



T.C.
İZMİR KATİP ÇELEBİ ÜNİVERSİTESİ
FEN BİLİMLER ENSTİTÜSÜ
YAZILIM MÜHENDİSLİĞİ ANABİLİM DALI
İKİNCİ ÖĞRETİM TEZSİZ YÜKSEK LİSANS DÖNEM PROJESİ

MAKİNE ÖĞRENMESİ TEKNİKLERİYLE
İSKONTO/İŞTİRA KREDİLERİNİN DEĞERLENDİRİLMESİ

İBRAHİM TOLGA AĞIM

Danışman
Prof.Dr. AYTUĞ ONAN

İZMİR – 2024

T.C.
İZMİR KATİP ÇELEBİ ÜNİVERSİTESİ
FEN BİLİMLER ENSTİTÜSÜ

TEZSİZ YÜKSEK LİSANS DÖNEM PROJESİ ONAY FORMU

Öğrenci No : 220234168
Adı ve Soyadı : İBRAHİM TOLGA AĞIM
Anabilim Dalı : Yazılım Mühendisliği
Tezsiz Yüksek Lisans Programı : Yazılım Mühendisliği İkinci Öğretim
Dönem Projesinin Başlığı : Makine Öğrenmesi Teknikleriyle
İskonto/İştira Kredilerinin
Değerlendirilmesi

Yukarıda belirtilen Tezsiz Yüksek Lisans öğrencisinin hazırladığı Dönem Projesi tarafımdan okunmuş, kapsamı ve niteliği açısından incelenmiş ve aşağıda belirtilen karara varılmıştır. Öğrencinin Dönem Projesi;

- Kapsamlı bir derlemedir
- Eleştirel bir rapordur
- Uygulamaya dönük bir projedir
- Deneysel bir çalışmadır

ve tarafımdan **BAŞARILI** bulunmuştur.

Prof.Dr. Aytuğ ONAN

16/07/2024

YEMİN METNİ

Uzaktan Öğretim Tezsiz Yüksek Lisans Dönem Projesi olarak sunduğum “Makine Öğrenmesi Teknikleriyle İskonto/İştira Kredilerinin Değerlendirilmesi” adlı çalışmanın, tarafımdan, akademik kurallara ve etik değerlere uygun olarak yazıldığını ve yararlandığım eserlerin kaynakçada gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve bunu onurumla doğrularım.

16/07/2024

İbrahim Tolga AĞIM

ÖZET

Uzaktan Öğretim Tezsiz Yüksek Lisans Dönem Projesi

Makine Öğrenmesi Teknikleriyle

İskonto/İştira Kredilerinin Değerlendirilmesi

İbrahim Tolga AĞIM

Katip Çelebi Üniversitesi

Fen Bilimleri Enstitüsü

Yazılım Mühendisliği Anabilim Dalı

Yazılım Mühendisliği Bölümü / Yüksek Lisans (Tezsiz)

Ticari işlemlerden doğan ve vadesi henüz gelmemiş kambiyo senedinin banka tarafından nakde dönüştürülmesi işlemine iskonto ve iştira kredileri adı verilmektedir. İskonto ve iştira işlemini gerçekleştirilen banka, senet borçlusunu ve varsa cirantaların ticari faaliyetlerine devam etme ve taahhütlerine uyum durumlarını kontrol ederek, araştırma sonuçlarının olumlu olması halinde işlemi kabul etmektedir.

Sağlam kredi riski yönetimi için kilit faktörün verimli kredi tahsisi olduğu bilinmektedir. Maruz kalınan kredi riskini kabul edilebilir çerçevede tutulması, kredi riski yönetiminin ana amacı konumundadır. Kredi riskinin minimize edilmesi için kredi tahsisi aşamasında müşterinin çok iyi analiz edilmesi gerekmektedir. Söz konusu analiz çalışması sırasında müşterinin sahip olduğu ekonomik güç, piyasadaki itibarı, ne kadar güvenilir olduğu, uygulanacak teminat yapısı gibi faktörlerin detaylı olarak değerlendirilmesi ve kredi tahsisinin bu analizler neticesinde gerçekleştirilmesi önem arz etmektedir. Bu çalışmada, iskonto ve iştira kredileri için yapılan kredi tahsis çalışmalarında makine öğrenmesi tekniklerinin kullanımının mümkün olup olmayacağı araştırılacaktır.

Anahtar Kelimeler: Ticari Kredi, İskonto ve İştira Kredileri, Kredi Risk Analizi, Yapay Zeka, Makine Öğrenmesi

ABSTRACT

Master's Graduate Project

Evaluation of Discount/Purchase Loans

With Machine Learning Techniques

İbrahim Tolga AĞIM

Katip Çelebi University

Department of Software Engineering / Master of Science (without Thesis)

The process of converting the bill of exchange notes arising from commercial transactions and which have not yet been due into cash by the bank is called discount and purchase loans. The bank, which has made the discount and purchase transaction, checks the status of the debtor of the promissory note and endorsers, if any, to continue their commercial activities and to comply with their commitments, and accepts the transaction if the results of the research are positive.

It is known that the key factor for sound credit risk management is efficient credit allocation. Keeping the exposed credit risk within an acceptable framework is the main purpose of credit risk management. In order to minimize the credit risk, it is necessary to analyze the customer very well at the credit allocation stage. During the analysis study in question, it is be important to evaluate factors such as the economic strength of the customer, its reputation in the market, how reliable it is, the collateral structure to be applied in detail and to carry out credit allocation as a result of these analyzes. In this study, it will be investigated whether it is possible to use machine learning techniques in credit allocation studies for discount and purchase loans.

Keywords: Commercial Credit, Discount and Purchase Loans, Credit Risk Analysis, Artificial Intelligence, Machine Learning

**MAKİNE ÖĞRENMEŞİ TEKNİKLERİYLE
İSKONTO/İŞTİRA KREDİLERİNİN DEĞERLENDİRİLMESİ**

İÇİNDEKİLER

DÖNEM PROJESİ ONAY SAYFASI	Hata! Yer işareti tanımlanmamış.
YEMİN METNİ	iii
ÖZET	iv
ABSTRACT	v
İÇİNDEKİLER	vi
KISALTMALAR	viii
TABLolar LİSTESİ	ix
ŞEKİLLER LİSTESİ	xi
GİRİŞ	1

BİRİNCİ BÖLÜM TEMEL KAVRAMLAR

1.1. KREDİ KAVRAMI	3
1.2. İSKONTO VE İŞTİRA KREDİLERİ	4
1.3. KREDİ TAHSİSİ	4

İKİNCİ BÖLÜM MAKİNE ÖĞRENMESİ

2.1. MAKİNE ÖĞRENMESİ KAVRAMLARI	6
2.2. MAKİNE ÖĞRENMESİ TÜRLERİ	8
2.2.1. Denetimli Öğrenme	8
2.2.2. Denetimsiz Öğrenme	9
2.3. MAKİNE ÖĞRENMESİNİN TEMEL ZORLUKLARI	9

ÜÇÜNCÜ BÖLÜM UYGULAMA

3.1. İHTİYACIN TESPİTİ	11
3.2. VERİNİN ANALİZİ	12
3.2.1. İşlem Talepleri	12
3.2.2. Çek Raporu	13

3.2.3. Risk Raporu	13
3.3. VERİNİN KEŞFİ	14
3.4. VERİ ÖN İŞLEME	19
3.4.1. Verinin Temizlenmesi	19
3.4.2. Veri Kategorilendirme	19
3.4.3. Veri Dönüşümleri	20
3.4.4. Anormal Gözlemlerin Tespiti	22
3.4.5. Öznitelik Seçimi	22
3.4.6. Verinin Dengelenmesi	23
3.4.7. Verinin Bölümlenmesi	24
3.5. MODELLEME	25
3.5.1. Sınıflandırma Metrikleri	25
3.5.2. Veri Setinin Sinir Ağları Algoritması ile Modellenmesi	27
3.5.3. Veri Setinin Karar Ağacı Algoritmaları ile Modellenmesi	29
3.5.3.1. C5.0 Algoritması	30
3.5.3.2. CHAID Algoritması	31
3.5.3.3. C&R Tree Algoritması	33
3.5.3.4. QUEST Algoritması	34
3.5.4. Veri Setinin Decision List Algoritması ile Modellenmesi	36
3.5.5. Veri Setinin Lojistik Regresyon Algoritması ile Modellenmesi	37
3.5.6. Veri Setinin Random Forest Algoritması ile Modellenmesi	39
3.5.7. Veri Setinin Random Trees Algoritması ile Modellenmesi	41
3.5.8. Veri Setinin XGBoost Algoritması ile Modellenmesi	42
3.6. DEĞERLENDİRME	44
SONUÇ	47
KAYNAKÇA	50

KISALTMALAR

AUC	Eđri Altında Kalan Alan (Area Under Curve)
BDDK	Bankacılık Dzenleme ve Denetleme Kurumu
GN	Gerçek Negatif
GP	Gerçek Pozitif
KKB	Kredi Kayıt Buros
ROC	İşlem Karakteristik Eğrisi (Receiver Operating Characteristic Curve)
TBB	Türkiye Bankalar Birliđi
XGBOOST	Ekstrem Gradyan Artırım (Extreme Gradient Boosting)
YN	Yanlış Negatif
YP	Yanlış Pozitif

TABLULAR LİSTESİ

Tablo 1: İşlem Taleplerinin Değişken Bilgileri	
s.12	
Tablo 2: Sınıflandırma Problemlerinde Kullanılan Metrikler	
s.26	
Tablo 3: Sinir Ağları Modeli Sonuçları (Tüzel)	
s.28	
Tablo 4: Sinir Ağları Modeli Sonuçları (Gerçek)	
s.28	
Tablo 5: C5.0 Modeli Sonuçları (Tüzel)	
s.30	
Tablo 6: C5.0 Modeli Sonuçları (Gerçek)	
s.30	
Tablo 7: CHAID Modeli Sonuçları (Tüzel)	
s.31	
Tablo 8: CHAID Modeli Sonuçları (Gerçek)	
s.32	
Tablo 9: C&R Tree Modeli Sonuçları (Tüzel)	
s.33	
Tablo 10: C&R Tree Modeli Sonuçları (Gerçek)	
s.33	
Tablo 11: QUEST Modeli Sonuçları (Tüzel)	s.34
Tablo 12: QUEST Modeli Sonuçları (Gerçek)	
s.35	
Tablo 13: Decision List Modeli Sonuçları (Tüzel)	
s.36	
Tablo 14: Decision List Modeli Sonuçları (Gerçek)	s.36

Tablo 15: Lojistik Regresyon Modeli Sonuçları (Tüzel)	
s.37	
Tablo 16: Lojistik Regresyon Modeli Sonuçları (Gerçek)	
s.38	
Tablo 17: Random Forest Modeli Sonuçları (Tüzel)	s.39
Tablo 18: Random Forest Modeli Sonuçları (Gerçek)	
s.39	
Tablo 19: Random Trees Modeli Sonuçları (Tüzel)	
s.40	
Tablo 20: Random Trees Modeli Sonuçları (Gerçek)	
s.41	
Tablo 21: XGBoost Modeli Sonuçları (Tüzel)	
s.42	
Tablo 22: XGBoost Modeli Sonuçları (Gerçek)	
s.42	
Tablo 23: Modellerin Test Verileri Üzerindeki Performans Sonuçları	
s.43	
Tablo 24: Algoritmaların Metrikler Özelinde Başarı Sıralamaları (Tüzel)	
s.44	
Tablo 25: Algoritmaların Metrikler Özelinde Başarı Sıralamaları (Gerçek)	s.44
Tablo 26: Algoritmaların Tüm Veri Seti Üzerindeki Performansı (Tüzel)	
s.47	
Tablo 27: Algoritmaların Tüm Veri Seti Üzerindeki Performansı (Gerçek)	s.47

ŞEKİLLER LİSTESİ

Şekil 1: Yeni Programlama Paradigması

s.7

Şekil 2: Makine Öğrenmesi Türleri

s.8

Şekil 3: Çeklerin Tutar Bazında Dağılımı

s.14

Şekil 4: Çeklerin Tutar ve Ödenme Durumu Bazında Dağılımı

s.15

Şekil 5: Çeklerin Banka Bazında Dağılımı

s.15

Şekil 6: En Yüksek Ödenmeme Oranına Sahip 10 Banka

s.16

Şekil 7: Çeklerin Keşide Edildiği İl Bazında Dağılımı

s.16

Şekil 8: En Yüksek Ödenmeme Oranına Sahip 10 İl

s.17

Şekil 9: Keşidecilerin Kişilik Nev'i ve Ödenme Durumu Dağılımı

s.17

Şekil 10: Çeklerin Vadesine Kalan Gün Sayısı Bazında Dağılımı

s.18

Şekil 11: Çeklerin Vadesine Kalan Gün Sayısı ve Ödenme Durumu Dağılımı

s.18

Şekil 12: Çeklerin Vadesinde Ödenme Durumunun Dağılımı

s.19

Şekil 13: Sayısal Veri Dönüşümü Uygulanan Değişkenlerin Dağılımı

s.20

Şekil 14: Anomaly Modelinin Sonuçları

s.22

Şekil 15: Verinin Dengelenmesi Sonrası Hedef Değişkenin Dağılımı (Tüzel)

s.23

Şekil 16: Verinin Dengelenmesi Sonrası Hedef Değişkenin Dağılımı (Gerçek)

s.24

Şekil 17: Veri Setinin Bölümlenmesi Sonrası Dağılım (Tüzel)

s.24

Şekil 18: Veri Setinin Bölümlenmesi Sonrası Dağılım (Gerçek)

s.25

Şekil 19: Karışıklık Matrisi

s.26

Şekil 20: Sinir Ağlarının Katman Yapısı

s.27

Şekil 21: Karar Ağlarının Yapısı

s.29

GİRİŞ

Bankaların topladıkları kaynakları, kaynak ihtiyacı duyan kişilere belirli bir vade ve tutarda kullandırılması işlemi kredi olarak tanımlanmaktadır. Bankalar kredi vererek müşterilerinden kredinin türü ve vadesine göre faiz ve komisyon geliri elde etmekteyken, bu işlem aynı zamanda bankalar için kredi kullandırılan müşterinin sözleşme kurallarına uymayarak alacağın tahsil edilememesi riskini ihtiva etmektedir (Öztürk, 2015).

Krediler farklı sınıflandırmalara tabi tutulabilmekte olup nakdi krediler grubunda bulunan kredi türlerinden birisi de iskonto ve iştirah kredileridir. Ticari işlemlerden doğan ve vadesi henüz gelmemiş kambiyo senedinin banka tarafından nakde dönüştürülmesi işlemine iskonto ve iştirah kredileri adı verilmektedir (Vurucu ve Arı, 2017).

Yüksek hacimde çek/senet portföyü bulunan, nakit akışı yoğun işletmelerin kısa vadeli finansman ihtiyaçlarını gidermek için başvurduğu kredi türlerinden birisi de iskonto ve iştirah kredileridir. BDDK tarafından yayımlanan 2021 yılı Haziran ayına ilişkin bankacılık sektör verileri incelendiğinde; bankaların iskontolu işlemlerden alacaklarının, toplam kredi hacmi içerisinde %4,14 oranında paya sahip olduğu görülmektedir.

