ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

OVERCOMING PAYMENT BEHAVIOR CHALLENGES: CLASSIFYING BUY NOW PAY LATER USERS WITH MACHINE LEARNING

M.Sc. THESIS

Ömür ÖZDOĞAN

Department of Data Engineering and Business Analytics

Big Data and Business Analytics Programme

AUGUST 2024

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

OVERCOMING PAYMENT BEHAVIOR CHALLENGES: CLASSIFYING BUY NOW PAY LATER USERS WITH MACHINE LEARNING

M.Sc. THESIS

Ömür ÖZDOĞAN
(528211073)

Department of Data Engineering and Business Analytics

Big Data and Business Analytics Programme

Thesis Advisor: Assist. Prof. Dr. Mehmet Ali ERGÜN

AUGUST 2024

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

ÖDEME ALIŞKANLIĞI ZORLUKLARINI AŞMAK: MAKİNE ÖĞRENİMİ İLE ŞİMDİ AL SONRA ÖDE KULLANICILARINI SINIFLANDIRMA

YÜKSEK LİSANS TEZİ

Ömür ÖZDOĞAN
(528211073)

Veri Mühendisliği ve İş Analitiği Anabilim Dalı

Büyük Veri ve İş Analitiği Programı

Tez Danışmanı: Dr. Öğr. Üyesi Mehmet Ali ERGÜN

AĞUSTOS 2024

Ömür ÖZDOĞAN, a M.Sc. student of ITU Graduate School student ID 528211073, successfully defended the thesis/dissertation entitled "OVERCOMING PAYMENT BEHAVIOR CHALLENGES: CLASSIFYING BUY NOW PAY LATER USERS WITH MACHINE LEARNING", which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Assist. Prof. Dr. Mehmet Ali ERGÜN** İstanbul Technical University	.....
Jury Members :	**Prof. Dr. Selim ZAİM** İbn Haldun University	.....
	Assoc. Prof. Dr. Ömer Faruk Beyca İstanbul Technical University	.....

Date of Submission : 23 May 2024
Date of Defense : 08 August 2024

To my family,

FOREWORD

I would like to express my heartfelt gratitude to those who have supported me in the completion of this thesis. Firstly, I would like to thank my thesis supervisor Assistant Professor Mehmet Ali Ergün. His expertise has enriched this thesis and fostered my growth in my academic journey. I am very grateful for his guidance.

I would like to thank my mother İlkay, father Uğur and brother Yiğit for their love and encouragement, as their unwavering belief in me has always fueled my determination. I am grateful for the love and joy he/she brings into my life every day. I would like to thank my better half Umur for his love have been my source of inspiration and I am grateful for the love and joy he brings into my life every day.

I would like to thank my colleagues and coworkers for their insightful discussions and shared experiences. Their contributions have added depth and perspective to this research. I would like to thank my dearest manager Nuri Yasin Külahçı, whom I deeply miss working with. His mentorship, leadership and guidance continue to inspire me, and I am grateful for the lessons learned under his tutelage.

Lastly, I would like to thank my beloved loyal companion Tarçın for her unconditional love and companionship over the years. Her presence has brought immeasurable joy and comfort to my life and I cherish the memories we've shared together.

June 2024

Ömür ÖZDOĞAN
(Sr. Data Analyst)

TABLE OF CONTENTS

Page

FOREWORD ..... **ix**
TABLE OF CONTENTS ..... **xi**
ABBREVIATIONS ..... **xiii**
LIST OF TABLES ..... **xiv**
LIST OF FIGURES ..... **xv**
SUMMARY ..... **xvii**
ÖZET ..... **xx**
1. INTRODUCTION **1**
2. LITERATURE REVIEW **3**
2.1 Financial Technology & Buy Now Pay Later 3
2.1.1 Overview of Fintech 3
2.1.1.1 Definition and evolution of Fintech 3
2.1.1.2 Today's financial landscape and the future of Fintech 4
2.1.2 Overview of BNPL 6
2.1.2.1 Definition of BNPL 6
2.1.2.2 BNPL users 7
2.1.2.3 Need for BNPL 8
2.1.2.4 Example of BNPL providers 8
2.2 Credit Risk 10
2.2.1 Overview of credit risk 10
2.2.2 Credit risk management 10
2.2.3 Importance of AI and Fintech in credit risk 12
2.2.3.1 Usage of ML in credit risk 14
2.3 Materials 14
2.3.1 Algorithms 15
2.3.1.1 Decision Tree 15
2.3.1.2 Random Forest 16
2.3.1.3 XgBoost 18
2.3.1.4 Deep Learning 19
2.3.2 Data preparation 19
2.3.3 Feature engineering and correlation matrix 21
2.3.4 K-fold cross validation 21
2.3.5 Model evaluation 22
2.3.5.1 Confusion matrix 22
2.3.5.2 Performance metrics 23
2.3.5.3 ROC curve and Kolmogorov-Smirnov chart 24
3. METHODOLOGY AND IMPLEMENTATION **25**
3.1 Data Analysis and Visualization 27
3.2 Model Preparation 29
3.2.1 Target details 29
3.2.2 Correlation matrix and features engineering 30
3.3 Model Application 32

3.3.1 Random Forest application 32
3.3.1.1 Model tuning 33
3.3.1.2 Feature importances 33
3.3.1.3 K-fold cross validation 34
3.3.1.4 ROC curve and weighted average AUC 34
3.3.1.5 Confusion matrix 35
3.3.2 XgBoost application 36
3.3.2.1 Model tuning 36
3.3.2.2 Feature importances 37
3.3.2.3 K-fold cross validation 37
3.3.2.4 ROC curve and weighted average AUC 37
3.3.2.5 Confusion matrix 38
3.3.3 Deep learning application 39
3.3.3.1 Model tuning 39
3.3.3.2 Model architecture 39
3.3.3.3 K-fold cross validation 40
3.3.3.4 ROC curve and weighted average AUC 40
3.3.3.5 Confusion matrix 41
4. RESULTS 43
5. CONCLUSION AND FUTURE RESEARCH 49
5.1 Conclusion 49
5.2 Future Research 50
5.2.1 Limitations of current study, new research questions, and hypotheses 50
5.2.2 Application areas and technological developments 51
REFERENCES 53
CURRICULUM VITAE 59

ABBREVIATIONS

AI	**:** Artificial Intelligence
AUC	**:** Area Under the Curve
BNPL	**:** Buy Now Pay Later
CNN	**:** Convolution Neural Networks
DL	**:** Deep Learning
DT	**:** Decision Tree
FINTECH	**:** Financial Technology
KS	**:** Kolmogorov-Smirnov
LSTM	**:** Long Short Term Memory
ML	**:** Machine Learning
MLP	**:** Multi-layer Perceptron
ROC	**:** Receiver Operating Characteristic
RT	**:** Random Forest
SMV	**:** Support Vector Machine
XgBoost	**:** Extreme Gradient Boosting

LIST OF TABLES

Page

Table 2.1: History of Fintech **4**
Table 3.1: Data summary **27**
Table 3.2: Random Forest model tuning hyperparameters **33**
Table 3.3: 5-fold cross validation for Random Forest..... **34**
Table 3.4: XgBoost model tuning hyperparameters..... **36**
Table 3.5: 5-fold Cross Validation for XgBoost **37**
Table 3.6: Deep Learning model tuning hyperparameters **39**
Table 3.7: 5-fold cross validation for Deep Learning **40**
Table 4.1: Random Forest model classification report..... **43**
Table 4.2: XgBoost model classification report..... **43**
Table 4.3: Deep Learning model classification..... **43**

LIST OF FIGURES

Page

Figure 1.1: Lending rates for selected G20 countries **1**
Figure 2.1: Total global funding value in Fintech between 2020 - 2023 ..... **5**
Figure 2.2: BNPL penetration by generation between 2019 - 2025 ..... **7**
Figure 2.3: Reasons for BNPL usage **8**
Figure 2.4: BNPL brand rankings in the U.S. **9**
Figure 2.5: General structure of Decision Tree **16**
Figure 2.6: General structure of Random Forest **17**
Figure 2.7: General structure of XgBoost **18**
Figure 2.8: General structure of Deep Learning **19**
Figure 2.9: Data preparation phase flow chart **20**
Figure 2.10: 10-fold cross-validation **22**
Figure 2.11: Example of multiclass classification problem confusion matrix **23**
Figure 2.12: Performance metrics for a multiclass confusion matrix **24**
Figure 3.1: Workflow diagram **25**
Figure 3.2: User count by age group **27**
Figure 3.3: User-based operator distribution **28**
Figure 3.4: Monthly credit amount and user count **28**
Figure 3.5: Data info **29**
Figure 3.6: User count by target group **30**
Figure 3.7: Correlation matrix for all features **31**
Figure 3.8: Correlation matrix of chosen features **32**
Figure 3.9: Random Forest feature importances **33**
Figure 3.10: Random Forest ROC curve graph **35**
Figure 3.11: Confusion matrix of Random Forest **36**
Figure 3.12: XgBoost feature importances **37**
Figure 3.13: XgBoost ROC curve graph **38**
Figure 3.14: Confusion matrix of XgBoost **38**
Figure 3.15: Deep Learning ROC curve graph **40**
Figure 3.16: Confusion matrix of Deep Learning **41**
Figure 4.1: Random Forest model confusion matrix **44**
Figure 4.2: XgBoost model confusion matrix **44**
Figure 4.3: Deep Learning model confusion matrix **44**

OVERCOMING PAYMENT BEHAVIOUR CHALLENGES: CLASSIFYING BUY NOW PAY LATER USERS WITH MACHINE LEARNING

SUMMARY

When the economies of developing countries are examined closely, issues such as high debt rates, limited access to funding and difficulties in accessing financing stand out. While Turkey is considered in the category of developing countries, various steps have been taken to strengthen financial stability within the scope of financial tightening policies in recent periods. These steps include steps such as increasing the risk weights of consumer loans, individual credit cards and vehicle loans, and reducing or eliminating the installments on credit cards in certain sectors.

Thus, consumers' financial access continues to be increasingly restricted. At this point, Fintech companies offer different alternatives as another option to the restrictions in the traditional financial system.

Fintech companies aim to offer low cost, fast and innovative solutions to their customers by using modern data analysis techniques and new generation technologies such as AI. Additionally, these companies aim to reach people who have limited access or are looking for alternatives, in addition to users who do not have access to the banking sector. In this study, first, the development and future of Fintech companies are mentioned and the main product of the study, Buy Now Pay Later, is explained in detail and examples from global BNPL providers are given. Then, credit risk and machine learning techniques in credit risk were also mentioned.

During the study, data analysis was performed using sample data of approximately 35,000 customers using the BNPL loan product developed by a London-based Fintech company for one of Turkey's leading retail market chains. A multi-class segmentation problem was designed for users, and popular machine learning methods were used to predict in 3 different classes whether the loans will be paid on time or not. During the experimental phase, Random Forest, Extreme Gradient Boosting and Deep Learning algorithms were tested and performance tables for the models were prepared. When

comparing model performances, values such as accuracy, f1-score, roc curve and confusion matrix were determined as priority metrics.

As a result, it has been determined that the use of machine learning modeling is quite advantageous when classifying credit risk. However, before deciding on model selection, companies' strategies and policies should always be taken into consideration, and at the same time, the bad loan risks they want to take should be evaluated.

Keywords: Fintech, BNPL, Credit Risk Model, Machine Learning

ÖDEME ALIŞKANLIĞI ZORLUKLARINI AŞMAK: MAKİNE ÖĞRENİMİ İLE ŞİMDİ AL SONRA ÖDE KULLANICILARINI SINIFLANDIRMA

ÖZET

Gelişmekte olan ülkelerin ekonomileri yakından incelendiğinde yüksek borçlanma oranları, kısıtlı fonlama erişimi ve finansman erişiminde zorluklar gibi maddeler ön plana çıkmaktadır. Türkiye de gelişmekte olan ülkeler kategorisinde değerlendirilirken ek olarak geçtiğimiz dönemlerde sıkılaştırma politikaları kapsamında finansal istikrarı güçlendirmeye yönelik çeşitli adımlar atılmıştır. Bu adımlar içerisinde ihtiyaç kredileri, bireysel kredi kartları, taşıt kredilerinin risk ağırlıklarının artırıldığı ve kredi kartlarında uygulanacak taksitlendirmelerin belirli sektörlerde azaltılacağı ya da tamamen kaldırılacağı gibi maddeler yer almıştır.

Böylelikle, tüketicilerin finansal erişimi giderek kısıtlanmaya devam etmektedir. Bu noktada, geleneksel finans sistemindeki kısıtlamalara başka bir seçenek olarak finansal teknoloji şirketleri farklı alternatifler sunmaktadır.

Fintech şirketleri modern data analizi teknikleri ve yapay zeka gibi yeni nesil teknolojileri kullanarak müşterilerine düşük komisyonlu, hızlı ve inovatif çözümler sunmayı hedeflerler. Ayrıca, bu şirketler bankacılık sektörü erişimi olmayan kullanıcılara ek olarak kısıtlı erişimi olan ya da alternatif arayan kişilere ulaşmayı amaçlarlar.

Bu çalışmada, öncelikle Fintech şirketlerinin gelişimi ve geleceğinden bahsedilmiş ve çalışmanın ana ürünü olan Şimdi Al Sonra Öde (BNPL) ürünü detaylı olarak açıklanarak küresel BNPL sağlayıcılarından örnekler verilmiştir. Ardından kredi riski ve kredi riskindeki makine öğrenmesi tekniklerinden de söz edilmiştir.

Çalışma esnasında Londra merkezli bir Fintech şirketinin Türkiye'nin önde gelen perakende market zincirlerinden biri için geliştirdiği BNPL kredi ürününü kullanan yaklaşık 35.000 müşterisinin örnek verisi kullanılarak bunun üzerinden veri analizi yapılmıştır. Kullanıcılar için çok sınıflı bir segmentasyon problemi tasarlanmış, ve kullanılan kredilerin zamanında ödenip ödenmeyeceğini 3 farklı sınıfta tahmin etmek

için popüler makine öğrenmesi yöntemleri kullanılmıştır. Deney aşamasında Random Forest, Extreme Gradient Boosting ve Deep Learning algoritmaları test edilerek modellere ilişkin performans tabloları hazırlanmıştır. Model performansları karşılaştırılırken accuracy, f1-score, roc curve gibi değerler ile birlikte confusion matrisi öncelikli metrikler olarak belirlenmiştir.

Kredi riski modellemesi alanında sürekli olarak yeni yaklaşımlar ve yöntemler geliştirilmektedir. Bu gelişmeler, hem teorik hem de uygulamalı araştırmalar için geniş bir araştırma alanı sunmaktadır. Akademik çalışmalarda bu konuya odaklanmak, finansal sektördeki uygulamalı sonuçlar üzerinde önemli ve değerli katkılar sağlamaktadır. Özellikle, yeni modelleme tekniklerinin ve algoritmalarının geliştirilmesi, kredi riski değerlendirmelerinde doğruluğu ve güvenilirliği artırmakta, bu da finansal kuruluşların risk yönetimi süreçlerinde daha etkin kararlar alabilmelerine olanak tanımaktadır. Ayrıca, akademik araştırmaların pratik uygulamalarla entegrasyonu, sektördeki mevcut zorluklara yenilikçi çözümler sunarak, finansal istikrarı ve sürdürülebilirliği desteklemektedir. Bu bağlamda, kredi riski modellemesi üzerine yapılan akademik çalışmalar, hem teorik bilgi birikimine hem de finansal uygulamalara yönelik somut faydalar sağlayarak, sektörün gelişimine ve inovasyonuna katkıda bulunmaktadır.

Sonuç olarak, kredi riski sınıflandırması yaparken makine öğrenmesi modellemelerinin kullanımının oldukça avantajlı olduğu tespit edilmiştir. Bununla birlikte model seçimine karar vermeden önce her zaman şirketlerin strateji ve politikaları göz önünde bulundurulmalı ve aynı zamanda almak istedikleri batık kredi riskleri değerlendirilmelidir.

Anahtar Kelimeler: Fintek, Şimdi al sonra öde, Kredi risk modeli, Makine öğrenmesi

1. INTRODUCTION

One of the major challenges of developing economies is high credit costs and sources of funds. It is seen that the recording rate of 2024 borrowing interest data for G20 countries is 11.75% for South Africa, 5.5% for the United Kingdom, 4.75% for the Euro Zone, 4.5% for South Korea and 46.5% for Turkey [1].

It is stated that one of the most important obstacles for low-income and developing countries is the high cost of borrowing and difficulty in accessing financing. Figure 1.1 below shows the lending rates of a few selected G20 countries.

