NONCONTRACTUAL CHURN ANALYSIS

FOR A PRIVATE BANK CUSTOMERS

BAHA ULUG

DECEMBER 2022

NONCONTRACTUAL CHURN ANALYSIS

FOR A PRIVATE BANK CUSTOMERS

A THESIS SUBMITTED TO THE

GRADUATE SCHOOL

OF

BAHÇEŞEHİR UNIVERSITY

BAHA ULUG

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR

THE DEGREE OF MASTER OF ARTIFICIAL INTELLIGENCE

IN THE DEPARTMENT OF GRADUATE SCHOOL OF ENGINEERING

DECEMBER 2022

T.C.

BAHCESEHIR UNIVERSITY

GRADUATE SCHOOL

MASTER THESIS APPROVAL FORM

Program Name:	Artificial Intelligence Graduate Program
Student's Name and Surname:	Baha Ulug
Name Of The Thesis:	Noncontractual Churn Analysis For Private Bank Customers
Thesis Defense Date:	18.01.2023

This thesis has been approved by the Graduate School which has fulfilled the necessary conditions as Master thesis.

Prof. Dr. Ahmet ÖNCÜ

Institute Director

This thesis was read by us, quality and content as a Master's thesis has been seen and accepted as sufficient.

	Title/Name	**Signature**
Thesis Advisor's	Prof. Dr. Süreyya Akyüz	
Member's	Doç. Dr. Tevfik Aytekin	
Member's	Doç. Dr. Burcu Tunga	

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Baha Ulug

Signature:

ABSTRACT

NONCONTRACTUAL CHURN ANALYSIS

FOR A PRIVATE BANK CUSTOMERS

Baha, Ulug

Master's Program in Artificial Intelligence

Supervisor: Prof. Dr. Sureyya AKYUZ

December 2022, 116 pages

The aim of this study is to examine the customer losses experienced by an interest-free bank and create an analysis process that gives the best churn approach. The study includes data for a bank in the interest-free finance sector for 2020 and 2021. 6,844 corporate segment customers of a private participation bank in Turkey who exchanged foreign currency and precious metals via a mobile banking application were investigated. For customer churn analysis, first the RFM model was developed, then the CLV value was calculated. A threshold was determined with the statistical method in the transactional data, which was unlabeled, and the customers below the threshold were labeled as "lost customers." After all, customers were labeled according to this rule, and supervised methods were used to predict the most probable churners.

Key Words: Customer Churn, RFM, CLV, CRM, Bank

ÖZ

SÖZLEŞMESİZ ÖZEL BANKA MÜŞTERİLERİNE YÖNELİK

MÜŞTERİ KAYBI ANALİZİ

Baha, Ulug

Yapay Zeka Yüksek Lisans Programı

Tez Danışmanı: Prof. Dr. Sureyya AKYUZ

Aralık 2022, 116 sayfa

Bu çalışmanın amacı, özel bir bankanın yaşadığı müşteri kayıplarını incelemek ve en iyi kayıp yaklaşımını veren bir analiz süreci oluşturmaktır. Çalışma faizsiz finans sektöründe faaliyet gösteren bir bankanın 2020 ve 2021 yıllarına ilişkin verilerini içermektedir. Türkiye'de faaliyet gösteren bir özel katılım bankasının mobil bankacılık uygulaması üzerinden döviz ve kıymetli maden takası yapan 6.844 kurumsal segment müşterisi incelenmiştir. Müşteri kayıp analizi için öncelikle RFM modeli geliştirilmiş, ardından müşterilerin CLV değeri hesaplanmıştır. Etiketlenmemiş işlemsel verilerde istatistiksel yöntemle bir eşik belirlenmiş ve eşiğin altındaki müşteriler "kayıp müşteri" olarak etiketlenmiştir. Sonrasında, en olası müşteri kayıplarını tahmin etmek için denetimli öğrenme yöntemleri kullanılmıştır.

Anahtar Kelimeler: Müşteri Kaybı, RFM, CLV, CRM, Bankacılık

To My Wife

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my thesis supervisor Prof. Dr. Sureyya Akyuz for her invaluable comments, support, advice, and guidance through all my studies.

I am extremely thankful to Yusuf Karaoglu, who is my colleague, for his endless support, understanding and advice during this time.

I also thank my friends for providing support and friendship that I needed. Finally, my very special thanks go to my family, who has always stood by my side, really deserves my great appreciation. This thesis is heartily dedicated to my family.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ALT	Gold
AUC	Area Under Curve
ATM	Automated Teller Machine
BG/NBD	The Beta-Geometric Negative Binomial Distribution
CART	Classification and Regression Trees
CRM	Customer Relationship Management
CLV	Customer Lifetime Value
EDA	Exploratory Data Analysis
FN	False Negative
FP	False Positive
GMS	Silver
KNN	K-Nearest Neighbors
KPI	Key Performance Indicator
LOFO	Leave One Feature Out
ML	Machine Learning
OLAP	On-line Analytical Processes
PLT	Platinum
RF	Random Forest
RFM	Recency-Frequency-Monetary
ROC	Receiver Operating Characteristics
ROI	Return of Investment
SOCP	Second Order Cone Programming
SQL	Structured Query Language
SVM	Support Vector Machines
TL	Turkish Lira
TP	True Positive
TN	True Negative
USD	United States Dollar

Chapter 1

Introduction

1.1 Theoretical Framework

The interest-free finance system is an alternative approach that aims to eliminate the confusion caused by the new order that emerged with the modern financial system. The system, which prohibits the determination of fixed interest and finds it appropriate to take a certain share of the income earned, is based on the principle of profit sharing. However, the fact that they cannot benefit from the flexibility provided by modern banking facilities has necessitated the establishment of interest-free financial institutions to maximize customer satisfaction. Because when it is not possible to compete financially, customer satisfaction will come to the fore. The easiest way to ensure customer satisfaction is to understand their needs and quickly meet them. One of the most important stages for analyzing these needs is churning analysis because it is an analysis aimed at preventing customers from leaving the institution. Statistical learning approaches come to the fore when these analyses are implemented quickly, effectively, and included in business processes. In this context, as well as primitive approaches such as RFM and probabilistic survival analyzes the concept of machine learning has become quite popular. Machine learning is a sub-branch of artificial intelligence and gives systems and computers the opportunity to learn from and improve on past experiences. Machine learning focuses on developing computer programs that can access data on their own and learn from it without human assistance. With machine learning, the outcome of a future event can be predicted, or patterns that are difficult to detect by humans can be revealed. Companies may benefit from statistical techniques such as RFM, survival models or machine learning to gain meaningful insights; they find relationships between events and predict future outcomes.

1.2 Statement of the Problem

The participation banking system is based on the basic principles of interest-free earnings and equity capital. This system has tried to survive with institutions carrying out similar functions in the past. Atici states that, in the Turkish monetary sector, participation banking is a unique and promising characteristic (2018). Unfortunately, these institutions, which could not keep up with the modern age, could not hold on. Today, there are many independent and dependent institutions working in accordance with the participation banking system, both in Islamic countries and in many other countries of the world. They are still moving towards becoming part of the world financial system and creating a seriously competitive environment.

In today's aggressive environment, acquiring new customers is more costly in many ways than retaining existing ones. Companies in the service sector focus on preventing customer loss. Churn analysis is a general concept for identifying the customers most likely to switch to another company. Churn analysis has led to the development of various techniques and models for loss estimation. Such models emphasize accuracy, flexibility, and lower implementation costs because each lost customer results in increased expenditures for the business. However, customer churn analysis, which is prepared for a company to overcome its concerns about losing customers and increasing costs, is an extremely difficult and comprehensive task. Various approaches to predictive modeling have been developed for churn, and these approaches can be performed in three steps by establishing a specific behavioral model for customers, or, in other words, customer profiling, segmentation, customer value calculation, CLV, and churn prediction, namely a classification model, through data mining processes. Customer segmentation may be planned efficiently by setting up the RFM model. With the help of numerous libraries established in programming languages such as Python or R in this area, CLV calculations can be performed using various techniques. By constructing a classification model, the third stage, churn prediction, may be carried out using standard and popular ML models.

1.3 Purpose of the Study

The aim of this thesis is to show that statistical techniques and machine learning approach can be applied effectively and successfully to customer churn prevention products that CRM departments need most in the banking industry. If the explainability of the prepared model is high, the factors that cause customers to separate from the bank will emerge. Customer loss will decrease, and customer satisfaction will reach the desired point, if the steps to improve these factors are successful.

This study focuses on customer churn forecasting and churns analysis in the private banking industry. First, relevant data was collected for churn prediction. It is of great importance to describe the data as accurately as possible because too much detail or incomplete data adversely affects the accuracy of the estimate. Decision tree-based models are used as appropriate estimation techniques according to the data structure. In the example of churn analysis, one of the key points of competition between banks, 100 thousand transactional order data, and 643,000-line customer data will be processed, in addition to 130 thousand foreign exchange buying and selling transactions, and the results will be analyzed.

This study is organized as follows: Chapter 2 presents a literature review and theoretical approaches to churn studies. Chapter 3 describes the methodology. It includes the different approaches to churn modeling, including RFM, CLV, and machine learning approaches, while Chapter 4 presents and concludes the experimental results of the model. And finally, Chapter 5 includes a discussion of findings, conclusion and recommendations.

1.4 Hypotheses / Research Questions

In Turkey, especially since 2013, the Turkish lira has been depreciating against foreign currencies. Especially since 2018, this depreciation has intensified. This process increased the volatility in the exchange rate and triggered the increase in inflation. It has directed companies carrying out their commercial activities in Turkey

to keep their money in foreign currency to take precautions against this situation. As a result of this situation, the depreciation of TL has become an exponentially increasing spiral.

As a result of foreign exchange buying and selling, banks have increased their profits more than ever before. This item has gained a very important place in the total balance sheets of banks. Therefore, in this service, where it is possible to obtain high profits with very little effort, the customer has a very important place for loss analysis. Due to the current economic conditions in Turkey, foreign currency has become an investment tool for customers.

When examining a customer's relationship with a product, the customer's relationship with the bank should also be considered. Because the banking world provides many financial services and opportunities to its customers. Access to these services provided by many banks is quite easy in today's mobile world. For this reason, customer loss can occur quite suddenly and quickly for any bank. Because in the mobile world, it is no longer necessary to physically go to a bank branch to open a new account with a bank or to move cash assets to a bank. The fact that a customer opens a new account with another bank may not mean that that customer is churned. What is important for churn analysis is whether the customer continues to use the bank's services. The customer may have opened an account with a different bank because of many obligations. This does not mean that he is not satisfied with his old bank. Customer transactions through the bank should be examined over time to determine churn situations. The purpose of this temporal analysis is to observe whether the customer's transactions have become more frequent compared to the past or whether the monetary equivalent of the realized transactions has increased. As soon as the graph drawn by the transaction frequency and the monetary value of the transactions over time starts to resemble the graph of the customers who were churned before, it may be a sign that the target customer is going to churn. To reflect the effect of this situation on the model, temporal features are derived. These features may be the volume of transactions over the last six months or even the frequency of transactions performed by customers. Melian et al. (2022) state that academics and businesses are very interested in the prediction of customer churn and many studies have been suggested for this process in a variety of industries, including financial services,

banking, retail, airlines, telecommunications, etc. As a result, several studies on churn prediction in service-based companies have been published. There isn't a common model, nevertheless, according to the research on churn prediction modeling methodologies. The industry in which the customer conducts its business, the geographic region in which the customer conducts its business, and the customer's experience in the industry may all contain information about the customer's behavior. Whether a commercial customer exports or imports can affect their foreign exchange buying and selling behavior. Because such customers need foreign currency in order to carry out their commercial activities, there is no such obligation for other customers. For the customer who continues its activities in the country, foreign currency is not as essential as the customer who imports. Features like educational background or citizenship can also be evaluated within this scope. According to Li and Zhou, which customers will leave based on profitability, and what are the primary causes of customer churn are the two key issues that drive the customer churn analysis study (2020). The customer churn framework was developed primarily using the traits of past data and is used to forecast consumers who have a high likelihood of going offline.

1.5 Significance of the Study

In addition to meeting needs quickly and easily, social, and religious concerns stand out among the most important reasons for the preference for the interest-free financial system, which is in practice in many European countries. According to Hardy, Islamic banking has undeniably become a worldwide phenomenon, and as the industry expands, the policy implications of its usage demand international attention. Over the previous decade, notably after 2008, participation banking has gained credibility in Turkey. The trend is expected to persist if Islamic businesses are regarded as safe alternatives to mainstream products and observant Muslims opt to adhere to religious principles (2012). Due to the interest-free finance system, which was created to facilitate the sale of real estate, vehicles, and workplaces, is becoming increasingly widespread in Turkey as well as in the world, the competition between the sector and conventional banking is increasing day by day. In today's competitive and flexible business world, the prediction of the customers who will leave them is very important

to ensure the continuity of the products released by the companies that pioneered the interest-free finance system, which cannot provide market flexibility in their products as easily as conventional banks and try to establish themselves in the sector. Iranmanesh et al. state that marketing techniques have changed from production to customer orientation in recent years, with many firms focusing on customer relationship management (2019). Machine learning applications should be actively used in this regard. These methods are not only effective but also very costly for companies. In addition, it requires serious know-how to be applied effectively and needs expensive technological infrastructure. In addition to the rapid development and transformation of the sector, the analysis and forecasting of the customers who will leave due to the great competition between companies are of vital importance in the interest-free finance sector. As a result, RFM and simple statistical methodologies also offer excellent outcomes for small and medium-sized businesses. All three approaches will be examined, considering their advantages and disadvantages. Because the loss of customers is a great opportunity for competitor companies to gain customers, this is true for all businesses, regardless of size. In a multi-competitive environment, companies must focus on retaining existing customers by meeting customer needs because attracting new customers is often more costly than retaining an existing customer. In recent days, many companies have started to focus on CRM to cope with intense competitive pressure and customer loss. Naz et al. claim businesses know their present customers, have a deep connection with them, and possess a massive quantity of data on them, which is essential for growing income and customer value (2018). The purpose of CRM processes can be summarized as not only shaping customer perception in the organization and creating customer information tables but also establishing sincere customer relations and providing an effective competitive advantage by making the purchasing behavior of customers more understandable.

Chapter 2

Literature Review and Theoretical Approach

2.1 Customer Churn Analysis

For the customer to remain within the designed business model, customers must first be satisfied with the company they are shopping with. According to Melian et al., (2019) many organizations know that their database of existing clients is their most important asset due to the severe price wars and saturated market in which they compete. This sort of development is especially prevalent with postpaid subscriptions. A significant portion of customers who have a bad customer service experience tend to become customers of a competitor very quickly. Melian et al. (2022) point out that it is critical to construct an effective and clean prediction model that functions as customer decision-maker assistance to efficiently manage customers with a high turnover rate within a firm. To do this, several modeling and prediction approaches may be used to identify consumers who are most likely to discontinue existing services. The data mining techniques employed, the selection procedure used, the variables contained in the model, and the time taken to complete the final churn forecasting models all have an impact on the predictive model that is implemented. It is a significant and growing problem that churn forecasting, or the early diagnosis of consumers who may discontinue the use of a service, affects many sectors. Knowing that these companies acquire a rising quantity of diverse data on customer traits and behaviors on a large scale, it is conceivable to adopt novel churn, prediction models. This firstly obliges the target of the customer service units to be 100% customer satisfaction. Customers expect prompt resolution and attention regarding their complaints. There should be an infrastructure that will inform the customer throughout the process, from the recording of the complaint to its closure. Complaint management solutions should automatically inform customers via channels such as e-mail and SMS. Seeing that his report is taken care of, and efforts are made to resolve the complaint will calm the customer and enable him to approach the organization with a more positive perspective. Customers can submit their complaints not only through the call center but also through different communication channels such as e-mail, websites, chat, and social media. Complaints coming from different communication channels

may be followed up independently by different teams within the organization. This process leads to disconnection and dissatisfaction in complaint management. When a customer who sends his complaint via email calls the call center to find out the status of his complaint, encountering a voice unaware of his complaint will increase dissatisfaction. These different channels should gather the complaints coming from the omnichannel infrastructure into one center. Thus, all complaints from customers will be traceable, whether via email, call center, or social media, through whichever channel they contact you.

2.1.1 Definition. Customer churn analysis is described as an investigation into the possibility of a customer abandoning a product or service (Celik & Osmanoglu, 2019). Churn of a customer is an unwished-for event for any business. Loss of customers occurs when the customer stops using the service. In fact, this is a great opportunity to improve processes, as service providers know why a customer leaves or abandons a service and returns. This process is called "churn analysis." Customer churn is divided into two categories based on whether the enterprise and the customer entered a contractual or non-contractual relationship. In a contractual scenario, the firm and the customer work together to reduce both parties' losses during the transaction process and to bind the contract. The contract specifies both parties' rights and obligations in detail. After signing a contract with the firm, the client must complete the appropriate responsibilities under the contract to enjoy the associated rights. In non-contractual interactions, however, the relationship between the firm and the client begins with the customer's initial transaction. Customers can freely choose to enter and exit the business' service, which means customers and businesses are not required to sign a contract, and the customer's trading and loss behavior is uncertain. Customers may continue to buy products for a long time after their initial transaction, or they may never do any transactions again (Xia & He, 2018). Loss of customers, in its simplest definition, means "customers giving up on choosing the company due to competition". The concept of customer loss has come to the fore with the increase in service businesses such as banks and telecommunication companies, the changes in the loyalty of their customers, and the increase in the customers' abandonment of the company because of these changes. It is very effective that the companies have a high potential to attract customers who are dissatisfied with the rival company by creating

very different services and a wide variety of campaigns to compensate for the high loss of customers. According to Cenggoro et al. (2021), churn prediction is mostly thought of as a type of binary classification issue. Customer churn rates in telecoms were 20%, 21% in online retail, 22% in financial/credit, 24% in general retail, 17% in travel, 11% in big-box electronics, and 25% in cable, according to a recent sectoral study on customer indexing. The churn concept is a critical application area of customer analytics in the broadest sense. Churn analysis is a subdomain of customer behavior modeling in customer analytics. Customer churn prediction may be challenging due to challenges such as unclean data, technological deficiencies, and censoring of churn events. Customer transaction patterns and other elements influencing the customer's selections are mostly utilized for churn analysis.

Generally, this method is applied in banking, telecommunications, insurance, and agencies. At this point, it is very important for us to identify the factors that cause customers to stay away from the business for reasons such as withdrawal, escaping, and wear. As Celik & Osmanoglu stated (2019), it is especially significant in CRM departments' strategies in industries with subscription-based income models, such as insurance, telecommunications, or banking. In today's competitive environment, gaining new consumers is up to ten times more costly than retaining existing clients, according to industry data. In cases where customer loss cannot be prevented, it also reduces the quality of the service provided and creates a chain effect. In terms of competitiveness and sustainability, this situation should be handled as a priority. Regardless of the institution, it offers its services to the public. Churn analysis is needed to detect situations that will cause a loss of customers. Various techniques are used to measure customer satisfaction. Considering these satisfaction levels, it is necessary to consider the complaints and demands that harm them, to keep the customer inspected, and, if necessary, to make changes in the processes.

With realistic targets, it is necessary to analyze the methods of retaining customers well and to review every situation that causes loss. This analysis is not as easy for institutions that are established with large investments, have millions of customers, and target significant turnover. In this thesis, it is mentioned about the situations and conditions that are needed to question at some points in the projects besides technical applications of statistical modeling for the companies that ask how

they will identify churning customers and act accordingly. A churn analysis that examines the behavior of customers in the period that has been determined will help in this regard.

