

T.C.
BAHÇEŞEHİR ÜNİVERSİTESİ

BİREYSEL KREDİLERİNİ ERKEN KAPATAN
MÜŞTERİLERİN VERİ MADENCİLİĞİ
YAKLAŞIMI İLE ANALİZİ

Yüksek Lisans Tezi

SEÇİL PAFTALI

İSTANBUL, 2021

**T.C.
BAHÇEŞEHİR ÜNİVERSİTESİ**

**LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİLGİ TEKNOLOJİLERİ YÜKSEK LİSANS PROGRAMI**

**BİREYSEL KREDİLERİNİ ERKEN KAPATAN
MÜŞTERİLERİN VERİ MADENCİLİĞİ
YAKLAŞIMI İLE ANALİZİ**

Yüksek Lisans Tezi

SEÇİL PAFTALI

Tez Danışmanı: DR. ÖĞR. ÜYESİ YÜCEL BATU SALMAN

İSTANBUL, 2021



BAHÇEŞEHİR ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

25/01/2021

YÜKSEK LİSANS TEZ ONAY FORMU

Program Adı:	BİLGİ TEKNOLOJİLERİ (TÜRKÇE, TEZLİ)
Öğrencinin Adı Soyadı:	SEÇİL PAFTALI
Tezin Adı:	Bireysel Kredisini Erken Kapanan Müşterilerin Veri Madenciliği Yaklaşımı ile Analizi
Tez Savunma Tarihi:	25.01.2021

Bu tezin Yüksek Lisans tezi olarak gerekli şartları yerine getirmiş olduğu Lisansüstü Eğitim Enstitüsü tarafından onaylanmıştır.

Doç. Dr. Burak KÜNTAY
Enstitü Müdürü

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

	Ünvanı, Adı Soyadı	İmza
Tez Danışmanı:	Dr. Öğr. Üyesi Yücel Batu Salman	
2. Üye :	Dr. Öğr. Üyesi Serkan Ayvaz	
3. Üye :	Dr. Öğr. Üyesi Atınç Yılmaz	



SEVGİLİ EŞİM OZAN'A
İTHAF EDİYORUM.

TEŐEKKÖR

Tez alıřmam boyunca benden tecrübelerini esirgemeyen, deęerli yardım ve katkılarıyla beni yönlendiren danışman hocam Sayın Dr. Öğr. Üyesi Yücel Batu Salman'a ve bu program süresince manevi desteęini hiç eksik etmeyip her zaman anlayışla yanımda olan sevgili eşim Ozan Paftalı başta olmak üzere tüm ailem ve arkadaşlarıma çok teşekkür ederim.

İstanbul, 2021

Seil PAFTALI



ÖZET

BİREYSEL KREDİLERİNİ ERKEN KAPATAN MÜŞTERİLERİN VERİ MADENCİLİĞİ YAKLAŞIMI İLE ANALİZİ

Seçil PAFTALI

Bilgi Teknolojileri

Tez Danışmanı: Dr. Öğr. Üyesi Yücel Batu SALMAN

Şubat 2021, 57 sayfa

Her geçen gün rekabetin arttığı bankacılık sektöründe, sürdürülebilir bir şekilde büyüebilmek için hem yeni müşteri kazanmaları, hem de mevcut müşterilerinin sadakatini sağlamak için faaliyetlerde bulunulması gerekiyor. Müşterilerin memnuniyetinin de sağlanabilmesi için doğru zamanda doğru teklifler sunularak müşteriler ile iletişim kurulmalıdır.

Bu tez çalışmasında özel bir bankanın kredisini erken kapatan müşterileri ile ilgili demografik bir analiz yapılması ve veri madenciliği yaklaşımları ile uygun modelin belirlenmesi hedeflenmiştir. Bu çalışma sonucunda elde edilecek verilere göre bankalar kredisini erken kapatabilecek potansiyel müşterileri işlem öncesinde tespit edebiliyor olacak ve onlara farklı ürünler, çapraz satış yapacak stratejiler geliştirebilecektir.

Anahtar Kelimeler: Bireysel Kredi, Erken Ödeme, Veri Madenciliği, Sınıflandırma Algoritmaları, SPSS

ABSTRACT

THE ANALYSIS WITH DATA MINING APPROACH OF CUSTOMERS WHO CLOSES OFF THEIR CONSUMER LOANS EARLY

Seçil PAFTALI

Information Technologies

Thesis Supervisor: Dr. Öğr. Üyesi Yücel Batu SALMAN

February 2021, 57 pages

In the banking sector, where competition is increasing day by day, it is necessary to engage in activities to gain new customers and to ensure the loyalty of existing customers in order to grow sustainably. In order to ensure customer satisfaction, customers should be contacted by presenting the right offers at the right time.

In this thesis, it is aimed to make a demographic analysis about the customers of a private bank who closed off their loan early and to determine the appropriate model with data mining approaches. According to the data obtained as a result of this study, banks will be able to identify potential customers who can close off their loans early before the transaction and develop strategies that will make them different products and cross-sell.

Keywords: Consumer Loan, Early Payment, Data Mining, Classification Algorithms, SPSS

İÇİNDEKİLER

TABLOLAR.....	ix
ŞEKİLLER.....	x
KISALTMALAR	xi
SEMBOLLER.....	xii
1. GİRİŞ	1
2. LİTERATÜR TARAMASI.....	4
3. VERİ MADENCİLİĞİ	8
3.1 VERİ MADENCİLİĞİ KAVRAMI.....	8
3.2 VERİ MADENCİLİĞİNİN TARİHSEL GELİŞİMİ	9
3.3 VERİ MADENCİLİĞİNİN KULLANIM ALANLARI VE FAYDALARI ..	10
3.4 VERİ MADENCİLİĞİNDE KARŞILAŞILAN PROBLEMLER	12
3.5 VERİ MADENCİLİĞİ SÜREÇLERİ	13
3.5.1 Problemin Tanımlanması	14
3.5.2 Verilerin Hazırlanması.....	14
3.5.3 Modelin Kurulması ve Değerlendirilmesi.....	14
3.5.4 Modelin Kullanılması ve İzlenmesi.....	15
3.6 VERİ MADENCİLİĞİ TEKNİKLERİ	15
3.6.1 Sınıflandırma	16
3.6.2 Kümeleme	21
3.6.3 Birliktelik Kuralları ve Ardışık Örüntüler	22
4. BİREYSEL KREDİ KAVRAMI	23
4.1 KREDİ KAVRAMI.....	23
4.2 KREDİ UNSURLARI	23
4.3 BİREYSEL KREDİ.....	24
4.4 BİREYSEL KREDİ ÇEŞİTLERİ.....	25
5. UYGULAMA	26
5.1 UYGULAMAYA GENEL BAKIŞ.....	26
5.2 UYGULAMADA KULLANILAN YAZILIM.....	26
5.3 PROBLEMİN TANIMLANMASI.....	27

5.4 VERİLERİN HAZIRLANMASI	28
5.4.1 Verilerin Toplanması.....	28
5.4.2 Veri Kümesinin Tanıtımı	29
5.4.3 Verilerin Temizlenmesi	34
5.4.4 Verilerin Dönüştürülmesi.....	34
5.5 MODELİN KURULMASI.....	36
5.5.1 Veri Setinin Uygulamaya Aktarılması ve Hedef Değişkenin Seçimi.....	36
5.5.2 Anormal Verilerin Belirlenmesi ve Veri Setinden Çıkartılması	39
5.5.3 Modelleme.....	42
5.5.3.1 C&R Tree algoritmasına ilişkin sonuçların yorumlanması	44
5.5.3.2 C5.0 algoritmasına ilişkin sonuçların yorumlanması	47
5.5.3.3 Quest algoritmasına ilişkin sonuçların yorumlanması	49
6. SAYISAL SONUÇLAR	52
7. TARTIŞMA VE SONUÇ	56
KAYNAKÇA.....	58
EKLER.....	Error! Bookmark not defined.

TABLolar

Tablo 3.1: Veri madenciliğinin tarihsel gelişimi	10
Tablo 4.1: Bireysel kredi kullandırım oranları (2020, %).	25
Tablo 5.1: Kategori edilmemiş ham veri.....	28
Tablo 5.2: Erken kapatma bilgisine göre dağılım.....	29
Tablo 5.3: Cinsiyet bilgisine göre dağılım	30
Tablo 5.4: Medeni hal değerleri	30
Tablo 5.5: Medeni hal bilgisine göre dağılım	30
Tablo 5.6: Meslek bilgisine göre dağılımda ilk 5	31
Tablo 5.7: İl bilgisine göre dağılımda ilk 5	32
Tablo 5.8: Yaş bilgisine göre veri dağılımda ilk 5	32
Tablo 5.9: Yaş gruplarına göre dağılım	34
Tablo 5.10: Bölge bilgisine göre dağılım.....	35
Tablo 5.11: Vade gruplarına göre dağılım	36
Tablo 6.1: Algoritmaların karşılaştırılması	55

ŞEKİLLER

Şekil 3.1: Veri madenciliğinin etkileşim içerisinde bulunduğu disiplinler	9
Şekil 3.2: Veri madenciliği süreçleri	13
Şekil 3.3: Veri madenciliği teknikleri	16
Şekil 3.4: X ve Y değişkenleri üzerinden uygulanan karar ağacı örneği	19
Şekil 3.5: Yapay sinir ağlarının katmanları	20
Şekil 5.1: SPSS Modeler ekran görüntüsü	27
Şekil 5.2: Verinin programdaki görüntüsü	37
Şekil 5.3: Değişkenlerin “Descriptive Statics” görüntüsü	38
Şekil 5.4: Veri aktarımı ve hedef değişken	39
Şekil 5.5: Anormallik testi sonrası SPSS Modeler ekranı	39
Şekil 5.6: Anormallik testi gruplara göre sonuç tablosu	40
Şekil 5.7: Anormallik testi özet tablosu	40
Şekil 5.8: “Feature Selection” modülünün eklenmesi sonrası SPSS Modeler ekranı	41
Şekil 5.9: Özellik Seçimi (Feature Selection) ekranı	41
Şekil 5.10: Eğitim ve test veri oranlarının belirlenmesi	42
Şekil 5.11: Auto Classifier işlemcisi	43
Şekil 5.12: Auto Classifier işlemcisi sonuçları	44
Şekil 5.13: C&R Tree algoritmasına göre önemli değişkenler	45
Şekil 5.14: C&R Tree algoritması ile oluşan ilk dal	46
Şekil 5.15: C5.0 algoritmasına göre önemli değişkenler	48
Şekil 5.16: Quest algoritmasına göre önemli değişkenler	50
Şekil 6.1: C&R Tree performans ölçüm sonuçları	53
Şekil 6.2: C5.0 algoritmasının performans ölçüm sonuçları	54
Şekil 6.3: Quest algoritmasının performans ölçüm sonuçları	55

KISALTMALAR

C&RT	:	Classification and Regression Tree
CHAID	:	Chi-squared Automatic Interaction Detector
IBM	:	International Business Machines
KKB	:	Kredi Kayıt Bürosu
MS Excel	:	MicroSoft Excel
QUEST	:	Quick, Unbiased, Efficient Statistical Tree
SPSS	:	Statistical package for the social sciences.
SQL	:	Structured Query Language

SEMBOLLER

Yüzde : %



1. GİRİŞ

Bankacılık rekabetin yüksek olduđu bir sektördür. Türkiye’de faaliyet gösteren yerli ve yabancı sermaye ile toplamda 48 banka bulunmaktadır. Bankacılıkta faiz gelirlerinin yasal kurumlar tarafından düzenlenmesi, müşterilerden alınan masraflara sınırlar getirilmesi ile karlılık geçmiş yıllara göre azalmıştır. Bu kadar fazla banka içersinde varolabilmek için müşteri ihtiyaçları doğrultusunda ve müşteri memnuniyeti düşünülerek yenilikleri hızlı yakalamak ve yüksek kaliteli ürünler üretmek, yüksek teknoloji sistemler kurmak ve tüm bunları hem yasalara uygun hem de güvenli şekilde yapmak öncelikli gereksinimlerdir.

Artan rekabet ortamı ile birlikte pazarlama açısından müşteriye odaklanılarak davranışları kapsamlı bir şekilde izlenilmesi zorunlu hale gelmiştir (Petersen 2011). Bunlardan ötürü de şuan ki ekonomide “müşterinin kral olduđu yer olan müşteri ekonomisi” ortamı yaşanmaktadır. Firmalar, ürüne değil müşteriye odaklanmaya geçmek zorundalardır (Kotler 2003).

Pazarlama stratejilerinin üretim merkezlilikten, müşterinin merkeze alınarak oluşturulmaya başlanması ile birlikte modern pazarlama kavramı doğmuştur. Müşteriyi memnun edebilmek için öncelikle müşteri tanınmalıdır. Davranış, satın alma alışkanlıkları ve geri bildirim verileri analiz edilerek pazarlama faaliyetleri, kampanyalar ve stratejiler oluşturulabilmektedir. Günümüz ekonomisinde veri petrol ile eş değeri tutulup, gelecekteki katkısının petrolden çok daha fazla olacağı ve onu geride bırakacağı düşünülmektedir (Namlı ve Murat 2019).

Teknolojinin gelişmesi ile elde edilen ve sistemlerde saklanan veriler de git gide çeşitlenmiş ve boyut olarak da büyümüştür. Saklanan verilerin kullanılabilmesi için anlamlı bilgiler haline dönüştürülmesi ihtiyacı ortaya çıkmıştır. Verilerin anlamlı hale getirilmesi aşamasında veri madenciliği yöntemi hayatımıza girmiş ve günümüzde yoğun olarak kullanılmaktadır. Rekabet ve bilginin öneminin her geçen gün artması ile birlikte

veri madenciliği de aynı yönlü önem kazanmış ve büyümeye devam etmiştir (Bose and Mahapatra 2001).

Veri madenciliği ile büyük veri depolarındaki değerli bilgi tespit edilebilmektedir. Birbiri ile ilişki içerisinde olan veri özelliklerinin tespitinden sonra asıl amaç bu modelin gelecek tahminlemesi için üst düzeyde performans göstermesidir. Performansın ölçülebilmesi için ellerindeki veriyi tanımları ve gelecek planlarını doğru oluşturmaları gerekmektedir. Hedefe doğru ilerlerken veri madenciliğinde kullanılan teknikleri de doğru sırasıyla uygulamak da sonucun başarısında önemli bir faktördür.

Veri madenciliği çalışmalarında problemin tanımlanması en önemli aşamadır. Hatalı yada eksik bir tanım tüm çalışmayı etkileyecek, tekrar tekrar başa dönmeye neden olabilecektir. Sonrasında, veri toplanması, bu verilerin ön değerlendirmeden geçerek uygulama sonucunda hatalı bir sonuca neden olacak veriler varsa temizlenir. Farklı kaynaklardan alınan veriler varsa birbirleri ile uyumlu çalışacağı şekilde birleştirilir. Gerekli görüldüğü durumlarda veriler dönüştürülür.

Problem tanımının ve verilerin hazırlanması aşamaları sonrasında model seçimi yapılır. Yukarıda bahsedilen veri düzenleme çalışmaları en uygun modeli bulunana kadar tekrar tekrar yapılabilmelidir. En iyi performansa sahip model seçilmelidir, en iyi performans olarak algoritmanın doğruluk oranının ne kadar yüksek olduğuna bakılarak genelde karar verilir.

Veri madenciliğinde kullanılan teknikler arasında sınıflandırma, kümeleme ve birliktelik kuralları yer almaktadır. Sınıflandırma algoritmalarında veri kümesinin bir kısmı eğitime kullanılır ve bu eğitim sonucuna göre de tahminleme yapılabilir. Kümeleme algoritmalarında ise veriler ortak özelliklerine göre gruplara ayrılır. Birliktelik kurallarında ise karakteristik olarak birlikte görülen veri özellikleri belirlenir (Katılmış 2018).

Tezin birinci bölümünde; literatürdeki kredi, kredi ödemeleri ve veri madenciliği ile ilgili çalışmalara yer verilmiştir. İkinci bölümde, veri madenciliği kavramları, özellikleri, süreçleri ve teknikleri, üçüncü bölümde ise kredi kavramı, unsurları ve bireysel kredi

kavramı ile çeşitleri açıklanmış, iki farklı yıla ait aynı ay için Türkiye’deki kredi kullanım oranlarındaki değişim değerlendirilmiştir. Dördüncü bölümde ise özel bir bankadan elde edilen kredi kullanan müşterilerinden kredisini erken kapatan müşteri verileri üzerine veri madenciliği teknikleri uygulanarak analiz çalışması yapılmış, sınıflandırma algoritmalarından uygun 3 model çalıştırılmış, model çıktıları değerlendirilerek en başarılı model belirlenmiştir.



2. LİTERATÜR TARAMASI

Literatürde, bankacılık sektöründeki kredi ödemeleri ile ilgili çalışmaların çok büyük bir bölümü risk belirleme yada kredi kullandırım amacıyla yapılan çalışmalar. Bu bölümde çalışmam sırasında incelediğim, faydalandığım ve bankacılık, kredi ve veri madenciliği başlıklarına ait literatürdeki bazı çalışmalara dair bilgiler yer almaktadır.

Çalış, çalışmasında bir banka şubesine ait bireysel kredi kullanan müşterilerin verilerini kullanarak, müşterilerin kredi geri ödeme performansına yönelik veri madenciliği kümeleme analizi tekniklerinden k-means algoritması ile mevcut müşterileri değerlendirmiş ve karar ağacı algoritmaları ile müşterilerin ödeme durumlarına göre sınıflandırılması sağlanarak, gelecekteki potansiyel müşteriler için çıkarım yapılmıştır. Kümeleme analizinde 3 küme oluşmuş ve örneğin, araç değişkeninin 1. ve 3. kümeler için önemli olup , 2. küme için ayırt edici bir niteliğe sahip olmadığı gibi her bir değişkenin tek tek kümeler üzerindeki değerlendirilmiştir. İkinci aşamadaki karar ağacı algoritmaları ile ilgili çalışma bölümünde ise, C&RT, C5.0, CHAID ve QUEST algoritmaları uygulanmış ve sonuçlar değerlendirilmiş, bu çalışma için en başarılı algoritmanın %95,5 doğruluk payı ile chaid algoritması olduğu sonucuna varılmış ve tüm modeller için sonuçlar değerlendirilmiştir (Çalış 2013).

Aksoy, çalışmasında bireysel kredilerin kullanımındaki riski analiz ederek ileriye yönelik tahminleme çalışması yapmayı amaçlamıştır. Bankaların 2018 yılı Ocak-Haziran dönemi bireysel kredi verilerinden yola çıkarak Haziran – Aralık aylarındaki bireysel kredi kullanımı ve riski verilerini gri tahmin $G(1,1)$ modeli kullanılarak tahminleme yapılmış ve analiz sonucunda 2018 yıl sonu itibarıyla bireysel kredi toplam nakdi tutarı kredi tipi bazında dağılımı ile birlikte tahminlenmiş, hangi kredi tipinin hangi tutarlarda daha riskli olduğu gibi sonuçlar değerlendirilmiştir (Aksoy 2019).

Budak ve Erpolat, çalışmalarında lojistik regresyon analizi ve yapay sinir ağları yöntemlerini kullanarak kredi almak isteyen kişilerin banka risklerinin hesaplanabilmesi için düzenli ödeme alışkanlıklarını değerlendirilmişlerdir. Çalışma özel bir bankanın bir

yıl öncesinde kredi kullanmış 1.639 müşterisine ait veriler kullanılarak yapılmıştır. Çalışma sonucunda yapay sinir ağı modelinin doğru sınıflandırma oranının % 70,3 lojistik regresyon analizinin doğru sınıflandırma oranının ise % 65,1 olduğu ve bunlara göre de yapay sinir ağlarının ödeme alışkanlıklarının düzeninin tahmininde lojistik regresyon analizine göre daha başarılı olduğu sonucuna varılmıştır (Budak ve Erpolat 2012).

Aşan, çalışmasında bir bankanın kredi kartı müşterileri içerisinde seçtiği bir örneklem kümesinde, kümeleme analizinin hiyerarsik olan yöntemlerinden ortalamalar bağlantı tekniği kullanılarak kredi kartı kullanan müşterilerin sosyo-ekonomik özellikleri ile ilgili 10 değişkene ait ölçümler yapılmış ve bu uygulama sonucunda bir grupta 31-40 yaş aralığında, medeni durumu evli, master/ doktora tamamlamış olan kadınlar yer alırken, diğer bir grupta erkek 45-50 yaşlarında, evli ve master/doktora diplomalı, üçüncü grupta ise, 51+ yaşlarında evli erkek ve üniversite mezunu olan müşteriler yer alacak şekilde 3 gruba ayrılmıştır. Bu grupların özelliklerine göre de ilerideki kredi kartı pazarlama stratejisi oluşturulması sağlanabileceği iletilmiştir (Aşan 2007).