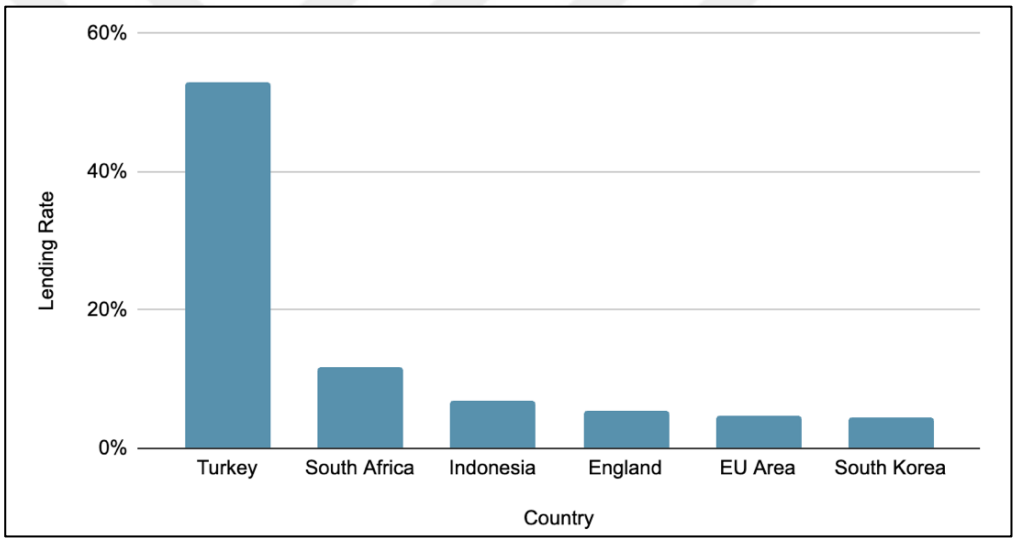

Figure 1.1: Lending rates for selected G20 countries (adapted from [1])

In addition, within the scope of the detailed tightening practices published in February 2024, the Banking Regulation and Supervision Agency (BRSA) has initiated tightening procedures in the credit policy to maintain financial stability; Among these steps, it is decided that the variability of risk weights and installment periods will not be applied to certain expenses in areas such as consumer loans, personal credit cards and vehicle loans [2]. At this point, Fintechs offer consumers savings, trade, insurance and investment services in addition to traditional banking services, thanks to new generation technologies, modern data analysis techniques and artificial intelligence development features They offer lower payments and faster loans while making

financial services more accessible to those not involved in the banking system or looking for an alternative. They aim to increase financial inclusion.

New business models that diversify credit risk forecasts are being created by using new generation technologies, big data analysis techniques and modern automation methods. At this point, Fintechs reveal their transformative effects on the financial system and reveal their potential to provide loans at lower costs while estimating credit risk with alternative data. At the same time, it promises to increase financial inclusion by including the population with difficult access to banking within the credit system. In addition, the risks brought by this new financial system transformed with Fintechs bring about various concerns on the part of economists and decision makers [3].

Machine learning models have found wide application in various industries in recent years. In particular, the predictive capabilities of these models and their ability to recognize patterns in complex data sets have led to a variety of uses in many fields. These uses are used to improve decision-making processes and increase efficiency in healthcare, finance, marketing, transportation and many other fields. In addition to all this, regulatory guidelines such as Basel and IFRS 9 should be considered for strategic implications. At the same time, care should be taken to ensure that the data sets used for these models comply with confidentiality and ethical rules [4].

In this article, data analysis was conducted on approximately 35,000 customers using the Buy Now Pay Later loan product, developed by a London-based Fintech company for one of Turkey's leading retail market chains. To segment this customer base and predict whether the loans used can be paid on time, the model training process is brightened by various lighting such as Xgboost, Random Forest and Deep Learning, In addition to metrics such as the date of first loan product use and the date of last loan product use, demographic information such as age was also included as a metric. The main purpose of the research is to divide the customer base into 3 different segments and prepare model performance tables for these models. Following the evaluation, recommendations for products are provided.

In the last decade, the extension from binary classification ($K = 2$) to multiple classification ($K > 2$) has become a research topic in its own right. Among the proposed methods, a basic distinction can be made between single optimization approaches and decomposition methods [5].

2. LITERATURE REVIEW

2.1 Financial Technology & Buy Now Pay Later

2.1.1 Overview of Fintech

2.1.1.1 Definition and evolution of Fintech

Financial technology companies, whose existence we feel strongly about today, especially after the end of the 2008 global economic crisis, actually date back to the past. It is possible to examine the joint development of finance and technology in three periods, starting from the 1800s. These periods are called Fintech 1.0, Fintech 2.0 and Fintech 3.0.

The period from 1866 to 1967 is called the Fintech 1.0 period, and the most critical determinant of this is the introduction of the telegraph system. This technology has allowed splintered local markets to grow into a national network.

The period between 1967 and 2008 is called Fintech 2.0. This period has evolved as online banking activities have increased and financial services have become more digital and global. The most important developments that emerged in the Fintech 2.0 period include Internet banking applications, credit cards and ATMs. In the early 21st century, banks' internal processes and interactions with customers became digital. During this period, fintech companies dominated the traditionally regulated financial services sector, primarily using technology to provide financial products and services. The Fintech 2.0 era continued until the end of 2008. However, with the global financial crisis, the curtain of the period began to fall. The causes of the crisis include factors such as uncontrolled growth of mortgage lending, unregulated markets, large amounts of consumer debt, the creation of "toxic" assets and the collapse of house prices. During this period, the national median home price fell by 29% between July 2006 and January 2009 [6,7].

The Fintech 3.0 period, which covers 2008 and later, refers to a period in which financial products and services began to be offered directly to businesses and the public, thanks to well-established technology companies and new initiatives. With the introduction of new-generation financial technologies, features such as reducing costs, transparency and "openness" have become an attraction to consumers. Startup growth has accelerated as third-party companies provide customers access to more financial services [6]. The historical development of fintech companies is shown in the Table 2.1 below.

Table 2.1: History of Fintech (adapted from [6,8]).

Fintech 1.0 | 1886-1967	Fintech 2.0 | 1967-2008	Fintech 3.0 | 2008 – present	Fintech 4.0 | Future
Transitioning from Traditional to Digital Platforms	The Evolution of Conventional Digital Financial Services	Rise of Startups	Future of Financial Technology
Telegraphs, railways, canals, and steamships supported cross-border financial connections, enabling the rapid transmission of financial information, transactions, and payments around the world. In 1866, the first transatlantic cable, Fedwire (1918) in the USA enabled the first electronic fund transfer. Additionally, the Use of Magnetic Stripe Cards began in the 1950s. The development of card technologies such as credit cards and debit cards is one of the factors leading to the digitalization of financial transactions.	This period, which started with the installation of Barclays Bank's ATM in 1967, continued with the NASDAQ establishment in 1971 and the SWIFT establishment in 1973. Towards 1984, the use of Bloomberg terminals gradually increased in financial institutions. While banks' internal processes and customer interactions were digitalized, this period ended with the 2008 Global Financial Crisis.	With the impact of the global crisis, the traditional banking system began to be seen as quite unsafe by the public. On this occasion, transparent solutions of Fintech and Startups began to be accepted as a ray of hope for consumers looking for a safe haven. A new era has begun with the adoption of smartphones and new generation digital solutions such as open banking. In 2009, Bitcoin made history as the first digital decentralized money. Google Wallet was introduced in 2011 and Apple Pay in 2014. Ant Financial, a China-based fintech company, was founded in 2014 and began offering financial services such as payment service, credit assessment and insurance.	Artificial Intelligence & Generative AI

2.1.1.2 Today's financial landscape and the future of Fintech

The global financial technology sector is rapidly growing and evolving to enhance financial activities using technology. Driven by market demands, traditional financial service providers are increasingly finding it difficult to meet increasing demands for digitalized and smart financial services. Therefore, the financial sector has significantly increased its investments in technology [8].

Global fintech funding between 2020 and 2023 is shown in the Figure 2.1 below. For 2023, 4.547 global fintech funding transactions worth $113.7 billion were recorded.

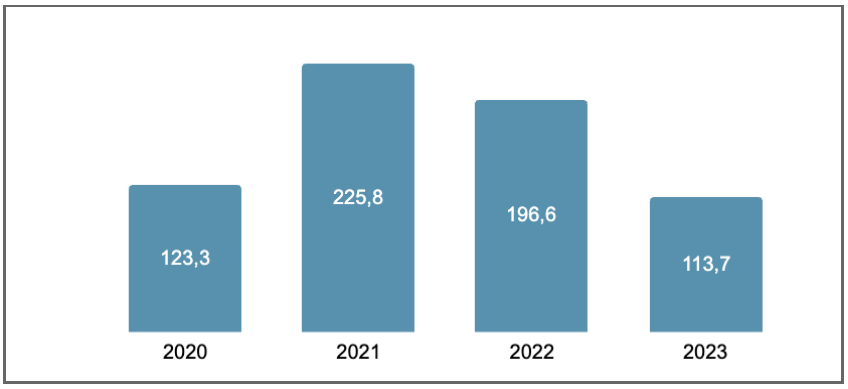

Figure 2.1: Total global funding value in Fintech between 2020 - 2023[1] (adapted from [11]).

At this stage, fintechs integrate the internet, finance and big data into their technologies and lead the overall change in the financial sector by using big data, blockchain, cloud computing, artificial intelligence and other emerging technologies.

Financial intelligence has a fast and accurate machine learning capability to enable intellectualization, standardization, and automation of large-scale business transactions. Thus, it can improve service efficiency and reduce costs. For example, AI technology combined with big data can integrate long-tail markets and improve the efficiency of fund allocation and financial risk management by reducing information asymmetry. Moreover, identity recognition and natural language processing technology could enable machines to replace workers and provide interactive services to customers. By combining financial intelligence with next-generation artificial intelligence, a "financial brain" can be built to provide inclusive financial services to meet the financial needs of ordinary people [9]. North America is the region producing the most fintech startups, with Asia in second place and followed by Europe in third place [10].

According to the KPMG H2'23 report, Fintechs stand out with their Payments, Insurtech, Regtech, Cybersecurity, Wealthtech, Blockchain/cryptocurrency and ESG/Greentech segments. When the payment field is examined in detail, it is observed that fintechs and investors have started to focus on regions such as Africa, South America and the Middle East, which are still developing in the field of digital payments. Real-time payments in the B2B field becoming visible, increases in BNPL

[1] Figure 2.1 data provided by PitchBook as of 31 December 2023.

product diversity, and various profitable payment companies working on public offerings are among the fintech trends for 2024 [11].

In addition, the use of Robo-advisors has increased and AI chatbots have become widespread to automate investment recommendations by reducing cost and increasing accessibility [12].

2.1.2 Overview of BNPL

2.1.2.1 Definition of BNPL

Buy Now Pay Later is a Fintech company loan product, which is a payment alternative that allows payments to be made in installments, often without being tied to traditional financial service providers. Customers can use this option at the payment stage and pay as a down payment or in installments when necessary. BNPL options have become very popular especially in online shopping in recent years. Different providers may offer a wide variety of products; For example, some may charge interest or additional fees, while others may be interest-free [13].

Since most BNPL providers do not charge interest or fees, they generate revenue through transaction fees, which they typically charge between 3% and 6% [14].

BNPL loan applications generally require being over 18 years of age, having a mobile phone number and having a debit card or bank account for payments. In addition, basic information such as name, surname, e-mail address, mobile phone number and date of birth is also frequently requested [15].

It is clear that BNPL is one of the pioneering and rapidly advancing products of the Fintech sector. BNPL is expected to represent 12% of all e-commerce sales on goods where it is currently focused and will be accounted for 680 billion dollars in transactions worldwide by 2025 [16].

It is observed that BNPL is generally common in young and disadvantaged geographies. Despite the rapid growth of the market and scrutiny from financial regulators, not much research has been done on BNPLs, and how they are used by consumers is not fully understood.

For example, there is no legal regulation regarding BNPLs yet, and many of them do not appear in credit debt files. In light of this information, it is predicted that BNPLs

may pose a credit risk for consumers. But debts to BNPL providers are increasing and there are doubts about customers' ability to repay [17].

In short, BNPL products aim to make customers feel like they are using a credit product but eliminate the need to manage and hold a credit card. In order to explain the rapid growth of BNPL, and to deeply understand the risks and opportunities that this growth will bring, it is necessary to understand why, for what and by whom BNPL products are used [18].

2.1.2.2 BNPL users

Although BNPL has increased critically in all generations between 2019 and 2021, there is a particularly high penetration in Generation Y and Z [19].

Consumer credit reporting agency TransUnion conducted an analysis on more than 4 million BNPL users in 2021. In this analysis, BNPL users were compared with average consumers, resulting in the following findings: 77% of BNPL users are in the 18-50 age group, i.e. a young age group; 69% have lower credit scores; They are generally 90+ days behind on their payments and are looking for more sources of credit [18,20].

According to the news in 2021, it was estimated that 45 million people aged 14 and over in the USA would use the BNPL service by the end of that year. In the chart Figure 2.2 below, it is estimated that by the end of 2025, almost 50% of Generation Z will have used BNPL services at least once.

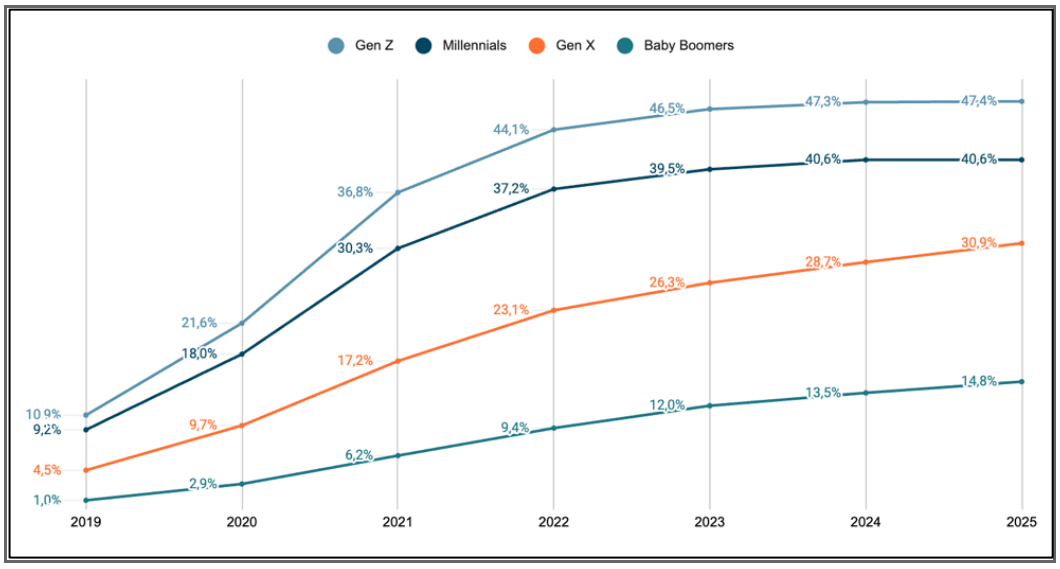

Figure 2.2: BNPL penetration by generation between 2019 - 2025

(adapted from [21]).

Note that Gen Z refers to individuals born between 1997 and 2012; Millennials refer to individuals born between 1981 and 1996; Generation X refers to individuals born between 1965 and 1980; Baby Boomers refer to individuals born between 1946 and 1964 [21].

2.1.2.3 Need for BNPL

According to a survey conducted by PYMNTS.com, 42% of respondents said they preferred BNPL due to "clarity of fees or interest rates." It was stated that the 5 categories in which survey participants preferred BNPL were "Clothing, Entertainment, Reading materials, Household goods and Food and Beverage" [22].

When the reasons why consumers use BNPL are examined, the most common reasons are "Avoiding Credit Card Interest" and "Shopping Outside of Budget and Insufficient Funds". In addition to these, reasons such as "Borrowing Without a Credit Check" constitute an important part of the reasons of the users. Additionally, other reasons such as "Credit Card Declined" are also detailed in Figure 2.3 below.

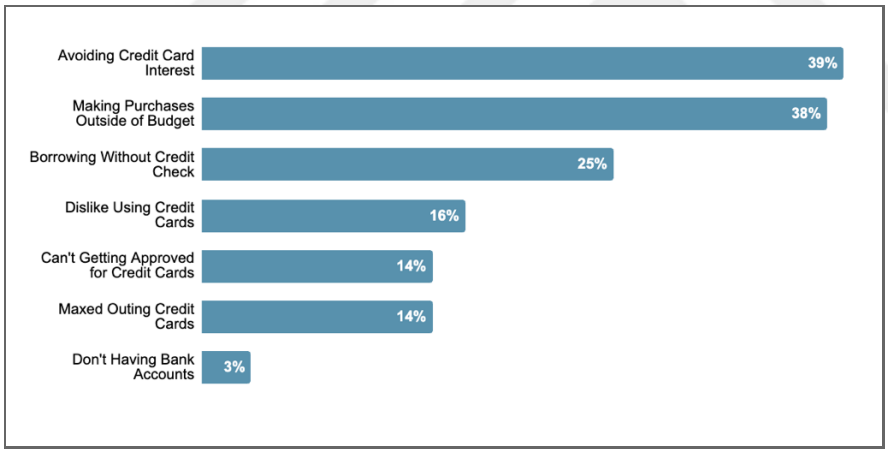

Figure 2.3: Reasons for BNPL usage (adapted from [21]).

2.1.2.4 Example of BNPL providers

The constant increase in living costs causes consumers to seek various online payment methods to meet their financial needs. In the BNPL awareness rankings in the USA, PayPal has made significant progress and ranked first, despite entering the market late. PayPal started offering its BNPL service in late 2020. Affirm, which has been providing BNPL services since 2014, ranks second in the same list.

In addition, Afterpay and Klarna companies, which have a very loyal customer base, also follow the first-place ranking closely as Figure 2.4 below [23].

