | | | | CALISMA_SEKLI = CALISMİYOR: 1 (3.0)
| | | | | | | | | CALISMA_SEKLI = EV_HANIMI: 1 (0.0)
| | | | | | | | | CALISMA_SEKLI = EMEKLI-KGMU_SEKTOR_UCRETLI: 1 (0.0)
| | | | | | | | | CALISMA_SEKLI = OGRENCI: 1 (0.0)
MEVDUAT > 0
| NAKIT_YONETIM_URUNLERI <= 0
| | MAAS_ARALIK = 2001-2500: 0 (2.0)
| | MAAS_ARALIK = 0
| | | YAS_ARALIK = 54-64: 0 (124.0/39.0)
| | | YAS_ARALIK = 47-53
| | | | HAYAT_SIGORTALARI <= 0
| | | | | CALISMA_SEKLI = EMEKLI: 0 (6.0/3.0)
| | | | | CALISMA_SEKLI = KGMU_SEKTOR_UCRETLI
| | | | | | KREDI_KARTLARI <= 0: 0 (23.0/7.0)
| | | | | | KREDI_KARTLARI > 0: 1 (13.0/2.0)
| | | | | CALISMA_SEKLI = OZEL_SEKTOR_UCRETLI
| | | | | | KREDI <= 0: 0 (82.0/33.0)
| | | | | | KREDI > 0: 1 (10.0/3.0)
| | | | | CALISMA_SEKLI = EMEKLI-OZEL_SEKTOR_UCRETLI: 1 (4.0/1.0)
| | | | | CALISMA_SEKLI = YAN_GELIR_SAHIBI: 0 (5.0)
| | | | | CALISMA_SEKLI = EMEKLI-SERBEST_MESLEK: 0 (4.0)
| | | | | CALISMA_SEKLI = CALISMİYOR: 0 (1.0)
| | | | | CALISMA_SEKLI = EV_HANIMI: 0 (0.0)
| | | | | CALISMA_SEKLI = EMEKLI-KGMU_SEKTOR_UCRETLI: 0 (0.0)
| | | | | CALISMA_SEKLI = OGRENCI: 0 (0.0)
| | | | | HAYAT_SIGORTALARI > 0: 1 (8.0/1.0)
| | | | | YAS_ARALIK = 65+: 0 (37.0/9.0)
| | | | | YAS_ARALIK = 41-46: 1 (191.0/73.0)
| | | | | YAS_ARALIK = 36-40
| | | | | | CINSIYET_KODU <= 0: 1 (82.0/27.0)
| | | | | | CINSIYET_KODU > 0
| | | | | | KREDI_KARTLARI <= 0: 0 (39.0/16.0)
| | | | | | KREDI_KARTLARI > 0: 1 (16.0/5.0)
| | | | | | YAS_ARALIK = 25-30: 1 (126.0/32.0)
| | | | | | YAS_ARALIK = 31-35: 1 (113.0/42.0)
| | | | | | YAS_ARALIK = 18-24: 1 (57.0/11.0)
| | | | | MAAS_ARALIK = 5001-8000: 0 (8.0)
| | | | | MAAS_ARALIK = 8001-12000: 0 (3.0)
| | | | | MAAS_ARALIK = 2501-5000
| | | | | | CALISMA_SEKLI = EMEKLI: 0 (0.0)
| | | | | | CALISMA_SEKLI = KGMU_SEKTOR_UCRETLI: 1 (3.0/1.0)
| | | | | | CALISMA_SEKLI = OZEL_SEKTOR_UCRETLI: 0 (15.0/2.0)
| | | | | | CALISMA_SEKLI = EMEKLI-OZEL_SEKTOR_UCRETLI: 0 (0.0)
| | | | | | CALISMA_SEKLI = YAN_GELIR_SAHIBI: 0 (0.0)
| | | | | | CALISMA_SEKLI = EMEKLI-SERBEST_MESLEK: 0 (0.0)
| | | | | | CALISMA_SEKLI = CALISMİYOR: 0 (0.0)
| | | | | | CALISMA_SEKLI = EV_HANIMI: 0 (0.0)
| | | | | | CALISMA_SEKLI = EMEKLI-KGMU_SEKTOR_UCRETLI: 0 (0.0)
| | | | | | CALISMA_SEKLI = OGRENCI: 0 (0.0)
| | MAAS_ARALIK = 1501-2000: 0 (12.0/1.0)

```

Şekil 4.5'deki J48 algoritma sonucuna göre dijital kanal kullanan müşterilerin özellikleri detaylı olarak aşağıda belirtilmiştir.