2.1.2 Why is customer churn important? According to Lalwani et al. (2022), financial institutions are organizations that handle monetary transactions such as business and private loans, consumer deposits, and investments. Due to the fact that clients are directly tied to revenues, financial firms must avoid losing customers while obtaining new ones. The importance of the churn concept in a company or business development is directly proportional to the growth of that business. Despite the fact that, in the financial services industry, the average sales and earnings for corporate customers are much higher than those for individual consumers, a variety of financial institutions lack a corporate customer churn analytical model, and most companies still rely on speculation based on the insights of salespeople or CRM department decision-makers to forecast customer behavior. As a result, this situation supports the idea that to retain consumers, financial institutions must build a churn forecasting model (Jung, & Lee, 2020). The addressee of the offered services, products and many other systems is the customer. Customer churn is a big worry for banks, according to Domingos et al. (2021), because it results in revenue loss for the sector. Banks would want to find the clients who are most likely to stop using their services for this reason. Churn prediction uses customer transaction data to estimate how likely it is that a customer will stop using a service. How popular and desired the product is also related to how many customers it reaches. Thanks to customer churn, it is observed to what extent the services delivered to the customers are used by real customers. Product or business success is determined not by the number of people reached but by the customers who consistently use these businesses. For example, a purchased application gains meaning with users who are constantly active within the application rather than with users who simply download it. Tiwari et al. (2010) claim that misclassifying non-churn as churn adds to the numbers that need contacting by the retention department, and as customers who never intended to churn are being contacted with expensive retention offers, real churners are left to defect. The most practical sources to employ are demographic information about customers, usage of services provided by the company, and billing data, for estimating customer churn. A customer who constantly uses the application

will indicate to the owner of the application that he is a permanent customer and that he will show continuity in return for the services provided. The situation of inactive customers in the application is dominated by the message that they are ready to delete that application at any time, so they will lose customers. The fact that the owners of the company make this examination and reveal the ratios ensures that the product or service is maintained efficiently and effectively.

2.1.3 What are the causes of churn? There are several main reasons for customer churn. These can be listed as low attendance, market shortage, and bad user experience. It is essential in terms of efficiency that all services prepared for the product are actively used by customers. Insufficient use of these services may mean customer churn based on the service provided. Santharam, & Krishnan, (2018) state that predicting customer churn is crucial since it helps identify clients who are most likely to discontinue the use of a service, subscription, etc. The organization will suffer a substantial loss if a client quits their memberships or services, and it will cost significantly and require a lot of effort to locate new consumers. Customer churn is a well-known issue that many businesses and sectors deal with since it results in lost sales and damages a company's reputation. It is troublesome to attract new consumers, and it is vital to develop trustworthy models that can forecast client turnover so that the business may avoid suffering substantial damage. The abandonment of all services for a product by the customer will also be a product churn. The subject of low attendance will be briefly summarized as follows: if a product released to the market is not used as often as it should be used and if all the features of the product are not active, it causes low participation for this product. On an application basis, the frequency of users visiting that application, which features they prefer within the application, etc. situations are participation. Detecting low participation is revealed by examining overall participation. These can be determined by in-app analysis methods or by surveys submitted to the customer. The surveys will reveal why the product is used incompletely and why it does not fulfill its main task. To prevent such situations, necessary improvements should be made according to the results of the survey and analysis. The main objective of banks is to make sure that existing clients are kept because it is not generally possible to attract new consumers in the ruthless banking industry. Customers' connections with banks are a goal for their long-term corporate

strategy because, like other service-oriented businesses, they are focused on delivering outstanding customer service. Because of the fiercely competitive and dynamic nature of the banking business, studies done in many nations for the finance industry have shown that guaranteeing customer happiness is a key policy to reduce client churn. Strong customer connections have the advantage of boosting customer satisfaction and, consequently, customer loyalty, which becomes the most crucial element for the sustainability, expansion, and revenue of organizations (Guliyev & Tatoğlu, 2021).

Another important factor after low attendance is the market shortage. The effect of market inadequacy on churn is related to non-essential products introduced to the market. It is a situation where the usage decreases and the products cannot be sold because of the production carried out without considering the customer's requests and needs. If the products are not used by the companies for which they are produced, if they do not find a necessary solution for any need but are still put on the market, it causes market insufficiency through these products. In the company, such situations are handled under the name of necessity in terms of making money and the continuity of the company. Reducing the product market inadequacy depends on the product development teams in the companies. These teams should prepare products that will really benefit and meet the needs within the framework of rational results. Companies should create a custom model for themselves. They should present their work and products to this customer profile. Otherwise, the company's production will be interrupted, and incorrect customer attraction will cause churn. It is possible to prevent this problem with healthy communication between the customer and the marketing team. The last notable category is customer experience. Although the products produced are the answers to the problems, they may not be easy enough to use, even if they do not have a problem in terms of functionality. Difficult interfaces and complicated tabs reduce the efficiency of that product and lead to the search for easier-to-use products. There are many different products on the market that solve the same problem. Choosing the one that does the most work among them is directly proportional to the user experience. Before purchasing a product or application, we always pay attention to the comments of people who have used that product before. It is extremely important to increase the positive impact of these comments and to make a good impression. Keeping the comments positive is about creating easy-to-

understandable products as well as products with high functionality. Thanks to a team that will ensure customer satisfaction and works towards the right needs, churn will be very low. Besides these three basic types of customer churn can be mentioned that are not just linked to experiences and satisfaction but also to both internal and external elements, some of which may be changed by customer engagement while others simply have to do with the clients' characteristics. Concepts for consumer happiness make up the first category. Second, the opportunity cost that the purchaser perceives is acknowledged as the switching cost. The relationship between consumer identities and their interrelations with some other clients is discussed in the third category. Customers' decisions and ties with the firm are influenced by these encounters and intimate relationships, which has a notable influence on how likely they are to abandon the company. (Méndez, 2021).

2.1.4 How does churn analysis benefits banks? Firms do not have an optimum churn rate. These rates vary by industry and can hardly be compared with each other to set a benchmark. It is recommended that the firm consider the customer drop rate as a reference point and plan and act to continuously improve it based on customer drop-driven analysis. Lima Lemos et al. (2021) state the efficiency of the retention action based on the avoidance forecasting model can achieve even more favorable outcomes in terms of cost-benefit ratios when it is applied with a methodology based on the return offered separately by each customer, if 20% of buyers are direct causes for 80% of the bank's results. The process of determining the profiles of customers who are likely to quit working with the company by reviewing the purchasing behavior of customers and predicting those who are likely to leave is called churn analysis. Considering that finding a new customer is much more costly than retaining existing customers, it can be easily understood how important a customer loss analysis is. This analysis has now become a tool frequently used by strategic decision-making and planning officials. In recent years, churn analysis has gained even more important in many sectors where competition is increasing, customers can easily find other alternatives, and companies are losing money. Churn analysis predicts customers who are likely to quit working with you, giving your company the opportunity to create a variety of campaigns aimed at keeping those customers from

leaving. Data mining, artificial intelligence, and machine learning techniques are used for churn analysis.

Churn rate analysis allows companies to see much more than a set of percentiles, setting it apart from many other analyses. This will allow you to understand that customers tend to quit working with the company and take preventive action before they even leave you. In this way, you can increase your loyal customer rate by preventing possible losses. Banks offer a several services to their many clients, including not only ATMs and physical branches but also, as part of their digitization strategies, online banking, and mobile applications. Customers that are more concerned with the quality of services might switch banks for their financial services due to a variety of factors, such as technical improvements, customer-friendly service, cheap interest rates, geographic proximity, and the availability of a wide range of services. A free-market economy develops when the alternatives available to customers for services increase. The upshot is that competition among banks greatly raises bank dependability and service quality, but it also stimulates the risk of client attrition. Machine learning methods make it feasible to anticipate who will churn and why. The development of a model that predicts client churn based on demographic, mental, and behavioral information is vital in the banking industry, as it is in many other industries (Guliyev & Tatoğlu, 2021). An effective churn analysis contributes to a better understanding of customer journeys. Lee et al. (2018) claim that The primary objective of churn prediction is to discover potential churners with extreme accuracy. However, a churn forecasting model should also consider ensuring higher financial gains from churn analysis. Based on the likelihood of customers leaving, companies can develop a range of CRM activities. Therefore, they may create a better customer experience that meets customer needs. According to De Caigny et al. (2020) due to the loss of up- and cross-selling opportunities caused by customer loss, it is more important than ever to recruit new clients. Therefore, it is essential to predict future churn behavior and take proactive measures to keep profitable clients. With churn analysis, it is an easy task to predict the likelihood of customers escaping with the predictions produced by high-success models. While performing these transactions, analysis can be made according to customer segments and loss amount. After this analysis, activities should be carried out to improve corporate communication and

increase loyalty to retain customers. Customer retention is a critical business goal in today's competitive landscape. As Khan et al. (2019) point out that most businesses in the present era continue to suffer greatly because of unhappy consumers switching to new direct competitors or quitting. Churn Analysis helps improve direct marketing efforts towards lost customers as well as enables them to offer proposals to re-engage those customers against the company. Designing and adapting a reliable model for predicting customer churn may significantly improve customer retention and boost revenue. (Yulianti & Saifudin, 2020). Big-tech companies that offer products and services that have become indispensable in daily life are working successfully for churn forecasting. YouTube, Netflix, and Spotify are the most well-known brands that offer music and video streaming services. Churn analysis is indispensable for online retail as well as banks and telecom companies. Ahn et al. (2006) claim that businesses must have a thorough understanding of a customer's behavioral churn route and the variables influencing it to effectively control customer churn in an efficient and comfortable manner. The literature hasn't adequately addressed these issues, though. At this point, questions like: how to start working with churn ratio estimation? What data is needed? What are the implementation steps? first comes to mind. Churn prediction models can be built using machine learning methodologies. As with any machine learning study, it needs data to be able to produce a service in this regard. Depending on the objective, what data should be collected should be defined. Then, the selected data is prepared, preprocessed, and transformed into a suitable form to build machine learning models. Finding the right methods, fine-tuning models, and choosing the best model is another important part of the job. Once a model that makes predictions with the highest accuracy is selected, it can be put into practice. The scope of the work to be carried out to create machine learning supported systems that can predict the churn rate is as follows: understanding a problem and end goal, data collection, data preparation and preprocessing, modeling, and testing, model deployment and monitoring, and finally action recommendations. The customer drop rate is obtained by dividing the rate of customers who turned their backs on the company in the specified period by the total number of customers for the month. A customer churn rate shows the number of customers that a business has lost in a specified period. For example, let's say a firm had 250 customers at the beginning of the month and 14 of them went down. It will now have 236 customers at the end of the

month. So, the customer churn rate is 14 divided by 250 times 100, which equals 5.6%. Two-way communication channels with the customer should always be open and used actively. Before the customer is lost, customers at risk of falling must be identified by conducting customer satisfaction surveys. The necessary technological and human resource investments must be made to satisfy them.

2.1.5 Types of customer loss and causes. Domingos et al. (2021) claim in their work that when a client is dissatisfied with a bank's service, customer churn occurs. Before churning occurs, banks and other financial institutions should regularly monitor the transactions of their clients to look for typical warning indications in a customer's behavior. There may be many customer types and characters, as well as the reasons for churn depending on them. There are two basic categories of churning customers: voluntary churners and un-voluntary churners. Since these consumers have had their services terminated by the business, non-voluntary churners are the simplest to spot. A corporation may terminate a customer's service for several reasons, including misuse of the service and non-payment of the service. According to the research of Hadden et al., it is more difficult to categorize a client as a "churned customer" in non-contractual relationships since this kind of churn occurs when a consumer intentionally decides to stop their connection with the supplier (2008). The important part here is that the person who voluntarily separated from the previous company and chose the competitor changed his choice. However, if various risk measures are taken in involuntary situations, a solution process can be created that can prevent this situation. Involuntary loss types are often ignored in statistical and data mining studies. The main reason for this is the inability to prevent the loss of customers. In the case of canceling the subscription, there may be issues with less communication with the customers and the desired situation cannot be determined, or in the event of the customer's death, they may be closed by their legal representatives. Closing the account may be due to the customer's own request, or it may be for mandatory reasons.

2.1.6 How to define churn threshold? The judgment of whether a client has churned is very arbitrary and subjective, and the definition of customer churn is often dependent on intuitive guidelines established by industry officials. As a result, churn parameters are greatly impacted by the modeler's personal know-how. The insights of

a business expert must choose, among other things, the behavioral indicator that will best assess and forecast standards. (Perišić, & Pahor, 2020). Rothmeier et al. (2020) claim that different labeling strategies for customer churn generate different outcomes since the identification of churn depends on the criteria utilized. Due to the hazy nature of the lost customer criterion and the lack of a defined line separating lost customers from non-lost customers, determining customer churn in non-contractual agreements is challenging. Based on the peculiarities of enterprise goods, e-commerce firms must establish the bar for customer turnover, be able to quickly identify the patterns of lost consumers, and then implement successful customer retention strategies (Xia & He, 2018). When calculating the churn period for businesses, it is useful to examine time series, and types of transactions of customers (saving for banks, spending...). Aleksandrova, (2018) points out that numerous businesses have difficulty determining when client churn has happened. This is problematic when it comes to non-contractual interactions. In these situations, business professionals must develop precise, crucial metrics for calculating real client attrition. Domain experts identify six months without a transaction as a credible measure of churn in the current case study. Although recency, frequency, and monetary values are simple to compute and comprehend, they only cover one component of client behavior. Data scientists require diverse data regarding client demands, views, socioeconomic traits, relationship data, and so on to create high-quality statistical models. Across many circumstances, such data is difficult to acquire since small and mid-sized organizations do not use a structured method of obtaining it. It may be useful to dwell on an example to express the importance of this situation. For a telecom company subscriber who talks an average of 1400 minutes a month, 200 minutes of talking this month may not be enough to be labeled "churn." because although it cannot be denied that there has been a serious decrease, the customer has continued to use the phone line for the last month, but it is also a fact that the customer speaks for 1/7 of the time compared to the normal situation. Apart from this, if a scenario from the banking sector is evaluated, it is seen that a Participation Bank customer, who has an average of 5000 TL in his account every month and uses his card for shopping 20 times in the last month, has deposited 1/3 of his salary into his gold account for savings purposes. With the new month, there is a noticeable decrease in all transactions made by the customer, who has only 200 TL left in his account, uses his card twice, and withdraws half of the gold in his gold

account at the end of the current month. How accurate would it be to add this customer to the loss list? As a result, if it is said that this customer will be churned, half of the customer's gold stops, and the customer continues to trade. Although it can be said that it has not disappeared, it cannot be denied that there has been a very serious decrease in the customer's transactions. As it becomes clear after this stage, making the modeling even more specific such that a customer will churn X% rather than not churning should also be considered as an option. Apart from this issue, there will be other points that need to be questioned. What customer data should be used in modeling? How many days in advance should the customer be monitored? The value of each customer to the company is different. Which customers should be prioritized in the churn account table? It is very important to approach modeling with these methods to obtain accurate results. Before starting the modeling, examining each question and any sub-questions in detail, and switching to modeling will contribute to saving time and money.

2.1.7 How to decrease churn rate? The ways of reducing the churn rate are generally discussed in this study. The first thing to do is to increase the data quality. The company should keep customer data in a regular and easily accessible manner and provide the necessary technological infrastructure to make it ready for processing. Following this, customer movements should be regularly examined, and measurements should be taken. As a result of these measurements, RFM analysis, CLV calculations, or machine learning methods can be used by using the available data. Campaigns should be organized using the outputs to be obtained from these analyses and presenting their services to the right customers. Lima Lemos et al. (2022) in their work found that customer usage of financial services, the amount of credit given to them (concessions), and product ownership exhibited better generalization performance than factors linked to transaction volume (balances). Consequently, developing customer relationships across product sales can be a beneficial method for client retention and churn reduction.

2.2 RFM Approach

To create a dataset suitable for analysis, first, customer segmentation should be done. Christy et al. (2021) claim customer segmentation customizes individual communications to engage more effectively with the desired groups. Location, age, gender, earning, culture, and past purchasing history are the most typical factors utilized in consumer segmentation." Customer segmentation is the first step of loss analysis. Behavioral customer segmentation groups customers who have similar desires and purchasing behaviors based on similarities among group members and differences between other categories (Hamzehei et al., 2011). Customers can be categorized in many ways. Grouping customers with common purchasing behaviors is one of the best-known methods. With this method, also known as RFM analysis, customers can be segmented quickly. RFM is short for three English words; recency (R) is the time since the last order placed by the customer. In other words, it is "the time since the last contact with the customer." Frequency (F) is the frequency with which the customer places an order. Monetary (M) is the total monetary value of orders placed by the customer. When customers are segmented with the mentioned data, the segment with the highest three values is the best customer group, and this is called the champion. With the determined percentage scales, many more different segments can be obtained. Each company can adapt this analysis to its own business structure by creating its own segments. Xia & He state that customer segmentation that is well-defined by the domain experts not only helps in the successful allocation of marketing resources. but also allows businesses to target a particular set of clients and aids in the development of healthy long-term relationships with customers (2018). After segmentation, possible loss rules are determined with templates according to the order ranges and order quantities of the customers. Again, similar customer behavior patterns can be used to determine at what point the customer tends to leave the company. After the rule sets are created, customers with churn potential can be listed in a report that will be created at the end of certain periods. Although these periods vary from company to company, weekly and monthly evaluations are usually made. The next process is all about designing campaigns to prevent customer churn.

It is very easy to find customers who churn in subscription based or contractual-based business models. The customer must fill out a form to unsubscribe. Here, the

desired campaign can be published, and the churn rate can be reduced, but customer behavior templates should be created in business models where non-subscription product-based services are sold, such as in the telecom and banking sectors, or in retail e-commerce applications. Campaigns can be organized for customers who follow the filters or rules in these templates. Unfortunately, such business models, like many other issues, require a very complex analysis process. Moving the current customer into the loyal customer category and reducing the churn rate should be among the main KPI goals of the companies. It is necessary to have some basic components, such as analytical CRM, business intelligence, and a data warehouse, to be able to make these analyses and achieve KPI targets.

For an organization, the knowledge and understanding it has about each of its customers is its most asset. To enhance client loyalty, business directors need to devise promotional campaigns that cater to the wants and needs of their target market. Due to the limited budget, it is nevertheless impossible to create promotions that will satisfy every consumer. Various individuals may have different desires. As a result, it's important to recognize the most lucrative consumer groups and concentrate on their requirements (Sabanci et al., 2020). Regardless of the number of customers, businesses that do not have sufficient information about their customers may become vulnerable to life-threatening risks and cause serious damage. Commercial success that will continue in the future by developing a sustainable business model can be achieved by understanding and analyzing this great power that can be obtained by knowing the customer and by constantly transforming the business according to the results of the analysis.

What distinguishes one client from another? What features stand out in customers' interactions with the company? Which product shows more interest? Why is the customer particularly interested in any product? Which product is purchased the most? Which customer has bought the most so far? So which ones were sold the least? What is the sales speed? RFM analysis is one of the most fundamental statistical tools in the analytical marketing world that asks all these questions and seeks answers.

There is a table commonly used in RFM analysis. Usually, two parameters are used in this table. These are the recency and frequency values. First, after the recency,

frequency, and monetary values are found, customers are divided into segments by giving values between 1 and 5. 5 means very good, and 1 means very bad. By looking at the parameter values, the segments of the customers are determined. This table is the one that explains the basic logic of RFM analysis. Each segment should be evaluated within itself, and policies appropriate for each segment should be determined.

For example, for the need-attention group with a recency value of 3 and a frequency of 3 in the table, by sending special emails, making promotions, and moving up the frequency values, the recency value can be moved up.

2.2.1 Recency, frequency and monetary values. Recency: The term "recency" describes the period that has passed since the most recent consumption activity. Fresh customers, in the opinion of many direct marketers, are more inclined to repurchase than less recent buyers (Birant, 2011). The closer the customer's last shopping date is, the easier it will be for them to remember that brand for future purchases. In other words, the more time that passes since the last purchase, the higher the probability of the customer forgetting a company. When investigating the probability of customers making a purchase at any 't' time, recent shoppers are more likely to repurchase than those months after their last purchase. This basic assumption can be used to get new shoppers to revisit the business and spend more. To not miss the customers who are relatively engaged, marketing communication can be done, and a special offer can be presented to them to remind them that it has been a while since their last transaction. The only action to be taken to calculate the recency value is to determine the difference in days between the day the customer made the last transaction and the day of the RFM analysis. The recency value is calculated by subtracting the customer's last dealing day from today. The lower the recency value in the RFM analysis, the higher the customer's score (Rs) should be.

