



T.C.
İSTANBUL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



Yüksek Lisans Tezi

**BANKACILIK SEKTÖRÜNDE KREDİ ÖDEMELERİNİN MAKİNE
ÖĞRENME ALGORİTMALARI ile ANALİZİ ve
TAHMİNLEMESİNİN YAPILMASI**

Enes GEZER

Enformatik Anabilim Dalı

Enformatik Programı

DANIŞMAN

Doç. Dr. Zümrüt ECEVİT SATI

Ağustos, 2019

İSTANBUL

*Hygundur
21.08.2019
Doç. Dr. Z. Sati
3. Ecevit*

Bu çalışma,21.08.2019 tarihinde ařağıdaki jüri tarafından Enformatik Anabilim Dalı,Enformatik Programında Yüksek Lisans tezi olarak kabul edilmiştir.

Tez Jürisi


Doç. Dr. Zümrüt ECEVİT SATI (Danışman)
İstanbul Üniversitesi
Fakülte


Prof. Dr. Sevinç GÜLSEÇEN
İstanbul Üniversitesi
Fakülte


Prof. Dr. Seniye Ümit OKTAY FIRAT
Marmara Üniversitesi
Fakülte



20.04.2016 tarihli Resmi Gazete’de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince; Bu Lisansüstü teze, İstanbul Üniversitesi’nin aboneli olduğu intihal yazılım programı kullanılarak Fen Bilimleri Enstitüsü’nün belirlemiş olduğu ölçütlere uygun rapor alınmıştır.

ÖNSÖZ

Tez çalışmam süresince her türlü destek ve yardımından dolayı tez danışmanım Doç.Dr. Zümrüt ECEVİT SATI'ya,

Yüksek lisans eğitimim süresince desteğini esirgemeyen Enformatik Bölüm Başkanı Prof.Dr. Sevinç GÜLSEÇEN'e,

Derslerim boyunca her zaman yanımda olan Enformatik Bölümü'ndeki hocalarım ve bölüm çalışanlarına teşekkürü borç bilirim.

Ağustos 2019

Enes GEZER



İÇİNDEKİLER

Sayfa No

ÖNSÖZ	iv
İÇİNDEKİLER.....	v
ŞEKİL LİSTESİ	viii
TABLO LİSTESİ.....	ix
SİMGE VE KISALTMA LİSTESİ	x
ÖZET	xi
SUMMARY	xii
1. GİRİŞ	1
2. GENEL KISIMLAR	3
2.1. KREDİ KAVRAMI	3
2.1.1. Kredi Tanımı	3
2.1.2. Kredinin Farklı Açılardan Türleri	3
2.1.2.1. Ticari Krediler	5
2.1.2.2. Bireysel Krediler.....	5
2.1.2.3. Bireysel Kredi Kullanımının Etkileri	6
2.1.3. Riskli Krediler	7
2.1.3.1. Takipteki Krediler	8
2.1.3.2. Temerrüte Düşen Krediler	9
2.1.4. Uluslararası Kredi Karşılaştırmaları	10
2.2. BANKACILIKTA RİSK KAVRAMI	12
2.2.1. Banka Sektöründe Risk Yönetimi	13
2.2.2. Karşılaşılan Risk Türleri.....	13
2.2.2.1. Kredi Riski	13
2.2.2.2. Likidite Riski	14
2.2.2.3. Piyasa Riski	15
2.2.2.4. Operasyonel Risk	16
2.2.2.5. Faiz Riski	17
2.2.2.6. Bilanço Riski	18
2.3. MAKİNE ÖĞRENMESİ	18
2.3.1. Makine Öğrenmesi Kavramı	19

2.3.2. Makine Öğrenmesi Süreci Adımları.....	20
2.3.2.1. Problemin Tanımlanması.....	20
2.3.2.2. Veri Toplanması.....	20
2.3.2.3. Verilerin Hazırlanması	21
2.3.2.4. Modelin Kurulması	23
2.3.2.5. Modelin Değerlendirilmesi	23
2.3.2.6. Modelin Kullanılması	23
2.3.3. Makine Öğrenmesi Yöntemleri	23
2.3.3.1. Denetimli Öğrenme.....	24
2.3.3.2. Denetimsiz Öğrenme.....	24
2.3.3.3. Yarı Denetimli Öğrenme.....	25
2.3.3.4. Pekiştirmeli Öğrenme	25
2.3.4. Makine Öğrenmesi Algoritmaları.....	25
2.3.4.1. K-NN Algoritması	26
2.3.4.2. Naive Bayes Algoritması.....	27
2.3.4.3. Karar Ağacı Algoritması	28
2.3.4.4. Rastgele Orman Yöntemi	30
2.3.4.5. Destek Vektör Makineleri	31
2.3.4.6. Lojistik Regresyon Algoritması.....	32
2.3.4.7. Yapay Sinir Ağları	33
3. MALZEME VE YÖNTEM.....	35
3.1. ARAŞTIRMA MODELİ.....	35
3.2. VERİLERİN TOPLANMASI.....	35
3.3. VERİLERİN ANALİZİ	36
3.4. ÇALIŞMANIN AMACI VE HEDEFİ.....	37
3.5. ÇALIŞMADA KULLANILACAK YÖNTEMLER.....	37
3.5.1. Azure Machine Learning Studio	37
3.5.2. Veri Ön İşleme	38
4. BULGULAR.....	41
4.1. ÖDEMESİ YAPILMAYAN KREDİ ORANLARI	41
4.2. KREDİ BAŞVURUSU YAPAN TÜKETİCİ ANALİZİ.....	42
4.3. MAKİNE ÖĞRENMESİ SONUÇLARI.....	45
4.3.1. Değerlendirme Ölçütleri	45
4.3.2. Değerlendirme Adımları.....	46

4.3.2.1. Tüm veri kümesi ile yapılan lojistik regresyon sonuçları.....	51
4.3.2.2. Tüm veri kümesi ile çalıştırılan karar ağacı sonuçları	52
4.3.2.3. Alt örnekleme metodu ile çalıştırılan lojistik regresyon sonuçları.....	52
4.3.2.4. Alt örnekleme metodu ile çalıştırılan karar ağacı sonuçları.....	53
4.3.2.5. Üst örnekleme metodu ile çalıştırılan lojistik regresyon sonuçları.....	54
4.3.2.6. Üst örnekleme metodu ile çalıştırılan karar ağacı sonuçları.....	56
5. TARTIŞMA VE SONUÇ	57
5.1. ANALİZ SONUÇLARI.....	57
5.2. BANKACILIK SEKTÖRÜ AÇISINDAN SONUÇLAR.....	61
5.3. ÖNERİLER.....	62
5.4. SINIRLILIKLAR.....	63
KAYNAKLAR.....	65
EKLER	71
EK 1. Azure ML Studio Programı Modelleme Ekran Görüntüsü	71
ÖZGEÇMİŞ	72

ŞEKİL LİSTESİ

Sayfa No

Şekil 2.1: Kredi türlerinin gelişimi (BDDK, 2018).....	5
Şekil 2.2: Takipteki krediler (BDDK, 2018).....	9
Şekil 2.3: Kişi başına kredi (dolar)(TBB, 2018).....	11
Şekil 2.4: Kredi ve mevduat oranı, yüzde birimde ifadesi (kredi/mevduat) (TBB, 2018).....	12
Şekil 2.5: Sorunlu krediler ve toplam krediler oranı (yüzde birimde ifadesi) (TBB, 2018).....	12
Şekil 2.6: Karar ağacı yapısı (Argüden ve Erşahin, 2008).....	30
Şekil 2.7: 2 sınıflı problem örneği (Ayhan ve Erdoğan, 2014).....	32
Şekil 2.8: Lojistik regresyon gösterimi (Kalaycı, 2018).....	33
Şekil 2.9: Yapay sinir ağları süreci (Zaloğlu, 2015).....	34
Şekil 3.1: Azure ML Studio önyüzü (Chappell, 2015).....	38
Şekil 4.1: FPD oran grafiği.....	41
Şekil 4.2: Kredi başvuru cinsiyet oranı.....	42
Şekil 4.3: Kredi başvuru medeni durum oranı.....	43
Şekil 4.4: Kredi başvuru eğitim grafiği.....	44
Şekil 4.5: Kredi başvuru ev durum grafiği.....	44
Şekil 4.6: Veri kümesi önizleme.....	47
Şekil 4.7: Seçilen kolonlar gösterimi.....	47
Şekil 4.8: Veri kategorikleştirme ekran görüntüsü.....	48
Şekil 4.9: Boş değerlere atama yapma ekran görüntüsü.....	48
Şekil 4.10: Veri normalleştirme ekran görüntüsü.....	49
Şekil 4.11: Eğitim-test verisi dağıtma ekran görüntüsü.....	51
Şekil 5.1: Model doğruluk puanları karşılaştırması.....	60
Şekil 5.2: Model F1 puanları karşılaştırması.....	60

TABLO LİSTESİ

Sayfa No

Tablo 2.1: Kredi türleri (Sungur,1999).....	4
Tablo 2.2: Yıl bazında kredi tutarı (BDDK,2018).....	6
Tablo 4.1: Karışıklık matrisi unsurları.	45
Tablo 4.2: Veri normalleştirme öncesi değerler.	49
Tablo 4.3: Veri normalleştirme sonrası değerler.	50
Tablo 4.4: Tüm veri kümesi ile yapılan lojistik regresyon sonuçları.	51
Tablo 4.5: Tüm veri kümesi ile çalıştırılan karar ağacı yöntemi sonuçları.	52
Tablo 4.6: Alt örnekleme metodu ile çalıştırılan lojistik regresyon sonuçları.	52
Tablo 4.7: Alt örnekleme lojistik regresyon öznitelik puanı.	53
Tablo 4.8: Alt örnekleme metodu ile çalıştırılan karar ağacı yöntemi sonuçları.....	53
Tablo 4.9: Alt örnekleme karar ağacı öznitelik puanı.	54
Tablo 4.10: Veri kümesinin SMOTE ile yüzdesi.	54
Tablo 4.11: Üst örnekleme metodu ile çalıştırılan lojistik regresyon sonuçları.	55
Tablo 4.12: Üst örnekleme lojistik regresyon öznitelik puanı.....	55
Tablo 4.13: Üst örnekleme metodu ile çalıştırılan karar ağacı algoritması sonuçları.	56
Tablo 4.14: Üst örnekleme karar ağacı öznitelik puanları.....	56
Tablo 5.1: Model değerlendirme kriterleri.	56

SİMGE VE KISALTMA LİSTESİ

Kisaltmalar	Açıklama
ABD	: Amerika Birleşik Devletleri
BDDK	: Bankacılık Denetleme ve Düzenleme Kurulu
GOÜ	: Gelişmekte Olan Ülkeler
GÜ	: Gelişmiş Ülkeler
ML	: Machine Learning
SMOTE	: Synthetic Minority Oversampling Technique
SVM	: Support Vektor Machine
S&P	: Standard and Poor's
TBB	: Türkiye Bankalar Birliği
TÜİK	: Türkiye İstatistik Kurumu

ÖZET

YÜKSEK LİSANS TEZİ

BANKACILIK SEKTÖRÜNDE KREDİ ÖDEMELERİNİN MAKİNE ÖĞRENME ALGORİTMALARI İLE ANALİZİ ve TAHMİNLEMESİNİN YAPILMASI

Enes GEZER

İstanbul Üniversitesi

Fen Bilimleri Enstitüsü

Enformatik Anabilim Dalı

Danışman : Doç. Dr. Zümrüt ECEVİT SATI

Kredi risk yönetimi bankaların yaşamlarını sürdürebilmesi açısından hayati önem taşımaktadır. Bu nedenle kredi başvuru ve kullandırım süreci öncesinde kredi kullanımı yapacak tüketicinin davranışını tahminleme çalışması, finans ve bankacılık sektöründe kritik bir yere sahiptir. Bu kapsamda kredinin ödenebilir olup olmadığının tespiti kesin bir ihtiyaçtır. Makine öğrenme algoritmaları ile sağlanan tahminleyici analizler doğrultusunda banka açısından sağlıklı krediler verilebilmesi öngörülmekte ve tamamlanması ile kredi başvuru sürecinin daha bilinçli ve hatasızca tamamlanması beklenmektedir.

Bu çalışmada bankaya kredi başvurusunda bulunmuş tüketicilerin kredilerinin takibe düşüp düşmeyeceğini, ödemede sorun yaşayıp yaşamayacağını veya makul bir doğrulukla kredi kullanıp kullanmayacağını tahmin etmek amaçlanmıştır. Bu amaç doğrultusunda istenilen, tüketici kredi taleplerine ilişkin herhangi bir karar vermeden önce bu tahminleme ve analizin etkin şekilde kullanılmasıdır. Çalışma sonucu oluşan modelin, bankanın kredi kaybının önlenmesi, performansının ve verimliliğinin artması, risk yönetimi ve ekonomik açıdan kaynakların doğru kullanılması açısından önem arz ettiği düşünülmektedir.

Ağustos 2019, 84 sayfa.

Anahtar kelimeler: Makine öğrenmesi, bankacılık sektörü, kredi ödemeleri.

SUMMARY

M.Sc. THESIS

ANALYSIS and PREDICTION of LOAN PAYMENTS by MACHINE LEARNING ALGORITHMS in BANKING SECTOR

Enes GEZER

İstanbul University

Institute of Graduate Studies in Sciences

Department of Informatics

Supervisor : Assoc. Prof. Dr. Zümürüt ECEVİT SATI

Credit risk management plays a vital role for survival of banks, therefore it is not only important, but also very critical to predict and manage the loan application and allocation process. It is necessary for the bank to understand whether it is going to be a non-performing or performing credit before granting the approval based on customer behaviour. As a result, it is aimed to offer a better and healthier credit appropriation by presenting the predictive analysis provided by machine learning algorithms. By the help of this project, it will be possible to manage the credit application process in more conscious and error free way.

In this study, it is aimed to estimate whether the credits of the consumers are going to be non-performing or performing and whether the credit loan allocations are applied in a controlled and proper way, or not. Applying the prediction analysis in the study in an efficient way is the main expectation of the project before making any decisions about credit, increasing the performance and efficiency, correct usage of the resources economically are going to be some of the results of the model that is built at the end of this work.

August 2019, 84 pages.

Keywords: Machine learning, banking sector, credit payments.

1. GİRİŞ

Finans sektöründe bankaların verdikleri krediler sürekli artış yaşamaktadır. Bu durum finans sektöründeki yatırım şirketleri için karlı gibi görünse de beraberinde riski de getirmektedir. Bu sebeple bankalar karlarını maksimuma çıkarmak ve kredilerden dolayı oluşabilecek riskleri minimuma indirmek için büyük ölçekte çalışmalar yapmaktadırlar. Çalışmaların bankaya sağlaması istenilen fayda var olan veya olacak müşterilerinin kredi borcuna girmemesi olarak düşünülmektedir. Bu gelişmeler sonucunda rahatlıkla söyleyebiliriz ki kredi ve risk modellemeleri oldukça kritik bir önem taşımaktadır.

Bu araştırmanın problemi “ Ödemede gecikmeye düşen kredilerin tahminlemesi başvuru öncesi makine öğrenme algoritmaları ile belirlenebilir mi?” sorusudur.

Çalışmanın amacı kredi başvurularının bankacılık sektöründeki kredi karar yetkilileri tarafından makine öğrenme algoritmaları kullanılmasıyla değerlendirilip yapılan kredi başvurularının etkinliğinin artırılmasıdır.

Tezin ilk bölümünde kredi, risk ve makine öğrenme kavramları incelenmiştir. Kredi kavramına baktığımız zaman kredi türleri ve konumuzu oluşturan riskli krediler ile ilgili bilgi verilmiştir. Risk bölümünde ise bankaların karşılaşılabileceği risk çeşitleri açıklanmış ve çalışmanın amaçlarından birisi olan kredi riskine sebep olabilecek durumlar kaynaklardan yararlanılarak anlatılmıştır.

Tezin malzeme ve yöntem kısmında, araştırma için kullanılan veri toplama ve analiz yöntemleri ve süreçleri açıklanmış, ayrıca yöntemin amacı ve hedefi tekrar belirtilmiştir. Ek olarak makine öğrenme algoritmalarını uygulayacağımız Azure ML Studio programına da değinilmiştir.

Tezin bulgular bölümünde toplanılan veri üstünden analizler yapılmış ve yorumlanmıştır. Yapılan analizler Türkiye üzerindeki sektör verileri ile de karşılaştırılmış ve verilerin doğru yorumlanıp yorumlanamayabileceği hakkında bilgi verilmiştir.

Tartışma ve sonuç başlığı altında da kullanılan makine öğrenme algoritmaları karşılaştırılmış ve birbiri arasındaki performansları değerlendirilmiştir. Hangi modelin uygulanabilirliği daha fazla, güçlü ve zayıf yönleri nelerdir, gibi sorulara cevap bulunmaya çalışılmıştır.

Çalışmamızın son bölümünde ise yapılan arařtırmalarda kullanılan kaynaklar ve ekler belirtilmiřtir.



2. GENEL KISIMLAR

2.1. KREDİ KAVRAMI

Kredinin kelime olarak anlamına bakacak olursak bir mal ya da paranın sonrasında geri iadesinin verilmesi koşuluyla alınmasıdır. Kredinin sözcük olarak kökenine baktığımız zaman Latincedeki inanma anlamında kullanılan “credere” sözcüğünden türemiştir. Latincede güven ve saygınlık anlamına gelmektedir (Usta, 2000). Kredinin güven ve güvenceye bağlı bir işlem olduğu söylenilebilir (Zarakolu, 1989).

2.1.1. Kredi Tanımı

Bankacılık açısından kredinin tanımına bakacak olursak; Bankacılık Kanunu’nda yer alan 5411 numaralı Bankacılık Kanunu’nun 48. Maddesinde yer alan ifadeye göre; bankalarca verilen teminat mektupları, kefaletler, aval, ciro, kabul gibi gayri nakdi krediler, satın alınan tahvil ve benzeri sermaye piyasası araçları, varlıkların vadeli satışından doğan alacaklar, vadesi geçmiş nakdi krediler, tahakkuk etmekle birlikte tahsil edilmemiş faizler, nakde tahvil olan gayrinakdi kredi bedelleri, ters repo uygulamalarından alacaklar, tevdiatta bulunmak suretiyle verilen ödünçler, ortaklık payları, vadeli işlem ve opsiyon sözleşmeleri ile benzeri diğer sözleşmeler nedeniyle üstlenilen riskler ve Bankacılık Düzenleme ve Denetleme Kurumu tarafından kredi olarak kabul edilen işlemlerin izlendikleri hesaba bakılmaksızın bu kanun uygulamasında kredi sayılmaktadır (Altıntaş, 2006).

2.1.2. Kredinin Farklı Açılardan Türleri

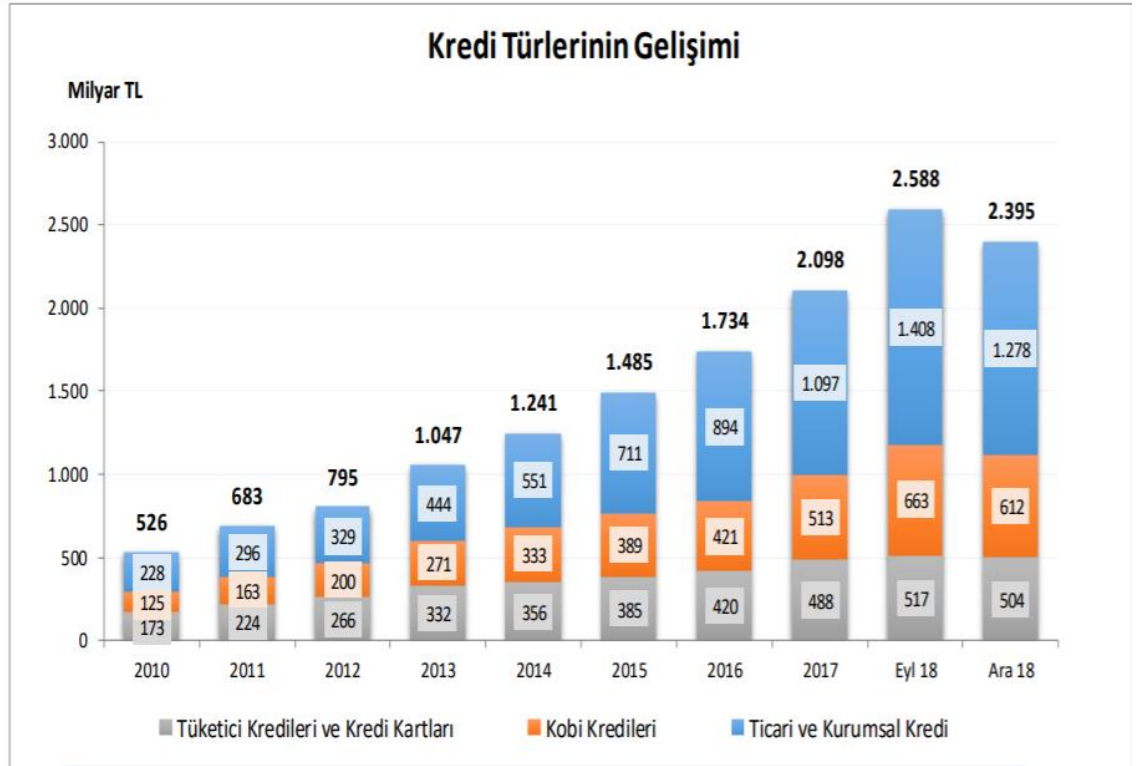
Krediler bir çok farklı türde sınıflandırılabilir. Tablo 2.1’de kredilerin özet olarak sınıflandırılması gösterilmiştir (Sungur, 1999).

Tablo 2.1: Kredi türleri (Sungur,1999).

Niteliklerine Göre Krediler	<ul style="list-style-type: none"> • Nakit Krediler • Gayri Nakit Krediler
Vadelerine Göre Krediler	<ul style="list-style-type: none"> • Vadesi Kısa Olan Krediler • Vadesi Orta Olan Krediler • Vadesi Uzun Olan Krediler
Teminatlarına Göre Krediler	<ul style="list-style-type: none"> • Teminatı Olmayan Krediler • Teminatlđ Krediler • Şahsi Teminatlđ Krediler • Maddi Teminatlđ Krediler
Kullanım Amacına Göre Krediler	<ul style="list-style-type: none"> • Yatırım Kredileri • Donatım Kredileri • İşletme Kredileri
Kaynaklarına Göre Krediler	<ul style="list-style-type: none"> • Banka Kaynaklarından Kullanılan • TCMB, Eximbank, Toplu Konut İdaresi Başkanlığı gibi yabancı kaynaklarından kullanılan krediler
Kullandırım Yetki Türüne Göre Krediler	<ul style="list-style-type: none"> • Şube Yetkili Krediler • Otorize Yetkili Krediler
İş Konusuna Göre Krediler	<ul style="list-style-type: none"> • Kurumsal Krediler • İhracat Kredileri • Tarım Kredileri • Konut Kredileri • Tüketici Kredileri

2.1.2.1. Ticari Krediler

Ticari krediler, işletme görevi gören ve ürün ve hizmet alım satımı yapan firmaların finansmanını karşılamak için kullanılan kredilerdir. Bu alınan krediyle işletmeler sermayelerini arttırabilir ve hacimlerini arttırabilirler. Küçük ve orta büyüklükteki işletmeler için bankalar tarafından ürün olarak sunulabilir. Kurumsal krediler ise büyük ölçekteki şirketler için tutarı yüksek olan kredilerdir. İşletmelerin piyasada kalıcılığını devam ettirebilmesi ve nakit ihtiyacının karşılanması için kullanılır (Sungur, 1999).



Şekil 2.1: Kredi türlerinin gelişimi (BDDK, 2018).

2.1.2.2. Bireysel Krediler

Bireysel Krediler, kişisel ihtiyaçların karşılanması amacıyla nakit olarak bankalar tarafından kullanılan kredilerdir. Bireysel kredilerin kullanıldığı alanları aşağıdaki şekilde sıralayabiliriz.