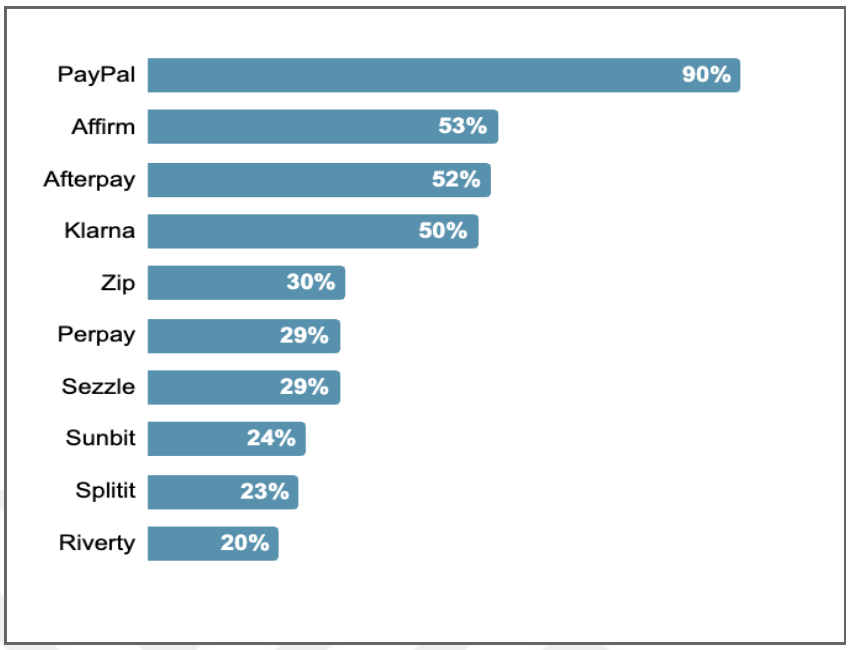

Figure 2.4: BNPL brand rankings in the U.S. (adapted from [23], November 2023; online survey; 1,249 respondents; 18 to 64 years).

Klarna

According to June 2023 forecasts, Klarna is expected to achieve $21.99 billion in BNPL payment value in the US by 2024, with 42.8 million users and an average spend of $513.47 per user. It is thought that Klarna's success in this field is due to its multi-channel approach and the popularity of its credit card and mobile application. In addition, Klarna's advertising campaigns with celebrities such as Lady Gaga and Paris Hilton were deemed very successful and brought Klarna popularity.

Affirm

Although Affirm's market share is smaller than Klarna's, its user spending is greater than Klarna's. According to eMarketer's 2024 forecast, 16.4 million users are expected to spend an average of $1,227.34. Affirm has also partnered with travel spending companies like Booking.com. The travel industry stands out as a rapidly growing segment for the BNPL product.

Afterpay

According to 2024 estimations, Afterpay is expected to reach $9.16 billion in payment value in the United States. To remain competitive, Afterpay focuses on AI-powered tools to deliver a personalized experience and shopping opportunities.

In addition to these examples, mobile wallet providers such as Google and Apple are also trying to capitalize on consumers' interest in BNPL. They strengthen and increase competition through various collaborations, mobile wallet integrations and personalized experiences [24].

2.2 Credit Risk

2.2.1 Overview of credit risk

The banking sector is more prone to be deeply affected by macroeconomic fluctuations and external shocks.

Credit risk is one of the main risks that banks and other financial institutions face in the markets. It can be expressed as the risk that the other party will not be able to fulfill its commitments [25]. When the expected principal and interest are not received, the risks of interruption of cash flows and increased costs for collection may also increase [26]. In addition to credit risk, there are also types of financial risks such as market risk, strategic risk or compliance risk. The credit risk management process generally includes identifying risks, assessing credit risk, processing the risk and implementing decided measures [27]. While banks are accepted as the cornerstone of economies, credit risk is one of the most important factors affecting the future of the banking sector [28].

2.2.2 Credit risk management

Banks can use various strategies to manage credit risk. For example, they may set certain criteria and standards in the lending process, and these requirements may include requiring borrowers to have a certain credit score. They can then periodically review their loan portfolios to track any changes in borrowers' creditworthiness and make appropriate adjustments as necessary [29]. According to the article of Gavlakova and Kliestik (2014), it is stated that managing credit sales has a direct impact on credit risk management. It can be stated that the most basic goal of credit risk management

is for credit providers to find the acceptable risk level before deciding to grant credit [30].

B2B and B2C activities of credit risk management can be examined in two types: Commercial Loans Risk Management and Consumer Loans Risk Management. The basic elements of credit risk management are stated as follows:

- Determining payment terms for customer segments,
- Determination and evaluation of customers' credit scores and worthiness,
- Evaluation of customers in critical situations and regular control of banned customers after negative experiences,
- Determining the basic principles for credit collection strategy and management

The credit risk management process can be examined under three main headings: "Checking Credit Limits", "Making the Credit Decision" and "Implementation of the Decision" [31].

Under the main heading of checking credit limits, credit rating agencies' reports, recommendations from other companies and other alternative data come to the fore for potential customer groups, while payment history and creditworthiness calculations made with in-company models gain importance for regular and loyal customer groups. By credit decision support systems; It is recommended to evaluate data such as determining credit limits, maximum maturity, payment method and collateral. Finally, it was mentioned that the role of IT should be evaluated during the implementation of decisions and the need to make current revisions of contracts and loan conditions.

Effective credit risk management, which is of critical importance for every company, should take into account the diversity of business partner creditworthiness. Credit standards set for independent consumer groups encourage an individual approach to customers. Customers can be assigned to predetermined credit groups based on objective criteria. In addition to previous experience, the knowledge level of new customers can also be obtained from existing credit records. Credit limits set for each customer class can help identify potential bankruptcy risks in a timely manner. Properly selected credit standards form the basis of effective credit management and will provide a powerful tool for managing the company's credit risk [32].

Loan returns have a distribution where there is a high probability of making a small profit but a low risk of a large loss. Therefore, credit risk management is a critical element that must be done comprehensively and accurately [25].

Financial credit risk assessment is very important to predict the risk category or bankruptcy of a company or bank in the coming years. It is generally accepted that two main aspects should be evaluated: credit rating/scoring and bankruptcy prediction. While credit risk measurement models can be classified as qualitative and quantitative; Details such as past payment performances, financial leverage, earnings fluctuations, collateral and interest rates are included in the qualitative classification. As a result of all these calculations, it is aimed to make a subjective evaluation as to whether users can use credit. Obtaining and statistically modeling credit scores with quantitative models can be difficult because default risks are generally small. For this reason, many different models have been put forward and tested for credit risk modeling. However, the statistical methods used since the beginning of the concept of credit risk are "gradually decreasing" in effectiveness because they are based on limited assumptions such as a large sample size, normally distributed independent variables and linear relationships between all variables [33].

Credit risk modeling made with traditional modeling methods may cause disadvantages such as financial losses that may occur due to the credit risk model, the cost of incomplete/incorrect estimations, and prolongation of the credit granting process/customer loss [34].

Traditional risk models often take a “one-size-fits-all approach” when evaluating borrowers. They often use a set of predetermined criteria to assess creditworthiness, which can lead to a standardized process that does not take individual cases into account. This approach can lead to missed opportunities and increased risks for banks.

2.2.3 Importance of AI and Fintech in credit risk

Artificial intelligence has initiated a rapid transformation in the banking sector, and this change has brought with it several advantages and disadvantages. Significant advances have been made, especially in areas such as machine learning technologies, personalized recommendations, 24/7 customer service and virtual assistants. These technologies are making advances in credit scoring and risk management by effectively analyzing big data [35].

In addition, Fintechs, which are already technology companies alongside banks, offer different advantages such as big data, artificial intelligence, blockchain and cloud computing, which are the basic technologies in financial risk management. For example, the edge computing[2] capabilities of big data and cloud computing enable gaining valuable insights by processing large and diversified data sets. This plays an important role in creating customer profiles and developing risk analysis. Edge Computing is a process in which data is analyzed and information deemed worthy of processing is quickly obtained. Particularly for financial institutions, the use of these technological capabilities enables better understanding and effective management of customer risks. Therefore, Edge Computing stands out as an important element of technological developments in financial risk management.

It is anticipated that financial technology companies can improve financial risk management with big data, artificial intelligence and other technologies. Big data and artificial intelligence can be seen as a good opportunity to conduct detailed risk analyzes and create customer-focused risk personas. In dynamic systems, support can be received from Fintechs to detect high-risk transactions and make sense of anomaly transactions. While risk management is improved with risk control models and optimizations, decreasing trends in credit and bankruptcy risks can be observed. In light of all this, with Fintechs and technological advances, resources can be used efficiently while risks in the financial system can be reduced.

Before the financial crisis in 2008, Fintech startups were measured to have no significant impact on the non-performing loan rate. When the process from 2008 to 2019 and from 2020 to the present is examined in detail, unlike before 2008, the impact of Fintech initiatives on the non-performing loan rate has become increasingly important. It can be stated that this situation is caused by the increasing risk awareness of the banking sector and the acceleration of the use of Fintech’s with the 2008 crisis. Finally, in the period between 2020-2021; It can be said that banks' willingness to invest in Fintechs has increased significantly, thanks to the increasing competition in

[2] Edge computing is an architecture that enables data to be processed closer to the network's edge rather than in central servers. This reduces latency and optimizes bandwidth usage. It is especially used in IoT devices and real-time applications.

the markets and the transformation trend of the industry with the impact of the Covid-19 epidemic [36].

2.2.3.1 Usage of ML in credit risk

With the rapid growth of financial data after the 2007-2008 global financial crisis, the role of big data analytics in accurate and effective default prediction has become increasingly important. With big data analytics, performance capacities have increased and various platforms that can perform deep analyzes for more accurate and reliable results have emerged.

Credit scoring plays an important role in determining companies' default prediction. Different machine learning-based solutions for credit scoring have been presented in the literature. Nowadays, the results of default and bankruptcy prediction models developed with machine learning algorithms such as DT, RF, SVM and MLP attract much attention. Depending on different company sizes and data volumes, various experimental results have been obtained and compared. Studies indicate that hybrid models perform better than single machine learning models. These developments highlight the effectiveness and importance of machine learning techniques in predicting default and bankruptcy [37].

In other words, credit risk can be calculated by analyzing non-linear relationships between financial data in balance sheets. When optimizing accuracy with models, a balance must be maintained between model predictions and reality according to the data science lifecycle [38].

In the IMF Working paper study, the most important machine learning applications that have achieved success in credit risk analysis are listed under 3 classes: tree-based models, support vector machines, and neural networks. Based on this, in this study, the preferred machine learning methods were tested and their performances were compared.

2.3 Materials

At this stage of the study, common machine learning algorithms such as Random Forest, XgBoost and DL, which are necessary for the application phase, were examined and the basic principles of these models and how they work were explained.

Then, detailed reviews were made for the entire process, from data preparation to comparison of model performance results.

In general, the problem of model evaluation and selection in machine learning becomes challenging due to the high diversity of various model types and hyperparameters. The emergence of Deep Learning has worsened this problem because deep learning has a high number of hyperparameters and trainable weights.

In multi-class classification problems, the default best model is often chosen based on a single metric; for example, on overall or average accuracy, the model that performs best on average over all classes is selected. However, it should be noted that the most appropriate model for each application may not necessarily be determined based on the selected metric. The costs of different classification errors may vary, and a single metric may not be sufficient to fully understand a model's results or elucidate classification errors. In real-world applications, model accuracy alone is not a selection criterion. Factors such as model complexity, interpretability, scalability, and processing time should also be taken into account. Therefore, the model with the highest accuracy may not always be the most suitable option, but the most appropriate model should be selected based on these criteria [39].

2.3.1 Algorithms

Machine learning is a relatively new scientific discipline developed to mimic human decision-making processes and generate new knowledge by learning from past experiences. This discipline draws on knowledge from various fields such as statistics, optimization, engineering and computer science, and this interdisciplinary approach has increased success in machine learning.

In recent years, the rapid development of statistically successful machine learning algorithms has made this technology indispensable in solving many complex problems in modern society. These developments have led to significant transformations in finance, healthcare, finance, automotive and many other sectors [40].

2.3.1.1 Decision Tree

Decision Trees, a widely used method for data classification and prediction, recursively partition multidimensional spaces in a nonparametric manner.

Based on the analysis results, they form a hierarchical tree-like structure as follows, and this structure contains branches and leaves. This framework is used to model decision-making processes and predict new cases in the fields of data science and machine learning. The general structure of the Decision Tree algorithm is shown in the Figure 2.5 below.

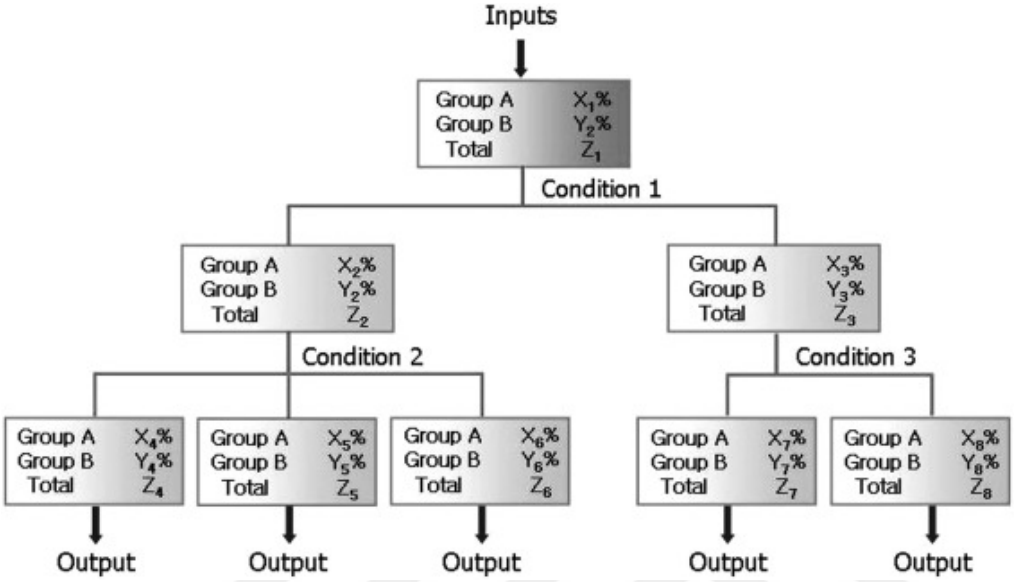

Figure 2.5: General structure of Decision Tree (adapted from [41]).

Decision Trees offer a number of advantages over other AI models. First, they can produce powerful results without requiring variable transformation or modification in nonlinear models. This is important to identify nonlinear relationships in the data set and integrate them into the model. They can also be trained in less time while achieving similar levels of accuracy compared to other more complex models. This is important to achieve fast and effective solutions on large data sets.

Decision Trees also reveal variable importance in a clear and analysable way. This is very important to understand the factors that affect the predictive ability of the model and to improve the performance of the model by focusing on important variables. As a result, Decision Trees have a wide range of applications and provide powerful and fast results that are easy to analyze [41].

2.3.1.2 Random Forest

Random Forest, a method proposed by Leo Breiman in the 2000s, creates an ensemble of predictors with a set of decision trees growing on randomly selected subspaces of the data [42].

The Random Forest classifier is created by combining a series of decision trees. Each tree is uniquely constructed using the training set and random vector k. The Random Forest algorithm forms a collection of classifiers named h(x, k) with k=1 independent of the input vector x and a different tree structure [43]. The general structure of the Random Forest algorithm is shown in the Figure 2.6 below.

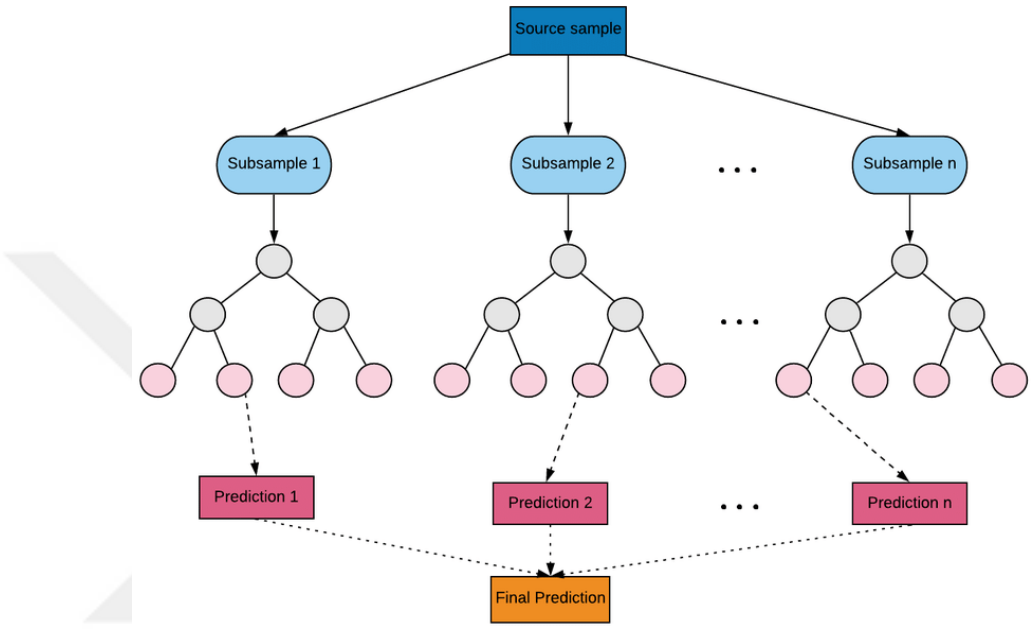

Figure 2.6: General structure of Random Forest (adopted from [44]).

Random Forest algorithm has the following advantages when used for classification problems:

1. **Overfitting Reduction:** Decision trees generally tend to overfit, but since Random Forest algorithms contain many decision trees and these trees are averaged without correlation, the risk of overfitting is reduced.
2. **Flexibility:** Random Forest can be used in both regression and classification models. Thanks to this feature, it is popular among data scientists.
3. **Feature Engineering:** Random Forest makes it easy to evaluate the contribution of variables to the model. This feature is valuable to data scientists when analyzing model performance.

These advantages enable Random Forest to be used effectively in classification problems [45].