Hsieh, çalışmasında bir bankadaki kredi kartı müşterilerinin bilgileri ile sinir ağları ve birliktelik kurallarını kullanarak davranışsal puanlama modeli oluşturmuş, ayrıca müşterileri üç büyük gruba ayırıp ilişkilendirme kuralları ile belirlenen müşterinin özelliklerine göre profil oluşturulmuştur. Çalışma sonucunda, müşterileri davranışsal puanlama modeli ile tanımlamanın müşterinin faydalı özellikleri olduğunu ve pazarlama stratejisi geliştirmeyi kolaylaştırdığı sonucuna varmıştır (Hsieh 2004).

Katılmış, çalışmasında özel bir bankanın 33.503 müşterinin 100.000 satır verisi üzerinden veri madenciliğinin sınıflandırma tekniğinin Navie bayes, Logistic, IBk ve J48 algoritmalarını kullanarak dijital kanal kullanım durumlarına göre değerlendirilmiştir. Çalışmanın sonucunda en yüksek doğruluk oranına sahip algoritma çıkan J48 algoritmasına göre; mevduat ürünü olup maaş almayan 25-30, 31-35, 41-46 yaş aralıklarındaki müşterilerin dijital kanalları daha yoğun kullandıkları sonucuna varmıştır. Oluşturulan profillere göre dijital kanal kullanmayan ama kullanabilecek profile olan müşterilerin dijital kanal kullanımına yönlendirilmesi amaçlanmıştır (Katılmış 2018).

Ranjan, Singh ve Bhatnagar, çalışmalarında bir ülkenin sigorta sektörü özelinde müşteri ilişkileri yönetiminin nasıl uygulanabileceğini veri madenciliği yöntemlerini kullanarak incelemişlerdir. Müşteri bilgilerinin doğru analiz edilmesi ile müşterilerin tanınabileceğini ve iyi tanınan müşteriye doğru sunumlar ile memnuniyet sağlanabileceği ve şirket karlılığının artırılabilmesi görülmüştür. Çalışma sonucunda veri madenciliği yöntemlerini kullanarak müşteri bilgileri üzerinden yapılan analiz ile karlılık nasıl artırılabilmesi belirtilmiştir (Ranjan ve diğ. 2011).

Tosun, çalışmasında özel bir bankanın kredi kartı müşterilerinin bazı bilgileri alınarak müşteri kaybetme sebepleri, hangi tür müşterilerin daha fazla gittikleri veri madenciliği teknikleri ile incelemiştir. Çalışmada karar ağacı algoritmaları kullanılmış, farklı eşik değerlere göre kurala uyma durumları değerlendirilmiştir. Karar ağaçlarında eşik değeri kullanılırsa ve kullanılan eşik değeri arttıkça, niteliklerin azalacağı, verinin alındığı aya yakın aylarda harcaması olan müşterilerin son 3 ayda hiç harcaması olmayanlara oranla kaybedilme oranının daha düşük olduğu, 11. ve 12. aylarda harcaması olan müşterilerden ilk hesap açtırma tarihi 1998 öncesinde ise kaybedilme olasılığını daha yüksek olduğu gibi sonuçlara ulaşılmıştır (Tosun 2006).

Kalaycı, web siteleri için kimlik hırsızlığı tespiti üzerinde bir çalışma yapmıştır. Adaboost, Decision Tree, Multilayer Perceptron, NaiveBayes ve Random Forest yöntemlerini kullanarak 9 farklı değişken ile toplam 1353 satırlık veri kümesi hazırlayarak web sitelerinin kimlik hırsızlığı yapıp yapmadıklarını tahmin etmek üzerine bir model hazırlamıştır. Sınıflandırma algoritmalarında yaygın olarak kullanılan eğitim ve test aşamalarını kendi tezinde de kullanan Kalaycı, yapmış olduğu çalışmada topluluk öğrenme yaklaşımını içeren bir karar ağacı modeli olan Random Forest yönteminin, diğer yöntemlerle karşılaştırıldığında en yüksek doğruluk oranını verdiğini görmüştür (Kalaycı 2018).

Kavcıoğlu, kurumsal kredilere ait veri kümesi ile kredi skorlamada makine öğrenmesi ve lojistik regresyon modellerinin performanslarını kıyaslayarak değerlendirmiştir. Yapay sinir ağları lojistik regresyona göre daha yüksek başarı sergilediği sonucuna varılmış olsa

da lojistik regresyonun daha tutarlı sonuçlar verdiđi görölmüştür. Diğer yandan yapay sinir ađlarının iterasyon aşamalarında optimizasyon ile daha anlamlı sonuçlar üreteceđi düşüncesine varılmıştır (Kavcıođlu 2019).



3. VERİ MADENCİLİĞİ

3.1 VERİ MADENCİLİĞİ KAVRAMI

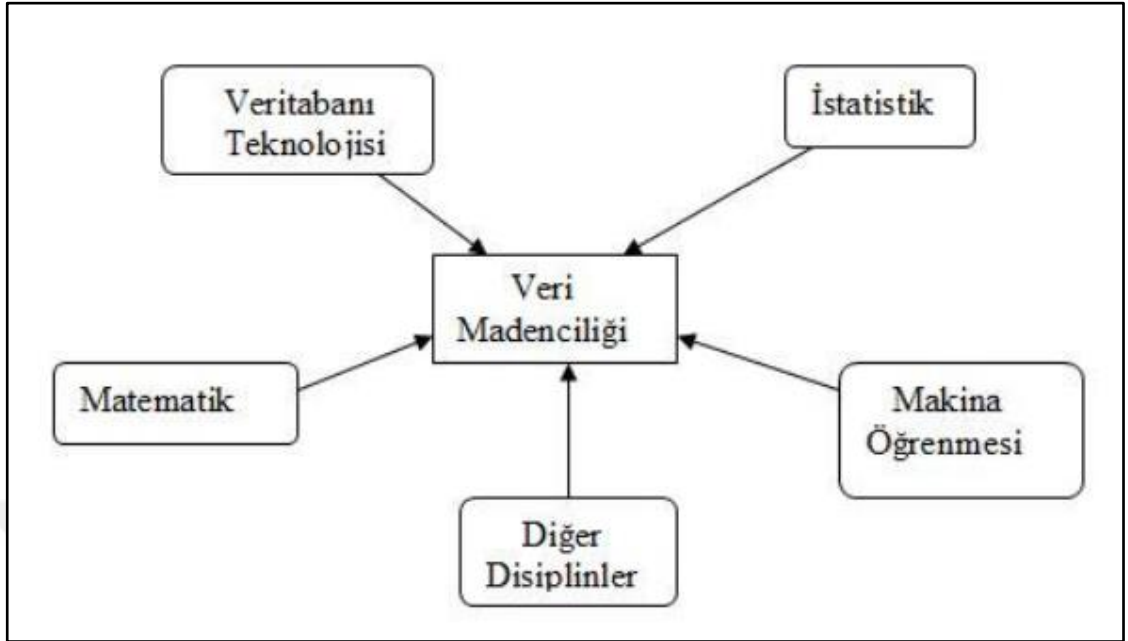
Veritabanı teknolojisinin ilerlemesi ile birlikte veri depolarında tutulan veri miktarı da artmıştır. Elde edilen verinin boyutunun ve karmaşıklığının artışı ile aralarındaki gizli örüntülerin bulunabilmesi için kullanılan geleneksel veri analiz yöntemleri yetersiz kalmıştır ve yapay sinir ağları, genetik algoritmalar, sınıflandırma algoritmaları vb. veri madenciliği yöntemlerin doğuşuna zemin oluşturmuştur.

Veri madenciliği ile geleneksel istatistiksel yöntemler arasındaki en büyük farklılıklardan biri veri kümesinin boyutudur. İstatistiksel yöntemlerde binler seviyesindeki veri ile çalışılabilirken veri madenciliği teknikleri ile milyarlar seviyesindeki veriler ile çalışabilmektedir (Bilekdemir 2010).

Veri madenciliği; bir işletmede tüm verilerin toplandığı yazılım ve donanımı ifade eden veri ambarlarında yer alan verilerdeki gizli eğilimleri, bu eğilimlerin birbirleriyle ilişkilerini, ortaya çıkan ilişkilerin nedenlerini ve verilerin nasıl bir seyir gösterdiğini ortaya çıkaran yöntemler topluluğu, modellemeler ve teknikler olarak görülebilir (Özmen 2009).

Şekil 3.1’de İstatistik, matematik, veri tabanı sistemleri, makine öğrenmesi, görselleştirme vb. veri madenciliğinin etkileşim içinde bulunduğu disiplinler yer almaktadır.

Şekil 3.1: Veri madenciliğinin etkilesim içerisinde bulunduğu disiplinler



3.2 VERİ MADENCİLİĞİNİN TARİHSEL GELİŞİMİ

Bilginin ve değerinin artması ile birlikte ona ulaşmak ve elde edilen bilginin kullanılabilir hale getirilerek analiz edilmesi konusunda gelişme ihtiyaç olmuştur.

Veri madenciliğinin tarihçesine baktığımızda temelini 1950 yıllarında matematikçilerin mantık ve bilgisayar bilimleri konularında çalışmaları ile yapay zeka, makine öğrenmesi alanlarındaki gelişmeler, 1960'lı yıllarda ise istatistikçilerin regresyon analizi, sinir ağları vb. algoritma çalışmaları oluşturmaktadır. Sonraki 30 yılda ise verilerin sınıflandırılması ve bu sınıflar arasında ilişkisel bağların kullanılması ile genetik algoritmalar (genetic algorithms) , “EM algorithms”, kümeleme yöntemleri (K-Means clustering methods) , karar ağaçları (decision tree) gibi algoritmalar oluşturulmuştur. Veri madenciliği terimi 1990'lu yıllarda kullanılmaya başlanmıştır (Kaya ve Köymen 2008).

Tablo 3.1'de veri madenciliğinin tarihsel sürecinde gelişim adımlarının hangi karakteristikler ile hangi teknolojiler kullanılarak hangi problemin cevaplandığı bilgileri yer almaktadır (Yapıcı ve diğ. 2010).

Tablo 3.1: Veri madenciliğinin tarihsel gelişimi

Gelişim Adımları	Cevaplanan Karar Problemi	Karakteristikler	Kullanılan Teknolojiler	Ürün Sağlayıcılar
Veri Toplama (1960'lar)	"Benim toplam karım geçen 5 yılda ne kadardı?"	Geriye dönük, statik veri dağıtımı	Bilgisayarlar, Teypler, Diskler	IBM, CDC
Veri Erişimi (1980'ler)	"Türkiye'de geçen Ekim ayında birim satışları ne kadardı?"	Kayıt düzeyinde geriye dönük, dinamik veri dağıtımı	İlişkisel Veritabanları, SQL, ODBC	Oracle, Sybase, Informix, IBM, Microsoft
Veri Ambarlama ve Karar Destek Sistemleri (1990'lar)	"İstanbul'da geçen Kasım ayında birim satışlar ne kadardı? İzmir ile karşılaştırıldığında fark nedir?"	Çoklu düzeylerde, geriye dönük dinamik veri dağıtımı	OLAP, Çok Boyutlu Veritabanı Sistemleri, Veri Ambarları	Pilot, Comshare, Arbor, Cognos, Microstrategy
Veri Madenciliği (Bugün)	"Gelecek ay Ankara'daki birim satışlar ne olabilir, niçin?"	Geleceğe dönük, proaktif informasyon dağıtımı	İleri Düzeyde Algoritmalar, Çok İşlemcili Bilgisayarlar, Büyük Veri Ambarları	Pilot, Lockheed, IBM, SGI, SPSS, SAS, Microsoft vb.

Kaynak: Yapıcı, A P; Özel, A; Ayça, C (2010). Oracle data miner ile kredi ödemeleri üzerine bir veri madenciliği uygulaması, Bitirme Ödevi.

3.3 VERİ MADENCİLİĞİNİN KULLANIM ALANLARI VE FAYDALARI

Veri madenciliğinin sektör bazında kullanım alanları ve faydaları aşağıdaki şekilde sıralanabilir:

Bankacılık:

- Müşteri profilleri oluşturularak pazarlama aktiviteleri profillere göre farklılaştırılabilir ve doğru müşteriye doğru iletişim ile doğru ürünün sunulması sağlanabilir. Bu sayede hem daha az operasyon ve iletişim maliyeti oluşur, hem de müşterilere gereksiz iletişim yapılmadığı için memnuniyet artar.

- b) Kredi, kredi kartı vb. müşterinin borçlandığı ürünlerde ödeme problemi yaşayacak müşteriler tespit edilebilir ve yeni müşterilerde daha başvuru aşamasında risk seviyesine göre hiç onay verilmemesi sağlanabilir.
- c) Dolandırıcılık işlemlerinin tespitinde erken uyarı sistemi oluşturulabilir.
- d) Mevcut müşteri davranışları incelenerek çapraz satış yapılması ve müşteri sadakati sağlanabilir.
- e) Müşteri kaybı analizi yapılarak müşteriler kaybetmemek için önlemler alınabilir.

Sigortacılık:

- a) Müşterilere doğru ürünü sunabilmek için müşteri profillerinin oluşturulması sağlanabilir.
- b) Mevcut poliçelerin yenilenme durumu tahminlerinde kullanılabilir.
- c) Sigorta dolandırıcılığının tespitinde kullanılabilir.
- d) Mevcut müşterilere çapraz ürün satışı yapılmasında fayda sağlayabilir.

Elektronik Ticaret:

- a) Web sayfalarındaki ziyaretlerin çözümlenmesinde kullanılabilir.
- b) Ekran tasarımlar ve davranışlarının müşterilerin beklentileri, kullanım alışkanlıkları analiz edilerek oluşturulması sağlanabilir.
- c) Müşterilerin birlikte aldıkları ürünler üzerinden pazar sepeti analizi yapılarak satın alma alışkanlıkları tespit edilip uygun ürün yelpazesi oluşturulması sağlanabilir.

Telekomünikasyon:

- a) Müşterilerin profilleri oluşturularak uygun hizmetlerin hazırlanmasında kullanılabilir.
- b) Mevcut müşterilere çapraz satış yapılmasında fayda sağlayabilir.
- c) Sim kart kopyalama, kaçak hat kullanımı gibi dolandırıcılık işlemlerinin tespit edilmesi sağlayabilir.
- d) Alt yapıda sorun yaşanan bölgelerin tespitinde kullanılabilir.

Tıp ve Sağlık:

- a) Önceki hastaların laboratuvar ve görüntüleme test sonuçları kullanılarak hastalıkların ön tanı ve erken teşhisinde fayda sağlanabilir.

- b) Hastalıklara ait yol haritası oluşturulmasında kullanılabilir.
- c) DNA içindeki gen sıralamalarının belirlenmesinde kullanılabilir.

Satış ve Pazarlama:

- a) Müşterilerin demografik bilgileri üzerinden profilleri oluşturularak pazarlama aktivitelerindeki içerik ve yöntem farklılaştırılabilir.
- b) Pazar analizi ile sektörün takip edilmesi, geride kalınmaması sağlanabilir.
- c) Stok tedarik yönetiminde mevcut veriler üzerinden geleceğin tahmin edilmesi kullanılabilir.

Yukarıda yer alan ana sektörler dışında; eğitim, turizm, borsa, üretim ve planlama, ulaşım, taşımacılık vb. birçok sektörde veri madenciliği analizleri kullanılabilmektedir.

3.4 VERİ MADENCİLİĞİNDE KARŞILAŞILAN PROBLEMLER

Veri madenciliğinde kullanılan ham datanın kalitesi ve hedefe uygun oluşu çok önemlidir. Veri madenciliği süreçlerinde karşılaşılan başlıca problemler aşağıda yer almaktadır (Akbulut 2006).

- a) Sınırlı bilgi: Veritabanları tasarlanırken veri madenciliği dikkate alınarak tasarlanmamaktadır. Bu nedenle de kullanılan veritabanlarında veri madenciliğinde öğrenmeyi kolaylaştıracak bazı özellikler yer almayabilir, sınırlı bilgi ile uygulamanın gerçekleştirilmesi gerekebilmektedir.
- b) Gürültü ve kayıp değerler: Veri girişi yada toplanması sırasında oluşan sistem dışı sorunlara gürültü adı verilmektedir. Gürültülü veri de amaca ulaşmayı engelleyebilmektedir.
- c) Belirsizlik: Ham verideki gürültünün derecesi ve hataların oranı ile ilgilidir. İleriye yönelik tahminleme yapılan veri madenciliği uygulamalarında önemlidir.
- d) Ebat, güncellemeler ve konu dışı sahalar: Veri tabanlarında zaman içerisinde yeni verilerin eklenmesi, silinmesi ile bilgiler değişebilmektedir. Değişen bilgilerin daha

önce uygulanmış modellerdeki öğrenme sistemine olumsuz etkisi olabileceği için takibi önemlidir.

3.5 VERİ MADENCİLİĞİ SÜREÇLERİ

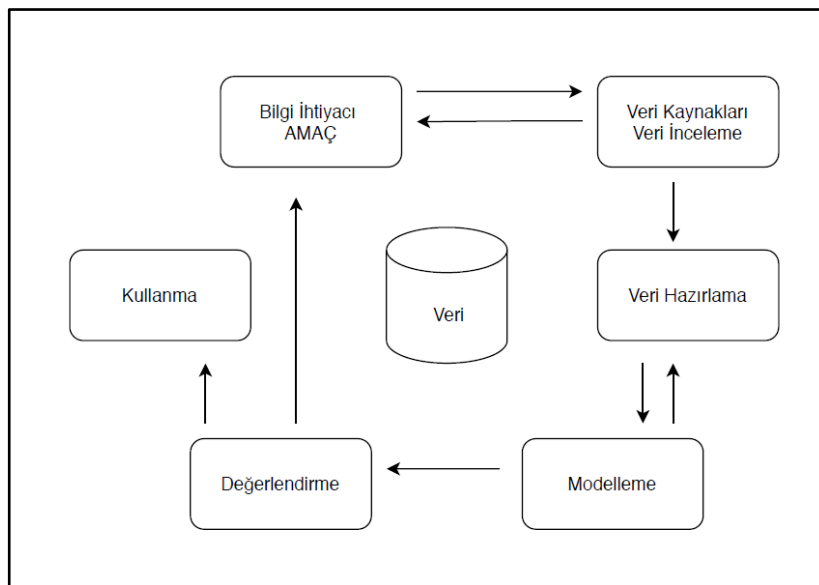
Veri madenciliğinde kullanılan algoritmanın başarıya ulaşması için mutlaka çalışma yapılan konunun ve kullanılan verinin özelliklerine hakim olunması gereklidir (H. Turgut 2012).

Veri madenciliği süreçleri beş ana başlıkta gösterilmiştir (Restivo 1999):

- a) Problemin Tanımlanması,
- b) Verilerin Hazırlanması,
- c) Modelin Kurulması,
- d) Modelin Değerlendirilmesi,
- e) Modelin Kullanılması ve İzlenmesi.

Şekil 3.2’de veri madenciliği süreçleri, bu süreçlerin birbirleri ile etkileşimleri yer almaktadır.

Şekil 3.2: Veri madenciliği süreçleri



3.5.1 Problemin Tanımlanması

Problemin tanımlanması veri madenciliğinin ilk ve en önemli aşamalarından biridir.

Bu aşamada çalışmanın amacı net bir şekilde belirtilmelidir. Ayrıca çalışma sonunda elde edilecek verilerin başarı düzeyinin neresi olduğu da belirlenmelidir. Doğru yapılan tahminlerde kazanılacak faydalar ve hatalı yapılan tahminlerde oluşacak maliyetlere yer verilmelidir. Fayda – maliyet analizi ile çalışmanın sağlayacağı değer analiz edilmelidir (Dumouchel 1999).

3.5.2 Verilerin Hazırlanması

Verilerin hazırlanması süreci, problem tanımına uygun olan verilerin kaynaklarının belirlendiği veri toplama adımı, birden fazla veri kümesi kullanılıyorsa birbiri ile uyumlulaştırıldığı ve birleştirildiği veri birleştirme adımı, kirli verinin kullanılacak çalışmaya uygun olacak şekilde temizlenmesi adımlarını kapsamaktadır.