- Taşıt Kredileri,
- Konut Kredileri,

- Kredi Kartları,
- İhtiyaç Kredileri.

Bireysel Krediler için BDDK'nın raporuna bakacak olursak Aralık 2018 döneminde bireysel kredi kartlarının toplam tutarı 105 milyar TL, ihtiyaç kredileri tutarı ise 205 milyar TL, konut kredileri de 188 milyar TL düzeyindedir. Önceki yıla göre %3 artış göstermiştir. Bireysel kredilerin rakamlarını incelediğimiz zaman ise ihtiyaç kredisi ile birlikte diğer tüketici kredilerinin payı %41, konut kredisinin oranı %37 ve kredi kartlarında ise %21 rakamlarına ulaşılmaktadır.

Tablo 2.2: Yıl bazında kredi tutarı (BDDK,2018).

Yıl	Kredi Tutarı (Milyar TL)
2010	173
2011	224
2012	266
2013	332
2014	356
2015	385
2016	420
2017	489
2018	505

2.1.2.3. Bireysel Kredi Kullanımının Etkileri

Bireysel Kredilerin bankalar üzerindeki etkisini inceleyecek olursak, bankalar yatırım hacimlerini bireysel kredilerle arttırabilir ve kredi risklerini bireysel seviyeye indirerek kredilerin batma ihtimalini azaltabilirler. Bankalar, ellerindeki yatırım fonlarını bireysel olarak

yüksek faiz getirisiyle değerlendirilip yüksek kar elde edebilirler. Piyasa dalgalanmalarından daha az etkilenen bu krediler, risk yönetiminde de etkin rol alabilir (Kaptan, 2011).

Bireyler açısından ele alacak olursak tüketici kendi ihtiyacını karşılayacağından kredi kullanımı kendi nakit yönetimi için uygundur. Tüketici hayat standartlarını alınan kredilerle yükseltebilir. Yapılan kredi sözleşmeleri ile tüketici uyulması kredi için gereken ödeme koşullarına uymalıdır.

Kredilerin bankalar açısından olumsuz etkilerine bakacak olursak, iyi yönetilmeyen bir kredi süreci sonucu tüketici, yüksek meblağlarda borç altına girebilir ve elindeki varlıkların likiditeye dönüşmemesi sonucu finansal varlıklara yatırım yapma yeteneği azalabilir. Ek olarak faiz faktörünün de tüketici ve banka arasında önemli bir faktör olduğunu söyleyebiliriz. Merkez bankası tarafından belirlenen faiz oranları, tüketicinin kredi talebi üzerinde büyük bir öneme sahiptir. Ek olarak fazla miktarda kullanılan bireysel krediler harcama taleplerini arttıracığı için piyasada enflasyon oluşmasına sebep olur.

Ekonomik yönetim açısından bakacak olursak, krediler piyasadaki paranın dolaşım hızı için pozitif etki gösterdiği kabul edilmiştir. Tüketicilerin kredi tutarının artması toplumun, bir nevi tüketim toplumu haline gelmesine sebep olur. Kredi taleplerindeki artış, toplam arzında artmasına ve üretimin de dolaylı olarak artışına sebep olur. Bu nedenle bireysel krediler, piyasadaki hacmi arttırıcı ve ekonomik olarak büyümeye sebep olan bir faktör olarak söylenebilir (Kaptan, 2011).

2.1.3. Riskli Krediler

Krediler, bankaların tüketicilere sağladığı en temel üründür. Kredileri karşılamak için de banka üzerindeki mevduatları kullandırırlar. Bunun sonucunda da karlarını maksimize etmek için çalışmalarda bulunurlar. Bu çalışmaların hedefi, kullanılan kaynaklardan dolayı oluşan risk ve gelirin dengesini kurmak ve riskin sebebini belirleyip önleyici bir yönetim sağlamaktır. Riskli krediler, borçlu firma ve bankalar arasında yapılmış olan geri ödeme planının bozulması sonucu oluşan krediler olarak tanımlanabilir. Kredilerin makul bir süre içerisinde geri ödenmemesi, tahsilatın gecikmesi ve zarar ihtimalinin ortaya çıkması durumları sorunlu kredileri ortaya çıkarmaktadır. Geri ödenmesinde herhangi bir sorun olmaması durumunda krediler bankalar

için oldukça karlı bir üründür. Bu nedenle oluşabilecek risklerin engellenmesi önem kazanır (Koyuncu ve Saka, 2011).

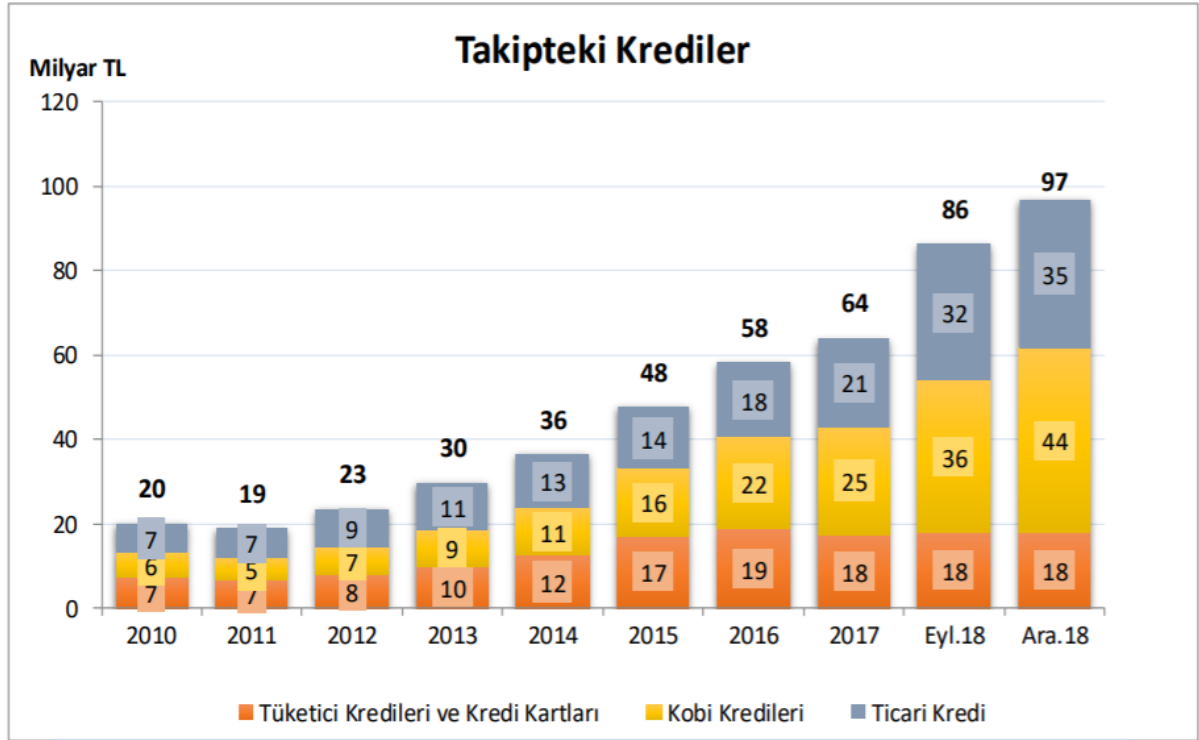
Riskli krediler başlığında Takipteki Krediler ve Temerrüte düşen krediler incelenecektir.

2.1.3.1. Takipteki Krediler

Takipteki krediler literatürde sorunlu krediler olarak adlandırılabilir (Koyuncu ve Saka, 2011). Kredi için başvuran ve anlaşmayı yapan banka ile tüketici arasındaki anlaşmanın bozulması sonucu ortaya çıkan kredilerdir. Kredinin vade ile paralel olarak ödemesinin yapılmaması veya ödemenin belirli bir zaman için gecikmesi ile krediler sorunlu durumuna gelmektedir. Kısaca tahsilatın ya da ödemenin gecikmesi bankalar açısından ilgili krediyi sorunlu kredi olarak adlandırması için yeterlidir (Aktaş, 2000). BDDK, 30.06.2001 tarihinde yayınlanmış olan Resmi Gazete’de ‘Bankalarca Karşılık Ayrılacak Kredilerin ve diğer Alacakların Niteliklerinin Belirlenmesi ve Ayrılacak Karşılıklara İlişkin Esas ve Usuller Hakkında Yönetmelik’ yayımlamıştır. Bu yönetmeliğe ‘Karşılıklar Yönetmeliği’ de denilmektedir. Yönetmelik içerisindeki 5. Maddede ‘tahsil kabiliyetine göre birinci ve ikinci grup dışında yer alan tüm kredi türleri bankalar açısından donuk alacak diğer bir ifadeyle takipteki kredidir.’ İbaresini kullanılmıştır.

Takipteki krediler için örnek vermek gerekirse bireysel kredilerde vadenin 1 gün gecikmesi veya kurumsal ve kobi kredilerinde anaparanın vadeden 90 gün geçmesi krediyi sorunlu sayabilir (Selimler, 2006).

Takipteki kredi değişimleri Şekil 2.2’de gösterilmektedir.



Şekil 2.2: Takipteki krediler (BDDK, 2018).

2.1.3.2. Temerrüte Düşen Krediler

Temerrüt; BDDK'nın, 24448. sayılı ve 30.06.2001 tarihinde yayınlanan Resmi Gazete'deki 'Bankalarca Karşılık Ayrılacak Kredilerin ve Diğer Alacakların Niteliklerinin Belirlenmesi ve Ayrılacak Karşılıklara İlişkin Esas ve Usuller Hakkında Yönetmeliği' ne göre kredi değerinin zayıflaması olarak tanımlanmıştır. Bankaların Temerrüt Kredisi (takipteki kredi olarak da adlandırılır), tüketicinin banka koşullarını yerine getirmede bir başarısızlık olduğunda ve vadesine ulaşan kredi koşullarına göre geri ödenemediğinde oluşmaktadır.

Bankalar açısından baktığımız zaman kredi riskini yönetmek karlılığı korumaya yardımcı olacaktır. Bu aşamada alınan kararların rasyonelliğini ölçmek için bazı bilimsel yöntemler bulunmaktadır. Eğer bankalar müşteriler tarafından oluşan yatırım risklerini iyi yönetebilirlerse, bütün olası sonuçları hesaplayabilirler. Bu doğrultuda, bankaların yatırım riskini minimum seviyede tutmak oldukça önem taşımaktadır. Krediler için yapılan hesaplamalar neticesinde bankalar olumlu ya da olumsuz sonuçlara karşı kendilerine bir pozisyon alıp bu sonuçlara karşı kendilerini maddi olarak hazır tutabilirler (Bromiley,2015). Bu nedenle ödenmeyen kredilerin oluşturduğu yük bankadaki paranın akışını bozabilir ve

bankanın iflasına neden olabilir. Bu ödenmeyen ve takip edilen kredilere de Temerrüt Kredileri olarak adlandırılmaktadır. Ayrıca bu ödenmeyen kredilerden oluşan riskleri etkileyen başka faktörler de bulunmaktadır. Bu faktörler arasında ülke riski ve kredi derecelendirme kuruluşlarının bankacılık sektöründeki etkileri gibi faktörleri de sayabiliriz (Yarız, 2012).

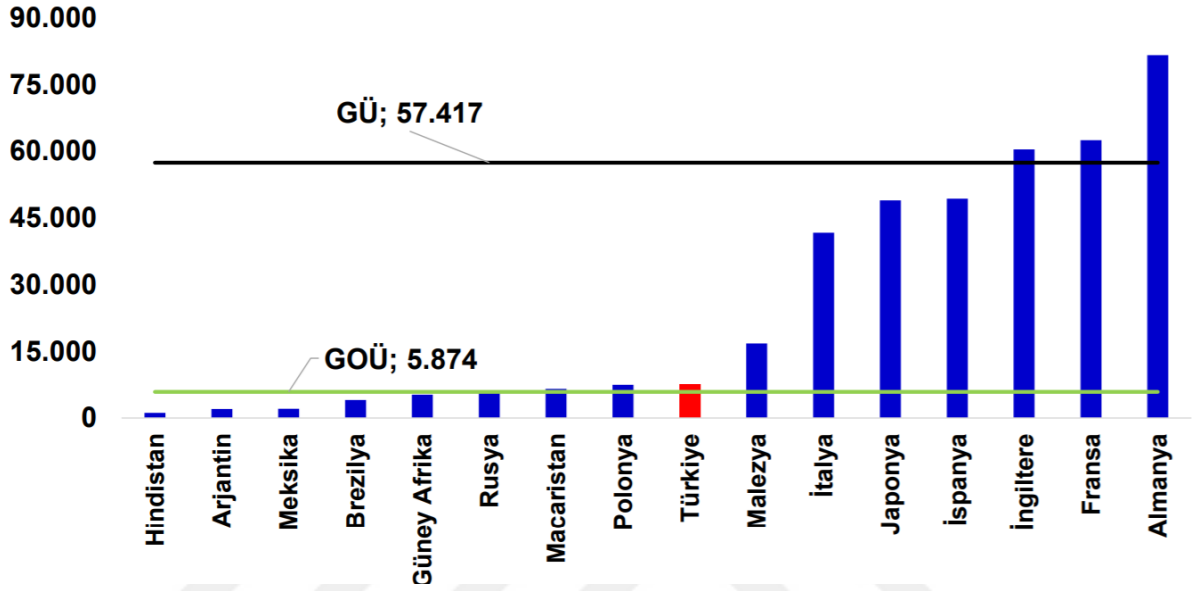
Bankalar, yalnızca bir temerrüdün gerçekleşip gerçekleşmeyeceğinin belirlenmesini ve aynı zamanda müşterilerin ne zaman kredilerini geri ödeyemeyecekleri ve temerrütlerin ne zaman ortaya çıkacağını da araştırmaktadırlar. Türkiye Bankalar Birliği'nin 2018 Eylül raporuna göre Temmuz-Eylül 2018 dönemi itibariyle yaklaşık 11 milyar TL olan takipteki krediler, tüketici kredileri ve konut kredileri toplamının yüzde 2,9'unu oluşturmuştur. Temmuz-Eylül 2018 döneminde kullanılan tüketici kredileri ve konut kredilerinden kanuni takibe alınan kredi miktarı, bir önceki yılın aynı dönemine göre yüzde 23 oranında azalarak 652 milyon TL olmuştur (TBB, 2018).

2.1.4. Uluslararası Kredi Karşılaştırmaları

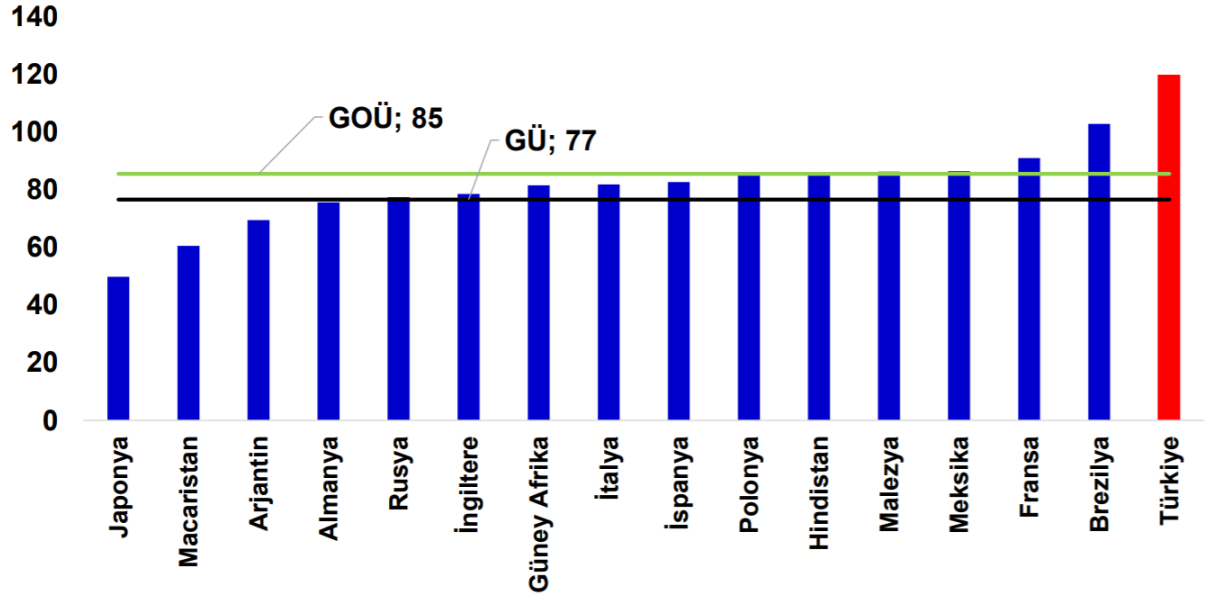
ABD bankaları kredi taleplerinin işletme ve gayrimenkul alanındaki azalmasına karşılık, işletmelere yönelik kredi standartlarını 2017 yılının 3'üncü çeyreğinde daha da esneterek bu azalmayı durdurmaya çalıştı. Ancak ticari gayrimenkuller için kredi standartlarında değişiklik olmadı. Euro Bölgesi'nin 2017 yılındaki 3'üncü çeyrek rakamlarına baktığımız zaman ise yüzdesel büyüme olarak yüzde 2,2 civarında olduğu tahmin edilmektedir (TBB, 2018). Ekonomik büyümede önemli rolü olan Avrupa bankacılık sektöründe, euro mevduatı ve krediler ise 2017 yılında büyüme gösterdi. Euro cinsinden olan mevduatlar 2017'ye yılına kadar olan 10 yıllık süreçte yüzde 4 artış yakaladı. Pozitif yönde olan bu büyüme ve değer artışının yanında ise sorunlu kredilerin birikmesi sebebi ile riskli bir dönemden geçiliyor ve karlılıkla ilgili sorunların devam ettiğini söyleyebiliriz (TBB, 2018).

2019 yılında dünyadaki en iyi data sağlayıcı olarak seçilen S&P Global Market Intelligence verilerini kullanarak TBB'nin bankacılık sektörünü uluslararası karşılaştırdığı rapora göre Şekil 2.3'de gelişmekte olan ülkelerin kredi rakamları gösterilmiştir (TBB, 2018). Türkiye, kişi başı kredi miktarında, gelişmekte olan ülkelerin en yükseklerinden biridir. Türkiye'nin kredi/mevduat oranında en yüksek ülke olduğunu grafikteki değerleri gözlemlediğimizde farkedebiliriz. Türkiye'nin, kredilerin aktif içindeki payı ortalamadan hayli yüksek iken, pasifler içinde ise mevduatların payı ortalamanın altında görülmektedir (TBB, 2018). Tahsili

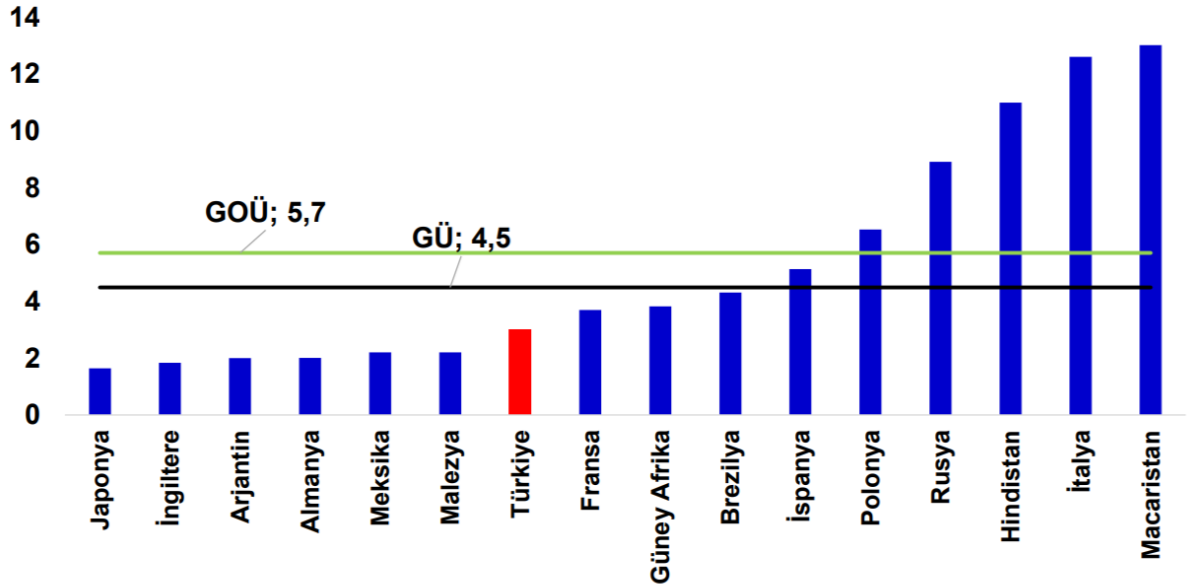
gecikmiş alacak ve takibe düşmüş kredilerin toplam kredilere oranı, gelişmekte olan ülke ortalamasının altındadır. Gelişmekte olan ülkeler arasındaki, takibe ya da gecikmeye düşmüş kredinin toplam krediye oranı verisine göre Türkiye ortalamasının altındadır.



Şekil 2.3: Kişi başına kredi (dolar)(TBB, 2018).



Şekil 2.4: Kredi ve mevduat oranı, yüzde biriminde ifadesi (kredi/mevduat) (TBB, 2018).



Şekil 2.5: Sorunlu krediler ve toplam krediler oranı (yüzde biriminde ifadesi) (TBB, 2018)

2.2. BANKACILIKTA RİSK KAVRAMI

Risk kelimesi zararlarla karşılaşma tehlikesi olarak adlandırılmaktadır. Zararlarla veya tehlikeyle karşılaşma tehlikesi olarak kullanılmaktadır. Riskle en çok sağlık sektöründe, çevresel alanlarda, güvenlik sektöründe, sigortacılık ve bankacılıkta karşılaşıldığı söylenebilir. BDDK'y

a göre ise ‘bir işleme ya da faaliyete ilişkin bir parasal kaybın ortaya çıkması veya bir giderin ya da zararın oluşması halinde ekonomik faydanın azalması ihtimali’ ifadesiyle tanımlanmıştır (Bozkurt, 2010). Literatürdeki tanımlara baktığımız zaman ise gelecekteki karşılaşılabilecek durumlara karşı günümüzde oluşacak belirsizliklerdir (Coyle, 2000). Bankacılık sektöründe risk gerçekleşecek kazanç ile beklediğimiz kazanç arasındaki farktır (Boyacıoğlu, 2003). Ayrıca bankacılıkta risk bankanın zarar etmesi ya da bankanın belirlediği hedeflerin ulaşılabilir olmanın uzağında kalmasıdır (Santamore, 1997).

2.2.1. Banka Sektöründe Risk Yönetimi

Bankacılık sektöründe karlılığın sürekliliğini sağlamak için banka içi çalışmaların ve tedbirlerin hepsine risk yönetimi denir. Risk tiplerini iç risk ve dış risk olmak üzere ikiye ayırabiliriz. İç riskler banka tarafından kontrol edilebilen risk türleridir. Örnek vermek gerekirse müşteri şikayetleri, çalışanlar tarafından yapılan işlemlerin riski, çağrı merkezlerinde yapılan işlemlerden dolayı oluşan risk, Kişisel Verileri Koruma Kanunu çerçevesi dışında yapılan işlemlerin riski gibi. Bu risk türleri iç denetlemelerle ve teftiş birimlerinin müdahaleleriyle minimuma indirgenebilir. Dış risk tipi için ise mevcut finansal piyasaların durumu, ülke yönetiminin etkisi ve Merkez Bankası politikalarını söyleyebiliriz (Teker, 2006).

2.2.2. Karşılaşılan Risk Türleri

Bankacılıkta risk ikiye ayrılmaktadır. Sistemik risk ve sistemik olmayan risk. Sistemik risk, sistem ve ekonomik koşullardan dolayı eldeki mevcut kıymetlerin değerinin değişme olasılığıdır. Sistemik olmayan risk ise bu koşullardan bağımsız bankadan bankaya bağımsızlık gösterebilen risklerdir (Mandacı, 2003).