Kredisi ve nakit yönetim ürünleri olan, özel sektörde çalışan, yaşı 54-64 aralığında, maaş almayan 10 kişinin 4'ü dijital kanal kullanmaktadır. Müşteri payı yüzde 0,53 ve kullanım oranı yüzde 40'dır.

Kredisi ve nakit yönetim ürünleri olan, özel sektörde çalışan, yaşı 54-64 aralığında, maaşı 8.001-12.000 olan 2 kişinin 2'si dijital kanal kullanmaktadır. Müşteri payı yüzde 0,11 ve kullanım oranı yüzde 100'dür.

Kredisi ve nakit yönetim ürünleri olan, özel sektörde çalışan, yaşı 54-64 aralığında, maaşı 2.501-5.000 olan, cinsiyeti kadın olan 2 kişinin 2'si dijital kanal kullanmaktadır. Müşteri payı yüzde 0,11 ve kullanım oranı yüzde 100'dür.

Kredisi ve nakit yönetim ürünleri olan, özel sektörde çalışan, yaşı 47-53 aralığında 77 kişinin 33'ü dijital kanal kullanmaktadır. Müşteri payı yüzde 4,08 ve kullanım oranı yüzde 42,9'dur.

Kredisi, kredi kartı ve nakit yönetim ürünleri olan, özel sektörde çalışan, yaşı 41-46 aralığında 72 kişinin 33'ü dijital kanal kullanmaktadır. Müşteri payı yüzde 3,82 ve kullanım oranı yüzde 45,8'dir.

Kredisi ve nakit yönetim ürünleri olan, özel sektörde çalışan, yaşı 36-40 aralığında, cinsiyeti erkek olan 58 kişinin 22'si dijital kanal kullanmaktadır. Müşteri payı yüzde 3,08 ve kullanım oranı yüzde 37,9'dur.

Kredisi ve nakit yönetim ürünleri olan, özel sektörde çalışan, yaşı 25-30 aralığında, maaş almayan 14 kişinin 4'ü dijital kanal kullanmaktadır. Müşteri payı yüzde 0,74 ve kullanım oranı yüzde 28,6'dır.

Kredisi ve nakit yönetim ürünleri olan, özel sektörde çalışan, yaşı 25-30 aralığında, maaşı 1.000-1.500 aralığında 7 kişinin 3'ü dijital kanal kullanmaktadır. Müşteri payı yüzde 0,37 ve kullanım oranı yüzde 42,9'dur.

Kredisi, hayat sigortası ve nakit yönetim ürünleri olan, özel sektörde çalışan, yaşı 31-35 aralığında, maaşı almayan 2 kişinin 2'si dijital kanal kullanmaktadır. Müşteri payı yüzde 0,11 ve kullanım oranı yüzde 100'dür.

Kredisi, kredi kartı ve nakit yönetim ürünleri olan, özel sektörde çalışan, yaşı 31-35 aralığında 28 kişinin 8'i dijital kanal kullanmaktadır. Müşteri payı yüzde 1,49 ve kullanım oranı yüzde 28,6'dır.

Kredisi ve nakit yönetim ürünleri olan, özel sektörde çalışan, yaşı 18-24 aralığında 10 kişinin 2'si dijital kanal kullanmaktadır. Müşteri payı yüzde 0,53 ve kullanım oranı yüzde 20'dir.