3.5.3 Modelin Kurulması ve Değerlendirilmesi

Bir probleme ait birden fazla çözüm yolu olabilmektedir ve başarılı bir sonuç için mümkün olan en uygun modelin seçilmesi hedeflenmektedir. Doğru model seçimi ise mümkün olduğunca fazla modelin denenmesi ve sonuçlarının analiz edilmesi ile mümkündür.

Model kurulması denetimli (supervised) ve denetimsiz (unsupervised) öğrenmenin kullanılmasına göre farklılık göstermektedir.

Denetimsiz öğrenme veriyi anlamaya yönelik olarak kullanılmakta ve sonraki uygulanacak yöntemler için fikir vermeyi amaçlamaktadır.

Denetimli öğrenme ise veriden bilgi ve sonuç çıkarmaya yönelik kullanılmaktadır. Denetimli öğrenme 2 aşamalıdır ve veriler iki gruba ayrılır. Birinci aşamada verilerin büyük olan kısmı modelin öğrenmesi için kullanılır. İkinci aşamada ise diğer veri grubu ile modelin geçerliliği test edilir.

Modelin doğruluğunun ölçümlemesinde en sık kullanılan yöntemlerden biri basit geçerlilik testidir. Veri seti, biri en fazla üçte biri olacak şekilde ikiye ayrılır. Büyük olan data ile modelin öğrenimi gerçekleştirildikten sonra kalan data ile model test edilir. Çıkan sonuçlarda doğru tahmin edilen olay sayısı tüm olay sayısına oranı doğruluk oranı, hatalı tahmin edilen olay sayısı tüm olay sayısına oranı ise hata oranıdır. Doğruluk oranı en yüksek olan model bizim için en iyi modeldir.

3.5.4 Modelin Kullanılması ve İzlenmesi

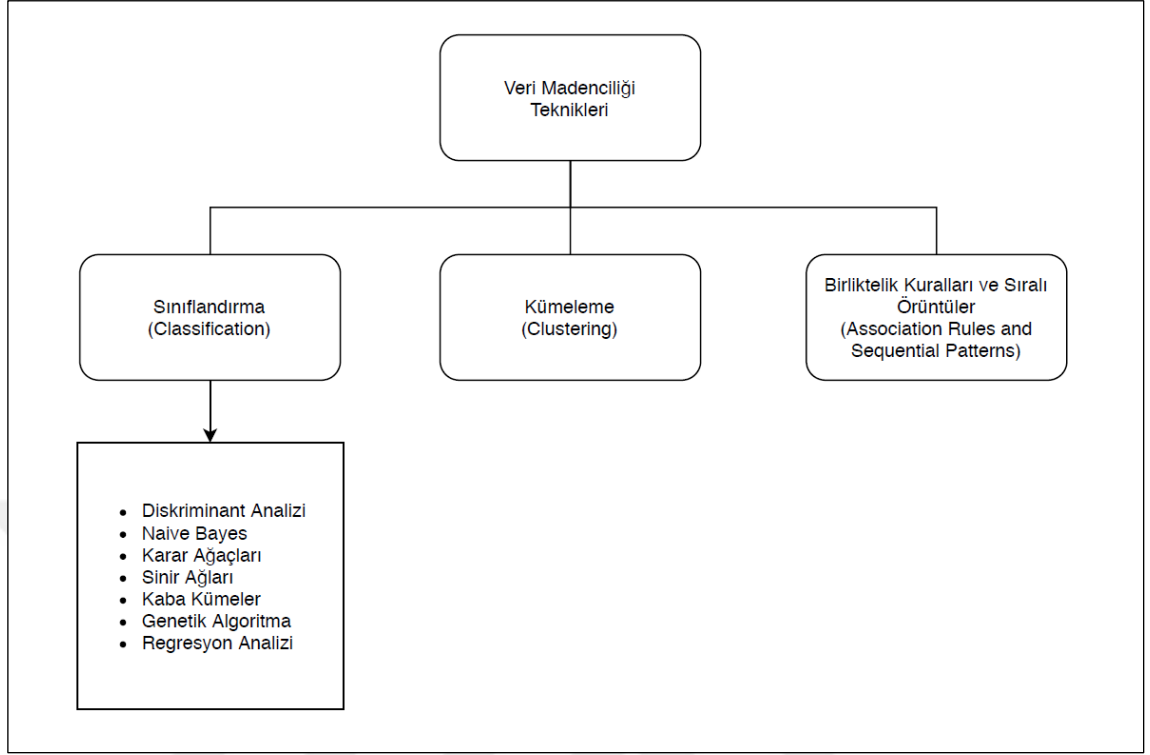
Kurulan ve başarılı olarak değerlendirilen model tek başına bir uygulama yada başka bir uygulamanın parçası olabilmektedir. Model kullanılmaya başlanıldığında izleme süreci devam etmelidir, veri eklenmesi, silinmesi, sistemin özelliklerinin değişmesi gibi durumlarda modelin geçerliliğinin devam etmesi kontrol edilmeli ve gerekli durumlarda model tekrar düzenlenmelidir.

3.6 VERİ MADENCİLİĞİ TEKNİKLERİ

Şekil 3.3’de de yer aldığı üzere, veri madenciliği tekniklerini işlevlerine göre 3 temel grupta toplanabilir;

- a) Sınıflandırma (Classification),
- b) Kümeleme (Clustering),
- c) Birliktelik kuralları ve sıralı örüntüler (Association rules and sequential patterns).

Şekil 3.3: Veri madenciliği teknikleri



3.6.1 Sınıflandırma

Sınıflandırma, kullanılacak veri kümesinin önceden belirlenmiş olan çıktılarına göre gruplanmasını sağlayan bir yöntemdir. Çıktıların önceden biliniyor olması sınıflandırmada denetimli (supervised) öğrenme kullanıldığını göstermektedir (Giudici 2003).

Sınıflandırma modeli elde edilecek bir eğitim veri setinde hangi sınıfa ait olduğu bilinen örnekler yer almaktadır. Söz konusu veri seti üzerinde sınıflandırma algoritmalarından biri uygulanarak bir sınıflandırma modeli oluşturulur ve ardından test veri setindeki örnekler için sınıf tahmini yapılarak, modelin tahmin sonucunun performans ölçüm değeri elde edilir.

Sınıflandırma yöntemleri, geleceğin tahminlenmesi çalışmalarında kullanılmaktadır. Veriler temizlenip modelde kullanmaya hazır hale getirildiğinde bir kısmı eğitim için ayrılır. Algoritma çalıştırıldığında ilk aşamada eğitim verisi içinde veri niteliklerine göre

sınıflara ayrılarak grup oluşturulur. Verinin doğru, kaliteli oluşu, oluşan veri sınıfı kalitesini de belirler. İkinci aşamada ise test verileri ile algoritma sınanır. Test verilerinde sınıfı belli olmayan veriler sınıflandırılmış olur ve model sınanmış olur (Salur 2016).

Temel sınıflandırma algoritmaları aşağıda yer almaktadır:

- a) Naive Bayes,
- b) Diskriminant analizi,
- c) Karar ağaçları,
- d) Sinir ağları,
- e) Kaba kümeler,
- f) Genetik algoritma,
- g) Regresyon analizi.

Bu algoritmalara ait genel bilgiler aşağıda yer almaktadır:

- a) Naive Bayes

Naive Bayes sınıflandırma algoritması, adını İngiliz Matematikçi Thomas Bayes'ten almaktadır.

Naive Bayes, hedef ve bağımsız değişkenler arasındaki ilişkiyi analiz eden tahminci ve tanımlayıcı bir algoritmadır (Hudairy 2004).

Teoremin çalışma prensibi, veri kümesindeki her bir özellik için olayların olma olasılığını hesaplanması ve olasılığı en yüksek olana göre sınıflandırma işlemi yapması şeklindedir. Naive Bayes teoremine göre sınıflandırma işlemleri aşağıda yer alan formül ile hesaplanarak yapılmaktadır.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2.1)$$

$P(A|B)$; B olayının olması koşuluyla A olayının olması olasılığıdır.

$P(B|A)$; A olayının olması koşuluyla B olayının olması olasılığıdır.

$P(A)$ ve $P(B)$; A ve B olaylarının olması olasılığıdır (Kuzucu 2015).

b) Diskriminant Analizi

Diskriminant analizi, çalışma öncesinde belirlenmiş olan sınıflara bir dizi gözlemi atayan bir tekniktir. Model, öğrenme kümesi üzerine kurulur. Öğrenme kümesine bağlı olarak doğrusal fonksiyonların bir kümesi oluşturulur, bu küme diskriminant fonksiyonları olarak da adlandırılır. Diskriminant fonksiyonu oluşturulduktan sonra yeni gözlemler olduğunda tüm diskriminant fonksiyonları hesaplanıp en yüksek olana atanır (Huberty 1994).

c) Karar Ağaçları

Görsel yapı itibarıyla da ağaca benzeyen karar ağacı algoritmaları, daha çok sınıflama ve tahminleme amacıyla kullanılmaktadır.

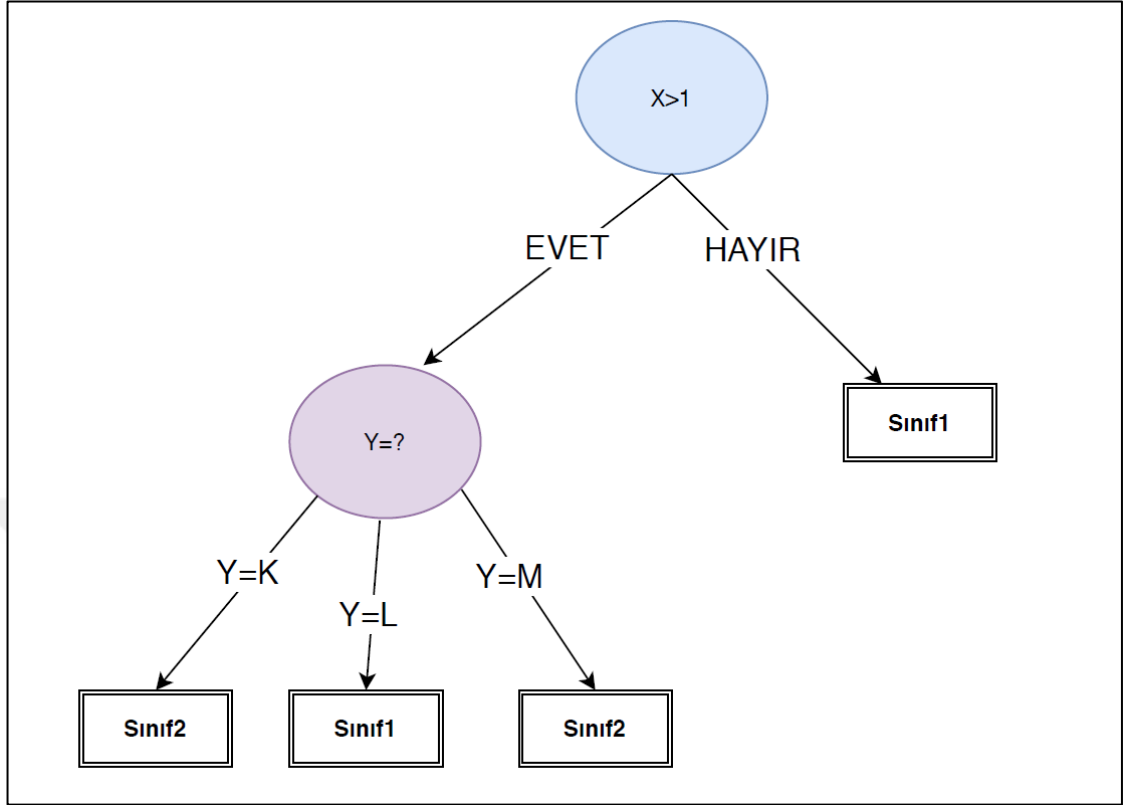
Yapısal olarak, düğüm (node) bir özellik üzerinde yapılan testi, dal ise testin çıktısını, yaprak düğüm ise sınıfları temsil etmektedir. En üstteki düğüm, kök düğüm olarak isimlendirilir. Karar ağaçları, kök düğümden yaprak düğüme doğru çalışır (Wei ve Chiu 2002) .

En fazla tercih edilen karar ağacı algoritmaları aşağıda yer almaktadır.

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

Şekil 3.4’de X ve Y isimli 2 değişken üzerinden uygulanan örnek bir karar ağacı görüntüsü yer almaktadır (Tekingöz 2016).

Şekil 3.4: X ve Y değişkenleri üzerinden uygulanan karar ağacı örneği



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

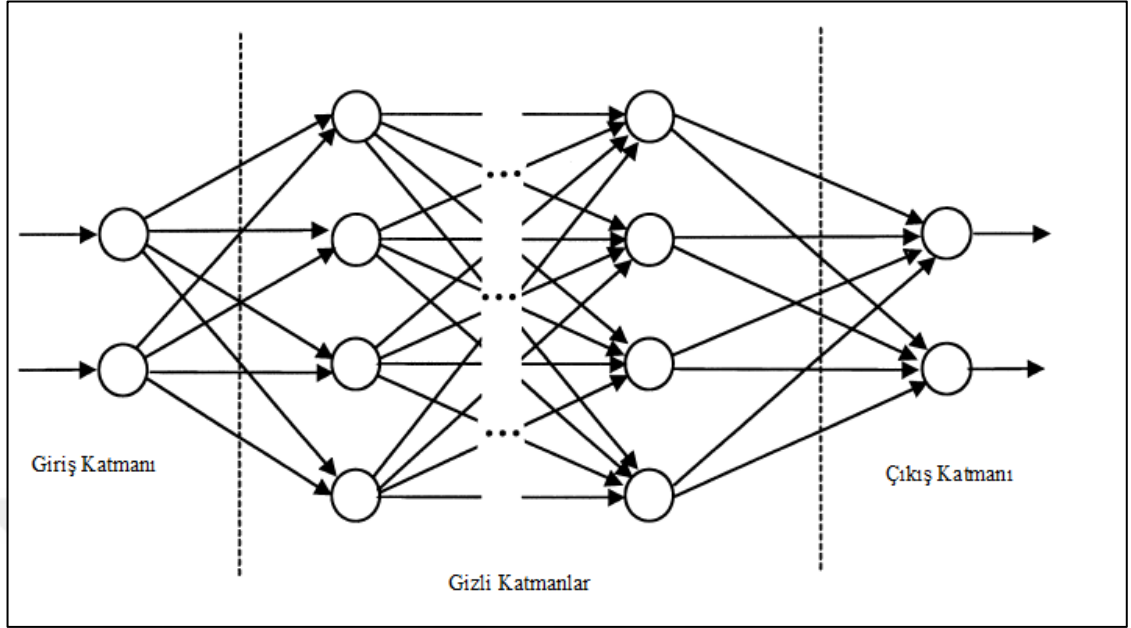
d) Yapay Sinir Ağları

Yapay sinir ağları, tanımlayıcı ve tahmin edici veri madenciliği algoritmalarındandır. Veritabanındaki ilişkileri, sınıflandırma ve tahminde kullanılması için genelleştirir.

Yapay sinir ağları, sinir sistemini model olarak yapılan çalışmalar sonucunda bulunmuştur. Öğrenme yeteneğine sahiplerdir ve sayısal veriler üzerinden çalışırlar. Karmaşık ve anlaması güç olan verilerin analizinin yapılması, anlam çıkartılmasında kullanılabilmektedir (Çetinyokuş 2008).

Yapı olarak yapay sinir ağları genellikle 3 katmandan oluşur; giriş katmanı, gizli katmanlar ve çıkış katmanıdır. En temel haliyle, her bir nöron bir önceki katmanlardaki diğer nöronlara, sinaptik ağırlıkları yoluyla bağlanmaktadır. Yapay sinir ağlarının katmanları Şekil 3.5’de gösterilmiştir (Kalogirou 2000).

Şekil 3.5: Yapay sinir ağlarının katmanları



Kaynak: Kalogirou, S A (2000). Applications Of Artificial Neural-Networks For Energy Systems

e) Kaba Kümeler

Kaba küme teorisi, 1970’li yıllarda Pawlak tarafından geliştirilmiştir. Bu teoride, yaklaştırma kümesi ve alt ve üst yaklaşımlar vardır. Alt yaklaşımlar kesin olan nesnelerin tanımlarıdır, üst yaklaşımlar ise alt yaklaşım kümesine dahil olması mümkün olabilecek nesnelerin tanımlarıdır. Üst ve alt sınırlar arasındaki nesnelere ise “Kaba Küme” denilmektedir (Pawlak 2002).

f) Genetik Algoritma

Genetik algoritmalar, John Holland’ın bilgisayarlara anlamadığı çözüm yöntemlerinin evrim süreci simüle edilerek anlaşılır hale getirileceği düşüncesinin bir sonucu olarak ortaya çıkmıştır (Çetinyokuş 2008).

Genetik algoritmalar evrim sürecini taklit etmektedir.

Genetik algoritmada öğrenmesi veri kümesi nüfus adı verilen bir çözüm kümesidir. Bir nüfus çalıştırılır ve çıktılar daha başarılı olacağı beklenen başka bir nüfus için kullanılır.

Yeni nüfus yaratma iterasyonu yani evrim süreci tamamlandığında sınıflandırma yapılmış ve bağımlılık kuralları belirlenmiş olur (Shah ve Kursak 2004).

g) Regresyon analizi

Regresyon analizi, hedef olan değişken ile bağımsız değişkenler arasındaki ilişki olup olmadığının analiz edildiği bir yöntemdir. Kolay kullanılabilmesi ve yorumlanabilmesi nedeniyle yoğun olarak kullanılmaktadır.

Doğrusal regresyonda hedef değişken sürekli bir değer alır ve bu hedef değişkenin gerçekleşme durumu tahminlenir. Diğer yandan lojistik regresyonda ise hedef değişken kesikli bir değer alır ve bu algorithmada hedef değişkenin alabileceği değerlerden birinin gerçekleşme durumu tahminlenir (Hui ve Jha 2000).

3.6.2 Kümeleme

Kümeleme algoritmalarında, veri kümesindeki birbirlerine benzer ama farklı özelliklerdeki kümelerin bulunup farklı gruplara ayrılması amaçlanmaktadır.

Kümelemede algoritmaları arasında k-ortalamlar algoritması ve Kohonen şebekesi algoritmaları daha yoğun kullanılmaktadır.

Seçilen kümeleme algoritmasından bağımsız olarak, işleyiş aynıdır. Veriler var olan kümeler ile karşılaştırılarak ona en uygun kümeye atanır ve kümeyi tanımlayan değer değişir (Hui ve Jha 2000).

Kümeleme algoritmaları arasında en fazla “k ortalamlar algoritması” kullanılmaktadır.

K ortalamlar algoritması: K ortalamlar algoritması (k means, simple k means), büyük miktardaki verilerin kümelенmesinde başarılı sonuçlar oluşturmaktadır. Yeni bir olay oluştuğunda, k ortalamlar algoritması tüm veriyi inceleyip bu yeni olaya benzer olaylar için bir alt küme oluşturup bu kümeyi sonucu tahminlemeye kullanır (Anand 2003).

K ortalamalar algoritmasında; ilk adım olarak veriler başlangıç kümelerine ayrılır. Örnekler, kümeler ile karşılaştırılıp kendilerine en yakın kümeye atanırlar ve bu işlem sonrasında küme merkezi tekrar hesaplanır. Bu işlem tüm verilerin atamaları bitene kadar devam eder.

3.6.3 Birliktelik Kuralları ve Ardışık Örüntüler

Birliktelik analizi, en basit haliyle 2 farklı olayın eş zamanlı birlikte gerçekleşme durumunu arayan bir tekniktir.

Birliktelik analizi, denetimsiz (unsupervised) veri madenciliği tekniğidir. Çoğunlukla perakende sektöründe satın alma eğilimleri üzerine kullanıldığı için “pazar sepeti analizi” olarak da isimlendirilir (Bland 2002).

Birliktelik kuralı ile ilgili örnekler aşağıda yer almaktadır (Akpınar 2000):

“Müşteriler bira satın aldıklarında yüzde 75 olasılıkla çocuk bezi de satın alırlar.”

“Düşük yağlı peynir ve yağsız süt alan müşteriler yüzde 85 olasılıkla diyet süt alırlar.”