Temel olan risk kavramlarına örnek vermek gerekirse; Kredi riski, piyasa riski ve operasyonel, davranışsal ve çevresel riskler (Anderson, 2001). Bankacılıkta risk yönetimini sağlamak için bankanın nasıl bir risk türü ile karşılaştığını öğrenmek gerekmektedir. Aşağıdaki kısımda kısa bu risk türlerini inceleyeceğiz. Bu çalışmada risk türleri 6 farklı kategoride incelenmektedir.

2.2.2.1. Kredi Riski

Kredi riskinin basit olarak açıklamak gerekirse kredi çeken müşterinin aldığı krediyi geri ödememesinden dolayı oluşan risk olarak söyleyebiliriz. Bu durumda banka verilen tutarın bir

kısmını veya tamamını kaybedebilir. Bunun sonucunda, oluşan risk durumu kredi risk başlığı altında incelenmektedir (Bolgen ve Akçay, 2009).

Kredi riski bankalar için risk türlerinin başında gelir. Tüketicilerin kredi çekerek yaptıkları anlaşma sonucu geri ödemeleri paranın faiziyle olur. Eğer bu geri ödeme işlem yapılmazsa kredi riski artar, ödenmeyen kredilerin oluşturduğu yük bankadaki paranın akışını bozabilir ve bankanın ekonomik olarak kötü bir şekilde etkilenmesine sebep olabilir. Bu ödenmeyen ve takip edilen kredilere de Temerrüt Kredileri olarak adlandırılmaktadır. Ayrıca bu ödenmeyen kredilerden oluşan riskleri etkileyen başka faktörler de bulunmaktadır (Yarız, 2012).

Ülkelerin mali durumları ve piyasalardaki durgunluk bankaların kredi risk yönetimi için daha fazla çaba harcamalarına neden olur. Bu durgunlukları şu sebeplerle açıklayabiliriz. Artan küresel ekonomik yarışla birlikte şirketler de daha fazla risk almaktadır. Şirketlerin alacağı bu türdeki finansal riskler de direkt olarak olmasa da bankaları da etkilemektedir. Çünkü ticari olarak yaptıkları yatırımların finansmanı bankalar tarafından karşılanmaktadır (Brown ve Moles, 2014).

Bankalar ayrıca kredi riskini yönetirken müşterilerin ne kadar sadık olduğuna da bakarlar. Geçmiş ödemelerin de herhangi bir problemle karşılaşmayan müşteriler için sonraki yaptığı kredi işlemlerinde de ödemelerinin yapıldığını tespit edip kredi analizlerini bu hesaplamalara göre yapabilirler. Eğer müşterinin gecikmeli ödemesi var ise yeni bir kredi çekme ihtimali daha az olasıdır. Ayrıca kredi işlemlerinde kefil gösterilirse o kişinin anapara veya faizi geri ödeme gücüne sahip olduğu kanıtlanabilir (Yılmaz, 2011).

2.2.2.2. Likidite Riski

Likidite, bankanın nakit olarak piyasa içinde ödemelerinin karşılayabilme gücüdür (Aloğlu, 2005). Likiditesi yüksek olan bir banka piyasa içindeki belirsizliklerden daha az etkilenebilir. Likidite riski ise bankaların nakit ihtiyacı doğması durumunda borçlanmak için başka enstrümanlar denemesi ve elindeki varlıkları, zararlı likit ihtiyacını karşılabilmek için nakite dönüştürme zorunluluğu olarak anlatılabilir (Aloğlu, 2005). Başka bir ifadeyle likidite riski bankaların taahhütlerini zamanında ve ya makul bir zamanda karşılayamama tehlikesidir (Altıntaş, 2006). Diğer bir çalışmaya göre ise bir bankanın ödeme gücü olmasına rağmen belirli bir zamandaki nakit sıkıntısı sebebiyle ödemesini gerçekleştirememesidir (Frexias ve Rochet, 1999).

BDDK'ya göre piyasa riski 'bankanın öngörülebilir ya da öngörülemez tüm nakit akışı gereksinimini, günlük operasyonları ya da finansal yapıyı etkilemeksizin gereğince karşılayamaması sonucu bankanın zarar etme olasılığını', piyasa likidite riski ise 'yetersiz piyasa derinliği veya piyasa şartlarının bozulması gibi nedenlerle piyasa fiyatını etkilemeden bir pozisyonun satılamaması veya kapatılamaması ile herhangi bir nedenle bir pozisyonun piyasa fiyatının oluşmaması riskini' şeklinde tanımlanmaktadır (BDDK, 2016).

Likidite riski bankaların kısa vadeli finansal talepleri karşılayamadığı ve bu nedenle para arzındaki kıtlıktan dolayı oluşabilecek riski ifade eder. Bu risk sonucunda oluşabilecek durumlar duran varlıklar, beklenmeyen gelir kaybı, ödemelerin gecikmesinden dolayı oluşan nakit kaybı gibi durumlardır. Bankaların uğraşması gereken üç tip kredi riski vardır (Buch, 2015).

- Merkez bankası likidite riski,
- Fonlama kredi riski,
- Piyasa likidite riski.

Merkez bankası likiditesi, finansal ortam için gerekli olan likiditenin merkez bankası tarafından arz edilebilme kapasitesidir. Merkez bankasındaki para tabanı değişimi baz alınarak ölçümü yapılmaktadır (Yıldırım, 2011).

Fonlama likiditesi, aracı kurumun nakit akışlarındaki uyumsuzluğun sözleşmeye bağlı koşullara yerine getirmesini engellemesi anlamına gelmektedir. Yüksek maliyetli fonlar sebebiyle fonlama likiditesi riski oluşmaktadır (Şimşek, 2007).

Piyasa likiditesi ise bir varlığın kısa sürede ve düşük maliyetle fiyatı en az etkilenecek şekilde alınıp satılması sonucu oluşan likiditedir. Herhangi bir nedenle bu varlığın satılamaması durumunda piyasa likiditesi riski oluşmaktadır (Yıldırım, 2011).

2.2.2.3. Piyasa Riski

Uluslararası Ödemeler Bankası, bankaların sermaye yeterliliğini ölçmek için İsviçre'de kurulmuş bir bankadır ve ülkelerin merkez bankalarıyla koordineli bir şekilde çalışmaktadır. Bu kuruma göre, negatif yöndeki hareketlenmelerden kaynaklanan bilanço içi ve bilanço dışı

pozisyon kayıpları piyasa riski olarak adlandırılmaktadır. Tüketicilerin piyasadaki varlıklarını yönetmek için yaptığı işlemler piyasalardaki genel durumu da etkileyebilir, belirsizlik yaratabilir ve bankaları riskle karşı karşıya bırakabilir. Bu türdeki risklere piyasa riski diyebiliriz (Lam, 2014).

Piyasaların mevcut istikrarını değiştiren ve piyasa üzerinde yarattığı belirsizlikte finansal kuruluşlar için finansal istikrarı bozacak ve risk oluşturabilecek değişken etkenler arasında aşağıdaki maddeleri sıralayabiliriz (Rankovic 2017).

- Faiz oranları,
- Sabit Döviz Kurları,
- Döviz Kurlarındaki Dalgalanmalar,
- Hisse Fiyatlarındaki Değişkenlik,
- Stoklardaki Dalgalanmalar.

Diğer risk türleri de piyasa riskini tetikleyebilir. Örneğin uluslararası kredi derecelendirme kuruluşları ülkenin kredi notunu düşürdüğünde yatırımcı varlıklarının likiditeye dönüştürmek isteyebilir. Bu şekilde yüksek likite sahip varlıklar satılmak istendiği zaman piyasanın dengesini değiştirebilir (Rankovic, 2017).

Piyasa riskini tahmin etmek veya proaktif davranmak için ilk yapılması gereken piyasa riskinin alt türünü belirlemektir. Bu şekilde bankalar piyasa riskinden dolayı oluşan kaybı en aza indirgeyebilir. Paranın oynaklığının artması daha fazla piyasa riskine sebep olur ya da hisse senetleri ne kadar az dalgalanırsa risk o kadar azalır. Bu göstergeleri takip edip ona göre firmaların piyasadaki pozisyonunu belirlemesi gerekmektedir.

2.2.2.4. Operasyonel Risk

Operasyonel risk, kurum içinde her şekilde oluşabilecek hatalardan, denetimsizlik dolayı oluşan usulsüzlerden, kurum çalışanlarının kurallara uygun hakeret etmemesinden kaynaklanan risk türüdür. Örnek verecek olursak, teknoloji yönetim sistemlerindeki aksaklıklar operasyonel risk başlığı altında tanımlanabilir (Uysal, 2009).

Bankacılık dünyası açısından operasyonel risk ise yanlış hesaplanan finansal varlıklar nedeniyle iflas etme riski ve beklenmeyen durumlar sonucu oluşan zararlar olarak açıklanabilir. Bankaların sermaye yeterliliğini denetleyen Basel 2 komitesi ise operasyonel riski “Zarar Riski” olarak tanımlamaktadır (BDDK, 2001). İnsanlardan, sistemlerden veya harici durumlardan dolayı oluşabilir. Riskin asıl sebebi insan olsa da sistemler, programlar, finansal dönemler, müşteri tercihleri, ticari büyüme, hızlı gelişen teknoloji ve bazı dış durumlar da bu riskin kaynağı olarak gösterilebilir (Beja, 2016).

Operasyonel riskler iki farklı seviyede ortaya çıkmaktadır: Birincisi, teknik bilgi sistemindeki eksiklikler; ikincisi, organizasyon seviyesi, raporlama ve izleme ile ilgili tüm kurallar ve politikalara uyulmaması.

2.2.2.5. Faiz Riski

Faiz riski, genel seviyedeki faiz oranları ile değişik faizlerin birbirlerine oranının değişmesinden dolayı kaynaklanan bir risk türüdür (Bessis, 2011).

Faiz riski kavramını daha da açacak olursak, işletmelerin ve bankaların karşılaştıkları faiz oranları değişiklikleri sebebiyle sermaye değerlerinin ve yatırım değerlerinin kaybı, nakit akışlarının aksaması, sonraki vadeler de ele geçecek likit değerinde yaşanabilecek azalış, kazanç miktarındaki düşüş gibi risklerle karşılaşmasıdır (Candan ve Özün, 2006).

Faiz riski, bankaların aktif ve pasif yönetimini risk altına sokan en önemli etkenlerden biridir. Oranın yükselmesiyle mevduat maliyetlerini arttırabilirken, oranını düşmesiyle de bankadaki mevduat sahibi müşterilerin kazançlarını azaltabilir ve mevduatlarını başka bir yatırım kaynağı olarak düşünmelerine sebep olabilir. Bu sebeple banka içindeki nakit miktarı azalacaktır. Faiz riski aşağıdaki değişkenlere bağlı olarak değişebilir (Altıntaş, 2006):

- Piyasalardaki fon arz talebine,
- Pazarın ekonomiden beklentilerine,
- Devletin ve işletmelerin izlediği politikalara,
- Borçlanma likiditesine.

Yukarıdaki deęişkenleri düřündüğümüz zaman bir bankanın faiz riskinden tamamen korunabileceğini söylemek mümkün deęildir (Altıntaş, 2006).

2.2.2.6. Bilanço Riski

Bilanço riski, bankaların finansal yapısındaki aktif ve pasif varlıkların dengede olmaması ile řirketlerin karşılaştığı bir risk türüdür (Öker, 2007). Pasifleri elde tutmanın maliyeti ve gelir sağlayacak aktiflerin vade yapılarının dengesiz olması, döviz riskinden kaynaklanan riskler, sermaye yönetimi gözetmeksizin alınan kararlar sonucu bankalar özkaynaklarından daha fazla borçlanmaya girebilir. Bu borçlanma sonucu oluşan risk bilanço riskidir. Temel neden aktif pasif varlıkların yönetiminin oluşabilecek deęişmelere karşı iyi yönetilememesidir. Aktif varlıkların büyümesiyle pasif varlıkların dengesinin nasıl kurulacağı en çok dikkat edilmesi gereken konudur. Maliyetlerin belirlenmesi de bu dengenin kurulması ve dalgalanmaların takibi de bilanço riski yönetiminin önemli maddelerinden birisidir. Aktif ve pasifler arasındaki vade uyumsuzluğu, varlıkların farklı para birimleri üzerinden olması, mali yapılarıdaki bozulma sonucu oluşan problemler banka genelinde krize neden olmaktadır. Faiz oranlı sözleşmelerin vadeli yapılması, bankaların piyasalarla yaptığı sözleşmeler, finansal enstrümanların karşılıklı deęişimi ve ya belirli bir fiyat üzerinden anlaşılması řirketleri bilanço riskinden koruyabilecek önlemlerdir (Öker, 2007).

2.3. MAKİNE ÖĞRENMESİ

Makine öğrenmesi, verinin anlaşılabilir ve anlamlı sonuçlara dönüřtürülmesini sağlayan kod algoritmalarının geliřtirmesi şeklinde tanımlanmaktadır (Cui, Wong ve Lui, 2006). Makine öğrenmesi; ulaşılabilir veri, istatistiksel metotlar ve gelişen hızlı hesaplama gücü olmak üzere üç ana başlık altında incelenmektedir (Lantz, 2013). Söz konusu üç başlık, birbirlerine baęlı olarak ilerlemekte ve büyümektedir.

Makine öğrenmesi, geçmiş veri seti veya örnek bir veri üzerinden bir performans kriterinin optimize edilmesidir (Alpaydın, 2004). Söz konusu verilere ait parametreleri optimize eden bir model geliřtirilip; bu model veri hakkında bilgi sağlamak veya veri üzerinden tahminler yapmak, bazen de her ikisi için kullanılabilir (Canepa, 2016).

2.3.1. Makine Öğrenmesi Kavramı

İnsanlığa günlük işlerinde yardımcı olmak için üretilen bilgisayarlar, başlangıçta sadece kendilerine verilen, önceden belirlenmiş görevleri yapacak şekilde tasarlanmışlardır. Yıllar içinde yaşanan gelişmelerle bilgisayarlar verilen işlerin yapılmasının yanısıra bu işlerle ilgili verilerin toplanması ve depolanmasını da sağlayacak şekilde geliştirilmiştir. Ancak depolanan bu veriler işlenip anlamlı hale getirilmedikçe, bir veri çöplüğü olmaktan öteye geçemiyordu. Bu ihtiyaç, bilgisayarların insanlar gibi eğitilerek; verinin içindeki örüntüleri bulan ve veriyi bir anlam ifade edecek hale getirmesi anlamına gelen "Makine Öğrenmesi" kavramını oluşturmuştur (Emir, 2013).

Öğrenme işini gerçekleştirenin makine veya insan olmasından bağımsız olarak, süreç üç temel adımda gerçekleşmektedir (Shwartz ve David, 2014):

- Girdi Veriler: Bu aşamada verinin daha sonra kullanılmak için hafızaya alınması sağlanır.
- Çıktı Veriler: Verinin geniş anlamda sunumuna imkân sağlayacak dönüşümün yapılmasıdır.
- Genelleştirme: Üzerinden karar alınabilecek hale getirilmiş verinin kullanılmasıdır.

Makine öğrenmesinde tasarlanacak model makinenin kararına bırakılmamaktadır. Eldeki geçmiş veri ve makinenin öğrenmesi amaçlanan görev modeli şekillendirmektedir (Barnes, 2015).

Veri seti üzerinden model geliştirmesi süreci eğitim, geliştirilen modelin gelecek verilere uygulanması süreci ise öğrenme olarak adlandırılmaktadır (Witten, Frank ve Hall, 2011). Makine, geçmişteki verilerin tamamı ya da bir kısmı üzerinden eğitime tabi tutulur. Eğitim sırasında kendisine sunulan verilerin içerisindeki yapıyı anlamaya çalışarak bir algoritma oluşturur. Oluşan bu algoritma gelecek veriler veya test verileri ile sınanarak sonuca ulaşılmaktadır.

2.3.2. Makine Öğrenmesi Süreci Adımları

Makine öğrenmesi kullanılarak çözülecek bir problemin model tasarım sürecinde izlenmesi gereken belirli adımlar vardır. Problemin istenilen şekilde ve başarılı bir şekilde çözülmesini sağlamak için bu adımları uygulamak büyük önem taşımaktadır. Bu süreçte uygulanacak adımlar aşağıdaki gibidir (Shearer, 2000):

- Problemin Tanımlanması,
- Veri Toplanması,
- Verilerin Hazırlanması,
- Modelin Kurulması,
- Modelin Değerlendirilmesi,
- Modelin Kullanılması.

2.3.2.1. Problemin Tanımlanması

Makine öğrenmesi sürecinde yapılması gereken ilk iş problemin doğru ve açık bir şekilde tanımlanması; ulaşılması istenen amacın net bir şekilde belirtilmesidir. İyi şekilde tanımlanmamış ve amacı belli olmayan bir proje; üzerinde çalışan araştırmacıyı karar alma konusunda belirsizliğe sokacaktır. Bu durum istenilen sonuca ulaşılmasını zorlaştırıp modelin tamamlanmasını engelleyecektir.

Bu adımda problemin ve mevcut durumun tanımıyla birlikte, ulaşılması istenen durum ve başarı kriterleri de açık şekilde belirlenmelidir. Makine öğrenmesi sonucunda alınan çıktıların doğru yorumlanabilmesi için bu adımın yukarıda anlatılan şekilde yapılması büyük önem taşımaktadır.

2.3.2.2. Veri Toplanması

Makine öğrenmesi sürecindeki ikinci adım probleme uygun verilerin toplanmasıdır. Bu aşamada gerekli veriler iki şekilde elde edilebilir:

- Mevcuttaki veri ambarları veya veri tabanlarından problemle ilgili istenilen veriler,

- Araştırmacının problemle ilgili olarak anket, ölçümler vb. çalışmalar yoluyla elde edeceği veriler kullanılabilir.

Araştırmacı bu aşamada edindiği verileri genel anlamda incelemelidir. Verinin formatı, veri sayısı gibi konular araştırmacı tarafından gözden geçirilmelidir. Ayrıca modelde kullanılacak olan bir parametrenin, toplanmış veri tarafından karşılanıp karşılanmadığı kontrol edilmelidir. Örneğin; yaş aralığının önemli olduğu bir çalışmada, toplanan veri tüm yaş aralıklarını yansıtmıyorsa; veri seti değiştirilmeli ya da tüm yaş aralıklarını yansıtacak şekilde eklentiler yapılmalıdır (Shearer, 2000).

2.3.2.3. Verilerin Hazırlanması

Modelin kurulmasında kullanılacak veri toplandıktan sonra makine öğrenme algoritmalarının kullanılabilmesi hale getirilmesi gerekmektedir. Bu amaçla eldeki veri bir takım ön işlemlerden geçirilmelidir. Bu işlemler bütünleştirme, dönüştürme ve indirgeme olarak tanımlanmaktadır. Eldeki veri setine göre bu işlemlerden hangilerinin kullanılacağına karar verilmelidir.

Veri temizleme işleminde iki ana işlem uygulanmaktadır. Her veri setinde hatalı ve tutarsız veriler bulunabilir. İlk olarak bu verilerin temizlenmesi işlemi bu aşamada yapılmaktadır. İkinci olarak kayıp değer olarak adlandırılan verilerin anlamlı bir ifadeye dönüştürülmesi; eğer bu mümkün değilse veri setinden atılması gerekmektedir. Kayıp değerler çalışmayı etkilemeyecek seviyede iste veri setinden çıkarılabilir. Ancak kayıp değerlerin kattığı anlamın fazla olması durumunda bu değerlere en çok tekrarlanan değer atanarak veya regresyon ile uygun değer bulunarak bu verilerin doldurulması sağlanabilir.

Veri dönüştürme ve bütünleştirme, veri seti içindeki ifadelerin tekil bir şekilde dönüştürülmesi aşamasıdır. Veri seti içinde aynı anlama gelen değerler farklı ifadelerle belirtilmiş olabilmektedir. Bu durum makine öğrenmesi aşamasında algoritmanın bunları farklı değerler olarak algılayıp işlem yapmasına sebebiyet vereceğinden; tüm farklı değerlerin, tek tip ifadeye dönüştürülmesi sağlanmalıdır.

Veri indirgeme, verilerin daha hızlı ve kolay işlenmesini sağlamak için sonuçları değiştirmeyecek şekilde veri sayısının veya değişkenlerin azaltılmasıdır. Eldeki veriye göre farklı veri indirgeme teknikler kullanılabilir.

Veri hazırlamada ön işlemler yapıldıktan sonra, verinin öznitelik kontrollerinin yapılması gerekmektedir. Öznitelik kontrollerinde normalizasyon ve veri ölçekleme öne çıkan işlemlerdendir.

Normalizasyon: Veri setindeki her bir sayısal değer farklı değer aralıklarında değişmektedir. Aralıkları büyük olan verilerin sonuç üzerindeki etkisi de büyük olacağından, veri normalizasyonu yapılmalıdır. Bu sayede değişkenlerin etkisi düzenlenerek en iyi sonuç hedeflenmektedir. Normalizasyonda, Z-Dönüşümü, 0-1 aralığı, minimum-maksimum gibi farklı metotlar kullanılmaktadır.

Veri Ölçekleme: En basit anlamıyla tüm verilerin sabit bir sayıya bölünmesi olarak tanımlanabilir. Ölçeklemede her bir veri grubu için değerler, o veri grubunun aralığına bölünerek; yeni bir aralık bulunmaktadır. Bu sayede yakın değerlere sahip giriş değerleri aynı aralıkta tutularak daha anlamlı bir veri seti oluşturulması sağlanır.

Veri Örnekleme: Dengesiz veri setlerinde veri örnekleme işlemleri yapılarak algoritmaların düzgün dağılım yapması sağlanır. Aksi takdirde modellerin doğruluk derecesi inandırıcı olmayan sonuçlar verebilir (He ve Garcia, 2009). Veri örnekleme iki şekilde yapılır. Veri üzerindeki hedef değişkenin daha fazla olduğu çoğunluk sınıfı ile azınlık sınıfının değerlerinin dengeli bir hale getirilmesi hedeflenir. Veri örnekleme iki şekilde yapılır.

Alt örnekleme(undersampling) yönteminde çoğunluk sınıfı olan veri sayısı, azınlık sınıfındaki veri sayısına yakınsamaya çalıştırılır. Bu azaltma işlemi rastgele yapılması halinde Rastgele Alt Örnekleme(Random undersampling) olarak adlandırılır (Shelke, Deshmukh ve Shandilya, 2017).

Üst örnekleme yönteminde ise azınlık sınıfı çoğunluk sınıfına yaklaştırılmaya çalışılır. Eğer bu işlemde azınlık sınıfı için sentetik olarak veri üretilirse SMOTE (Synthetic Minority Oversampling Technique), rastgele veriler azınlık sınıfında üretilirse Rastgele Üst Örnekleme(Random Oversampling) olarak tanımlanır (Chawla, Bowyer, Hall ve Kegelmeyer, 2002).

2.3.2.4. Modelin Kurulması

Veri seti üzerinde hangi yöntemin ve algoritmanın en iyi sonuca ulaştıracağına, deneme yanılma yöntemiyle karar verilmesi süreci model kurulumu olarak adlandırılmaktadır (Hornick, Marcadé ve Venkayala, 2007). Bu aşamada veri setinde uygulanmak üzere birden fazla algoritma ve teknik belirlenerek model üzerine uygulanarak en iyi sonuca ulaşılmaya çalışılır. Uygulanacak algoritma ve teknik sayısı arttıkça model verimliliği de doğru orantılı olarak artacaktır.