2.3.1.3 XgBoost

XgBoost is one of the well-known algorithms used to solve regression and classification problems. It's recognized as a gradient boosting decision tree implementation and is particularly preferred for faster and more competitive performance [46].

XGBoost is a decision tree ensemble suitable for large-scale applications. It is gradient-boosted and optimizes the objective function cumulatively. This function can be integrated into the split criteria of decision trees, thus ensuring simpler trees.

Additionally, other strategies can be employed to reduce the complexity of trees. Moreover, random sampling techniques are applied in XgBoost, reducing overfitting and increasing training speed.

XgBoost focuses on reducing computational complexity in split finding algorithms to increase the training speed of decision trees. Specifically, it utilizes a compressed column-based structure to avoid repeatedly sorting the data at each node. This structure enables parallel discovery of the best split. Moreover, XgBoost employs a method that tests only a subset to find splits, reducing computational costs [47]. The general structure of the XgBoost algorithm is shown in the Figure 2.7 below.

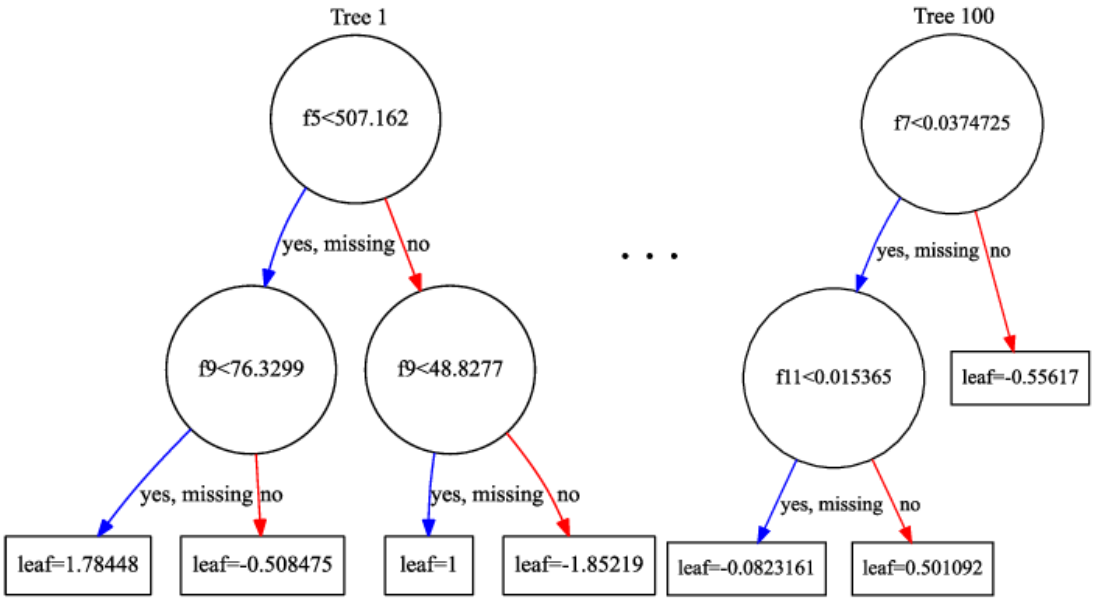

Figure 2.7: General structure of XgBoost (adapted from [48]) (xgboost-based algorithm interpretation and application on).

2.3.1.4 Deep Learning

Deep learning is a technique used in machine learning. This technique is based on identifying different aspects of the input data using multiple layers to extract features from the raw data. Deep learning techniques include convolutional networks, recurrent neural networks, and deep neural network.

In the past, the use of machine learning was limited by its inability to process raw input data. However, deep learning has the ability to operate on large data sets to overcome this limitation, making it an effective and useful technique for machine learning. Additionally, deep learning has gained momentum due to advances in computer hardware. Deep experience is required in feature extraction to transform raw data into a suitable form. This helps subsystems recognize and classify raw data [49].

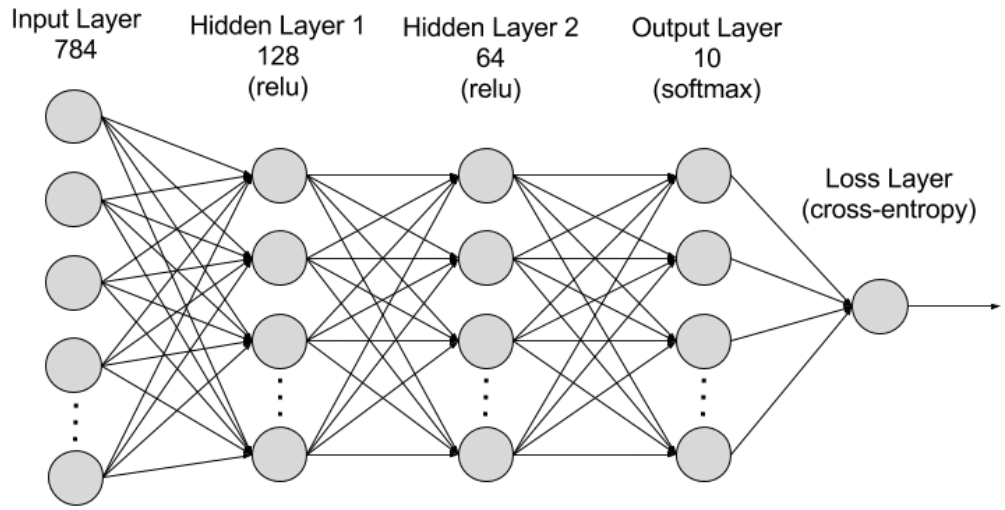

Figure 2.8: General structure of Deep Learning (adapted from [50]).

2.3.2 Data preparation

The data preparation phase is very important before testing model algorithms to obtain accurate predictions. This stage is an important step to reduce performance losses that may be caused by missing or outlier values.

The data preparation process generally consists of four basic steps, namely: separating features and target, checking and filling missing values, handling outliers, and splitting the dataset into training and testing dataset as shown in the Figure 2.9 below.

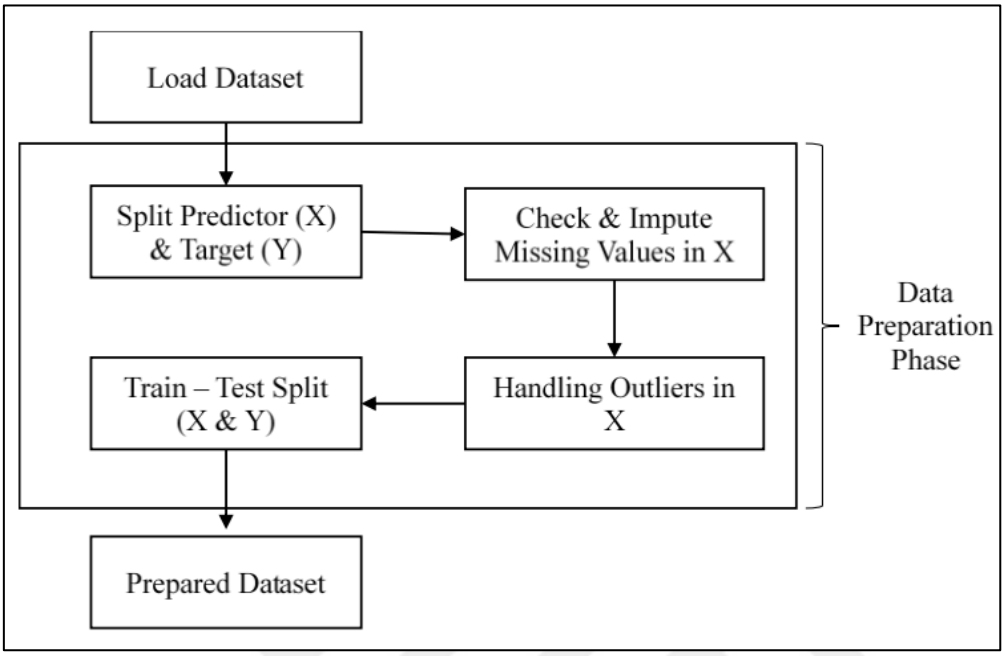

Figure 2.9: Data preparation phase flow chart (adopted from [51]).

Separating Features and Target: The dataset is processed by separating the features (independent variables) and the target variable (dependent variable) of the observations. This step allows the model to analyze the relationship between the independent variables and the dependent variable.

Checking and Filling Missing Values: It is checked whether there are missing values in the data set and if so, these values are filled. Filling in missing values increases the accuracy of the model and is important to obtain reliable results.

Handling Outliers: Outliers are values that are generally significantly different from other observations in the data set. These values may affect the performance of the model and cause biased results. Therefore, it is important to identify and address outliers.

Separating the Dataset into Training and Testing Dataset: The dataset is divided into two parts for training and testing the model: training dataset and testing dataset.

The training dataset is used in the learning process of the model, while the testing dataset is used to evaluate the performance of the model.

Correct execution of these steps increases the accuracy of the model and helps to obtain reliable results. Therefore, care should be taken in data preparation.

2.3.3 Feature engineering and correlation matrix

At this stage, it is aimed to reduce the number of features to be used in the low modeling process. It is an important step in using computer resources efficiently and obtaining fast results [51].

The correlation matrix is a (K×K) square and symmetric matrix that shows the relationship between columns i and j of X. High values in this matrix indicate serious multicollinearity among the variables of interest. However, the absence of extreme correlations does not mean that linearity is lacking.

For example, for multiple regression, regressor variables may be highly multicollinear with each other, even if no pairwise correlation has a large value. A variable may be approximated as a linear function of four other variables, but there may be no pairwise correlation between these variables. Therefore, pairwise correlations are inadequate for multicollinearity diagnoses.

Examining the eigenvalues and eigenvectors of the correlation matrix provides a more robust method for detecting multicollinearity. This forms the basis for the number of conditions [52]. Since it can be easily implemented in Python or any programming language, it is one of the most popular statistical techniques preferred to choose which data and features will have the most impact in the created machine learning models [53].

2.3.4 K-fold cross validation

One of the most frequently used methods to prevent overfitting in machine learning prediction models is cross validation [54].

With this method, the data set is divided into k different clusters and calculations are made with random samples without repetition. Here, the model is trained using k-1 subsets. Afterwards, the model is applied on the last remaining subset and this process is repeated [55]. It has been stated that the most common K values are 5 and 10. It has been observed that when the K value is small, the outputs obtained are dependent on the K value. Similarly, when the K value is chosen large, the calculation becomes costly [56].

For instance, k=10, the process is shown in the Figure 2.10. For the first layer, the cluster in the first row is used as the test data set and the others are used as training data. Similarly, in the second layer, the second subset serves as the validation set and the others serve as the training set. For the 10-fold cross validation algorithm, this process is repeated 10 times.

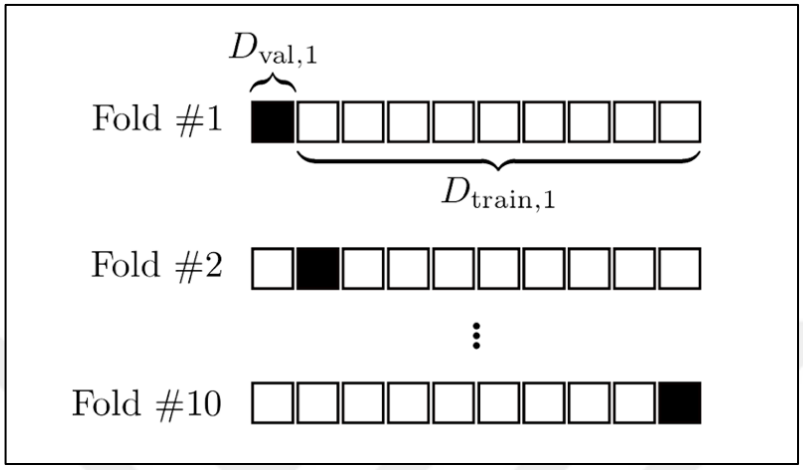

Figure 2.10: 10-fold cross-validation (adapted from [55]).

In addition, The StratifiedKFold method can be used during k-fold cross-validation to address unbalanced class distributions in datasets. This method provides a more reliable assessment of model performance by preserving class distributions within each layer and addressing the problem of random sampling [57].

2.3.5 Model evaluation

Classification models are evaluated by creating confusion matrix, accuracy, sensitivity, specificity, ROC curve and Kolmogorov-Smirnov chart.

2.3.5.1 Confusion matrix

Since this study was prepared to find a solution to a 3-class multi-class classification problem, a 3x3 dimensional confusion matrix was created. Each column represents the classes into which the predictions made are classified, while each row represents examples of the actual classes. Cell (i, j) in the confusion matrix indicates the number of samples where the predicted class is j and the true class is i. In this context, the confusion matrix shows what errors a classification model makes when making predictions. Along with these errors, it also provides insight into the type of errors that occur [58].

Example of multiclass classification problem confusion matrix table is shown as Figure 2.11 below.

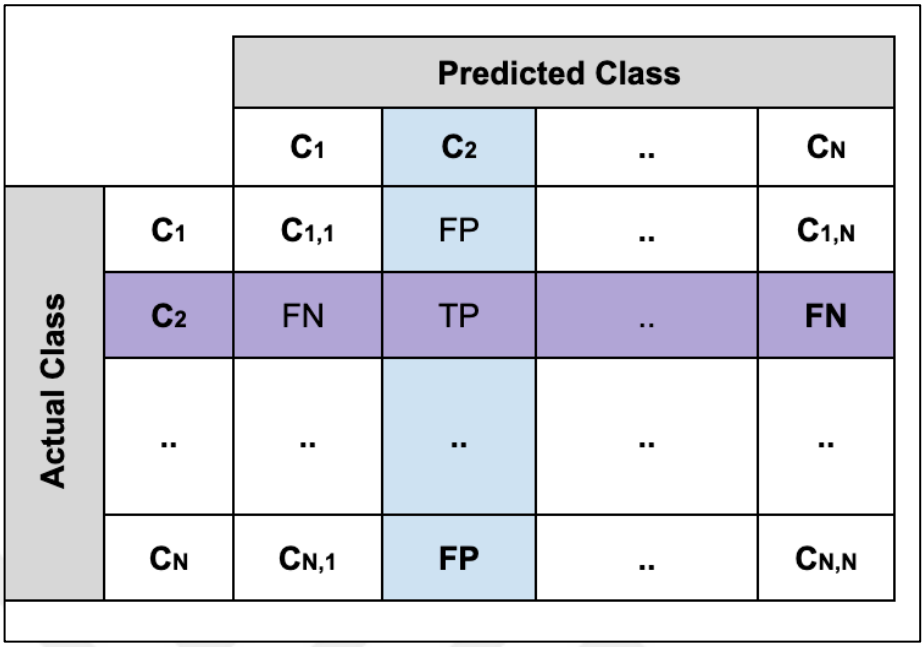

Figure 2.11: Example of multiclass classification problem confusion matrix (adapted from [58]).

As seen in the figure above, in the case of multiple classification, the metrics defined for binary classification are not fully valid. The multiclass confusion matrix size is NxN, where N is denoted as the number of different class labels.

2.3.5.2 Performance metrics

The figure below provides an overview of the metrics calculated for multi-classification problems. Specifically, accuracy, recall, precision, and F1 score were noted.

Similarly, performance metrics for a multiclass confusion matrix figure is shown as Figure 2.12 below.

Metric	Formula
Accuracy	$Acc(Areduced) = \sum_{i=1}^{N} TP(Ci) / \sum_{i=1}^{N} \sum_{j=1}^{N} Ci,j$
Recall of Class C*i*	$TPR(Ci) = TP(Ci)/(TP(Ci) + FN(Ci)))$
Precision of Class C*i*	$PPV(Ci) = TP(Ci)/(TP(Ci) + FP(Ci)))$
F1-Score of Class C*i*	$F1(C1) = 2.TPR(Ci).PPV(Ci)/TPR(Ci).PPV(Ci)$

Figure 2.12: Performance metrics for a multiclass confusion matrix (adapted from [58]).

2.3.5.3 ROC curve and Kolmogorov-Smirnov chart

One of the preferred methods to show how well the model can predict each class and to evaluate binary classification models is the ROC Curve. Similarly, the Kolmogorov-Smirnov (KS) score is a scoring system that measures how different the prediction is between two classes. If the model can correctly separate the data between two classes, the KS score is 100; If it cannot produce a significant prediction, the KS score is 0. This score is visualized by two lines representing each class in the dataset [59].

When evaluating multiple classification models, metrics such as ROC and KS used in binary classification needs to be adapted to be more understandable. Methods such as OvR (One-vs-Rest) and OvO (One-vs-One) can be used [60].

In this study, using the OvR strategy, a method that compares multi-class models with all the others simultaneously for statistical performance evaluations was preferred. This was calculated by converting the multiclass classification output into binary classification output. Process is repeated for each class in the data set. As a result, 3 different OvR scores are created, and an OvR score is obtained by calculating their weighted average.

3. METHODOLOGY AND IMPLEMENTATION

This study has been prepared for a multi-class segmentation solution designed to solve a business problem after it is identified. In the methodology section, the approach and steps taken to solve this problem are explained in detail.

Additionally, the study workflow diagram is summarized and stated in the Figure 3.1 below.