Mevduat ve kredi kartı ürünleri olan, kamu sektöründe çalışan, yaşı 47-53 aralığında, maaş almayan 13 kişinin 2'si dijital kanal kullanmaktadır. Müşteri payı yüzde 0,69 ve kullanım oranı yüzde 15,4'dür.

Mevduat ürünü olan, özel sektörden emekli, yaşı 47-53 aralığında, maaş almayan 4 kişinin 1'i dijital kanal kullanmaktadır. Müşteri payı yüzde 0,21 ve kullanım oranı yüzde 25'dir.

Mevduat ve hayat sigortası ürünleri olan, yaşı 47-53 aralığında, maaş almayan 8 kişinin 1'i dijital kanal kullanmaktadır. Müşteri payı yüzde 0,42 ve kullanım oranı yüzde 12,5'dir.

Mevduat ürünü olan, yaşı 41-46 aralığında, maaş almayan 191 kişinin 73'ü dijital kanal kullanmaktadır. Müşteri payı yüzde 10,13 ve kullanım oranı yüzde 38,2'dir.

Mevduat ürünü olan, yaşı 36-40 aralığında, maaş almayan, cinsiyeti erkek olan 92 kişinin 27'si dijital kanal kullanmaktadır. Müşteri payı yüzde 4,88 ve kullanım oranı yüzde 29,3'dür.

Mevduat ve kredi kartı ürünleri olan, yaşı 36-40 aralığında, maaş almayan, cinsiyeti kadın olan 16 kişinin 5'i dijital kanal kullanmaktadır. Müşteri payı yüzde 0,85 ve kullanım oranı yüzde 31,3'dür.

Mevduat ürünü olan, yaşı 25-30 aralığında, maaş almayan 126 kişinin 32'si dijital kanal kullanmaktadır. Müşteri payı yüzde 6,68 ve kullanım oranı yüzde 25,4'dür.

Mevduat ürünü olan, yaşı 31-35 aralığında, maaş almayan 113 kişinin 42'si dijital kanal kullanmaktadır. Müşteri payı yüzde 5,99 ve kullanım oranı yüzde 37,2'dir.

Mevduat ürünü olan, yaşı 18-24 aralığında, maaş almayan 57 kişinin 11'i dijital kanal kullanmaktadır. Müşteri payı yüzde 3,02 ve kullanım oranı yüzde 19,3'dür.

Mevduat ürünü olan, kamu sektöründe çalışan, maaşı 2501-5000 aralığında 3 kişinin 1'i dijital kanal kullanmaktadır. Müşteri payı yüzde 0,16 ve kullanım oranı yüzde 33,3'dür.

Mevduat ve nakit yönetim ürünleri olan 980 kişinin 213'ü dijital kanal kullanmaktadır. Müşteri payı yüzde 51,99 ve kullanım oranı yüzde 21,7'dir.

Algoritma detaylarını incelediğimizde nakit yönetimi ve kredi ürünlerine sahip, özel sektörde çalışan müşterilerin yaş, maaş durumu ve cinsiyetten bağımsız olarak 282 kişisinden 115'i dijital kanal kullanmaktadır. Mevduat ürünü olup maaş almayan müşterilerin yaş ve cinsiyetten bağımsız olarak 620 kişisinden 194'ü dijital kanal kullanmaktadır.

Bulguların detaylarından yararlanılarak dijital kanal kullanımına yönlendirilecek müşterilerin profilleri de ortaya çıkmıştır.

5. TARTIŞMA VE SONUÇ

Veri madenciliği çok büyük verilerden anlamlı bilgiler sağlayarak, birden çok sektöre fayda sağlamaktadır. Bu anlamlı bilgiler geleneksel yöntemler ile ortaya çıkarılamayacağı için veri madenciliğine ihtiyaç duyulmuştur. Bankacılık sektörü de diğer sektörler gibi veri madenciliği uygulamalarını kullanmaktadır. Özellikle müşterilerin profilleri doğrultusunda ihtiyaçlarını önceden tahmin etme ve ürün önerme konusunda sektörde tercih edilen bir yöntemdir.