2.3.2.5. Modelin Değerlendirilmesi

Problemin tanımı aşamasında belirlenen hedef ve amaçların ne kadarının karşılandığının ölçüldüğü; modelin kalite ve etkinliğinin, model kullanıma geçmeden önce değerlendirildiği aşamadır. Probleme özellikle belirtilen noktaların ne denli dikkate alındığı ve modelin bu problemlere uygulanabilir bir çözüm getirip getirmediği ile ilgili net kararlar verilmelidir (Demircioğlu ve Bilge, 2015).

Bu aşamada model kurulumunda kullanılan farklı algoritma ve tekniklerin sonuçları karşılaştırılarak en uygun model seçimi yapılır. Seçilen modelin kendi içinde tutarlılığı test edilip; modelin bir genellemeye dönüştürülüp dönüştürülemeyeceği değerlendirilir. Bu değerlendirmeyi yapmaya yarayan yöntemlerden biri de k-kat çapraz değerlendirmedir. Bu yöntemde veri kümesinde k eşit parçaya bölünür. K adet testin her bir parçasında bir küme test verisi k-1 küme ise eğitim verisi olarak kullanılır. Yapılan k adet testin her birinden bulunan hata oranı genele yansıtılarak genel modelin hata oranı bulunur.

2.3.2.6. Modelin Kullanılması

Problemin çözümü için geliştirilip test edilen model kullanıma alındıktan sonra mevcut problem için kullanılacak ana modeldir. Probleme benzer olan konularda yapılan araştırmalarda geliştirilen model kullanılıp istenilen sonuçlar alınabilir. Modelin uzun vadedeki başarı oranı, bu süre içindeki kullanım oranına bağlı olarak artacaktır.

2.3.3. Makine Öğrenmesi Yöntemleri

Makine öğrenmesi yöntemler sınıflandırması çalışmalar arasında farklılık göstermiştir. Temelde gözetimli ve gözetimsiz olmak üzere iki grup altında incelenmektedir. Ancak

teknolojinin ilerlemesi ile yıllar içinde bu iki gruba bazı eklemeler yapılmıştır. Gelen eklemeler, yarı denetimli, pekiştirmeli, akıl yürütme, çevrimiçi ve aktif öğrenme olarak sınıflandırılabilir. Güncel olarak makine öğrenmesi yöntemleri denetimli, denetimsiz, yarı denetimli, pekiştirmeli öğrenme olarak 4 başlık altında sınıflandırılmaktadır (Brownlee, 2014).

2.3.3.1. Denetimli Öğrenme

Denetimli öğrenme önemli ölçüde eğitim verilerinden bir model sağlar. Potansiyel ve belirsiz veriler üzerinde tahminlerde bulunuruz. Bu yaklaşım bir öğretmenin öğrenciye öğretmesine benzer. Öğretmen öğrenciye iyi örnekler sunar ve öğrenci de bu örnekleri ezberleyerek seçilen örneklerin üzerinden genel kurallar çizer ve sonrasında yapılacağı sınavlardan bu ezberlediği örnekten çıkardığı kurallara göre yorum yapabilir.

Denetimli makine öğreniminde, bu girdilerden bilgi çıkarmak için sisteme girdi olarak giren dataların etiketlenmesi gerekir. Girdiler ve bunlara verilen etiketler arasındaki ilişkiler denetimli makine öğrenme tekniklerinde incelenmektedir. Sınıflandırma, her kayıt için verilen etiketin bu kaydın ait olduğu sınıfı temsil ettiği, en çok kullanılan denetimli veri madenciliği tekniklerinden biridir. Ardından, sınıflandırıcılar özniteliklerin o kaydı karakterize eden değerleri ile kaydın üyesi olarak etiketlendiği sınıf arasındaki ilişkileri çıkarır. Bu bilgi daha sonra onlar için bir sınıfı öngörmek üzere sınıflandırılmamış yeni kayıtlara uygulanır. Bu tahmin, o sınıftaki kayıtların genel özelliklerine bağlı olarak, o yeni kaydın gelecekteki davranışını tahmin etmeye yardımcı olabilir (Suthaharan, 2014).

2.3.3.2. Denetimsiz Öğrenme

Denetimsiz makine öğrenmesi, olduğu gibi veri kümelerinden bilgileri çıkarmak için kullanılır; burada, çıkartılmış bilginin veri kümesindeki kayıtların özellik değerleri arasındaki ilişkileri temsil ettiği görülür. Bu tür bir makine öğrenmesi, veri setindeki kayıtlara etiketleme gerektirmez, çünkü bu algoritmalar, her bir kaydı karakterize eden değerlere bağlı olarak, kayıtlar arasındaki ilişkileri bulma eğilimindedir. Kümeleme, veri kümesindeki kayıtların, belirtilen değerlerine göre gruplara dağıtıldığı, denetlenmeyen en popüler veri madenciliği tekniklerinden biridir. Bu gruplarda, her kayıt, bu gruplardaki diğer kayıtlara, diğer gruplardaki diğer kayıtlardan daha benzer. Böylece kümeleme homojen kayıt grupları oluşturulur (Liu ve d'Aquin, 2017).

Ek olarak, grup sayısı kümelenme sürecinin performansını etkileyen ana faktördür ve burada daha fazla sayıda küme, kümelerdeki kümelenmelere gerçek bir fayda olmadan, veri kümesindeki kayıtları işlemek için gereken süreyi arttırır. Daha az sayıdaki kümeler ise anlamsız kümeler oluşturulabilir. Bu nedenle, veri kümesindeki öznitelik değerlerinin dağılımına bağlı olarak kayıtları optimum küme sayısına kümelemek önemlidir. Bu optimal sayı, insanlar tarafından kümeleme algoritmasına sağlanabilir veya bazı optimal sayıdaki küme seçim teknikleri kullanılarak, bu teknikler, belirli bir faktöre bağlı olarak, bu veri kümesi için optimum küme sayısını seçmek amacıyla her küme seçim tekniği test edilir (Hançer ve Karaboga, 2017).

2.3.3.3. Yarı Denetimli Öğrenme

Yarı denetimli öğrenme, hem etiketlenmiş verileri hem de etiketlenmemiş verileri içerir. Denetimli ve denetimsiz öğrenme, yarı denetimli öğrenmenin iki parçasıdır (Dasgupta ve Nath, 2014).

2.3.3.4. Pekiştirmeli Öğrenme

Pekiştirmeli makine öğrenmesi, diğer makine öğrenimi görevlerinden oldukça farklıdır. Öğrenim sistemine ajan denir ve öğrenilen stratejiye politika denir. Sistemin verdiği kararlara dayanarak, ödül ve ceza alır. Pekiştirmeli makine öğrenmesi yürüme robotlarında kullanılır (Géron, 2017).

2.3.4. Makine Öğrenmesi Algoritmaları

Makine öğrenmesi insanların gerçek problemleri çözmek için makineye açıkça bir kodlama yazmadan nasıl yapılacağını gösterdiği yapay zekanın bir alt dalıdır. Genellikle bu öğrenme çeşitlerinin üstünden yukarıdaki konularda bahsetmiştik. Makine öğrenmesinde genellikle kullanılan algoritmalar aşağıdaki gibidir.

- K-NN Algoritması,
- Naive Bayes Algoritması,
- Karar Ağacı Algoritması,

- Rastgele Orman,
- Destek Vektör Makineleri,
- Lojistik Regresyon Algoritması,
- Yapay Sinir Ağları.

2.3.4.1. K-NN Algoritması

K-NN algoritması K En Yakın Komşu Algoritması olarak da adlandırılır. En basit öğrenme algoritmalarından biridir. Sınıflandırma problemlerinde, K kullanıcı tanımlı bir parametredir. Bir sorgu örneğinin etiketini tanımlamak için, sorgu örneği ile etiketlerin bulunduğu eğitim örnekleri arasındaki mesafe hesaplanır. Sorgu örneğinin etiketi seçilen K değerine bağlıdır. Örneğin K 1 olarak alınmışsa sorgu örneğinin etiketi en yakın eğitimin etiketiyle aynı olacaktır. K 1'den büyükse, sorgu örneği ile eğitim arasındaki mesafeler bilinen etiketlere sahip örnekler hesaplanır. Sonra sorgu noktası olarak sınıflandırılır. Sorguya en yakın olan K eğitim numarası en yakın komşu sınıfına atanır. Eğitim örneklerinin çok boyutlu uzayda etiketli vektörler olduğuna dikkat edilmelidir. Bir en popüler mesafe fonksiyonlarından biri Öklid mesafesidir (Mitchel, 1997). Bir K'ye en yakın komşu algoritmasında oluşabilecek sorun overfitting sorunudur. Aşırı uyum gösterme(overfitting), eğitim sürecinin eğitim için kullanılan veri kümesindeki değerlere fazlasıyla bağlı kalarak öğrenen yöntemin genelleme yapma yeteneğinin düşmesi durumudur (Tuffery, 2011).

En yakın komşu algoritmaları henüz belirlenmemiş özelliklerdeki verileri kendilerine en yakın özelliklerdeki verilerle aynı sınıflara atar (Gaganis, Pasiouras, Spathis ve Zopounidis, 2007). Bu belirttiğimiz tanım basit olarak görünebilir ama en yakın komşu algoritmaları birçok sektörde kritik işler üzerinde kullanılmaktadır.

- Fotoğraf ve medya oynatım gibi görsel uygulamalarda,
- Ses tanıma sistemlerinde,
- Kullanıcılara uygun tercih yapılma işlemlerinde(dizi, film),
- Hastalıkların önceden tanımlanabilmesi için genetik verilerin sınıflandırılmasında

başarıyla kullanılmaktadır (Lantz, 2013; Nahayo ve Arı, 2016).

En yakın komşu algoritmasının güçlü yönleri olarak basit bir algoritma olduğunu söyleyebiliriz. Öğrenme aşaması gayet hızlıdır. Verinin dağılımıyla ilgili herhangi bir tahminlemede bulunmaz. Zayıf yönlerine baktığımız zaman ise veriler arasındaki sınıflandırmada yeni çıkarımlarda bulunmaz. Sınıflandırma aşaması yavaştır. Bozuk veriler için verilerin düzeltilmesi ve tekrardan işlem yapılması gerekir (Lantz, 2013). Bu duruma örnek vermek gerekirse nominal veriler olarak tanımlanan verilerin sayısal veri olarak değiştirilmesi gerekebilir. Bu değiştirilme işlemi kukla kodlama adıyla yapılmaktadır.

2.3.4.2. Naive Bayes Algoritması

Naive Bayesian yaklaşımı, olasılıkları hesaplamak için kullanılan iki rastgele olayın koşullu olasılıklarıdır. Maksimum olasılık prensibine dayanan bir teoremdir. Naive Bayes, Bayes'in olasılık teoreminden elde edilen bir makine öğrenme algoritmasıdır. Yapay zeka alanında Naive Bayesian modeli olarak tanınır (Monhamady, 2018). Model basit ve yapımı kolaydır ancak özellikler öngörücüleri arasındaki bağımsızlık varsayımına dayanır.

Naive Bayes algoritmasını kullanan sınıflandırmalarda, eğitim verisindeki değerler gözlemlenerek ve veri setininin değerleri temel alınarak gözlenmiş olasılıklar her bir sınıf için hesaplanır. Test verisi üstünde çalıştığı zaman bu verilerin özelliklerini sınıflandırırken daha önce test verisi üstünde hesaplanmış olduğu olasılıklara dayanarak sınıflandırmasını gerçekleştirmektedir. Naive Bayes algoritması basit bir algoritma olarak bilinse de kompleks algoritmaların sonuçlarıyla benzer düzeyde sınıflandırma yapabilmektedir (Cichosz, 2015; Barber, 2012).

Naive Bayes algoritması günümüzde aşağıdaki işlemler için kullanılabilir.

- Yazı sınıflandırma işlemleri için (mail, kitap, yazar sınıflandırma),
- Network üzerindeki normal olmayan giriş ve işlemlerin tespitinde,
- Finans sektöründe tahminleme işlemlerinde,
- Gözlemlenen hasta verilerine dayanarak hastalığın önceden belirlenmesinde

başarıyla kullanılmaktadır (Bell ve Bala, 2015).

Naive Bayes algoritmasının temel olarak formülünü açıklayacak olursa aşağıdaki şekilde gösterebiliriz.

$$\text{Naive Bayes} = P(A|B) = P(B|A) * P(A) / P(B)$$

$P(A|B)$ A olayının B şartlı olasılığı

$P(B|A)$ = B olayının A şartlı olasılığı

$P(A)$ = A olayının gerçekleşme olasılığı

$P(B)$ = B olayının gerçekleşme olasılığı

Naive Bayes algoritmasının güçlü yanlarından bahsedecek olursak, ilk olarak basit, hızlı ve iyi sınıflandırma sonuçları verebilen bir algoritma olduğunu söyleyebiliriz. En yakın komşu algoritmasının tersine kayıp ve bozuk verilerle ilgili daha iyi sonuçlar verebilir. Eğitim verisinde çok sayıda örnek ya da az sayıda örnek olması Naive Bayes algoritmasının sonuçlarını diğer algoritmalara göre daha az etkiler. Naive Bayes algoritmasının zayıf yönlerini açacak olursak sayısal verilerin fazla olduğu örnekler için en uygulanabilir algoritma değildir. Tahmin edilen olasılıklar, test verisinden çıkan sınıflara göre daha az güvenilirdir (Lantz, 2013; Han ve Kamber, 2006; Cichosz, 2015).

2.3.4.3. Karar Ağacı Algoritması

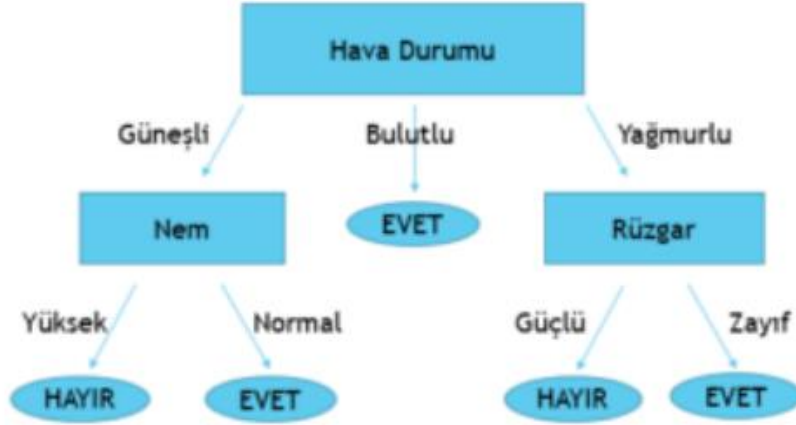
Karar ağacı, verilerin dağılımına dayalı bir makine öğrenme prosedürüdür. Karar ağacı algoritması denetimli bir makine öğrenme algoritmasıdır. Çalışılan probleme göre sınıflandırma metodu veya ilgili bir regresyon tahmincisi olarak görev işlevini yapabilirler. Bir kök düğümden başlar ve verileri dallara ayırarak bir yaprak düğüme ulaşana kadar örneklerin bölüdüğü verilere dayanır. Kök düğümün yani ağacın en üst noktasındaki veri ağaçtaki akışa göre sınıflandırılarak ilgili veriye ait yaprağa ulaşır ve sınıflandırması karar ağacı modeli ile yapılmış olur (Witten, Frank ve Hall, 2011; Zafarani, Abbasive ve Liu, 2014). Karar ağacı yinelemeli bölümlenme yaparak ilerler ve veri setini en küçük parçalara ayırarak bölünmeyi gerçekleştirir.

Karar ağacı algoritması bir akış olarak ilerlediğinden dolayı karar alma mekanizmalarının şeffaf olması gerekmektedir. Bu sebeple karar ağacı yöntemi aşağıdaki işlemler için başarıyla uygulanmaktadır (Bell, 2015; Lantz, 2013).

- Kullanıcıların kredi risk modellerindeki skorlarının ölçülmesi,
- Finans sektöründeki fiyatlama departmanlarının yaptığı çalışmalar,
- Müşteri memnuniyet ve şikayeti üzerinden yapılan müşteri ilişkileri yönetimi,
- Tıp alanındaki laboratuvar çalışmaları.

Karar ağacı algoritması çok çeşitli alanlardaki farklı ihtiyaçları giderebilen bir yöntemdir (Tuffery, 2011). Sayısal ve bozuk verilerle de sınıflandırma yapabilen bir yöntemdir. Akış şemasında düğümler arasında en önemli belirleyici özellikleri kullanır. Model sonuçlarının incelenmesi ve çıkarımlar yapılması diğer algoritmalara göre daha basittir. Ayrıca diğer modellere göre veri üstünde daha etkili sonuçlar çıkarabilmektedir. Karar ağacı yöntemine sahip modellerde verilerin en önemli özelliklerini ve özniteliklerini baz alarak sınıflandırma yaptığından dolayı çok sayıda farklı özelliklere sahip veriler için sonuçları isabetli olmamaktadır. Modeldeki akış üzerinde yaşanan bir sorun ile birlikte overfitting ya da underfitting sorunları ile karşılaşmaktadır. Eğitim verisinin üzerinde yapılan değişikliklerle algoritma sonucunda büyük değişiklikler yaşanmaktadır. Geniş veya çok sayıdaki özelliklere sahip ağaçlarda sonuç beklendiği gibi çıkmayabilir ve yorumlanması kolay olmamaktadır (Ocak, Şeker, 2013; Tuffery, 2011; James, Witten, Hastie ve Tibshirani, 2013).

Örnek olarak bir karar ağacı şeması gösterimi aşağıdaki şekilde olmaktadır. Bu gösterimde hava durumu verisinin nasıl dallanabileceği gösterilmektedir.



Şekil 2.6: Karar ağacı yapısı (Argüden ve Erşahin, 2008).

2.3.4.4. Rastgele Orman Yöntemi

Rastgele Orman, torbalama üzerine geliştirilen bir topluluk yöntemidir (Breiman, 2001). Torbalı ağaçlara benzer şekilde, çok sayıda ağaç oluşturur ve nihai kararı vermek için verimlerini toplar. Ancak, torbalanmış ağaçların aksine, ağaç oluşturulurken, her bölme için rastgele bir özellik örneği seçilir ve bu bölme için yalnızca bu özellikler kullanılabilir. Bu rastgele özellik seçimi, bazı özelliklerin diğerlerine hakim olmasını engeller ve ağaçları birbirinden ayırarak varyansı azaltmaya yardımcı olur. Rastgele orman algoritması Leo Breiman tarafından geliştirilmiştir (Breiman, 2001).

Aşağıdaki adımlar rastgele orman algoritmasının adımlarıdır.

- Verilerinden bir bölüm seçilir.
- Seçilen örnekler ve nitelikler dikkate alınarak en iyi ayırım noktası Gini Safsızlık Kriteri baz alınarak belirlenir.
- Ayırım noktasından itibaren iki yeni düğüm oluşur.
- Hedef düğüm sayısı için önceki iki adım sırayla tekrarlanır.
- Hedef ağaç sayısı için önceki maddeler sırayla tekrarlanır.

Random forest yönteminin üstün yönlerinden bahsedecek olursak çok fazla sayıda kayıp ve bozuk veri ile de başarılı sonuçlar verebilmektedir. Kategorik veya kategorik olmayan iki veri

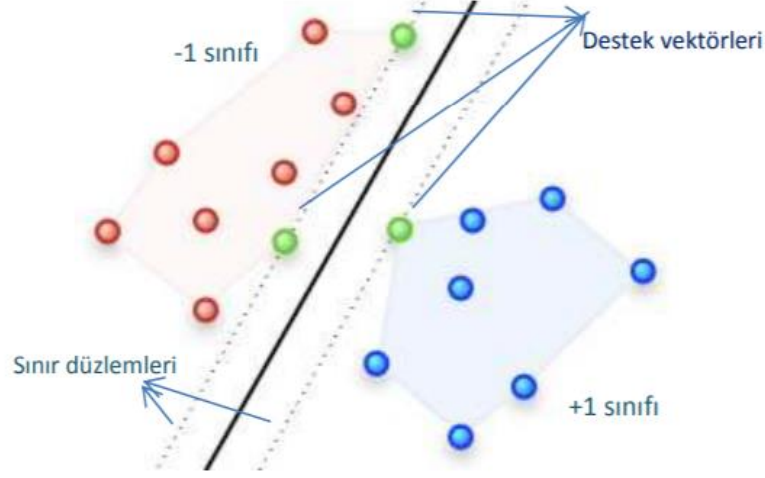
türü için de başarılı bir algoritmadır. Karar ağacı yöntemine nazaran çok daha fazla sayıda değişken içeren örnekler için de uygun bir yöntemdir. Sınıflara için olan örnek sayısı gerçekçi değilse, sınıfları daha dengeli bir şekilde dağıtabilir. Ayrıca değişkenlerin önem derecesi de hesaplanabilmektedir. Zayıf yönlerini sayacak olursak bazı model sonuçlarında oluşan güven aralığı (yapay sinir ağları, lojistik regresyon) rastgele karar ağacında verilmemektedir. Bu yöntemin düşük bellekli bilgisayarlar üstünde uygulanması zordur. Ayrıca model sonucu işlemlerin adımları model karmaşıklığından dolayı görülmemektedir (Akman, Genç ve Ankaralı, 2011).

2.3.4.5. Destek Vektör Makineleri

En popüler öğrenme algoritmalarından biri, optimum bir hiper düzlem bularak verileri iki kategoriden ayıran Destek Vektör Makineleri (SVM) 'dir. Optimum hiper düzlem, iki sınıfta birbirlerine en yakın nokta olan destek vektörleri kullanılarak oluşturulur. Hiper düzlem ve destek vektörleri arasındaki boşluğa marj denir ve SVM daha iyi ayrılma için maksimum marjı bulmayı amaçlar (Ng, 2000).

Destek Vektör Makine algoritmasının avantajlarından bahsedecek olursak modelde kareli optimizasyon yaparak sınıflandırmayı daha basit bir şekilde yapmaktadır. Bunun sonucunda öğrenme verisi üstünde yapılan işlemler daha azaldığı için sonuca diğer makine öğrenme yöntemlerine göre daha hızlı ulaşmaktadır (Osowski, Siwekand ve Markiewicz, 2004). Büyük hacimli veriler üstünde de rahatlıkla çalışabilen bir algoritmadır. Ayrıca sonuç hesaplama değerlerine baktığımız zaman diğer algoritmalara göre daha iyi performans gösterir (Nitze, Schulthess ve Asche, 2012).

İki boyutlu bir sınıflandırma için Destek Vektör Makine algoritmasının iki boyutlu düzlemde gösterimi aşağıda paylaşılmıştır. Destek Vektör Makine algoritması destek vektörleri arasındaki yer alan hiper düzleminin bulunmasını amaçlamaktadır.



Şekil 2.7: 2 sınıflı problem örneği (Ayhan ve Erdoğan, 2014).

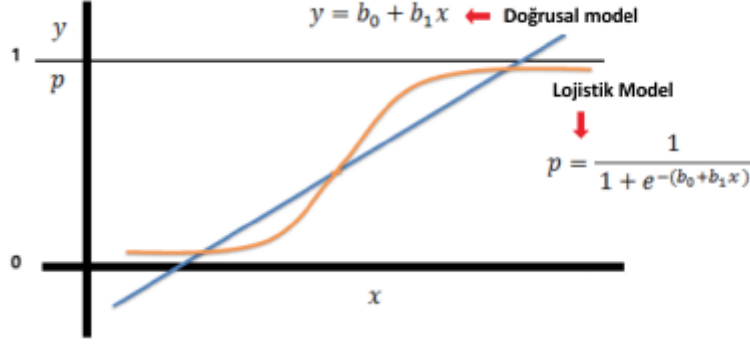
2.3.4.6. Lojistik Regresyon Algoritması

Lojistik Regresyonun sınıflandırma algoritmalarında kullanılan ana algoritmalarından biridir. Lojistik regresyon ile ilgili literatür, bir model olarak Lojistik Regresyon'un sınıflandırma ve tahmin olaylarında yaygın olarak kullanıldığını iddia eder (Thomas, 2000; Menard, 2002). Dolayısıyla, hedef değişkenin iyi olma ihtimalinin doğrusal olmayan bir işlevi olduğu doğrusal bir regresyondur (Thomas, 2000). Ayrıca, çalışmaya göre, Lojistik Regresyon modelinin sınıflandırma sonuçları bağımsız değişkenler arasındaki korelasyona duyarlıdır. Regresyon katsayıları genellikle maksimum olabilirlik tahmini kullanılarak tahmin edilir (Menard, 2002).