Frequency: The number of transactions done by a consumer in a certain time period is referred to as "frequency." This metric is based on the idea that customers who make more purchases are more likely to make additional purchases than customers who make fewer purchases (Birant, 2011). Purchasing frequency varies by industry and product type. If a customer's purchasing cycle can be predicted, then

future sales turnover can be predicted by reminding the customer of their next purchase. In fact, this reminder can also be positioned as a "service" that will make the customer's life easier. In the RFM analysis, the total number of purchase transactions for each customer must be found in the customer database to calculate the frequency value. Although many products are made at the same time, this should be considered a single purchase. For example, if the number of products purchased by the customer in the last 2 years, which is the subject of the RFM analysis, is 9, but the number of purchases is 3, the customer's frequency value is 3. The higher the frequency value in the RFM analysis, the higher the client's score (Fs) should be.

Monetary: The accumulated total amount of money spent by a certain client is referred to as their "monetary" (Birant, 2011). It is inevitable that the customers who spend the most money are the marketing focus of the business. However, if marketing spends all its attention encouraging these customers to continue their purchases, it can miss potential customers and customers who haven't seen their real buying potential yet. Still, with a simple ROI calculation, it's only natural that the most marketing investments go to the customers who are the most profitable (or have the most earning potential). While calculating the monetary value in the RFM analysis, the total sales turnover of each of the customers is determined separately. The higher the monetary value in the RFM analysis, the higher the customer's score (Ms) should be.

2.2.2 What does RFM do? RFM analysis allows making comparisons between existing customers by making a differentiation between them. Kabasakal states that RFM is a simple-to-use approach for consumer segmentation that breaks down buying patterns into a small number of factors and finds customer groups with members that have similar buying habits (2020).

Accordingly, marketing activities for existing customers can be planned based on focus segments. It also creates an opportunity to develop segment-based sales tactics for potential customers. By making RFM analysis, companies decide which customers should invest more, which customers should remember their brands, and which customers should try to make them feel more special. Or they realize who the customers they are about to lose are and determine that they need to take proactive measures against them.

2.2.3 Why is RFM analysis useful? RFM helps to get rid of the high costs of mass marketing by reaching the right customers at the right time. A more trustworthy viewpoint could be provided by customer segmentation, which is generated by taking additional factors into account. This may help businesses better focus on their strategy. If businesses had more devoted clients, for instance, one way to keep them would be to provide more individualized promotions. (Dogan et al., 2018). Examining the past purchasing behavior of customers and segmenting them helps to get to know the customers more closely and increase sales. Thanks to targeted and personalized marketing campaigns with RFM, investment income is increased, customer loyalty and retention rate increase, and therefore a higher CLV is achieved. The values of three factors—recency, frequency, and monetary—have been found to be directly linked to a customer's lifespan and maintenance (Christy et al., 2021).

RFM analysis is based on a concrete, simple, measurable factor related to customers. Birant states that client segments with comparable RFM values are discovered so that a variety of marketing tactics may be implemented for distinct customer segments. Besides, classification rules are developed by employing demographic information (gender, ethnicity, education level, and so on). RFM values for consumer segments enable better forecasting of future customer behaviors and targeting of customer profiles. (2020). Thanks to RFM analysis, the most important condition for carrying out predictable and sustainable marketing and sales activities is provided by identifying customers with segment labels. To reduce marketing costs and increase ROI (return on investment), it was stated that before defining the campaigns for the customers, the behavior of the customers should be examined, and they should be segmented. Segmented, targeted, and triggering (automatic) campaigns constitute a significant part of the marketing ROI. According to the Dogan et al. (2018) Improved client segmentation is critical for retail businesses. Because categorizing clients based on comparable requirements, desires, and behavioral patterns enables organizations to better comprehend their potential customers, as a result, businesses might take part in activities such as tailored branding, pricing management, promotional offers, boosting client points of contact, and so on. Instead of reaching the entire target audience with non-personalized and non-targeted campaigns, the right message should be delivered at the right time to customers segmented and divided into small clusters with RFM.

Thus, sales and profits can be greatly increased. In every study, it was discovered that recency, frequency, and monetary values—both as raw figures and as scores—were major determinants of customer turnover (Aleksandrova, 2018).

2.2.4 How is RFM applied? Customer-based scoring is performed on the recency, frequency, and monetary metrics. Considering the behavior of all customers during the same period, the rank of the customer is determined by these three metrics. The quantizing of the continuous values is used in the RFM scoring procedure to quantify consumer behavior. Regarding recency, the first quintile with the greatest values (and smallest) is denoted as 5. Four represented the following quintile, and so on. Resultantly, it is possible to represent all customers by the numbers 555, 554, 553... 112, 111. The best and worst customer groups are 555 and 111, respectively (Ernawati et al., 2021). For example, a customer who bought a product 3 days ago gets a 5 in the recency score; another customer who made the last purchase 120 days ago can get a score of 1. Generally, five different intervals are determined according to the recency values of all customers. More customized ranges can also be determined according to the company sector, product, and customer distribution. Similarly, scoring is done for frequency and monetary. Customers who shop more and spend more get better scores on these metrics.

2.3 Customer Lifetime Value Prediction Approach

Customer Lifetime Value (CLV) means the lifetime value of a brand to the customer. It means the monetary value of the value that will emerge in the future for the relationship of any brand with the consumer, in the present time. Customer lifetime value is also evaluated as the value that website visitors create for the company after they become customers. According to Sabbeh, CRM systems target the most profitable consumer segment by using predictive analytics and statistical approaches in order to increase client retention rates. Based on an analysis of customers' behavioral, geographical, and social data, those models may forecast clients who have a high likelihood of leaving a business in order to create individualized and client-focused marketing campaigns to increase client happiness. Identification, attraction, retention, and development are the four major phases that make up the customer-company interaction lifecycle. (2018).

2.3.1 What benefit does CLV provide to businesses? If the CLV value is increased by making the correct calculation, it provides great benefits to businesses. To summarize these benefits, listing them gives you an idea of what to improve and how to do so. As a result, you can't improve anything you don't measure. CLV calculations also provide you with great clues in this area. As Binh et al. (2021) state in their work, because the true worth of each customer is measured, CLV not only makes a substantial contribution to market segmentation but also assists promotional strategies. In particular, because of the calculation, each component can be evaluated among itself. In which areas will there be cost restrictions? Which areas will have a higher income? It is directly related to the costs that will be incurred in acquiring customers. Keeping existing customers and adding new ones to the organization will be possible thanks to CLV. CLV measurements don't just indicate the amount of profit that will be made. The expenses incurred for this profit, including personnel expenses, advertising work, etc., are everything. Without these measurements, there is a risk of under or over-expenditure. These risks cause expenses to be carried out unplanned and without estimation.

2.3.2 Why is CLV important for organizations? Customers who are in constant interaction with different channels, especially mobile, want to maintain their loyalty for a lifetime. Of course, the quality of the products or services you offer plays a role in the formation of this situation. However, you can increase customer value with other investments you will make. A few examples can explain why CLV is important for a brand. The organization, may devise appropriate strategies to retain current customers and find and entice new clients, based on this categorization and identification of customers. To accomplish an effective CRM strategy, firms must thus look for assurance of outstanding quality, and the effort should start with people who are in close contact with consumers. Eventually, effective marketing initiatives, price reductions, and incentives for lucrative clients will be methods for luring and keeping clients. (Farzanfar & Delafrooz, 2016). Since CLV is focused on retaining existing customers, the amounts the company will earn will be much higher. Also, when compared in terms of revenue, the profit generated by the new customer will be lower than that of the existing customers. It paves the way for a further increase in brand value and awareness. In fact, your brand becomes more popular among other family

members, even over generations. Promoting the brand to new customers and ensuring their loyalty is not just about spending money. Companies will also need a certain amount of time for customers to believe in their brand and make purchases. However, it is a situation where they will not spend a lot of effort on the loyal customers added to their structure. According to Libai et al. (2020), The customer conversion rate will rise with more accurate and effective targeting, which will move down the cost of gaining new customers. An accurate CLV calculation will show the customer base that generates the most revenue. Special campaigns and promotions for this audience will be very effective in increasing the companies' sales.

2.3.3 How to calculate customer lifetime value for CRM? One of the most important criteria to be considered while developing marketing strategies is "customer lifetime value," which provides the opportunity to create sustainable income for businesses. However, as Koike et al. state that even though LTV is a crucial marketing concept, there is no agreed-upon definition, and it differs from source to source (2022). The profit made from the first purchase a customer makes to the last time they stop buying from you gives the metric "customer lifetime value." The CLV computation is often done within a particular time frame or starting from the date of the initial transaction for the customer (Dachyar et al., 2019). The resulting final situation is used in acquiring new customers, increasing the basket size with direct and cross-selling techniques, monitoring customer behavior, and making strategic decisions. Unfortunately, if the Customer Lifetime Value metric, which is not yet calculated for many businesses but is the most important performance criterion, is calculated correctly, the size of the costs to be spent on acquiring a customer can be determined. In this way, it can be predicted that it will be more profitable for companies to acquire new customers or increase the income obtained from existing customers. Monitoring customer lifetime value is helpful for focusing on what needs to be done to retain and satisfy existing customers. A business model with a high customer lifetime value can be expected to be both sustainable and more successful. The average lifetime revenue

from a customer depends on the ratio of average order quantity to recurring orders and the estimated time to retain the customer.

$$CLV = \sum_{t=0}^{T} \frac{(p_t - c_t)r_t}{(1+i)^t} - AC \tag{2.1}$$

Where;

p_t = *price paid by a customer at time t,*

c_t = *direct cost of servicing the customer at time t,*

i = *discount rate or cost of capital for the firm,*

r = *probability of customer repeat buying or being* alive *at time t,* t

AC = *acquisition cost of the customer, and* T = *time horizon for estimating CLV.*

2.3.4 How to increase customer lifetime value? There can be many ways to increase LTV in different industries. Up-selling, cross-selling, and client retention are the three fundamental actions required to improve customer value. Upselling is selling the same types of things that a client has already purchased, whereas cross-selling is selling products that the consumer has never purchased. (Hwang, 2004). But basically, a CRM approach is possible. The first of the ways to do this is to talk to customers with the highest LTV. Missing points can be learned by talking to customers with the highest LTV. For example, when communicating with the customer, the main purpose should be to clarify the following points: Which feature of the product is used more? What is the product used for? How did they find out about the product? How did the process progress after starting to use the product? Why did they continue to use the product? By using the feedback containing this information, valuable tips can be obtained to increase customer satisfaction. Apart from the dialogue with customers with high LTV, dividing LTV value into groups and plans is another approach that stands out. By grouping the users according to the plans their usage, an inference can be made as to which plan generates more returns. For example, lower subscription plans often have a lot more active users, but they remain low in terms of LTV. On the

other hand, users with high membership plans are more inclined to continue, so the churn rate will be low, and the returns will be high. As a result, it can be concluded that it is more appropriate to devote more time and resources to acquiring and increasing users in the mid-high segment. Apart from all this, the main target should be increasing the average revenue per user. However, one should not focus on the cost of acquiring customers in the first place, as it is likely to result in a single purchase action. CLTV focuses on customers who are in constant contact with the brand or show potential to be; it is a longer process. The main goal should be to reduce the customer churn rate. As a positive effect of this, the LTV value will also increase automatically. The longer the time that users can continue to use the company's service, the higher the LTV value can be.

2.3.5 Recognizing customer differences. Traditional segmentation work is inherently tedious, but it is often easy to organize. Jasek et al. point out that e-commerce is still lacking more comprehensive comparison assessments of CLV models done on many datasets using current empirical data. (2019). Difficulties in merging and updating customers' data and resolving existing inconsistencies cause inefficient database formation. However, there is no marketing without data. The more up-to-date and easily editable the customer data, the more consistent the marketing strategies will be. Given enough data and an easily organized environment, marketers can integrate customer data, transactional data, and even metadata to analyze customer behavior, develop strategies specific to each channel, and implement marketing practices that allow them to achieve a higher campaign conversion rate. Customer data from different sources should be collected in an integrated manner. After checking the consistency, accuracy, and validity of this data, unnecessary copies are eliminated, and it is seamlessly integrated with existing applications. Multidimensional dynamic segmentation can be applied by using a combination of customer information, tags, and digital behaviors. There are many different formulas and approaches used to calculate lifetime customer value. The average lifetime revenue from a customer is based on the ratio of average order quantity to recurring orders and the anticipated length of time to retain the customer. In a monthly subscription model, a customer's lifetime value can be easily calculated by multiplying the subscription price by the average subscription period. For an e-commerce site that sells products, the recurring

order rate can be considered instead of the subscription period. The maximum expenditure to be made to acquire customers should be planned by evaluating the average amount of shopping that an average customer will make from the company throughout the entire relationship process, together with the profitability ratio. Monitoring lifetime customer value helps provide the information needed to focus on what needs to be done to retain and keep existing customers happy. Finally, it's important to remember that a customer's lifetime value to the company may not be limited to the earnings it generates through their relationship. Recency is a considerably bigger factor in determining CLV for customers who have made a lot of purchases in the past than it is for customers who have made few interactions (Fader et al., 2005).

2.3.6 CLV with BGNBD & Gamma-Gamma Model. For calculating customer value, the CLV model implementation utilizing accessible historical data might be useful. The key challenge faced by e-commerce enterprises is the effective selection of a competent CLV model as a decision-making platform for marketing management appropriate for the circumstances of the organization (Jasek et al., 2019). In the e-commerce or retail industry, the relationship between customers and companies is a non-contractual one. In such a contract-less business world, churning (or losing) users is usually a silent action, and it is difficult to tell whether these users are asleep or churning. There is no agreement in the literature even though CLV indicators and techniques of assessment are crucial instruments for successfully building business strategies and acquiring a competitive edge over the market leader. The fundamental negative binomial distribution (NBD), Markov chains, regression, and RFM techniques are a few of the methods used to quantify CLV (Hajipour & Esfahani, 2019).

To more effectively generate CLV, models such as RFM and probabilistic, economic, or persistence models have been created. For starters, RFM models are deemed unsatisfactory since they outperform other CLV approaches and merely attempt to forecast behavior in the following period. However, RFM may be used as an input for numerous probability models, regarding the RFM model's dissatisfying outcomes. Probability models correspond to representations in which observable behavior is understood as the comprehension of a process guided by latent behavioral

features that vary between people (Do Ruibin et al., 2018). With the models to be built with the past shopping habits of the users, it is possible to predict which customers will come back in certain time periods, the most profitable customers, and the CLTV values of the customers. Recent customers are almost certainly alive. Customers who shop frequently but haven't visited recently are users who go towards becoming churn. The more these users have shopped in the past, the more likely they are to churn. Modeling CLV in this situation is made more difficult by the ignorance of customer defections. (Singh & Jain, 2017). Before moving on to the time projected probabilistic CLTV estimation, customer value, as stated earlier, is equal to the number of purchases multiplied by the average earnings per purchase. Customer lifetime value is also equal to customer value multiplied by profit margin. We can now operate on the probabilistic states of these values for the time projected probabilistic CLTV estimation. The expected CLTV is equal to the expected number of purchases multiplied by the expected profit margin. This description will provide the opportunity to make predictions using probability distributions. Essentially, the formula is expressed by the conditionally expected number of transactions and conditionally expected average profit values. Predicting the volume of transactions and the average customer lifetime is the main problem involved with CLV analysis in non-contractual contexts. (Mzoughia & Limam, 2014). The purchasing behavior of the whole audience is modeled with a probability distribution for the conditionally expected number of transactions. Then, the behavior patterns modeled with this probability distribution are conditionally shaped according to the individual, and the expected number of purchases (expected transaction numbers) for each person is estimated. After learning the general purchasing behavior of customers with probability distribution and establishing a probabilistic model, the expected number of purchases for each person is estimated by shaping the features of that model for the individual. Likewise, for the conditional expected average profit, after modeling the average profit margin of the whole audience as probabilistic, when the person characteristics are entered using this model, the expected average profit margin for each person can be calculated, conditionally fed from the distribution of the main audience for the customer. The CLTV value can be calculated with the two models used in these processes. Accordingly, CLTV will be equal to the product of the BG/NBD model and the Gamma-Gamma model.

This also creates difficulties with the Customer Lifetime Value account. For non-contract business models, CLTV can be evaluated with BG/NBD and Gamma-Gamma models. According to the Jasek et al. Mostly all forecasting models have shown generally positive and reliable outcomes on a large percentage of datasets, both at the customer level and for the metrics for the entire customer base. In all assessment criteria, no model outperformed the others (2019). Although there are many alternatives to the BG/NBD model, it is one of the most widely used. According to the research on the comparison is concentrated on the domains of B2C interactions and online commerce by Jasek et al., and the major conclusion is that, among all evaluation criteria, no single model showed better performance than the rest. The BG/NBD and Pareto/NBD models, although not the best in all the chosen criteria, produced solid and consistent results while falling short of the ideal model for online retailers. (2019).

As a result of this model, the expected average profit and expected number of transactions for each person can be calculated by reducing the general behavior characteristic of the audience to each individual and asking the model to incorporate the individual characteristics. The advanced and time-projected version of the CLTV value can be calculated by multiplying these values. Companies are unable to determine whether a client has defected or plans to continue doing business with them under CLV models for non-contractual scenarios. (Singh & Jain, 2017). Thus, CLTV calculations can be performed for each person by using the correct formatting technique to accommodate the characteristics of the general audience carried by the conditional expected expression.

BG/NBD Model (Beta Geometric / Negative Binomial Distribution)

These models assume that the number of transactions made by each customer follows a Poisson process with a heterogeneity in transaction rates across customers following a gamma distribution. These assumptions give a Negative binomial distribution (NBD) for modeling the number of transactions made by the customer while customer is alive. Beta geometric distribution and negative binomial distribution are discrete probability distributions that are studied in statistics and probability

sciences (Mzoughia & Limam, 2014). Using these distributions, the expected number of purchase transactions can be calculated. First, the random variable is defined as the variable that takes its values from the results of an experiment, but the expression "expected" here is used to express the expected value of a random variable. The expected value of a random variable is the mean of that random variable. The main purpose is to model the purchasing behavior of the general audience with probability distributions and reduce it to individual decisions. Singh et al. stated in their work, consumers are expected to defect right away after making a purchase, and the total number of purchases a customer makes over the course of her lifetime is dispersed over the population according to a Beta-Geometric distribution in the BG/NBD approach (2009). In addition, the BG/NBD model is a probabilistic model that is also used as a stand-alone sales forecasting model. BG/NBD probabilistically models two processes for the expected number of transactions: the buying and dropout processes. The BG/NBD requires three historical measures for each customer: *x, tx* and *T*. The measure *x* is the total number of transactions made during the period of observation; in our case we need the number of transactions *xi* in each time interval *ti*. Indeed, we consider that the time intervals have the same duration *u* (e.g., week or month). The measures *tx* and *T* are also assumed to be multiples of the time interval *u*. For each customer, let us assume that the number of transactions *xi* which occurred during each period *ti* during the period (0, T] is known (Mzoughia & Limam, 2014).

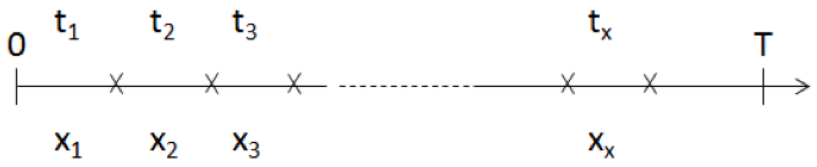

Figure 1. The diagram depicts a timeline of a customer's transaction number and time interval within period T (Mzoughia & Limam, 2014).