Bu çalışmada verinin tanımı, veri madenciliğinin tanımı, aşamaları ve yöntemleri, veri madenciliğinde karşılaşılan sorunlar detaylı olarak ele alınmıştır. Aynı zamanda bankacılık sektöründe dijital kanalların kullanımı, avantajları ve tercih sebepleri incelenmiştir. Buna göre dijital kanal ürünü kullanan banka müşterilerinin profilleri belirlenip dijital kanal kullanmayan müşterilere ürün önerilmesi hedeflenmiştir.

Uygulama Weka programı kullanılarak yapılmıştır. Veri Türkiye’de hizmet veren özel bir bankanın dijital kanal kullanan müşterilerinden oluşmaktadır. Aynı zamanda müşterilerin yaş, cinsiyet, çalışma şekli, maaş bilgisi ve en çok kullandıkları beş farklı ürünün detaylarını içermektedir. Verideki özellikler çok geniş aralıklara sahip olduğu için gruplamalar yapılarak algoritmalar çalıştırılmıştır.

Bu çalışmada Navie bayes, Logistic, IBk ve J48 algoritmaları kullanılmıştır. En yüksek doğruluk oranı yüzde 82,96 ile J48’dir. J48’in karar ağacı sonucuna göre; mevduat ve nakit yönetim ürünleri olan, mevduat ürünü olan yaşı 41-46 aralığında maaş almayan, mevduat ürünü olan yaşı 25-30 aralığında maaş almayan, mevduat ürünü olan yaşı 31-35 aralığında maaş almayan müşterilerin dijital kanal ürünlerini yoğun olarak kullandıkları görülmüştür. Buna göre ilgili özelliklere sahip 1.410 müşterinin yüzde 25,5’i olan 360 müşterinin dijital kanal kullandığı tespit edilmiştir. Bu profile uygun diğer müşterilere dijital kanal önerilebileceği ön görülmüştür.

Günümüzde birçok bankacılık işlemi (fatura ödeme, havale, kredi kartı borcu ödeme, vb.) dijital kanallar üzerinden online olarak yapıldığında bankaların bazı maliyetlerini

(personel, kırtasiye, vb.) minimize etmektedir. Bu çalışmada dijital kanalları kullanan müşterilerin özelliklerinin tespiti yapılmış ve aynı profile sahip kullanmayan müşterilerin belirlenerek, kullanmalarına teşvikinin sağlanması yönünde kuruma faydası amaçlanmıştır.



KAYNAKÇA

Kitaplar

- Akküçük, U., 2011. *Veri Madenciliği Kümeleme ve Sınıflama Algoritmaları*. 1.Baskı. İstanbul: Yalın Yayıncılık.
- Ersöz, F., 2013. *Veri Madenciliği ve Uygulamaları*. 1.Baskı. Ankara: Sage Yayıncılık.
- Michael J.A. Berry & Gordon S. Linoff. (2004). *Data Mining Techniques For Marketing, Sales, and CRM*, 2.bs, Indianapolis: Wiley Publishing Inc.
- Özkan, Y., 2016. *Veri Madenciliği Yönetimleri*. 3.Baskı. İstanbul: Papatya Yayıncılık.
- Rokach, L. & Maimon, O., (2008). *Data Mining with Decision Trees Theory and Applications*. Machine Perception and Artificial Intelligence, 81, 100.
- Silahtaroglu, G., 2016. *Veri Madenciliği Kavram ve Algoritmaları*. 3.Baskı. İstanbul: Papatya Yayıncılık.