King ve Zeng (2001) tarafından yazılan makalede, birçok araştırmacı olayların nadir olduğu veriler için geleneksel lojistik regresyonun yasal olarak kullanılıp kullanılmayacağından endişe duyduğu belirtilmektedir. Buradaki zorluk özellikle olayların nadirliği değil, iki sonucun olağandışı olması üzerine birkaç vaka olması ihtimalidir.

Lojistik regresyon yönteminin güçlü yönlerine bakacak olursak çok sayıda değişken türü ile çalıştırılabilen bir algoritmadır. Sınıflandırma başarısı yüksek bir yöntemdir. Örnek sayısının az olması ile de iyi performans göstermektedir. Birbirine bağımlı olmayan değişkenler arasında da iletişimi mümkün kılmaktadır. Model sonucu güven aralığı verebilmektedir. Lojistik regresyon yönteminin zayıf yönleri olarak, model üzerindeki açıklayıcı değişkenler bağımsız olmalıdır. Bozuk ya da kayıp veriler sonucu algoritma çalışamaz duruma gelebilir. Değerlerin çok yüksek ya da fazla olmasında karşı duyarlıdır (Tuffery, 2011).

Lojistik regresyon modelinin matematiksel gösterimi aşağıdaki şekilde yapılmıştır (Kalaycı, 2018).



Şekil 2.8: Lojistik regresyon gösterimi (Kalaycı, 2018).

2.3.4.7. Yapay Sinir Ağları

Bu algoritma, bir nöronun dendritlerden girdi aldığı, hücre gövdesinde topladığı ve akson aracılığıyla diğer nöronlara çıktısını veren insan beynindeki sinir ağından ilham alır. Beyindeki nöronlara benzer şekilde, yapay bir nöron, bir dizi nörondan girdisini alır, bazı işlemleri yapar ve çıktısını verir (Budama ve Locascio, 2017).

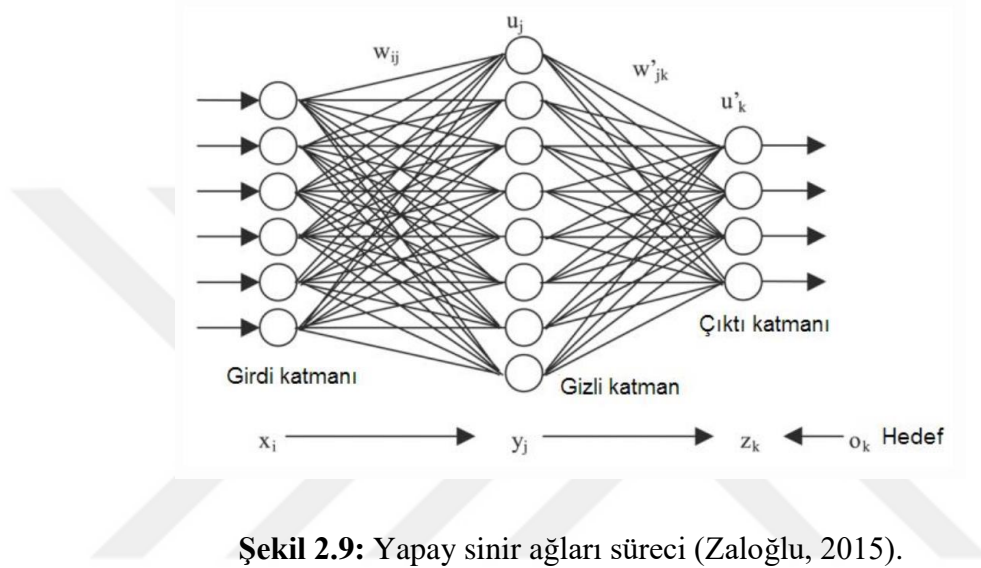
Sinir Ağları veya Yapay Sinir Ağları olarak bilinen algoritma görüntü tanıma, konuşma analizi ve metin işlemede de uygulamaların gözlendiği güçlü bir öğrenme yöntemi olarak ortaya çıktı. Yapay sinir ağları katmanlardan oluşur ve her katman, karşılığında bir aktivasyon işlevi içeren birbirine bağlı düğümlerden oluşur. Bir sinir ağında üç ana katman vardır. İlk katman, verilerin ağa beslendiği girdiler katmanı olarak adlandırılır. İkinci katman, bir veya daha fazla katmandan oluşan gizli katman(lar) olarak adlandırılan ve gerçek işlemin düğümler arasındaki bağların ağırlandırılıp bir işlemle gösterildiği katmandır. Son katman, ağın son cevabını sağlayan çıktı katmanıdır. Ağlar boyunca iletilen sinyaller ve hatalar, ağırlık katsayısını, geri yayılma olarak bilinen dinamik bir algoritma ile düzeltmek ve ayarlamak için geri gönderilir. Yapay sinir ağları sınıflandırma ve tahminleme yöntemleri için kullanılan bir algoritmadır (Lantz, 2013).

Yapay sinir ağları aşağıdaki uygulamalar için kullanılmaktadır (Tuffery, 2011).

- Sesli konuşma ve yazı sınıflandırma ve tanımlama işlemleri,

- Akıllı ofis, ev ve araçların otomasyonları,
- İklim koşullarının modellenmesi, akışkanlar dinamiği değerlendirmeleri,
- Birçok sektörde koşulların değerlendirilmesi.

Yapay sinir ağları sürecini aşağıdaki gibi bir grafikte gösterebiliriz (Zaloğlu, 2015)



Şekil 2.9: Yapay sinir ağları süreci (Zaloğlu, 2015).

3. MALZEME VE YÖNTEM

Bu bölümde çalışmada kullanılacak olan verinin toplanması, veri toplanmasında kullanılan araçlar, çalışma yapılacak kitlenin belirlenmesi ve analizi konusunda bilgiler verilmiştir.

3.1. ARAŞTIRMA MODELİ

Bu çalışma, bir bankadan elde edilen yaklaşık 400 bin kayıttan oluşan gerçek bir veri seti kullanılarak ilk kredi ödemesini yapmamış (FPD) kredisi başvuru sahiplerinin analizindeki bir veri madenciliği çalışmasıdır. Veriler, bir bankanın müşteri bilgilerinin tahminlenmesiyle oluşturulmuştur. Çalışmanın veri seti, yalnızca Ocak 2018 - Kasım 2018 arasındaki döneme ait tahsis edilen kredileri ilgilendiren temel tüketici kredileri (tüm bireysel krediler değil) bilgilerinden oluşmaktadır. Veri kümesi, her biri tüketici kredisinin ayrıntılarını temsil eden 45 sütun ve 400 bin satırdan oluşmaktadır. Her sütun, cinsiyet, yaş, eğitim, meslek, evlilik durumu, konut durumu, hanehalkı geliri ve bankanın bir müşterinin sınıflandırılmasını ve ayrıca kredi detaylarını içerir. Müşterinin ilk kredi ödememesini gösteren bilgi ise 'fpd_flag' kolonu ile belirtilmiştir. Bu kolonun değeri 1 ise tüketicinin ilk taksitte ödenmeyen kredisi vardır. 0 ise tüketicinin ödemelerini yaptığını belirtir. Tablo için örnek veri veri ön izleme bölümünde ekran görüntüsü olarak paylaşılacaktır.

3.2. VERİLERİN TOPLANMASI

Veriler bir banka veritabanı sistemi üzerinden çekilmiştir. Direkt olarak SQL programı kullanılarak veriye ulaşılmıştır. Sonrasında ise veriler üzerindeki kişisel veriler değiştirilerek müşteri bazlı herhangi bir bilginin kişisel verilerden dolayı erişilmemesi sağlanmıştır. Veriler toplanırken kullanıcıların modellerde değerlendirilmesi düşünülen tüketici kredisi bilgileri baz alınarak veri toplanmıştır. Toplanan verilerin bulunduğu tablo ilgili veritabanında tüketici bilgilerinin olduğu bir tablo olarak yer almaktadır. Örnek olarak kullanılacağı için 400000 satır alınmıştır. Verinin niteliklerini içeren bilgiler ve kolonların hepsi alınmış ve 45 kolon üzerinden işlem yapılmıştır. Kolonlar veri ön işleme metoduyla incelenmiş ve sonrasında algoritma içinde kullanılmıştır.

3.3. VERİLERİN ANALİZİ

Çalışma yapılacak veri seti üstünde en sık kullanılan ve önem arz eden kolonlar aşağıdaki bilgileri bulunduran kolonlardır.

Veri setinde en sık kullanılan kolonlar aşağıda bilgileri içermektedir.

- Müşterinin kredi talep numarası,
- Kredinin verildiği tarih,
- Kredi tutarı,
- Kredinin vadesi,
- Kredinin aylık ödeme tutarı,
- Kredinin kaç gün gecikmede olduğu,
- Kredinin statusu,
- Kredinin yüzde olarak faizi,
- Kredi talep eden tüketicinin cinsiyeti,
- Müşterinin cinsiyeti,
- Müşterinin medeni durumu,
- Müşterinin eğitim durumu,
- Müşterinin ev sahiplik durumu,
- Müşterinin bulunduğu sektör (kobi,mikro,bireysel),
- Müşterinin krediye başvurduğu kanal (şube,mobil),
- Müşterinin kredisinin gecikmede olup olmadığı durumu.

3.4. ÇALIŞMANIN AMACI VE HEDEFİ

Çalışmanın hedefi tüketici kredi başvurularının değerlendirilme aşamasında karar destek sistemlerine yardımcı olmaktır. Bu hedef doğrultusunda amaçlanan kredi başvurularının makine öğrenme algoritmaları ile değerlendirilip gecikmeye girmeyecek kredi başvurularının tahminlenmesidir. Bu hedef sonucunda aşağıdaki sonuçların çıkması beklenmektedir.

- Bankanın likidite ve kredi riskinin düşürülmesi,
- Tüketici kredisine başvuran kişilerin borçlanmasının engellenmesi,
- Finansal sistemdeki kredi başvurularının kişi bazlı değil kriterler baz alınarak değerlendirilmesi.

3.5. ÇALIŞMADA KULLANILACAK YÖNTEMLER

Verilerin tahminlenmesinde iki algoritma kullanılacaktır. Bu algoritmalar karar ağacı ve lojistik regresyon algoritmalarıdır. Bu iki algoritmayı çalışmada kullanmamızın sebebi sahip olduğumuz veri seti üzerinde rahatlıkla sınıflandırma yapabilecek yöntemler olmalarıdır. Ayrıca bu algoritmaların güçlü yönlerine bakacak olursak model sonuçlarının yorumlanması daha basittir ve bozuk veriler üstünde de etkili bir şekilde çalışabilirler. Ek olarak bu algoritmaların sınıflandırma başarılarının da yüksek olduğunu söyleyebiliriz.

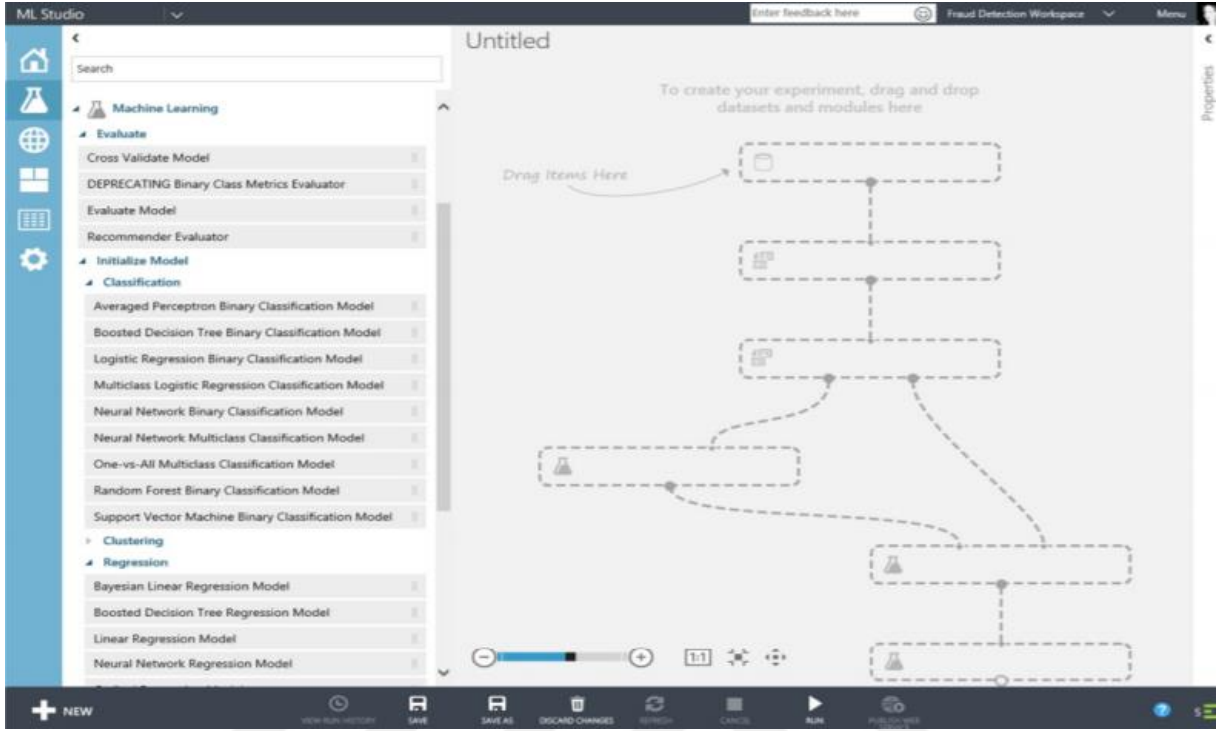
3.5.1. Azure Machine Learning Studio

Veri üstünde çalıştırılacak algoritmalar Azure ML Studio üzerinden çalıştırılmaktadır. Azure ML Studio, kullanıcıların makine öğrenme sürecini yürütmelerine yardımcı olan bir bulut hizmetidir. Adından da anlaşılacağı gibi halka açık bir bulut platformu olan Microsoft Azure'da çalışıyor. Bu nedenle, çok büyük miktarda veriye dünyanın her yerinden erişilebilir. Azure ML için internet tarayıcı üstünden erişim sağlanabilir (Chappell, 2015).

Azure ML Studio'nun kullanıcılara sağladığı hizmetleri aşağıdaki gibi sıralayabiliriz.

- Veri ön işleme modülleri,
- Makine öğrenme algoritmaları,
- Kodlamalarla daha önce Azure ML üstünde yapılan modellemelerden yapılan uygulamalar,
- Akışların başından sonuna kadar kontrol edilmesini sağlayan grafiksel araçlar.

Azure ML Studio'nun önyüzü aşağıdaki gibidir (Chappell, 2015) .



Şekil 3.1: Azure ML Studio önyüzü (Chappell, 2015).

3.5.2. Veri Ön İşleme

Araştırma projesinin ilk aşamasında, hedefler arasında veri temizleme ve veri kalitesi değerlendirmesi yer almaktadır. Toplanan verilerin analizinde genel bir veri profilinin ve kalite değerlendirmenin oldukça önemli olduğunu belirtmekte fayda var. Çalışmada kullanılan veri setinin kalitesi çok iyi değil. Kayıp değerleri ve bazı veri hataları var. Veri ön işleme aşaması, sonuçların güvence kalitesi için önemlidir. Düşük kaliteli veri, tüm çabalarda yanlış sonuçlara yol açacaktır. Bu nedenle modelleri uygulamadan önce veriler temizlendi.

Aşağıda aşama aşama veri ön işleme ve temizlenmesinin detayları belirtilmiştir.

- İlk adım, % 25' ten fazla eksik değere sahip özellikleri kaldırmaktı ve bunun sonucunda 5 değişken kaldırıldı. İş Telefonu, İşin Şehri, Net Aylık Harcama, Askerlik Durumu, Personel Girişi Kredi Kodu gibi bilgileri içeren değişkenler eksik değerlerden kurtulmak için düşürüldü. Bu adımdaki amacımız modelin tahmin performansını etkileyecek olan bilgileri çıkartmaktır. Eğer bu bilgiler çıkarılmayıp yerinde kukla

değişkenler atama işlemi yapılırsa bu kukla değerler algoritma sonucu çıkan değerleri şaşırtabilirler.

- Kredi Tipi, Müşteri Tipi ve İzlenen Kredi Tipi sütunları tüm krediler için aynıdır, dolayısıyla hariç tutulurlar. Bu bilgileri çıkarmamızın sebebi verimizde sadece tüketici kredi bilgilerinin olmasıdır. Farklı türde bir kredi tipi olursa ancak bu bilgilerin analize eklenmesi mantıklı olabilir.
- Kredinin Açık Kapalı Durumu, Kredinin Gecikme Gün Sayısı, Kredinin Hangi aşamada olduğunu içeren sütunlar, önceden müşterilerin kredileriyle ilgili bilgilerini gösteren değişkenlerdir. Yeni bir tüketici krediye başvurduğunda bu bilgilere sahip değildir. Bu nedenle bu bilgiler de veri üzerinden çıkarılır.
- Kredinin Faizli Tutarı ile Kredi Tutarı bilgileri aynı bilgiyi vermektedir ve bu iki kolon birbiriyle yüksek oranda ilişkilidir.
- Tüketicilerin başvuru yaptığı kanallar üzerinden yapılan araştırmalarda da başvuru yapılan kanal bilgilerinin paralel olarak yazıldığı değişkenler bulunmaktadır. Burdaki gerekli olmayan değişkenler de çıkarılmıştır. Bu değişkenler Kredinin Distribütör Kodu, Kredi Başvurusunun yapıldığı kanal, Kredinin Türü, Kredinin Kanal Kategori Kodu gibi bilgilerdir.
- Bazı eksik değerleri Bilinmeyen / Diğer değeri ile değiştirdik veya özelliğin sayısal anlamına gelecek şekilde değişmesini sağladık.
- Kayıp veya boş olan veriler 'Diğer' olarak değiştirildi ya da sayısal veya kategorik hale gelecek şekilde değiştirildi.
- Müşterinin İkametini içeren veriler içindeki eksik değerler '0' ile değiştirilmiştir.
- Kredinin Başvurulduğu Tarih ve Kredi Başvurusunu alan çalışan bilgisi de sonuç ile herhangi bir ilişki içermediğinden dolayı düşürülmüştür.
- Şubeler üzerinde başvuru kredi sayısı yaklaşık değerdedir bu nedenle başvuru yapılan şubenin bilgisinin istatistiksel olarak veri üzerinde yer alması anlamsızdır ve çıkarılmıştır.

- Çok fazla kategoriye sahip olan deęişkenlerin kategori sayısı arttırılmıştır. Kredinin Geciken Gün Sayısı ve Vade bilgisini içeren deęişkenler manipüle edilmiştir.
- Bazı deęişkenlerin içedięi deęerler çok sayıda olmaktadır. Bu nedenle eşit bir aralıkta deęişkenleri deęerlendirebilmek için veri normalleştirme işlemi yapılmıştır. Zscore yöntemi ile normalleştirme yapılarak bu çok sayıdaki deęerlerin 0 ile 1 arasında bir deęer alması sağlanmıştır. Bu deęişkenlere örnek olarak Kredinin Aylık Ödeme Bilgisi, Tüketicinin Aylık Geliri, Kredinin Faiz oranı, Kredinin Tutarı gibi bilgileri içeren deęişkenleri belirtebiliriz.



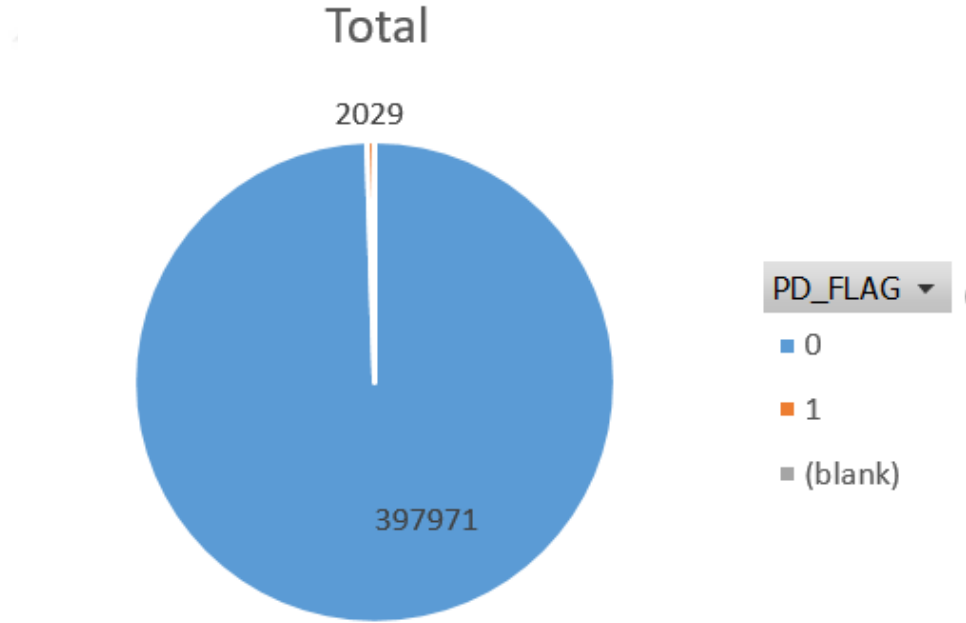
4. BULGULAR

Bu bölümde yapılan veri analizleri neticesinde analizlerin gerçek verilerle karşılaştırmasına ve verilerin makine algoritmaları ile işlenmesinin sonuçlarına yer verilmiştir.

4.1. ÖDEMESİ YAPILMAYAN KREDİ ORANLARI

Araştırmada kullanılan tüm değişkenlerin tanımlayıcı analizi, en ayırt edici ve alakalı değişkenleri tanımlamak için veri analizinin değişkenler bazında yapılması gereklidir.. Tanımlayıcı analizden sonra hangi özelliklerin modele dahil edileceği planlanabilir. Ödenmeyen kredi sayısının (Hedef değişken), zamanında / erken ödenmiş kredi sayısından çok daha düşük olduğu kabul edildi.

Hedef değişkenimiz olan FPD_FLAG kolonunun değerlerinin veri setimizden incelediğimiz zaman aşağıdaki sonuçlara ulaşabilir. Bildiğimiz gibi veriler dengesiz ve aşağıdaki grafik de bunu doğruluyor.



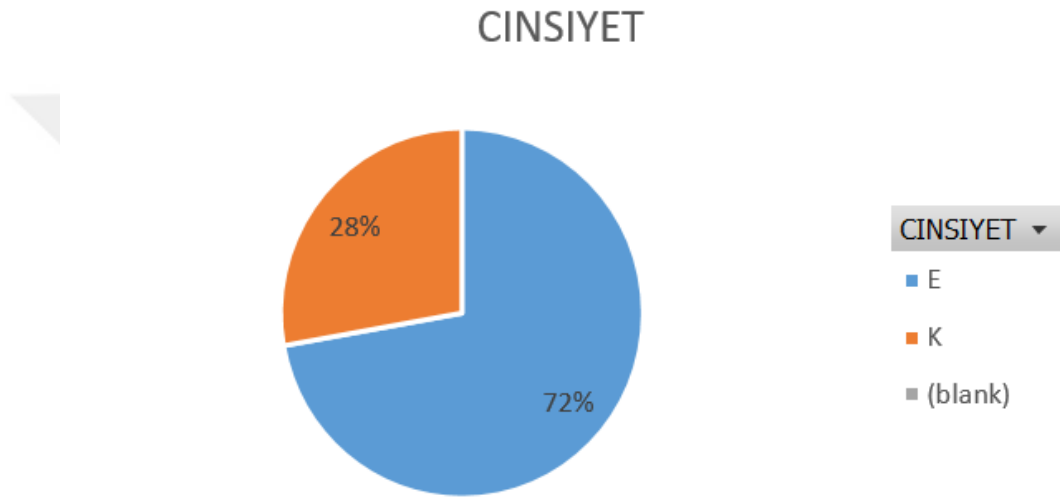
Şekil 4.1: FPD oran grafiği.