Kredi müşterisinin sözleşme şartlarına uymaması, farklı bir deyişle bankanın krediden kaynaklı alacağını zamanında ve/veya tam olarak tahsil edememesi kredi riski olarak adlandırılmaktadır. Üstlenilen kredi riskinin düşürülmesi için kredinin tahsisi aşamasında müşterinin kredibilitesinin iyi ölçümü zorunludur (Yaslıdağ, 2007).

Kredi riskinin etkin yönetimi için tüzel kişilerin değerlendirilmesi sürecinde kullanılan yöntemlerin başında mali analiz gelmektedir. Müşterinin mali verilerinden yola çıkılarak taşımakta olduğu riskin bankalar tarafından objektif kriterlere bağlı olarak değerlendirilmesi süreci mali analiz olarak adlandırılmaktadır (Yazıcı, 2018).

Diğer sektörlerde olduğu gibi finans sektöründe de otomasyonla birlikte veri miktarı çok ciddi artış göstermiştir. Günümüz rekabet ortamında söz konusu verilerin işlenmesi suretiyle bilgiye, bilgiden kazanılmış bilgiye ve son olarak bilgeliğe dönüştürülmesi işletmeler için kritik önemi haizdir. Bu kapsamda finansal analiz

yöntemlerinde yapay zeka teknolojilerinin kullanımı da tercih edilen yöntemlerden biri haline gelmiştir (Çavdar ve Aydın, 2018).

Bilgisayar veya bilgisayar denetiminde bir makinenin, insanoğluna ait fonksiyonları yerine getirme yeteneği yapay zekâ olarak adlandırılmaktadır (Nabiyev, 2021).

Makine öğrenmesi ise doğrudan verilmiş yönlendirmeler olmaksızın, bilgisayarın öğrenmesine yardımcı olmak amacıyla matematiksel model kullanması olup aynı zamanda yapay zekanın da alt kümesidir. Başarılı bir makine öğrenmesi uygulaması ile önemli miktarda kaynak tasarrufu elde edilerek çalışanların daha önemli işlere odaklanması sağlanabilmektedir (Microsoft Azure, 2021).

İskonto ve iştirah kredilerini değerlendirmek üzere yapılmakta olan mali analiz sürecine alternatif olarak işleme konu kambiyo senedinin vade sonunda ödenip ödenmeyeceğinin makine öğrenmesi teknikleri kullanılarak tahmin edilebileceği ve bu sayede söz konusu işletmelerin mali analizi için harcanan kaynaktan tasarruf elde edilebileceği düşünülmektedir.

Bu araştırmada farklı bir karar destek sistemi ya da kredi analisti kullanılmadan, sadece keşidecinin geçmiş dönem çek ödeme performansı ile kredi ödeme geçmişi bilgileri üzerinden çekin vadesinde ödenip ödenmeyeceğine ilişkin tahminde bulunan bir yapı oluşturulması amaçlanmıştır.

İş Faktoring A.Ş. tarafından Kişisel Verilerin Korunması Kanunu ve diğer yasal düzenlemelere uygun bir biçimde anonim halde sağlanan, 01.08.2017 ile 05.10.2021 tarihleri arasında değerlendirilmiş çek iskonto/iştirah talepleri arasından; inceleme tarihi itibarıyla vadesi dolmuş, tutarı 1.000-100.000 TL, vadesi 10-365 arasında değişen 139.981 adet çek araştırmaya konu edilmiş ve çeklerin vadesinde ödenme durumunu tahmin eden modeller geliştirilmiştir.

BİRİNCİ BÖLÜM

TEMEL KAVRAMLAR

1.1. KREDİ KAVRAMI

Finansal sistem içerisinde aracılık görevi sunan kurumların en başında bankalar yer almaktadır. Piyasaların vazgeçilmez kurumları arasında olan bankaların, toplam finansal işlem hacminin yaklaşık yüzde doksanına sahip olduğu tahmin edilmektedir. Mevduat/katılım fonu kabulü, fon transferi, saklama hizmetleri, kredi/banka kartları vasıtalarının ihracı, çek ve diğer kambiyo senetlerinin iştirası gibi önemli rollere sahip olan bankalar tarafından icra edilen faaliyetlerin en önemlilerinden biri de kredi kullanıdır (Uzunođlu, 2020).

Bankaların özkaynaklarının yanında, mevduat sahiplerinden topladığı Türk Lirası ve döviz mevduatları ile dışarıdan sağlanan fonların belli bir süre sonunda geri alınmak üzere yasal ve dahili mevzuat gözetilerek ihtiyaç sahiplerine borç olarak verilmesi ya da bir taahhüdün ve/veya taahhütten doğacak borcun garanti edilmesi işlemi kredi olarak tanımlanmaktadır (Yaslıdağ, 2007).

Verilen krediler üzerinden elde edilen gelirlerin diğer bankacılık ürünlerinden sağlanan gelirlere kıyasla çok daha yüksek olması nedeniyle bankalar, ticari faaliyetlerini sürdürebilmek için sağlıklı bir kredi portföyüne sahip olmayı istemektedir. Karlı bir işlem olan kredilendirme süreci aynı zamanda riskli de bir işlemdir. Müşterilere verilen kredilerin vadelerinde geri ödenmemesi ya da donuklaşması, kâr amacı güden bir bankanın bu hedefe ulaşamamasının yanında, zamanla bankanın faaliyetlerin sonlandırılmasına varabilecek düzeyde risk ihtiva etmektedir (Öztürk, 2015).

Bankalar için kaynakların verimli kullanımının ne kadar önemli olduğu, dünya genelinde etkileri uzun süre devam eden 2008 yılı finansal krizinde bir kez daha hatırlanmıştır. Kredi müşterisinin banka ile yaptığı sözleşme gerekliliklerine tamamen ya da kısmen uymayarak yaratacağı zarar olasılığı, bankalar için kredi risk olarak tanımlanır. Kaynaklarını verimli kullanmak isteyen ve buna bağlı olarak taşıdığı kredi riskini düşürmek isteyen bankalar haliyle en yüksek kredibiliteye sahip müşterilerine kredi kullanıdır (Yazıcı, 2018).

1.2. İSKONTO VE İŞTİRA KREDİLERİ

Bankaların Türk Ticaret Kanunu'nda yer alan "tacir" statüsüne sahip gerçek ya da tüzel kişilere, işletme sermayesi ya da yatırım finansmanı amacıyla kullandığı kredilere "ticari kredi" adı verilmektedir. Ticari krediler, farklı kriterlere göre farklı sınıflandırmalara tabi tutulabilmekte olup nakdi krediler grubunda bulunan kredi türlerinden birisi de iskonto ve iştirah kredileridir.

Ticari işlemlerden doğan ve vadesi henüz gelmemiş kambiyo senedinin banka tarafından nakde dönüştürülmesi işlemine iskonto ve iştirah kredileri adı verilmektedir. Kambiyo senedi için işlemi yapan banka şubesi ile senedin ödeme yerinin aynı il belediye sınırları içerisinde olması durumunda işlem iskonto kredisi, diğer durumda ise iştirah kredisi olarak tanımlanmaktadır (Vurucu ve Arı, 2017).

İskonto ve iştirah kredilerinde işlem konusu kambiyo senedinin hukuki açıdan şekil şartlarını taşıması, taraflar arasında gerçek bir ticari ilişki bulunması ve senedin hatır/ortak senedi olmaması önem taşımaktadır (Öztürk, 2015).

İskonto ve iştirah işlemi gerçekleştirilen banka, senet borçlusunu ve varsa cirantaların ticari faaliyetlerine devam etme ve taahhütlerine uyum durumlarını kontrol ederek, araştırma sonuçlarının olumlu olması halinde işlemi kabul etmektedir. Borçlunun vadesinde senet bedelini ödemesi ile birlikte bedel kullanılan krediye mahsup edilerek kredi kapatılır. Bedelin vadesinde ödenmemesi durumunda ise banka senet borçlusunu ve/veya cirantalar için "Kambiyo Senetlerine Mahsus Haciz Yolu ile Takip" işlemi başlatacağı gibi kredi sözleşmesi kapsamında bedelin ödenmesi için kredi borçlusuna da başvurabilir (Vurucu ve Arı, 2017).

1.3. KREDİ TAHSİSİ

Finansal piyasalar içerisinde bilinen en yaygın ve eski risk türü kredi riski olarak tanımlanmaktadır. Sistem içerisinde bulunan tarafların hepsi bu risk türünden etkilenmektedir. Sistemik bir risk türü olan kredi riski ekonomik koşullardan etkilenmekte olup bu özelliği sebebiyle son derece konjonktürel bir risktir. Ekonomik daralma ve durgunluk dönemlerinde yükselme eğilimi göstermektedir (Alper, 2011).

Sağlam kredi riski yönetimi için kilit faktörün verimli kredi tahsisi olduğu bilinmektedir. Maruz kalınan kredi riskini kabul edilebilir çerçevede tutulması, kredi riski yönetiminin ana amacı konumundadır. Kredi borçlusunun vadesinde borcu ödeyemeyeceğini tahmin etmek, kredi riski yönetimi sanatı olarak adlandırılır (Akkaya, 2021).

Kredi riskinin minimize edilmesi için kredi tahsisi aşamasında müşterinin çok iyi analiz edilmesi gerekmektedir. Söz konusu analiz çalışması sırasında müşterinin sahip olduğu ekonomik güç, piyasadaki itibarı, ne kadar güvenilir olduğu, uygulanacak teminat yapısı gibi faktörlerin detaylı olarak değerlendirilmesi ve kredi tahsisinin bu analizler neticesinde gerçekleştirilmesi önem arz etmektedir (Öztürk, 2015).

Kredi tahsis çalışmaları 1950'li yıllardan bu yana gerçekleştirilmekte olup kredi analistlerinin en önemli görevleri, kredi talebi bulunan kişilerin kredi geçmişlerinde yola çıkarak uygun kredi tutarı ve koşulunu belirlemektir. Yıllar itibarıyla kredi hacminde yaşanan büyümeye paralel olarak kredi tahsis süreçlerinde daha isabetli sonuçlar üretilmesi amacıyla farklı yöntemler denenmiştir. Yapay zekâ teknolojilerinin kullanımına başlanması, bankalar başta olmak üzere finans kuruluşların kredilendirme sürecinde daha isabetli tahminlerde bulunmasına imkân sağlamıştır (Çavdar ve Aydın, 2018).

İKİNCİ BÖLÜM MAKİNE ÖĞRENMESİ

2.1. MAKİNE ÖĞRENMESİ KAVRAMLARI

Bilgisayar veya bilgisayar denetiminde bir makinenin; akıl yürütme, genelleme, anlam çıkarma, deneyimden öğrenme gibi insanoğluna ait fonksiyonları yerine getirme yeteneğine yapay zekâ adı verilmektedir (Nabiyev, 2021).

Dünya üzerindeki en kompleks makinenin insan beyni olduğunu söylemek mümkündür. Sayısal bir problemi belirli bir süre sonunda gerçekleştirebilen beyin, idrak etme işlemini ise çok kısa bir zamanda sonuçlandırmaktadır. Bilgisayarlar ise kompleks problemleri kısıtlı bir zaman içerisinde çözebilmelerinin yanında, idrak etme ve deneyimden elde edilen bilgiyi kullanma gibi hususlarda oldukça yetersiz durumdadır. Bu kapsamda yapay zekâ, insanoğlu ya da diğer bazı canlılarda bulunan idrak etmek, deneyimden elde edilen bilgiyi kullanma gibi zihinsel prosesleri bilgisayarlar yardımıyla icra etme yeteneği olarak ifade edilmektedir (Elmas, 2021).

Yapay zekâ alanında araştırma yapanlar arasında en önemli yere sahip kişilerden birisi olan Alan Turing, 1950 yılında yayımladığı bir makalesinde zekanın makineleştirilmesi olasılığını ele almıştır. Turing, farklı amaçlara hizmet eden farklı makineler yerine tüm amaçları gerçekleştirme potansiyeline sahip evrensel makinelerin önemi üzerinde durmuştur. Makinelerin düşünme ve zekâ yeteneklerinin test edilebildiği “Turing Testi” de yine Alan Turing tarafından ortaya konmuştur (Balta, 2021).

Makineler ile insanlar arasındaki ayrımın hızlı bir şekilde bulanıklaştığı döneme adım adım yaklaşmaktayız. İnsanoğlu günlük işleriyle uğraşırken, bilgisayarlar milyonlarca parametreden oluşan derin öğrenme modelleri oluşturma ve bu modeller insanların kavrayabileceğinin çok üzerinde bir kapsama sahip olmaktadır (Gürsakal, 2017).

İstatistiksel ve matematiksel teknikler kullanarak, veriler üzerinden çıkarımlarda bulunan ve böylelikle bilinmeyene dair tahminde bulunan yönteme ise makine öğrenmesi adı verilmektedir. Bu kapsamda makine öğrenmesi tekniklerinin temel amacının doğru tahminde bulunmak olduğu söylenebilir (Akay, 2020).

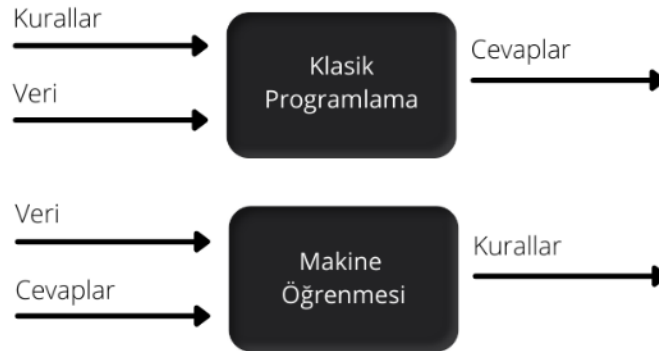
Matematik, yapay zekâ ve istatistik alanlarının birlikte çalışmasıyla makine öğrenmesi kavramı ne kadar zor görünse de mümkün hale gelebilmiştir (Nilsson, 2010).

Yapay zekâ ve makine öğrenmesi kavramları sıklıkla karıştırılmakla birlikte ikisi farklı şeyleri ifade etmektedir. Makinelerin programlanmadan, veri üzerinden öğrenimini sağlayan yapı makine öğrenimi olarak tanımlanırken, insanoğlunun düşünme kabiliyeti ve davranışlarını taklit etmek üzere akıllı sistem ya da makine oluşturmak gibi daha büyük bir kavramı ise yapay zekâ temsil etmektedir (Javatpoint, 2020).

Bir performans çıktısının optimize edilmesi amacıyla önceki deneyimler ya da eldeki verilerin kullanılması suretiyle yapılan yazılımlar ile makine öğrenmesi algoritmaları oluşmaktadır. Asıl amacın istatistikî metotlardaki gibi örneklemeler üzerinden çıkarımlar yapılması olması nedeniyle modellerin oluşturulması aşamasında istatistik teorisinden faydalanılmaktadır (Alpaydın, 2016).

Makine öğrenmesi modelleri programlanmak yerine eğitime tabi tutulmakta olup verilen örneklerden yola çıkarak istatistiksel örüntüleri tespit ederek verilen görevi otomatikleştirmeye yönelik kuralları ortaya koyarlar. Problemin çözümünde klasik yöntemlerin aksine girdi olarak veri ve cevaplar ortaya konularak yanıt olarak kurallar elde edilir (Chollet, 2019).