In the context of modeling the purchasing process, the number of transactions to be performed by a customer in a certain period is distributed with the transaction rate parameter. In other words, customers will continue to make random purchases around

their transaction rate if they are alive. Transaction rates vary for each customer and are gamma-distributed for the entire audience. On the other hand, the dropout process models the user's abandonment process. Each customer has a dropout rate with probability p. A customer drops with a certain probability after making a purchase.

$$E(Y(t)|X = x, t_x, T, r, \alpha, a, b) \quad 0.2)$$
$$= \frac{(a + b + x - 1)}{(a - 1)}$$
$$\times \frac{[1 - (\frac{\alpha + T}{\alpha + T + t})_2^{r+x} F_1(r + x, b + x; a + b + x - 1; \frac{t}{\alpha + T + t})]}{1 + \delta_{(x>0)} \frac{a}{b + x - 1} (\frac{\alpha + T}{\alpha + t_x})^{r+x}}$$

Gamma-Gamma Model:

It is used to estimate the average profit a customer can generate per transaction, and the monetary value of the customer's transactions is randomly distributed around the average of the transaction values. The average transaction value may change between users over time, but not for a single user. The average transaction value is gamma distributed among all customers.

$$E(M|p, q, \gamma, m_x, x) = \frac{(\gamma + m_x x)p}{px + q - 1} = (\frac{q - 1}{px + q - 1}) \frac{\gamma p}{q - 1} \quad 0.3)$$
$$+ (\frac{px}{px + q - 1}) m_x$$

Customers made less than 1 purchase at a rate of 0.205 percent. These customers are not included in the calculation. Transactional buying and selling data are separated based on the business key, which is the unique id of each transaction performed for each customer, and R, F, M, and T values are calculated. This data will be used in the BG/NBD calculation, and the probability of customers trading again will be calculated for any future period. The most important issue in this section is the need to estimate the potential value of new customers. Whether the customer has just come and bought very few products or is a highly profitable customer with whom good relations have been established for many years, it is the desire of all companies to assign a value to this customer. Therefore, it is necessary to make an estimation of whether this customer is very valuable or insignificant. If this situation can be foreseen, we can

follow a marketing strategy with the customer accordingly. In this case, making CLTV predictions for the future is a very important point in CRM studies. Another requirement here is that although the customer has an average income and frequency of purchases, the reason for not being able to predict the future is because there is no analysis for the company about the behavioral distribution of all customers. Therefore, companies start off without this information. In addition, it is not possible to make an estimation for individuals according to this probability distribution. With BG-NBD and Gamma-Gamma models, this problem will be solved, and forward-looking predictions will be made.

2.4 Machine Learning Approach

Machine learning is a field that we hear and read about all the time and that is referenced in the background of many artificial intelligence projects around the world. Since it is so important, it is getting stronger every day with new algorithms. Another alternative to be used for churn analysis instead of simple rule-based applications or probabilistic calculations is statistical learning-based approaches. Here, different learning methods can be used according to the business model. For example, as stated before, there is no need for labeling or specifying a threshold for the churn since the churn target value is known for each customer in subscription-based cases. Regression and classification are the two fundamental issues that fall within the field of supervised learning. The target attribute's domain in categorization is limited and categorical. In other words, the learning algorithm is aware of a limited number of types or groups that may be used to forecast a sample. When a classifier is taught using a set of training data, it must assign a class to an unknown sample. The purpose of classification is to distinguish certain cases from others, with the primary goal of making accurate predictions. Accurate forecasts for future events may be generated if a model matches the historical data (Batista & Monard, 2002). Classification can be made with any supervised model. As Lima Lemos et al. point out that in the financial industry, either random forests or the ensemble might be utilized to guide CRM initiatives to support customer loyalty, maintenance, and long-lasting connections with these institutions' customer bases (2022). However, in non-contractual situations, many different

alternatives can be used. Customer churn prediction is one of the most effective machine learning applications in the business world. It's a very simple problem if you have a ton of high-quality, appropriate data to work with. Tackling customer churn may be a wonderful way to boost profitability (Singh et al., 2022). By using a certain set of variables with unsupervised methods, the activity status of customers can be divided into clusters. Each cluster formed here can be handled as a customer segment, and different strategic actions can be developed for them. Or a threshold covering all customers is drawn, considering the behavior of customers who do not act for a certain period, and the customers are labeled using a statistical method. Apart from all these, according to de Lima Lemos et al., due to the temporal nature of customer churn analysis, predicting customer turnover is a difficult challenge that adds to the complexity of analysis and interpretation. The results do, however, demonstrate that machine learning may help banks automate their comprehension of their clients' behavior, enabling them to take proactive measures to avoid future customer churn and minimize potential financial losses (2022). A solution to the problem is sought using a time series approach using only transactional data. As a final alternative, churn analysis can also be treated as an anomaly detection problem.

2.4.1 Exploratory data analysis. Extracting or identifying trends and patterns in raw data is a crucial step for machine learning applications. Exploratory data analysis (EDA) allows for handling it using statistics or graphical representations (Anitha & Patil, 2019). It is an approach to analyzing datasets to summarize their key features, often by visual methods. EDA is also used to scan a dataset for a variety of hidden factors that influence the effectiveness of prediction (Javed et al., 2021). In 1962, in his article "The Future of Data Analysis," John Tukey expressed his concerns about the lack of progress in the field of statistics. He stated that there is too much focus on statistical mathematics, which is insufficient for data analysis, and he envisions a movement as a solution to the inflexibility that characterizes the discipline. At a statistical meeting, Tukey took the first revolutionary step by referring to himself as a data analyst, not a statistician. EDA provides a summary of the data set's row and column counts, missing values, data types, and preview. Correct inaccurate numbers, manage missing data, and clean up contaminated data. In EDA, data distributions are represented graphically using bar charts, histograms, and box plots. Another

application of EDA is creating a heat map to represent the relationships (correlations) between the features included in the dataset (Sahoo et al.,2019). EDA presented a fresh, assumption-free, non-parametric problem-solving approach that uses self-training techniques. Data scientists make decisions on which data to include, how to process and display data, and how to supplement data during the so-called exploratory data analysis phase, when the emphasis is on data understanding and data selection (Vogelsang & Borg, 2019).

2.4.2 Data preprocessing. The collected data may be incomplete, inconsistent, or out of date. Pre-processing should be done for quality data. According to Sahoo et al., it involves finding inaccurate data, eliminating unnecessary information from the data, and then restoring the incorrect data. Correcting mistakes and validating the data is the real process of data cleaning. Cross-checking the data might help to remove the mistake. Validation can be used to address problems (2019). It is one of the most essential steps in building machine learning models. The steps to clean, transform, and fit the data are a large part of the work to be done. Data collected from multiple sources is often found in an unorganized form. This is a situation that affects the prediction performance of models. Therefore, raw data must be modified before training, evaluating, and using machine learning models. It is one of the most essential steps in building machine learning models. The steps to clean, transform, and fit the data are a large part of the work to be done. Data collected from multiple sources is often found in an unorganized form. This is a situation that affects the prediction performance of models. Therefore, raw data must be modified before training, evaluating, and using machine learning models. Several activities are part of this step. For instance, data cleaning deals with removing noise and inconsistent data. Data integration, or the process of merging many data sources into one, is a further step. The process of transforming and consolidating data into formats suitable for certain data management or aggregation activities is another crucial component. The next step is data reduction, which involves choosing and extracting characteristics and examples

from databases (García et al., 2015). The data preprocessing steps were carried out as follows:

1- Zero Analysis: Zero analysis means that you are finding all the zeroes or solutions to a particular function. The usual functions that you will work with are polynomials of the following form.

2- Rare Analysis: Adding an extra column for rows that are rarely seen before encoding the data creates a structure that unnecessarily complicates the model in terms of performance. To prevent this, different rare variables are gathered under the rare class.

3- Missing Observation Analysis: Missing data is the absence of a value in any cell in the data set. What is meant by "absence" here is not that the value is zero but that there is no response to the observation. As Fan et al., (2021) states that when creating operational data, there are two standard approaches to handling missing numbers. As most machine learning algorithms cannot easily handle missing values, the first option is to simply exclude data samples with missing values. Only when the fraction of missing values is negligible is such an approach appropriate. The second approach is to use missing value imputation techniques to substitute inferred values for missing data. K-nearest neighbor method has been selected to impute missing values. Missing value is filled by the mean value of the corresponding column of the nearest neighbor of corresponding row that have no missing values in KNN method. The nearest neighbor can be defined in terms of Euclidean distance (Zainuri et al., 2015). The K-Nearest Neighbor method works with an algorithm that matches a selected value with its nearest neighbors in a multidimensional plane. The k-nearest neighbor technique is suggested by Batista et al., (2015) as a method for estimating and filling in for missing data. The key benefits of this approach are that k-nearest neighbors can predict both discrete and continuous qualities and that it does not necessitate creating a predictive model for each attribute with missing data. The k-nearest neighbor algorithm does not yield explicit models (such as random forests). As a result, by simply modifying which attributes are used in the distance metric, the k-nearest neighbor method may be readily adjusted to work with any attribute as a class. Also, this approach can easily treat examples with more than one missing value. This

method calculates the distance between the missing observation and the k nearest observations and assigns the average of these distances to the missing observation. The "k" observation value chosen in the KNN method is critical. A high k value considers observations that differ significantly from the missing value, while a low k value excludes significant values for a good estimate. For this reason, by trying different “k” values, the most suitable observations for the data set can be found. Its advantages are as follows: It can also be used for all lost data mechanisms. It can be used for all data types (continuous, discrete, categorical, etc.). It can make highly accurate predictions with enough data. The disadvantages can be listed as follows: Since the KNN method makes a distance-based calculation, the data set must be normalized to obtain the most accurate estimates. It works slowly on crowded datasets with many variables.

4- Outlier Observation Analysis: An observation that is significantly different from other observations is called an outlier. According to García et al., (2015) Outlier identification is the method or process of identifying unusual occurrences that deviate significantly from general trends or expected behavior. These are known as outliers or anomalies. Outliers behave differently from the rest of the dataset and therefore attract attention. Outliers can be caused by many different reasons. For example, errors during data entry or measurement may be caused by corruption. In addition, it is possible to encounter an outlier performance in real situations without any errors. like an athlete who broke many records in a sport or abnormal demand formations in the market. It is critical to understand that not every outlier is necessarily an error. Most statistical approaches can find abnormalities in quantitative research (Fan et al., 2021). In some cases, outliers are a harbinger of high variance in the data set.

5- Variance Analysis: The standard deviation is a measure of how far data values are from the mean. The standard deviation for a population is expressed as σ; for sample data, it is denoted as s or s' for standard deviation. The standard deviation gives us information about the prevalence of the distribution. Accordingly, when the standard deviation value is small, the data are close to the mean, the larger the data, the further away from the mean. The size of the standard deviation indicates that our data has a high degree of variability. Therefore, the distribution is not homogeneous, and the difference between the observations in the data set is large. A standard deviation of zero means that all data are equal. The variance is the square of the

standard deviation and is denoted by the notation σ^2. Accordingly, the standard deviation is the square root of the variance. When calculating the variance with the squaring process, we give more weight to the values that are far from the mean.

6- Feature Scaling: Another stage of model data preparation is data transformation. Feature scaling is frequently required to maintain machine learning model validity, particularly when the input features have various scales. (Fan et al., 2021). Scaling process is applied to reduce the values of continuous variables in the data set with varying scales to between 1 and 0. The measuring unit chosen might have an impact on modeling of data. All the characteristics should be stated in the exact measuring units and on the same scale or range. Normalizing data has the purpose of providing equal weight to all qualities, and it is very important in statistical learning approaches (Garcie et al., 2015).

7- Correlation Analysis: Correlation analysis is a statistical method that provides information about the relationship between variables and the direction and severity of this relationship. The mathematical relationship between two or more variables is analyzed by "regression analysis," and the direction and degree of the relationship are examined by "correlation analysis." The correlation coefficient is a coefficient that shows the strength of the relationship between the dependent variable and the independent variables. For example, the correlation coefficient can be used to examine whether there is a relationship between the student's study time and the statistical grade he receives, or the relationship between the daily return (X) of a stock traded in the stock market in a certain period and the daily return (Y) of an index in which it is included. The correlation coefficient gives information about the direction of the variables and how they interact. The correlation coefficient is defined as the ratio of explained variance to unexplained variance. The correlation coefficient is the measure of the linear relationship between two variables, is independent of the units of the variables studied, and is between -1 and 1, so it can't be 5 or -5. When the correlation coefficient approaches 0, it indicates the existence of a weak relationship between the variables. If the variables increase or decrease together, there is a positive relationship, and if one variable increase while the other decreases, there is a negative relationship.

2.4.3 Feature extraction and feature engineering. As stated earlier, a "feature" is any measurable input that can be used in a predictive model. Apart from feature selection, which picks only relevant features from existing variables, feature extraction seeks to create new features based on linear or nonlinear mixtures of dimensions on an analyzed dataset. (Fan et al., 2021). At the same time; It can also find its place in the literature as a dimension, attribute, and dependent variable. It is interested in identifying and generating additional features that are computed as a function of the base features in feature extraction. (Garcia et al., 2015). The number of features is as important as the features themselves; if the features in the feature vectors that represent real-world data given to the model are not sufficient, the model will not be able to fulfill its main task; if there is a feature vector with redundant and irrelevant features, the model will still not produce correct results. In the machine learning process, not only the model but also the features to represent the data are selected. According to Huang et al., (2012) As one of the most important factors, feature extraction can influence the performance of predictive models in the terms of prediction rates (high TP and low FP). If a robust set of features can be extracted by this phase, the prediction rates of TP and FP can be significantly improved and reduced, respectively. Apart from the data at hand, in line with the requirements of the business area that the product addresses, the features should be subjected to different processes so that they do not produce bias. These are pre-processes such as localization and standardization. Xu et al. state that data processing requires feature engineering. High reliability in machine-learning activities requires careful feature engineering and selection (2021). Well-processed features improve the ability of the resulting model to complete the desired task; however, if features are not created properly, a much more complex model may be required to achieve the same level of performance.

2.4.4 Feature selection. For real-life problems with much more data and dimensions, the purpose of feature selection is to discover the smallest number of attributes such that the resultant probability distribution of the data output classes is as near to the original distribution as feasible. It not only improves understanding of the extracted pattern but also accelerates the learning period by decreasing model complexity. (Garcia et al., 2015). Feature selection is a useful way to simplify the model. By focusing on the attributes that are thought to be most helpful to a model in

predicting the target variable, feature selection approaches aim to decrease the number of data points. (Brownlee, 2019). With feature or variable selection, the overfitting situation that may occur in the model can be prevented. Chai et al. claim that supervised feature selection is frequently used to solve classification issues, and its core idea is the relevance or link between the attribute and the class label (2018). In a model with more than enough variables, the training performance can be extremely good. However, if the situation is reversed when we test, there may be variables that cause overfitting somewhere. If these are removed from the model, the actual accuracy of the model can be increased. Another benefit is that some of the variables in a data set may have a high degree of multicollinearity. In this case, it is understood that both variables will provide the same amount of explanation for the output variable (target) in the model. The research by Yulianti and Saifudin suggests employing feature selection to identify significant features that might increase the efficacy of customer churn forecasting models. The dataset utilized for customer churn estimations contains numerous characteristics; however, some of them may be duplicated or unnecessary, resulting in a drop in prediction accuracy (2020). Therefore, deleting one of them means getting rid of an unnecessary variable while increasing the success of the model. In addition, performance savings are achieved. Al-Tashi et al., (2020) claim that Diminished overfitting results in less duplicated data, resulting in lower chances for noise-based decision-making. The fundamental advantages of feature selection include greater accuracy and fewer misleading data points. Additional advantages of feature selection include increased modeling precision, reduced training time, fewer data points, lower algorithm complexity, and faster algorithm training. Wrapper methods, which is one of the popular ways to select best features, create many models with different subsets selected from the input variables in the data set. Thus, it determines the best performing model together with the variables in the selected subset. For example, assuming that there are X inputs (independent) and one output (dependent) variable, combinations of variable sets that may be different for X inputs are created using one of the spiral methods. And the model is run for each combination of variable sets. According to Cai et al., (2018) The accuracy rate of classification model prediction is used as the feature assessment benchmark in wrapper models. Because the learning technique is involved in feature selection, the outcome of feature selection is frequently produced concurrently with that of the learning model. When compared

to filtering approaches, the wrapper model achieves greater classification accuracy. It has a reduced subset size; nevertheless, it has poor generalization capacity and a high level of time complexity. As expected, wrapper methods are disadvantageous compared to the simple filtering methods in terms of time and cost. Wrapper methods try to capture the most appropriate data pattern by trial and error by including or excluding variables within the model. There are three types of wrapper methods used.

1. Forward Selection: The process is started by selecting the variable that best contributes to the performance of the model with the forward selection. After the first selected variable, the second variable is also selected with the same method. And the cycle continues until the limit of the determined performance measure.

2. Backward Selection: The backward selection method proceeds in the opposite way to the forward selection method, where the aim is to eliminate the most negatively affecting the model among all variables and reach the best ones. In this method, in the first step, the variable with the highest P value is removed from the model data set, if it is above 0.05. This subtraction process continues until the P values in the entire data set are below 0.05.

3. Stepwise Selection: Stepwise selection has a structure in which each variable is evaluated together with other variables. Therefore, we can say that this method is a synthesis of both forward and backward selection methods.

Another alternative for the feature selection is leave one feature out (LOFO) method, which includes modeling in its structure, works integrated with LightGBM by default. Cross-validation is applied, the number of which is determined by the user, and effective variable selection is aimed with sensitive models. Evaluating the variables in terms of categorical and non-categorical variables, LOFO passes the variables through the correlation matrix. Scored features are evaluated according to whether they are below or above the specified threshold. This results in considering, altering, or choosing the most crucial variables rather than those that might have a detrimental effect on the model's performance or ability to forecast the future. Because fewer resources were required to train the model, there were advantages in terms of

reduced over-fitting, increased accuracy, and improved performance (Díaz Méndez, 2021).

In a recent study, all the features were trained on a simple neural network, and an ensemble of features was generated. Using second-order cone programming, redundant features are eliminated, the remaining features are trained on the same neural network, and the performance of the pruned ensemble is observed (Guldogus et al., 2022); this technique is called feature selection with second-order cone programming (FSOCP). This is accomplished by utilizing the outputs of each neural network trained on different datasets and single characteristics to produce a probability class distribution and a prediction. Using neural networks as embedded classifiers, this technology enables the application of SOCP to identify the attributes that are most important for accurate categorization.

2.4.5 Model creation. Machine learning is based on the idea of enabling machines to learn these processes while performing certain operations in people's daily lives. We can talk about many of the methods of machine learning used in the algorithms we're researching. Machine learning, popularly known as "ML," is a category of artificial intelligence (AI) that comprises readily available computers with the ability to be learned without even being exhaustively coded. ML is interested in computer program extensions that can modify themselves when exposed to fresh data. Reinforcement learning, supervised learning, and unsupervised learning are three types of ML algorithms. (Praveena & Jaiganesh, 2017). The two most well-known and preferred methods are supervised learning and unsupervised learning.

Based on labeled instances containing the correct answer, supervised learning discovers connections between input-output pairs. This form of machine learning is used in commercial processes today, but it necessitates high-quality data and a greater standard of technological infrastructure (Vogelsang & Borg, 2019). Supervised learning is a machine learning technique that generates a function over training data. In other words, this learning technique produces a function that maps between inputs (labeled data) and desired outputs. Training data consists of both inputs and outputs.

The function can be determined by classification or regression algorithms. The worst part of supervised learning is generating training data. A function is generated with a machine learning method (there are hundreds of methods) thanks to the training data. With this function, new incoming data is tried to be estimated. The most time-consuming part of supervised learning is the preparation of this training data. Supervised learning is divided into two. These are regression and classification. While the aim of regression is to predict the value of the target variable, the main purpose of classification is to predict the class label of the observation and to determine the class in which the observation will take place in accordance with the predetermined class labels. Classification can be between two options or more than two options.