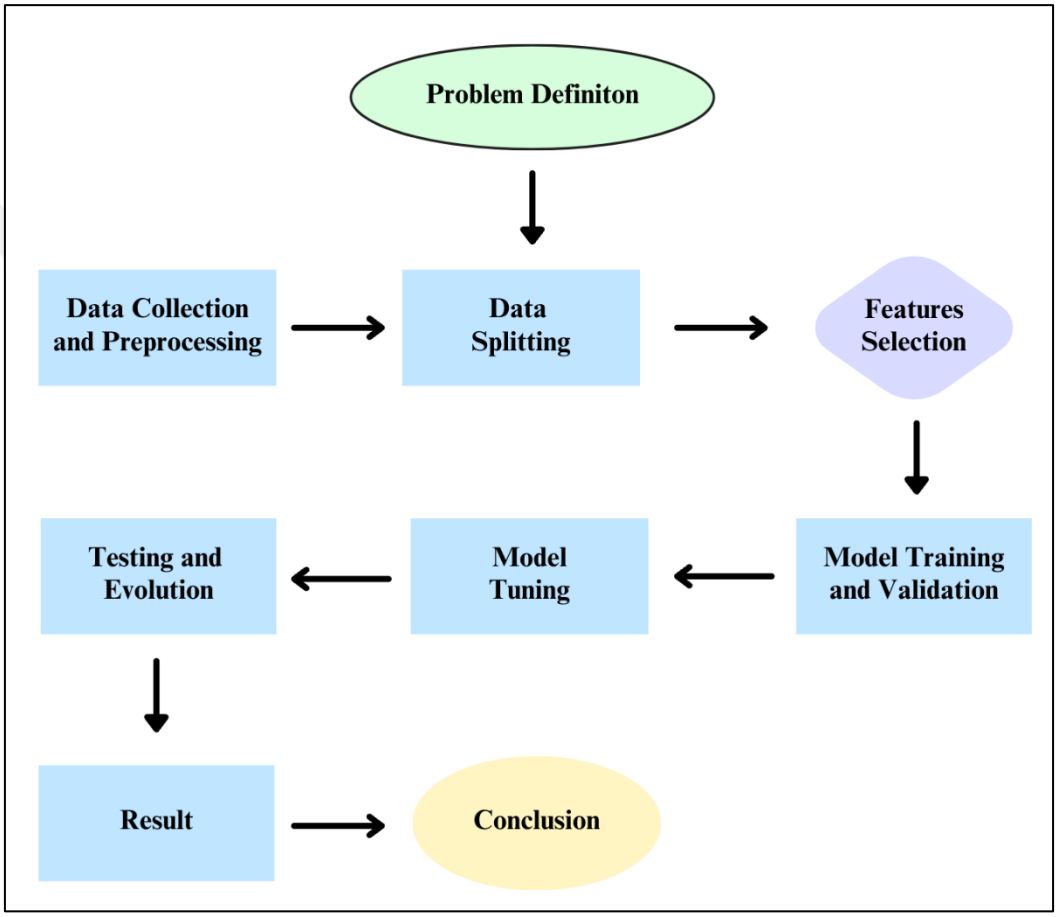

Figure 3.1: Workflow diagram (prepared by the authors).

The data preparation process generally consists of 8 basic steps, namely: problem definition, data collection and preprocessing, data splitting, features selection, model training and validation, model tuning, testing and evolution, result and final conclusion.

First, the data collection process and the data sources used are explained.Then, detailed data analysis stages are explained, along with the methods used to process the data and extract features.

In addition, the reasons for choosing the Random Forest, XgBoost and Deep Learning machine learning models used in the experiment phase, how these models were applied and the finalization process were mentioned.

While preparing the data for this research, a data sample of a Fintech company providing services as a BNPL provider in Turkey was used, taking into account all legal compliance. Generally, users can download mobile applications developed by Fintech companies and register by entering the necessary information (GSM number, e-mail, ID number, etc.).

A scoring algorithm runs for customers and financial risk is calculated. If the customer completes the KYC process, he can use some or all of his calculated limit in designated businesses. The information required for legal compliance to complete KYC processes may vary. The BNPL loan product used is for 30 days and customers must pay within 30 days. Approximately 7 months of shopping data between August 2023 and March 2024, in cooperation with Turkey's retail company and Fintech, were included.

Python, an open source programming language that offers a broad data science ecosystem, was used for data analysis and implementation of machine learning models. While commonly used libraries such as pandas and NumPy were used for data processing and feature engineering, matplotlib for visualization and libraries such as scikit-learn and TensorFlow were preferred for the application of machine learning models.

The basic infrastructure of the study is created using Python and these libraries which were used in almost every step of the analysis process.

3.1 Data Analysis and Visualization

Before moving on to the modeling process, a data set with 20 variables is created from the raw data and defined as in the Table 3.1 below.

Table 3.1: Data summary.

Order	Features	Definition
1	operator_flag	Classifies Turkish mobile numbers by operator.
2	age	Contains users' age.
3	email_flag	Indicates whether users share email information with the company.
4	regandact_diff	Difference between registration and first credit date.
5	actandlatest_diff	Difference between first and latest credit date.
6	total_credit	Total number of credits used by users is included.
7	credit_id	Unique ID is assigned to each credit.
8	credit_date	Date of credit usage.
9	user_id	Unique ID is assigned to each user.
10	online_flag	Indicates whether the loan used is online or physical.
11	amount	Total amount of the credit is included.
12	payment_amount	Total amount of the payment is included.
13	latefee	Total amount of the calculated late fee is included. (for unpaid overdue credit)
14	servicefee_flag	Indicates whether a service fee is charged from credit users.
15	credit_diff	Difference between the users' credit date and '2024-05-01'
16	due_diff	Difference between the users' due date and '2024-05-01'
17	payment_diff	Difference between the users' payment date and '2024-05-01'
18	max_payment_amount	Highest loan amount the user has ever used at one time.
19	max_datediff_dueandpayment	Maximum number of days difference between the user's loan and payment date.
20	min_datediff_dueandpayment	Minimum number of days difference between the user's loan and payment date.
21	target	Target is divided into 3 classes: users who pay on time, those who pay later and those who do not pay at all.

Analysis was made on a total of 36,157 different BNPL users, 423,241 BNPL loans and 181,489,277 TL BNPL volume. To better understand the data, a detailed analysis was carried out using several different graphs. According to verified user demographics, the youngest user is 18 years old and the oldest user is 91 years old. While the age group distribution of BNPL loan users is shown in the Figure 3.2 below, it is understood that most of the users are customers between the ages of 26-35.

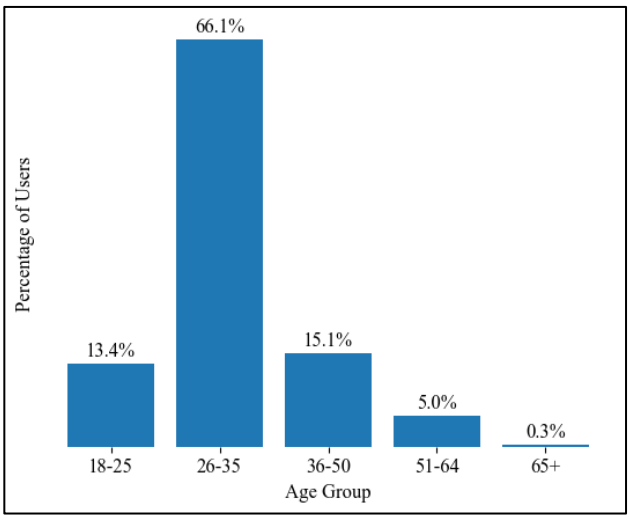

Figure 3.2: User count by age group.

In another chart, 4 different operator classifications for users were examined, and it is understood that the line origins of approximately half of these users were Turkcell.

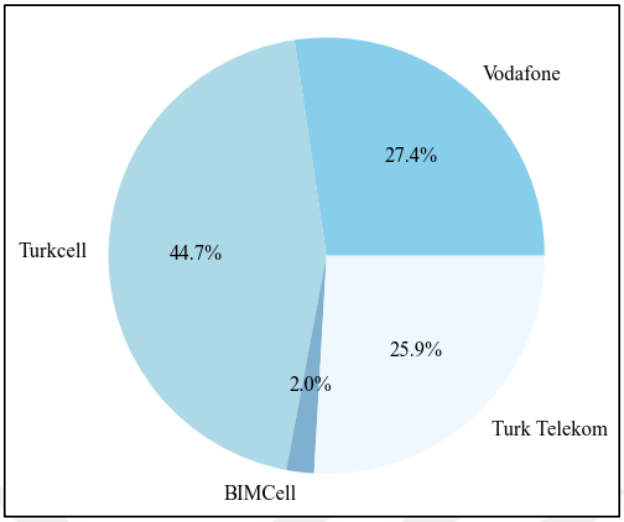

Figure 3.3: User-based operator distribution.

While evaluating the loans used since August, loans under 10 TL were excluded from the data as they were test data. For this reason, the lowest loan amount is 10 TL and the highest loan amount is 10,000 TL. (Shopping with BNPL using more than 10,000 TL is not allowed.) Monthly credit usage and credit user graphs are given in Figure 3.4 below.

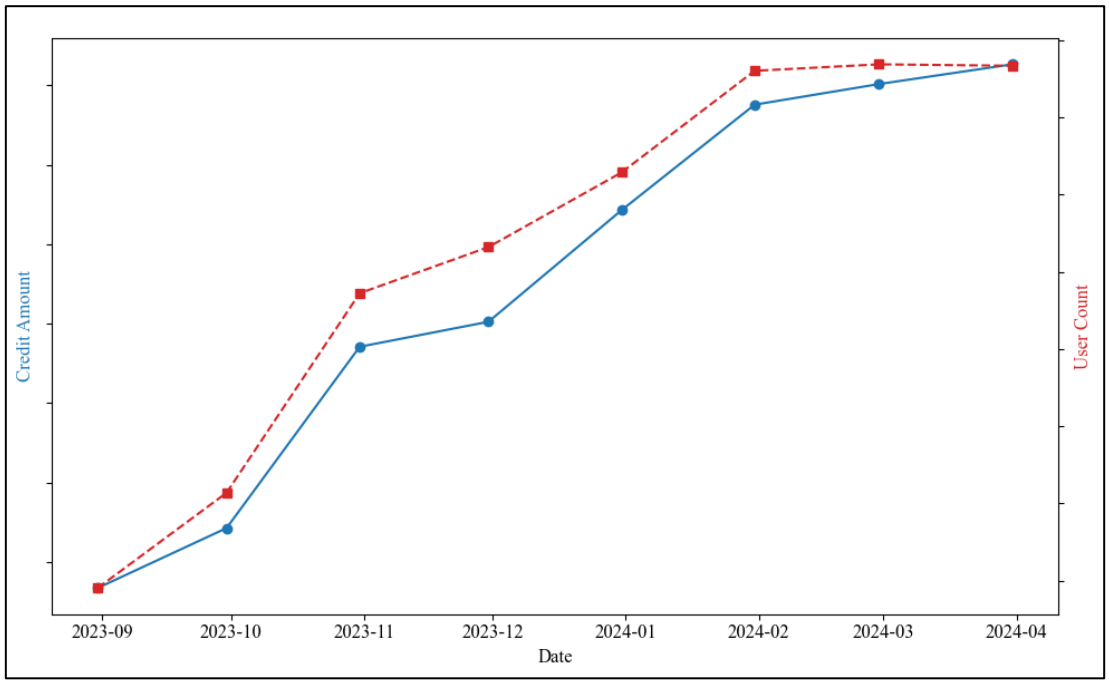

Figure 3.4: Monthly credit amount and user count.

It has been observed that BNPL usage has been increasing both on a user basis and in volume since August 2023. Since there is no legal regulation policy specific to the BNPL product, this increase also causes an increase in the risk of users not being able to repay.

3.2 Model Preparation

The statistical values of the data to be used in model training are shown in the Figure 3.5 below. With this summary, basic statistical values such as count, mean, std (standard deviation), min (minimum value), 25%, 50% (median), 75% and max can be accessed for the numerical columns in the data set.

	count	mean	min	25%	50%	75%	max	std
operator_flag	423240.0	2.248993	0.0	1.0	2.0	3.0	4.0	1.105715
age	423238.0	36.117804	18.0	28.0	35.0	43.0	91.0	10.376598
email_flag	423240.0	0.082525	0.0	0.0	0.0	0.0	1.0	0.275164
regandact_diff	423238.0	60.052181	0.0	1.0	8.0	46.0	1188.0	126.37097
actandlatest_diff	423238.0	556.173838	0.0	411.0	502.0	722.0	1307.0	225.998244
total_credit	423238.0	110.576475	1.0	39.0	78.0	145.0	1129.0	108.829379
credit_id	423240.0	10816005.564401	10143635.0	10480333.25	10816877.0	11151895.75	11485576.0	387501.2337
credit_date	423240	2024-01-01 04:37:15.939892224	2023-08-01 00:00:00	2023-11-16 00:00:00	2024-01-09 00:00:00	2024-02-18 00:00:00	2024-03-31 00:00:00	NaN
online_flag	423240.0	0.613836	0.0	0.0	1.0	1.0	1.0	0.486869
amount	423240.0	428.807292	10.0	125.9	256.28	512.25	10000.0	578.079209
payment_amount	398458.0	443.272636	0.01	131.25	267.07	532.0475	10761.76	590.995554
latefee	423240.0	6.008196	0.0	0.0	0.0	0.15	1858.12	34.630981
serviceefee_flag	423240.0	0.683059	0.0	0.0	1.0	1.0	1.0	0.465285
credit_diff	423240.0	120.807454	31.0	73.0	113.0	167.0	274.0	57.855453
due_diff	423240.0	90.753709	0.0	43.0	83.0	137.0	244.0	57.833703
payment_diff	396657.0	99.289779	0.0	47.0	90.0	148.0	272.0	63.096467
max_payment_amount	421578.0	2175.856751	0.25	801.17	1340.99	2284.69	95380.0	4800.421677
max_datediff_dueandpayment	421472.0	27.87365	0.0	24.0	29.0	30.0	673.0	10.879105
min_datediff_dueandpayment	421472.0	29.275629	0.0	3.0	11.0	29.0	937.0	55.91457
target	423240.0	1.667215	0.0	1.0	2.0	2.0	2.0	0.589624

Figure 3.5: Data info.

3.2.1 Target details

Segmentation was done for model development. This segmentation process created a basic framework for model developments and identified three distinct target classes: PAID, LATE, and UNPAID. These classes represent users who pay their BNPL loan on time, pay after the payment date, and those who have not made a payment.

The PAID class includes those who make their loan payments on time, the LATE class includes those who pay after the payment date, and the UNPAID class includes those

who still have not paid after the payment date. This segmentation is an important step to understand loan payment behavior and develop appropriate strategies.

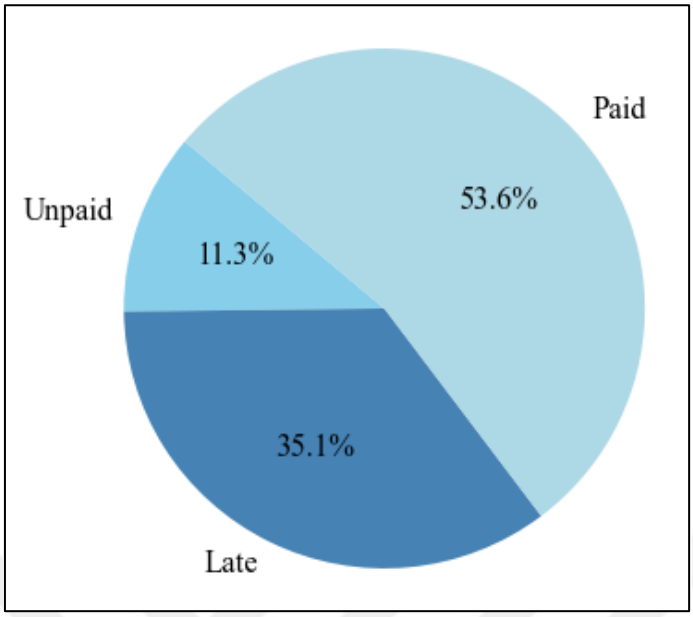

Figure 3.6: User count by target group.

3.2.2 Correlation matrix and features engineering

In the model training process, feature selection is an important step. After completing the steps of transforming existing features, creating derived features, and processing missing data, a correlation matrix was drawn to understand the relationships between the selected features. While this correlation matrix, seen in the Figure 3.7, visualizes the relationship of each feature with the others, the features were selected by taking into account the potential multicollinearity that may affect the performance of the model.