Şekil 1: Yeni Programlama Paradigması



Kaynak: Yazar tarafından derlenmiştir.

2.2. MAKİNE ÖĞRENMESİ TÜRLERİ

Denetimli (supervised) ve denetimsiz (unsupervised) olmak üzere makine öğrenmesinin temelde iki farklı türü bulunmaktadır. Verilerin önceden etiketlenerek öğrenme sürecinin gerçekleştirildiği yöntem denetimli öğrenme olarak tanımlanmaktayken, denetimsiz öğrenmede ise etiketlenmemiş veriler kullanılmaktadır. Öte yandan, denetimli öğrenmede modelin neyi kestireceği ve hedefi belliyken, denetimsiz öğrenmede hedef belirli değildir. Bu kapsamda denetimsiz öğrenme tekniğinin insanoğlunun öğrenme tekniğine daha yakın olduğu söylenmektedir (Gürsakal, 2017).

Şekil 2: Makine Öğrenmesi Türleri



Kaynak: Yazar tarafından derlenmiştir.

2.2.1. Denetimli Öğrenme

Verilerin büyük bir bölümünü içeren eğitim verileri üzerinden algoritma öğrenme sürecinin işletildiği makine öğrenmesi teknikleri denetimli öğrenme olarak tanımlanmaktadır. Denetimli öğrenme problemlerinde bir ya da birden fazla bağımsız değişken ile bağımlı değişkeni içeren bir eğitim verisi kullanılarak tahminde bulunmakta olup tahmin ve kestirim problemlerine odaklanılmaktadır. Eğitim verisi kullanılarak modelin eğitilmesinin ardından verinin kalan kısmından oluşan test verisi üzerindeki bağımsız değişkenler kullanılarak bağımlı değişkene ilişkin kestirimleri yapmak üzere modelin tahminde bulunması amaçlanmaktadır. Denetimli makine

öğrenmesi teknikleri regresyon ve sınıflandırma olmak üzere iki grupta ele alınmaktadır (Akay, 2020).

Denetimli öğrenmede temel olarak regresyon ve sınıflandırma kullanılsa da dizi oluşturma, sözdizimi ağaç tahmini, nesne tespiti, resim segmentasyonu gibi farklı çeşitleri de bulunmaktadır (Chollet, 2019).

2.2.2. Denetimsiz Öğrenme

Denetimsiz öğrenme yönteminde modelin neyi kestireceği yani hedefi belli olmayıp kümeleme ve birliktelik analizleri bu teknik ile gerçekleştirilmektedir (Gürsakal, 2017).

Veri bilimi sürecinin ekmeği, suyu gibi tanımlanan denetimsiz öğrenme teknikleri, genellikle denetimli öğrenme tekniklerinden önce test verisini anlamlandırmak üzere kullanılan bir süreçtir (Chollet, 2019).

Eğitim verilerinin etiketsiz olduğu öğrenme yöntemi denetimsiz öğrenme olarak adlandırılmakta olup bu yöntemde herhangi bir öğretene olmadan model eğitime çalışmaktadır. Bir internet sayfasının ziyaretçileri ile ilgili çok sayıda verinin olması durumunda kümeleme algoritması yardımıyla ziyaretçileri benzer özellikler gösterecek şekilde küçük gruplara ayırmak mümkündür. Böyle bir örnekte algoritma ziyaretçilerin hangi grup içinde olduğu bilmez ve gruplar arasındaki benzerlikler model tarafından tespit edilir. Hiyerarşik kümeleme algoritması yardımıyla her grup daha ufak alt kümelere de ayrılabilir (Geron, 2021).

Araştırma kapsamında ele alınmamakla birlikte denetimli ve denetimsiz öğrenme yöntemlerine ilave olarak literatürde denetimli ve denetimsiz öğrenme yöntemlerinin karışımından oluşan “Yarı Denetimli Öğrenme” ile “Pekiştirmeli Öğrenme” teknikleri de kullanılmaktadır (Gürsakal, 2017).

Pekiştirmeli öğrenme tekniği uzun yıllar araştırma alanı olarak kalmış ve pratikte oyunlar dışında bir başarısı olmamıştır (Chollet, 2019).

2.3. MAKİNE ÖĞRENMESİNİN TEMEL ZORLUKLARI

Makine öğrenmesi problemlerinde karşılaşılabilecek ana problemlerden ilki yetersiz sayıda eğitim verisine sahip olmaktır. Oldukça basit problemlerin çözümünde

bile binlerce gözlem verisine ihtiyaç duyulmakta olup karmaşık problemlerde bu sayı milyonlara ulaşabilir.

Diğer bir zorluk ise ana kütleyi temsil yeteneğine sahip olmayan eğitim verileri ile çalışmaktır. Genelleştirme yapılabilmesi için eğitim verilerinin yeni gözlemleri temsil edebilmesi çok önemli ancak görüldüğünden daha zor bir konudur.

Düşük kalitede verilere sahip olmak da makine öğrenmesi problemlerinde karşılaşılabilecek zorluklar arasındadır. Bu tip verilerle sistemin iyi bir performans düzeyine ulaşması ihtimali düşüktür. Veri bilimi problemlerinde araştırmacılar zamanlarının önemli bir kısmını bu sorunu gidermek yani verileri temizlemek için harcamaktadır.

Makine öğrenmesi projelerinin en kritik başarı faktörlerinden birisi de yeterli sayıda ve ilgili özniteliklerle, çok sayıda alakası bulunmayan özniteliğin yer almadığı bir öğrenme süreci gerçekleştirmektir. Öznitelik mühendisliği adı verilen bu süreçte mevcut veriler arasından en kullanışlı öznitelikleri belirleme, daha kullanışlı öznitelikler için mevcutları birleştirme ve yeni verilerle yeni öznitelikler oluşturma işlemleri yapılmaktadır.

Eğitim verilerine aşırı ya da yetersiz uydurma da makine öğrenmesi problemlerinde karşılaşılabilecek problemler arasındadır. Eğitim verilerinin yetersiz sayıda ya da gürültülü olması halinde model gürültü desenlerini takip ederek genelleştirilemeyecek örüntüler tespit edebilir. Öte yandan, model verilerinin örüntülerin tespit edilmesinde yetersiz kalması durumuyla da karşılaşılabılır. Oluşturulan modellerin ne çok kompleks ne de çok basit olması beklenmektedir (Geron, 2021).

ÜÇÜNCÜ BÖLÜM

UYGULAMA

Bu bölümde, araştırma kapsamında gerçekleştirilen analiz çalışmasına ilişkin bilgilere yer verilecektir. Araştırmacının amacı ve kapsamı hakkında bilgilerin aktarılmasının ardından modelleme çalışmasında kullanılan veri seti tanıtılacak ve ardından söz konusu veri setine makine öğrenmesi teknikleri uygulanacaktır.

Araştırmaya konu veriler, İş Faktoring A.Ş. (Şirket) tarafından Kişisel Verilerin Korunması Kanunu ve diğer yasal düzenlemelere uygun bir biçimde anonim halde sağlanmış ve sadece araştırma kapsamında kullanılmış olup uygulama için IBM SPSS Modeler programı kullanılmıştır.

3.1. İHTİYACIN TESPİTİ

Şirket, çek iskonto/iştira taleplerini kullanmakta olduğu karar destek sistemi ve/veya kredi/istihbarat analistleri tarafından yapılan değerlendirmeler neticesinde kabul etmekte ya da reddetmektedir. Bu çalışmada, çek iskonto/iştira kredilerinin vadesinde ödenme durumuna etki eden faktörlerin belirlenerek, söz konusu kredi türleri için gerçekleştirilen tahsis çalışmalarında sadece makine öğrenmesi tekniklerinin kullanımının mümkün olup olmadığı araştırılacaktır.

Şirket tarafından 01.08.2017 ile 05.10.2021 tarihleri arasında değerlendirilmiş çek iskonto/iştira talepleri arasından; inceleme tarihi itibarıyla vadesi dolmuş, tutarı 1.000-100.000 TL, vadesi 10-365 arasında değişen 139.981 adet çek araştırmaya konu edilmiştir. Tüm çeklerin vadesinde ödenme durumu bilinmektedir.

Şirket, tahsis çalışmaları sırasında KKB ve TBB Risk Merkezi üzerinden çek keşidecinin geçmiş dönem çek ödeme performansı ile kredi ödeme geçmişine ilişkin bilgileri temin etmektedir. Araştırmada farklı bir karar destek sistemi ya da kredi analisti kullanılmadan, sadece keşidecinin geçmiş dönem çek ödeme performansı ve kredi ödeme geçmişi bilgileri üzerinden çekin vadesinde ödenip ödenmeyeceğine ilişkin tahminde bulunan bir model oluşturulmasına çalışılacaktır.

3.2. VERİNİN ANALİZİ

Şirket tarafından veriler; “İşlem Talepleri”, “Çek Raporu”, “Risk Raporu” olmak üzere üç farklı veri kaynağı üzerinden sağlanmıştır. “İşlem Talepleri” tablosunda şirketin geçmişte değerlendirmeye konu ettiği çek iskontosu taleplerine konu çekle ilgili temel bilgiler, “Çek Raporu” tablosunda keşideci için Kredi Kayıt Bürosu üzerinden yapılmış Çek Raporu Sistemi sorgu sonuçları, “Risk Raporu” tablosunda ise yine keşideci için gerçekleştirilen Risk Raporu Sistemi sorgu sonuçları yer almaktadır. Çek Raporu Sistemi keşidecinin geçmiş dönem çek ödeme performansına ilişkin bilgiler verirken, Risk Raporu ise keşidecinin ticari ve varsa bireysel nitelikli kredi ödeme performansına yönelik bilgiler sağlamaktadır.

Veri seti içerisinde keşidecinin hiçbir kişisel verisi bulunmamaktadır. İşleme konu çek keşidecisinin T.C. Kimlik Numarası/Vergi Kimlik Numarası ile çek numarası ve sorgu tarihi bilgilerinin birleşiminden oluşan benzersiz ve maskelenmiş “uniqueID” değişkeni tüm tablolarda bulunmakta olup veriler bu değişken aracılığıyla ilişkilendirilmiştir.

3.2.1. İşlem Talepleri

Çeke ilişkin temel bilgiler ile modelleme çalışmasında hedef değişken olan, çekin vadesinde ödenme/ödenmeme durumunu gösteren “risk” değişkeni bu tabloda yer almaktadır. Tabloda yer alan verilere ve bunlara ilişkin açıklamalara aşağıda yer verilmiştir.

Tablo 1: İşlem Taleplerinin Değişken Bilgileri

DEĞİŞKEN ADI	AÇIKLAMA
uniqueID	Farklı veri kaynakları üzerinden gelen verilerin birleştirilmesi amacıyla oluşturulmuş; keşidecinin T.C. Kimlik Numarası/Vergi Kimlik Numarası, çek numarası ve sorgu tarihi bilgilerinin birleşimi.
cekTutar	Keşide edilen çek tutarı.
cekBankasi	Çek yaprağını basan banka kodu.
cekKesideYeri	Çekin keşide edildiği il bilgisi.

kesideciTur	Çeki keşide eden kişinin gerçek ya da tüzel kişi olma durumu. (T: Tüzel, G: Gerçek)
vadeyeKalanGun	Çekin iskonto edilme talebinin iletildiği gün ile çekin vade tarihi arasındaki farkın gün olarak karşılığı.
risk	Çekin vade tarihinde ödenip/ödenmediği bilgisi. (0: Ödendi, 1: Ödenmedi)

Kaynak: Yazar tarafından derlenmiştir.

3.2.2. Çek Raporu

Çek Raporu, çek muhataplarının çeki kabul etmesiyle aldığı riski göstermeyi amaçlayan ve bu kararın somut çıktılara dayanmasını sağlamaya yarayan bir raporlama sistemidir. TBB Risk Merkezi'ne üye bankalar tarafından günlük olarak bildirilen ve kural olarak rapor tarihinden önceki gün güncellenen çek bilgileri, üye banka ve finansal kuruluşlar tarafından bildirilen bilgilerin kullanılması suretiyle üretilmektedir.

Keşidecinin geçmiş çek ödeme performansının yanında bankacılık sistemine girilmiş ileri vadeli çek bilgileri ile ihtiyati tedbir kararı alınmış çeklere ilişkin bilgiler de sağlanmaktadır. Sorgu yapmaya yetkili kuruluşlar web servis teknolojisi kullanarak bir tür XML yapısında sorgu giriş değerlerini göndermekte ve yine aynı şekilde XML yapısında çıktılara ulaşmaktadır.

Çek raporu içerisinde yer alan çek tutarları, çekin ibraz edildiği yıla ait ortalama TCMB döviz satış kuru dikkate alınarak TL'ye çevrilmektedir. Tarih formatındaki verilerin modellemede bu şekilde kullanılamaması sebebiyle talep tarihi ile aralarındaki farkın gün olarak karşılıkları alınmıştır.

3.2.3. Risk Raporu

Risk Raporu, gerçek ve tüzel kişilerin bankalardan kullanmış oldukları kredi ürünlerine ilişkin geçmiş ödeme performansını göstermek üzere bankalar ve finansal kuruluşların düzenli olarak bildirdikleri ticari ve bireysel nitelikli kredi bilgileri kullanılmak suretiyle üretilmektedir. Finans kuruluşlarının karşı tarafın kredibilitesi konusunda doğru ve kapsamlı bilgiler edinmesini mümkün kılmaktadır. Risk raporu sadece negatif durumları değil, olumlu bilgileri de gösteren bir rapor olup bu kapsamda

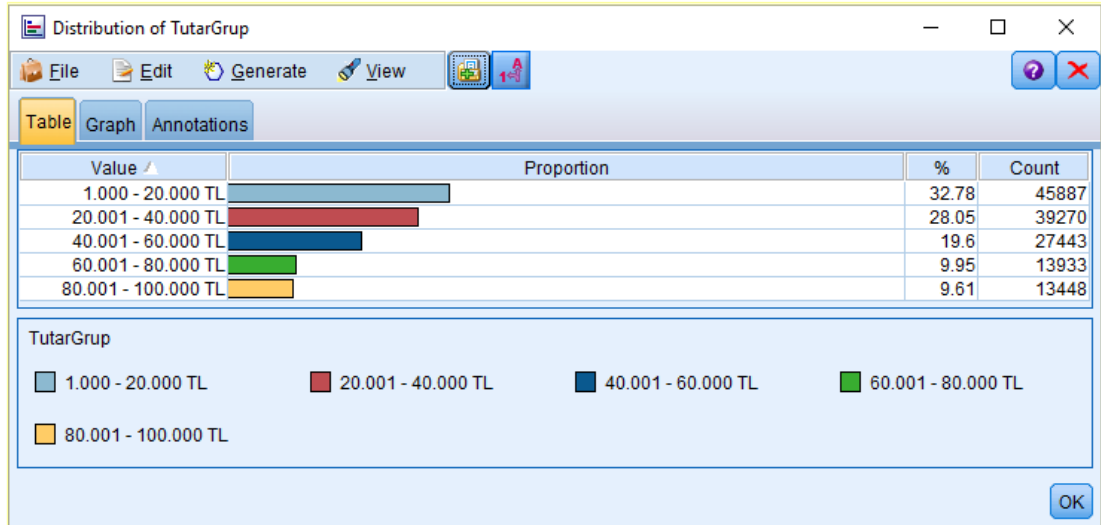
sadece gecikmiş ödemeler ya da takibe aktarılan krediler değil, zamanında ödenen kredi bilgileri de paylaşılmaktadır.