On the other hand, in unsupervised learning, the outputs are not found in the study. According to Vogelsang and Borg, without labels, unsupervised learning may identify patterns in data. A typical use for this kind of ML is clustering. Reinforcement learning, which was motivated by the reward-penalty idea, trains the agent using a reward signal in a defined environment with a specific action set (2019). The observed units are grouped together according to their similar characteristics. Unsupervised learning uses machine learning algorithms to analyze and cluster unlabeled datasets. These algorithms classify data without the need for human intervention. Unsupervised learning models are used for three main tasks: clustering, association, and dimension reduction. Clustering is a data mining technique used to group unlabeled data based on their similarities or differences. For example, K-means clustering algorithms assign similar data points to groups, where K represents the size and granularity of the grouping. Association is another unsupervised learning method that uses different rules to find relationships between variables in each dataset. These methods are frequently used in market basket analysis and recommendation engines to generate "customers who bought this product also bought" recommendations. Dimension reduction is a learning technique used when the number of features (or dimensions) in each dataset is too high. It reduces the number of data entries to a manageable size while maintaining data integrity.

2.4.5.1 Classification and regression trees. The representation for the CART model is a binary tree. This is a binary tree derived from algorithms and data structures. Each root node represents a single input variable (x) and a split point on that variable

(assuming the variable is numeric). The leaf nodes of the tree contain an output variable (y), which is used to make a prediction. Tree-based learning algorithms are the most widely used supervised learning algorithms. They can be used to solve all the problems (classification and regression) considered. Grange et al. (2019) state that by dividing data into two homologous categories, the binary recursive classification technique used by decision trees produces nodes that are "highly purified." The procedure is recursive, so splitting is carried out again until node homogeneity is reached. A tree is the collective name for the complete network of splits, often known as nodes or branches separately. If the trees are permitted to reach their maximum depth, the recursive process will always categorize incoming data properly. Methods such as decision trees, random forests, and gradient boosting are widely used in all kinds of data science problems. They are easy to understand and interpret. The tree structures used can be visualized. It requires little data preparation. The cost of the tree used is logarithmic to the number of data points used to train the tree. It can handle both numeric and categorical data. They can handle multi-output problems. It is possible to validate a model using statistical tests. In a work of Borg (2021) it is claimed that they are extremely basic in nature and serve as a sequential depiction of every option that might be made and its accompanying goal. The root node serves as the algorithm's starting point and encourages the formation of further branches, each of which has a corresponding node. Each node will be associated with a certain decision, and based on the decision made, multiple paths are taken that result in other nodes. The leaf node, often referred to as the terminating node, will be at the base of the tree and will provide the ultimate expected result. Simple tree methods developed over time and diversified with bagging and boosting methods. By employing several trees that are formed in parallel structures for training, bagged decision trees enhance the effectiveness of decision trees. Due to the weak learning capabilities of a single decision tree and its extreme sensitivity to training patterns, it can cause the model to be overfitted by overfitting some patterns in the dataset. The prediction against fresh observations is generated from the overall detection outputs of all the decision trees (Javed et al., 2021). Here, an ensemble of decision trees is trained from bootstrap samples of the training set. Bootstrap sampling means drawing random samples from a training set with replacement. However, boosting refers to this general problem of producing a very accurate prediction rule by combining rough and moderately

inaccurate rules-of-thumb. Vafeiadis et al., (2015) point out that boosting's primary objective is to enhance classification efficiency by combining judgments from various classification models. Customer attrition in retail has been accurately predicted using boosting. The weak learners in boosting are decision trees with a single split. Boosting works by weighting the observations, putting more weight on difficult to classify instances and less on those already handled well. New weak learners are added sequentially that focus their training on the more difficult patterns. A collection of weak learners is being made into a stronger learning group. It is proposed that a weak algorithm will outperform random guessing. Any algorithm that can produce at least marginally above-random solutions is thus considered a weak learner (Sabbeh, 2018).

2.4.5.1.1 Random forests. According to the Borg (2021), decision trees provide several benefits, including being simple to comprehend and represent. However, decision trees frequently exhibit a strong coupling to the data, leading to overfitting and producing wildly divergent results on various test datasets. Other techniques were created, like random forest, to deal with the variability problem seen in decision tree ensemble approaches. An ensemble model called RF has been shown to be effective in predicting churn (Rijnen et al., 2018). RF is a bagging algorithm. It draws random bootstrap samples from the training set. However, in addition to the bootstrap samples, also drawn were random subsets of features for training the individual trees; in bagging, each tree was provided with the full set of features. Due to the random feature selection, the trees are more independent of each other compared to regular bagging, which often results in better predictive performance (due to better variance-bias trade-offs), and it's also faster than bagging because each tree learns only from a subset of features. Lalwani et al., (2022) states that the random forest approach creates many decision trees, and each one is trained by picking any random subset of features from the predictor data set. Depending on the characteristics or factors provided, each tree develops to the fullest degree possible. For the estimate, the overall decision tree is created mostly using weighted averages. Vast numbers of input parameters may be handled by it without any being deleted. In the data set used to train the predictive model, it can also deal with missing values. The term "random decision forest" was first proposed in 1995 by Tin Kam Ho. He developed a formula to use random data to create predictions. Then in 2001, Leo Breiman extended the algorithm and created

random forests as we know them today. Also, Adele Cutler has also put forward important studies in this field (Breiman & Cutler, 2005). This means this technology and the math and science behind it are still relatively new. A "random forest" is a group of decision trees. However, there are some differences between the two. A decision tree tends to create rules, which it uses to make decisions. A random forest will randomly choose features and make observations, build a forest of decision trees, and then average out the results. Borg (2021) states that the random selection of training data samples during the building of decision trees and the random selection of a portion of the whole feature vector during node splits are the two main causes of the term "random" in random forest. Each decision tree will vote for a class, and the final class will be determined by majority voting. Bagging, otherwise known as bootstrapping aggregation, allows individual decision trees to randomly sample from the dataset and replace data, creating very different outcomes in individual trees. This means that instead of including all the available data, each tree takes only some of the data. These individual trees then make decisions based on the data they have and predict outcomes based only on these data points. That means that in each random forest, there are trees that are trained on different data and have used different features to make decisions. This provides a buffer for the trees, protecting them from errors and incorrect predictions. The process of bagging only uses about two-thirds of the data, so the remaining third can be used as a test set. According to Jaiswal & Samikannu, (2017) class variables, also referred to as decision variables or predictor variables, can be used for pattern recognition, such as handwritten digits, a yes-or-no choice to predict a cancer diagnosis or loan approval, junk mail or not in the case of emails, even excellence (bad, normal, good) for the product attributes, and more.

2.4.5.1.2 Gradient boosting. A bagging ensemble algorithm like Random Forest is one example but boosting ensemble learning is yet another. Instead of independent model training, sequential model training is used in boosting. A variation of boosting ensemble learners known as "gradient boosting decision trees" (GBDT) has been mentioned in several works on churn prediction (Borg, 2021). Gradient boosting is a machine learning technique for regression and classification problems. This creates a model that typically consists of decision trees by combining weak predictive models. The purpose of it, as with any supervised learning algorithm, is to identify and

minimize a loss function. According to Raeisi & Sajedi, (2020) There are three components to the machine learning technique known as "gradient boosting," which is used for regression and classification. These include a weak learner, a loss function, and appending weak learners. The error function must be differentiable to optimize it. The fact that this technique can be applied to any loss function without the need for a new boosting algorithm is a benefit. Forecasts are made using the weaker learners. Decision trees are employed in gradient boosting techniques for weak learners. An appending weak learner is included in the model as a step of the gradient descent technique after the gradient descent method has been applied to minimize the cost function.

The intuition behind the gradient boosting algorithm is to repeatedly use patterns in residuals to strengthen and improve a model with poor predictions. Once it reaches a stage where the residuals have no patterns that can be modeled, GBM can stop modeling the residuals (otherwise, it can lead to overfitting). Algorithmically, the loss function is minimized so that the test loss is minimal.

2.4.5.1.3 XGBoost. XGBoost is a learning algorithm based on decision trees and using gradient boosting. Recently, it has dominated both tournaments for structural or tabular data and analytical projects conducted by CRM departments. Gradient-boosted decision trees are implemented using XGBoost, a fast and efficient method. However, the term "XGBoost" in fact stands for the engineering objective of pushing the computational resource ceiling for tree-based boosted algorithms. Therefore, XGBoost is used by numerous researchers and tech companies. The algorithm's implementation was designed to use memory and computing resources efficiently. (Celik, & Osmanoglu, 2019). If unstructured data, such as sounds, texts, or pictures, is used in model training, deep learning with artificial neural networks is preferred, but otherwise, one of the most accurate alternatives is XGBoost. The logic behind is like GBM. Guliyev & Tatoglu also mention that the decision tree approach with gradient boosting employed by XGBoost utilizes fresh models to calculate the residuals of previously used models, which are then combined to get the final prediction. (2021). The advantage of XGBoost over GBM is both system (hardware and software) optimization and algorithmic enhancements. There are many reasons for choosing this algorithm. In order to train trees in parallel, XGBoost utilizes a sparsity-aware

algorithm. Real-world data sets frequently involve sparse datasets, and in certain circumstances, feature engineering techniques like one-hot encoding are used to intentionally create this sparsity. One-hot encoding boosts the number of dimensions, which entails longer training durations for the standard GB model (Borg, 2021).

$$E(M|p,q,\gamma,m_x,x) = \frac{(\gamma + m_x x)p}{px+q-1} = \left(\frac{q-1}{px+q-1}\right)\frac{\gamma p}{q-1} + \left(\frac{px}{px+q-1}\right)m_x \quad (0.4)$$

$$\Omega(h) = \gamma T + \frac{1}{2}\lambda||w||^2, \quad (0.5)$$

2.4.5.1.4 LGBM. Data is growing day by day, and the need for new methods that are faster, focused on accuracy, and produce high performance results has increased. The LGB algorithm was developed by Microsoft for these purposes. The LGB algorithm is a decision tree-based ensemble learning algorithm. Most decision tree algorithms use level-wise division. Given that it takes a lot of work to train many decision trees using large datasets, the GB approach is not adaptable. Therefore, the Light Gradient Boosting Machine, which employs the unique approaches of Exclusive Feature Bundling and Gradient-based One-Side Sampling, was developed with great effort by a Microsoft team. (Borg, 2021). LGBM uses leaf-wise splitting as follows: Leaf-wise finds and selects the leaves that will reduce loss the most, according to level-wise. It just splits that leaf. It uses depth-first search (DFS) instead of breadth-first

search (BFS). BFS visits the nearest neighbors. DFS, on the other hand, aims to go as far as it can go, then return and wander around the graph.

2.4.5.1.5 CatBoost. In addition to increasing the performance of Gradient Boosting, it is an open-source machine learning algorithm based on Gradient Boosting developed by Yandex as an alternative to XGBoost and LightGBM. Although CatBoost is a GB decision tree technique as well, it makes use of symmetric trees to shorten prediction times. It changes the base model to deliver better results after computing the residuals that are named "pseudo." Lalwani et al. state that CatBoost's main improvement is that it incorporates some of the most highly regarded pre-processing techniques, such as one-hot encoding, label encoding, etc., which reduces the pre-processing work. It does not, however, provide all statistical tools for data pre-processing (2022).

2.4.6 Data splits - cross validation. Before modeling or estimating, the data set must be split into a training and a test set. Thus, it is possible to train the model on the training set and make predictions on the test set. The train set is the dataset on which the model is trained. A test set is a data set used to evaluate the model developed on a training set. The larger our training set, the better the model will learn. As the test set grows, the success of the evaluation metrics will increase, and tighter confidence intervals will emerge.

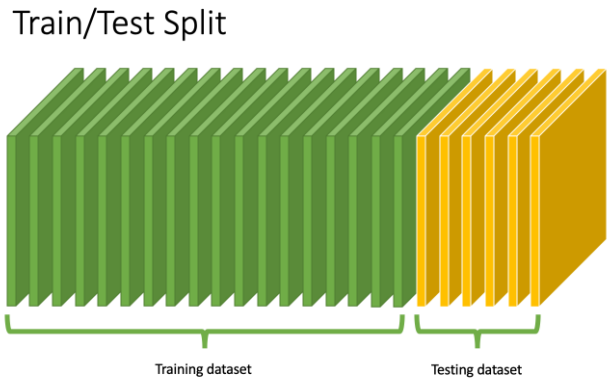

Figure 1. Depicts how the logic of the train test is split in the cross validation technique.

In the cross-validation step, it is aimed to train the model by touching all points in the data set by using the cross-validation method. Vogelsang & Borg mention that it is essential to describe activities that guarantee that training data genuinely correlates to real data since an ML system's behavior depends on the data it was trained on. Because real-world data properties can change over time, requirements validation must be done continuously while the system is in use (2019). Cross-validation is a statistical re-sampling method used to evaluate the performance of a machine learning model objectively and accurately on data it has not seen. Díaz Méndez, states that cross-validation approaches include the K-fold, stratified cross-validation, and leave-one-out cross-validation. In the first step, k samples are created from the given dataset; each k acts as a training data, and the k - 1 samples represent as model validation samples. (2021). Following this repeated procedure, the estimate error is calculated and averaged to yield the model's estimation error. The stratified cross-validation method is an extension of the k-fold cross-validation method. The basic goal is to try to balance the classes when faced with an unbalanced situation, which implies one of the classes of dependent characteristics is notoriously larger than the other (Bharat Rao & Fung, 2008). The cross-validation method is one of the most used. The model will be more generalizable with this model because both test and train operations are performed on all data. As a result, the trained model's success on new future raw data will improve. Cross validation necessitates more computing power as it uses multiple train and test splitting operations. Fagerholm points out that because a more complicated model may readily overfit the learning data and reduce the training error to zero, training success normally rises with the model's complexity. As a result, different methodologies must be employed in order to obtain a valid approximation of the real out-of-sample outcomes. (2022). Also, as expressed in the study of Fagerholm by dividing the available data into multiple data sets for training and testing, cross-validation may be used to assess the model's performance on previously unexplored data (2022).

2.4.7 Model evaluation for classification tasks. Confusion Matrix: A confusion matrix is a table often used to describe the performance of a classification model on a set of test data for which the actual values are known.

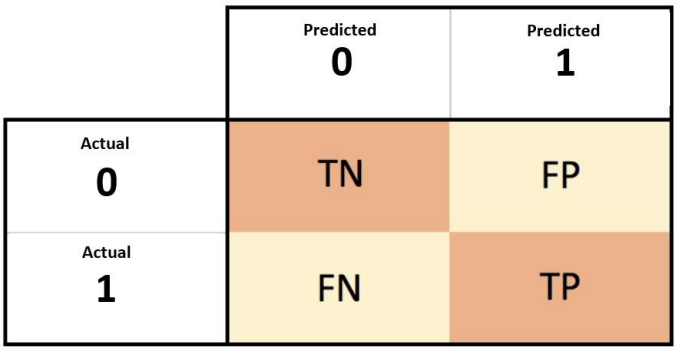

Figure 2. A confusion matrix-style table depicting the relationship between predicted and actual values.

Accuracy, Precision, Recall and F1 Score: The accuracy value is calculated by the ratio of correctly predicted areas in the model to the total data set. It is generally a measure of how often the classifier guesses correctly. (TP + TN) / Total. Accuracy is a metric that is widely used to measure the success of a model but does not appear to be sufficient on its own. Therefore, other metrics such as recall, precision, and F1 score should also be considered and interpreted according to the distribution of the target variable. Precision shows how many of the positively predicted values are actually positive. The precision value is especially important when the cost of false-positive estimation is high. Precision is a measure of how accurately all classes predict. It should be as high as possible. also known as "positive predictive value". TP / (TP + FP). Recall, on the other hand, is a metric that shows how many of the transactions that need to be predicted positively are predicted positively. The sensitivity value is also a helpful metric when the cost of estimating something as "false negative" is high. The F1 Score value shows the harmonic average of the Precision and Recall values. The reason why it is a harmonic mean instead of a simple mean is that extreme cases should not be ignored. It is a measure of how well the classifier is performing and is often used to compare classifiers. The main reason for using the F1 Score value instead of accuracy is to avoid making an incorrect model selection in unevenly distributed

data sets. In addition, the F1 Score is of critical importance, as there is a need for a measurement metric that will include all error costs instead of just false negatives or false positives.

ROC AUC: Ullah et al. mention that the average performance versus all feasible cost ratios between FP and FN is represented by the ROC area. The prediction is accurate if the ROC area value is greater than 0.9. Similar to this, the intervals 0.5–0.6, 06–0.7, 0.7–0.8, and 0.8–0.9 stand for poor, moderate, good, and superior random prediction, respectively. Other ROC region values suggest a problem in the predicted values (2019). AUC-ROC measures discrimination performance, defined as the ability to separate ones from zeros (Sofaer et al., 2019). The area under the ROC curve is called the AUC (Area Under the ROC Curve). It shows how much it can separate the positive class from the negative class. The larger the area, the greater the discrimination ability. On the other hand, the ROC curve is basically the curve that shows the values that FPR will take in case of an increase in TPR, that is, in case of 1' convergence. Here is a situation where we want the TPR to converge to 1, but we want the FPR to remain low when this TPR converges to 1. This is a chart used to summarize the classifier's performance over all possible values. It is generated by plotting the ratio of true positive values (x-axis) versus the ratio of false positive values (specificity) (y-axis) when you change the threshold for assigning observations to a particular class. The ROC curve is used to generate a sensitivity / specificity report. The Area Under Curve (AUC) is a measure of how well a parameter can be distinguished between two classes. The area under the ROC curve, or AUC, is the most significant statistic related to ROC curves. Given that the curve is within the unit square, AUC = 0 if the classifier rates all positives higher than all negatives, and AUC = 0 if the classifier rates all negative higher than all positives. The anticipation that a randomly drawn positive would obtain a higher ranking than a randomly drawn negative is the extremely helpful statistical interpretation of AUC, which is the normalized version of the Wilcoxon-Mann-Whitney test (Flach, 2016).

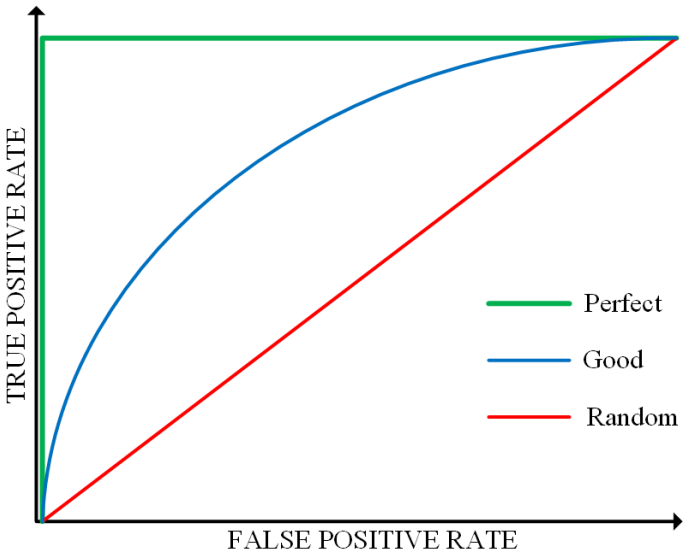

Figure 3. Graph showing the change of different ROC curves in accordance with their success

2.4.8 Hyper parameter optimization. Model creation with default hyperparameters makes direct predictions in various algorithms without any operations, so that it can be seen how much error has decreased after the operations have been done. A variety of options must be investigated to create the best ML model. "Hyper-parameter tuning" is the process of creating the perfect model architecture with the best possible hyper-parameter configuration (Yang & Shami, 2020). Machine learning models are basically based on mathematical functions that represent the relationship between dependent and independent variables (target features). To illustrate, linear regression assumes a linear relationship between dependent and independent variables, which can be expressed with the $y = mx + b$ function. During the training of the linear regression model, the most appropriate m and b values are determined by using various optimization techniques. These values are the model parameters. Unlike parameters, hyperparameters are not learned during model training. Hyperparameters are different from the internal model parameters, such as the neural network's weights, which can be learned from the data during the model training phase. Before the training phase, we would like to find a set of hyperparameter values which achieve the best performance on the data in a reasonable amount of time. They are determined by the data scientist before the modeling phase. For example, the KNN

algorithm, which is one of the non-parametric classification algorithms, makes classifications by looking at the k nearest neighbors to the value to be estimated. Here, the number k (n_neighbors:) and the distance metric (metric:) to be used are hyperparameters that should be specified by the data scientist before modeling and increase the performance of the model.