Verilerde sadece %0,05 FPD Kredisi varken %99,5 Normal Kredidir, veri kümesi kredi ödemesini yapmayan yani FPD_F=1 olan tüketicilerin çokluğundan dolayı dengesizdir. ML algoritmalarını uygulamadan önce, veri kümelerini eşitlemek için veri örnekleme yapılmıştır.

4.2. KREDİ BAŞVURUSU YAPAN TÜKETİCİ ANALİZİ

Makine öğrenme sürecindeki adımlar dikkate alınarak verilerin genel olarak analizleri yapılmıştır (Shearer, 2000). Tüketici ile ilgili bilgi içeren kolonların incelemeleri ile veri setinin keşfedici incelemesinden, kredilerin özelliklerini karakterize edebilecek bazı keşifler aşağıda gösterilmiştir.

- Kredi başvurusunda bulunanların çoğunluğu (yaklaşık %72) erkektir.

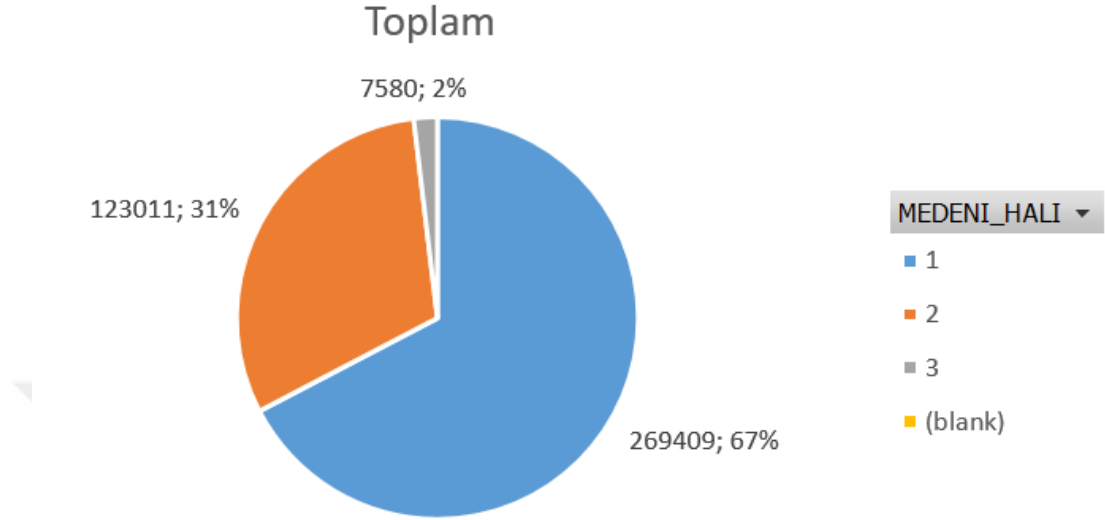


Şekil 4.2: Kredi başvuru cinsiyet oranı.

TÜİK tarafından yayımlanan Şubat 2018 tarihinde yayınlanan İşgücü İstatistikleri'ne göre, işgücüne katılım oranı erkekler için %71,5, kadınlar için %33,3 idi. Bu istatistik, erkeklerin aile ekonomisini yönettiğini ve erkek egemen toplumun bir yansıması olduğunu göstermektedir (TÜİK, 2018).

- Kredi başvurusunda bulunanların % 67'den fazlası evlidir.

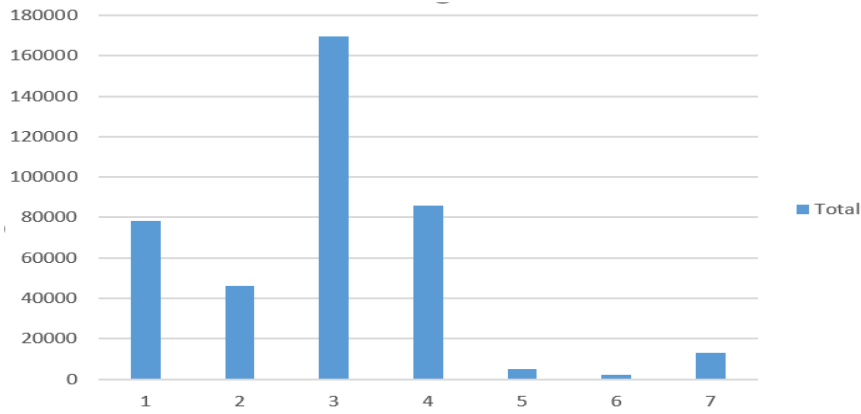
Şekil 4.3’de 1 rakamı evli, rakamı bekar, 3 rakamı da boşanmış kişilerin yapmış oldukları kredi başvurularını göstermektedir.



Şekil 4.3: Kredi başvuru medeni durum oranı.

TÜİK’in yayınladığı sonuçlara göre 2017 yılında, 15 yaşından büyük nüfusun yaklaşık 61 milyon olduğu ve bu rakamın %63,37’sinin evli olduğu bilinmektedir (TÜİK,2017). Bu oran kredi başvurusunda bulunanlarla aynı değerdedir.

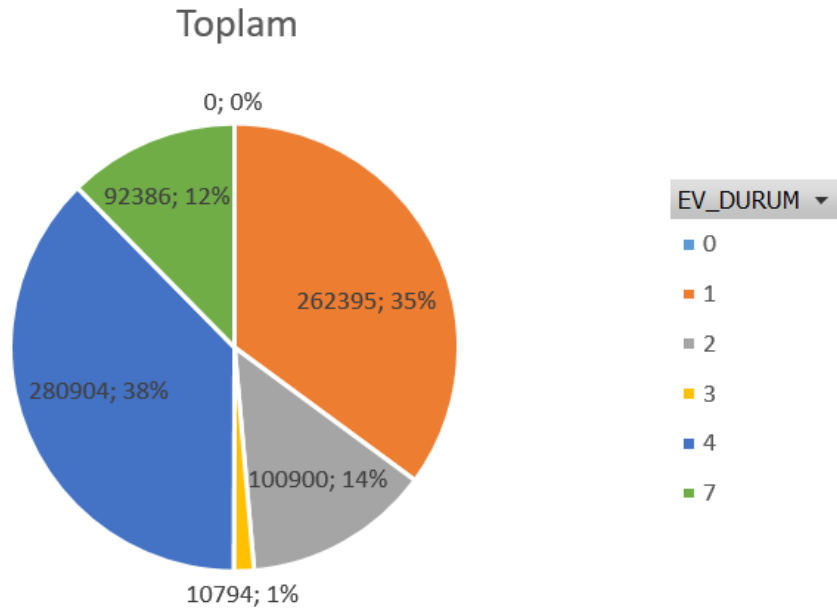
- Kredi başvurularının yaklaşık %42’si lise mezunudur. Şekil 4.4’te 2 ve 3 numaralar sırasıyla ortaokul mezunu ve lise mezunu kişilerin yaptıkları kredi başvuru rakamlarını göstermektedir.



Şekil 4.4: Kredi başvuru eğitim grafiği.

TÜİK tarafından sağlanan Milli Eğitim İstatistikleri Veri Tabanı'na göre 2017 yılında, iki yüksek oranlı eğitim seviyesi istatistiği, % 38 ile İlköğretim veya İlköğretim ve % 23 ile Yüksek ve Meslek Lisesi'nden oluşmaktadır (TÜİK, 2017). Bir kez daha, kredi başvurusunda bulunanların eğitim oranları Türkiye istatistiklerini temsil etmektedir.

- Kredi başvurusunda bulunanların yaklaşık % 60'ı kendi evinde yaşıyor.



Şekil 4.5: Kredi başvuru ev durum grafiği.

Veri analizi, bankanın müşterilerini daha iyi anlamasını sağlar ve ve müşterilerin ortak özelliklerini özetlemektedir. Bu özellikler bankaya ve müşteriye özeldir. Ayrıca, veri madenciliğinin sonuçlarını açıklamaya ve yorumlamaya yardımcı olur. Ayrıca, gelecekteki

finansal ürün tasarımı için kullanışlıdır. Veri kümesinin yapısına daldığımız ve krediler ile değişkenler arasındaki ilişkiyi incelediğimiz söylenebilir.

4.3. MAKİNE ÖĞRENMESİ SONUÇLARI

Genel olarak, çalışmalar sınıflandırma algoritmalarının dengeli bir eğitim seti üzerinde eğitildiğinde daha iyi performansa sahip olduğunu göstermiştir. Kullandığımız veri setine baktığımız zaman ise verimizin dengesiz olduğunu söyleyebiliriz. Bu nedenler veri kümesini daha da dengeli hale getirmek için örnekleme yöntemleri kullanılır. Bu tekniklerle veri üzerinde değişkenlerin kaldırılması veya eklenmesi yapılırken sınıf değişkenlerinin değeri dikkate alınmaz. Bu yöntemlerin uygulanması diğer algoritmalara göre daha basittir. Bu örnekleme yöntemleri veri örnekleme konusu altında alt örnekleme ve üst örnekleme olarak değinmiştik.

4.3.1. Değerlendirme Ölçütleri

Birden fazla model kullandığımız için herbir model farklı sonuçlar verecektir. Bu nedenle modellerin sonuç performanslarını da incelemek gerekecektir. Oluşan sonuçlar verinin uygulanabilirliğine ve algoritma yöntemine bağlıdır. Uygulamada, modeller eğitim verileriyle eğitilir ve test verileri ve karışıklık matrisleriyle değerlendirilir. Tahmini performans, hepsi bir karışıklık matrisinde bulunan Recall ve Precision gibi doğruluk ölçümleri kullanılarak değerlendirilir. Karışıklık matrisi unsurları şu şekildedir;

Tablo 4.1: Karışıklık matrisi unsurları.

Veri Seti	Model Tahminleme	
	FPD	FPD Olmayan Krediler
FPD	GP= Gerçek Pozitif FPD olarak tahminlenen FPD kredi sayısı	YN= Yanlış Negatif FPD olmayan kredi olarak tahminlenen FPD kredi sayısı
FPD Olmayan Krediler	YP= Yanlış Pozitif FPD olarak tahminlenen FPD olmayan kredi sayısı	GN= Gerçek Negatif FPD olmayan kredi olarak tahminlenen FPD olmayan kredi sayısı

Her modelin performansı, model sonucunda oluşan doğru tahminler hesaplanarak yapılmıştır. Algoritmalarda hesapladığımız doğruluk bize modelin doğru tahminlerinin bir kısmını verir (Doğru tahminlerin sayısı / Toplam tahminlerin sayısı). Sahip olduğumuz veri setinin performansları ölçülürken verinin dengesizliğini de göz önünde bulundurmanız gereklidir. Bu sebeple farklı ek ölçümlere de bir göz atıyoruz. Modelleri, kesinlik ve duyarlılık gibi bilinen ve en sık kullanılan önlemlere ve bunları F1-Skoru gibi birleştiren istatistiklere dayanarak karşılaştırdık. Bu metriklerin her biri Gerçek Pozitif (DP), Yanlış Pozitif (YP) ve Yanlış Negatif (FN) oranları kullanılarak hesaplanır. Hesaplamalar aşağıdaki şekilde yapılmaktadır.

- Doğruluk= $(GP+GN) / (GP+GN+YN+YP)$
- Kesinlik = $GP / (GP + YP) \rightarrow$ Pozitif olarak tahmin edilen bir durumdaki başarıyı gösteren durum
- Duyarlılık = $GP / (GP + GN) \rightarrow$ Pozitif durumların ne kadar başarılı tahmin edildiğini gösterir.
- F1 Score= $2 * ((Kesinlik * Duyarlılık) / (Kesinlik + Duyarlılık)) \rightarrow$ F1 skoru hem kesinlik hem de duyarlılık oranlarını dikkate alır ve daha dengeli bir görünüm sunar.

4.3.2. Değerlendirme Adımları

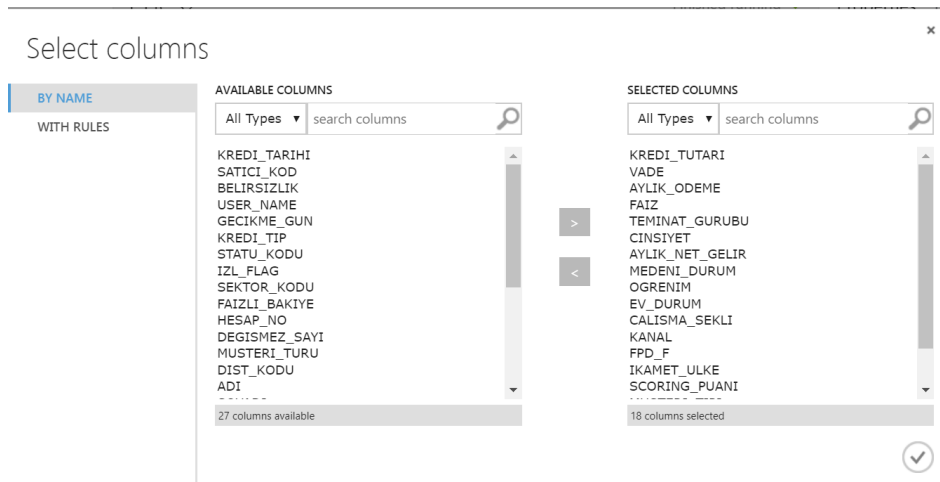
Bu başlık altında Azure ML üstünde algoritmaların nasıl uygulanacağı adım adım başlıklar halinde açıklanmıştır.

Data Visualize: Azure ML Studio üzerinde 'new' butonundan veri kümemiz ilgili ortama alınır ve veriye sağ tıklayarak Visualize seçilir. Bu sayede verinin ön izlemesi ve tutarlılığı kontrol edilebilir.

KREDI_TUTARI	VADE	AYLIK_ODEME	SATICI_KOD	BELIRSIZLIK	USER_NAME	GECIKME_GUN	KREDI_TIP	STATU_KODU	IZL_FL
4000	36	160.72	418-14	K	FINANS	-5	IHT-35	KTA	0
2699	24	139.28	153-351	K	FINANS	54	IHT-867	KTA	0
2000	36	81.29	209-14	K	FINANS	11	IHT-2001	KTA	0
21000	48	647.3	433-14	K	FINANS	0	IHT-2048	KTA	0
10267.11	48	324.46	1148-270	K	FINANS	0	IHT-1129	KTA	0
2999	24	154.76	153-351	K	FINANS	54	IHT-867	KTA	0
1300	24	68.02	279-351	K	FINANS	50	IHT-867	KTA	0
5000	12	477.81	686-14	K	FINANS	0	IHT-2031	KTA	0
3500	9	424.48	17-335	K	FINANS	0	IHT-833	KTA	0
7500	48	234.64	78-291	K	FINANS	4	IHT-28	KTA	0

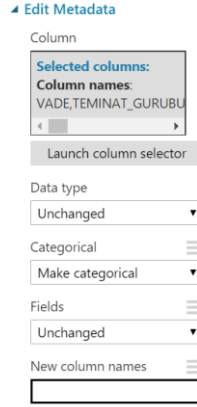
Şekil 4.6: Veri kümesi önizleme.

Select Columns in Dataset: Veri Ön işleme adımındaki bağımsız değişkenlerin tahminlenmesinde bir payı olmayacağı düşünülen değerlerin algoritmaya alınmaması için oluşturduğumuz adımdır. Bu adımda tüm kolonlardan seçili olacakları “Selected Columns” bölümüne ekliyoruz.



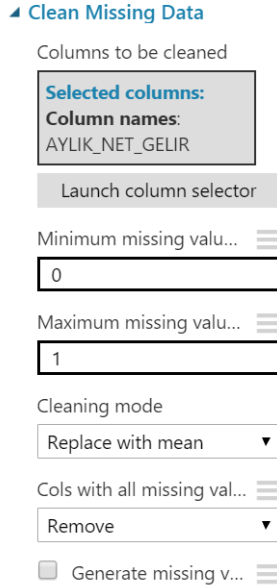
Şekil 4.7: Seçilen kolonlar gösterimi.

Edit Metadata: Kategorik olmayan ama kategorik olması gereken değerleri kategorik hale getiriyoruz. Bu adımı yapmamızın sebebi kullandığımız algoritmalarda kategorik değişkenlerin daha doğru sonuç vermesini sağlamaktır.



Şekil 4.8: Veri kategorikleştirme ekran görüntüsü.

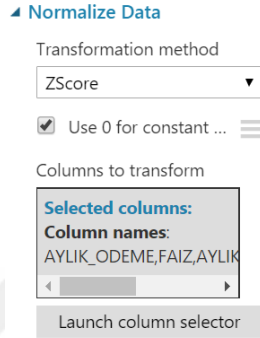
Clean Missing Data: Sayısal değer olan net gelir bilgisinin boş olduğu durumlarda verisetinin genel ortalaması ile boş olan veriler doldurulmuştur. Boş içeren verilerin tahminlemede doğru sonucu etkileyeceği düşünüldüğü için bu adım uygulanmıştır. Minimum missing value değeri 0 maximumu missing value değeri 1 olarak atanarak hiç boş değer kalmaması sağlanmıştır.



Şekil 4.9: Boş değerlere atama yapma ekran görüntüsü.

Normalize Data: Veri ön işleme adımında uygulamada bir teknik olan normalizasyon yönteminin uygulandığı adımdır. Normalleştirilmenin amacı veri aralığındaki sayısal sütunların değerlerini tüm değerler aralığından farkları bozmadan ya da bilgi kaybetmeden ortak bir ölçek kullanmak üzere değiştirmektir. Bizim veri setimizde de birbirini domine edebilecek sayısal

veriler olduğundan, kullandığımız algoritmaların verileri doğru şekilde modelleyebilmesi için normleştirme de gereklidir. Veri setindeki tüm sayısal veriler için uygulanmalıdır. Normleştirme yapılırken Zscore methodu kullanılmıştır. Bu metod tüm değerleri Z puanına dönüştürür. Bu yöntemde ortalama ve standart sapma her sütun için ayrı ayrı hesaplanır ve ana kütle standart sapması kullanılır.



Şekil 4.10: Veri normleştirme ekran görüntüsü.

İlgili veri kümemizdeki sayısal değerleri içeren aylık ödeme, aylık gelir, faiz ve kredi tutar bilgileri Zscore değerine göre normalize edilmiştir. Tablo 4.2 ve Tablo 4.3’de normleştirilen verilerin aralıklarının nasıl değiştiği görülebilir.

Tablo 4.2: Veri normleştirme öncesi değerler.

	Toplam değer	Farklı olan değerler	En küçük değer	En büyük değer	Ortalama değer	Ortalama sapma	Standart sapma
Kredi tutarı	40000	35362	496	2000000	9.239	7699,260.88	13.621,9756
Aylık ödeme	40000	77251	19,68	233.306,1	449	312,027.051	1019,579.55
Faiz	40000	47	0	1,99	1,455.173	0,189672	0,293737
Aylık gelir	40000	6245	0	391.800	2892,662.99	1593,151.19	3.608,118.40

Tablo 4.3: Veri normalleştirme sonrası değerler.

	Toplam değer	Farklı olan değerler	En küçük değer	En büyük değer	Ortalama değer	Ortalama sapma	Standart sapma
Kredi tutarı	40000	35362	-0,641832	146,143.52	0	0,56521	1.000.001
Aylık ödeme	40000	77251	-0,420786	228,386.01	0	0,306035	1.000.001
Faiz	40000	47	-4.954.002	1,820.769	0	0,645721	1.000.001
Aylık gelir	40000	6245	-0,801711	107,786.87	0	0,441547	1.000.001

Normalleştirme işlemi sonrasında sayısal olarak tutulan tutar ve faiz gibi değerler ortalama değeri 0 ve standart sapması 1 olacak şekilde değiştirilmiştir. Bu işlemle birlikte birbirinin ölçeğini ezebilecek rakamlar yerine ortalama bir ölçek kullanılıp sonucun daha isabetli olması sağlanmıştır.

Split Data : Modeldeki verileri eğitim ve test veri seti olarak ikiye ayırmak için kullandığımız aşamadır. Kredi skorlamada genel olarak %70 eğitim verisi %30 test verisi kabul gören orandır. (Lin, Hu, Tsai 2011). Bu yüzden bu aşamada rastgele dağıtılacak şekilde oluşturulan bu oranı uygulanmıştır.

Split Data

Splitting mode
Split Rows

Fraction of rows in the f...
0.7

Randomized split

Random seed
100

Stratified split
False

Şekil 4.11: Eğitim-test verisi dağıtma ekran görüntüsü.

Veri ön işleme çalışmaları yapıldıktan sonra uygulama üstünde Karar Ağacı ve Lojistik Regresyon yöntemleri uygulanmıştır.

4.3.2.1. Tüm veri kümesi ile yapılan lojistik regresyon sonuçları

Veri kümesinde örnekleme metodu kullanılmadan çalıştırılan ve kaynak olarak veri ön işleme yapılmış ham veriyi kullandığımız uygulamada sonuçlar aşağıdaki gibi gerçekleşmiştir. Veri kümesi 4000 satır ve 18 kolondan oluşmaktadır. Model sonuçları Tablo 4.4’de paylaşılmıştır.

Tablo 4.4: Tüm veri kümesi ile yapılan lojistik regresyon sonuçları.

Gerçek Pozitif	Yalancı Negatif	Yalancı Pozitif	Gerçek Negatif	Doğruluk	Kesinlik	Duyarlılık	F1 Puanı
0	608	0	118675	0.995	1.000	0	0

Çıkan sonuçlara baktığımız zaman duyarlılık ve F1 puanının 0 olduğunu görüyoruz. Duyarlılık değerinin yüksek olması FPD değerine sahip kişilerin bulunmasının düşük olduğu tahminini çıkarabiliriz. Diğer taraftan doğruluk değeri de yüksek çıkmaktadır. Bu kişiler FPD olmayan yani kredisini zamanında ödemiş kişilerin tahmin doğruluğunun yüksek bir yüzde de yapıldığını göstermektedir.

4.3.2.2. Tüm veri kümesi ile çalıştırılan karar ağacı sonuçları

Örnekleme yapılmadan veri ön işleme yapılmış veri kümesi üstünde karar ağacı yöntemi çalıştırılmıştır. Veri kümesi 4000 satır ve 18 sütundan oluşmaktadır. Çalıştırılan model sonuçları Tablo 4.5’de görülmektedir.

Tablo 4.5: Tüm veri kümesi ile çalıştırılan karar ağacı yöntemi sonuçları.

Gerçek Pozitif	Yalancı Negatif	Yalancı Pozitif	Gerçek Negatif	Doğruluk	Kesinlik	Recall	F1 Puanı
15	596	8	119381	0.995	0.652	0.025	0.047

Yukarıdaki çıkan değerlere bakarsak lojistik regresyona göre doğruluk değeri yine yüksek çıkarken, kesinlik değeri ise lojistik regresyona göre değerlendirilince daha düşük çıkmıştır.

4.3.2.3. Alt örnekleme metodu ile çalıştırılan lojistik regresyon sonuçları

Alt örneklemede (undersampling) çalıştırılan Lojistik Regresyon algoritmasına ait modelleme sonuçları Tablo 4.6’de paylaşılmıştır. 4058 satır ve 18 sütundan oluşmaktadır.

Tablo 4.6: Alt örnekleme metodu ile çalıştırılan lojistik regresyon sonuçları.