Sorgu yapmaya yetkili kuruluşlar web servis teknolojisi kullanarak bir tür XML yapısında sorgu giriş değerlerini göndermekte ve yine aynı şekilde XML yapısında aşağıdaki çıktılara ulaşmaktadır. Tarih formatındaki verilerin modellemeye bu şekilde kullanılamaması sebebiyle talep tarihi ile aralarındaki farkın gün olarak karşılığı alınmıştır.

3.3. VERİNİN KEŞFİ

Toplam 139.981 adet çekin tutar bazında dağılımı incelendiğinde; çeklerin çok büyük bir bölümünün (%60,83) 40.000 TL ve altındaki çeklerden oluştuğu görülmektedir. En düşük pay ise %9,61 ile 80.001-100.000 TL arasındaki çeklere aittir.

Şekil 3: Çeklerin Tutar Bazında Dağılımı



Kaynak: Yazar tarafından derlenmiştir.

Çeklerin tutar grupları içerisinde ödenme durumları incelendiğinde %7,12 ile en çok 60.001-80.000 TL aralığındaki çeklerin karşılıksız kaldığı, oransal olarak en çok 1.000-20.000 TL aralığındaki çeklerin ödendiği sonucuna ulaşılmaktadır.

Şekil 4: Çeklerin Tutar ve Ödenme Durumu Bazında Dağılımı

	TutarGrup	0	1	Oran (%)
1	60.001 - 80.000 TL	12941	992	7.120
2	40.001 - 60.000 TL	25747	1696	6.180
3	80.001 - 100.000 TL	12639	809	6.016
4	20.001 - 40.000 TL	37221	2049	5.218
5	1.000 - 20.000 TL	44636	1251	2.726

Kaynak: Yazar tarafından derlenmiştir.

Çek yaprağını düzenleyen banka bazında dağılım incelendiğinde; veri seti içerisinde Halk Bankası, Yapı Kredi, Ziraat Bankası, Garanti Bankası ve İş Bankası çeklerinin %52'lik paya sahip olduğu görülmektedir.

Şekil 5: Çeklerin Banka Bazında Dağılımı

Value	Proportion	%	Count
Diğer		13.45	18821
Türkiye Halk Bankası A.Ş.		12.61	17650
Yapı ve Kredi Bankası A.Ş.		10.23	14315
Türkiye Cumhuriyeti Ziraat Bankası A.Ş.		9.97	13963
Türkiye Garanti Bankası A.Ş.		9.82	13747
Türkiye İş Bankası A.Ş.		9.45	13231
Akbank T.A.Ş.		8.32	11641
Türkiye Vakıflar Bankası T.A.O.		7.38	10327
Finansbank A.Ş.		6.84	9569
Denizbank A.Ş.		6.19	8662
Türk Ekonomi Bankası A.Ş.		5.75	8055

Kaynak: Yazar tarafından derlenmiştir.

Değerlendirilen çekler arasında çek bankasına göre karşılıksız kalma durumu incelendiğinde en yüksek paya sahip 10 banka şu şekilde gerçekleşmiştir.

Şekil 6: En Yüksek Ödenmeme Oranına Sahip 10 Banka

	Banka	0	1	Oran (%)
1	Turkish Bank A.Ş.	12	5	29.412
2	Alternatifbank A.Ş.	362	44	10.837
3	Burganbank A.Ş.	344	33	8.753
4	Anadolubank A.Ş.	1376	119	7.960
5	Fiba Bank A.Ş.	1269	103	7.507
6	Turkland Bank A.Ş.	116	9	7.200
7	Türkiye Vakıflar Bankası T.A.O.	9696	631	6.110
8	Oyak Bank A.Ş. - ING BANK	1177	75	5.990
9	Türk Ekonomi Bankası A.Ş.	7579	476	5.909
10	Şekerbank T.A.Ş.	1902	113	5.608

Kaynak: Yazar tarafından derlenmiştir.

Çeklerin keşide edildikleri yere göre dağılımları incelendiğinde; İstanbul ilinin %33 ile hâkim paya sahip olduğu, İstanbul, Ankara ve İzmir keşideli çeklerin toplam çeklerin %52'sini oluşturduğu belirlenmiştir.

Şekil 7: Çeklerin Keşide Edildiği İl Bazında Dağılımı

Value	Proportion	%	Count
İstanbul		33.36	46699
Diğer		33.36	46695
Ankara		12.16	17026
İzmir		6.25	8753
Bursa		5.54	7754
Kocaeli		3.64	5093
Antalya		3.2	4482
Adana		2.49	3479

Kaynak: Yazar tarafından derlenmiştir.

İl bazında karşılıksız kalma durumu incelendiğinde en yüksek paya sahip 10 ilin aşağıdaki illerimiz olduğu sonucuna ulaşılmıştır.

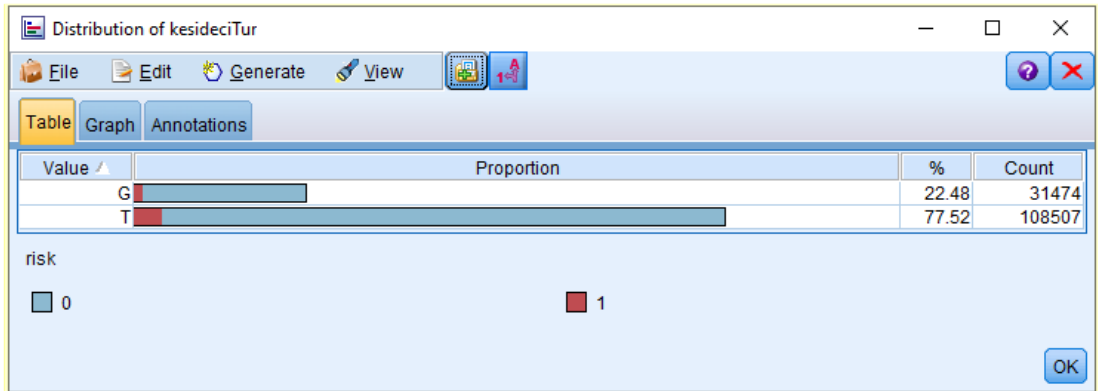
Şekil 8: En Yüksek Ödenmeme Oranına Sahip 10 İl

	il	0	1	Oran (%)
1	Mardin	230	24	9.449
2	Siirt	117	11	8.594
3	Osmaniye	372	34	8.374
4	Bingöl	108	9	7.692
5	Çanakkale	771	63	7.554
6	Batman	382	31	7.506
7	Karabük	219	16	6.809
8	Sakarya	1943	141	6.766
9	Hakkari	98	7	6.667
10	Şanlıurfa	796	55	6.463

Kaynak: Yazar tarafından derlenmiştir.

Çeklerin %77,5'inin tüzel, %22,5'inin ise gerçek kişi tacirler tarafından keşide edildiği görülmektedir. Gerçek ve tüzel kişi keşidecilerin çek ödememe oranında ise bir farklılaşmanın bulunmadığı gözlemlenmiştir.

Şekil 9: Keşidecilerin Kişilik Nev'i ve Ödenme Durumu Dağılımı

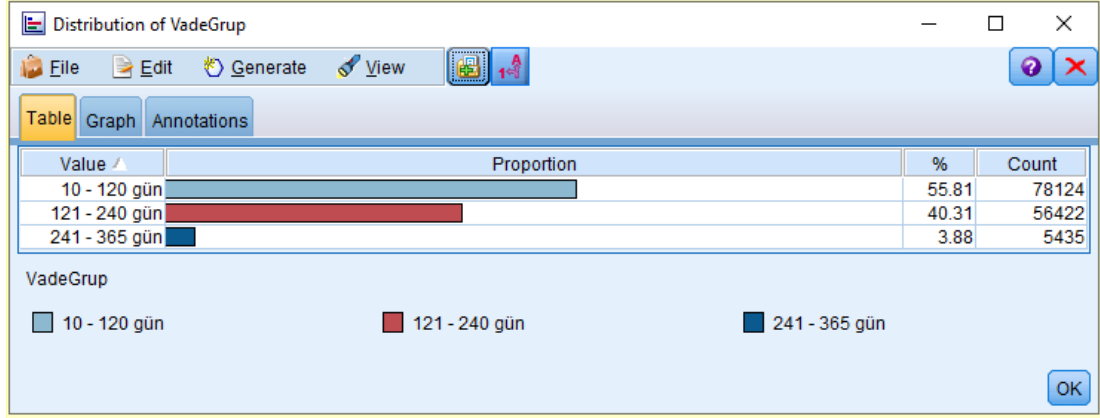


	kesideciTur	0	1	Oran (%)
1	T	103238	5269	4.856
2	G	29946	1528	4.855

Kaynak: Yazar tarafından derlenmiştir.

Çeklerin vadelerine kalan gün bazında dağılımları incelendiğinde; en yüksek payın %55,8 ile 10-120 gün arası vadeli çeklere ait olduğu görülmektedir.

Şekil 10: Çeklerin Vadesine Kalan Gün Sayısı Bazında Dağılımı



Kaynak: Yazar tarafından derlenmiştir.

Vade gruplarına göre çeklerin karşılıksız kalma oranları incelendiğinde; vadenin arttıkça ödenmeme olasılığının da artış gösterdiği sonucuna ulaşılmaktadır.

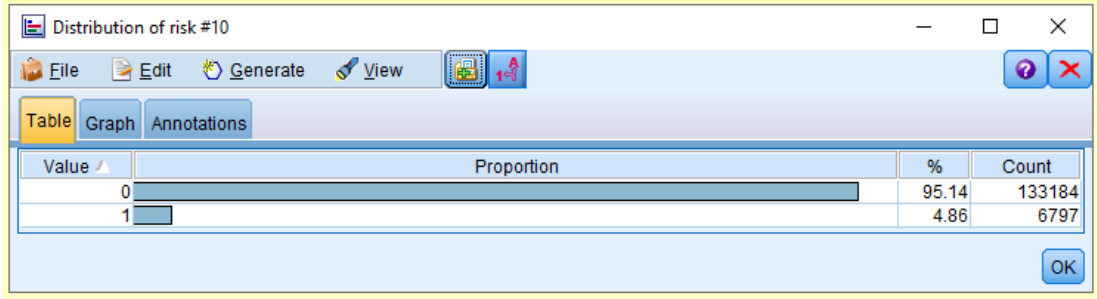
Şekil 11: Çeklerin Vadesine Kalan Gün Sayısı ve Ödenme Durumu Dağılımı

	VadeGrup	0	1	Oran (%)
1	241 - 365 gün	5068	367	6.753
2	121 - 240 gün	52911	3511	6.223
3	10 - 120 gün	75205	2919	3.736

Kaynak: Yazar tarafından derlenmiştir.

Çeklerin vadesinde ödenme durumunu gösteren ve aynı zamanda araştırmanın hedef değişkeni konumunda olan “risk” değişkeninin dağılımı incelendiğinde ise çeklerin %95,14 oranında vadesinde ödendiği anlaşılmaktadır.

Şekil 12: Çeklerin Vadesinde Ödenme Durumunun Dağılımı



Kaynak: Yazar tarafından derlenmiştir.

3.4. VERİ ÖN İŞLEME

Veri bilimi projelerinin en çok emek ve zaman gerektiren bölümlerinden olan veri ön işleme aşaması aynı zamanda projenin en önemli aşaması olarak tanımlanmaktadır. Veri temizleme, dengeleme, dönüştürme, kategorileştirme, bölümlenme, anormal gözlem tespiti gibi süreçler bu aşamada gerçekleştirilmektedir.

3.4.1. Verinin Temizlenmesi

Veri madenciliği sürecinde kullanılan verinin temiz olması, başarılı bir tahmin modeli oluşturmanın kritik faktörlerinin başında gelmektedir. Bu nedenle modelleme öncesinde veri seti içinde yer alan kayıp ve geçersiz değerlerin minimum seviyeye indirilmesi gerekmektedir. Veri seti içerisinde bu nitelikte gözlemler olup olmadığı detaylı olarak incelenmiş ve hatalı değer içerdiği sonucuna ulaşılan gözlemler (Örneğin; negatif kredi bakiyesi gibi) veri seti içerisinde çıkarılmıştır. Daha önce 139.981 olan gözlem sayısı, verinin temizlenmesi aşamasından sonra 137.992 olmuştur.

3.4.2. Veri Kategorileştirme

Yüksek performans değerine sahip bir modelin oluşturulması için çarpık dağılıma sahip verilere veri dengeleme yöntemlerinin uygulanması gerekmektedir. Veri seti içerisinde bulunan sürekli değişkenler için kategorik değişkenler oluşturmak, kullanılabilir veri dengeleme yöntemlerinden birisidir. Veri kategorileştirme işleminin avantajlarının yanında dezavantajları da bulunmaktadır. Kategorileştirme

dağılım ile ilgili sorunları çözerken kategorik yapıya geçen veri taşıdığı bilgiyi tam olarak aktaramayabilir.

Veri seti içerisinde yer alan sürekli değişkenlerin histogram grafikleri yardımıyla dağılımları incelenerek kategorileştirmeye uygun olanlar için kategorik değişkenler üretilmiştir.

3.4.3. Veri Dönüşümleri

Çarpık dağılıma sahip verileri dengelemekte kullanılabilecek yöntemlerden diğeri de sayısal veri dönüşümleridir. Kategorileştirmeye uygun olmayan değişkenlerin dağılımları incelenerek, en uygun dönüşümün logaritmik dönüşüm olduğu sonucuna ulaşılmıştır. Logaritmik dönüşümün ardından değişken dağılımlarının normal dağılıma uyma hususunda gelişme kat ettiği gözlemlenmiştir. Dönüşüm uygulanan değişkenler ile bunların önceki ve sonraki dağılımlarına aşağıda yer verilmiştir.