- Witten, I. & Frank, E. & Hall, M & Pal, Christopher J. (2011). *Data Mining Practical Machine Learning Tools And Techniques*. 2nd ed. San Francisco: Morgan Kaufmann.

Sürelî Yayınlar

Academia (online), Lojistik Regresyon, 2013,

http://www.academia.edu/11479607/Lojistik_Regresyon (erişim tarihi 23.04.2018)



Diğer Yayınlar

- Ayığılıoğlu, A. (2016). İnternet Bankacılığının Başarısını Etkileyen Faktörlerin Tespiti: Türkiye ve Çin Karşılaştırması. *Yayınlanmamış Yüksek Lisans Tezi*. Ankara: Gazi Üniversitesi SBE.
- Bilen, H. (2009). Bankacılık Sektöründe Personel Seçimi Ve Performans Değerlendirilmesine İlişkin Veri Madenciliği Uygulaması. *Yayınlanmamış Yüksek Lisans Tezi*. Ankara: Gazi Üniversitesi FBE.
- Bilgin, Ş (2013). Banka Müşterilerinin İnternet Bankacılığına İlişkin Yaklaşımlarının Veri Madenciliği Teknikleri İle İncelenmesi. *Yayınlanmamış Yüksek Lisans Tezi*. İstanbul: İstanbul Üniversitesi SBE.
- Boyacı, A. (2017). Öğretmenlerin Algılanan Örgütsel Destek Ve Örgütsel Özdeşleme Düzeylerinin Veri Madenciliği İle Analizi. *Yayınlanmamış Yüksek Lisans Tezi*. Çorum: Hitit Üniversitesi SBE.
- Coşkun, C. (2010). Veri Madenciliği Algoritmaları Karşılaştırılması. *Yayınlanmamış Yüksek Lisans Tezi*. Diyarbakır: Dicle Üniversitesi FBE.
- Çalış, A. (2013). Veri Madenciliği Yaklaşımı İle Bireysel Müşterilerin Kredi Ödeme Performanslarının Değerlendirilmesi. *Yayınlanmamış Yüksek Lisans Tezi*. Kocaeli: Kocaeli Üniversitesi FBE.
- Çığışar, B. (2017). Kredi Risklerinde Veri Madenciliği Sınıflandırma Algoritmaları. *Yayınlanmamış Yüksek Lisans Tezi*. Adana: Çukurova Üniversitesi FBE.
- Çil, F. (2010). Banka Yatırım Fonu Müşteri Hareketlerinin Belirlenmesine Yönelik Bir Veri Madenciliği Uygulaması. *Yayınlanmamış Yüksek Lisans Tezi*. Ankara: Gazi Üniversitesi FBE.
- Demirel, A.C. (2017). Dijital Bankacılık Ve Türkiye’deki Mevcut Durumun Analizi. *Yayınlanmamış Yüksek Lisans Tezi*. Ankara: Başkent Üniversitesi SBE.
- Dikkatli, M. (2016). Alternatif Dağıtım Kanallarının Müşteriler Üzerine Etkisi: Şube Ve

- İnternet Bankacılığı Arasında Bir Karşılaştırma. *Yayınlanmamış Yüksek Lisans Tezi*. İstanbul: Beykent Üniversitesi SBE.
- Erkuş, S. (2015). Veri Madenciliği Yöntemleri İle Kardiyovasküler Hastalık Tahmini Yapılması. *Yayınlanmamış Yüksek Lisans Tezi*. İstanbul: Bahçeşehir Üniversitesi FBE.
- İşler, A.U. (2015). Bankacılık Sektöründe İnternet Bankacılığı Ve Mobil Bankacılık Uygulamalarının E-Hizmet Kalitesinin Kullanıcıları Tarafından Değerlendirilmesi: Kuveyt Türk Örneği. *Yayınlanmamış Yüksek Lisans Tezi*. Karabük: Karabük Üniversitesi SBE.