Gerçek Pozitif	Yanlış Negatif	Yanlış Pozitif	Gerçek Negatif	Doğruluk	Kesinlik	Duyarlılık	F1 Puanı
548	77	72	512	0.877	0.884	0.877	0.88

Makine öğrenmesinde doğruluk değeri üzerinden baktığımız zaman bu değer 0.877 olduğu gözlenmiştir. Ayrıca duyarlılık ve F1 puanı değerlerinin de örnekleme yapılmadan önceki değerlerine göre daha yüksek yüzdeye sahip olduğu gözlemlenmiştir. Bu sebeple model karşılaştırmalarında sadece doğruluk değerininin karşılaştırılmasının bize yeterli bir sonuç vermeyeceğini düşünebiliriz.

Model sonucuna göre en önemli beş öznitelik önem sırasına göre Tablo 4.7’deki gibidir.

Tablo 4.7: Alt örnekleme lojistik regresyon öznitelik puanı.

Başvuru_kanalı:	0.133875
Vade:	0.084522
Skor:	0.047358
İl:	0.040817
Öğrenim_durumu:	0.036346

Skorlama yaparken ve krediler değerlendirilirken bu algorithmada krediye başvuru kanalı bilgisinin modeldeki tahmini etkileyen en önemli öznitelik olduğunu görmekteyiz. Kanal bilgisinin yüksek değer alması başvuru türünün kredi geri ödemelerinde bir kriter olabileceğini bize göstermektedir. Sonuçlara göre başvuru kanallarının diğer özniteliklerle ve hedef değişkenimiz olan FPD koşuluyla en yüksek bağlantıya sahip olması sebebiyle, kredi risk yönetimini yapan bankalar tarafından ayrıca değerlendirilmesi gerektiği söylenebilir.

4.3.2.4. Alt örnekleme metodu ile çalıştırılan karar ağacı sonuçları

Alt örneklemede (undersampling) çalıştırılan Lojistik Regresyon algoritmasına ait modelleme sonuçları Tablo 4.8’de paylaşılmıştır. 4058 satır ve 18 sütundan oluşmaktadır.

Tablo 4.8: Alt örnekleme metodu ile çalıştırılan karar ağacı yöntemi sonuçları.

Gerçek Pozitif	Yalancı Negatif	Yalancı Pozitif	Gerçek Negatif	Doğruluk	Kesinlik	Duyarlılık	F1 Puanı
548	77	72	512	0.877	0.884	0.877	0.88

Çıkan sonuçlara baktığımız zaman modelin örnekleme yapılmadan önceki verilere göre daha gerçekçi sonuçlara verdiği gözlenmiştir. Alt örnekleme yaptığımız zaman FPD bilgisini tutan verinin dengeli hale gelmesi bu modelimizin daha başarılı bir şekilde çalışmasını sağlamıştır.

Model sonuçlarına baktığımız zaman ise en çok etkileyen beş öznitelik Tablo 4.9’da paylaşılmıştır.

Tablo 4.9: Alt örnekleme karar ağacı öznelik puanı.

Başvuru_kanalı:	0.133875
Vade:	0.084522
Skor:	0.047358
İl:	0.040817
Öğrenim_durumu:	0.036346

Yukarıdaki açıklanan sonuçlara göre teminat ve kredi başvurusu yapılan kanal bilgilerinin veri kümesi içindeki diğer değerlere göre daha fazla ağırlıklandırıldığını söyleyebiliriz.

4.3.2.5. Üst örnekleme metodu ile çalıştırılan lojistik regresyon sonuçları

Üst örnekleme yapılarak verinin dengesizliği azaltılmıştır. Üst örnekleme yöntemi olarak SMOTE kullanılmıştır. SMOTE yöntemi veri kümemiz 805800 olmuş ve azınlık sınıfının değeri yani FPD olmayan kredi ödemesi olmayan müşterilerin sayısı çoğunluk sınıfına yakınsanmıştır.

SMOTE, azınlık sınıfını gözlemlenenlerin kümesinde sentetik örnekler üretmektedir. Fikir aynı sınıftan örnekler arasında araya girerek yeni azınlık örnekleri oluşturmaktır. Bunun, her azınlık gözleminin etrafında kümeler oluşturma etkisi vardır. SMOTE yaygın bir örnekleme tekniği olarak kullanılmaktadır.

SMOTE kullanıldığında veri sınıflarının değerlerini ve oranlarını Tablo 4.10'da görebiliriz.

Tablo 4.10: Veri kümesinin SMOTE ile yüzdesi.

	PD=0	PD=1	Toplam
Veri Seti	397971 %99.49	2029 %0.05	400000
SMOTE uygulanmış veri seti	397971 %50.56	407029 %49.44	805800

Veri kümesi 18 sütundan oluşmaktadır. Lojistik regresyon algoritmasının çalışması sonucu oluşan değerleri Tablo 4.11’de görebilirsiniz.

Tablo 4.11: Üst örnekleme metodu ile çalıştırılan lojistik regresyon sonuçları.

Gerçek Pozitif	Yalancı Negatif	Yalancı Pozitif	Gerçek Negatif	Doğruluk	Kesinlik	Duyarlılık	F1 Puanı
96937	25498	28284	89916	0.776	0.774	0.792	0.783

Çıkan sonuçlara baktığımız zaman doğruluk, duyarlılık değerlerinin alt örnekleme sonucu çıkan değerlere göre daha alt seviyede olduğunu görebiliriz. Modeller sonuç bölümünde değerlendirilirken doğruluk, duyarlılık, kesinlik gibi değerler üzerinden değerlendirilecektir. Doğruluk değerinin kıstas alınması durumunda üst örnekleme ile çalışan lojistik regresyon metodunun etkili bir yöntem olduğunu söyleyebiliriz.

Model sonuçlarını en çok etkileyen beş öznelik Tablo 4.12’de paylaşılmıştır.

Tablo 4.12: Üst örnekleme lojistik regresyon öznelik puanı.

Başvuru_kanalı:	0.133875
Vade:	0.084522
Skor:	0.047358
İl:	0.040817
Öğrenim_durumu:	0.036346

Başvuru kanalı bilgisinin üst örnekleme lojistik regresyonda diğer özneliklere göre daha etkili olduğu söylenebilir. İl ve öğrenim durumu gibi demografik ve kişisel bilgilerin ise daha az etkili olduğu gözlemlenmiştir.

4.3.2.6. Üst örnekleme metodu ile çalıştırılan karar ağacı sonuçları

Karar ağacı yönteminin SMOTE yöntemi ile arttırılmış veri kümesi üstünde uygulanması sonucu oluşan değerler Tablo 4.13’de paylaşılmıştır. Veri kümesi 805000 satır ve 18 kolondan oluşmaktadır.

Tablo 4.13: Üst örnekleme metodu ile çalıştırılan karar ağacı algoritması sonuçları.

Gerçek Pozitif	Yalancı Negatif	Yalancı Pozitif	Gerçek Negatif	Doğruluk	Kesinlik	Duyarlılık	F1 Puanı
121409	1422	3249	115660	0.981	0.974	0.988	0.981

Çıkan sonuçlara baktığımızda tüm değerlendirme kriterlerinin üst örnekleme metodu ile çalıştırılan lojistik regresyon sonuçlarına göre daha büyük çıktığı gözlenlenmiştir. Bu sebeple üst örnekleme yapılmış modellerle karar ağacı yönteminin uygulanması daha isabetli bir karar olabilir.

Model sonuçlarını en çok etkileyen beş öznelik aşağıda paylaşılmıştır.

Tablo 4.14: Üst örnekleme karar ağacı öznelik puanları.

Başvuru_kanalı:	0.133875
Vade:	0.084522
Skor:	0.047358
İl:	0.040817
Öğrenim_durumu:	0.036346

Öznelikler arasında karar ağacı skorlarını en çok etkileyen bilgi vade bilgisi olarak göze çarpıyor. Önceki değerlendirmelere baktığımızda kredi başvuru kanal bilgisinin karar ağacı için yine ayırıcı bir öznelik olduğunu yukarıdaki değerlere bakarak söyleyebiliriz.

5. TARTIŞMA VE SONUÇ

Çalışmanın sonucunda elde edilen bulguların değerlendirilmesine, bulguların değerlendirilmesi ile elde edilen sonuçlara, sonuçların değerlendirilmesine ve benzer çalışmalar yapacak araştırmacılara yönelik öneriler ve kısıtlamalara bu bölümde yer verilmiştir.

Yapılan analiz sonuçlarının değerleri ile hangi yöntemin daha uygulanabilir olması tartışılmıştır. Yapılan örnekleme teknikleri ile çalıştırılan model sonuçlarındaki farklılıklar gözlenmiştir. Doğru model seçiminde hangi kriteri kıstas olarak ele almamız gerektiği konusunda fikir verilmiştir. Ayrıca veri kümesinin dengesiz olması nedeni ile yapılan işlemlere ve veri analizi aşamalarında hangi adımların uygulandığı belirtilmiştir.

Bankacılık açısından çalışmanın sonucu ayrı bir başlıkta değerlendirilmiştir. Bu konu altında bankacılıkta risk yönetiminin amacı olan riski minimize etmek için uygulanacak yöntemlerden biri olarak makine öğrenme algoritmaları ilişkilendirilmiştir. Müşterilerin kredi ödeme tahminlemesinin sektör açısından nasıl bir fayda sağlayabileceği açıklanmış ve kredi değerlendirme ekiplerinin riskli krediler yönetiminde nasıl adımlar atabileceği tartışılmıştır.

Öneriler alt başlığında çalışmanın sonraki adımları olarak neler uygulanabileceğine değinilmiştir. Model implementasyonlarında farkle türde yöntemler ile farklı tipteki kredilerin değerlendirilebileceği açıklanmıştır.

Sınırlandırmalar bölümünde veri kümesi ile ilgili geliştirilebilecek özelliklerden bahsedilmiştir. Ayrıca mevcut sistemde bulunan kredi kayıtlarının başvurusu kabul olmuş müşteri verisini içermesinden dolayı sonuçların veri genişletildiği takdirde daha doğru sonuçlar verebileceği açıklanmıştır.

5.1. ANALİZ SONUÇLARI

Bu bölümde model sonuçlarını değerlendirdik. Lojistik Regresyon ve Karar Ağacı olmak üzere iki ana sınıflandırma algoritmasını kullanıyoruz. Bu modelleri uygularken uyguladığımız adımlar aşağıdakilerden oluşur:

- Veri setinin çalışmaya uygun hale gelmesi için veri manipülasyonu,
- Veri keşfi ve özet hale getirilmiş verilen görselleştirilmesi,

- Özellik seçimi,
- Veri örnekleme metodları,
- Makine öğrenmesi sınıflandırma algoritmalarını uygulama,
- Tüm sonuçları ölçümlerle değerlendirme ve veri kümesine farklı bir açıdan bakma.

Ayrıca veri setini anlamada ve modelleri uygulamada bazı adımlarda göz önünde tutulmuştur.

- Veri analizi ve tanımlayıcı istatistikler,
- Veri keşfi ve görselleştirme,
- Veri normalleştirme,
- Kukla değişkenler oluşturma.

Daha önce de belirttiğimiz üzere, FPD veri kümesinde kredilerde FPD olan kredilerin sayısı az olduğundan dengesiz bir veri kümesi olduğunu söyleyebiliriz. Bu, analitik modellerimizin performansını düşüren faktörlerden biri olabilir. Şimdiye kadar yapılan araştırmalarda, dengesiz veri setleri kullanıldığında, modelleri eğitmek için, en azından yanlış sınıflandırma hatası ile tam olarak sınıflandırma yapılamamıştır. Geleneksel makine öğrenme algoritmaları, çoğunluk sınıfını tercih etme eğiliminde olan sınıflandırıcıyı etkileyebilir ve azınlık sınıfına göre daha az miktarları nedeniyle zayıf doğruluk üretebilir. Bu nedenle, yeniden örnekleme yöntemlerini kullanmaya karar verdik.

Aşağıdaki altı model veri kümesi üzerinde uygulanmıştır ve sonuçları Tablo 5.1'de gösterilmiştir.

- MODEL 1: Tüm veri kümesi ile Lojistik Regresyon
- MODEL 2: Tüm veri kümesi ile Karar Ağacı
- MODEL 3: Undersampling metodu ile Lojistik Regresyon
- MODEL 4: Undersampling metodu ile Karar Ağacı

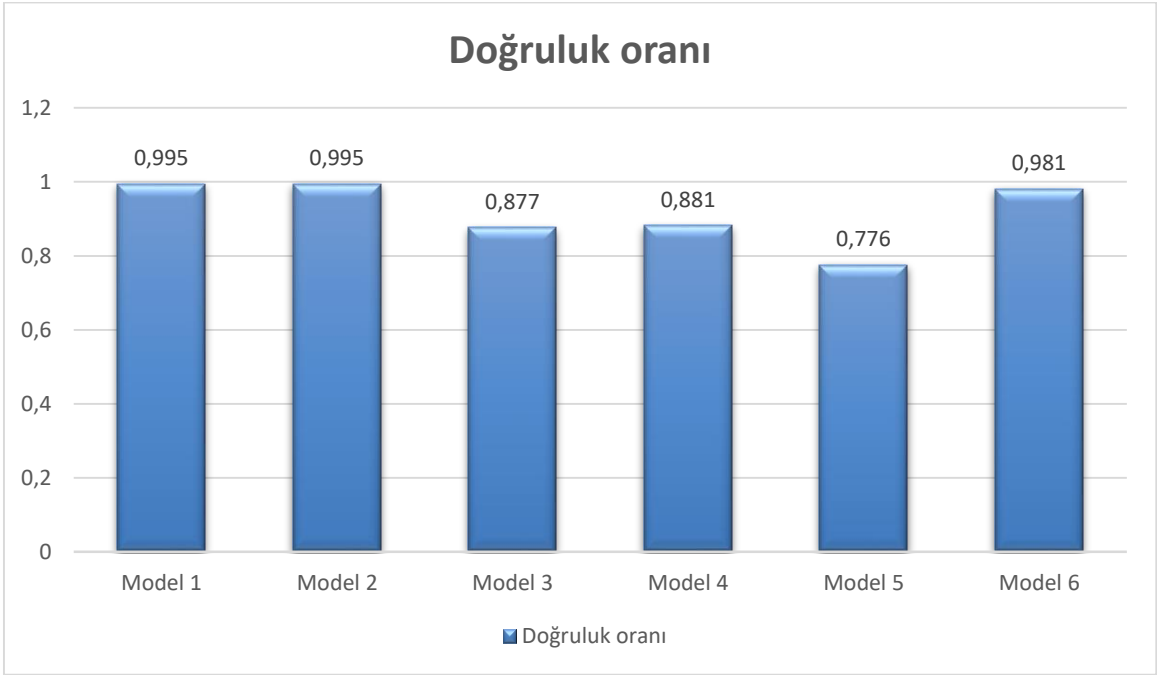
- MODEL 5: Oversampling Metodu ile Lojistik Regresyon
- MODEL 6: Oversampling Metodu ile Karar Ağacı

Tablo 5.1: Model değerlendirme kriterleri.

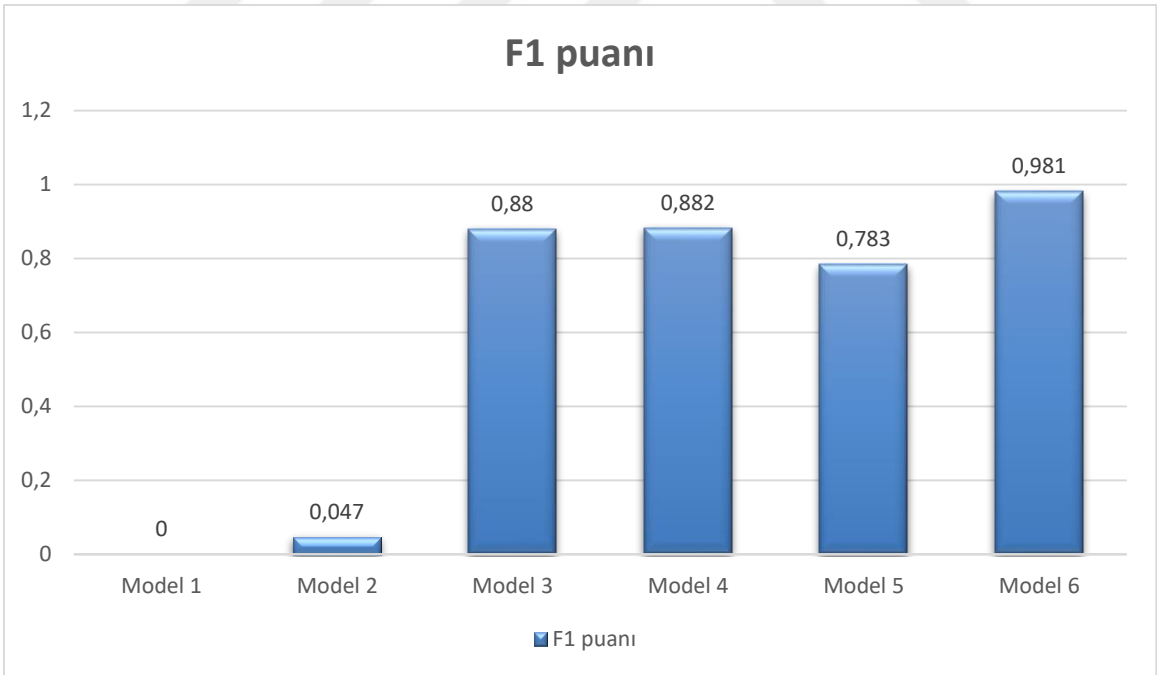
Modeller	GN	GP	YN	YP	Doğruluk	Kesinlik	Duyarlılık	F1 Puanı
MODEL 1	118675	0	608	0	0.995	1.000	0	0
MODEL 2	119381	15	596	8	0.995	0.652	0.025	0.047
MODEL 3	512	548	77	72	0.877	0.884	0.877	0.88
MODEL 4	530	542	86	59	0.881	0.902	0.863	0.882
MODEL 5	89916	96937	25498	28284	0.776	0.774	0.792	0.783
MODEL 6	115660	121409	1422	3249	0.981	0.974	0.988	0.981

Her şeyden önce, sonuçların ve modellerin performansının model sonuçlarına baktığımız zaman bize fikir verebilecek değerlere sahip olduğunu söyleyebiliriz. Model bulgularını kontrol ettiğimiz zaman sektör açısından faydalı çıkarımlar yapılabilir ve kredi karar vermede iyi birer öngörücü olabilir. Sonuçların değerlendirilmesinde F1 puanı baz alındığında Karar Ağacı modelinin diğer tüm modellere göre daha iyi bir performansa sahip olduğunu göstermiştir. Karar ağacı modellerinin belirli sınırlamalara tabi tutulduğundan, ağaç modellerinin overfitting eğiliminde olduğu unutulmamalıdır. Eğitim setinde olan tren seti ve test seti doğruluk skorlarına bakacak olursak, üst örnekleme tekniği kullanarak olan Karar Ağacı modelinin veriyi ezberleyebileceğini görebiliriz. Bu sebeple modelin çalışacağı veri kümesi farklılaştığında aynı oranda yüksek sonuçları bize vermeyebilir. Bunun sonucunda veri kümesi her farklılaştığında kredilerin gecikmeye kalmasının tespiti giderek zorlaşır.

Model sonuçlarına ait model değerlendirme kriter verilerine ait grafikler Şekil 5.1 ve Şekil 5.2'de gösterilmiştir.



Şekil 5.1: Model doğruluk puanları karşılaştırması.



Şekil 5.2: Model F1 puanları karşılaştırması.

Sınıflayıcıları eğitmek için hangi örnekleme yönteminin daha iyi olduğunu değerlendirdiğimizde, undersampling modellerinin daha iyi doğruluk puanlarına sahip

olduğunu söyleyebiliriz. Uygulamada, oversampling, nadir sınıftaki gözlem sayısının küçük olduğu aşırı durumlarda daha iyidir. Bunun nedeni, veri setindeki diğer sınıfların dağılımı ile ilgili önemli bilgileri kaybetmek istemememizdir. Bununla birlikte, iki veri örnekleme tekniğinin performansı ayırt edilemez. Yeniden örnekleme teknikleri veri seti hakkında yeni bilgiler vermemektedir. Karar ağacı modeli hariç, modellerimizin undersampling veri ile öğrendiklerinin yüksek doğruluk ve F1-Puanları'na sahip olduğu söylenebilir. Doğruluk puanı düşük olan modellerde negatif tahminlenen verilerin değerlendirilmesi konusunda eksiklik olduğu söylenebilir.

Analiz sonuçlarına göre üst örnekleme yapılmış karar ağacı yönteminin değerleri, diğer modellerle karşılaştırıldığında daha uygulanabilir ve daha doğru bir yöntem olduğu gözlemlenmiştir. Modeller üzerinde doğruluk açısından en yüksek puana sahip olmayabilir ama F1 puanı sebebiyle bu yöntem tercih edilebilir. Eğer sadece doğruluk puanı en yüksek olan modeli seçseydik FPD olan kredilerin tespiti önemini kaybedecektir. Bunun sonucunda da kredi risk değerlendirmelerinde sorunlu kredi tespiti zorlaşacaktır.

5.2. BANKACILIK SEKTÖRÜ AÇISINDAN SONUÇLAR

Türkiye ve Dünya'da bankacılık sektöründe meydana gelen gelişmeler ve bu gelişmelerin insanların sosyal ve ekonomik yaşantılarına ve ülke politikalarına etkisi incelendiğinde risk yönetimi ve kredi risk adı altında kredi değerlendirmelerinin önemi açıkça görülebilmektedir. Bankacılıkta iyi yönetilemeyen süreçler neticesinde çıkan bir kriz, sadece ilgili banka değil aynı zamanda devletin ekonomik durumunu ve tüketicilerin de sosyo-ekonomik hayatını zora sokmaktadır. Bu kriz ve risk yönetiminde oluşabilecek zararı minimuma indirmek için çalışmada oluşturduğumuz kredi modeli gibi bir model oluşturulmaktadır ve sorunlu kredilerin tespiti yapılmaktadır. Bunun sonucunda sağlıklı bir ekonomik süreç amaçlanmaktadır.

Bu çalışmadaki tüm modeller, modelde girdi olarak yer alan kredinin sorunlu olup olmadığını değerlendirmektedir. Yapılan model değerlendirmeleri sonucunda makul bir doğrulukla sorunlu krediye sahip tüketicinin olup olmadığını öngörmeyi sağlamaktadır. Bu müşterilere karşı herhangi bir karar vermeden önce bankanın yararına olacaktır. Hedef, kredi zararı riskini en aza indirmektir. Finans açısından, sınıflandırma modeli, bankanın kredi kaybını önlemesine, performansı artırmasına, verimliliği arttırmasına ve topluma değer vermesine yardımcı olacaktır.

Bankalar açısından baktığımızda, bankalar yalnızca bir temerrütün yani kredi ödeme gecikmesinin gerçekleşip gerçekleşmeyeceğini belirlemekle ilgilenmiyor, müşterilerin ne zaman kredilerini geri ödeyemeyecekleri ve temerrütlerin ortaya çıkacağı zamanlar da ilgileniyorlar. Ayrıca, uzun vadeli krediler için “iyi” ve “kötü” krediler arasındaki farkın belirlenmesi de bankalar için tek amaç değildir. Bankalardaki yönetim ekipleri için bir tavsiye olarak, bankaların riskli müşterilere kredi tahsis etmekten kaçınmaları gerekecektir. Ayrıca krediler sürekli ilgili kredi izleme ekipleri tarafından kontrol edilmelidir.

Bankada çalışan analist ekipler tarafından sonuca baktığımız zaman ise algoritmaların nasıl kullanılacağı ve test verilerinin nasıl değerlendirileceği öğrenilmiştir. Bu sayede mevcut veri veya gelebilecek başka verilerden başka algoritmalar da uygulanıp sonuçlar gözlemlenebilir. Veri analizi, ön işleme, normalizasyon ve hedef değişkenin belirlenmesinin önemi bu çalışma ile anlaşılmıştır. Ayrıca hedef değişken olan kredi gecikmesi bilgisinin diğer bilgilerle bir ilişkisinin olmaması çalışmanın daha da tutarlı olmasını sağlamıştır. Bu bölüm öznitelik değerlendirme bölümünde ele alınmıştır.