Hyperparameter optimization is the process of finding the most appropriate combination of hyperparameters according to the success metric determined for a machine learning algorithm. A balance between overfitting and underfitting can be achieved by combining model complexity equilibration with hyperparameter optimization. The overfitting problem, which arises from the flexibility of the model, can be solved with the constraints imposed by the hyper-parameters. According to Yu and Zhu, grid search is the fundamental optimization technique. The user-specified list of hyper-parameters is subjected to an exhaustive search. Because these factors create all candidates, data scientists must have a fundamental comprehension of them (2020). In the case of a small search space, grid search is appropriate for numerous hyper-parameters.

Chapter 3

Methodology

3.1 Research Design

Among all customers, the corporate segment with the highest profitability was selected. Within this segment, there are macro, micro, and commercial customer groups. Within this customer segment, those whose customer age is older than 1 year and those who have shopped through the mobile application at least once were preferred. For this, the FXTransactions_tbl table, where each transaction is logged, the Order_tbl table with the information of the orders given by the customers via the mobile application, the Customer table with the information of the customers, and the FEC_tbl table with the currency pairs and current exchange rates are used. RFM and CLV analyses were performed with the data in the transaction table. For the machine learning approach, temporal columns were obtained from the Order_tbl and FXTransactions_tbl tables and merged into the Customer table. In addition, the FECTABLE table was used to standardize transactions by converting them to daily exchange rates. Only Numpy and Pandas libraries are sufficient for RFM calculations. BG/NBD and Gamma-Gamma models were used for CLV calculation. For the machine learning approach, RF and CART from the Sklearn library, GBM models, as well as XGBoost, LightGBM, and CatBoost libraries, were used. Li and Zhou, (2020) state that the corporation may devise a hierarchical customer retention plan that is critical to the retention and value improvement of the company's existing customers and is based on a comprehensive analysis of behavioral data on the churn of customers and the output of each client's churn likelihood and churn attributes. Hierarchical customer retention solutions can dramatically lower customer retention costs and increase customer retention effectiveness when compared to typical customer retention tactics.

3.2 Setting and Participants/Target Population and Participants

The foreign exchange buying and selling data of legal customers over the mobile application for 2 years was kept in a temp table. Then, RFM and CLV calculations were made for the general behavioral analysis of these customers. In this way, customers were segmented, and the most profitable customers were determined. Customers who are likely to shop in the future were also determined by probabilistic calculations. For the Churn classification prediction, another data set was created by considering the common data of all customers. This dataset is labeled to contain the target variable. Accordingly, customers who did not shop for 147 days were expressed as churn. The created data set consists of 6844 rows, and each row corresponds to a customer. The churn column, which is the dependent variable of the data set, is a binary column consisting of data expressing the churn and non-churn states, which correspond to the values 1 and 0. For the estimation of this column, firstly, data preprocessing steps are defined. The final data set prepared was divided into 80:20 ratios for training and testing. Train data is used to train the model. The test data was used to measure the success of the trained model. Many artificial intelligence models that are popular today have been used. According to the preliminary results, it is striking that tree-based methods make for much more successful estimation. Among the tree-based methods, boosting methods come to the fore. For this reason, finally, hyperparameter tuning was performed using tree-based methods, and the most suitable parameters were determined using different search methods.

3.3 Procedures

3.3.1 Data collection instruments. Database operations with Python have an important position in programming. Python supports an MSSQL connection to store data in the database and use it when needed. Very basic SELECT, INSERT, UPDATE, and DELETE commands can be implemented easily. For example, many libraries have been prepared and made available to developers as open source to easily perform database operations with MSSQL. Access to ready databases is provided with a username and password via SQL server authentication. In the database, the stored

procedure was run, and the result produced was written to a temporary table, and the data in this table was transferred to the Python environment.

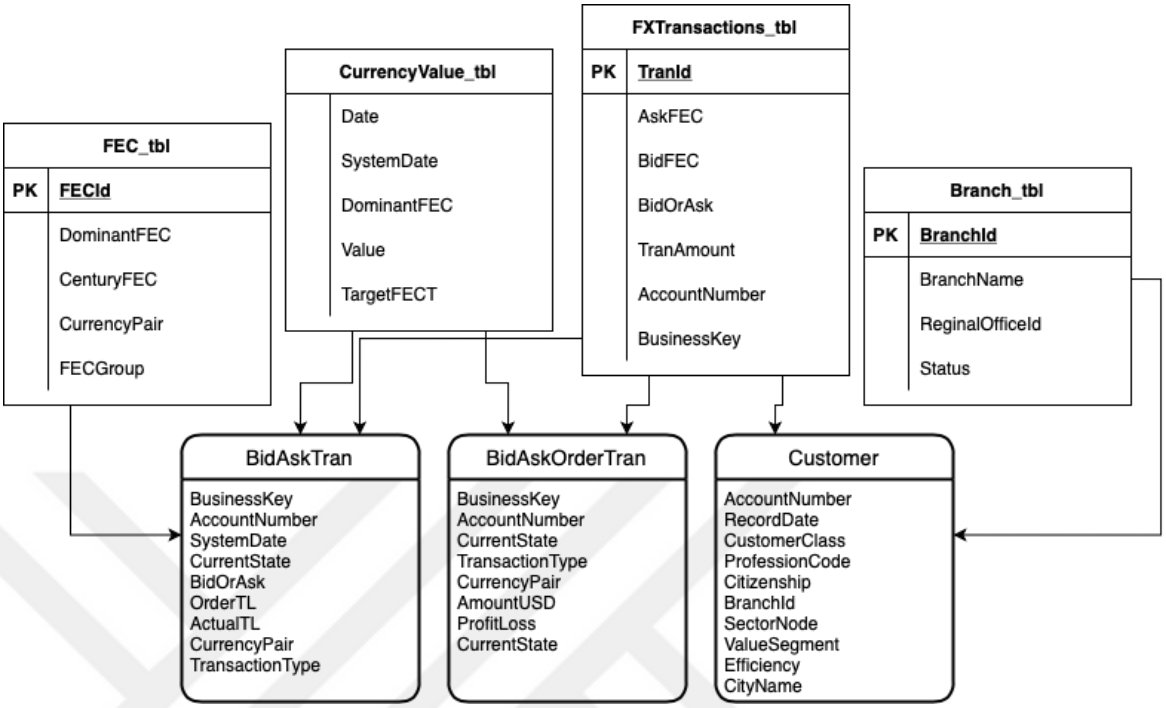

Figure 4. A representation of a relational database that forms summary tables.

3.3.2 Data collection procedures. The data of corporate customers aged more than 1 year for the years 2020 and 2021 of the mobile foreign exchange trading application, which is the digital channel of a bank operating in Turkey, has been brought together in 3 different tables. In the first of these, the transactional data of the customers in question was used. There are approximately 130,000 transactions. In addition to the information on which customer made the transaction, on which date, and when, the volume of the transaction, between which foreign currency types it took place, and the profit earned by the bank from this transaction are also included. In the other dataset, there are orders given by the same customers to perform transactions in this application in the same period. Similarly, it has a transactional structure. Roughly, there is information about who gave the order, when it was given, the size of the transaction, and whether the order was executed or not. The third and final dataset

includes the demographic and behavioral data of the customer. Information such as age, sector, productivity category, and segment can be given as examples.

3.3.3 Dataset descriptions. The data obtained from the SQL database was brought together in three different summary tables. These are the "BidAskTran," "BidAskOrderTran," and "Customer" tables. The customer ID field is present in all three tables, and the relationship between them will be analyzed using the ID column. Customers in the “Customer” table are one-of-a-kind.

3.3.3.1 Transactional bid or ask data. The dataset contains 208249 bid or ask transactions made by 6844 users. The corporate clients that conducted the transactions in the data set were engaged in various business activities. Commercial, macro, and micro are the three categories into which customers are split. The data set has eight columns: account number, business key, system date, private branch cost currency, transaction type, currency pair, transaction amount in USD, and profit/loss.

Table 1

Transactional Bid or Ask Dataset Variable Description

Variable	Description	Data Type
Account Number	The unique customer numbers.	Integer
Business Key	The unique number of transactions	Integer
System Date	The date the transaction took place.	Date
Private Branch Cost Currency	Flag that indicates whether there is a special cost rate defined for the customer.	Bool
Transaction Type	Flag that indicates whether the transaction is a foreign currency or precious metal transaction.	Bool
Currency Pair	The currency pair in which the transaction takes place.	Text
Amount (USD)	The amount of the transaction in USD.	Integer
Profit Loss	Profit and Loss According to the Rate Calculated in the Market Average TL	Integer

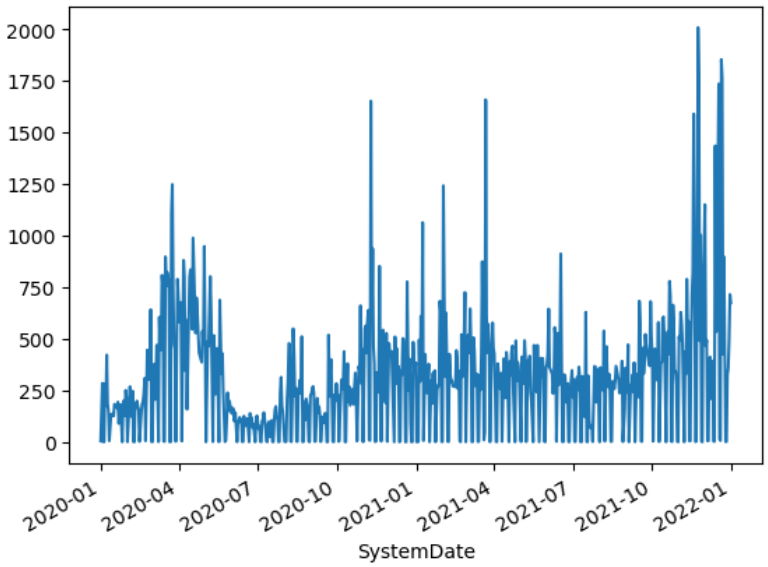

Figure 5. Bar graph showing sum of amount (USD) performed by day

3.3.3.2 Transactional bid or ask order data. In this data collection, there are 92876 bids or asks for orders submitted by 3291 consumers using mobile applications. If a customer places an order through the mobile application, it indicates that the customer is keeping up with the market. In this regard, reviewing consumer purchase and sell orders via the mobile application will contribute positively to the model's success. Account Number, Bid Ask Order Id, System Date, Current State, Bid or Ask, Order in TL, Actual TL, and Currency Pair are the eight attributes in total. The extracted features were added to the final data set after this data set was used under the feature extraction title.

Table 2

Transactional Bid or Ask Dataset Variable Description

Variable	Description	Data Type
Account Number	The unique customer numbers.	Integer
BidAsk Order Id	The unique number of transactions	Integer
System Date	The date the transaction took place.	Date
Current State	The flag that indicates whether the given order has been fulfilled or not.	Bool
Bid or Ask	The flag that indicates whether the given order is buy or sell.	Bool
OrderTL	The amount of the order given in TL.	Integer
ActualTL	The amount of the performed order in TL.	Integer
CurrencyPair	The currency pair in which the transaction takes place.	Text

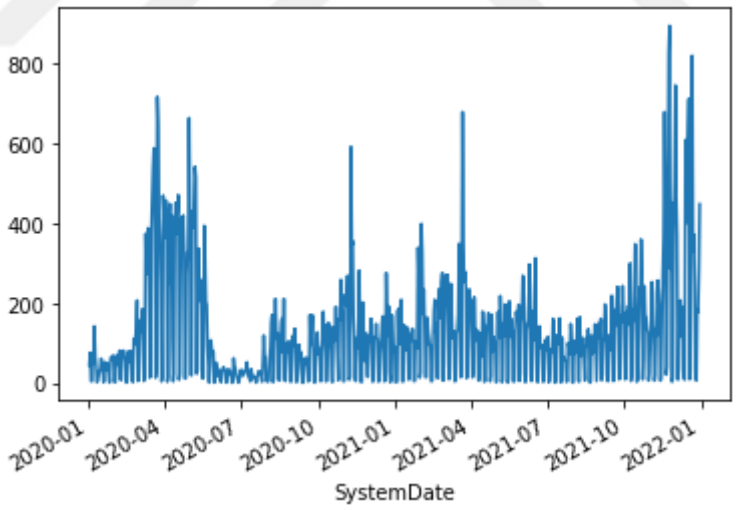

Figure 6. Bar graph showing count of order given by day

3.3.3.3 Customer data. The data set, which contains 6844 rows, includes behavioral, characteristic, and demographic variables for each client. In addition to the features such as time spent with a customer, province, customer segment, productivity, and treasury value, there are 28 more.

Table 3

Transactional Bid or Ask Dataset Variable Description

Variable	Description	Data Type
Account Number	The unique customer numbers.	Integer
Recording Time	The unique number of transactions.	Integer
System Date	The date the transaction took place.	Date
Customer Class	The flag that indicates whether the given order has been fulfilled or not.	Bool
Profession Code	The client's profession	Bool
Citizenship	Whether the customer is a domestic or foreign customer.	Integer
Branch Id	The identity number of the customer's branch of the bank.	Integer
City Name	The city where the customer lives.	Text
Customer Class	The class that the customer belongs to.	Text
Sector Node	The sector in which the customer carries out its commercial activities.	Text
Importer	The flag that indicates whether the customer is importing or not.	Bool
Exporter	The flag that indicates whether the customer exports or not.	Bool
Foreign Trader	The flag of the customer engaged in international trade.	Bool
Total FX Amount TL	The total FX trading volume made by the customer during the analysis date.	Integer
FX Tran Efficiency	The total FX trading efficiency that the customer has made within the analysis date.	Integer
Education Level	The education level of the customer.	Integer

Table 4 (cont'd)

Variable	Description	Data Type
Related Banks Count	The number of bank accounts other than the bank where the customer is	Integer
Real Estate Count	The number of real estates owned by the customer.	Integer
Vehicle Count	The number of vehicles owned by the customer.	Integer
Value Segment	The classification of the customer's valuation.	Integer
Is Private Customer	The flag that indicates whether there is a private customer or not.	Bool
Is Corporate Customer	The flag that indicates whether it is a corporate customer or not.	Bool
Is Commercial Customer	The flag that indicates whether there is a commercial customer or not.	Bool
Active Products Count	The number of products that the customer actively uses in the bank.	Integer
Open Products Count	The total number of products owned by the customer in the bank.	Integer
Open Treasury Products Count	The total number of treasury products used by the customer.	Integer
Active Treasury Products Count	The total number of active treasury products used by the customer.	Integer
Last Year Efficiency	The efficiency of the total transactions made by the customer during the analysis date.	Integer
Last Year Treasury Efficiency	The total treasury transaction efficiency of the customer during the analysis date.	Integer
Efficiency	The classification of efficiency according to the profit and loss status of the customer within the total period of time in the bank.	Integer

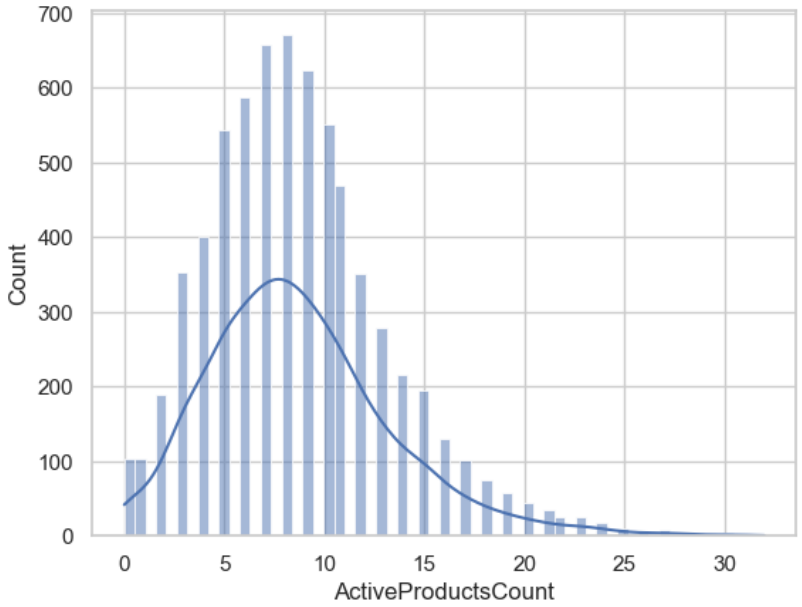

Figure 7. Bar graph showing the active products count by customer

It may be useful to examine the distribution of some data in the data set on customers. For example, active product counts have a normal distribution among customers. A significant portion of customers use between five and fifteen bank products.

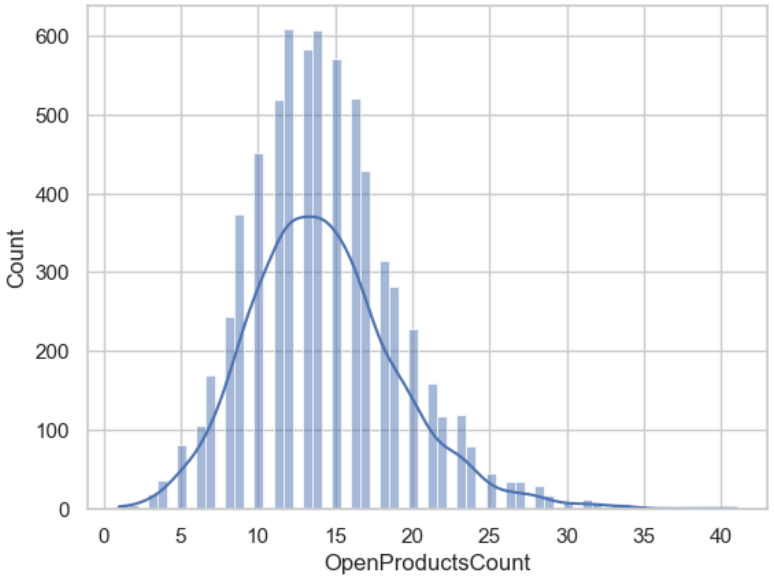

Figure 8. Bar graph showing the open products count by customer

Similarly, the open product count does not show an abnormal distribution. It is also striking that it has a more normal distribution than the active product count. In this respect, it can be said that it will benefit the success of the model.

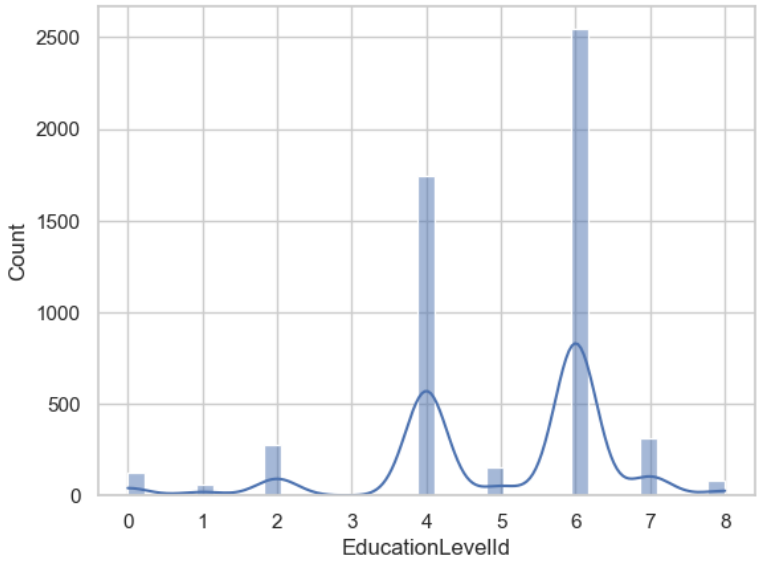

Figure 9. Bar graph showing the education level of customers

The clients' education level rises in tandem with their education level. From this perspective, it appears that clients usually have high levels of education.