Ardışık analizde ise, yine 2 farklı olayın birbiriyle bağlantısı tahminlenmeye çalışılır fakat birliktelik analizinden farklı olarak olayların eş zamanlı değil birbirini takip eden dönemlerde gerçekleşmesi durumundaki ilişki tanımı için kullanılır.

Ardışık analiz ile ilgili örnekler aşağıda yer almaktadır (Akpınar 2000):

“Çadır alan müşterilerin yüzde 10’u bir ay içerisinde sırt çantası almaktadır.”

“A hissesi yüzde 15 artarsa üç gün içinde B hissesi yüzde 60 olasılıkla artacaktır.”

4. BİREYSEL KREDİ KAVRAMI

4.1 KREDİ KAVRAMI

Kredi kelimesinin sözlük anlamı güven, saygınlık ve itibar olup, Latince “inanma” anlamındaki “credere” kelimesinden gelmiştir (Usta 1995).

5411 Bankacılık Kanunu’na göre kredi;

“Madde 48.-Bankalarca verilen nakdi krediler ile teminat mektupları, kontrgarantiler, kefaletler, aval, ciro, kabul gibi gayrinakdî krediler ve bu niteliği haiz taahhütler, satın alınan tahvil ve benzeri sermaye piyasası araçları, tevdiatta bulunmak suretiyle ya da herhangi bir şekil ve surette verilen ödünçler, varlıkların vadeli satışından doğan alacaklar, vadesi geçmiş nakdi krediler, tahakkuk etmekle birlikte tahsil edilmemiş faizler, gayrinakdî kredilerin nakde tahvil olan bedelleri, ters repo işlemlerinden alacaklar, vadeli işlem ve opsiyon sözleşmeleri ile benzeri diğer sözleşmeler nedeniyle üstlenilen riskler, ortaklık payları ve kurulca kredi olarak kabul edilen işlemler izlendikleri hesaba bakılmaksızın bu kanun uygulamasında kredi sayılır. Birinci fıkrada belirtilenlere ilave olarak, kalkınma ve yatırım bankalarının finansal kiralama yöntemiyle sağladığı finansmanlar ile katılım bankalarının taşınır ve taşınmaz mal ve hizmet bedellerinin ödenmesi suretiyle veya kar ve zarar ortaklığı yatırımları, taşınmaz, ekipman veya emtia temini veya finansal kiralama, mal karşılığı vesaikin finansmanı, ortak yatırımlar veya benzer yöntemlerle sağladıkları finansmanlar da bu kanun uygulamasında kredi sayılır.” (TBB, 5411 Sayılı Bankacılık Kanunu 2011)

4.2 KREDİ UNSURLARI

Literatürde kredi için dört ana unsur yer almaktadır. Detayları aşağıda yer almaktadır.

Zaman Unsuru; kredinin krediyi veren kuruma başta belirlenmiş olan vade sonunda ödenmesi gerekliliğidir. Geri ödeme süresinin daha uzun olduğu krediler geleceğin tahmin edilememesinden dolayı daha yüksek riskli, daha kısa olduğu krediler ise daha düşük risklidir (Söğütçü 2011).

Güven Unsuru; verilen kredinin geri ödenme koşulu nedeniyle güven önemli bir unsurdur. Banka tarafından yapılan değerlendirmede krediyi kullanan kişinin güvenilir olması önemlidir (Aras 1996).

Risk Unsuru'nu verilen kredinin anlaşılan vadede geri ödenmeme ihtimali ortaya çıkarır. Bu unsuru minimize etmek için bankalar kredinin büyüklüğüne yada tüketicinin profiline göre ek teminatlar talep edebilmektedirler (K. Turgut 2019).

Gelir Unsuru; bankalar işleyişlerini sürdürebilmek, mevduat topladıkları müşterilerine ve banka ortaklarına gelirlerinin önemli bir bölümünü kredilerin faiz ve komisyonlarından karşıladıkları için gelir de önemli bir unsurdur (Aras 1996).

4.3 BİREYSEL KREDİ

Bireysel kredi; bankaların kanıtlanabilir geliri olan gerçek kişilere bir mal veya hizmet satın alınması amacıyla verdikleri kredilerdir. Kredi verilmeden önce hangi faiz ve nasıl bir plan ile ödeneceği belirlenir (Şakar 2014).

Bireysel krediler, Türkiye'de Avrupa ve Amerika'dan yaklaşık 30 yıl sonrasında 1980'lerin sonlarında kullanılmaya başlanılmış, 90'larda yaygınlaşmıştır (Güney 2017).

Kredi kullanım oranları gün geçtikçe artış göstermektedir. Türkiye'deki banka ve banka haricindeki finansal kurumların 2020 Ocak ayında kullandıkları bireysel krediler bir önceki yılın aynı ayındaki kullandırmalara göre yüzde 16,8 artmış ve 632 Milyar TL'ye ulaşmıştır.

2020 verilerine göre kullandırılan bireysel kredilerin çeşitlerine göre kullanım oranları yüzdesel olarak Tablo 4.1'de belirtilmiştir (TBB, Risk Merkezi, Aylık Bülten 2020).

Tablo 4.1: Bireysel kredi kullandırım oranları (2020, %)

Kredi Türü	Kredi Kullanım Oranı
İhtiyaç kredisi	39
Konut kredisi	36
Taşıt kredisi	6
Kredi kartı	20

4.4 BİREYSEL KREDİ ÇEŞİTLERİ

Krediyi kullandıracak kuruluşun türü, ekonomik şartlar, tüketicinin ihtiyacı vb. nedenlere göre farklı çeşitlerde bireysel krediler bulunmaktadır. Kredinin kullanım nedenlerine göre bireysel krediler, konut, araç, eğitim, tatil, kariyer, onarım, inşaat tamamlama, kredili mevduat hesapları ve kredi kartları gibi çeşitlere ayrılmaktadır. Bankalar tarafından kullanılan krediler genel olarak 4 kategoriye ayrılır:

İhtiyaç Kredisi: Bankaların kullandırım öncesinde vade ve faizini belirledikleri, gerçek kişilere kullandırdıkları bir kredi çeşididir (Çelik 2014). İhtiyaç kredisi kullandırılırken herhangi bir mal yada hizmet satın alınması şartı yoktur (Vurucu ve Arı 2017).

Konut Kredisi : Kullanılabilir durumdaki yeni ve belirlibir yaşa kadar olan konutlar için kullanılabilen bir kredidir. Konutun ekspertiz ile değeri belirlendikten sonra banka bu tutarın tamamı olmamak kaydıyla belli bir oranında konut kredisi kullandırır (Güney 2017).

Taşıt Kredisi: Tüketicinin yeni yada ikinci el bir araç almak için kullandığı kredi türüdür. Krediyi kullanan müşteriye kredi nakit olarak verilmez, alıcının hiç elinden geçmeden satışa istinaden satıcıya gönderilir (Vurucu ve Arı 2017).

Kredi Kartı: Tüketicinin belirlenen kart limiti dahilinde nakit çekim yapabildiği veya alışveriş esnasında kullanabildiği ödeme aracıdır (Çırpan 2000). Kredi kartı diğer kredilerden farklı olarak harcama sonrası ödeme yapıldıkça yeniden başvurulmadan kullanımına devam edilebilen bir bankacılık ürünüdür.

5. UYGULAMA

5.1 UYGULAMAYA GENEL BAKIŞ

Uygulama bölümünde, özel bir bankanın kredi kullanmış müşterilerinin ödeme durumları incelenip kredisinin vadesi gelmeden tamamını ödeyerek erken kapatma eğilimi olan müşterilerin tahminlenebilmesine yönelik veri madenciliği uygulamasına yer verilmiştir.

Çalışma kapsamında kredi kullanmış müşterilerin mevcut kredilerinin ve kredili ürünlerinin ödeme geçmişlerine ait bilgilere ek olarak müşteri numaraları üzerinden cinsiyet, yaş, eğitim, meslek, çalışma durumu ve yaşadıkları şehir bilgilerine ulaşılmıştır. Gizlilik prensipleri nedeniyle müşteri numaraları değiştirilmiştir. Eksik ve hatalı veriler temizlenmiştir. Kullanılan veriler kategorik hale getirilmiştir. Uygulamada veri madenciliği tekniklerinden sınıflandırmaya yer verilmiştir.

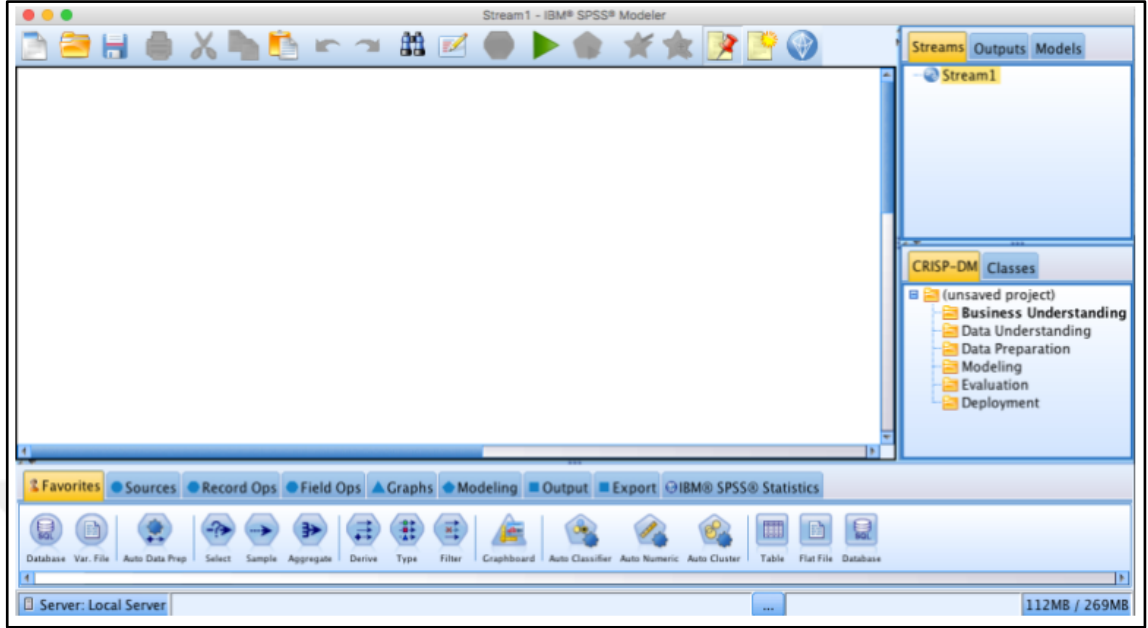
21.484 müşteri tarafından kullanılmış kredi verisi üzerinden çalışma yapılmıştır.

5.2 UYGULAMADA KULLANILAN YAZILIM

Çalışmada IBM firmasına ait olan Statistical Package for the Social Sciences (SPSS) Modeler yazılımı kullanılmıştır. SPSS Modeler, veri madenciliği sınıflandırma tekniklerini kullanmada oldukça güçlü, kullanıcı dostu ara yüzleri olan ve sonuçları görsel olarak gerçekleştirebilen bir araçtır.

Tüm fonksiyonlar ekranın en altında bir palette ikon şeklinde yer almaktadır. Modelin kurulması ve analiz için ikonların ana panele sürüklenmesi, bunların oklarla birleştirilmesi gerekmektedir. Şekil 5.1’de yazılıma ait bir ekran görüntüsü yer almaktadır.

Şekil 5.1: SPSS Modeler ekran görüntüsü



5.3 PROBLEMİN TANIMLANMASI

Bankalar kredi verirken müşterileri ile karşılıklı bir anlaşma içindelerdir. En ideal yapıda, müşteriler aldıkları kredileri zamanında ödemeleri beklenilir. Müşterilerin kredilerini erken kapatmaları durumunda, bankalar verdikleri üründen en başta anlaştıkları faiz gelirini kazanamamış olmakta ve müşterinin krediyi kapatması ile müşteriye kaybedebilmektedir. Diğer yandan rekabetin yüksek olduğu bankacılık sektöründe sürdürülebilir bir şekilde büyüebilmek için müşterilerin ihtiyaçları olan ürünler tespit edilip doğru zamanda doğru müşteriye doğru ürünü sunmak hem yeni müşteri kazanımında hem de mevcut müşteri memnuniyetinin artırılması ile sadakatin sağlanmasında önemlidir.

Bu çalışma ile kredi kullanmış olup erken kapatmış ve kapatmamış olan müşterilerden oluşan 21.484 satırlık veri kümesindeki müşterilerin demografik özelliklerinin incelenmesi ve hangi özellikteki müşterilerin hangi özellikteki kredisini vadesinden önce erken kapatma eğiliminde olduğunu tahminleyecek modelin tespit edilmesi amaçlanmaktadır. Buna bağlı olarak, banka bu modeli kullanarak tespit ettiği müşterilerine bireysel kredilerin yanı sıra bankanın diğer ürünlerinin de satışa

sunulabilmesi, mevcut ve potansiyel yeni kredi kullanacak müşteriler yine tespit edilerek onlara özel faiz indirimi sağlanan kredi kampanyaları sunulabilmesi gibi ileriye dönük pazarlama stratejisinin oluşturulmasında fayda sağlanacaktır.

5.4 VERİLERİN HAZIRLANMASI

Structured Query Language (SQL) server aracılığı ile elde edilen veri MS Excel'e aktarıldı ve datadan gereksiz olan verilerin çıkarılarak gerekli olanların birleştirilmesi ve dönüştürülmesi adımları bu bölümde detaylandırılmıştır.

5.4.1 Verilerin Toplanması

Özel bir bankanın dijital ve fiziki kanallarından kullanılmış olan kredilerden erken ödeme yapılarak kapatılmış ve erken ödeme ile kapatılmamış 21.484 kredisine ait banka ve Kredi Kayıt Bürosu (KKB)'dan alınan bilgiler ve krediyi kullanan müşterilerinin bankadaki demografik bilgileri Oracle veri tabanından alınmıştır. Hassas veri içeren müşteri bilgileri kullanılmadan anonimleştirilmiş değerler içeren veriler kullanılmıştır. Veri kümesinin ilk versiyonu olan ham veri hali aşağıdaki Tablo 5.1'de gösterilmiştir.

Tablo 5.1: Kategori edilmemiş ham veri

Değişken Adı	Müşteri - 1	Müşteri - 2	Müşteri - 3
Cinsiyet	E	E	K
Eğitim	İlkokul	Lisans	Lisans
Medeni Hal	1	2	1
Meslek	İşçi	Teknisyen	Bankacı
İl	08	03	34
Yaş	48	27	36
Çalışma Durumu	Ücretli - Özel	Ücretli - Özel	Ücretli -Kamu
Erken Kapatma	0	1	1
Vade	36	24	36
Tutar	2.000	2.000	5.000
Ara Ödeme	0	0	0

Toplem Erken Taksit	36	10	8
Kredi Kartı	2	6	6
24 Ayda Kapalı Kredi	1	5	14
Toplam Kredi	5	6	22
Toplam Kredi Kartı Ödeme	1.470	1.405	5.356
Toplam Ödenmiş Kredi	1.953	4.441	25.056
Son 6 Ay Ödeme Gecikme	0	0	0

5.4.2 Veri Kümesinin Tanıtımı

Veri kümesi, 21.484 müşteri tarafından kullanılmış krediye ait 18 özellik içerecek şekilde toplanmıştır. Veri kümesine ait bilgiler aşağıda detaylandırılmıştır.

Erken Kapatma: Kredinin erken mi, zamanında mı kapatıldığını gösteren alandır. Hedef değişkendir. Erken kapatılan krediler “1”, erken kapatılmamış krediler ise “0” ile gösterilmiştir. Tablo 5.2’de yer aldığı üzere, veri kümesinin yüzde 69,33’ü erken kapatılan kredilerden oluşmaktadır.

Tablo 5.2: Erken kapatma bilgisine göre dağılım

Erken Kapatma	Veri Satır Sayısı	Veri Dağılımı
0	6.590	30,67%
1	14.894	69,33%
Toplam	21.484	100,00%

Cinsiyet: Müşterilerin cinsiyet bilgisidir. Erkekler için “E”, kadınlar için “K” değerlerini almaktadır. Cinsiyet bilgisine göre data dağılımı Tablo 5.3’de yer almaktadır. Veri kümesinin yüzde 84,53’ü erkek müşterilerden oluşmakta iken onların da yüzde 58,73’ü kredisini erken kapatanlardan oluşmaktadır. Kadınların ise yüzde 10,60’sı kredisini erken kapatmışken yüzde 4,87’sinin erken kapatmamış olduğu görülmektedir.

Tablo 5.3: Cinsiyet bilgisine göre dağılım

Cinsiyet	Erken Kapatma Durumu	Veri Sayısı	Veri Dağılım Oranı	Toplam
Erkek	Evet	12.617	58,73%	84,53%
	Hayır	5.543	25,80%	
Kadın	Evet	2.277	10,60%	15,47%
	Hayır	1.047	4,87%	
Toplam		21.484	100,00%	100,00%

Eğitim: Müşterilerin eğitim bilgisini gösteren alandır. “İlköğretim”, “Lise”, “Lisans ve Üzeri” değerlerini almaktadır.

Medeni Hal: Müşterilerin medeni hallerini gösteren alandır. Tablo 5.4’te yer aldığı üzere “Evli” ve “Bekar” olarak 2 grupta toplanmıştır. Veri kümesindeki medeni duruma göre dağılımının yer aldığı Tablo 5.5’te görüldüğü üzere, evli müşterilerin yüzde 42,48 oranında erken kapatanlar, yüzde 17,65 erken kapatmayanlar olmak üzere toplam yüzde 60,13 oranla kümenin büyük bölümünü oluşturduğu görülmektedir.

Tablo 5.4: Medeni hal değerleri

Medeni Hal	Açıklama
1	Evli
2	Bekar

Tablo 5.5: Medeni hal bilgisine göre dağılım

Medeni Hal	Erken Kapatma Durumu	Veri Sayısı	Veri Dağılım Oranı	Toplam
Bekar	Evet	5.767	26,84%	39,87%
	Hayır	2.799	13,03%	
Evli	Evet	9.127	42,48%	60,13%
	Hayır	3.791	17,65%	
Toplam		21.484	100,00%	100,00%

Meslek: Müşterilerin mesleklerini gösteren alandır. Müşterilerin bankanın müşterisi olurken, bankanın daha önceden belirlediği kategoriler arasından seçim yapılarak oluşturulmuştur. EK A1’de veri kümesindeki müşterilerin meslek ya da meslek gruplarına göre dağılımları yer almaktadır. Tablo 5.6’da ise meslek grubu bazında veri sayısı en fazla olan ilk 5 meslek bilgileri ve bunlara ait erken kapatma oranları yer almaktadır. En yüksek, yüzde 27,04 oranı ile müşteri bankanın sunduğu seçeneklerde kendi mesleğini kategorize edememiş ve “Diğer” seçeneğini seçmiş olan müşterilerin meslekleri görünmektedir. Farklı meslekleri içinde barındırıyor olabilmesi nedeniyle, bu belirleyici bir bilgi değildir. İkinci sırada yüzde 12,76 ile “İşçi” mesleğine sahip müşteriler veri kümesindeki üçüncü sırada “Muhasebeci/Mali Müşavir” mesleğine sahip müşteriler vardır.

Tablo 5.6: Meslek bilgisine göre dağılımda ilk 5

Meslek	Erken Kapatma Durumu	Veri Sayısı	Veri Dağılım Oranı	Toplam
Diğer	Evet	3.972	18,49%	26,63%
	Hayır	1.749	8,14%	
İşçi	Evet	1.904	8,86%	12,33%
	Hayır	745	3,47%	
Muhasebeci/Mali Müşavir	Evet	725	3,37%	5,09%
	Hayır	368	1,71%	
Pazarlama/Satış Personeli	Evet	653	3,04%	4,23%
	Hayır	256	1,19%	
Memur	Evet	680	3,17%	3,99%
	Hayır	177	0,82%	

İl: Müşterilerin ikamet ettikleri il bilgisini göstermektedir. EK A.2’de illere göre müşteri dağılımları ve il bazında erken kapatma oranları yer almaktadır. Ve görüldüğü üzere veri kümesinde Türkiye’deki 81 il ve Kıbrıs’ta ikamet eden müşteriler yer almaktadır. İllere göre veriler gruplandığında en fazla kredi kullanılan ilk 5 il için bilgiler Tablo 5.7’de yer

almaktadır. Tabloya göre ilk 4 sırada yer alan iller nüfusa göre illeri sıraladığımızda da ilk 4 sırada yer alan illerdir.