- Karadeniz, N. (2008). Müşteri İlişkileri Yönetimi Açısından Veri Madenciliği Yöntemi Ve Hizmet Sektörü Üzerine Bir Uygulama. *Yayınlanmamış Yüksek Lisans Tezi*. İstanbul: Marmara Üniversitesi SBE.
- Karadirek, G (2014). Tüketicilerin Elektronik Hizmet Kalitesi Algılamaları: Bireysel İnternet Bankacılığı Müşterileri Üzerine Bir Uygulama. *Yayınlanmamış Yüksek Lisans Tezi*. Gümüşhane: Gümüşhane Üniversitesi SBE.
- Karahan Adalı, G. (2017). Veri Madenciliğinde Birliktelik Yöntemleri Ve Müşteri İlişkileri Yönetimine İlişkin Bir Uygulama. *Yayınlanmamış Doktora Tezi*. İstanbul: İstanbul Üniversitesi FBE.
- Mocan, G. (2016). Perakendecilikte Veri Madenciliği Uygulamaları Ve Sorunları. *Yayınlanmamış Yüksek Lisans Tezi*. İstanbul: Yıldız Teknik Üniversitesi FBE.
- Odabaş, Ö. (2017). Veri Madenciliği Teknikleri İle Telekom Sektöründe Ayrılan Müşteri Analizi. *Yayınlanmamış Yüksek Lisans Tezi*. İstanbul: İstanbul Ticaret Üniversitesi FBE.
- Özbal, T (2011). Bankacılık Sektöründe Müşteri İlişkileri Yönetimi Ve İnternet Bankacılığında Müşterilerin Seçimlerine Etki Eden Faktörler. *Yayınlanmamış Yüksek Lisans Tezi*. Ankara: Ufuk Üniversitesi SBE.
- Özoğlu, B (2010). Tüketici Odaklı Yaklaşım İle İnternet Bankacılığı Faaliyetlerini Değerlendirmeye Yönelik Bir Uygulama. *Yayınlanmamış Yüksek Lisans Tezi*. Kayseri: Erciyes Üniversitesi SBE.
- Sivri, E. Ş. (2015). Veri Madenciliği/E-Ticaret İçin Ürün Tavsiye Sistemi Geliştirilmesi.

- Yayınlanmamış Yüksek Lisans Tezi.* İstanbul: İstanbul Ticaret Üniversitesi FBE.
- Tekingöz, Ö. (2016). Dijital Yayıncılıkta İçerik İzleme Oranlarına Göre Müşteri Kümeleneşmesi. *Yayınlanmamış Yüksek Lisans Tezi.* İstanbul: Bahçeşehir Üniversitesi FBE.
- Toksöz, C. (2016). Veri Madenciliği Süreçleri İle Tavuk Yumurtalarında Cinsiyet Ayrımı. *Yayınlanmamış Yüksek Lisans Tezi.* Isparta: Süleyman Demirel Üniversitesi FBE.
- Ulukoş, E. (2011). Bireysel Bankacılıkta Pazarlama Stratejileri. *Yayınlanmamış Yüksek Lisans Tezi.* İstanbul: Kadir Has Üniversitesi SBE.
- Uyumaz, Ö. (2017). Bankacılık Sektöründe Pazarlama Stratejilerinin Belirlenmesinde Sınıflandırma Ve Veri Madenciliği. *Yayınlanmamış Yüksek Lisans Tezi.* İstanbul: Beykent Üniversitesi FBE.
- Ünsal, U (2013). Türkiye’de İnternet Bankacılığı Kullanımını Etkileyen Faktörlerin Belirlenmesi. *Yayınlanmamış Yüksek Lisans Tezi.* İstanbul: Bahçeşehir Üniversitesi FBE.
- Yurdakul, S. (2015). Veri Madenciliği İle Lise Öğrenci Performanslarının Değerlendirilmesi. *Yayınlanmamış Yüksek Lisans Tezi.* Kırıkkale: Kırıkkale Üniversitesi FBE.