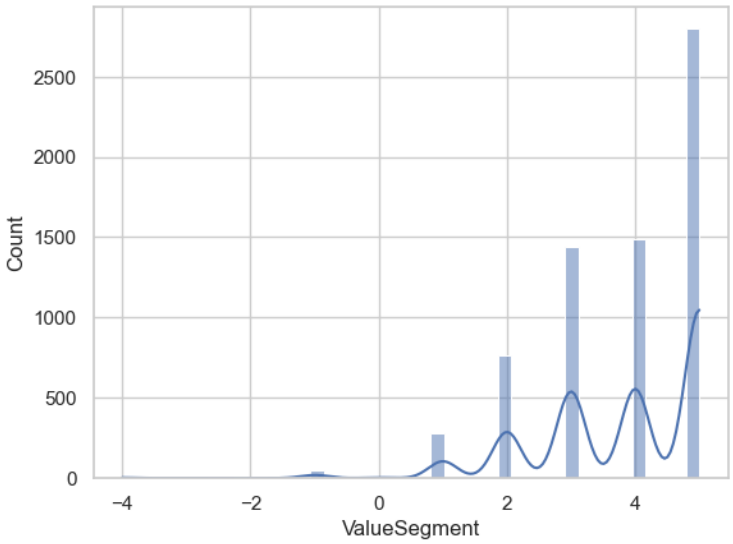

Figure 10. Bar graph showing the value segment of customers

Customers' value segment score is obtained automatically by the bank after reviewing the data including the full history of the customer's connection with the bank. It is not just dependent on currency bid-ask transactions by the bank. A right-skewed distribution was seen.

3.4 Limitations

Due to legal obligations and regulations to which the bank is subject, restrictions were encountered while determining the sample from the data set. Over a longer period, more reliable and consistent historical data could be obtained by using all the data accumulated from previous years. In addition, all customer contact points with the bank could be identified. Thus, for each customer, the frequency of usage of all services provided by the bank could be determined and measured.

3.5 RFM and CLV application

A loss reduction study begins with definitions such as active customers, inactive customers, and lost customers. The concept of loss is one that varies between sectors. The restaurant has probably lost a customer who has not been to a restaurant near the workplace for 3–4 weeks. If it is a store that specializes in winter sports, it should carry two more concepts in the second year besides "missing" or "lost," which are active and inactive customers. The customer is active first, shopping with the brand. If he no longer does the shopping that he used to do frequently, it can be said that he has become inactive. If this situation continues for a certain period, the customer is now a lost customer.

3.5.1 RFM modeling. With the RFM method, customers will be segmented according to their shopping frequency. For this, a transactional data set will be used. The RFM model will serve to classify customers at a very primitive level. After the transactional data is aggregated as follows, each customer will be segmented.

Table 5

Recency, Frequency and Monetary value of first five customer

Customer Id	Recency	Frequency	Monetary
1	2	28	151161.00
2	312	10	39119.60
3	2	122	187747.40
4	328	219	444011.60
5	81	83	2195888.80

Then, by using the CLV calculation, the frequency and recency values of these customers, as well as the monetary information, will be used to list the customers who will make possible shopping from the largest to the smallest.

3.5.2 CLV modeling with BG/NBD & Gamma-Gamma. To summarize, as stated in the theoretical part, when customers are evaluated on an annual basis, for example, the date and number of orders placed throughout the year, as well as how much the customer spends on these orders, are sufficient to calculate the customer lifetime value for each customer.

Table 6

Total Frequency, Recency, T and Monetary Values for Randomly Selected Customer

Customer Id	Frequency	Recency	T	Monetary
23	22.0	728.0	728.0	6830.0
12	8.0	390.0	700.0	4814.8
325	73.0	645.0	645.0	2501.0
12	59.0	311.0	637.0	7479.0
876	50.0	649.0	728.0	43008.0
3	31.0	462.0	658.0	9643.9

3.6 Machine Learning Models Application

For the classification models of the three different data sets used in this study to be trained correctly, it is necessary to convert them to the necessary format and perform the preprocessing steps. These steps are respectively for data preparation: feature creation, rare analysis, missing observation analysis, outlier observation analysis, variance analysis, feature scaling, correlation analysis, and feature analysis. After data preparation, the data set is divided into two: test and train. Models were trained with the cross-validation method. Then, the grid search method was used for hyperparameter optimization. After the best parameters were selected, the final models were created. The final success of the models was measured using the test data, and the predictions for the customers included in the test data were examined.

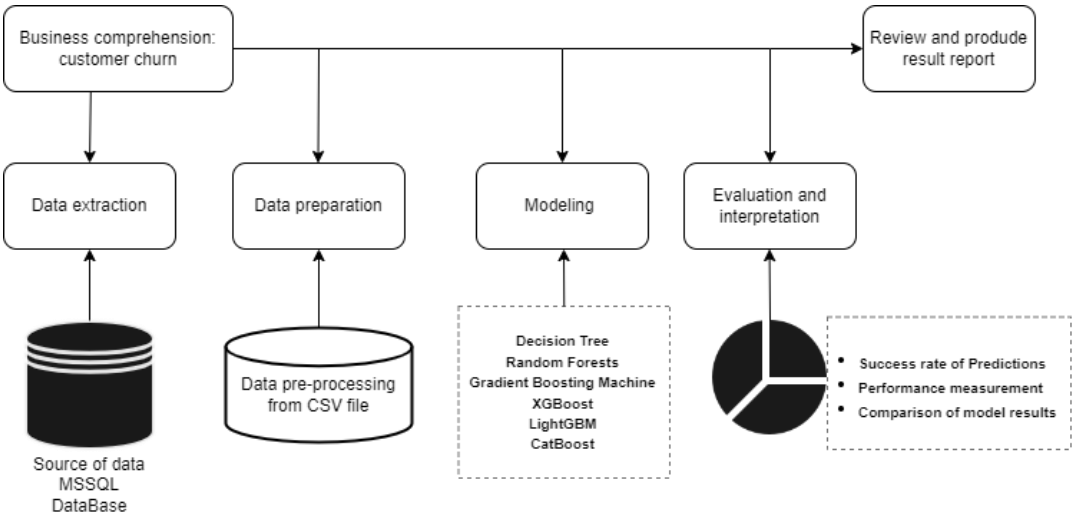

Figure 11. The diagram depicts all of the stages involved in the machine learning approach.

3.6.1 Data preparation. Each transaction of the customers used for analysis within our sample is included in the data set as a transaction. Unlike the so-called contractual or subscription-based situation in this data, there is no labeled target variable. For this reason, a label will be made, considering the behavior of the customers. As such, it can be subjected to classification algorithms.

3.6.1.1 Calculating churn thresholds. It is critical to determine the ideal churn time based on a statistical method. Generally, CRM teams in companies determine a period that they instinctively determine as a churn limit without considering any statistical method. For example, this period can be 3 months, 6 months, or 1 year. Without specifying any mathematical method, such periods are not applicable to the entire customer base, so they are unlikely to be useful. To determine the churn period, the difference between the recurring purchases of each customer is subtracted. Listed below are the days' difference between the first five purchases for a customer.

Table 7

The Days Difference Between the First Five Purchases for a Customer

Customer Id	System Date	Difference Between Last Shopping
1	2020-01-03	NaT
1	2020-01-04	11 Days
1	2020-01-14	1 Day
1	2020-01-15	1 Day
1	2020-02-03	18 Days

When this shopping habit is combined with the cumulative sum method for all customers, a graph like the one below is drawn. In this way, how most customers behave can be expressed in a very simple and informative way. A significant portion of the customers made repetitive purchases within the first 100 days after their last purchase.

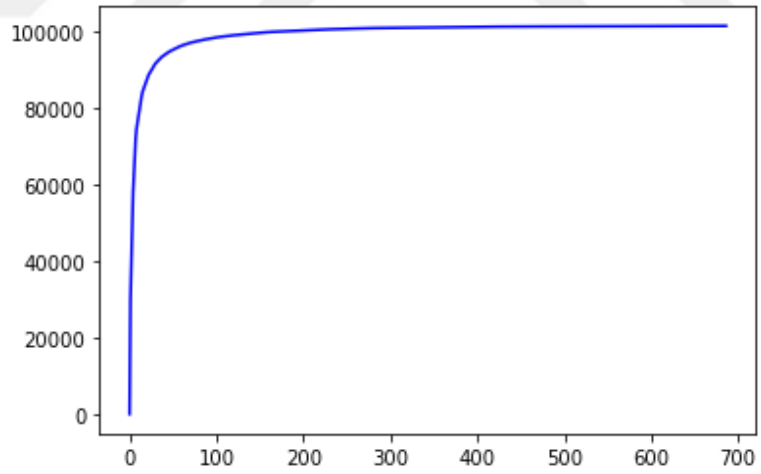

Figure 12. Line graph showing count of customers' repetitive purchases by day

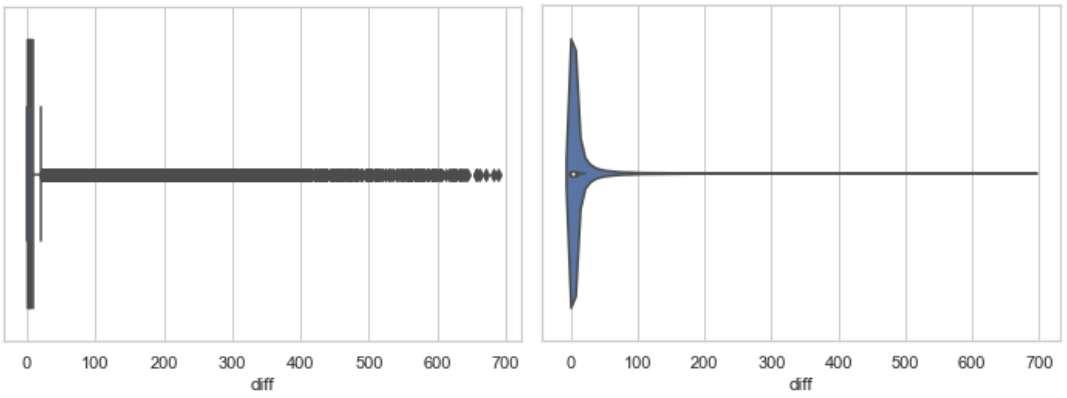

Figure 13. Box and violin plots showing the difference in days between repetitive purchases by customers.

According to the Pareto analysis, which is a statistically common theory, when a period is determined by considering the behavior of the top 80 percent of the customers, a customer must be labeled "churning"; the time to pass since the last shopping was calculated as 147 days, that is, 5 months on average. The time elapsed since each customer's last transaction is listed above (Figure 12). As stated before, these values also mean recency.

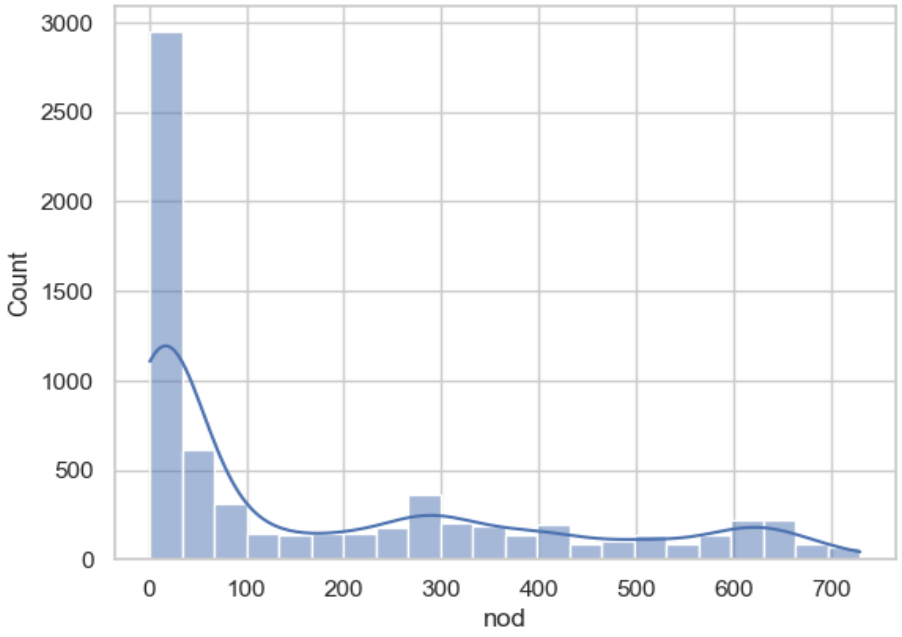

Figure 14. Bar graph showing number of days (nod) of customers since their last transaction.

Table 8

The Days Elapsed Since Each Customer's Last Transaction

Customer Id	System Date	Number of Day
1	2021-12-31	1 Day
2	2021-02-24	311 Days
3	2021-12-31	1 Day
4	2021-02-08	327 Days
5	2021-10-13	80 Days

Customers are labeled according to the determined churn period. An example of this is listed in the table below (Table 8).

Table 9

Churn Label of Customers According to the Determined Churn Period

Customer Id	System Date	Number of Days	Number of Months	Churn
1	2021-12-31	1	0	False
2	2021-02-24	311	10	True
3	2021-12-31	1	0	False
4	2021-02-08	327	11	True
5	2021-10-13	80	3	False

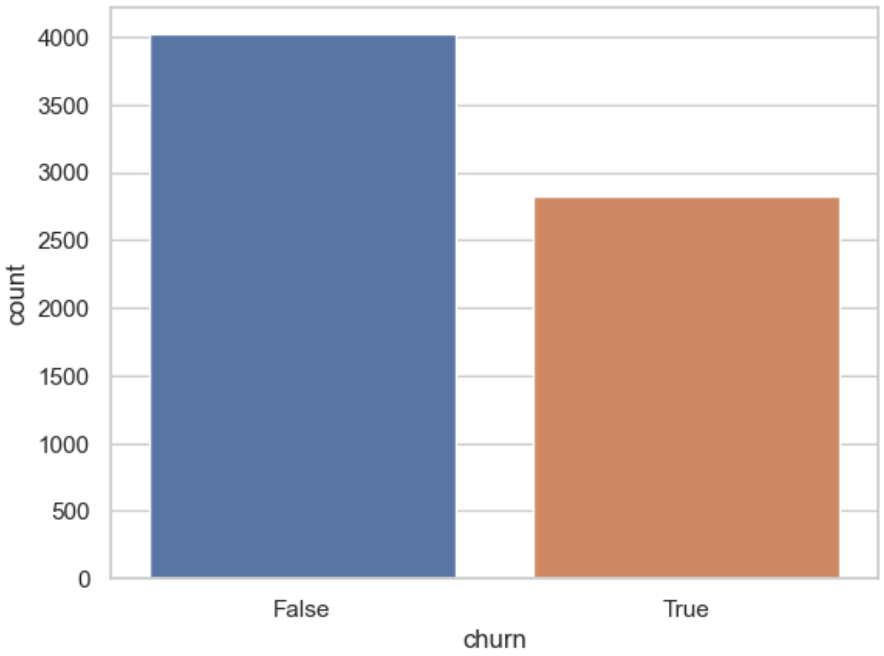

Figure 15. Bar graph depicting the distribution of customers based on churn or no churn labels.

3.6.1.2 Feature creation. To train a classification model, the data must be converted from a transactional format to a customer-based structure. It is important for the success of the model to obtain the temporal data from the transactional data and then add it as a feature to the final data frame. All these steps are detailed under the title "feature engineering."

3.6.1.2.1 Bid or ask order based features. In the "Bid or Ask Order" data frame, there are foreign exchange buying and selling orders of the customers via the mobile application. In addition to buying and selling foreign currency, a customer's ability to give orders by following the market situation and the sectoral development in the coming days is an informative feature of the customer's character. For this reason, the number of orders given by the customer and the number of their executions have been

added to the data set as a new feature. In addition, the total dollar value of these orders was added to the data set as two separate columns.

In the table below, the comparison of both the count and amount of the orders executed and the canceled order for each customer is listed.

Table 10

Amount and Count of the Performed and Canceled Orders for each Customer

Customer Id	Performed Order Amount	Canceled Order Amount	Performed Order Count	Canceled Order Count
1	39251.50	NaN	3.0	NaN
2	20173.90	15911.50	1.0	3.0
3	17416.30	10840.90	21.0	15.0
4	27164.60	83357.10	25.0	47.0
5	50482.20	51643.00	3.0	4.0

3.6.1.2.2 Currency pair based features. The total number of each customer's preferred currency exchange according to each currency pair has been added to the data set as separate columns. Thus, the change in their behavior according to the currency pair they are buying and selling and its effect on the churn status were intended to be reflected in the model.

Table 11

Currency Pairs by Their Total Amount (USD)

Currency Pair	Amount (USD)
ALT(gr)/TL	46323
PLT(gr)/TL	6079
USD/TL	55108
EUR/TL	19366
ALT(gr)/USD	3274
GMS(gr)/TL	63421
GMS(gr)/USD	7035
Total	2948861040

The number of transactions made by each customer with certain currency pairs is as follows.

Table 12

The Count of Transactions Performed by Each Customer with Certain Currency Pair

Customer Id	ALT(gr)/TL	PLT(gr)/TL	USD/TL	EUR /TL	GMS(gr) /TL	GMS(gr) /USD
1	4.0	0.0	24.0	0.0	0.0	0.0
2	8.0	0.0	2.0	0.0	0.0	0.0
3	28.0	0.0	6.0	27.0	56.0	2.0
4	133.0	0.0	0.0	1.0	85.0	0.0
5	58.0	0.0	16.0	8.0	1.0	0.0

3.6.1.2.3 Temporal features derived from customer behavior. The pattern of shopping behaviors that the customer reveals as he moves toward churn is also an important information-providing factor. The monthly change in the customer's shopping volume and number of purchases in the last 6 months have been added to the model as separate columns. The table below lists the monthly changes in the total Amount USD values of the customers in the last 6 months. These derived features will also be presented to the model for training.

Table 13

Total Amount of Monthly Transactions for Each Customer (USD)

Customer Id	L1 Month Amount	L2 Month Amount	L3 Month Amount	L4 Month Amount	L5 Month Amount	L6 Month Amount
1	3000.0	0.0	0.0	0.0	0.0	0.0
2	3261.0	0.0	0.0	0.0	0.0	0.0
3	6490.0	12774.0	3868.0	2645.0	9311.0	13359.0
4	2787.0	12799.0	55344.0	53035.0	0.0	2591.0
5	20074.0	15175.0	0.0	0.0	0.0	0.0

3.6.1.2.4 RFM based features. In the RFM analysis, the recency, frequency, and monetary values of each customer were scored and recorded as separate data. These values will be evaluated as features related to customer segmentation, and they will be presented to the model for training.

3.6.2 Rare analysis. It was determined that there was rare data in five different columns. These columns are: "IsCorporateCustomer," "SectorNode," "CustomerClassName," "CityName," and "Profession". The threshold value was determined as 0.01; those below this limit were aggregated under the rare variable.

3.6.3 Missing observation analysis. One of the major issues that must be addressed during data preparation is the difficulty of incorporating missing features into huge collections of data. In practical applications, many datasets have incomplete information, that is, a small or substantial portion of the dataset is missing. There are classifiers that display resilient patterns in datasets with missing values, so the decision's ultimate outcome is unaffected by the missing values. On the other hand, some classifiers, such as neural networks, require careful treatment of partial data (Alexandropoulos et al., 2019). The critical threshold for the missing value analysis was set at 25 percent. Since the column consisting of more null values was given to the train model, it would cause too much noise, so the columns that exceeded this threshold value were removed from the data set. In the list below, the top 10 columns with the emptiest values are listed.