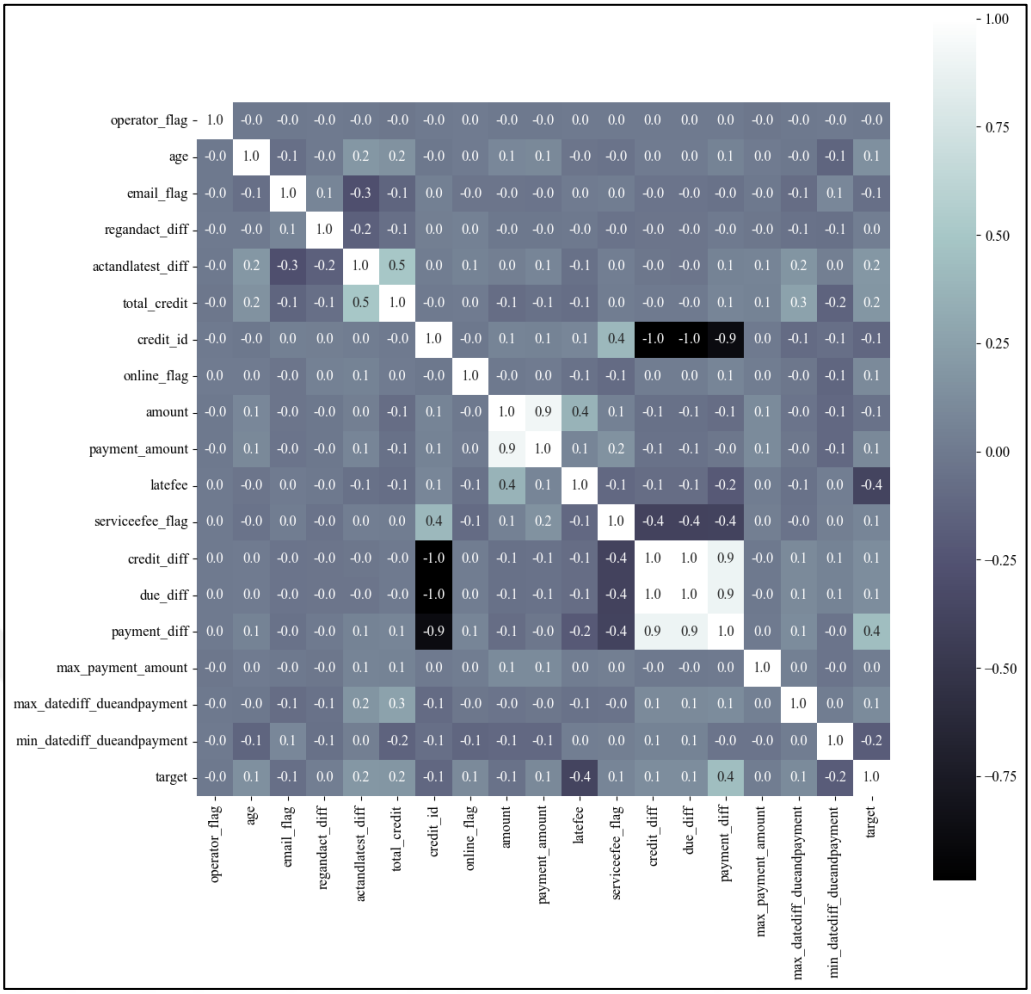

Figure 3.7: Correlation matrix for all features.

At the beginning of the study, there are approximately 20 features. Only a few features that are strongly related to each other have been selected to improve model training and performance and to eliminate unnecessary complexity. It is decided to test the model with 12 features for the experimental phase. The correlation matrix of these features has been recreated in the Figure 3.8 below.

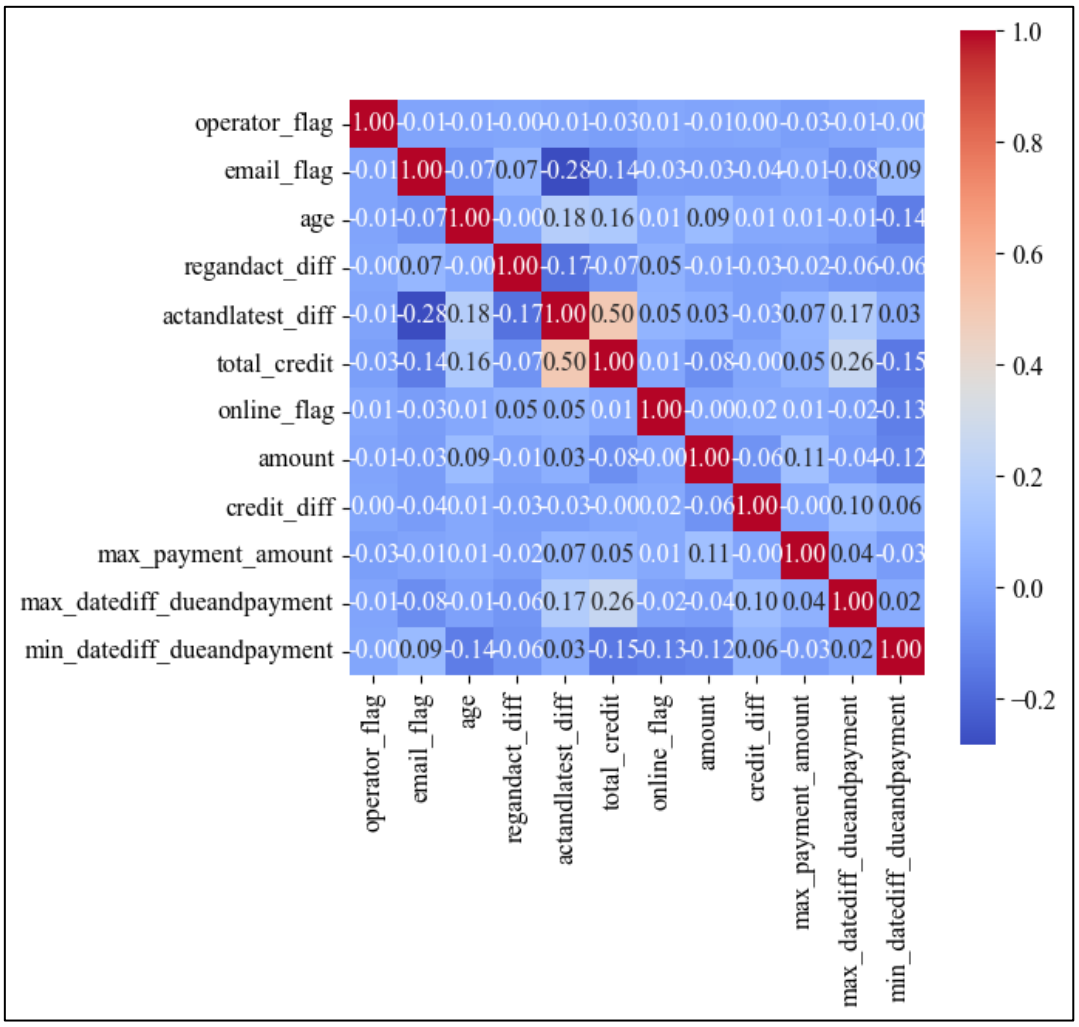

Figure 3.8: Correlation matrix of chosen features.

3.3 Model Application

In this section, model applications are mentioned. Practical applications of machine learning models developed to solve a specific problem or increase understanding on a specific subject constitute one of the main focuses of this study.

3.3.1 Random Forest application

In this section, the Random Forest application, the first algorithm used to solve the problem, is closely examined. While working with a wide feature set, it was preferred because it is both resistant to overfitting and has a high ability to tolerate imbalance.

3.3.1.1 Model tuning

The grid search method was used to decide the hyperparameters of the Random Forest model. Hyperparameter tuning often allows the model to make more accurate and reliable predictions. This increases the likelihood that the model will achieve better results on real-world data. The hyperparameter values used for this model are shown in Table 3.2 below.

Table 3.2: Random Forest model tuning hyperparameters.

Hyperparameter	Description	Used Values
n_estimators	Number of decision trees to be created	50, 100, 200
max_depth	Max. depth of the decision trees	"None", 10, 20
min_samples_split	Min. number of samples required to split a node	2, 5, 10
min_samples_leaf	Min. number of samples required to be at a leaf node	1, 2, 4
max_features	Max. number of features for each decision tree	"auto", "sqrt", "log2"

As a result of the grid search process, the hyperparameter combination that provides the best performance was determined as *{'max_depth': None, 'max_features': 'auto', 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 200}*. Determining these best parameters contributed to achieving the aim of the study by enabling the model to make more accurate and reliable predictions.

3.3.1.2 Feature importances

The performance impact of each feature contributing to the predictions of the Random Forest model in the Figure 3.9 below is shown. According to the graph below, the features that contributed the most to the model were *'credif_diff'* and *'min_datediff_dueandpayment'*. Similarly, it was understood that the least effective feature was *'email_flag'*.

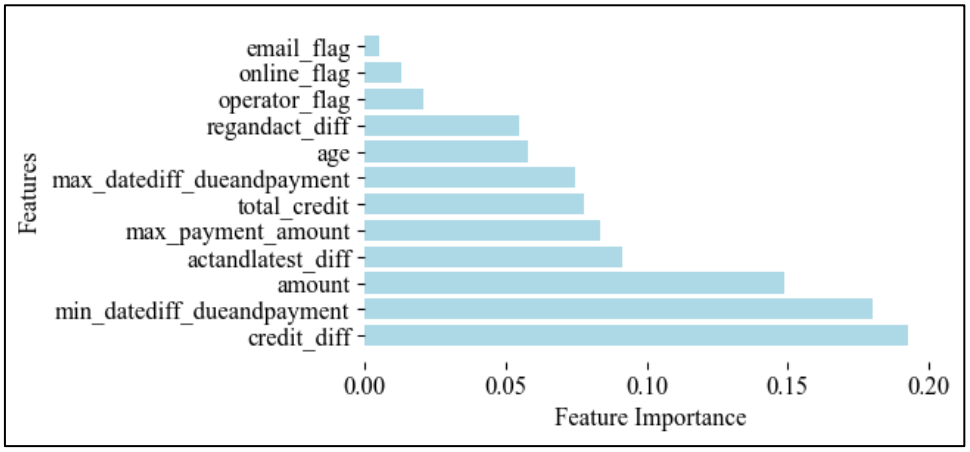

Figure 3.9: Random Forest feature importances.

3.3.1.3 K-fold cross validation

The 5-fold cross validation method was used to evaluate the performance of the random forest model. The scores obtained at each fold indicate how well the model performs on different data sets.

The mean score and standard deviation provide more information about the generalization ability of the model. Table 3.3 below shows the scores and statistics obtained in the 5-fold cross validation process of the model.

Table 3.3: 5-fold cross validation for Random Forest.

Fold Number	Score
1	0.8761592
2	0.87515358
3	0.87530715
4	0.87478735
5	0.87592146

With these results, the average score was calculated as 0.8754657495529757 and the standard deviation: 0.0005042558712073656.

3.3.1.4 ROC curve and weighted average AUC

In this section, ROC curves were drawn for multi-class classification problems using the OvR approach. The OvR approach is a common way to address multi-class classification problems. This approach creates a binary classification problem for each class separately from other classes and then plots a separate ROC curve for each class. The ROC curve shows the relationship between sensitivity (true positive rate) and specificity (true negative rate).

The ROC curve graph drawn for the Random Forest model is shown in the Figure 3.10 below. In addition, the weighted average AUC was calculated as 0.937820765813097.

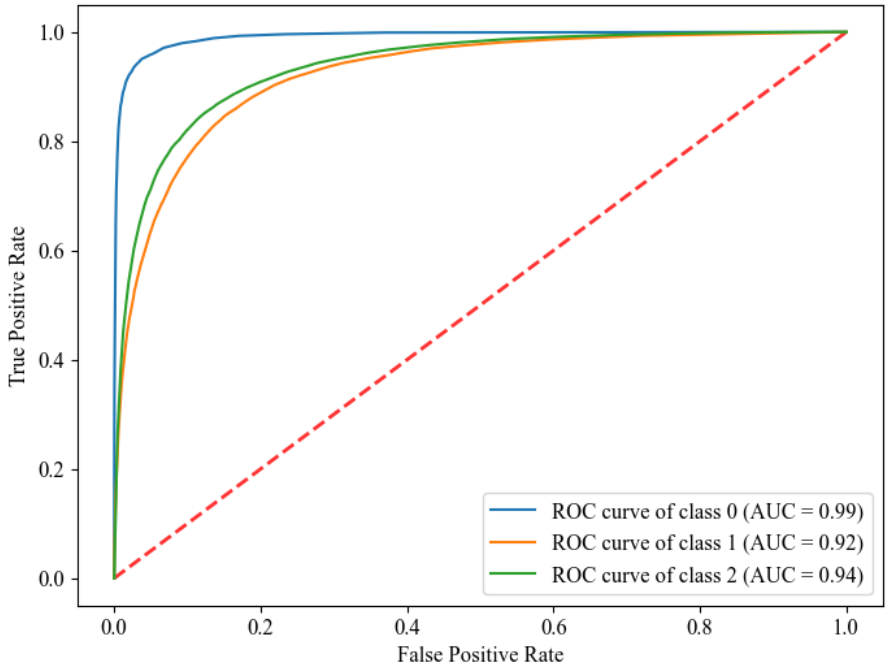

Figure 3.10: Random Forest ROC curve graph.

3.3.1.5 Confusion matrix

In this study, confusion matrix was used to evaluate the performance of the classification model. Confusion matrix is an effective tool used to show how accurate and inaccurate predictions the model makes for each class. Particularly in multi-class classification problems, it provides valuable information about the strengths and weaknesses of the model and provides important guidance for improving the model. In Figure 3.11 below, a 3x3 confusion matrix has been created for the 3-class segmentation model.

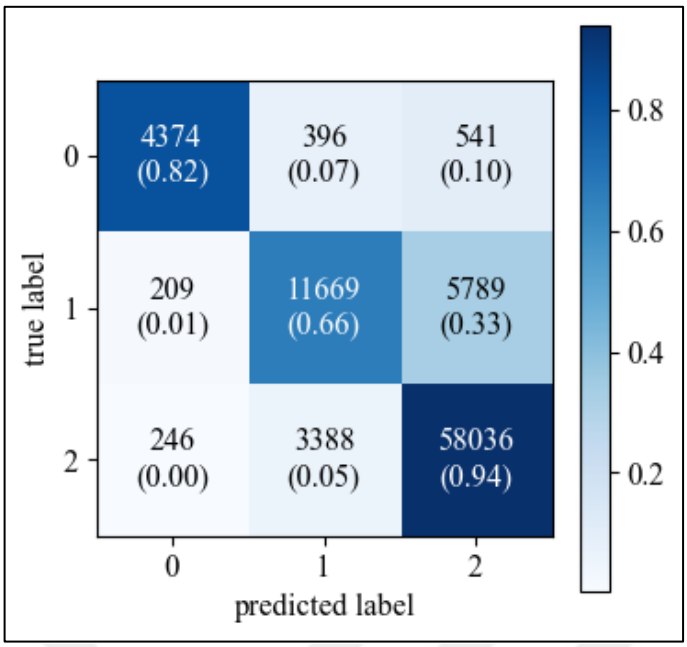

Figure 3.11: Confusion matrix of Random Forest

3.3.2 XgBoost application

In this section, a machine learning XgBoost algorithm, which is included in the gradient boosting framework and exhibits high performance especially on structured data sets, was evaluated. Providing high speed, scalability and precisely unified decisions, this algorithm has a wide range of uses for modeling complex relationships and making accurate predictions.

3.3.2.1 Model tuning

The parameters used in the grid search method for the XgBoost model are stated in Table 3.4 below.

Table 3.4: XgBoost model tuning hyperparameters.

Column A	Column B	Column C
n_estimators	Number of trees to be built	100, 200, 300
max_depth	Maximum depth of each tree	3, 6, 9
learning_rate	Learning rate	0.1, 0.01, 0.001
gamma	Minimum loss reduction required to make a split	0, 0.1, 0.2
subsample	Subsample ratio of the training instances	0.6, 0.8, 1.0
reg_alpha	L1 regularization term	0, 0.1, 0.5
reg_lambda	L2 regularization term	0, 0.1, 0.5

As a result of the grid search process, the hyperparameter combination that provides the best performance *{'gamma': 0, 'learning_rate': 0.1, 'max_depth': 9, 'n_estimators': 300, 'reg_alpha': 0.1, 'reg_lambda': 0, 'subsample ': 0.8}* was found.

3.3.2.2 Feature importances

The impact of each feature on the model performance that contribute to the predictions of the XgBoost model is shown in the Figure 3.12 below. According to the graph below, the features that contributed the most to the model were *'min_datediff_dueandpayment'* and *'actandlatest_diff'*. Similarly, it was understood that the least effective feature was *'operator_flag'*.

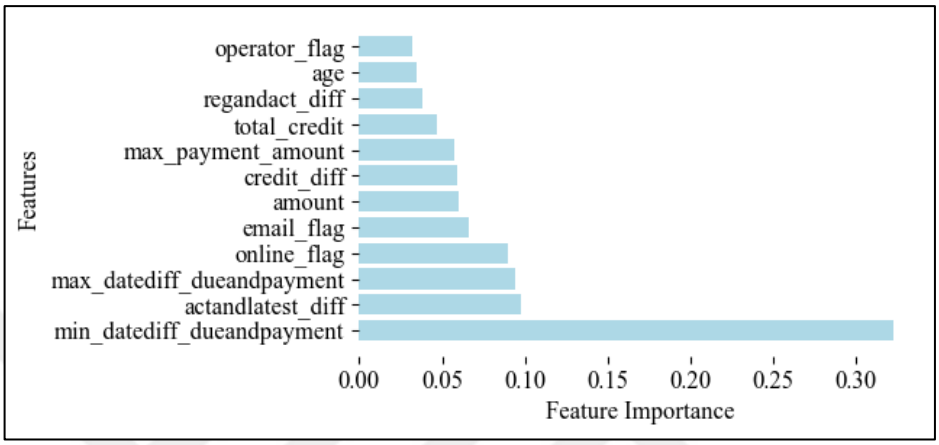

Figure 3.12: XgBoost feature importances.

3.3.2.3 K-fold cross validation

In order to better comment on the generalization ability and performance of the XgBoost model, 5-fold cross validation values were calculated and stated in Table 3.5 below.

Table 3.5: 5-fold Cross Validation for XgBoost.

Fold Number	Score
1	0.80570355
2	0.80537048
3	0.80361024
4	0.80394103
5	0.80522871

With these results, the average score was calculated as 0.8047708021369617 and the standard deviation: 0.0008336346205260454.

3.3.2.4 ROC curve and weighted average AUC

Similarly, ROC curves were drawn with the OvR approach to address multi-class classification problems. The weighted average AUC value obtained for the XgBoost

model was calculated as 0.8802563323910162. ROC curves are as in Figure 3.13 below.