Şekil 13: Sayısal Veri Dönüşümü Uygulanan Değişkenlerin Dağılımı

Field	Selected Transfor...	Current Distribut...	Inverse	LogN	Log10	Exponential	Square Root
cekTutar							
vadeyeKalanGun							
toplamCekSayisi							
toplamOdenmisCekSayisi							
toplamOdenmisCekTutari							
acikCekAdedi							
toplamVadeliCekAdedi							
sistemeDonmeyenCekAdedi							
sonAyOdenenCekTutari							
son3AyOdenenCekTutari							
son12AyOdenenCekSayisi							
son12AyOdenenCekTutari							

gunceYilKesideEdilenEnDusukTutar		8817.644 (21182.493)	0.000 (0.002)	8.577 (0.947)	3.725 (0.411)	1211838146676153...	78.445 (48.638)
gunceYilKesideEdilenEnYuksektutar		347823.437 (15414...	0.000 (0.000)	11.700 (1.347)	5.081 (0.585)	3408297705213388...	425.252 (408.638)
gunceYilKesideEdilenortalamaCekTutari		48148.792 (62083.5...	0.000 (0.000)	10.451 (0.845)	4.539 (0.367)	2239411168193500...	194.795 (101.014)
VadeliToplamCekTutari1Ay		4062506.058 (4126...	0.000 (0.000)	12.805 (1.848)	5.474 (0.802)	1.000 (0.000)	896.980 (1819.581)
VadeliToplamCekAdedi3Ay		63.545 (295.828)	0.136 (0.211)	2.878 (1.410)	1.250 (0.612)	7110703624727046...	5.482 (5.788)
VadeliToplamCekTutari3Ay		7707988.593 (6977...	0.000 (0.000)	13.373 (1.852)	5.808 (0.822)	1.000 (0.000)	1326.565 (2438.905)
VadeliToplamCekAdedi6Ay		77.153 (323.045)	0.112 (0.191)	3.115 (1.415)	1.353 (0.615)	2908692784646421...	6.212 (6.210)
VadeliToplamCekTutari6Ay		8887076.760 (7247...	0.000 (0.000)	13.644 (1.889)	5.925 (0.821)	1.000 (0.000)	1512.439 (2568.979)
VadeliToplamCekAdedi12Ay		80.184 (328.342)	0.109 (0.188)	3.156 (1.420)	1.371 (0.617)	3362864529025632...	6.352 (6.311)
VadeliToplamCekTutari12Ay		9176219.889 (7278...	0.000 (0.000)	13.697 (1.893)	5.949 (0.822)	1.000 (0.000)	1554.166 (2600.161)
VadeliToplamCekTutari12AyUstu		40728.325 (103388...	0.000 (0.000)	11.986 (1.782)	5.205 (0.774)	1.000 (0.000)	26.322 (200.089)
son1AyTahsilEdilenCekTutar		1088338.935 (1224...	0.000 (0.000)	12.167 (1.826)	5.258 (0.791)	1101543281951045...	300.701 (998.962)
son3AyTahsilEdilenCekTutar		3277569.599 (3673...	0.000 (0.000)	12.770 (1.961)	5.546 (0.852)	6266409687419604...	552.207 (1724.137)
son3AyKarsiliksizCekTutar		16279.118 (137925...	0.000 (0.000)	11.071 (1.325)	4.808 (0.576)	2220210478858138...	31.783 (123.588)
son12AyTahsilEdilenCekTutar		12425767.280 (140...	0.000 (0.000)	13.544 (2.193)	5.882 (0.952)	6455016422379063...	1121.374 (3341.911)
son12AyKarsiliksizCekTutar		60533.826 (439267...	0.000 (0.001)	11.429 (1.468)	4.963 (0.638)	2533349779215617...	84.542 (231.056)
son12AyOdenmiseDonenCekTutar		21022.717 (146634...	0.000 (0.000)	10.950 (1.421)	4.755 (0.617)	2683872770714472...	40.184 (139.313)
tkGecikmisBakiyeToplami		22382.826 (943954...	0.033 (0.133)	7.426 (3.164)	3.225 (1.374)	1753368482629032...	20.745 (148.164)
tkToplamLimit		62641776.524 (357...	0.000 (0.013)	15.316 (2.188)	6.652 (0.950)	2952285065500734...	4028.919 (6812.483)
tkToplamRisk		28493503.604 (191...	0.000 (0.014)	14.497 (2.237)	6.296 (0.971)	6527722486901434...	2683.577 (4614.333)
tkToplamNakdiLimit		56800575.176 (482...	0.000 (0.016)	15.028 (2.264)	6.527 (0.983)	3680082727623641...	3528.090 (6659.841)
tkToplamNakdiRisk		20513677.891 (131...	0.002 (0.041)	14.049 (2.544)	6.101 (1.105)	3955729051765164...	2207.485 (3954.845)
tkToplamGayriNakdiLimit		26514741.514 (210...	0.000 (0.005)	14.008 (2.299)	6.084 (0.998)	2884278988562208...	2349.709 (4581.895)
tkToplamGayriNakdiRisk		7953806.507 (8762...	0.000 (0.007)	12.874 (2.209)	5.591 (0.959)	3814695664740836...	1223.961 (2540.821)
bkToplamLimit		35378.218 (124937...	0.000 (0.014)	11.267 (1.368)	4.853 (0.594)	9439458469696448...	75.734 (172.171)
bkToplamRisk		21799.234 (88302.2...	0.001 (0.029)	10.554 (1.697)	4.584 (0.737)	1871673375797969...	55.350 (136.879)
bkGeciktirdigiBakiyeToplami		326.870 (2177.649)	0.008 (0.060)	7.116 (1.669)	3.091 (0.725)	3307528348969749...	4.609 (17.482)

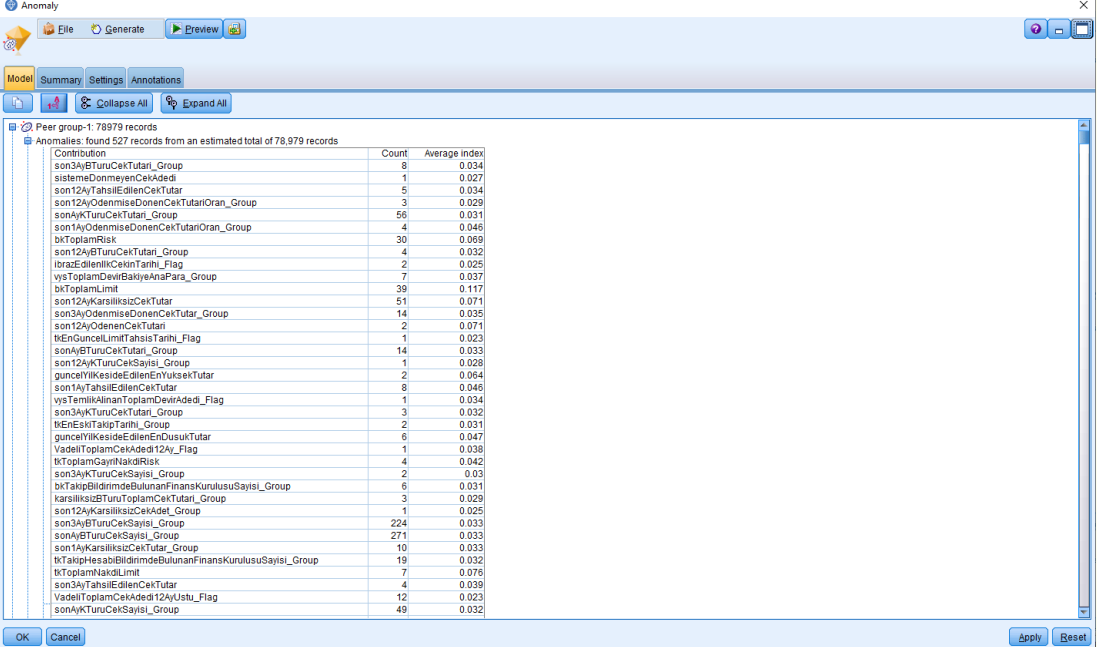
Kaynak: Yazar tarafından derlenmiştir.

3.4.4. Anormal Gözlemlerin Tespiti

Veri bilimi projelerinde veri seti içinde yer alan anormal gözlemlerin tespit edilerek gerekli aksiyonların alınması gerekmektedir. Büyük veri setlerinde anormal değerlerin modelleme üzerinde etkisinin sınırlı olduğu düşünülmeyle birlikte daha az sayıda gözlem içeren veri setlerinde ise yüksek etki gözlemlenmektedir.

Oluşturulacak modelde herhangi bir performans sorunu yaşanmaması adına verinin genel karakteristiğine uymayan gözlemlerin belirlenmesi için Two-Step kümeleme algoritması kullanılmıştır. Anomaly modelinin örnek bir grup için çıktısına aşağıda yer verilmiştir.

Şekil 14: Anomaly Modelinin Sonuçları



Contribution	Count	Average Index
son3AyBTuruCekTutar_Group	8	0.034
sistemeDonmeyenCekAdedi	1	0.027
son12AyTahsilEdilenCekTutar	5	0.034
son12AyOdenmiseDonenCekTutarOran_Group	3	0.029
sonAyKTuruCekTutar_Group	56	0.031
son1AyOdenmiseDonenCekTutarOran_Group	4	0.046
bkToplamRisk	30	0.069
son12AyBTuruCekTutar_Group	4	0.032
ibradzEdilenCekinTarih_Flag	2	0.025
vySToplamDevirBakijeAnaPara_Group	7	0.037
bkToplamLimit	39	0.117
son12AyKarsiliksizCekTutar	51	0.071
son3AyOdenmiseDonenCekTutar_Group	14	0.035
son12AyOdenenCekTutar	2	0.071
tkEnGuncelLimitTahsisTarih_Flag	1	0.023
sonAyBTuruCekTutar_Group	14	0.033
son12AyKTuruCekSayisi_Group	1	0.028
guncelYikeseEdilenEnYuksektutar	2	0.064
son1AyTahsilEdilenCekTutar	8	0.046
vySTemlikAlinanToplamDevirAdedi_Flag	1	0.034
son3AyKTuruCekTutar_Group	3	0.032
tkEnEskitakipTarih_Group	2	0.031
guncelYikeseEdilenEndusuktutar	6	0.047
VadeliToplamCekAdedi12Ay_Flag	1	0.038
tkToplamSartliRisk	4	0.042
son3AyKTuruCekSayisi_Group	2	0.03
bkTakipBildirimdeBulunanFinansKurulusuSayisi_Group	6	0.031
karsiliksizBTuruToplamCekTutar_Group	3	0.029
son12AyKarsiliksizCekAdet_Group	1	0.025
son3AyBTuruCekSayisi_Group	224	0.033
sonAyBTuruCekSayisi_Group	271	0.033
son1AyKarsiliksizCekTutar_Group	10	0.033
tkTakipHesabiBildirimdeBulunanFinansKurulusuSayisi_Group	19	0.032
tkToplamNispetLimit	7	0.076
son3AyTahsilEdilenCekTutar	4	0.039
VadeliToplamCekAdedi12AyUstü_Flag	12	0.023
sonAyKTuruCekSayisi_Group	49	0.032

Kaynak: Yazar tarafından derlenmiştir.

Anormal değerlerin veri setinden çıkarılmasının ardından daha önce 137.992 olan gözlem sayısı 136.613 olarak gerçekleşmiştir.

3.4.5. Öznitelik Seçimi

Çoğu algoritma çok sayıda değişken ile iyi performans gösterse de oluşan çıktılar karmaşık ve yorumlaması zor olabilmektedir. Bu nedenle modelleme çalışması öncesinde değişkenlerin incelenerek yüksek oranda aynı gözlemden oluşan ya da

kayıp değere sahip özniteliklerin veri setinden çıkarılmasında fayda görülmektedir. Veri seti içerisinde bu nitelikte bir öznitelik olup olmadığı sorgulanmış ve hedef değişken ile ilişkisi bulunmadığı tespit edilen değişkenler çıkarılmıştır.

Tüm bu çalışmalar neticesinde toplam 157 adet bağımsız, 1 adet hedef değişken ile 136.613 gözlemden oluşan veri seti elde edilmiştir.

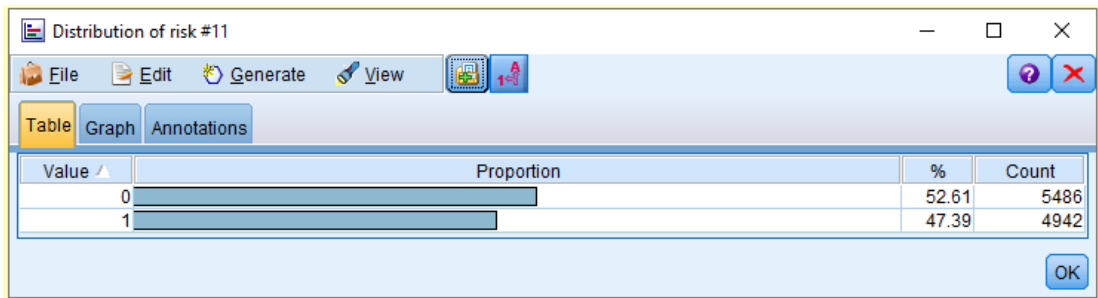
3.4.6. Verinin Dengelenmesi

Hedef değişkenin dağılımının çarpık olması durumunda öngörülse modellerde problem yaşanması riski bulunmakta olup veri setimiz içerisinde ödenen çeklerin hâkim payı nedeniyle çarpıklık söz konusudur.

Problemin giderilmesi için az gözlem sayısına sahip kategoriye göre azaltma (reduce) işlemi gerçekleştirilerek veri seti dengelenmiştir. Ödenen çeklerin çok fazla olması nedeniyle dengeli bir eğitim süreci gerçekleşmesi için yarı yarıya dağılım yerine ödenen çeklerin bir miktar daha fazla olduğu bir dağılım tercih edilmiştir.

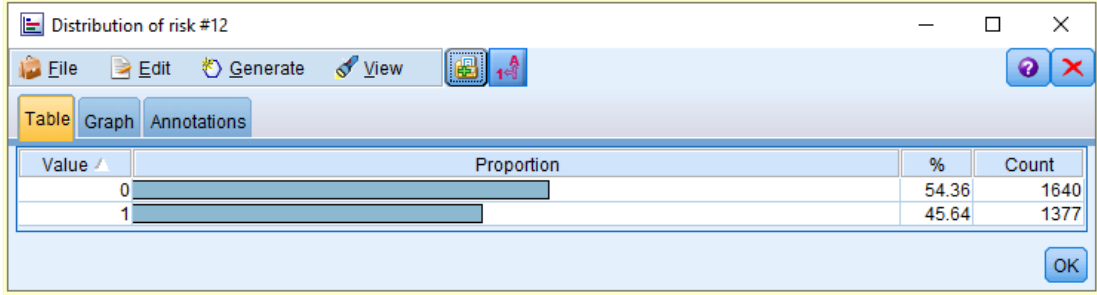
Veri seti içerisinde yer alan bazı değişkenlerin (Örneğin; bireysel kredi bakiyesi) sadece gerçek kişiler için sonuç vermesi nedeniyle gerçek ve tüzel kişi keşideciler için ayrı modeller oluşturulacak olup bu sebeple veri setimiz gerçek ve tüzel kişiler olmak üzere ikiye ayrıştırılmış, veri dengeleme ve veri bölümlenme süreçleri söz konusu veri setleri için ayrı ayrı gerçekleştirilmiştir.

Şekil 15: Verinin Dengelenmesi Sonrası Hedef Değişkenin Dağılımı (Tüzel)



Kaynak: Yazar tarafından derlenmiştir.