5.3. ÖNERİLER

Çalışmanın devamı olması durumunda ilgili veri üzerinden farklı algoritmalar denenebilir. Örneğin diğer sınıflandırma algoritmaları olarak kullanılabilir olan Naive Bayes ve Destek Vektör algoritmaları kullanılabilir. Ayrıca örnekleme yöntemlerinde SMOTE üst örnekleme(oversampling) ve rastgele örnekleme(undersampling) uygulanmıştır. Farklı sonuçlar almak için diğer örnekleme yöntemleri kullanılabilir. Bizim uyguladığımız yöntemler gibi bu yöntemler de veri kümesindeki dengesizliği ortadan kaldıracaktır. Yapılabilecek farklı çalışmalar sonucunda kredi ödeme skorlamasını en çok etkileyen öznitelikler tekrar değerlendirilip bu öznitelikler arasında ayrı bir analiz kurgulanabilir. Bu sayede kredi gecikmesine sebep olan farklı etkenler de bulunabilir.

Veri kümesi olarak sadece tüketici kredileri kullanan bu çalışma farklı kredi tipleri için de uygulanabilir ve farklı çıkarımlar yapılabilir. Ayrıca yaklaşık bir yıllık kredi başvurularını değerlendirdik. Tarih aralığı daha da arttırıp tarihsel veriler üstünden de algoritmalar çalıştırılabilir.

Oluşturulan bu modelin otomasyonu sağlanabilir. Çıkarılan kredi modelleri belirli periyotlarda yeniden çalıştırılabilir ve yeni gelen veriler üstünden değerlendirmeler yapılabilir. Bu

değerlendirme sonucu, yeni gelen veriler üzerinden kredi başvurularının değerlendirmesi hakkında fikir sahibi olunabilir ve değerlendirme aşamasında bu modellerden destek alınabilir. Sorunlu kredilerden oluşabilecek riskler bu otomasyon çalışmasıyla azaltılabilir ya da öngörülebilir ve bir kriz ortamı için önlem alınabilir.

Tüketicilerin ilgili ödeme bilgilerinin bulunduğu veri yerine kredi başvurusu kabul edilmeden önce yaptığı başvuru bilgileri de kullanılarak model kurgulanabilir. Bunun sonucunda kredi başvuru süreci zamanı azaltılabilir ve risk ekiplerinin değerlendirmesi yerine sistemsal verilerin model üzerinden işleyişi ile başvuru sonuçları değerlendirilebilir. Başvurulara hem hızlı yanıt verilebilir hem de kredilerden dolayı oluşabilecek risk azaltılabilir.

Ek olarak oluşturulan kredi modellerinin standardizasyonu sağlanabilir. Bu süreç içerisinde kredi başvuru süreçlerini denetleyen bankacılık kurumları olan TBB, BDDK ve KKB ile iletişime geçilip ülke içerisinde bir regülasyon oluşturulabilir. Netice olarak bir banka içerisinde yapılan bir model tüm finans sektöründe çalışan şirketler için uygulanabilir ve ülke ekonomisine katkı sağlayabilir.

5.4. SINIRLILIKLAR

Çalışmada bazı sınırlandırmalar bulunmaktadır. Üzerinde çalışılan veri kümesi sadece kredi başvurusu yapılmış ve kabul edilmiş verileri içermektedir. Kredi başvurusu kabul edilmeyen veriler de eklenerek verinin büyüklüğü artırılabilir ve daha ikna edici sonuçlar elde edilebilir.

Veri üzerinde kredi başvurusunun yapıldığı şube bilgisi gibi veriler 350'den fazla değere sahiptir. Bu sayıdaki değerler kategorik olarak sınıflandırmada iyi bir tahminleyici olamayabilir. Eğer şube yerine bölge bilgisi eklenip bu veri de analizlere eklenebilirse model sonucunu etkileyebilir ve modelin tahmin performansını arttırabilir.

Modelde değerlendirilen niteliklere baktığımızda bozuk ve kayıp veriler olduğu gözlemlenmiştir. Model kurulurken bu değerler sonuca etki edeceği için model kurulum aşamasında çıkarılmıştır. Bozuk ve kayıp verilerin sisteme girişi incelenebilir ve belirli kriterler getirilip doğru şekilde girmesi sağlanabilir. Örneği, kimlik numarası hanesinde sadece rakam girilmelidir gibi bir kriter belirlenerek yanlış şekilde harf verisi girişi engellenebilir.

Teknik olarak kısıtlamalar da model kurulununun yapılması için ilgili veri kümesinin, hem analiz hem de kredi başvurusu deęerlendirmesini yapan ekiplerce birlikte deęerlendirilmesi gerekmektedir. Model kurulumu sadece veri üzerinden ıkarılan fikirler baz alınarak yapılmıřtır. Süreci daha iyi anlamak ve ek gereksinimleri ıkartmak için kredi operasyon ekipleri ile görüřülebilir.



KAYNAKLAR

- Akkaya, M., & Azimli, T., 2018, *Türk Bankacılık Sektöründe Likidite Riski Yönetimi. Finans Politik & Ekonomik Yorumlar*, 55(638), 35-48.
- Akman, M., Genç, Y., & Ankaralı, H., 2011, *Random Forests Methods And An Application In Health Science. Turkiye Klinikleri Journal Of Biostatistics*, 3(1), 36.
- Aktaş, R., 2000, Sorunlu Krediler. *Tbb Eğitim Ve Tanıtım Grubu Seminer Notları*, İstanbul, 25-26.
- Aloğlu, Z. T., 2005, *Bankacılık Sektörünün Karşılaştığı Riskler Ve Bankacılık Krizler Üzerindeki Etkileri*, Tcmb Bankacılık Ve Finansal Kuruluşlar Genel Müdürlüğü Uzmanlık Yeterlilik Tezi, Ankara.
- Altıntaş, M. A., 2006, *Bankacılıkta Risk Yönetimi Ve Sermaye Yeterliliği: 5411 Sayılı Bankacılık Kanunu*, Basel-I Ve Basel-II Çerçevesinde. Turhan Kitabevi.
- Anderson, A., 2001, *Riskler ve risk yönetimi*. Seminer Notları. Ankara.
- Argüden, Y. Ve Erşahin B., 2008, *Veri Madenciliği: Veriden Bilgiye. Masraftan Değere*, Arge Danışmanlık Yayınları, (10).
- Ayhan, S., & Erdoğan, Ş., 2014, Destek Vektör Makineleriyle Sınıflandırma Problemlerinin Çözümü İçin Çekirdek Fonksiyonu Seçimi, *Eskişehir Osmangazi Üniversitesi İktisadi Ve İdari Bilimler Dergisi*, 9(1), 175-201.
- Barber, D., 2012, *Bayesian Reasoning And Machine Learning*, Cambridge University Press.
- Barnes, J., 2015, *Microsoft Azure Essentials Azure Machine Learning*, Microsoft Press.
- BDDK, 2001, Bankaların İç Denetim ve Risk Yönetim Sistemleri Hakkında Yönetmelik, *Bankalar Kanunu 4389/9 Madde 2*, Resmi Gazete, 24312 (8.2.2001): 2.
- BDDK, 2018, *Türk Bankacılık Sektörü Temel Göstergeler Raporu Aralık 2018*, <https://www.bddk.org.tr/Haberler/Turk-Bankacilik-Sektoru-Temel-Gostergeler-Raporu-Aralik-2018-yayinlanmistir/38>, [Ziyaret tarihi: 10 Mayıs 2019].
- Beja, A., 2016, *Operational Risk Management And The Case Of Albania. European Journal Of Multidisciplinary Studies*, 1(3), 65-72.
- Bell, S., & Bala, K., 2015, *Learning Visual Similarity For Product Design With Convolutional Neural Networks*, *Acm Transactions On Graphics (Tog)*, 34(4), 98.
- Bessis, J., 2011, *Risk Management In Banking*, John Wiley & Sons.
- Bolgün, K. E., & Akçay, M. B., 2009, *Risk Yönetimi: Gelişmekte Olan Türk Finans Piyasasında Entegre Risk Ölçüm Ve Yönetim Uygulamaları*, Scala Yayıncılık.

- Boyacıođlu, M. A., 2003, 1980 Sonrası Türk Bankacılık Sektöründeki Gelişmeler, Krizlerin Sektör Üzerindeki etkileri Ve İyileştirici öneriler, *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (9), 523-538.
- Bozkurt, C., 2010, Risk, Kurumsal Risk Yönetimi ve İç Denetim, *Denetim*, (4), 17-30.
- Breiman, L., 2001, Random Forests, *Machine Learning*, 45(1), 5-32.
- Bromiley, P., Mcshane, M., Nair, A., & Rustambekov, E., 2015, Enterprise Risk Management: Review, Critique, And Research Directions, *Long Range Planning*, 48(4), 265-276.
- Brown, K., & Moles, P., 2014, *Credit Risk Management*, Credit Risk Management, 16.
- Brownlee, J., 2014, An Introduction To Feature Selection, *Machine Learning Process*, 6.
- Buch, C. M., & Goldberg, L. S., 2015, International Banking And Liquidity Risk Transmission: Lessons From Across Countries, *Imf Economic Review*, 63(3), 377-410.
- Buduma, N., & Locascio, N., 2017, *Fundamentals Of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms*. O'reilly Media, Inc.
- Candan, H., & Özün, A., 2006, *Bankalarda Risk Yonetimi Ve Basel II*. Türkiye İş Bankası Yayınları.
- Canepa, G. A., 2016, *What You Need To Know About Machine Learning*, Birmingham: Packt Publishing.
- Chappell, D., 2015, *Introducing Azure Machine Learning. A Guide For Technical Professionals*, Sponsored By Microsoft Corporation.
- Chawla, N. V., Bowyer, K. W., Hall, L. O. ve Kegelmeyer, W. P., 2002, SMOTE: synthetic minority over-sampling technique, *Journal of artificial intelligence research*, 16, 321–357.
- Cichosz, P., 2015, *Data Mining Algorithms: Explained Using R*, John Wiley & Sons Incorporated.
- Coyle, B., 2000, *Framework For: Credit Risk Management*, Global Professional Publishi.
- Cui, G., Wong, M. L., & Lui, H. K., 2006, Machine Learning For Direct Marketing Response Models: Bayesian Networks With Evolutionary Programming, *Management Science*, 52(4), 597-612.
- Dalı, İ. B., 2016, *Geleneksel Bankalar İle Katılım Bankalarının Finansal Performanslarının Basel Iı Kriterleri Açısından Kıyaslanması*.
- Dasgupta, A., & Nath, A., 2014, *Classification Of Machine Learning Algorithms*.
- Demirciođlu, H., & Bilge, H., 2015, Yumurtalık Kanseri Veri Kümesindeki Gen İfadelerinin Veri Madenciliđi İle Analizi. *Marmara Fen Bilimleri Dergisi*, 27(4), 125-134.

- Drenovak, M., Ranković, V., Ivanović, M., Urošević, B., & Jelic, R., 2017, Market Risk Management In A Post-Basel Iı Regulatory Environment, *European Journal Of Operational Research*, 257(3), 1030-1044.
- Emir, Ş., 2013, *Yapay Sinir Ağları Ve Destek Vektör Makineleri Yöntemlerinin Sınıflandırma Performanslarının Karşılaştırılması: Borsa Endeks Yönünün Tahmini Üzerine Bir Uygulama*, İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, Yayınlanmamış Doktora Tezi, İstanbul.
- Frexias X., Rochet J.C., 1999, *Microeconomics of Banking*, MIT Press, U.S.A..
- Gaganis, C., Pasiouras, F., Spathis, C., & Zopounidis, C., 2007, A Comparison Of Nearest Neighbours, Discriminant And Logit Models For Auditing Decisions, *Intelligent Systems In Accounting, Finance & Management: International Journal*, 15(1-2), 23-40.
- Géron, A., 2017, *Hands-On Machine Learning With Scikit-Learn And Tensorflow: Concepts, Tools, And Techniques To Build Intelligent Systems*. O'reilly Media, Inc.
- Hancer, E., & Karaboga, D., 2017, A Comprehensive Survey Of Traditional, Merge-Split And Evolutionary Approaches Proposed For Determination Of Cluster Number, *Swarm And Evolutionary Computation*, 32, 49-67.
- He, H. ve Garcia, E. A., 2009, *Learning from Imbalanced Data*, *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284. doi:10.1109/TKDE.2008.23
- Hornick, M. F., Marcade, E., & Venkayala, S., 2007, *Java Data Mining: Strategy, Standard, And Practice: A Practical Guide For Architecture, Design, And Implementation*.
- İnsel A. ve Sungur N., 2003, *Sermaye Akımlarının Temel Makro-Ekonomik Göstergeler Üzerindeki Etkileri: Türkiye Örneği*. Discussion paper.
- Jaiwei, H., & Kamber, M., 2006, *Data Mining: Concepts And Techniques*, Ed: Morgan Kaufmann San Francisco.
- James, G., Witten, D., Hastie, T., & Tibshirani, R., 2013, *An Introduction To Statistical Learning (Vol. 112, P. 18)*, New York: Springer.
- Kalaycı, S., Kamasak, M., & Arslan, S., 2018, *Credit Risk Analysis Using Machine Learning Algorithms*, In 2018 26th Signal Processing And Communications Applications Conference (Sıu)(Pp. 1-4), Ieee.
- Koyuncu, C., & Saka, B., 2011, Takipteki Kredilerin Özel Sektöre Verilen Krediler Ve Yatırımlar Üzerindeki Etkisi. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, (31).
- King, G., & Zeng, L., 2001, *Logistic Regression In Rare Events Data*. *Political Analysis*, 9(2), 137-163.
- Lantz, B., 2013, *Machine Learning With R*. Packt Publishing Ltd.

- Lin, W. Y., Hu, Y. H., & Tsai, C. F., 2011, Machine learning in financial crisis prediction: a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(4), 421-436.
- Liu, S., & D'aquin, M., 2017, *Unsupervised Learning For Understanding Student Achievement In A Distance Learning Setting*, In 2017 Ieee Global Engineering Education Conference (Educon) (Pp. 1373-1377). Ieee.
- Mandacı, P. E., 2003, *Türk bankacılık sektörünün taşıdığı riskler ve finansal krizi aşmada kullanılan risk ölçüm teknikleri*.
- Menard, S., 2002, *Applied Logistic Regression Analysis (Vol. 106)*, Sage.
- Mitchell, T. M., 1997, Does Machine Learning Really Work?, *AI Magazine*, 18(3), 11-11.
- Monhamady, K. K., 2018, *Developing Machine Learning Methods for Business Intelligence* (Doctoral dissertation, Abdullah Gul University).
- Nahayo, Y., & Arı, S., 2016, Birleşik Modellemeli Ve Modellemesiz Metin-Bağımsız Konuşmacı Tanıma İçin Svm, K-Nn Ve Nbc Sınıflandırıcıların Başarımı. *Sakarya University Journal Of Science*, 20(1), 1-6.
- Ng, A., 2000, Cs229 Lecture Notes, *Cs229 Lecture Notes*, 1(1), 1-3.
- Nitze, I., Schulthess, U., & Asche, H., 2012, Comparison Of Machine Learning Algorithms Random Forest, Artificial Neural Network And Support Vector Machine To Maximum Likelihood For Supervised Crop Type Classification, *Proceedings Of The 4th Geobia*, Rio De Janeiro, Brazil, 79, 3540.
- Ocak, I., & Seker, S. E., 2013, Calculation Of Surface Settlements Caused By Epbm Tunneling Using Artificial Neural Network, Svm, And Gaussian Processes, *Environmental Earth Sciences*, 70(3), 1263-1276.
- Osowski, S., Siwek, K., & Markiewicz, T., 2004, *Mlp And Svm Networks-A Comparative Study*, In Proceedings Of The 6th Nordic Signal Processing Symposium, NORSIG 2004. (Pp. 37-40). Ieee.
- Öker, A., 2007, *Ticari Bankalarda Kredi Ve Kredi Riski Yönetimi-Bir Uygulama*, Marmara Üniversitesi Sosyal Bilimler Enstitüsü Yayınlanmamış Doktora Tezi, İstanbul.
- Santamore A.M., 1997, *Commercial Bank Risk Management: An Analysis Of The Process*, The Wharton Financial Institutions Center, U.S.A..
- Selimler, H., 2006, *Türk Bankacılık Sektöründe Sorunlu Kredilerin Varlık Yönetim Şirketlerince Tasfiyesi, Seçilmiş Ülkeler Ve Türkiye Uygulaması*, Marmara Üniversitesi Bankacılık Ve Sigortacılık Enstitüsü, Bankacılık Anabilim Dalı, Doktora Tezi.
- Shalev-Shwartz, S., & Ben-David, S., 2014, *Understanding Machine Learning: From Theory To Algorithms*, Cambridge University Press.

- Shearer, C., 2000, The Crisp-Dm Model: The New Blueprint For Data Mining, *Journal Of Data Warehousing*, 5(4), 13-22.
- Shelke, M. S., Deshmukh, P. R. ve Shandilya, V. K., 2017, A Review on Imbalanced Data Handling Using Undersampling and Oversampling Technique, *International Journal of Recent Trends in Engineering and Research*, 3(4), 444-449.
- Suthaharan, S., 2014, Big Data Classification: Problems And Challenges İn Network İntrusion Prediction With Machine Learning, *Acm Sigmetrics Performance Evaluation Review*, 41(4), 70-73.
- Şimşek, K. Ç., 2007, Bankacılıkta Risk ve Risk Ölçüm Yöntemleri. *Ankara Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı, Ankara.*
- TBB, 2018, *Uluslararası Karşılaştırmalar İtibariyle Bankacılık Sektörü*, https://www.tbb.org.tr/Content/Upload/Dokuman/7561/Uluslararası_Karsilastirmalar__2017.pdf, [Ziyaret tarihi: 29 Mayıs 2019].
- Teker, S., Akçay, B., & Turan, M., 2006, Measuring Credit Risk Of A Bank's Corporate Loan Portfolio Using Advanced Internal Ratings Base Approach, *Journal Of Transnational Management*, 11(1), 17-40.
- Tufféry, S., 2011, *Data Mining And Statistics For Decision Making*, John Wiley & Sons.
- TÜİK, 2018, *Temel İstatistikler*, http://www.tuik.gov.tr/PreIstatistikTablo.do?istab_id=1532, [Ziyaret tarihi: 29 Mayıs 2019].
- Usta, M., 2000, *Temel Kredi Bilgileri*, Pamukbank Eğitim Müdürlüğü Yayınları.
- Uysal, E. U., 2009, Operasyonel Risk Yönetiminde Senaryo Analizi, *Bankacılar Dergisi*, 69, 73-85.
- Van Suntum, U., Kaptan, M., & Ilgmann, C., 2011, Reducing The Lower Bound On Market Interest Rates, *Economic Analysis And Policy*, 41(2), 134.
- Wei, Y., Lam, P. T. I., Chiang, Y. H., Leung, B. Y. P., & Seabrooke, W., 2014, An Exploratory Analysis Of İmpediments To China's Credit Control On The Real Estate İndustry: An İstitutional Perspective, *Journal Of Contemporary China*, 23(85), 44-67.
- Witten, I. H., Frank, E., & Hall, M. A., 2011, *Data Mining: Practical Machine Learning Tools And Techniques*, Burlington, Ma 01803.
- Yarız, A., 2011, *Bankacılıkta Risk Yönetimi: Risk Matrisi Uygulaması*.
- Yıldırım, B. D., 2011, Türkiye'nin finansal piyasa likiditesi, ölçümü ve analizi, *Central Bank Review*, 11(1), 11.
- Yıldız, O. T., & Alpaydın, E., 2004, *Learning Rules From Data*, Boğaziçi University.
- Yılmaz, S., 2011, Ticari Bankalarda Kredi Portföyü Ve Kredi Riski Yönetimi-Bankacılık Sektöründe Bir Uygulama.

Zafarani, R., Abbasi, M. A., & Liu, H., 2014, *Social Media Mining: An Introduction*, Cambridge University Press.

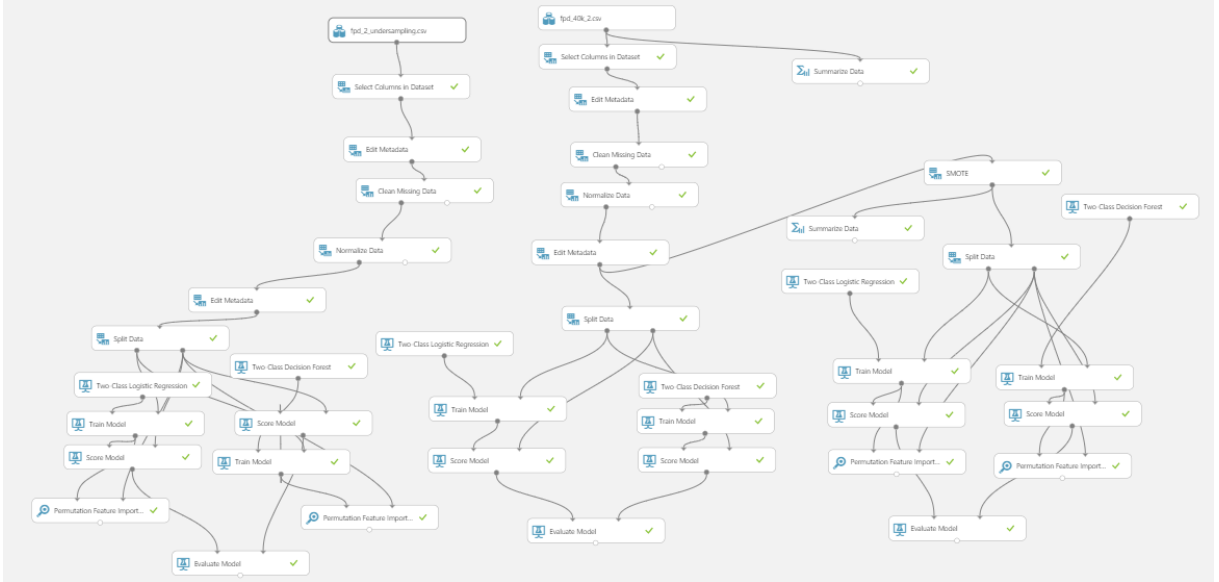
Zalođlu, 2015, *Dünyayı Deđiřtirmekte Olan Yapay Sinir Ađları Nedir?*, <https://Bilimfili.Com/Dunyayi-Degistirmekte-Olan-Yapay-Sinir-Aglari-Nedir/>, [Ziyaret tarihi: 15 Mayıs 2019].

Zarakolu, A., 1989, A New Development in the Turkish Banking Sector: Banking without Interest. *Banka Huk. Dergisi*, 15, 3.



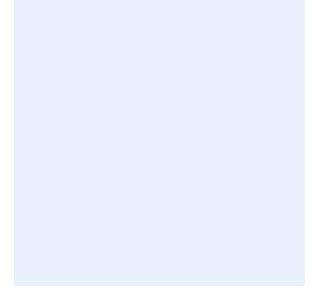
EKLER

EK 1. Azure ML Studio Programı Modelleme Ekran Görüntüsü



ÖZGEÇMİŞ

Kişisel Bilgiler	
Adı Soyadı	Enes GEZER
Doğum Yeri	Bolu
Doğum Tarihi	06.09.1989
Uyruğu	<input checked="" type="checkbox"/> T.C. <input type="checkbox"/> Diğer:
Telefon	0555 559 34 29
E-Posta Adresi	enesgezer89@gmail.com
Web Adresi	



Eğitim Bilgileri	
Lisans	
Üniversite	Yıldız Teknik Üniversitesi
Fakülte	Makine
Bölümü	Endüstri Mühendisliği
Mezuniyet Yılı	13.02.2014

Yüksek Lisans	
Üniversite	İstanbul Üniversitesi
Enstitü Adı	Fen Bilimleri Enstitüsü
Anabilim Dalı	Enformatik
Programı	Enformatik