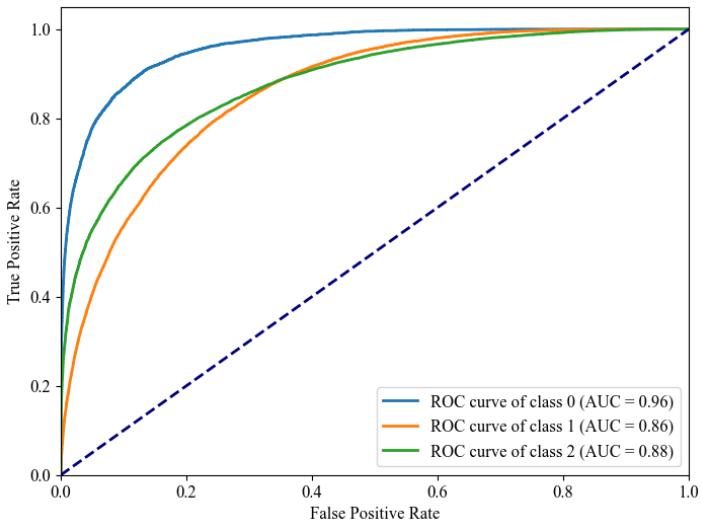

Figure 3.13: XgBoost ROC curve graph.

3.3.2.5 Confusion matrix

The 3x3 dimensional confusion matrix for the XgBoost model is shown in Figure 3.14 below.

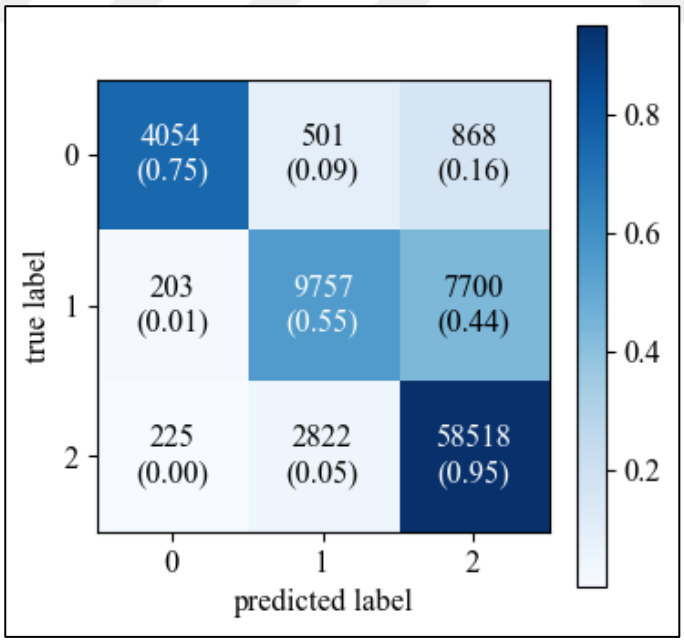

Figure 3.14: Confusion matrix of XgBoost.

3.3.3 Deep learning application

At this stage of the study, the details of the model developed using deep learning methods are mentioned. Deep learning is a powerful artificial intelligence technique used to analyze complex data structures and automatically extract their features.

3.3.3.1 Model tuning

Parallel to other modeling, the parameters used in the grid search method for the XgBoost model are stated in Table 3.6 below.

Table 3.6: Deep Learning model tuning hyperparameters.

Hyperparameter	Description	Used Values
optimizer	Optimization algorithm	'adam', 'sgd', 'rmsprop'
activation	Activation function	'relu', 'sigmoid', 'tanh'
dropout_rate	The fraction of the input units to drop to prevent overfitting during training.	0.1, 0.2, 0.3
batch_size	The number of training examples used in one iteration.	32, 64, 128
epochs	The number of complete passes through the training	50, 100
neurons	The number of neurons in a neural network layer.	128, 64, 32

As a result of the grid search process, the hyperparameter combination that provides the best performance was found to be *{'activation': 'relu', 'batch_size': 64, 'optimizer': 'adam'}*.

3.3.3.2 Model architecture

This model is designed for a multi-class classification problem and is trained with target variables encoded with one-hot encoding. Its architecture is a simple artificial neural network with three layers. The first layer has 128 neurons, and the input size corresponds to the number of features in the dataset. Next, a 20% dropout is applied to both layers to help prevent overfitting.

Dropout is the random shutdown of a certain proportion of neurons in any layer of the network, thus increasing the generalization ability of the model. The model has two denser layers containing 64 and 32 neurons consecutively. In the last layer, the model has 3 neurons and a softmax activation function for classification, because this is a multi-class classification problem.

The model is compiled with the 'adam' optimizer and trained using the *'categorical_crossentropy'* loss. Training is performed for 10 epochs and the performance of the model is evaluated using the accuracy value on the validation dataset.

3.3.3.3 K-fold cross validation

To better understand the overall performance and generalization ability of the deep learning model, the performance of the model on different data sets was evaluated using the 5-fold cross validation method. This method helps evaluate the generalization ability of the model by testing it on various pieces of data. The performance metrics achieved in each fold are mentioned in Table 3.7 below.

Table 3.7: 5-fold cross validation for Deep Learning.

Fold Number	Score
1	0.7692080736160278
2	0.7687207460403442
3	0.7679051160812378
4	0.7674769163131714
5	0.7701644897460938

With these results, the average score was calculated as 0.768695068359375 and the standard deviation: 0.0009519403298320819.

3.3.3.4 ROC curve and weighted average AUC

Similarly, ROC curves were drawn with the deep learning model using the OvR approach, as in other models. The weighted average AUC value obtained to evaluate the performance of the model was calculated as 0.8321108371303028. ROC curves show the sensitivity and specificity relationship between different classes and are used as an important indicator to evaluate the classification ability of the model.

Figure 3.15 below shows the ROC curves drawn for the Deep Learning model.

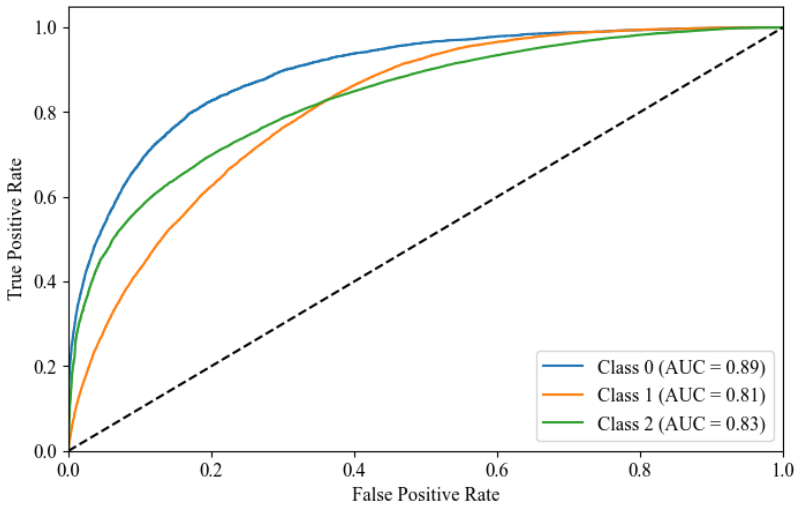

Figure 3.15: Deep Learning ROC curve graph.

3.3.3.5 Confusion matrix

The 3x3 dimensional confusion matrix for the Deep Learning model is shown in Figure 3.16 below.

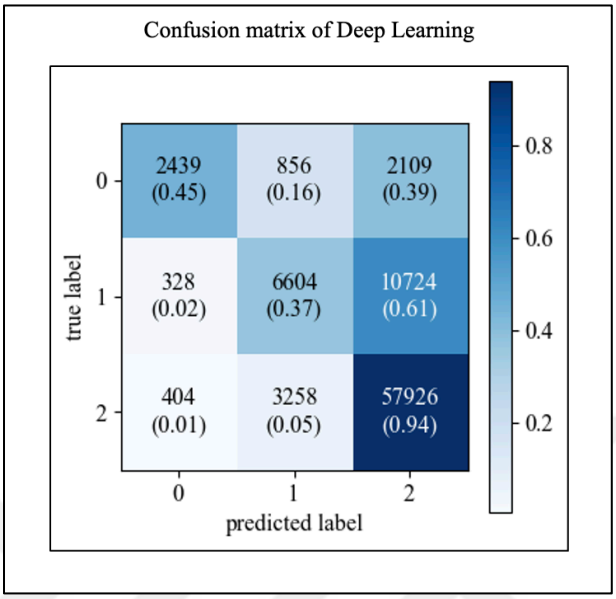

Figure 3.16: Confusion matrix of Deep Learning.

4. RESULTS

The conclusion part of the thesis is very important for the research. At this stage, the focus is on comparing the credit risk prediction performances between RF, XgBoost and Deep Learning models. Criteria such as accuracy, sensitivity, specificity and F1 score were used to evaluate the performance of each model, and the models were trained on the same data set, keeping the conditions the same. The classification reports for Random Forest, XgBoost and Deep Learning are given as Table 4.1, Table 4.2 and Table 4.3, respectively.

Table 4.1: Random Forest model classification report.

class	precision	recall	f1-score	support
0	0.91	0.82	0.86	5311
1	0.76	0.66	0.70	17667
2	0.90	0.94	0.92	61670
Accuracy			0.87	84648

Table 4.2: XgBoost model classification report.

class	precision	recall	f1-score	support
0	0.88	0.70	0.78	6588
1	0.73	0.49	0.59	16492
2	0.86	0.95	0.90	61568
Accuracy			0.84	84648

Table 4.3: Deep Learning model classification report.

class	precision	recall	f1-score	support
0	0.77	0.45	0.57	5404
1	0.62	0.37	0.47	17656
2	0.82	0.94	0.88	61588
Accuracy			0.79	84648

As a result of the comparative analysis, it was determined that the Random Forest model had the highest accuracy score. It appears that this machine learning model has the ability to better learn more complex relationships and patterns in credit risk

prediction. In addition, the information on how the model predicts the classes was examined in detail.

The confusion matrix approach was evaluated for this examination method. In the following Figure 4.1, Figure 4.2 and Figure 4.3 are shown for Random Forest, XgBoost and Deep Learning, respectively.

	Random Forest		
	UNPAID	LATE	PAID
UNPAID	0.82	0.07	0.10
LATE	0.01	0.66	0.33
PAID	0.01	0.05	0.94

Figure 4.1: Random Forest model confusion matrix.

	XgBoost		
	UNPAID	LATE	PAID
UNPAID	0.48	0.16	0.36
LATE	0.02	0.42	0.56
PAID	0.00	0.05	0.95

Figure 4.2: XgBoost model confusion matrix.

	Deep Learning		
	UNPAID	LATE	PAID
UNPAID	0.45	0.16	0.39
LATE	0.02	0.37	0.61
PAID	0.01	0.05	0.94

Figure 4.3: Deep Learning model confusion matrix.

When the figures are examined, it is seen that the Random Forest (RF) model has the highest ability to predict the credit group with UNPAID status. Similarly, for loans in the LATE category, the RF model made the best prediction, while the XgBoost and Deep Learning models show similar performance and remain at the same level in terms of prediction accuracy. It has been observed that all three models have high prediction rates for loans in PAID status. While making these comparisons, it is considered that the distribution of the target column was 53% PAID, 33% LATE and 14% UNPAID, that is, it did not exhibit a uniform distribution.

These findings show that the Random Forest model is superior to other models, especially in predicting loans in UNPAID and LATE statuses. In PAID status, all three models demonstrated high performance. However, the unbalanced distribution of the target column should be considered as an important factor when evaluating the performance of the models. In particular, the fact that the PAID category had the highest rate in the dataset may have made it easier for the models to achieve high prediction accuracy in this category. Evaluating the performance in the LATE and UNPAID categories requires a more complex process, because the number of examples in these categories is less than in the PAID category. This may naturally affect the performance of models in these categories.

In the predictive model I developed, there are three categories: LATE, PAID, and UNPAID. When evaluating the model's performance, it is crucial to consider the margin of error that could arise from incorrect predictions. Since the models perform similarly in the PAID category, it is more beneficial to focus on the more critical categories, LATE and UNPAID.

In this context, incorrectly predicting loans in the LATE and UNPAID categories as PAID could have adverse effects on cost calculations and cash flow. To better analyze this situation, confusion matrices were examined, and it was observed that the XgBoost and Deep Learning models have a higher likelihood of predicting loans in the LATE and UNPAID categories as PAID. This finding is a critical point to consider in model selection and performance evaluation.

During the evaluation of model performance, an additional cost metric was created to aid in the assessment. When developing the cost metric, it was assumed that a service fee of 10 TL is charged for each paid loan, loans classified as LATE and paid after 60

days incur a monthly late fee of 10%, and these loans are paid off after 60 days. The calculations based on the confusion matrices involved summing the service fees and late fees, then subtracting the amount of lost income from expected but unpaid loans.

The cost calculations were made as follows:

Service Fee Cost: TP x 10 TL

Late Fee Cost: FN x 12.1 TL

Lost Income from Unpaid: FP (LATE, UNPAID → PAID) x 10 TL

(4.1)

For example, the calculations for the **Random Forest** model:

Service Fee Cost: 58,589 x 10 TL = 585,890 TL

Late Fee Cost: 2,914 x 12.1 TL = 35,259.4 TL and 6,003 x 12.1 TL = 72,636.3 TL

Lost Income from Unpaid: 921 x 10 TL = 9,210 TL, 6,003 x 10 TL = 60,030 TL, and 331 x 10 TL = 3,310 TL

The total cost: 585,890 + 35,259.4 + 72,636.3 − 72,550 = 571,235.7 TL

(4.2)

Similarly, for the **XgBoost** model:

Service Fee Cost: 58,695 x 10 TL = 586,950 TL

Late Fee Cost: 2,531 x 12.1 TL = 30,625.1 TL and 8,070 x 12.1 TL = 97,547 TL

Lost Income from Unpaid: 1,432 x 10 TL = 14,320 TL, 8,070 x 10 TL = 80,700 TL, and 342 x 10 TL = 3,420 TL.

The total cost: 586,950 + 30,625.1 + 97,547 − 98,420 = 616,702.1 TL

(4.3)

For the **Deep Learning** model:

Service Fee Cost: 58,897 x 10 TL = 588,970 TL

Late Fee Cost: 2,137 x 12.1 TL = 25,859.7 TL and 12,089 x 12.1 TL = 146,276.9 TL

Lost Income from Unpaid: 3,841 x 10 TL = 38,410 TL, 12,089 x 10 TL = 120,890 TL, and 644 x 10 TL = 6,440 TL.

The total cost: 588,970 + 25,859.7 + 146,276.9 − 165,740 = 595,366.6 TL

(4.4)

In conclusion of these calculations, the cost for the Random Forest model was found to be 571,235.7 TL, the cost for the XgBoost model was 616,702.1 TL, and the cost for the Deep Learning model was 595,366.6 TL. These results indicate that when the cost metric is included in the model selection process, the Random Forest model is the least costly, while the XgBoost model is the most costly.

As a result, the RF model demonstrated superior prediction performance, especially in the UNPAID and LATE categories, despite the unbalanced data distribution, while all models reached high accuracy rates in the PAID category. These data reveal that the Random Forest model is a preferable model in credit status estimation and provides more reliable results, especially for critical situations.

5. CONCLUSION AND FUTURE RESEARCH

Measuring and predicting the probability of default is of great importance for financial institutions. Credit risk modeling is used to identify these possibilities and related risks. Credit risk modeling helps financial institutions take more conscious and accurate steps in their lending decisions. Various studies and recent advances in machine learning further improve this process and allow institutions to manage credit risks more effectively.

5.1 Conclusion

New approaches and methods are constantly being developed in the field of credit risk modelling. These innovations offer significant advantages for financial institutions, both theoretically and practically. Focusing on this issue in academic studies provides valuable contributions to practical results in the financial sector. In particular, the development of new modeling techniques and algorithms increases the accuracy and reliability of credit risk assessments, allowing financial institutions to make more effective decisions in risk management processes. The integration of academic research with practical applications supports financial stability and sustainability by providing innovative solutions to current challenges in the sector. In this context, academic studies on credit risk modeling contribute to the development and innovation of the sector by providing concrete benefits for both theoretical knowledge and financial applications.

When evaluating model performances and before choosing a model, it should not be forgotten that models have their own behaviors and areas of use. Each model has certain advantages and limitations, so the needs and strategies of financial institutions should be considered when deciding which model to use. In addition, the risks that financial institutions want to take and their tolerance level towards these risks also play an important role in the model selection process. The strategies of institutions and organizations should be taken into consideration and action should be taken in line with these strategies in decision-making processes. In this way, more effective and sustainable results can be achieved in credit risk management.

5.2 Future Research

This section will discuss directions for future research based on the findings and limitations of the current study. These studies aim to overcome the limitations of the current study and provide a broader perspective on the subject.

5.2.1 Limitations of current study, new research questions, and hypotheses

The current study has some limitations as follows, however, new research questions and hypotheses have emerged in light of the findings of this study. Future research offers significant opportunities to overcome the limitations of the current study and provide a broader perspective. The suggestions mentioned in this section can be quite instructive in directing this process.

- **Time Interval:** In this study, only usage details within a certain time period were examined. Research conducted in different time periods can evaluate the effects of variables in a broader perspective. Considering that default risk estimates vary on a monthly, quarterly and seasonal basis, it may be useful to examine these time periods in future research.
- **Limited Time:** Some detailed analyzes could not be carried out due to time constraints during the thesis writing process. These limitations also affected the model development processes. Although it was observed that the Random Forest model came to the fore in the current study, more studies can be done on Deep Learning models.

Deep learning models can provide more effective and high-performance results with larger data sets and longer running times. After detailed examination of the model architecture, it is possible to develop models with much higher performance by using different modeling structures. In particular, studies can be conducted on how deep learning techniques such as CNN and LSTM networks can be used in default risk predictions.