Şekil 16: Verinin Dengelenmesi Sonrası Hedef Değişkenin Dağılımı (Gerçek)



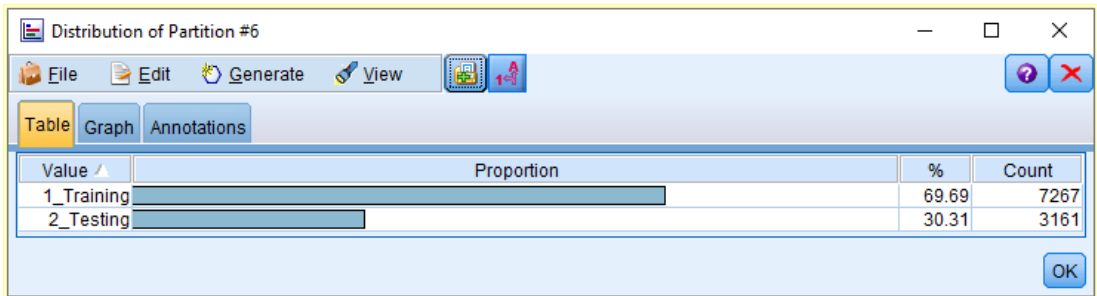
Kaynak: Yazar tarafından derlenmiştir.

3.4.7. Verinin Bölümlemesi

Veri bilimi çalışmalarında tüm veri setinin modelleme çalışmasında kullanılmaması ve verilerin bir bölümünün test çalışması sırasında kullanılması gerekmektedir. Bu sayede oluşan model daha önce hiç görmediği bir veri seti üzerinde tahminlerde bulunarak bize performansı hakkında bilgi vermektedir.

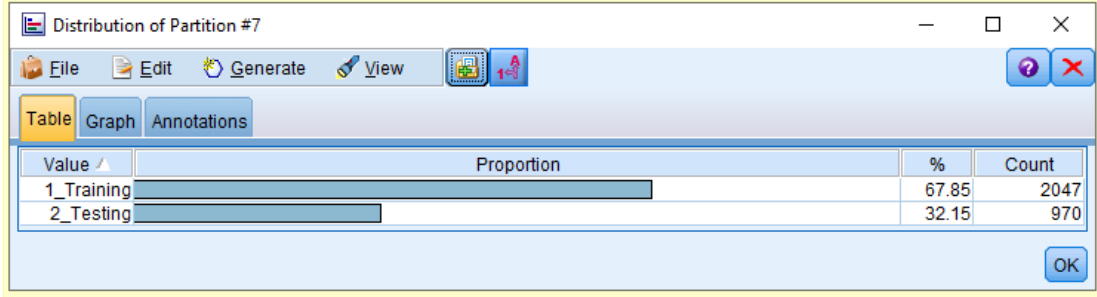
Her iki veri seti içerisinde yer alan gözlemlerin yaklaşık %70'i eğitim, %30'u test verisi olarak ayrıştırılmış ve modellerin eğitim aşamasında sadece eğitim verileri kullanılmıştır.

Şekil 17: Veri Setinin Bölümlemesi Sonrası Dağılım (Tüzel)



Kaynak: Yazar tarafından derlenmiştir.

Şekil 18: Veri Setinin Bölümlenmesi Sonrası Dağılım (Gerçek)



Value	Proportion	%	Count
1_Training		67.85	2047
2_Testing		32.15	970

Kaynak: Yazar tarafından derlenmiştir.

3.5. MODELLEME

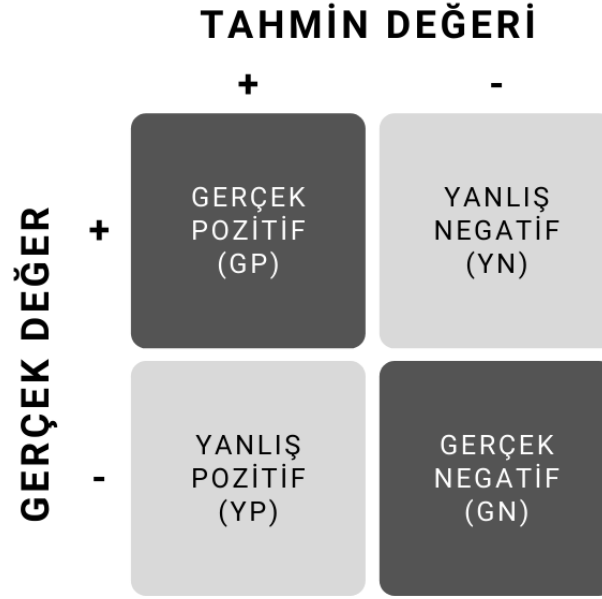
Daha önce belirtildiği üzere; veri seti içerisinde bulunan Risk Raporu verileri arasındaki bazı değişkenlerin (Örneğin; bireysel kredi bakiyesi) sadece gerçek kişiler için sonuç dönmeye sebebiyle gerçek ve tüzel kişi keşideciler için ayrı modeller hazırlanmasının daha uygun olacağı sonucuna ulaşılmış ve bu kapsamda söz konusu veri setleri ayrıştırılmıştır.

Bu bölümde kullanılan sınıflandırma problemlerinde kullanılan performans metriklerine, araştırmada kullanılan makine öğrenmesi algoritmalarına ilişkin genel esaslara ve bu algoritmalar kullanılarak oluşturulan modellerin performans sonuçlarına yer verilmiştir.

3.5.1. Sınıflandırma Metrikleri

Sınıflandırma problemlerinde modelin performansı hakkında sonucunda bilgi sahibi olmak için farklı metrikler bulunmakta olup ikili sınıflandırma problemlerinde bunun için karışıklık matrisi (confusion matrix) kullanılmaktadır (Amidi, 2018).

Şekil 19: Karışıklık Matrisi



Kaynak: Yazar tarafından derlenmiştir.

Modellerin performans ölçümünde kullanılan metriklere ve bunlara ilişkin açıklamalara aşağıda yer verilmiştir.

Tablo 2: Sınıflandırma Problemlerinde Kullanılan Metrikler

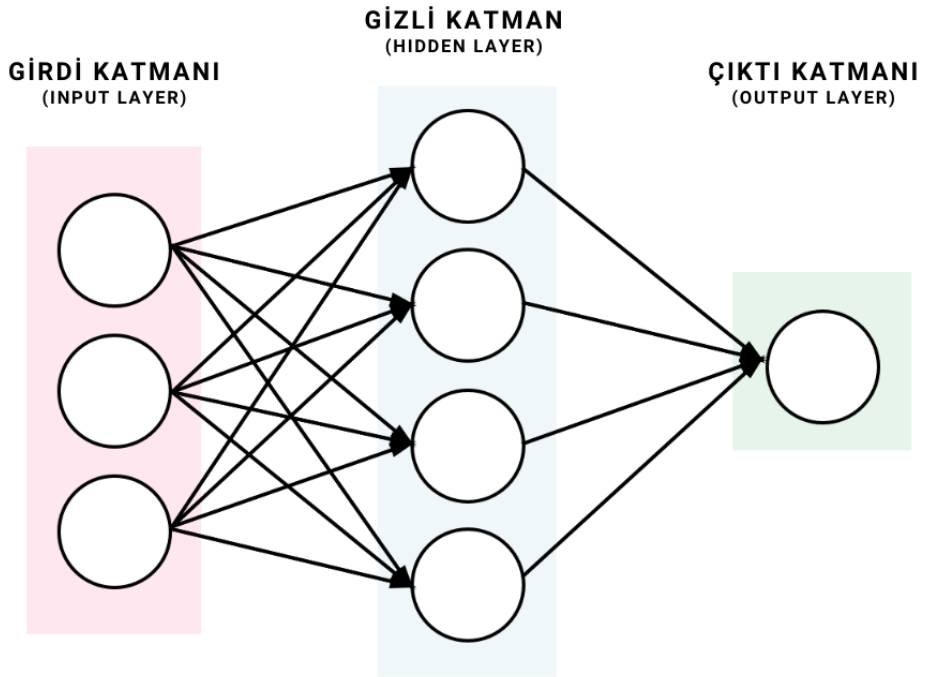
Metrik	Açıklama	Formül
Doğruluk (<i>Accuracy</i>)	Toplam tahmin içinde doğru tahminlerin payı.	$\frac{GP + GN}{GP + YN + YP + GN}$
Kesinlik (<i>Precision</i>)	Doğru tahminlerin kesinliği.	$\frac{GP}{GP + YP}$
Geri Çağırma (<i>Recall</i>)	Gerçek pozitiflerin oranı.	$\frac{GP}{GP + YN}$
Özgüllük (<i>Specificity</i>)	Gerçek negatiflerin oranı.	$\frac{GN}{YP + GN}$
F1 Skoru (<i>F1 Score</i>)	Dengesiz sınıflarda kullanılan hibrit bir metrik.	$\frac{2GP}{2GP + YP + YN}$
Eğri Altında Kalan Alan (<i>AUC</i>)	İşlem karakteristik eğrisi (ROC) altında kalan alan olup modelin ayırt ediciliğini temsil eder.	

Kaynak: Amidi, 2018

3.5.2. Veri Setinin Sinir Ağları Algoritması ile Modellenmesi

İnsan beyninin çalışma sistemine benzeyen, güçlü ve iş dünyasında sıklıkla tercih edilen modelleme yöntemlerinden biridir. Sinir ağları, çeşitli katmanlarda bulunan ve birden çok girdi toplamının bir aktivasyon fonksiyonundan geçirilmesiyle meydana gelen nöronlardan oluşur ve her bir nöronun görevi bulunmaktadır. Sinir ağlarında girdi (input), gizli (hidden) ve çıktı (output) katman olmak üzere üç katman bulunmaktadır.

Şekil 20: Sinir Ağlarının Katman Yapısı



Kaynak: Yazar tarafından derlenmiştir.

Girdi katmanında tüm bağımsız değişkenler bulunmaktayken, çıktı katmanında ise hedef değişken yer alır. Gizli katmanlar ise girdi katmanından gelen bilgileri işleyerek süreç tamamlandığında çıktı katmanına iletirler. Son gizli katmanda nöronlar birleştirilerek tahmin değerine ulaşılır. Karmaşık yapısı ve yapılan tahminlerin dayanak noktasının tam olarak açık olmaması sinir ağları ile ilgili eleştirilen hususlardan biridir. Sinir ağları algoritması kullanılarak oluşturulan model sonuçları aşağıdadır.

Tablo 3: Sinir Ağları Modeli Sonuçları (Tüzel)

EĞİTİM VERİSİ			TAHMİN			
			0		1	
	GERÇEK	0	2.881		947	
	1	897		2.542		
Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC	
0,7463	0,7626	0,7526	0,7392	0,7576	0,8290	
TEST VERİSİ			TAHMİN			
			0		1	
	GERÇEK	0	1.210		448	
		1	430		1.073	
	Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC
0,7222	0,7378	0,7298	0,7139	0,7338	0,8060	

Kaynak: Yazar tarafından derlenmiştir.

Tablo 4: Sinir Ağları Modeli Sonuçları (Gerçek)

EĞİTİM VERİSİ			TAHMİN			
			0		1	
	GERÇEK	0	825		282	
	1	195		745		
Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC	
0,7670	0,8088	0,7453	0,7926	0,7757	0,8250	
TEST VERİSİ			TAHMİN			
			0		1	
	GERÇEK	0	388		145	
		1	111		326	
	Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC
0,7361	0,7776	0,7280	0,7460	0,7519	0,8000	

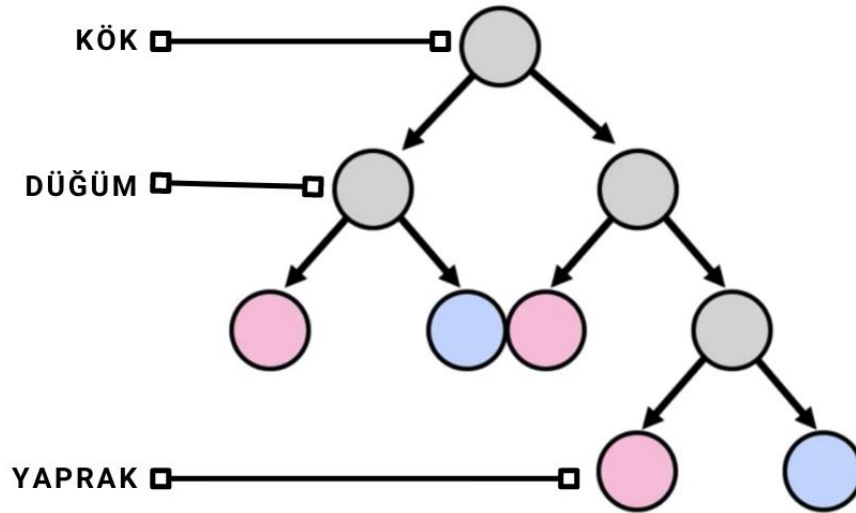
Kaynak: Yazar tarafından derlenmiştir.

Model sonuçları incelendiğinde; tüzel kişi keşidecilere ilişkin eğitim veri setinde %74,6 oranında doğru tahminde bulunurken, test veri setinde bu oranının %72,1 olarak gerçekleştiği görülmektedir. Gerçek kişi keşideciler için oluşturulan modelde ise eğitim veri setinde %76,45 olan doğru tahmin oranının, test verisinde %73,09 olduğu gözlemlenmektedir.

3.5.3. Veri Setinin Karar Ağacı Algoritmaları ile Modellenmesi

Karar ağaçları, karmaşık veri setleri üzerinde gerek regresyon gerekse sınıflandırma problemlerinde uygulanabilen ağaç tabanlı algoritmalarıdır. Kök hücresiyle başlanarak alınacak koşul değerine göre düğümlere inilmekte ve son olarak yapraklar aracılığıyla sonuca ulaşılmaktadır (Akca, 2020).

Şekil 21: Karar Ağaçlarının Yapısı



Kaynak: Yazar tarafından derlenmiştir.

Karar ağacı algoritmalarında düğümlerin ayırım kriteri olarak gini ya da entropi gibi metrikler kullanılmaktadır. Gini tesadüfi olarak seçilen bir örneğin yanlış saptanma sıklığını ölçmek üzere kullanılan metrikken, tesadüfi bir değişkenin

belirsizliğinin ölçüsü ise entropi olarak tanımlanmaktadır. Araştırmada C5.0, CHAID, C&R Tree ve QUEST olmak üzere dört farklı karar ağacı algoritması kullanılmıştır.

3.5.3.1. C5.0 Algoritması

C5.0 algoritması çoklu bölünme uygulanabilen, hedef değişkenin kategorik, bağımsız değişkenlerin sürekli ya da kategorik olabildiği ve bağımsız değişken seçiminde istatistiksel olmayan teknikler kullanılan bir karar ağacı yöntemidir. C5.0 algoritması kullanılarak oluşturulan modellerin performans çıktıları aşağıdadır.

Tablo 5: C5.0 Modeli Sonuçları (Tüzel)

EĞİTİM VERİSİ			TAHMİN			
			0		1	
	GERÇEK	0	3.053		775	
		1	628		2.811	
Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC	
0,8069	0,8294	0,7975	0,8174	0,8132	0,8580	
TEST VERİSİ			TAHMİN			
			0		1	
	GERÇEK	0	1.162		496	
		1	392		1.111	
Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC	
0,7191	0,7477	0,7008	0,7392	0,7235	0,7670	

Kaynak: Yazar tarafından derlenmiştir.

Tablo 6: C5.0 Modeli Sonuçları (Gerçek)

EĞİTİM VERİSİ			TAHMİN	
			0	1
	GERÇEK	0	907	200
1		253	687	

	Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC
	0,7787	0,7819	0,8193	0,7309	0,8002	0,8420

TEST VERİSİ			TAHMİN			
			0			1
	GERÇEK	0	416		117	
		1	156		281	
Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC	
	0,7186	0,7273	0,7805	0,6430	0,7529	0,7520

Kaynak: Yazar tarafından derlenmiştir.