Tablo 5.7: İl bilgisine göre dağılımda ilk 5

İl Kodu	İl	Erken Kapatma Durumu	Veri Sayısı	Veri Dağılım Oranı	Toplam
34	İstanbul	Evet	4.961	23,09%	34,65%
		Hayır	2.484	11,56%	
06	Ankara	Evet	1.536	7,15%	9,64%
		Hayır	535	2,49%	
35	İzmir	Evet	978	4,55%	6,42%
		Hayır	401	1,87%	
16	Bursa	Evet	711	3,31%	4,64%
		Hayır	285	1,33%	
41	Kocaeli	Evet	628	2,92%	4,05%
		Hayır	243	1,13%	

Yaş: Müşterilerin yaş bilgisidir. Doğum tarihleri baz alınarak müşterilerin yaşları hesaplanmıştır. Veri kümesinde en küçük 19 en büyük 69 yaşında müşteri verisi yer almaktadır. EK A.3'te yaş bazında kredi erken kapatma durumuna göre dağılım yer almaktadır. Veri kümesinde yaşlara göre kayıtlar gruplandığında ilk 5 yaşa göre bilgiler Tablo 5.8'de yer aldığı üzere, en fazla 31 yaşındakiler 783 kredi ile kümede yer alırken onu 744 kredi ile 29 yaşındakiler, 730 kredi ile 35 yaşındakiler takip etmektedir. İlk 5'te yer alanların yaş grubunun 29-35 arasında olduğu görülmektedir.

Tablo 5.8: Yaş bilgisine göre veri dağılımda ilk 5

Yaş	Erken Kapatma Durumu	Veri Sayısı	Veri Dağılım Oranı	Toplam Oran
31	Evet	717	3,337 %	4,62 %
	Hayır	276	1,28 %	
29	Evet	664	3,091 %	4,32 %

	Hayır	264	1,23 %	
35	Evet	654	3,044 %	4,23 %
	Hayır	255	1,19 %	
33	Evet	702	3,268 %	4,31 %
	Hayır	225	1,05 %	
34	Evet	624	2,904 %	4,15 %
	Hayır	268	1,25 %	

Çalışma Durumu: Müşterilerin çalışma durumlarını gösteren alandır. “Çalışan Özel”, “Çalışan Kamu”, “Çalışmıyor” değerleri yer almaktadır.

Vade: Kullanılan kredinin vadesini ay cinsinden göstermektedir.

Kullanılan Tutar: Kullanılan kredinin ana parasıdır, faiz dahil değildir.

Toplam Erken Taksit: Erken ödenen taksit sayısını göstermektedir.

Toplam Kredi Kartı: Müşterinin sahip olduğu tüm bankalardaki kredi kartı sayısını göstermektedir.

Son 24 Ayda Kapatılmış Kredi: Son 24 ay içerisinde müşterinin ödemesini tamamlayarak kapattığı kredi sayısını göstermektedir.

Toplam Kapalı Kredi: Müşterinin tüm kapalı kredi sayısını göstermektedir.

Toplam Kredi Kartı Ödeme: Müşterinin son dönem kredi kartına yaptığı ödemeyi göstermektedir.

Toplam Ödenmiş Kredi Ana Para: Tüm kredileri için o güne kadar toplam ödenmiş ana para tutarını göstermektedir.

Son 6 Ayda Gecikme Sayısı: Son 6 ay içerisinde geç ödenmiş kredili ürün sayısını göstermektedir.

5.4.3 Verilerin Temizlenmesi

Bu aşamada veri toplama aşamısında veri kümesine dahil edilen tüm alanlar için veriler tek tek değerlendirilmiştir.

Müşterilerin eğitim durumları veri kümesine dahil edilmiş, analiz esnasında tüm datanın yüzde 10'a yakın bir oranında bu verinin boş olduğu gözlemlenmiştir, değişkenin doldurulması hataya neden olabileceği için bu değişken çalışmadan çıkartılmıştır. Ayrıca meslek bilgisi boş değer içeren kayıtlar tespit edilmiş, bankadaki diğer ürünlerinde müşteri tarafından beyan edilen meslek bilgileri alınarak güncellenmiştir.

5.4.4 Verilerin Dönüştürülmesi

Uygulamada kullanılan veri setinde bazı özellikler üzerinde veri dönüştürme işlemleri yapılmıştır. Yapılan veri dönüştürme işlemlerine ait adımlar aşağıda belirtilmiştir.

Yaş aralığı 19 ile 69 arasında bulunan veri setindeki yaş bilgisi bankacılık yasal raporlarında kullanılan yaş aralıkları ve uzman görüşü ile 19-25, 26-35, 36-55, 56-69 olmak üzere 5 gruba ayrılmıştır. Tablo 5.9'da yaş verisinin dönüştürülmesi sonucunda elde edilen yeni veri seti yer almaktadır. Tabloya göre, 36-55 yaş aralığındaki müşteri sayısı ve erken kapatma oranı en yüksek aralıktır, hem müşteri sayısında hem de erken kapatma oranında ikinci sırada ise 26-35 yaş grubundaki müşteriler yer almaktadır.

Tablo 5.9: Yaş gruplarına göre dağılım

Yaş Grubu	Erken Kapatma Durumu	Veri Sayısı	Veri Dağılım Oranı	Toplam Oran
19-25	Evet	609	2,83%	4,42%
	Hayır	340	1,58%	
26-35	Evet	6.108	28,43%	39,72%

	Hayır	2.425	11,29%	
36-55	Evet	7.086	32,98%	48,19%
	Hayır	3.268	15,21%	
56-69	Evet	1.091	5,08%	7,67%
	Hayır	557	2,59%	
Toplam		21.484	100,00%	100,00%

İl bilgisinde Türkiye'deki 81 il ve Kıbrıs'a ait olan veriler EK A.4'te yer aldığı üzere Türkiye için 7 coğrafi bölge ve Kıbrıs için 1 bölge olmak üzere 8 gruba ayrılmıştır. Tablo 5.10'da bölge dönüştürülmesi sonrasında elde edilen veri dağılımı yer almaktadır.

Tablo 5.10: Bölge bilgisine göre dağılım

Bölge	Erken Kapatma Durumu	Veri Sayısı	Veri Dağılım Oranı	Toplam Oran
Bolge1	Evet	1.265	5,89%	8,69%
	Hayır	602	2,80%	
Bolge2	Evet	490	2,28%	3,20%
	Hayır	197	0,92%	
Bolge3	Evet	1.858	8,65%	12,30%
	Hayır	784	3,65%	
Bolge4	Evet	161	0,75%	1,13%
	Hayır	81	0,38%	
Bolge5	Evet	2.603	12,12%	16,73%
	Hayır	992	4,62%	
Bolge6	Evet	1.171	5,45%	7,75%
	Hayır	494	2,30%	
Bolge7	Evet	7.344	34,18%	50,20%
	Hayır	3.440	16,01%	
Bolge8	Evet	2	0,01%	0,01%
	Hayır	0	0,00%	
Toplam		21.484	100,00%	100,00%

Banka, kredi vadelerini önceden belirleyerek müşterilere belirlediği değerleri seçim için sunmaktadır. Yasal kurumların belirlediği vade üst sınırları zaman zaman değişiklik göstermesi, müşterilerin zaman zaman ara toplu ödemeler yaparak krediyi yapılandırması sonucu vadelerin belirlenen kategoriden farklılaşması söz konusudur. Tüm bunlar göz önüne alınarak, uzman görüşleri de doğrultusunda vade verisi 1 aydan 12 aya kadar olan vadeler için “1-12”, 13 aydan 24 aya kadar vadeler için “13-24”, 25 aydan 36 aya kadar olan vadeler için “24-36” ve şuanda üst sınır 36 olsa da daha önce 60’a kadar krediler kullandırılabilceği için de 36 ay ile 60 ay arasındaki vadeler için “37-60” olarak şekilde 4 kategoriye ayrılmıştır. Tablo 5.11’de vadeye göre dağılım yer almaktadır.

Tablo 5.11: Vade gruplarına göre dağılım

Vade Aralığı	Veri Satır Sayısı	Veri Dağılımı
1-12	5.618	26,15%
13-24	3.467	16,14%
25-36	7.561	35,19%
37-60	4.838	22,52%
Toplam Satır Sayısı	21.484	100,00%

5.5 MODELİN KURULMASI

5.5.1 Veri Setinin Uygulamaya Aktarılması ve Hedef Değişkenin Seçimi

Kredi kullanmış 21.484 müşterinin satır verisi, SPSS Modeler programının “Source / Excel” modülünden yararlanılarak programa aktarılmıştır. Programa aktarılan verinin tablo şeklinde gösterimini sağlamak için “Output/ Table” modülünden yararlanılmıştır. Şekil 5.2’de Modeller programına aktarılmış olan verinin tablo gösterimi yer almaktadır.

Şekil 5.2: Verinin programdaki görüntüsü

Table	Annotations	ONSİYET	MEDENİ_HAL	MESLEK	İL_BOLGE	YAS_GRUBU	ÇALIŞMA_S	ERKEN_KAPAMA_FLAG	VADE_GRUBU	KUL_TUTAR	KULL_FAZL_ORANI	TOPLAM_ERKEN_TAKSIT	TOPLAM_KK	24_KAPALI_KREDİ	TOPLAM_KAPALI_KREDİ	TOPLAM_KK_ÖDEME	TOPLAM_OD
1	Ernek	Evl	Bankacı	Bankacı	Bolge7	36-55	Çalışan - Özel 0	1-12	500	3.50	1	3	5	16	38642	2705424	
2	Ernek	Evl	Bankacı	Bankacı	Bolge7	36-55	Çalışan - Özel 0	1-12	500	3.50	1	3	5	16	38642	2705424	
3	Ernek	Evl	Bankacı	Bankacı	Bolge7	36-55	Çalışan - Özel 0	1-12	500	3.50	1	3	5	16	38642	2705424	
4	Ernek	Evl	Bankacı	Bankacı	Bolge7	36-55	Çalışan - Özel 0	1-12	500	3.50	1	3	5	16	38642	2705424	
5	Ernek	Evl	Muhasebeci/Mali Müşavir	Muhasebeci/Mali Müşavir	Bolge7	36-55	Çalışan - Özel 1	37-60	10000	2.09	36	4	23	31	60	6093	
6	Ernek	Evl	Muhasebeci/Mali Müşavir	Muhasebeci/Mali Müşavir	Bolge7	36-55	Çalışan - Özel 0	13-24	8000	2.09	24	4	23	31	60	6093	
7	Kadın	Evl	Bankacı	Bankacı	Bolge7	36-55	Çalışan - Özel 1	25-36	5000	2.09	8	6	14	22	5356	25056	
8	Ernek	Evl	Bankacı	Bankacı	Bolge7	36-55	Çalışan - Özel 0	1-12	500	3.50	1	5	0	3	10711	20732	
9	Ernek	Evl	Bankacı	Bankacı	Bolge7	36-55	Çalışan - Özel 0	1-12	500	3.50	1	3	2	5	12215	140559	
10	Ernek	Evl	İşçi	İşçi	Bolge6	36-55	Çalışan - Özel 0	25-36	2000	3.19	36	2	1	5	1470	1953	
11	Ernek	Evl	Diğer	Diğer	Bolge6	36-55	Çalışan - Özel 0	25-36	7200	1.89	8	3	2	4	2520	14422	
12	Ernek	Evl	Diğer	Diğer	Bolge7	56-69	Çalışmıyor 0	25-36	6250	1.89	2	3	1	11	1627	11820	
13	Ernek	Evl	İşletmeci	İşletmeci	Bolge7	36-55	Çalışan - Özel 1	1-12	2000	1.79	1	9	0	11	7726	54353	
14	Ernek	Evl	Diğer	Diğer	Bolge5	26-35	Çalışan - Ka. 1	25-36	2000	3.39	2	3	5	7	81	9360	
15	Ernek	Evl	Öğretmeni/Akademisyen	Öğretmeni/Akademisyen	Bolge1	36-55	Çalışan - Ka. 0	37-60	25000	1.99	24	4	1	7	12560	72155	
16	Ernek	Evl	Akser	Akser	Bolge5	36-55	Çalışan - Ka. 1	25-36	2000	2.99	2	5	7	24	3377	17309	
17	Ernek	Evl	Diğer	Diğer	Bolge5	56-69	Çalışmıyor 0	25-36	2000	2.99	35	3	8	29	2096	37449	
18	Ernek	Evl	Şoför	Şoför	Bolge3	36-55	Çalışan - Ka. 1	25-36	2000	3.39	1	2	1	2	350	5019	
19	Ernek	Evl	İşçi	İşçi	Bolge5	36-55	Çalışmıyor 1	25-36	4500	1.89	25	7	21	32	1965	1365	
20	Kadın	Evl	Diğer	Diğer	Bolge7	36-55	Çalışan - Özel 1	25-36	15000	2.09	2	1	4	13	3730	27275	
21	Ernek	Evl	Memur	Memur	Bolge7	56-69	Çalışan - Ka. 1	25-36	9000	2.29	2	4	14	36	8392	59990	
22	Ernek	Evl	Memur	Memur	Bolge7	56-69	Çalışan - Ka. 1	25-36	2000	3.39	1	4	14	36	8392	59990	
23	Ernek	Evl	Memur	Memur	Bolge7	56-69	Çalışan - Ka. 1	25-36	6250	1.99	2	4	14	36	8392	59990	
24	Ernek	Evl	Memur	Memur	Bolge7	56-69	Çalışan - Ka. 1	25-36	6250	2.29	1	4	14	36	8392	59990	
25	Ernek	Evl	Memur	Memur	Bolge7	56-69	Çalışan - Ka. 1	37-60	8700	2.29	2	4	14	36	8392	59990	
26	Ernek	Evl	Memur	Memur	Bolge7	56-69	Çalışan - Ka. 1	37-60	9500	2.29	6	4	14	36	8392	59990	
27	Ernek	Evl	Memur	Memur	Bolge7	56-69	Çalışan - Ka. 1	1-12	7000	2.29	2	4	14	36	8392	59990	
28	Ernek	Bekar	Diğer	Diğer	Bolge5	36-55	Çalışmıyor 1	25-36	4500	1.89	3	3	4	13	15	13526	
29	Ernek	Bekar	Diğer	Diğer	Bolge5	36-55	Çalışmıyor 1	25-36	3250	1.89	1	3	4	13	15	13526	
30	Ernek	Evl	Mimar/Mühendis	Mimar/Mühendis	Bolge5	36-55	Çalışan - Ka. 1	13-24	2000	2.29	7	5	13	19	2100	6794	
31	Ernek	Evl	Diğer	Diğer	Bolge3	56-69	Çalışmıyor 1	37-60	2500	1.99	2	7	5	26	900	3158	
32	Ernek	Evl	Diğer	Diğer	Bolge3	56-69	Çalışmıyor 1	25-36	2000	2.29	2	7	5	26	900	3158	
33	Ernek	Evl	Diğer	Diğer	Bolge3	36-55	Çalışan - Özel 1	25-36	2000	2.39	7	2	9	13	0	14537	
34	Ernek	Evl	Öğretmeni/Akademisyen	Öğretmeni/Akademisyen	Bolge1	56-69	Çalışan - Özel 0	25-36	5000	1.89	21	5	8	13	7700	77853	
35	Ernek	Evl	Tesisatçı/Tamirci/Camcı	Tesisatçı/Tamirci/Camcı	Bolge7	56-69	Çalışan - Özel 0	13-24	8000	1.99	15	4	6	17	905	8992	
36	Ernek	Evl	Teknisyen	Teknisyen	Bolge3	26-35	Çalışan - Özel 1	13-24	2000	1.89	10	6	5	8	1495	4441	
37	Ernek	Evl	Memur	Memur	Bolge7	56-69	Çalışan - Ka. 0	25-36	2000	2.99	32	6	8	16	3412	29889	
38	Kadın	Evl	Mimar/Mühendis	Mimar/Mühendis	Bolge7	36-55	Çalışan - Özel 1	25-36	6185	2.09	4	3	2	12	1900	31404	
39	Kadın	Evl	Mimar/Mühendis	Mimar/Mühendis	Bolge7	36-55	Çalışan - Özel 1	37-60	2500	2.09	8	3	2	12	1900	31404	
40	Ernek	Evl	Muhasebeci/Mali Müşavir	Muhasebeci/Mali Müşavir	Bolge1	56-69	Çalışan - Özel 0	1-12	2000	1.79	10	4	6	13	4300	44468	
41	Ernek	Evl	Esnaf/Tüccar	Esnaf/Tüccar	Bolge6	36-55	Çalışan - Özel 1	25-36	9250	2.29	1	7	1	15	8050	87587	
42	Ernek	Evl	Mimar/Mühendis	Mimar/Mühendis	Bolge5	36-55	Çalışan - Ka. 0	1-12	2000	1.79	1	7	8	17	4244	30124	
43	Ernek	Evl	Mimar/Mühendis	Mimar/Mühendis	Bolge5	36-55	Çalışan - Ka. 1	1-12	2000	2.39	2	7	8	17	4244	30124	
44	Ernek	Evl	İşçi	İşçi	Bolge6	36-55	Çalışan - Ka. 1	37-60	2500	1.99	10	5	4	7	1160	13926	
45	Ernek	Evl	İşçi	İşçi	Bolge6	36-55	Çalışan - Ka. 1	25-36	2000	3.19	17	5	4	7	1160	13926	
46	Ernek	Evl	Asp	Asp	Bolge7	26-35	Çalışan - Özel 0	1-12	2000	2.39	5	3	2	8	550	1317	
47	Ernek	Evl	Diğer	Diğer	Bolge5	36-55	Çalışmıyor 1	25-36	30000	3.65	7	6	8	18	14889	18848	
48	Kadın	Evl	Diğer	Diğer	Bolge7	36-55	Çalışan - Özel 1	25-36	5250	1.79	8	3	10	18	725	5892	
49	Ernek	Evl	Teknisyen	Teknisyen	Bolge3	36-55	Çalışan - Özel 1	25-36	2000	2.99	20	7	6	8	3600	19720	
50	Ernek	Bekar	Reklam/Halkla İlişiler	Reklam/Halkla İlişiler	Bolge1	36-55	Çalışan - Özel 1	37-60	13500	2.29	19	6	2	8	914	34356	
51	Ernek	Evl	Asp	Asp	Bolge6	36-55	Çalışan - Özel 1	25-36	6500	1.89	1	5	6	24	5148	710	
52	Ernek	Evl	İşletmeci	İşletmeci	Bolge7	36-55	Çalışan - Özel 1	25-36	2000	3.19	6	4	4	20	1951	57328	
53	Ernek	Evl	Diğer	Diğer	Bolge5	36-55	Çalışmıyor 1	25-36	2000	2.39	10	9	3	20	1500	8458	

Şekil 5.3’de model çalıştırmadan önce verilerin istatistiksel bilgilerini, data içerisindeki dağılımlarını açıklayıcı istatistikler bölümü (descriptive statics) görüntüsü yer almaktadır. Değişkenlerin özellikleri bu aşamada değerlendirilmiştir.

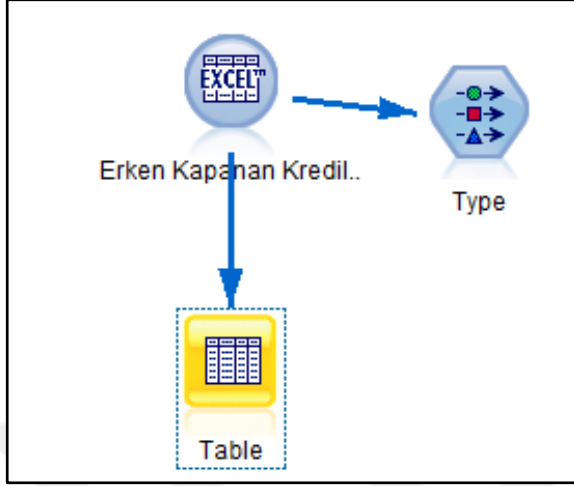
Şekil 5.3: Değişkenlerin “Descriptive Statics” görüntüsü

Data Audit of [16 fields] #1									
File Edit Generate									
Audit Quality Annotations									
Field	Sample Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
MUSTERI_NO		Continuous	1.000	17314.000	8470.253	4958.375	0.037	—	21484
CINSIYET		Flag	—	—	—	—	—	2	21484
MEDENI_HAL		Flag	—	—	—	—	—	2	21484
MESLEK		Nominal	—	—	—	—	—	63	21484
IL_BOLGE		Nominal	—	—	—	—	—	8	21484
YAS_GRUBU		Nominal	—	—	—	—	—	4	21484
CALISMA_SEKLI		Nominal	—	—	—	—	—	3	21484
ERKEN_KAPAMA_FLAG		Flag	0.000	1.000	—	—	—	2	21484
VADE_GRUBU		Nominal	—	—	—	—	—	4	21484
KUL_TUTAR		Continuous	100.000	35000.000	6086.287	6234.941	1.897	—	21484
TOPLAM_ERKEN_TAKSIT		Continuous	1.000	40.000	6.665	7.863	1.920	—	21484
TOPLAM_KK		Nominal	—	—	—	—	—	17	21484
24_KAPALI_KREDİ		Nominal	—	—	—	—	—	55	21484
TOPLAM_KAPALI_KREDİ		Nominal	—	—	—	—	—	89	21484
6AY_ÖDEME_GECİKME...		Nominal	—	—	—	—	—	14	21484

Not: Erken kapatılan krediler bordo renk, erken kapatılmayanlar rmavi renk ile gösterilmektedir.