Additionally, hyperparameter optimization and model fine-tuning can improve the model's accuracy and generalization ability. Therefore, studies conducted over a longer period of time will allow for more in-depth analyzes and the application of more sophisticated modeling techniques.

- **Additional Features:** Studies can be repeated with new features to be added to the data set. For example, different results can be obtained by adding details such as location data where the product is used and online device category.

5.2.2 Application areas and technological developments

Discussions can be held on practical applications of the study and how these applications can be improved. Subsequently, technological developments in this field may offer various opportunities for future studies.

- **Industrial Applications:** How research findings can be used in the financial industry and in which areas further research is needed can be analyzed and specified in detail.
- **Policy Development:** The importance of research results for policy makers and regulators can be reconsidered for further research in this context.
- **New Data Collection Techniques:** Thanks to developing artificial intelligence technologies, more accurate and comprehensive data collection methods can be included in the research.
- **Collaborations:** More comprehensive and in-depth research can be conducted by collaborating with experts from different disciplines, especially in sectors such as banking and marketing.
- **Cross-Disciplinary Research:** The ways in which the topic intersects with other scientific disciplines and how these intersections can be examined can be studied academically as another research subject

REFERENCES

[1] **Lending Rate - Countries - List | G20.** (n.d.). Tradingeconomics.com. Retrieved May 16, 2024, from https://tradingeconomics.com/country-list/lending-rate?continent=g20

[2] **BloombergHt.** (2023, August 1). BDDK'dan sıkılaşma adımları. BloombergHT. Retrieved from: https://www.bloomberght.com/kredi-kartina-taksit-sinirlamasi-2335767

[3] **Bazarbash, M.** (2019). FinTech in Financial Inclusion: Machine Learning Applications in Assessing Credit Risk. IMF Working Papers, 19(109), 1. Retrieved from: https://doi.org/10.5089/9781498314428.001

[4] **Baesens, B., & Smedts, K.** (2023). Boosting credit risk models. The British Accounting Review, 101241. Retrieved from: https://doi.org/10.1016/j.bar.2023.101241

[5] **Waegeman, W., Verwaeren, J., Slabbinck, B., & De Baets, B.** (2011). Supervised learning algorithms for multi-class classification problems with partial class memberships. Fuzzy Sets and Systems, 184(1), 106–125. Retrieved from: https://doi.org/10.1016/j.fss.2010.11.012

[6] **Arner, D. W., Barberis, J. N., & Buckley, R. P.** (2015). The Evolution of Fintech: a New Post-Crisis Paradigm? SSRN Electronic Journal, 47(4).

[7] **The Housing Downturn in the United States 2009 First Quarter Update.** (2009). Retrieved from: http://demographia.com/db-ushsg2009q1.pdf

[8] **Chen, Y., kumara, E. K., & Sivakumar, V.** (2021). Investigation of finance industry on risk awareness model and digital economic growth. Annals of Operations Research, 1(1), 1–22. Retrieved from: https://link.springer.com/article/10.1007/s10479-021-04287-7

[9] **Zheng, X., Zhu, M., Li, Q., Chen, C., & Tan, Y.** (2019). FinBrain: when finance meets AI 2.0. Frontiers of Information Technology & Electronic Engineering, 20(7), 914–924. Retrieved from: https://doi.org/10.1631/fitee.1700822

[10] **Kagan, J.** (2023, December 20). Financial Technology (Fintech): Its Uses and Impact on Our Lives. Investopedia. Retrieved from: https://www.investopedia.com/terms/f/fintech.asp

[11] **Haji, A. R.** (2024, February 4). Global Insights - KPMG Global. KPMG. Retrieved from: https://kpmg.com/xx/en/home/insights/2024/02/pulse-of-fintech-h2-2023-global-insights.html

[12] **Kagan, J.** (2023, December 20). Financial Technology (Fintech): Its Uses and Impact on Our Lives. Investopedia. Retrieved from: https://www.investopedia.com/terms/f/fintech.asp

[13] **Akhtar, F., Mobarak, M., Ali Akbar Shaikh, & Adel Fahad Alrasheedi.** (2023). Interval valued inventory model for deterioration, carbon emissions and selling price dependent demand considering buy now and pay later facility. Ain Shams Engineering Journal, 102563–102563. Retrieved from: https://doi.org/10.1016/j.asej.2023.102563

[14] **Consumer Financial Protection Bureau Opens Inquiry Into "Buy Now, Pay Later" Credit.** (n.d.). Consumer Financial Protection Bureau. Retrieved from: https://www.consumerfinance.gov/about-us/newsroom/consumer-financial-protection-bureau-opens-inquiry-into-buy-now-pay-later-credit/

[15] **What is a Buy Now, Pay Later (BNPL) loan? | Consumer Financial Protection Bureau.** (2021, December 2). Consumer Financial Protection Bureau. Retrieved from: https://www.consumerfinance.gov/ask-cfpb/what-is-a-buy-now-pay-later-bnpl-loan-en-2119/

[16] **Desk, F. N.** (2020, September 22). Buy Now Pay Later Digital Spend, Led by Klarna, PayPal & Afterpay, to Double by 2025: Reaching $680 Billion - Kaleido Intelligence. GlobalFinTechSeries. Retrieved from: https://globalfintechseries.com/fintech/buy-now-pay-later-digital-spend-led-by-klarna-paypal-afterpay-to-double-by-2025-reaching-680-billion-kaleido-intelligence/

[17] **Consumer Financial Protection Bureau Opens Inquiry Into "Buy Now, Pay Later" Credit.** (n.d.). Consumer Financial Protection Bureau. Retrieved from: https://www.consumerfinance.gov/about-us/newsroom/consumer-financial-protection-bureau-opens-inquiry-into-buy-now-pay-later-credit/

[18] **Lux, M., School, H., & Epps, B.** (2022). M-RCBG Associate Working Paper Series | No. 182 Grow Now, Regulate Later? Regulation urgently needed to support transparency and sustainable growth for Buy-Now, Pay-Later. Retrieved from: https://www.hks.harvard.edu/sites/default/files/centers/mrcbg/files/182_AWP_final.pdf

[19] **Shevlin, R.** (n.d.). Buy Now, Pay Later: The "New" Payments Trend Generating $100 Billion In Sales. Forbes. Retrieved May 7, 2024, from https://www.forbes.com/sites/ronshevlin/2021/09/07/buy-now-pay-later-the-new-payments-trend-generating-100-billion-in-sales/?sh=42c2b2e02ffe

[20] **TransUnion Study Examines the Risk Profile of BNPL Applicants and the Financial Inclusion Opportunities that Exist for Both Consumers and Lenders.** (n.d.). TransUnion Study Examines the Risk Profile of BNPL Applicants and the Financial Inclusion Opportunities That Exist for Both Consumers and Lenders. Retrieved from: https://newsroom.transunion.com/transunion-study-examines-the-risk-profile-of-bnpl-applicants--and-the-financial-inclusion-opportunities-that-exist-for-both-consumers-and-lenders/

[21] **Almost 75% of BNPL users in the US are Gen Z or millennials.** (n.d.). Insider Intelligence. Retrieved from: https://www.emarketer.com/content/almost-75-of-bnpl-users-us-gen-z-millennials

[22] **Howarth, J.** (2022, February 2). 12 Amazing BNPL Stats (2022). Exploding Topics. Retrieved from: https://explodingtopics.com/blog/bnpl-stats

[23] **Statista - The Statistics Portal.** (n.d.). Statista. Retrieved May 7, 2024, from https://www.statista.com/insights/consumer/brand-profiles/2/4/buy-now-pay-later/united-states/#contentBox1

[24] **King, J.** (2024, March 6). Guide to buy now, pay later: Industry trends, regulation, and top companies explained. EMARKETER. Retrieved from: https://www.emarketer.com/insights/buy-now-pay-later-industry-challenges/

[25] **Musa, H., Klieštik, T., Frajtová-Michalíková, K., & Debnárova, L.** (2015). The Tradition Approach to Credit Risk and its Estimation for Selected Banks in Slovakia. Procedia Economics and Finance, 24, 451–456. Retrieved from: https://doi.org/10.1016/s2212-5671(15)00702-9

[26] **The Fed - Supervisory Policy and Guidance Topics - Credit Risk Management.** (2010). Federalreserve.gov. Retrieved from: https://www.federalreserve.gov/supervisionreg/topics/credit_risk.htm

[27] **Roeder, J.** (2021). Alternative Data for Credit Risk Management: An Analysis of the Current State of Research. AIS Electronic Library (AISeL) (Association for Information Systems). Retrieved from: https://doi.org/10.18690/978-961-286-485-9.13

[28] **Prof. Jia Ruan, & Jiang, R.** (2024). Does digital inclusive finance affect the credit risk of commercial banks? Finance Research Letters, 62, 105153–105153. Retrieved from: https://doi.org/10.1016/j.frl.2024.105153

[29] **BIS.** (2020). Credit risk management Principles for the Management of Credit Risk. Retrieved from: https://www.bis.org/publ/bcbsc125.pdf

[30] **Gavlakova P., Kliestik T.,** (2014). Credit Risk Models and Valuation. 4th International Conference on Applied Social Science, Information Engineering Research Institute, Singapore, pp. 139-143.

[31] **Buc D., Kliestik T.,** (2013). Aspects of statistics in terms of financial modelling and risk. Proceeding of the 7th International Days of Statistics and Economics, pp. 215- 224, Prague.

[32] **Anna, S., Boris, K., & Ivana, W.** (2015). Impact of Credit Risk Management. Procedia Economics and Finance, 26, 325–331. Retrieved from: https://doi.org/10.1016/s2212-5671(15)00860-6

[33] **Luana Cristina Rogojan, Andreea Elena Croicu, & Laura Andreea Iancu.** (2023). Modern Approaches in Credit Risk Modeling: A Literature Review. Proceedings of the ... International Conference on Business Excellence, 17(1), 1617–1627. Retrieved from: https://doi.org/10.2478/picbe-2023-0145

[34] **GiniMachine.** (2020, November 30). How to Improve Credit Risk Management: AI for Banking. GiniMachine. Retrieved from: https://ginimachine.com/blog/ai-credit-risk-management/#how_does_ai_assist_credit_risk_in_banks

[35] **Yusof, S. A. B. M., & Roslan, F. A. B. M.** (2023). The Impact of Generative AI in Enhancing Credit Risk Modeling and Decision-Making in Banking Institutions. Emerging Trends in Machine Intelligence and Big Data,

15(10), 40–49. Retrieved from https://orientreview.com/index.php/etmibd-journal/article/view/30

[36] **Wang, H., Mao, K., Wu, W., & Luo, H.** (2023). Fintech inputs, non-performing loans risk reduction and bank performance improvement. International Review of Financial Analysis, 90, 102849. Retrieved from: https://doi.org/10.1016/j.irfa.2023.102849

[37] **Yıldırım, M., Okay, F. Y., & Özdemir, S.** (2021). Big data analytics for default prediction using graph theory. Expert Systems with Applications, 176, 114840. Retrieved from: https://doi.org/10.1016/j.eswa.2021.114840

[38] **Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J.** (2020). Explainable Machine Learning in Credit Risk Management. Computational Economics, 57.

[39] **Theissler, A., Thomas, M., Burch, M., & Gerschner, F.** (2022). ConfusionVis: Comparative evaluation and selection of multi-class classifiers based on confusion matrices. Knowledge-Based Systems, 247, 108651. Retrieved from: https://doi.org/10.1016/j.knosys.2022.108651

[40] **Zhang, T.** (n.d.). Mathematical Analysis of Machine Learning Algorithms. Retrieved from: https://tongzhang-ml.org/lt-book/lt-book.pdf

[41] **Vayssières, M. P., Plant, R. E., & Allen-Diaz, B. H.** (2000). Classification trees: An alternative non-parametric approach for predicting species distributions. Journal of Vegetation Science, 11(5), 679–694. Retrieved from: https://doi.org/10.2307/3236575

[42] **Biau, G., & Fr, G.** (2012). Analysis of a Random Forests Model. Journal of Machine Learning Research, 13, 1063–1095. Retrieved from: https://www.jmlr.org/papers/volume13/biau12a/biau12a.pdf

[43] **Goh, C.-H., Mahbuba Ferdowsi, Gan, M., Kwan, B.-H., Wei Yin Lim, Yee Kai Tee, Roshaslina Rosli, & Maw Pin Tan.** (2024). Assessing the Efficacy of Machine Learning Algorithms for Syncope Classification: A Systematic review. MethodsX, 12, 102508–102508. Retrieved from: https://doi.org/10.1016/j.mex.2023.102508

[44] **Random Decision Forest in Reinforcement learning.** (2018, July 6). Retrieved from: https://www.mql5.com/en/articles/3856

[45] **IBM.** (2023). What is Random Forest? | IBM. https://www.ibm.com/topics/random-forest#:~:text=Random%20forest%20is%20a%20commonly

[46] **Yan, A.** (2020). A Survey on Tree-Based Machine Learning Methods: The Path to XGBoost and Generalized Random Forests. Retrieved from: https://andongyan.com/uploads/survey2020.pdf

[47] **Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G.** (2019). A Comparative Analysis of XGBoost. Retrieved from: https://arxiv.org/pdf/1911.01914/1000

[48] **Chen, M., Liu, Q., Chen, S., Liu, Y., Zhang, C.-H., & Liu, R.** (2019). XGBoost-Based Algorithm Interpretation and Application on Post-Fault Transient Stability Status Prediction of Power System. IEEE Access, 7,

13149–13158. Retrieved from: https://doi.org/10.1109/ACCESS.2019.2893448

[49] **Mishra, R. K., Reddy, G. Y. S., & Pathak, H.** (2021). The Understanding of Deep Learning: A Comprehensive Review. Mathematical Problems in Engineering, 2021, 1–15. Retrieved from: https://doi.org/10.1155/2021/5548884

[50] **Derin Öğrenme Nedir? - Derin Öğrenmeye Ayrıntılı Bakış - AWS.** (n.d.). Amazon Web Services, Inc. Retrieved May 11, 2024, from https://aws.amazon.com/tr/what-is/deep-learning/#:~:text=Deep%20learning%20is%20a%20method

[51] **Hanif, I.** (2020). Implementing Extreme Gradient Boosting (XGBoost) Classifier to Improve Customer Churn Prediction. Proceedings of the Proceedings of the 1st International Conference on Statistics and Analytics, ICSA 2019, 2-3 August 2019, Bogor, Indonesia. Retrieved from: https://doi.org/10.4108/eai.2-8-2019.2290338

[52] **Correlation Matrix - an overview** | ScienceDirect Topics. (n.d.). Www.sciencedirect.com. Retrieved from: https://www.sciencedirect.com/topics/mathematics/correlation-matrix

[53] **Wagavkar, S.** (2023, March 17). Introduction to The Correlation Matrix | Built In. Builtin.com. https://builtin.com/data-science/correlation-matrix

[54] **Japkowicz, N., & Shah, M.** (2011). Evaluating Learning Algorithms: A Classification Perspective. In Google Books. Cambridge University Press. https://books.google.com.tr/books?hl=tr&lr=&id=VoWIIOKVzR4C&oi=fnd&pg=PR7&ots=5z34TFOBOM&sig=IgVYMWjdJZyL9v6qxuRcQekw-9E&redir_esc=y#v=onepage&q&f=false

[55] **Roitberg, B. D. (n.d.). Reference Module in Life Sciences.** (2016) In Google Books. Retrieved from https://books.google.com.tr/books/about/Reference_Module_in_Life_Sciences.html?id=OmYnnQAACAAJ&redir_esc=y

[56] **Arlot, S., & Lerasle, M.** (2015, October 11). Choice of V for V-Fold Cross-Validation in Least-Squares Density Estimation. ArXiv.org. Retrieved from: https://doi.org/10.48550/arXiv.1210.5830

[57] **Stratified K Fold Cross Validation.** (2020, August 6). GeeksforGeeks. Retrieved from: https://www.geeksforgeeks.org/stratified-k-fold-cross-validation/

[58] **Markoulidakis, I., Rallis, I., Georgoulas, I., Kopsiaftis, G., Doulamis, A., & Doulamis, N.** (2021). Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem. Technologies, 9(4), 81. Retrieved from: https://doi.org/10.3390/technologies9040081

[59] **Srivastava, T.** (2019, August 6). 12 Important Model Evaluation Metrics for Machine Learning Everyone Should Know (Updated 2023). Analytics Vidhya. Retrieved from:

https://www.analyticsvidhya.com/blog/2016/02/7-important-model-evaluation-error-metrics

[60] **Trevisan, V.** (2022, March 24). Multiclass classification evaluation with ROC Curves and ROC AUC. Medium. Retrieved from: https://towardsdatascience.com/multiclass-classification-evaluation-with-roc-curves-and-roc-auc-294fd4617e3a

CURRICULUM VITAE

Name Surname **: Ömür ÖZDOĞAN**

EDUCATION **:**

- **B.Sc.** **:** 2020, Middle East Technical University, Faculty of Arts and Sciences, Mathematics
- **M.Sc.** **:** 2024, Istanbul Technical University, Graduate School, Big Data and Business Analytics

PROFESSIONAL EXPERIENCE AND REWARDS:

- 2023-Present, Colendi - Senior Data Analyst
- 2022-2023, QNB Finansbank - Data Analyst
- 2021-2022, QNB Finansbank - Junior Data Analyst
- 2020-2022, Development Investment Bank of Turkey - Assistant Specialist
- 2019-2020, Middle East Technical University - Student Assistant
- 2018, Central Bank of the Republic of Türkiye - Intern