Tüzel kişi keşidecilere ilişkin eğitim veri setinde %80,69 oranında doğru tahminde bulunulduğu ancak test veri setinde bu oranın %71,91'e gerilediği görülmektedir. Gerçek kişi keşidecilerde ise benzer bir şekilde eğitim verilerinde %77,87 olan doğru tahmin oranının test verilerinde %71,86'ya gerilediği ölçümlenmiştir.

C5.0 algoritması kullanılarak gerçekleştirilen modelleme çalışmalarında torbalama (bagging) yöntemi kullanılarak eğitim veri seti üzerinde %98'lere varan doğru tahmin oranına ulaşılabildiği ancak aşırı öğrenme nedeniyle aynı modelin test verisinde %78 oranında doğru tahminde bulunulduğu gözlemlenmiştir.

3.5.3.2. CHAID Algoritması

CHAID algoritması çoklu bölünme uygulanabilen, hedef ve bağımsız değişkenlerin sürekli ve/veya kategorik olabildiği, bağımsız değişken seçiminde istatistiksel teknikler kullanılan bir karar ağacı tekniğidir. CHAID algoritması ile oluşturulan modellerin performans çıktılarına aşağıda yer verilmiştir.

Tablo 7: CHAID Modeli Sonuçları (Tüzel)

EĞİTİM VERİSİ			TAHMİN			
			0		1	
	GERÇEK	0	2.978		850	
1		936		2.503		
Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC	
0,7542	0,7609	0,7780	0,7278	0,7693	0,8320	

TEST VERİSİ			TAHMİN			
			0		1	
	GERÇEK	0	1.179		479	
1		449		1.054		
Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC	
0,7064	0,7242	0,7111	0,7013	0,7176	0,7650	

Kaynak: Yazar tarafından derlenmiştir.

Tablo 8: CHAID Modeli Sonuçları (Gerçek)

EĞİTİM VERİSİ			TAHMİN			
			0		1	
	GERÇEK	0	908		199	
1		169		771		
Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC	
0,8202	0,8431	0,8202	0,8202	0,8315	0,9040	
TEST VERİSİ			TAHMİN			
			0		1	
	GERÇEK	0	394		139	
1		135		302		
Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC	

	0,7175	0,7448	0,7392	0,6911	0,7420	0,7800
--	--------	--------	--------	--------	--------	--------

Kaynak: Yazar tarafından derlenmiştir.

CHAID algoritması kullanıldığında doğru tahmin oranının; tüzel kişi keşidecilerin eğitim veri setinde %75,42, test veri setinde %70,64 olduğu, gerçek kişi keşidecilerde eğitim veri setinde %82,02, test veri setinde ise %71,75 olduğu görülmüştür.

3.5.3.3. C&R Tree Algoritması

C&R Tree algoritması ikili bölünme uygulanabilen, hedef ve bağımsız değişkenlerin sürekli ya da kategorik olabildiği, bağımsız değişken seçiminde istatistiksel olmayan teknikler kullanılan bir karar ağacı yöntemidir. Oluşturulan modellerin performans çıktıları aşağıdaki gibidir.

Tablo 9: C&R Tree Modeli Sonuçları (Tüzel)

EĞİTİM VERİSİ			TAHMİN			
			0		1	
	GERÇEK	0	2.901		927	
	1	868		2.571		
Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC	
0,7530	0,7697	0,7578	0,7476	0,7637	0,8030	
TEST VERİSİ			TAHMİN			
			0		1	
	GERÇEK	0	1.222		436	
		1	379		1.124	
	Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC
0,7422	0,7633	0,7370	0,7478	0,7499	0,7910	

Kaynak: Yazar tarafından derlenmiştir.

Tablo 10: C&R Tree Modeli Sonuçları (Gerçek)

EĞİTİM VERİSİ			TAHMİN			
			0		1	
	GERÇEK	0	820		287	
		1	163		777	
	Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC
0,7802	0,8342	0,7407	0,8266	0,7847	0,8320	

TEST VERİSİ			TAHMİN			
			0		1	
	GERÇEK	0	369		164	
		1	104		333	
	Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC
0,7237	0,7801	0,6923	0,7620	0,7336	0,7860	

Kaynak: Yazar tarafından derlenmiştir.

Model sonuçları incelendiğinde; tüzel kişi keşidecilerin eğitim verisinde %75,3 olan doğru tahmin oranı, test veri setinde %74,22 olarak gerçekleşmiştir. Gerçek kişi keşideciler için oluşturulan modelde ise eğitim veri setinde %78,02 olan doğru tahmin oranı, test verisinde %72,37 olarak ölçümlenmiştir.

3.5.3.4. QUEST Algoritması

QUEST algoritması ikili bölünme uygulanabilen, hedef değişkenin kategorik, bağımsız değişkenlerin sürekli ya da kategorik olabildiği, bağımsız değişken seçiminde istatistiksel teknikler kullanılan bir karar ağacı tekniğidir. Oluşturulan modellerin performans çıktıları aşağıdadır.

Tablo 11: QUEST Modeli Sonuçları (Tüzel)

EĞİTİM VERİSİ			TAHMİN	
			0	1
	GERÇEK	0	2.296	1.532
1		808	2.631	

	Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC	
		0,6780	0,7397	0,5998	0,7650	0,6624	0,7120
TEST VERİSİ			TAHMİN				
			0			1	
	GERÇEK	0	979		679		
		1	344		1.159		
	Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC	
	0,6764	0,7400	0,5905	0,7711	0,6568	0,7070	

Kaynak: Yazar tarafından derlenmiştir.

Tablo 12: QUEST Modeli Sonuçları (Gerçek)

			TAHMİN			
			0			1
EĞİTİM VERİSİ	GERÇEK	0	827		280	
		1	253		687	
Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC	
0,7396	0,7657	0,7471	0,7309	0,7563	0,7520	
TEST VERİSİ			TAHMİN			
			0			1
	GERÇEK	0	409		124	
		1	145		292	
Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC	

	0,7227	0,7383	0,7674	0,6682	0,7525	0,7170
--	--------	--------	--------	--------	--------	--------

Kaynak: Yazar tarafından derlenmiştir.

QUEST algoritması kullanıldığında tüzel kişi keşidecilerin eğitim verisinde %67,8 olan doğru tahmin oranının test verisinde %67,64 olduğu, gerçek kişilerde ise eğitim verisinde %73,96 olan doğru tahmin oranının test verisinde %72,27 olduğu tespit edilmiştir.

3.5.4. Veri Setinin Decision List Algoritması ile Modellenmesi

Karar ağaçları gibi kural çıkarım tekniklerinden biri olan Decision List algoritması, karar ağacı oluşturmadan tekrarlı karar kuralları yaklaşımı ile hedef alan için tahminde bulunur. Decision List algoritması çoklu bölünme uygulanabilen, hedef ve bağımsız değişkenlerin kategorik olabildiği, bağımsız değişken seçiminde istatistiksel tekniklerin kullanıldığı bir algoritmadır. Sürekli bağımlı değişkenler işlemci tarafından otomatik olarak eşit büyüklükteki kategorilere ayrıştırılmaktadır.

Söz konusu algoritma kullanılarak oluşturulan modellerin performans çıktıları aşağıda yer almakta olup doğru tahmin oranının oldukça düşük seviyede olduğu gözlemlenmiştir.

Tablo 13: Decision List Modeli Sonuçları (Tüzel)

EĞİTİM VERİSİ			TAHMİN			
			0		1	
	GERÇEK	0	531		3.297	
	1	1.433		2.006		
Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC	
0,3491	0,2704	0,1387	0,5833	0,1834	0,6420	
TEST			TAHMİN			
			0		1	
	GERÇEK	0	237		1.421	
	1	658		845		

	Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC
	0,3423	0,2648	0,1429	0,5622	0,1857	0,6490

Kaynak: Yazar tarafından derlenmiştir.

Tablo 14: Decision List Modeli Sonuçları (Gerçek)

EĞİTİM VERİSİ			TAHMİN				
			0	1			
	GERÇEK	0	215	892			
		1	544	396			
	Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC	
	0,2985	0,2833	0,1942	0,4213	0,2304	0,7050	
TEST VERİSİ			TAHMİN				
			0	1			
	GERÇEK	0	129	404			
		1	233	204			
	Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC	
	0,3433	0,3564	0,2420	0,4668	0,2883	0,6560	

Kaynak: Yazar tarafından derlenmiştir.

3.5.5. Veri Setinin Lojistik Regresyon Algoritması ile Modellenmesi

Sürekli hedef değişken ile sürekli ya da kategorik bağımsız değişkenler arasındaki ilişkiyi ortaya koymak üzere kullanılan doğrusal regresyonun aksine lojistik regresyonda hedef değişkenin kategorik olması şartı bulunmaktadır. Bir olayın gerçekleşme ya da gerçekleşmeme ihtimalini temsil eden olasılık oranını kullanan lojistik regresyon, olasılık oranının doğal logaritmasını (\ln) esas alır. Lojistik regresyonun amacı; minimum sayıda değişken kullanarak, bağımlı ve bağımsız

değişkenler arasındaki ilişkiyi en iyi tanımlayan ve uyuma sahip modeli kurmaktır (Özcan, 2020).

Oluşturulan modellerin performans çıktılarına aşağıda yer verilmiştir.

Tablo 15: Lojistik Regresyon Modeli Sonuçları (Tüzel)

			TAHMİN			
			0		1	
EĞİTİM VERİSİ	GERÇEK	0	3.026		802	
		1	797		2.642	
	Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC
	0,7800	0,7915	0,7905	0,7682	0,7910	0,8620
			TAHMİN			
			0		1	
TEST VERİSİ	GERÇEK	0	1.261		397	
		1	384		1.119	
	Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC
	0,7529	0,7666	0,7606	0,7445	0,7635	0,8340

Kaynak: Yazar tarafından derlenmiştir.

Tablo 16: Lojistik Regresyon Modeli Sonuçları (Gerçek)

EĞİTİM VERİSİ			TAHMİN	
			0	1
GERÇEK	0	914	193	
	1	159	781	

	Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC
		0,8280	0,8518	0,8257	0,8309	0,8385
TEST VERİSİ	TAHMİN					
			0			1
	GERÇEK	0	413		120	
		1	136		301	
	Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC
	0,7361	0,7523	0,7749	0,6888	0,7634	0,8070

Kaynak: Yazar tarafından derlenmiştir.

Lojistik regresyon algoritması kullanıldığında tüzel kişi keşidecilerin eğitim veri setinde %78 olan doğru tahmin oranı, test veri setinde %75,23, gerçek kişi keşidecilerin eğitim veri setinde %82,8 olan doğru tahmin oranı ise test veri setinde %73,4 olarak ölçülmüştür.

3.5.6. Veri Setinin Random Forest Algoritması ile Modellenmesi

“Kalabalığın bilgeliği” fikrine dayanan Random Forest algoritması, birden fazla tahmincinin toplamının, en iyi bireysel tahminden daha iyi tahminde bulunacağı düşüncesiyle hareket etmektedir (Johnson, 2021).

Algoritma adından da anlaşılabilir üzere çok sayıda ve grup halinde hareket eden karar ağaçlarından oluşmaktadır. Karar ağaçlarının bağımsız olarak tahminde bulunduğu ve en çok oyu alan sınıfın modelin tahmini oluşturduğu bir kurgusu bulunmaktadır (Yiu, 2019).

Random Forest kullanılarak oluşturulan model sonuçları aşağıdadır.

Tablo 17: Random Forest Modeli Sonuçları (Tüzel)

EĞİTİM	TAHMİN	
	0	1

	GERÇEK	0	3.821		7	
		1	43		3.396	
	Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC
	0,9931	0,9889	0,9982	0,9875	0,9935	1,0000
TEST VERİSİ			TAHMİN			
			0		1	
	GERÇEK	0	1.375		283	
		1	407		1.096	
	Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC
	0,7817	0,7716	0,8293	0,7292	0,7994	0,8670

Kaynak: Yazar tarafından derlenmiştir.

Tablo 18: Random Forest Modeli Sonuçları (Gerçek)

EĞİTİM VERİSİ			TAHMİN			
			0		1	
	GERÇEK	0	1.106		1	
		1	25		915	
Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC	
0,9873	0,9779	0,9991	0,9734	0,9884	1,0000	
TEST VERİSİ			TAHMİN			
			0		1	
	GERÇEK	0	437		96	
		1	155		282	
	Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC
0,7412	0,7382	0,8199	0,6453	0,7769	0,8180	

Kaynak: Yazar tarafından derlenmiştir.

Eđitim veri setlerinde ok yksek dođruluk payına (Tzel: %99,31, Gerek: %98,73) sahip olan model, test verilerinde tzel kiřilerde %78,11, gerek kiřilerde ise %73,92 oranında dođru tahminde bulunmuřtur.

3.5.7. Veri Setinin Random Trees Algoritması ile Modellenmesi

Sınıflandırma ve regresyon ađacı metodolojisi zerine kurulan ve ađaç tabanlı bir diđer tahmin algoritması olan Random Trees, birden ok karar ađacından oluřan bir topluluk modeli yapısı kullanmaktadır. Belirli sayıda tahmincinin rastgele seilmesinin ardından C&R Tree algoritmasına benzer bir yapıda her blme alanı iki dalla sonulanmaktadır. Random Trees algortimasında her yaprak dđm tek bir kayıt ierene kadar bymekte ve buna bađlı olarak ađacın derinliđi ok byk olabilmektedir (IBM, 2020).

Tablo 19: Random Trees Modeli Sonuları (Tzel)

EĐİTİM VERİSİ			TAHMİN			
			0		1	
	GEREK	0	3.126		702	
	1	381		3.058		
Dođruluk	Kesinlik	G. ađırma	zgllk	F-1 Skoru	AUC	
0,8510	0,8914	0,8166	0,8892	0,8524	0,9280	
TEST VERİSİ			TAHMİN			
			0		1	
	GEREK	0	1.184		474	
	1	294		1.209		
Dođruluk	Kesinlik	G. ađırma	zgllk	F-1 Skoru	AUC	
0,7570	0,8011	0,7141	0,8044	0,7551	0,8320	

Kaynak: Yazar tarafından derlenmiřtir.

Tablo 20: Random Trees Modeli Sonuçları (Gerçek)

EĞİTİM VERİSİ			TAHMİN			
			0		1	
	GERÇEK	0	1.014		93	
		1	47		893	
Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC	
0,9316	0,9557	0,9160	0,9500	0,9354	0,9850	
TEST VERİSİ			TAHMİN			
			0		1	
	GERÇEK	0	385		148	
		1	111		326	
Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC	
0,7330	0,7762	0,7223	0,7460	0,7483	0,8080	

Kaynak: Yazar tarafından derlenmiştir.

Random Trees algoritması kullanıldığında tüzel kişi keşidecilerin eğitim verisinde %85,1 olan doğru tahmin oranı test verisinde %75,7'ye gerilemiş, gerçek kişilerde ise eğitim verisinde %93,16 olan doğru tahmin oranı test verisinde %73,3'e gerilemiştir.