Tüm verilerin SPSS Modeler programına aktarılmasından sonra değişkenler içinden hedef bir değişkenin seçilmesi gerekmektedir. Çalışmada “ERKEN_KAPAMA_FLAG” değişkeni yani kredinin erken ödenip ödenmediğini gösteren alan hedef değişken olarak belirlendi. Hedef değişkenin belirlenmesinde Modeler programının “Field Ops/Type” modülünden yararlanıldı. Bu modülün işletilmesiyle programda hedef değişken “Out” olarak işaretlendi. Modeler programının ekran görüntüsü Şekil 5.4’deki gibi oluştu.

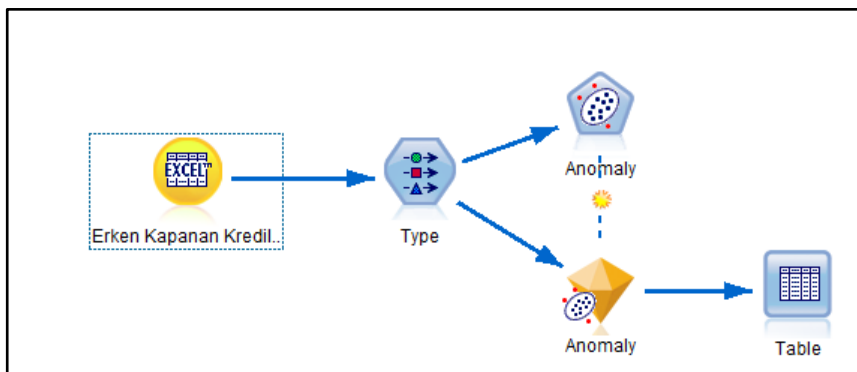
Şekil 5.4: Veri aktarımı ve hedef değişken seçimi sonrası Modeler ekran görüntüsü



5.5.2 Anormal Verilerin Belirlenmesi ve Veri Setinden Çıkartılması

Modeler programında, veri集中的 normal olmayan verilerin belirlenmesi için “Modeling/Anolmany” modülünden yararlanılır. Bu modül “Type” modülüne bağlanarak modeli çalıştırır ve normal olmayan verileri ortaya çıkarır. Şekil 5.5’de anormallik testi sonrasında Modeler sayfasında oluşan görüntü yer almaktadır.

Şekil 5.5: Anormallik testi sonrası SPSS Modeler ekranı



“Anolmany” modülü ile veriler program tarafından gruplanarak ayrı ayrı incelenmektedir. Bunun sonucunda her gruptaki farklılıklar anormallik indeksindeki değerlere göre belirlenmekte ve indeks değeri yüksek olan veriler ana veri setinden çıkartılmaktadır. Çalışmada kullanılan veri seti için Anolmany modülünün işletilmesi ile

birinci grupta 8, ikinci grupta 93, üçüncü grupta 113 olmak üzere toplam 214 anormal veri belirlenmiştir. Bu anormal veriler, veri setinden çıkartılmıştır. Yapılan analizler bu veri üzerinden gerçekleştirilmiştir. Şekil 5.6’da anormallik testinin gruplara göre sonuçları yer almakta, “Anomaly” alanında “T” değerini alan satırlar anormal kayıtları göstermektedir.Şekil 5.7’de ise tüm veri seti üzerinden gruplar olmadan toplam kaç kayıt anormal olarak ayrıldı, kaç kayıt ile devam edildiği yer almaktadır, burada yer aldığı üzere 214 anormal veri çıkartılmış ve veri setinde bulunan toplam veri sayısı 21.270’e inmiştir.

Şekil 5.6: Anormallik testi gruplara göre sonuç tablosu

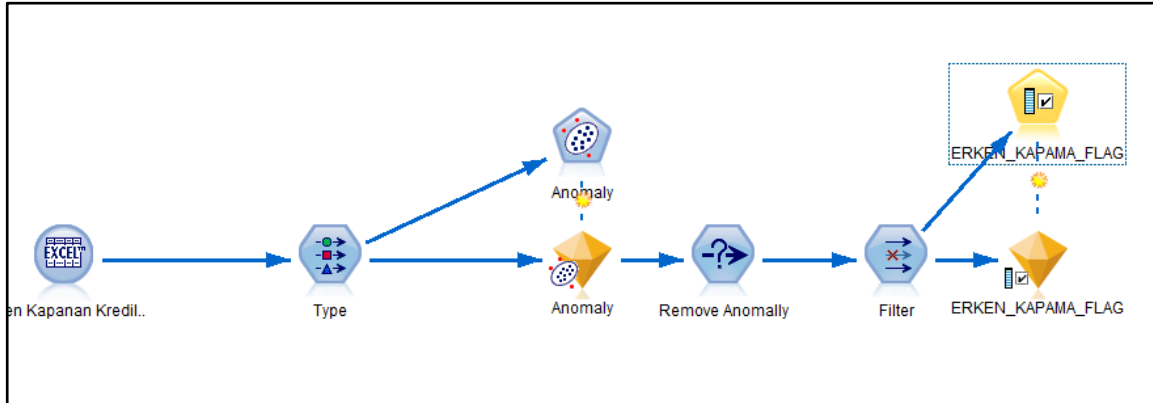
	\$O-Anomaly	\$O-PeerGroup	Record_Count
1	F	1	11402
2	T	1	8
3	F	2	5362
4	T	2	93
5	F	3	4506
6	T	3	113

Şekil 5.7: Anormallik testi özet tablosu

	\$O-Anomaly	Record_Count
1	T	214
2	F	21270

Anormallik testi sonrasında veri setindeki değişkenlerin özelliklerine bakılarak önemsiz, sorunlu, fazla eksik veri içeren yararlı olamayacak kadar çok yada az varyansı olan alanlar kaldırılması, kalan verilerin önem derecesine göre sıralanması ve sonraki modellerde kullanılacak özelliklerin alt kümesinin tanımlanması aşaması olan Özellik Seçim (Feature Selection) işlemi şekil 5.8’deki gibi oluşur.

Şekil 5.8: “Feature Selection” modülünün eklenmesi sonrası SPSS Modeller ekranı



Şekil 5.9’da ise özellik seçimi işleminin veri setimize uygulandığındaki çıktı ekran görüntüsüdür. Bu aşamada 2 değişkenin kullanılmamasına karar verildi. CINSIYET ve CALISMA_SEKLI alanları modellerde kullanılmayacaktır.

Şekil 5.9: Özellik Seçimi (Feature Selection) ekranı

Rank	Field	Measurement	Importance	Value
1	VADE_GRUBU	Nominal	Important	1.0
2	TOPLAM_ERKEN_T...	Continuous	Important	1.0
3	TOPLAM_KAPALI_K...	Nominal	Important	1.0
4	24_KAPALI_KREDİ	Nominal	Important	1.0
5	KUL_TUTAR	Continuous	Important	1.0
6	MESLEK	Nominal	Important	1.0
7	TOPLAM_KK	Nominal	Important	1.0
8	6AY_ÖDEME_GECİ...	Nominal	Important	1.0
9	YAS_GRUBU	Nominal	Important	1.0
10	MEDENI_HAL	Flag	Important	1.0
11	IL_BOLGE	Nominal	Important	1.0
12	CINSIYET	Flag	Unimport...	0.742

Selected fields: 11 Total fields available: 13

> 0.95 <= 0.95 < 0.9

1 Screened Fields

Field	Measurement	Reason
CALISMA_SEKLI	Nominal	Single category too large

5.5.3 Modelleme

Bu aşamada programın modeller arasından gerçek sonuçlara en uygun olan modelinin belirlenmesi hedeflenmektedir. Veri setinin hazırlığı bittikten sonra sıra modelin kurulması aşamasındadır.

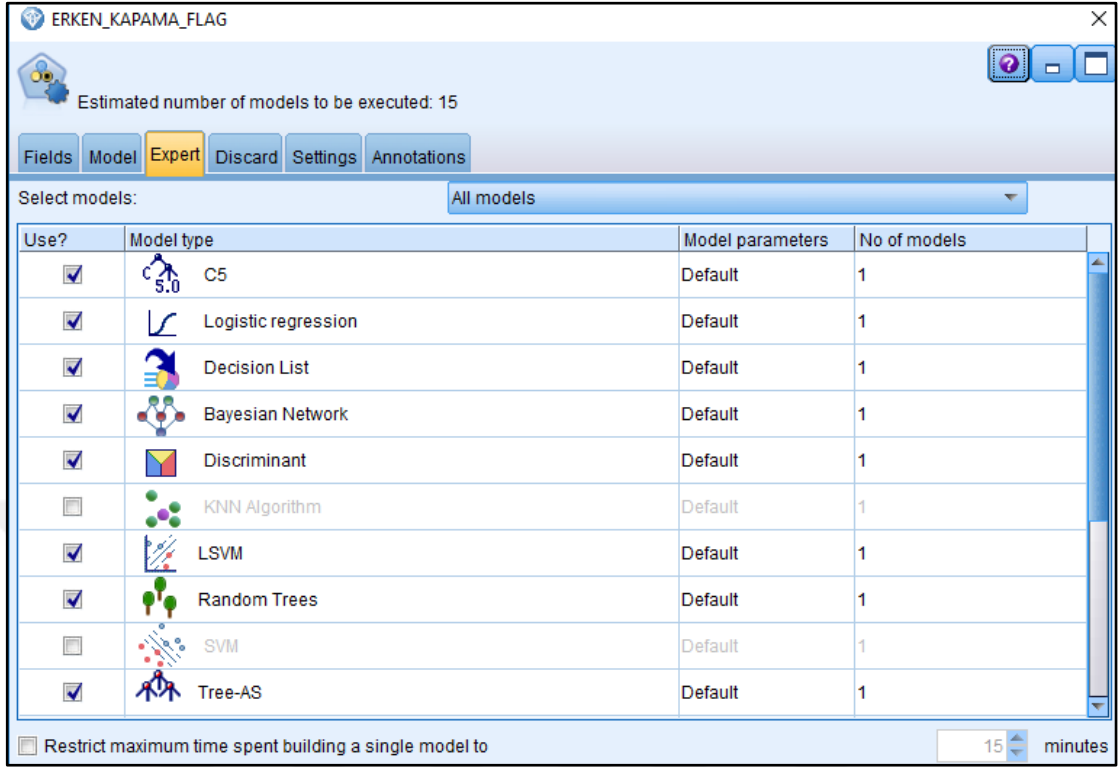
İlk olarak Partition Node’u kullanılarak data Şekil 5.10’da yer aldığı üzere yüzde 70 eğitim verisi (training), yüzde 30 ise test verisi (testing) olacak şekilde ayrıştırıldı.

Şekil 5.10: Eğitim ve test veri oranlarının belirlenmesi

Partition field:	Partition		
Partitions:	☒ Train and test ☐ Train, test and validation		
Training partition size:	70	Label: Training	Value = "1_Training"
Testing partition size:	30	Label: Testing	Value = "2_Testing"
Validation partition size:	0	Label: Validation	Value = "3_Validation"
Total size:	100%		
Values:	☐ Use system-defined values ("1", "2" and "3")		
	☒ Append labels to system-defined values		
	☐ Use labels as values		

Bir sonraki aşamada programın “Auto Classifier” modülü kullanılarak en uygun modelin seçilmesi işlemi yapıldı. Mümkün olan en fazla modeli çalıştırıp veri seti için en uygun modeli tespit etmesi beklendi. Şekil 5.11’de yer aldığı üzere 15 farklı model çalıştırılmıştır.

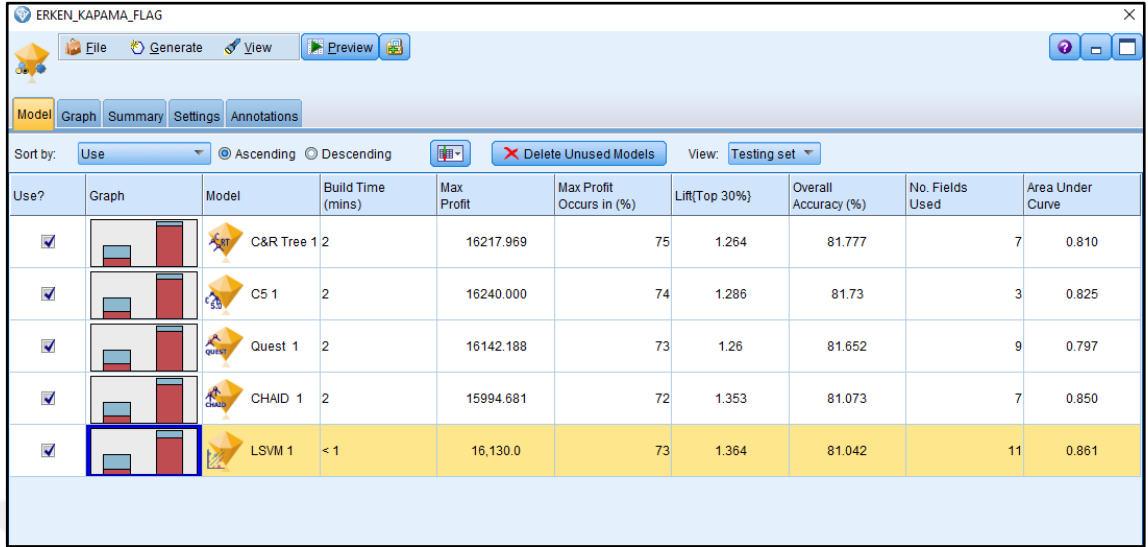
Şekil 5.11: Auto Classifier işlemcisi

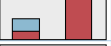


Şekil 5.12’de Auto Classifier işlemcisinde modeller çalıştırıldıktan sonra en yüksek doğruluk oranına sahip ilk 5 modele ait bilgiler yer almaktadır. Bu sonuçlara istinaden veri kümesini kullanılması ile kıyaslanan modellerin genel doğruluk oranlarına bakıldığında C&R Tree karar ağacı modeli yüzde 81,777 ile en yüksek genel doğruluk oranına sahiptir. Doğruluk oranına göre sonraki algoritmalar ise yine karar ağacı algoritmaları olan C5.0 yüzde 81,73, Quest yüzde 81,655, Chaid yüzde 81,073 ve yüzde 81,042 oranı ile linear SVM olduğu gözlemlenmiştir. Sonraki bölümde en yüksek oranlı ilk 3 model detaylı açıklanacaktır.

Kullanılacak modele karar verilirken Şekil 5.12’de yer alan “Overall Accuracy” alanı dikkate alınmıştır. Bu alan modelin doğru tahmin etme yüzdesini göstermektedir.

Şekil 5.12: Auto Classifier işlemcisi sonuçları



Use?	Graph	Model	Build Time (mins)	Max Profit	Max Profit Occurs in (%)	Lift (Top 30%)	Overall Accuracy (%)	No. Fields Used	Area Under Curve
☒		C&R Tree 1	2	16217.969	75	1.264	81.777	7	0.810
☒		C5 1	2	16240.000	74	1.286	81.73	3	0.825
☒		Quest 1	2	16142.188	73	1.26	81.652	9	0.797
☒		CHAID 1	2	15994.681	72	1.353	81.073	7	0.850
☒		LSVM 1	< 1	16,130.0	73	1.364	81.042	11	0.861

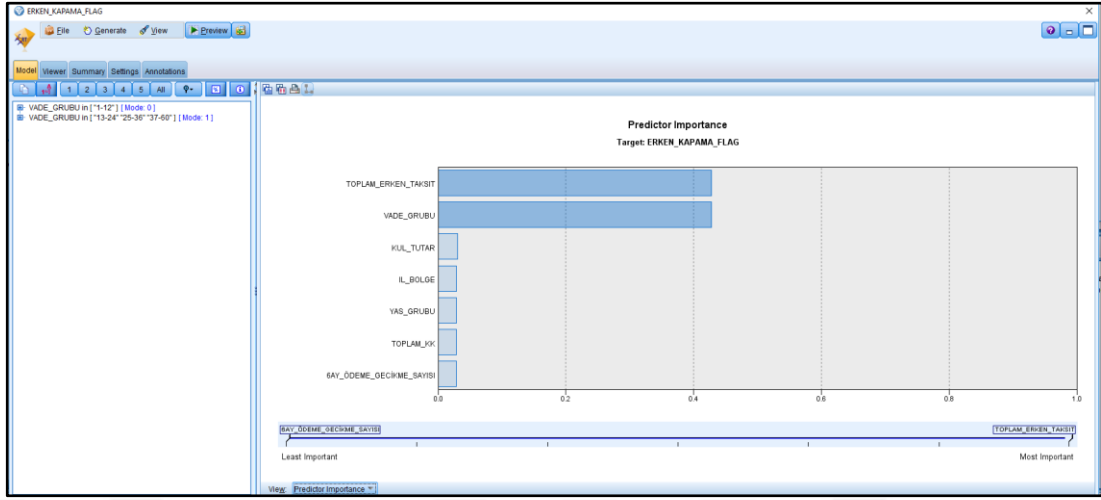
5.5.3.1 C&R Tree algoritmasına ilişkin sonuçların yorumlanması

C&R Tree (Classification and Regression Trees), Sınıflandırma ve Regresyon Ağaçları olarak tanımlanır. Karar ağacı oluşturmada kullanılan algoritmalarından biridir.

Hem sayısal hem de nominal türdeki veriler ile çalışabilen, sınıflandırma ve regresyon problemlerinde bir çözüm olarak kullanılan bir algoritmadır. C&RT ağacı, ikili olarak bölünen bir yapıdadır. Dallara ayrılma Gini indeksinden kriter olarak yararlanmaktadır. Ağaç kurulurken hiç bir durma kuralı olmadan devamlı bölünerek büyümektedir. Bölünme işlemi tamamlandıktan sonra ters yönlü uçtan köke budama başlatılır. Olabilecek en başarılı C&RT karar ağacı her budamadan sonra rastgele seçilen bir test verisi ile değerlendirilerek tespit edilmeye çalışılır (Sezer, ve diğerleri 2010).

C&R Tree modelinde Şekil 5.13’de yer aldığı üzere, model tarafından 7 farklı değişken seçilmiştir. Seçilen değişkenler arasında “Toplam Erken Ödenen Taksit Sayısı” ve “Vade Grubu” en önemli değişkenler olarak seçilmişken bunlara ek olarak “Kullanılan Tutar” alanı modelde kullanılmıştır. Diğer 4 değişkenin önem dereceleri 0,05’in altında olduğundan anlamlı bir fark yaratmayacağı için modelde kullanılmamıştır.