3.5.8. Veri Setinin XGBoost Algoritması ile Modellenmesi

Gradient boosting algoritmalarının optimize edilen ve yüksek performans gösteren bir türü olan Ekstrem Gradyan Artırım (XGBoost), aşırı öğrenmenin önüne geçerek başarılı tahmin performansına sahip olması nedeniyle oldukça popüler ve karar ağacı tabanlı algoritmalar arasından en iyisi olarak kabul edilen algoritma konumundadır. İlk tahminin yapılmasının ardından bu tahminin ne kadar başarılı olduğu incelenerek, hataları tahminleyen bir karar ağacı oluşturulur. Amaç, hatalar üzerinden öğrenerek başarılı tahmin değerine ulaşmaktır. (Muratlar, 2020).

Tablo 21: XGBoost Modeli Sonuçları (Tüzel)

EĞİTİM VERİSİ			TAHMİN			
			0		1	
	GERÇEK	0	3.168		660	
		1	490		2.949	
	Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC
0,8418	0,8660	0,8276	0,8575	0,8464	0,9240	
TEST VERİSİ			TAHMİN			
			0		1	
	GERÇEK	0	1.240		418	
		1	321		1.182	
	Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC
0,7662	0,7944	0,7479	0,7864	0,7704	0,8460	

Kaynak: Yazar tarafından derlenmiştir.

Tablo 22: XGBoost Modeli Sonuçları (Gerçek)

EĞİTİM VERİSİ			TAHMİN			
			0		1	
	GERÇEK	0	1.032		75	
		1	68		872	
	Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC
0,9301	0,9382	0,9322	0,9277	0,9352	0,9830	
TEST VERİSİ			TAHMİN			
			0		1	
	GERÇEK	0	413		120	

		1	129		308	
	Doğruluk	Kesinlik	G. Çağırma	Özgüllük	F-1 Skoru	AUC
	0,7433	0,7620	0,7749	0,7048	0,7684	0,8180

Kaynak: Yazar tarafından derlenmiştir.

XGBoost algoritması tüzel kişiler veri setinde eğitimde %84,18, test verisinde %76,62 oranında, gerçek kişilerin eğitim veri setinde %93,01, test veri setinde ise %74,33 oranında başarılı olmuştur.

3.6. DEĞERLENDİRME

Modellerin test verileri özelindeki tahminlerine ilişkin performans sonuçları aşağıda özetlenmiştir.

Tablo 23: Modellerin Test Verileri Üzerindeki Performans Sonuçları

Algoritma	G/T	Doğ.	Kes.	G.Çağ.	Özgül.	F-1	AUC
Sinir ağları	Tüzel	0,7222	0,7378	0,7298	0,7139	0,7338	0,8060
Sinir ağları	Gerçek	0,7361	0,7776	0,7280	0,7460	0,7519	0,8000
C5.0	Tüzel	0,7191	0,7477	0,7008	0,7392	0,7235	0,7670
C5.0	Gerçek	0,7186	0,7273	0,7805	0,6430	0,7529	0,7520
CHAID	Tüzel	0,7064	0,7242	0,7111	0,7013	0,7176	0,7650
CHAID	Gerçek	0,7175	0,7448	0,7392	0,6911	0,7420	0,7800
C&R Tree	Tüzel	0,7422	0,7633	0,7370	0,7478	0,7499	0,7910
C&R Tree	Gerçek	0,7237	0,7801	0,6923	0,7620	0,7336	0,7860
QUEST	Tüzel	0,6764	0,7400	0,5905	0,7711	0,6568	0,7070
QUEST	Gerçek	0,7227	0,7383	0,7674	0,6682	0,7525	0,7170
Decision List	Tüzel	0,3423	0,2648	0,1429	0,5622	0,1857	0,6490
Decision List	Gerçek	0,3433	0,3564	0,2420	0,4668	0,2883	0,6560
Logistic Reg	Tüzel	0,7529	0,7666	0,7606	0,7445	0,7635	0,8340
Logistic Reg	Gerçek	0,7361	0,7523	0,7749	0,6888	0,7634	0,8070

Algoritma	G/T	Doğ.	Kes.	G.Çağ.	Özgül.	F-1	AUC
Random Forest	Tüzel	0,7817	0,7716	0,8293	0,7292	0,7994	0,8670
Random Forest	Gerçek	0,7412	0,7382	0,8199	0,6453	0,7769	0,8180
Random Trees	Tüzel	0,7570	0,8011	0,7141	0,8044	0,7551	0,8320
Random Trees	Gerçek	0,7330	0,7762	0,7223	0,7460	0,7483	0,8080
XGBoost	Tüzel	0,7662	0,7944	0,7479	0,7864	0,7704	0,8460
XGBoost	Gerçek	0,7433	0,7620	0,7749	0,7048	0,7684	0,8180

Kaynak: Yazar tarafından derlenmiştir.

Algoritmaların her bir metrikte başarı sıralaması ise aşağıdaki gibi gerçekleşmiştir.

Tablo 24: Algoritmaların Metrikler Özelinde Başarı Sıralamaları (Tüzel)

Algoritma	Doğ.	Kes.	G.Çağ.	Özgül.	F-1	AUC
XGBoost	2	2	3	2	2	2
Random Forest	1	3	1	7	1	1
Random Trees	3	1	6	1	4	4
Lojistik Reg.	4	4	2	5	3	3
C&R Tree	5	5	4	4	5	6
Sinir Ağları	6	8	5	8	6	5
C5.0	7	6	8	6	7	7
QUEST	9	7	9	3	9	9
CHAID	8	9	7	9	8	8
Decision List	10	10	10	10	10	10

Kaynak: Yazar tarafından derlenmiştir.

Tablo 25: Algoritmaların Metrikler Özelinde Başarı Sıralamaları (Gerçek)

Algoritma	Doğ.	Kes.	G.Çağ.	Özgül.	F-1	AUC
XGBoost	1	4	3	4	2	2
Random Forest	2	8	1	8	1	1
Sinir Ağları	3	2	7	2	6	5
Lojistik Reg.	4	5	4	6	3	4
Random Trees	5	3	8	3	7	3

C&R Tree	6	1	9	1	9	6
C5.0	8	9	2	9	4	8
QUEST	7	7	5	7	5	9
CHAID	9	6	6	5	8	7
Decision List	10	10	10	10	10	10

Kaynak: Yazar tarafından derlenmiştir.

Test veri seti üzerindeki model performansları karşılaştırıldığında; hem tüzel, hem de gerçek kişiler için en iyi sonuçların Random Forest ve XGBoost algoritmalarıyla elde edildiği görülmektedir.

SONUÇ

Kurumlarda yapısal, yarı yapısal ve yapısal olmayan sorunlar vardır ve her bir sorunun çözümü için farklı seviye yönetici/çalışan tarafından farklı tip kararların alınması gerekmektedir. Söz konusu kararların alınması sürecinde en alt seviyeden, en üst seviyeye kadar her yöneticinin kendi yetkisindeki kararları doğru alabilmesi için bilişim sistemlerini kullanması önem arz etmektedir. Yazılımcılar, yöneticilere karar verme sürecinde destek olacak yazılım çözümleri geliştirerek, her işi mümkün olduğunca analiz edip ortaya koyar ve bu çözümlerle iş süreçlerinin daha etkin ve verimli yürütülmesini sağlarlar.

Bu çalışmada, çek iskonto/iştira kredilerinin vadesinde ödenme durumuna etki eden faktörlerin belirlenerek, söz konusu kredi türleri için gerçekleştirilen tahsis çalışmalarında sadece makine öğrenmesi tekniklerinin kullanımının mümkün olup olmadığı araştırılmış ve ilgili işlemlerin değerlendirme sürecinde çalışan/yönetici kararına ihtiyaç duyulmaksızın kendi kendisine işletilen bir yapı oluşturulmasının mümkün olup olmadığı irdelenmiştir.

TBB Risk Merkezi tarafından yayımlanan 2021, Haziran ayına ilişkin Bankalara İbraz Edilen ve Karşılıksız İşlemi Yapılan Çek Bilgileri Raporu'na göre; karşılıksız işlemi yapılan çekler, ibraz edilen toplam çekler içerisinde adet olarak %2,2 oranında paya sahiptir (TBB Risk Merkezi, 2021). Araştırmaya konu veri setinin içinde ise karşılıksız işlemi yapılan çeklerin payı %4,9 olarak ölçülmüştür.

Veri setinde ödenmemiş çeklerin toplam çekler arasında sınırlı bir yere sahip olması nedeniyle modellerin eğitim sürecinden önce az sayıda gözleme sahip kategoriye göre azaltma işlemi gerçekleştirilerek veri seti dengelenmişti. Oluşturulan modellerin nihai performans sonucunu öğrenmek için farklı bir ölçüm yöntemi

denenmiş ve eğitilen modellerin veri dengeleme aşaması öncesindeki tüm veri seti için tahminde bulunması sağlanmış olup elde edilen sonuçlara aşağıda yer verilmiştir.

Tablo 26: Algoritmaların Tüm Veri Seti Üzerindeki Performansı (Tüzel)

Algoritma	Kabul Oranı (%)	Vadesinde Ödenen (%)	Vadesinde Ödenmeyen (%)
Random Forest	77,62	99,41	0,59
Lojistik Regresyon	69,31	98,33	1,67
Sinir Ağları	68,18	97,99	2,01
XGBoost	67,98	98,83	1,17
C&R Tree	67,48	98,06	1,94
CHAID	66,23	97,7	2,30
Random Trees	64,63	98,91	1,09
C5.0	62,15	98,47	1,53
QUEST	49,57	97,93	2,07
Decision List	17,61	86,61	13,39

Kaynak: Yazar tarafından derlenmiştir.

Tablo 27: Algoritmaların Tüm Veri Seti Üzerindeki Performansı (Gerçek)

Algoritma	Kabul Oranı (%)	Vadesinde Ödenen (%)	Vadesinde Ödenmeyen (%)
Random Forest	75,12	99,08	0,92
C5.0	69,28	97,79	2,21
XGBoost	68,07	99,04	0,96
Lojistik Regresyon	67,88	98,45	1,55
CHAID	64,63	98,43	1,57
Random Trees	64,19	99,16	0,84
QUEST	63,42	97,94	2,06

Algoritma	Kabul Oranı (%)	Vadesinde Ödenen (%)	Vadesinde Ödenmeyen (%)
Sinir Ağları	62,26	98,33	1,67
C&R Tree	59,56	98,36	1,64
Decision List	29,34	87,33	12,67

Kaynak: Yazar tarafından derlenmiştir.

Tüm veri seti üzerindeki durum incelendiğinde; en yüksek kabul oranı ve kabul edilenler arasında en yüksek ödenme oranına sahip algoritmanın Random Forest olduğu sonucuna ulaşılmıştır.

Elde edilen sonuçların 2022, Haziran ayı itibarıyla ülkemiz geneli ve araştırmaya konu veri seti içerisindeki karşılıksız çek oranı ile karşılaştırılması sonucunda iskonto/iştir kredileri için gerçekleştirilen tahsis çalışmalarında makine öğrenmesi tekniklerinin kullanımının mümkün olduğu sonucuna ulaşılmaktadır.

KAYNAKÇA

Akay, E. Ç. (2020). *Ekonometride Büyük Veri ve Makine Öğrenmesi*. İstanbul: Der Yayınları.

Akca, M. F. (2020). *Karar Ağaçları (Makine Öğrenmesi Serisi-3)*. <https://medium.com/deep-learning-turkiye/karar-agaclari-makine-ogrenmesi-serisi-3-a03f3ff00ba5>, (25.10.2021).

Akkaya M. (2021). *Finansal Risk Yönetimi*. Ankara: Seçkin Yayıncılık.

Alpaydın, E. (2016). *Machine Learning: The New AI*. Cambridge: The MIT Press.

Alper, D. (2011). *Kredi İflas Takası CDS*. Bursa: Ekin Yayınevi.

Amidi, S. (2018). *Machine Learning Tips and Tricks Cheatsheet*. <https://stanford.edu/~shervine/teaching/cs-229/cheatsheet-machine-learning-tips-and-tricks>, (24.10.2021).

Balta, B. K. (2021). *Eser Kavramı ve Yapay Zekâ Ürünleri*. Ankara: Seçkin Yayıncılık.

Chollet, F. (2019). *Python ile Derin Öğrenme*. Ankara: Buzdağı Yayınevi.

Çavdar Ş. Ç., Aydın A. D. (2018). *Finans Alanında Yapay Zekâ ve Ekonometrik Uygulamalar*. Ankara: Seçkin Yayıncılık.

Elmas, Ç. (2021). *Yapay Zekâ Uygulamaları*. Ankara: Seçkin Yayıncılık.

Geron, A. (2021). *Scikit-Learn, Keras ve TensorFlow ile Uygulamaları Makine Öğrenmesi*. Ankara: Buzdağı Yayınevi.

Gürsakal, N. (2017). *Makine Öğrenmesi ve Derin Öğrenme*. İstanbul: Dora Yayınevi.

IBM (2020), *Random Trees Node*. <https://www.ibm.com/docs/en/cloud-paks/cp-data/3.5.0?topic=modeling-random-trees-node>, (27.10.2021).

Javatpoint (2020), *Difference Between Artificial Intelligence and Machine Learning*. <https://www.javatpoint.com/difference-between-artificial-intelligence-and-machine-learning>, (23.10.2021).

Johnson, D. (2021). *R Random Forest Tutorial with Example*. <https://www.guru99.com/r-random-forest-tutorial.html>, (26.10.2021).

Microsoft Azure (2021). *Makine Öğrenmesi Nedir?* <https://azure.microsoft.com/tr-tr/overview/what-is-machine-learning-platform/#benefits>, (21.10.2021).

Muratlar, E. R. (2020). *XGBoost Nasıl Çalışır? Neden İyi Performans Gösterir?* <https://www.veribilimiokulu.com/xgboost-nasil-calisir/>, (27.10.2021).

Nabiyev, V. (2021). *Yapay Zekâ*. Ankara: Seçkin Yayıncılık.

Nilsson, N. J. (2010). *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. Cambridge: Cambridge University Press.

Özcan, K. A. (2020). *Sınıflandırma Problemleri: Utadis ve Sıralı Lojistik Regresyon Modelleri Karşılaştırması*. Ankara: Gazi Kitabevi

Öztürk, K. (2015). *Kredi Politikası ve Değerlendirilmesi*. Ankara: Siyasal Kitabevi.

TBB Risk Merkezi (2021). *Bankalara İbraz Edilen ve Karşılıksız İşlemi Yapılan Çek Bilgileri - 2021*. <https://www.riskmerkezi.org/tr/istatistikler/23>, (29.10.2021).

Uzunođlu, S. (2020). *Bankacılıđa Giriş*. İstanbul: Literatür Yayıncılık

Vurucu, M., Arı M. U. (2017). *Bankacılıkta Ürün ve Hizmetleri*. Ankara: Seçkin Yayıncılık.

Yaslıdađ, B. (2007). *Kredi Derecelendirme Sistemi*. İstanbul: Literatür Yayıncılık.

Yazıcı, M. (2018). *Bankacılıkta Kredi Tahsisi*. Ankara: Seçkin Yayıncılık.

Yiu, T. (2019). *Understanding Random Forest*. <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>, (26.10.2021).