Şekil 5.13: C&R Tree algoritmasına göre önemli değişkenler

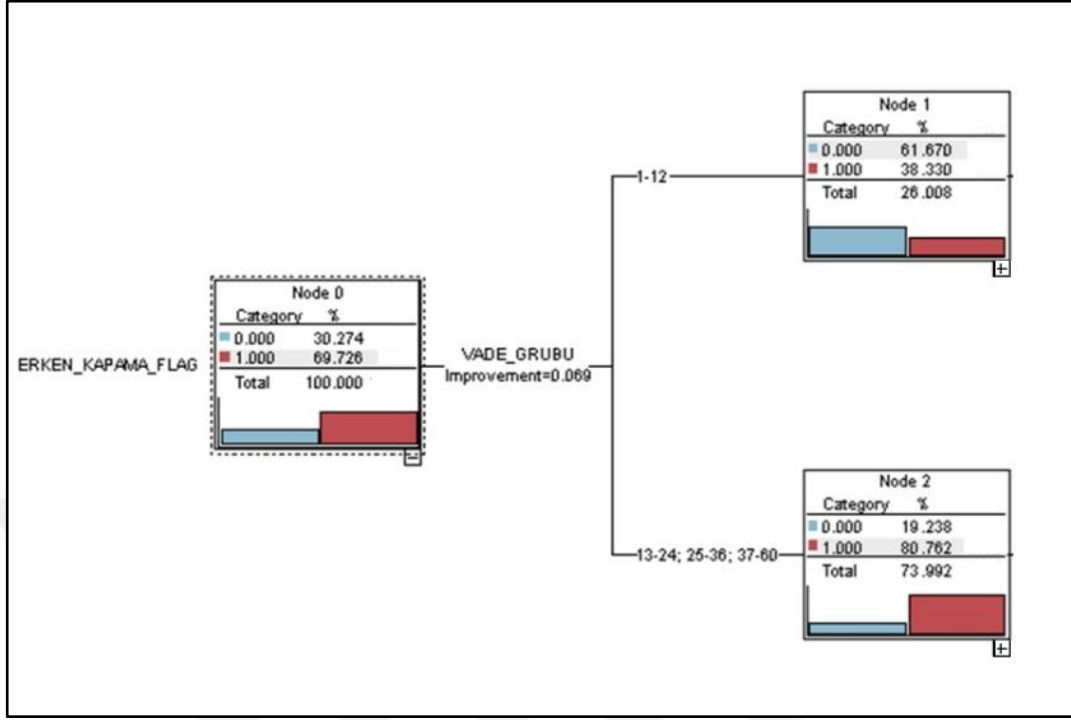


C&RT algoritmasına ilişkin karar ağacı EK A.5’de yer almaktadır.

Ağaç derinliği 5 olan karar ağacı, kök düğüm haricinde toplam 16 düğümden oluşmaktadır.

Şekil 5.14’te yer aldığı üzere, karar ağacında ilk dalın oluşumunda müşterilerin erken ödemeleri açısından kullanılan kredinin vadesinin yer aldığı vade grubu alanı kullanılmıştır. Model vadeye göre ilk dalı oluşturmuştur. 1-12 vade aralığına sahip veri yüzdesi 26,008 iken bu düğümdeki müşterilerin yarısından fazla bölümü erken kapatmamıştır. Diğer yandan 13-24, 25-36, 37-60 ay vade aralıklarına sahip kredilerin oluşturduğu düğüm ise kullanılan veri kümesinin yüzde 73,992’sini oluşturmakta ve bu sınıfın yüzde 80,762’si ise erken kredisini kapatan müşterilerdir.

Şekil 5.14: C&R Tree algoritması ile oluşan ilk dal



Vadeye göre 13-24, 25-36, 37-60 ay sınıfında yer alan müşteriler için karar ağacının bir sonraki dalında toplam erken ödenen taksit sayısı belirleyici değişken olmuştur ve 16.5 altında ve üstünde taksite sahip olan müşterileri içerecek şekilde 2 grup oluşmuştur. 16.5 ve daha altında erken ödenen taksit sayısına sahip müşterilerin yüzde 87,74'ü kredisini erken kapatmıştır.

Bu daldaki 16.5'tan daha yüksek sayıda taksiti olan müşteriler için bir sonraki dal vade grubuna göre oluşmuştur ve 37-60 ay vade grubu önemli bir sınıf oluşturmuştur. Bu sınıftaki müşterilerin yüzde 81,319'u kredisini erken kapatmıştır. Diğer yandan 13-24, 25-36 vadelerindeki müşterilerin sınıfı için karar ağacındaki bir sonraki dalında toplam erken ödenen taksit sayısı önemli bir ayırıcı kriterdir. 31.5'tan büyük erken kapanan taksit sayısına sahip müşterilerin yüzde 91,286'sının oluşturduğu çok büyük bölümü kredisini erken kapatmamıştır. 31,5'den küçük olanlarda ise vade grubuna göre tekrar bir dal oluşur ve 13-24 vade aralığındakilerin büyük bölümü erken kapatmamıştır, buradaki erken kapatma oranı ise yüzde 22,727'dir. 25-36 vade grubundaki müşterilerden kredilerini erken kapatanlar ile erken kapatmayanların oranı birbirine yakın olmakla birlikte erken kapatan müşteri oranı yüzde 58,416 ile daha yüksektir.

İlk dallanmada düğüm 1’de yer alan kredi vadesi 1-12 ay vade aralığında olan müşteriler için bir sonraki dal toplam erken ödenen taksit sayısına göre oluşmuştur. 8,5’tan daha fazla taksiti erken kapatan müşterilerin yüzde 88,997’si erken kapatmamışken, 8,5’tan daha az kalan taksiti olan 750 TL’ye eşit yada altında kredi kullanan müşterilerin hiçbiri erken kapatma yapmamıştır. 750 TL’den yüksek tutarda kredi kullanan müşteriler için belirleyici bir erken kapatma oranı olmayıp kapatanlar ile kapatmayanlar birbirlerine çok yakın orandadırlar. Bu sınıf tekrar toplam erken ödenen taksit sayısına göre gruplandığında 2.5 ve daha az taksiti olanlar arasında yüzde 67,485 oranı işe erken kapatanların sayısı fazla iken 2,5’tan daha fazla taksiti olanlar arasında erken kapatanlar yüzde 38,189 oranı ile erken kapatmayanlara göre daha azdır.

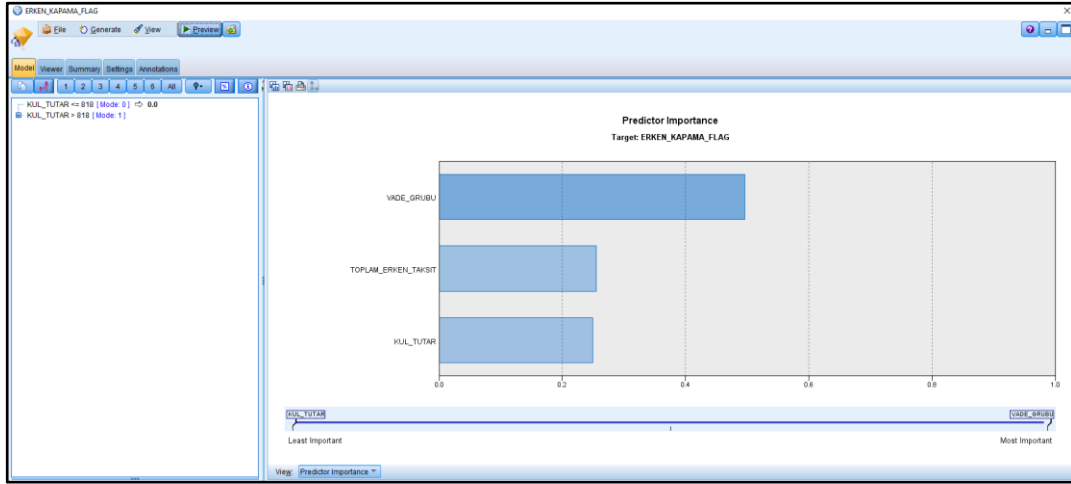
5.5.3.2 C5.0 algoritmasına ilişkin sonuçların yorumlanması

C5.0 algoritması C4.5 algoritmasının geliştirilmesiyle oluşmuş, daha çok büyük veri kümelerinde kullanılan, hızlı ve hafızayı başarılı kullanan bir karar ağacı algoritmasıdır. Karar ağacı algoritmalarından en çok kullanılanı C4.5 algoritmasıdır. C5.0 algoritması ise C4.5 algoritmasının geliştirilmesiyle oluşmuş, daha çok büyük veri kümelerinde kullanılmaktadır. Ayrıca C5.0 algoritması C4.5’e göre çok daha hızlıdır. C5.0 algoritması doğruluk oranını yükseltmek için boosting algoritmasını kullandığı için boosting ağaçları olarak da kullanılır (Sancak 2008).

C5.0 karar ağacında her bir düğümden çıkan çoklu dallar ile karar ağacı oluşur. Tahmin edinin kategori sayısı dalların miktarına eşittir.

C5.0 algoritması modelinde Şekil 5.15’deki önemli değişken olarak 3 farklı değişken seçilmiştir. En önemli görünen değişken “Vade Grubu” seçilmiş, önem sırasına göre sonrakiler “Toplam Erken Ödenen Taksit Sayısı” ve “Kullanılan Tutar” değişkenleridir.

Şekil 5.15: C5.0 algoritmasına göre önemli değişkenler



Algoritma uygulandığında oluşan karar ağacı EK A.6'da yer almaktadır.

C5.0 Karar ağacı algoritmasına göre tüm katılımcılar kök düğümden yaprak düğümlere kadar 24 alt grupta toplanmıştır.

Kök düğümde müşterilerin erken ödemeleri açısından kullanılan kredinin tutarı karar ağacında ilk dalın oluşumunda kullanılmış ve ayırt edici bir özellik olduğu görülmektedir. Kredi tutarı 818 TL'ye eşit yada daha küçük olan kredilerin tamamı erken kapatılmamış kredilerden oluşmaktadır.

Diğer yandan 818 TL'den büyük olan kredilerin erken kapatma oranı kök düğümdeki orana yakın olup bu sınıf için bir sonraki dal erken kapatılan taksit sayısına göre oluşmuştur. 8 taksit yada daha az taksiti olan krediler için erken kapatılma oranı bu sınıfın büyük orandaki verisini kapsamaktadır. Bir sonraki dalda bu düğüm vadeye göre ikiye ayrılmakta. 13-24, 25-36 ve 37-60 vade gruplarının yer aldığı sınıftaki verilerin çok büyük oranı erken kapatmışken, 1-12 vade grubu için erken kapatılan kredi sayısı ile erken kapatılmayan kredi sayısı hemen hemen eşittir. 1-12 vade grubunda yer alan düğümde erken ödeme yapılan taksit sayısının 1'den büyük olduğu sınıfta erken kapatılmayan krediler erken kapatılanlara göre fazla iken sonraki dal yine erken kapatılan taksit sayısına göre oluşturulmuş ve 4'ten fazla olduğu durumda erken kapatılmayan taksit sayısı daha fazladır. 4 yada daha az taksit için oluşan grupta erken kapatılan ve kapatılmayanların sayıları birbirlerine yakındır. Bu sınıf için oluşan son dal kullanılan tutara göre

oluşturulmuştur. 2.185 TL'ye eşit yada daha düşük tutarlı kredilerde erken kapanma oranı, 2.185 TL'den büyük tutarlı kredilerdeki erken kapatma oranına göre daha düşüktür. İkinci dalda oluşan toplam erken kapatılan taksit sayısı 8'den büyük olduğundaki düğüm 2 sınıfı bir sonraki kırılımda vade grubuna göre 4'e ayrılmıştır. Oluşan gruplarda; 1-12 vade grubundaki düğüm 13'te hemen hemen hiç biri kredisini erken kapatmamıştır. 13-24 vade grubundaki düğüm 14'te ise, erken kapatılmayan kredi oranı kapatılmalardan fazla ama birbirlerine yakın oranlara sahiplerdir. 25-36 vade grubundaki düğüm 19'da, yine erken kapatılan krediler ile erken kapatılmayan kredilerin oranları birbirlerine yakın olup, bu grupta erken kapatanların sayısı fazladır. Bu düğümdeki son grup olan 37-60 vade grubunda çok yüksek oranda erken kapatılan krediler yer almaktadır.

Düğüm 14 bir sonraki dalda toplam erken ödenen taksit sayısının 18'den büyük ya da küçük eşit olma durumuna göre sınıfları oluşturmuştur. 18'den büyük olduğunda oluşan düğüm 15'te, sınıftaki verilerin büyük oranı erken kapanmamış kredilerden oluşurken, 18'e eşit ya da daha küçük olduğu durumda erken kapanma durumuna göre oranlar birbirine yakın ama erken kapatılmış kredi oranı fazladır. Düğüm 15'ten sonra oluşan dal toplam erken ödenen taksit sayısının 14'ten büyük ya da küçük ve eşit olma durumuna göre oluşmuştur. 14'ten küçük yada eşit toplam erken ödenen taksit sayısı olan düğüm 16'da erken kapatanların sayısı erken kapatmayanlara göre fazladır.

Düğüm 19 bir sonraki dalda toplam erken ödenen taksit sayısına göre dallanmıştır. 29'dan büyük olduğunda oluşan düğüm 20'de kredisini erken kapatan sayısı fazla iken bu sınıfta kullanılan tutarın 15.875 TL'den büyük mü yoksa küçük veya eşit mi olması durumuna göre yeniden dallanmıştır. 15.875 TL'ye eşit yada ondan daha küçük olduğunda oluşan düğüm 21'de erken kapanan kredi oranı yüksekken, 15.875 TL'den büyük olanların oluşturduğu düğümde 22'de erken kapatan ve erken kapatmayan müşteri sayıları eşite yakındır.

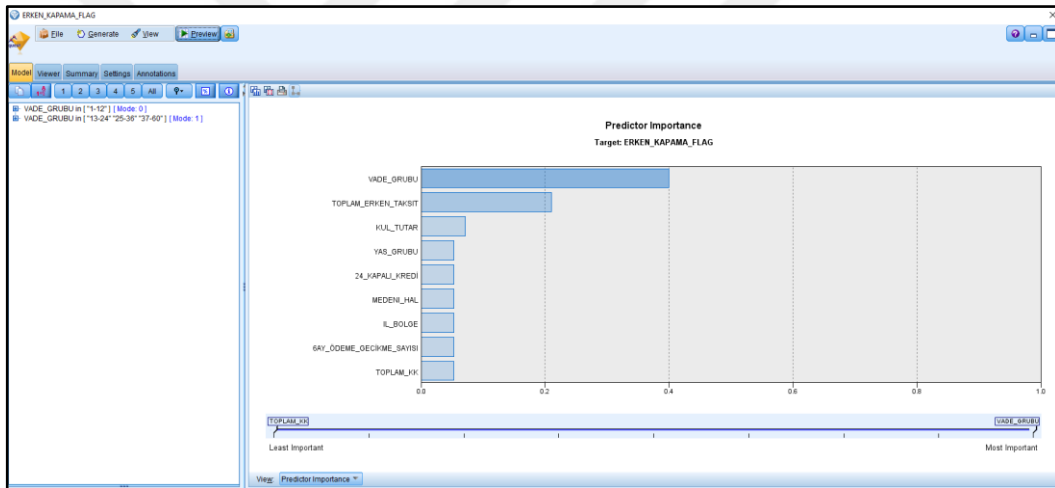
5.5.3.3 Quest algoritmasına ilişkin sonuçların yorumlanması

Sınıflandırma algoritmalarından QUEST algoritmasını, 1997 yılında geliştirmişlerdir. Budama ve doğrudan durma kuralı gibi yöntemleri kullanabildiği için ikili ağaç yapısını

kullanabilmektedir. CHAID ve C&RT algoritmalarında değişken seçilmesi ve bölünme aynı zamanda yapılmakta iken QUEST algoritması bunlara ayrı zamanlar ayırmaktadır. Dalların oluşması sırasındaki önyargılı seçimin daha genel hale getirilmesi ve hesaplama maliyetlerinin azaltılması için geliştirilmiştir (Sancak 2008).

QUEST modelinde Şekil 5.16’da yer aldığı üzere, ilk önce 9 farklı değişken seçilmiştir. Seçilen değişkenler arasından “Vade Grubu” en önemli değişken olarak belirlenmiş, ayrıca “Toplam Erken Ödenen Taksit Sayısı” ve “Kullanılan Tutar” değişkenleri de önemli bulunmuş ve modelde kullanılmıştır. Diğer 6 değişkenin önem dereceleri 0,05’in altında olduğundan anlamlı bir fark yaratmayacağı için modelde kullanılmamıştır.

Şekil 5.16: Quest algoritmasına göre önemli değişkenler



EK A.7’de görüldüğü üzere, bu algoritma sonucunda oluşan karar ağacı kök düğüm dahil edildiğinde 17 düğümden oluşmaktadır. Karar ağacında oluşan dallanmaların vade grubu ile başladığı görülmektedir.

13-24, 25-36, 37-60 vade gruplarına göre oluşan düğüm 2’deki grubun büyük bir kısmı kredisini erken kapatılmış kredilerden oluşmaktadır. Bu gruptaki sonraki dallanmalarda toplam erken ödenen taksit sayısı ve vade grubu değişkenleri önemli değişkenler olmuş ve dallanma bu değişkenlere göre devam etmiştir. Toplam erken kapatılan taksit sayısı 17,39’a eşit yada daha küçük olan kredilerin tamamına yakını erken kapatılmıştır. Düğüm 14’te yer aldığı üzere, toplam erken kapatılan taksit sayısı 22,57’den büyük olanlar daha çok erken kapatılmayacağı eğilimini göstermektedir. 22,57’den küçük olanlar ise tekrar

vadeye göre dallanmış ve 13-24 vade grubundakiler arasında erken kapatmayanlar büyük payı oluştururken, 25-36 ve 37-60 vade grubundaki krediler erken kapatılma eğilimindedirler.

Düğüm 1’de 1-12 vade grubu sınıfında ise toplam erken ödenen taksit sayısı 3,95’ten büyük olanların erken kapatmama oranı yüksektir, 3,95’ten küçük yada eşit olanlar içinse 4.102,17 TL’den daha yüksek tutarlı olanlarda erken kapatma eğitimi varken, 1.809,22 TL’ye eşit yada altındaki tutarlardaki kredilerin büyük bölümü erken kapatılmıyor. 1.809,22 TL – 4.102,17 TL aralığındaki kredilerden erken kapatılan taksit sayısı 2,32’den büyük olanlar erken kapatılmama eğiliminde, 2,32’ye eşit yada küçük olanların oluşturduğu grupta erken kapatılan krediler erken kapatılmayanlardan fazladır.

6. SAYISAL SONUÇLAR

Sınıflama modellerinde hata oranı hesaplanması için yanlış olarak sınıflanmış veri sayısının, tüm veri sayısına bölünmesi ile, doğruluk oranı ise doğru olarak sınıflanmış veri sayısının yine tüm veri sayısına bölünmesi ile hesaplanır (Akpınar 2000).

Bir önceki bölümde detayları yer aldığı üzere ham verimiz incelenip, temizleme ve dönüştürme süreçleri uygulandıktan sonra 3 farklı sınıflandırma algoritması ile yüzde 70 eğitim verisi, yüzde 30 test verisi olacak şekilde çalıştırılmıştır.

C&R Tree algoritması ile oluşan karar ağacı EK A.1’de, C5.0 algoritması ile oluşan karar ağacı EK A.2’de, Quest algoritması ile oluşan karar ağacı EK A.3’te yer almaktadır.

C&RT algoritmasının performans ölçüm sonuçlarına göre Şekil 6.1’de yer aldığı üzere, modelin eğitimdeki doğru tahminleme oranı yüzde 82,31’dir. 14.877 kredi eğitim modelinde kullanılmış ve bunların 12.246 tanesi için doğru tahminleme yapılmış , yüzde 17,69 oranına sahip 2.631 kredi içinde hatalı tahmin yapılmıştır. Eğitim detay verilerine göre de 11.370 kredinin erken kapanacağı varsayımında bulunulmuş ama bunların 9.535 tanesi erken kapatılmış, diğer yandan 3.507 kredinin erken kapatılmayacağı sonucuna varsa da gerçekte bunların 2.711 tanesi erken kapatılmamış, 796 kredi ise erken kapatılmıştır.

Şekil 5.15’deki test verilerine göre, 6.393 kredi verisi kullanılmış ve yüzde 81,78 oranında 5.228 kredi için doğru tahminleme sonucu elde edilmiştir. Hatalı tahmin edilen 1.165 kredi ise test verisinin yüzde 18,22’sini oluşturmaktadır. Test detay sonuçlarına göre ise; 4.856 kredinin erken kapanması beklenirken 806 tanesi erken kapanmamış, 4.050 tanesi erken kapanmıştır. 1.537 kredinin ise erken kapatılmayacağı tahminlenmişken 1.178 tanesi doğru tahminlenmiş, 359 tanesi ise hatalı tahminlenmiş ve erken kapatılmış kredilerdir.

Şekil 6.1: C&R Tree performans ölçüm sonuçları

Individual Models

Comparing \$R-ERKEN_KAPAMA_FLAG with ERKEN_KAPAMA_FLAG

'Partition'	1_Training		2_Testing	
Correct	12,246	82.31%	5,228	81.78%
Wrong	2,631	17.69%	1,165	18.22%
Total	14,877		6,393	

Coincidence Matrix for \$R-ERKEN_KAPAMA_FLAG (rows show actuals)

'Partition' = 1_Training	0.000000	1.000000
0.000000	2,711	1,835
1.000000	796	9,535

'Partition' = 2_Testing	0.000000	1.000000
0.000000	1,178	806
1.000000	359	4,050

Şekil 6.2'deki C5.0 algoritmasının performans ölçüm sonuçlarına ait tabloda yer aldığı üzere, modelin eğitimdeki doğru tahminleme oranı yüzde 82,99 dur. 14.877 kredi eğitim modelinde kullanılmış ve bunların 12.346 tanesi için doğru tahminleme yapılmış , yüzde 17,01 oranına sahip 2.531 kresi içinde hatalı tahmin yapılmıştır. Eğitim detay verilerine göre de 11.174 kredinin erken kapanacağı varsayımında bulunulmuş ama bunların 9.487 tanesi erken kapatılmış, diğer yandan 3.703 kredinin erken kapatılmayacağı sonucuna varsa da gerçekte bunların 2.859 tanesi erken kapatılmamış, 844 kredi ise erken kapatılmıştır.

Şekil 6.2'deki test verilerinde ise, 6.393 kredi verisi kullanılmış ve yüzde 81,73 oranında 5.225 kredi için doğru tahminleme sonucu elde edilmiştir. Hatalı tahmin edilen 1.168 kredi ise test verisinin yüzde 18,27'sini oluşturmaktadır. Test detay sonuçlarına göre ise; 4.773 kredinin erken kapanması beklenirken 766 tanesi erken kapanmamış, 4.007 tanesi erken kapanmıştır. 1.620 kredinin ise erken kapatılmayacağı tahminlenmişken 1.218 tanesi doğru tahminlenmiş, 402 tanesi ise hatalı tahminlenmiş ve erken kapatılmış kredilerdir.

Şekil 6.2: C5.0 algoritmasının performans ölçüm sonuçları

The screenshot displays the performance metrics for the C5.0 algorithm, comparing predicted values (\$C-ERKEN_KAPAMA_FLAG) with actual values (ERKEN_KAPAMA_FLAG). It includes a table for training and testing results, and a coincidence matrix for both partitions.

'Partition'	1_Training		2_Testing	
Correct	12,346	82.99%	5,225	81.73%
Wrong	2,531	17.01%	1,168	18.27%
Total	14,877		6,393	

'Partition' = 1_Training		0.000000	1.000000
0.000000		2,859	1,687
1.000000		844	9,487

'Partition' = 2_Testing		0.000000	1.000000
0.000000		1,218	766
1.000000		402	4,007

QUEST algoritmasının performans ölçüm sonuçlarına ait Şekil 6.3’de görüldüğü üzere, modelin eğitimdeki doğru tahminleme oranı yüzde 82,05’tir. 14.877 kredi eğitim modelinde kullanılmış ve bunların 12.207 tanesi için doğru tahminleme yapılmış, yüzde 17,95 oranına sahip 2.670 kredi içinde hatalı tahminleme yapılmıştır. Eğitim detay verilerine göre de 11.007 kredinin erken kapanacağı varsayımında bulunulmuş ama bunların 9.334 tanesi erken kapatılmış 1.673 tanesi erken kapatılmamıştır, diğer yandan 3.870 kredinin erken kapatılmayacağı sonucuna varsa da gerçekte bunların 2.873 tanesi erken kapatılmamış, 997 kredi ise erken kapatılmıştır.

Şekil 6.3’de yer alan test verilerinde ise, 6.393 kredi verisi kullanılmış ve yüzde 81,65 oranında 5.220 kredi için doğru tahminleme sonucu elde edilmiştir. Hatalı tahmin edilen 1.173 kredi ise test verisinin yüzde 18,35’ini oluşturmaktadır. Test detay sonuçlarına göre; 4.708 kredinin erken kapanması beklenirken 736 tanesi erken kapanmamış, 3.972 tanesi erken kapanmıştır. 1.685 kredinin ise erken kapatılmayacağı tahminlenmişken 1.248 tanesi doğru tahminlenmiş, 437 tanesi ise hatalı tahminlenmiş ve erken kapatılmış kredilerdir.

Şekil 6.3: Quest algoritmasının performans ölçüm sonuçları

Individual Models				
Comparing \$R-ERKEN_KAPAMA_FLAG with ERKEN_KAPAMA_FLAG				
'Partition'	1_Training		2_Testing	
Correct	12,207	82.05%	5,220	81.65%
Wrong	2,670	17.95%	1,173	18.35%
Total	14,877		6,393	
Coincidence Matrix for \$R-ERKEN_KAPAMA_FLAG (rows show actuals)				
'Partition' = 1_Training		0.000000	1.000000	
0.000000		2,873	1,673	
1.000000		997	9,334	
'Partition' = 2_Testing		0.000000	1.000000	
0.000000		1,248	736	
1.000000		437	3,972	

Kullanılan karar ağacı algoritmalarının sonuçları Tablo 6.1’de yer almaktadır. Buna göre doğruluk oranının en yüksek olduğu model yüzde 81,78 oranla C&R Tree algoritması ile elde edilmiştir.

Tablo 6.1: Algoritmaların karşılaştırılması

	C&RT	C5.0	QUEST
Doğru sınıflandırılan müşteri sayısı	5.228	5.225	5.220
Hatalı sınıflandırılan müşteri sayısı	1.165	1.168	1.173
Modelin Doğruluk Oranı (%)	81,78%	81,73%	81,65%

7. TARTIŞMA VE SONUÇ

Veri madenciliği, büyük veri kümelerindeki gizli kalmış olan anlamlı bilgilere ulaşmayı sağlayan bir süreçtir. Pazarlama, satış, sağlık, sigorta, elektronik ticaret gibi alanlarda olduğu gibi bankacılık sektöründe de karar verme süreçlerinde ihtiyaç duyulmasından dolayı yaygın olarak kullanılmaktadır.

Sürdürülebilir bir şekilde büyüebilmek için işletmelerin öncelikle müşterinin tanınması gerekmektedir. Müşterinin davranışları, satın alma alışkanlıkları ve geri bildirim verileri incelenerek işletmeler hem yeni müşteri kazanmaları hem de mevcut müşterilerinin sadakatini sağlamaları için faaliyetlerde bulunulmalıdır.

Bu çalışmada, bankacılık sektöründe kredisini vadesinden önce erken ödeyerek kapatan müşterilerin tahminlenmesine yönelik bir uygulama gerçekleştirilmiştir. Çalışma kapsamında SPSS Modeler programı ile veri kümesi yüzde 70 eğitim verisi (training), yüzde 30 ise test verisi (testing) olacak şekilde ayrıştırılarak C&R Tree, C5.0 ve Quest karar ağacı algoritmaları ile müşteriler sınıflandırılmış ve geleceğe yönelik tahminlerde bulunabilmek için kural çıkarımı sağlanmıştır.

Yapılan analiz çıktılarındaki doğruluk oranları dikkate alındığında, C&R Tree algoritması ile elde edilen modelin yüzde 81,78 doğruluk oranı ile geçerliliği, diğer algoritmalara göre daha yüksektir. C&R Tree algoritmasındaki oluşan karar ağacı incelendiğinde vadesi 13-24, 25-36, 37-60 aralığında olup kalan taksit sayısı 16 ve daha düşük olanların yüzde 87,74 oranında erken kapattıkları ve vade grubu 37-60 arasında olup 16'dan fazla erken kapatılacak taksiti olanların yüzde 81,3 oranında kredilerini erken kapattıkları görülmektedir.

Bu konuda literatürde farklı çalışmalar yer almaktadır. Örneğin, Çalış, çalışmasında bir banka şubesine ait bireysel kredi kullanan müşterilerin verilerini kullanarak, müşterilerin kredi geri ödeme performansına yönelik veri madenciliği kümeleme analizi tekniklerinden k-means algoritması ile mevcut müşterileri değerlendirmiş ve karar ağacı

algoritmaları ile müşterilerin ödeme durumlarına göre sınıflandırılması sağlanarak, gelecekteki potansiyel müşteriler için çıkarım yapılmıştır. Kümeleme analizinde 3 küme oluşmuş ve örneğin, araç değişkeninin 1. ve 3. kümeler için önemli olup , 2. küme için ayırt edici bir niteliğe sahip olmadığı gibi her bir değişkenin tek tek kümeler üzerindeki değerlendirilmiştir. İkinci aşamadaki karar ağacı algoritmaları ile ilgili çalışma bölümünde ise, C&RT, C5.0, CHAID ve QUEST algoritmaları uygulanmış ve sonuçlar değerlendirilmiş, bu çalışma için en başarılı algoritmanın yüzde 95,5 doğruluk payı ile chaid algoritması olduğu sonucuna varılmış ve tüm modeller için sonuçlar değerlendirilmiştir (Çalış 2013). Çalış çalışmasını sadece müşterilerin demografik bilgileri üzerinden geçerkleştirmiştir.

Sonuç olarak, bu tez çalışmasındaki veriler ışığında C&R Tree algoritması, mevcut kullanılmış krediler ve yeni kullanılacak olan krediler için farklı çalışmalarda kullanılabilir. Kredi başvuru modüllerindeki kurallar belirlenirken yada kredi teklifi sunmak için müşteri grupları belirlenirken müşterilerin belirli özellikleri modele yerleştirilip kullanılacak kredilerin erken kapatma ihtimalleri değerlendirilir ve erken kapatacağı sonucu elde edilen müşteriler için daha indirimli özel faiz oranı, daha esnek kısa vadede ödemelere özel ödeme planları uygulanabilir ve bankaya bağlılığı sağlanabilir.

Mevcut kredi kullanmış müşterilerin modelde kullanılan özellikleri üzerinden kredisini erken kapatacak olan müşteriler belirlendikten sonra bu müşterilerin bankanın herhangi bir kanalı ile temasında farklı çarpaz ürün satışları yapılabilir. Verimliliği artırılabilmek amacıyla bankanın diğer departmanlarında da benzer uygulamalar gerçekleştirilebilir. Geçmiş dönem ve anlık verileri kullanarak, tüketici davranışları eğilimlerini tahmin edebilen gerçek zamanlı uygulamalar geliştirilebilir.

KAYNAKÇA

Kitaplar

- Aras, G. 1996. *Ticari Bankalarda Kredi Portföyünün Yönetimi*. 1. Baskı. Ankara: Sermaye Piyasası Kurulu.
- Çelik, F. 2014. *Kredi Talepleri*. İstanbul: Türkmen Kitapevi.
- Çırpan, B. 2000. *Kredi Kartları*. 1. Baskı. Ezgi Kitapevi.
- Güney, A. 2017. *Banka İşlemleri*. 10. Baskı. İstanbul: Beta Yayım.
- Giudici, P. 2003. *Applied Data Mining: Statistical Methods for Business and Industry*. Chichester: John Wiley & Sons Inc.
- Hsieh, N. 2004. *An integrated data mining and behavioral scoring model for analyzing. Expert Systems With Applications*.
- Huberty, C. 1994. *Applied Discriminant Analysis*. Canada: John Wiley & Sons Inc.
- Kotler, P. 2003. *Marketing Insights from A to Z*. New Jersey.
- Özmen, Ş. 2009. *Ağ Ekonomisinde Yeni Ticaret Yolu: E-Ticaret*. İstanbul: Bilgi Üniversitesi Yayınları.
- Petersen, M. A. 2011. *Agile Marketing*. New York: A Press.
- Şakar, B. 2014. *Banka Kredileri ve Yönetimi*. 5. Baskı. İstanbul: Beta Yayıncılık.
- Usta, M. 1995. *Temel Kredi Bilgileri*. İstanbul: Pamukbank T.A.Ş. Eğitim Yayınları.
- Vurucu, M., & Arı, M. 2017. *Güncel Gelişmeler Işığında Bankacılıkta Kredi Bilgisi*. 1. Baskı. Ankara: Seçkin Yayıncılık San. ve Tic. A.Ş.

Sürekli Yayınlar

- Akpınar, H. 2000. Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, **29** (1), ss. 1-22.
- Aşan, Z. 2007. Kredi kartı kullanan müşterilerin sosyo ekonomik özelliklerinin kümeleme analizi ile incelenmesi. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*. **17**, ss. 256-267.
- Bose, I., & Mahapatra, R. 2001. Business data mining - a machine learning perspective. *Information & Management*. **39** (3), pp. 211-225.
- Budak, H., & Erpolat, S. 2012. Kredi Riski Tahmininde Yapay Sinir Ağları ve Lojistik Regresyon Analizi Karşılaştırılması. *Online Academic Journal of Information Technology*. **3**, pp. 23-30
- Dumouchel, W. 1999. Bayesian Data Mining in Large Frequency Tables With an Application to the FDA Spontaneous. *American Statistician*. **53** (3), pp. 190-196.
- Hui, S., & Jha, G. 2000. Application Data Mining for Customer Service Support. *Information and Management*. **38**, pp. 1-13.
- Kalogirou, S. A. 2000. Applications of Artificial Neural-Networks for Energy Systems. *Applied Energy*. pp. 17-35.
- Kaya, H., & Köymen, K. 2008. Veri Madenciliği Kavramı Ve Uygulama Alanları. *Doğu Anadolu Bölgesi Araştırmaları Dergisi*. ss. 159-164.
- Kalaycı, T. 2018. Kimlik hırsızlı web sitelerinin sınıflandırılması için makine öğrenmesi yöntemlerinin karşılaştırılması. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*. **24** (2), ss. 870-888.
- Kavcıoğlu, Ş. 2019. Kurumsal Kredi Skorlamasında Klasik Yöntemlerle Yapay Sinir Ağı Karşılaştırması. *İstanbul İktisat Dergisi*. **69** (2), ss. 207-245.

- Namlı, E., & Murat, S. 2019. Müşteri Odaklı Pazarlama Stratejileri İçin Veri Madenciliği Teknikleri Kapsamında Perakende Sektöründe Kümeleme Analizi Uygulaması. *Gümüşhane Üniversitesi Sosyal Bilimler Enstitüsü Elektronik Dergisi*. **10**, ss. 317-327.
- Pawlak, Z. 2002. Rough Sets Decision Algorithms and Bayes Theorem. *European Journal of Operation Research*. **136** (1), pp.181-189.
- Ranjan, J., R. Singh, ve V. Bhatnagar. 2011. Analytical Customer Relationship Management in Insurance Industry Using Data Mining: A Case Study of Indian Insurance Company. *International Journal of Networking & Virtual Organisations*. **9**(4), pp. 331-366.
- Restivo, K. 1999. The Drill on Data Mining. *Computer Dealer New*. **15** (14), pp. 25-30.
- Sezer, E. A., Bozkır, A. S., Yağız, S., & Gökçeoğlu, C. 2010. Karar Ağacı Derinliğinin Cart Algoritmasında Kestirim Kapasitesine Etkisi: Bir Tünel Açma Makinesinin İlerleme Hızı Üzerinde Uygulama. *Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu*. Kayseri. ss. 21-24.
- Shah, S., & Kursak, A. 2004. *Data mining and Genetic Algorithms Based Gene/ SNP Selection*. Artificial Intelligence in Medicine. **31**, pp. 183-196.
- TBB. 2011. *5411 Sayılı Bankacılık Kanunu*. İstanbul: Türkiye Bankalar Birliği Yayın No:273.
- TBB, Risk Merkezi, Aylık Bülten, 2020. [Çevrimiçi] https://www.riskmerkezi.org/Content/Upload/istatistikiraporlar/ekler/1709/Risk_Merkezi_Aylık_Bulten_Ozeti_Ocak_2019.pdf [erişim tarihi 2 Kasım 2020]. https://www.riskmerkezi.org/Content/Upload/istatistikiraporlar/ekler/2222/Risk_Merkezi_Aylık_Bulten_Ozeti_Ocak_2020.pdf [erişim tarihi 2 Kasım 2020].
- Wei, C., & Chiu, T. 2002. *Turning Telecommunications Call Details to Churn Prediction: a Data Mining Approach*. Expert Systems with Applications. **23**, pp. 103-102.

Diğer Yayınlar

- Akbulut, S., (2006). Veri Madenciliği Teknikleri ile Bir Kozmetik Markanın Ayrılan Müşteri Analizi ve Müşteri Segmentasyonu. *Yüksek Lisans Tezi*. Ankara: Gazi Üniversitesi, Fen Bilimleri Enstitüsü.
- Aksoy, E., (2019). Bankalarda bireysel kredilerin riskinin gm(1,1) modeli ile tahmin edilmesi. *Yüksek Lisans Tezi*. Isparta: Süleyman Demirel Üniversitesi, Sosyal Bilimler Enstitüsü.
- Anand, A., (2003). A study and comparison of data clustering techniques. *Master Thesis*. Texas: Faculty of The Graduate School of The University of Texas El Paso.
- Bilekdemir, G., (2010). Veri Madenciliği Tekniklerini Kullanarak Üretim Süresi Tahmini Ve Bir Uygulama. *Yüksek Lisans Tezi*. İzmir: Dokuz Eylül Üniversitesi.
- Bland, C., (2002). The Discovery of Multiple Level Profile Association Rules. *Doctoral Thesis*. Mississippi: Graduate School of Computer and Information Science of University of Mississippi.
- Çalış, A., (2013). Veri madenciliği yaklaşımı ile bireysel müşterilerin kredi ödeme performanslarının değerlendirilmesi. *Yüksek Lisans Tezi*. Kocaeli: Kocaeli Üniversitesi, Fen Bilimleri Enstitüsü.
- Çetinyokuş, T., (2008). Veri Küplerinin Bütünleşik Kullanımına Yönelik Yeni Bir OLAP Mimarisi. *Doktora Tezi*. Ankara: Gazi Üniversitesi, Fen Bilimleri Enstitüsü.
- Hudairy, H., (2004). Data mining and decision making support in the governmental sector. *Master Thesis*. Kentucky: Faculty of Graduate School of The University
- Katılmış, S., (2018). Veri Madenciliği Yöntemleri İle Banka Dijital Kanallarına Eğilimli Müşterilerin Belirlenmesi. *Yüksek Lisans Tezi*. İstanbul: Bahçeşehir Üniversitesi Fen Bilimleri Enstitüsü.

- Kuzucu, K., (2015). Müşteri memnuniyetini belirlemek için metin madenciliği tabanlı bir yazılım aracı. *Yüksek Lisans Tezi*. İstanbul: Maltepe Üniversitesi Fen Bilimleri Enstitüsü.
- Salur, M., (2016). Büyük Veri Araçlarından Hadoop Kullanarak Veri Madenciliği. *Yüksek Lisans Tezi*. Denizli: Pamukkale Üniversitesi, Fen Bilimleri Enstitüsü.
- Sancak, S., (2008). Saldırı tespit sistemleri tekniklerinin karşılaştırılması. *Yüksek Lisans Tezi*. Kocaeli: Gebze İleri teknoloji Enstitüsü, Sosyal Bilimler Enstitüsü.
- Söğütçü, M., (2011). Tüketici ve İhtisas Kredilerinin Bazı Makroekonomik Göstergeler Üzerindeki Etkisinin Analizi: Türkiye Örneği. *Yayımlanmamış Yüksek Lisans Tezi*. Konya: Dumlupınar Üniversitesi- Sosyal Bilimler Enstitüsü.
- Tekingöz, Ö., (2016). Dijital Yayıncılıkta İçerik İzleme Oranlarına Göre Müşteri Kümelenmesi. *Yüksek Lisans Tezi*. İstanbul: Bahçeşehir Üniversitesi, Fen Bilimleri Enstitüsü.
- Tosun, T., (2006). Veri madenciliği teknikleriyle kredi kartlarında müşteri kaybetme analizi. *Yüksek Lisans Tezi*. İstanbul: İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü.
- Turgut, H., (2012). Veri Madenciliği Süreci Kullanılarak Alzhemimer Hastalığı Teşhisine Yönelik Bir Uygulama. *Yüksek Lisans Tezi*. Isparta: Süleyman Demirel Üniversitesi, Fen Bilimleri Enstitüsü.
- Turgut, K., (2019). Banka Kredi Hacmi ile İktisadi Büyüme Arasındaki İlişki: Türkiye Örneği. *Yayımlanmamış Yüksek Lisans Tezi*. Ankara: Hacettepe Üniversitesi- Sosyal Bilimler Enstitüsü.
- Yapıcı, A, A Özel, & C Ayça., (2010). Oracle data miner ile kredi ödemeleri üzerine bir veri madenciliği uygulaması. *Bitirme Ödevi*. İstanbul: İstanbul Teknik Üniversitesi