Table 14

Top 10 Columns with the Emptiest Values

Columns Name	Number of Missing Row	Ratio
RealEstateCount	6031	88.12
RealisedOrderCount	4605	67.29
CanceledOrderCount	4422	64.61
FXTranEfficiency	3298	48.19
RelatedBanksCount	3145	45.95
VehicleCount	2987	43.64
Profession	1546	22.59
EducationLevelId	1546	22.59
TotalFXAmountTL	530	7.74

Eight of these columns were removed from the data set and were not included in the training.

3.6.4 Outlier observation analysis. Two different perspectives have been adopted to fight against deviant values. The first is the local outlier factor, which is applied by evaluating all the data together. The second one was applied by changing the values outside the interquartile range with a threshold on a column basis.

The local outlier factor method was used, and five rows were removed from the data set due to outlier values.

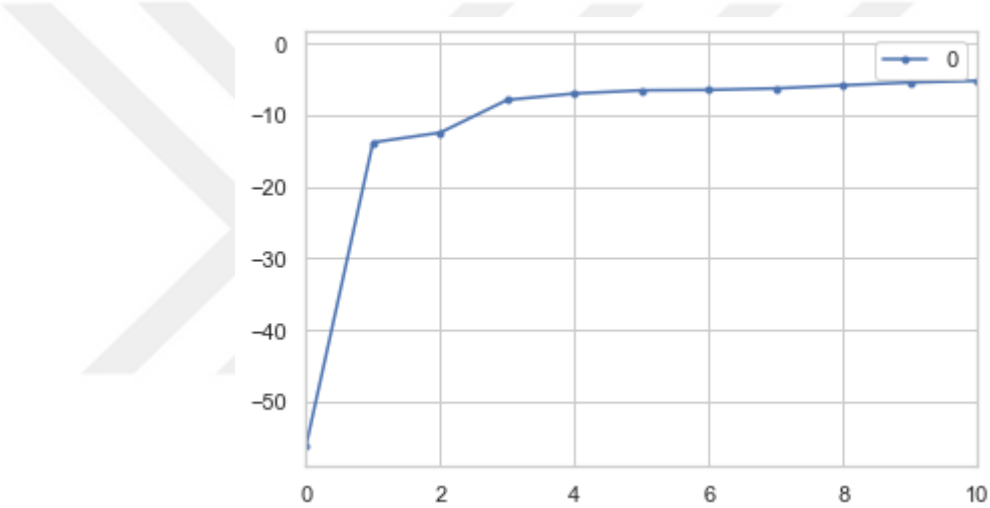

Figure 16. Plotting of the elbow method to detect the threshold for the local outlier factor to label a row as an outlier

As the second step, interquartile ranges were determined and the values outside this range are listed as follows.

Table 15

The Columns Having the Outlier Values

Columns Name	Number of Outlier Rows
TotalFXAmountTL	886
EducationLevelId	124
ValueSegment	62
ActiveProductsCount	241
OpenProductsCount	138
OpenTreasuryProductCount	39
ActiveTreasuryProductCount	31
LastYearEfficiency	790
LastYearTreasuryEfficiency	878
Efficiency	911
AmountUSD	894
ALT(gr)/TL	974
PLT(gr)/TL	340

The above values have been replaced with the q1 and q3 values so that the model will cause

3.6.5 Variance analysis. An analysis of variance evaluates the effect of a column on the model. If all rows in a column carry the same information, it does not mean a change for the model. For example, if everyone in the dataset is the same age, the age variable is not a determining factor for our model. The threshold value was determined as 0.05, and finally, the PLT(gr)/TL, ALT(gr)/USD, and GMS(gr)/USD columns were removed from the data set.

3.6.6 Feature scaling. A robust scaler is preferred as a feature scaling method. Thus, it is aimed at increasing the success of the model and ensuring that the model works.

3.6.7 Correlation analysis. Overfitting is an issue that happens when a model becomes overly reliant on its training data. When exposed to new data, an overfit model performs badly. To illustrate, the model may be very accurate overall despite classifying in a wrong way classes when applied to real-world data. (Behnke et al., 2021). As stated in Chapter 2, correlation analysis is done in order not to include

columns with similar information in the model unnecessarily. Because the inclusion of similar data both creates bias for the model and reduces its success, it negatively affects its working performance. In the correlation analysis, the threshold value was determined as 0.95, but since there was no relationship between any of the columns exceeding this value, no feature elimination was made from the data set at this step.

3.6.8 Feature elimination. For feature selection, SOCP and LOFO were used, as well as wrapper methods. The most successful method was determined to be LOFO when the results were evaluated, but its operating time is much longer than SOCP. In order to compare the two methods, different models will be trained with the feature sets produced, and their successes and predictions will be compared.

Table 16

Accuracy Score Derived from Variety of Feature Selection Methods (LightGBM Acc on Test)

Feature Selection Method	Accuracy Score	Time (Sec)
Leave One Feature Out (LOFO)	72.30	452.5
SOCP	71.80	165.2
Forward Selection	69.60	98.1
Backward Selection	70.40	101.5
Stepwise Selection	71.20	145.6

In LOFO, the effects of each variable on the success of the regression were evaluated based on the accuracy score using the 10-fold cross-validation method. As a result, the 39 variables had a negative effect on the success of the model. These 39 variables were not considered for the final models and were removed from the dataset using the "leave one feature out" method. While, in the SOCP method, 45 features were eliminated.

3.6.9 Data splits - cross validation. The average accuracy rate of the algorithm that will work for each data segment is also calculated. The cross validation (CV) value is given as 10. This means that it will divide the data set into 10 equal parts in each for

loop and give a different section to the algorithm as a test. Thus, an algorithm will have 10 different accuracy rates.

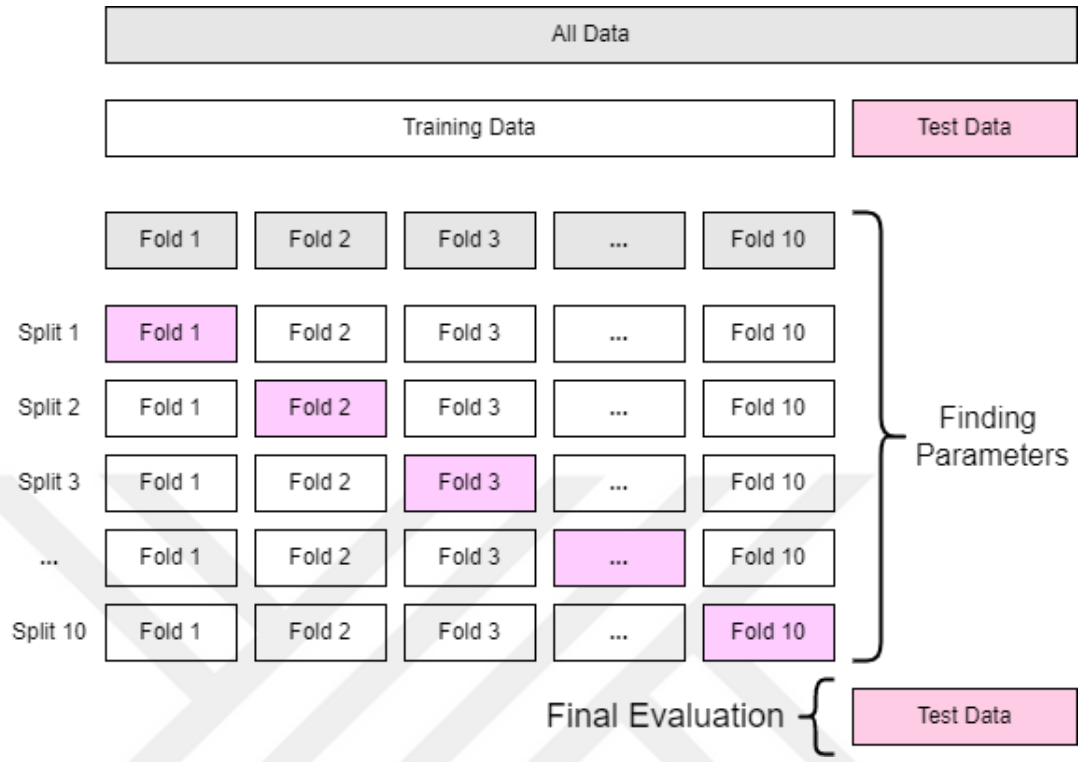

Figure 17. Diagram summarizing the applied method for cross validation and train test split

Chapter 4

Findings

4.1 RFM Approach Results

Segmenting the customers according to the RF scores obtained will be the first thing to do. The map below shows the corresponding scores and divisions for this classification. This task can be done very quickly in Python using regular expressions. It is possible to examine each customer within a segment after each customer's RFM score is calculated.

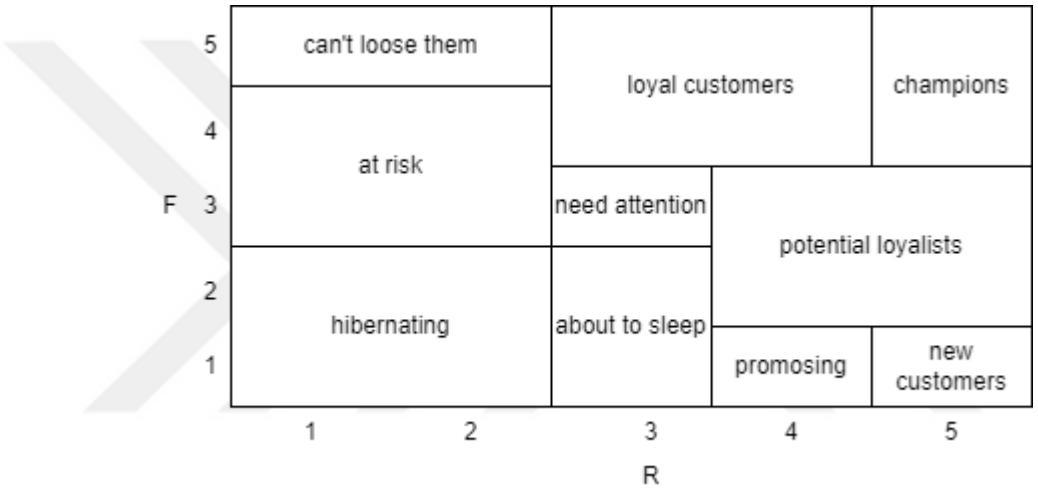

Figure 18. The mapping used to convert RFM scores into segments.

Both customer behavior and a strategy for the future can be determined by using RFM scores. These customers, whose RFM scores have been determined, can be grouped together as they have unique behavioral characteristics. The final version of the table where customers are divided into segments. Therefore, both in terms of action decisions and the structure of the points, they can be interpreted.

Table 17

Division of Customers into Segments According to Their RFM Scores

Customer Id	Recency Score	Frequency Score	Monetary Score	RFM Score	Segment
1	5	4	4	544	Champions
2	2	3	3	233	At Risk
3	5	5	4	554	Champions
4	2	5	5	255	Can’t Lose
5	3	5	5	355	Loyal Customer

The number of customers in each segment is shown in the table below. The segment with the highest number of customers is Hibernating, with 1424 customers. The segment with the fewest customers is the New Customer Segment, with 135 customers.

Table 18

The Number of Customers in Each Segment

Customer Segment	Number of Customer
Hibernating	1424
Loyal Customers	1165
At Risk	1035
Potential Loyalists	964
Champions	831
About to Sleep	543
Can’t Lose	276
Need Attention	276
Promising	195
New Customers	135

For the analysis of the total and average distribution of the recency, frequency, and monetary values in the segments according to the main segments, the table below has been prepared. Accordingly, the most profitable customers are those in the "Champions" group. This group performed an average of 1,508,600 TL transactions per person during the period of analysis. On the contrary, the lowest volume per capita,

with an average of 22,681 TL, was observed in the "promising" segment. Customers in the "Champions" segment ranked first in frequency, with the highest average of 105 transactions per person. The group with the lowest frequency value was the "promising" segment, with an average of 1.20 transactions. A low recency value is a good sign for the company. While the audience with the lowest recency value was the "Champions" segment with an average of 4.8 days, on the other hand, the group with the highest recency value was the customers in the "Hibernating" segment with an average of 450 days.

4.2 CLV Approach Results

The CLTV account has many calculation metrics. Basically, the metrics to focus on are average order value, purchase frequency, profit margin, churn rate, customer value, and customer lifetime value. The recency value between the columns created is equal to the time between a customer's first purchase and their last purchase. The value of T (customer's age) is equal to the time between a customer's first purchase in the dataset and their last day. Frequency is the number of times a customer makes repeat purchases. On the other hand, monetary value is the profitability value of the customer's relationship with the company.

4.2.1 The BG/NBD model. For small datasets in frequency/recency analysis using the BG/NBD model, the parameters can be incredibly large, so adding a l2 penalty to the probability can control how large these parameters can be. For this case, it can be made to affect the model by entering the penalizer coefficient value while creating the model. In typical applications, punishments between 0.001 and 0.1 are effective.

Table 19

Penalizer Coefficient that Controls How Large the Parameters Can be

	Coefficient	se(coef)	Lower 95% Bound	Upper 95% Bound
r	0.558563	0.010479	0.538025	0.579101
alpha	8.170005	0.221731	7.735412	8.604597
a	0.416228	0.022306	0.372509	0.459948
b	6.569472	0.580774	5.431154	7.707790

A customer who has been shopping from a company continuously and consistently for a very long time—for example, if no interaction has been made for a month—what is the probability that this customer is "alive" for the company? It can be expected to be quite low. On the other hand, another customer makes a purchase every three months and has done his shopping last quarter, so it can be said that this customer is still alive.

The number of transactions a customer is expected to complete in the next time slot can be visualized using the result matrix divided by frequency and recency. Frequency (the number of repeat purchases) and Recency (last shopping date—first shopping date)

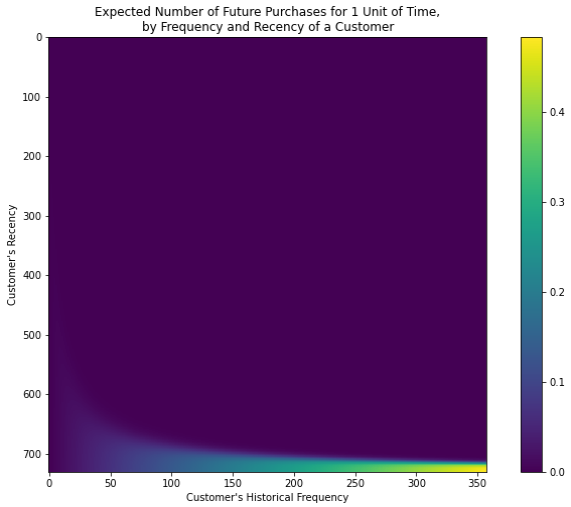

Figure 19. Expected number of future purchases for 1 unit of time by frequency and recency of a customer.

Customers who have made many purchases and have recently made these purchases will be the best customers in the future.

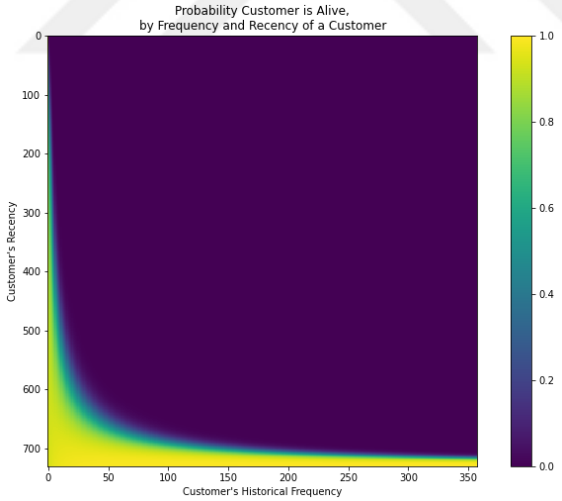

Figure 20. Probability customer is alive by frequency and recency of a customer.

Recent customers are almost certainly "alive." Customers who shop frequently but haven't visited recently are users who are heading towards churn. The more these users have shopped in the past, the more likely they are to churn. (Figure 20.)

When customers are ordered from most to least according to their probability of purchasing the most, the result will be as follows: The model is used to predict the expected future purchases of customers based on their past shopping habits.

Table 20

Total Amount of Monthly Transactions for Each Customer (USD)

Customer Id	Frequency	Recency	T	Monetary	Probability Alive	Predicted Purchases
349	249.0	408.0	408.0	126968.0	0.9983	17.70
351	91.0	149.0	149.0	76364.0	0.9957	16.79
11	357.0	648.0	648.0	243679.0	0.9988	16.17
72	181.0	326.0	326.0	30164.0	0.9977	15.97
196	258.0	484.0	484.0	115013.0	0.9984	15.54

Listed above are the top 5 customers the model expects to make a purchase the next day.

"Predicted Purchases" represents the number of expected purchases, while the other three columns represent current RFM metrics. The BG/NBD model predicts that these people will make more purchases soon because they are the current top customers. It is observed that the success of the established model is satisfactory.

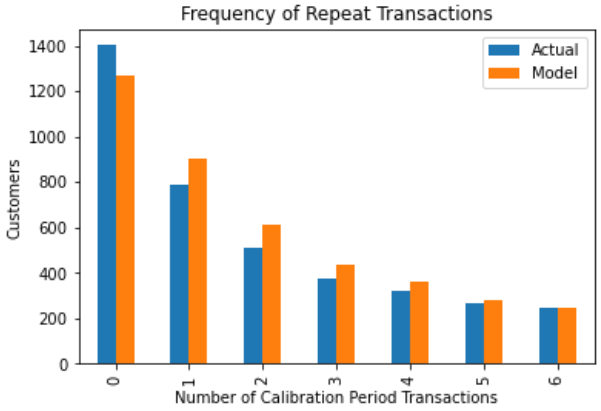

Figure 21. Bar graph showing actual and predicted number of transactions during calibration period to compare.

To increase the generalizability of the CLV model, the cross-validation method was applied.

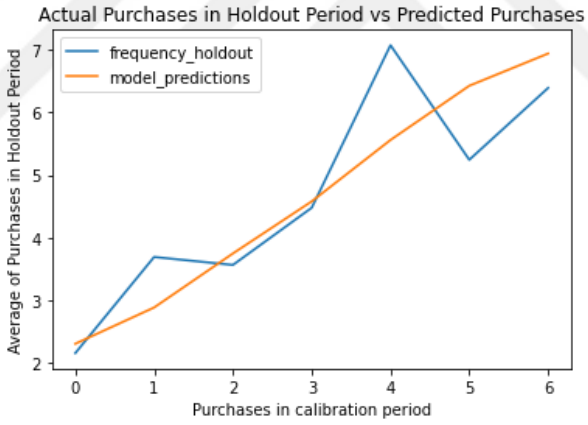

Figure 22. Line graph showing actual and predicted purchases in holdout period.

The chart above shows the data separated by train and test set. The end of the training period is "2020-12-31," while the end of the testing period is "2021-12-31." The graph groups all customers in the train period by the number of repeat purchases (x-axis) and then averages their repeat purchases in the waiting period (y-axis). The blue and orange lines indicate the train and test sets, respectively. As can be seen, the

model can accurately predict customer behavior with a high success rate. However, the success of the forecasts declines after future purchases.

Except for the analyses made on the entire customer base, the need for customer-based forecasting for customers that are important to companies will be met. Based on customer history, it is now possible to predict an individual's future purchases. For example, the model predicts that the future transaction of a randomly selected customer with customer number 567 is 0.35396828232219774 in 10 days.

Viewing customer probability histories can also play a critical role in CRM studies. Periodic visualization provides a wide perspective and an opportunity to gain flexibility in accordance with the conditions of the day. For this need, given the transaction history of a customer, the probability of their past survival according to the trained model can be easily calculated. For example, the transaction history of customer number 347, one of the top customers, can be viewed, and the probability of survival can be periodically examined.

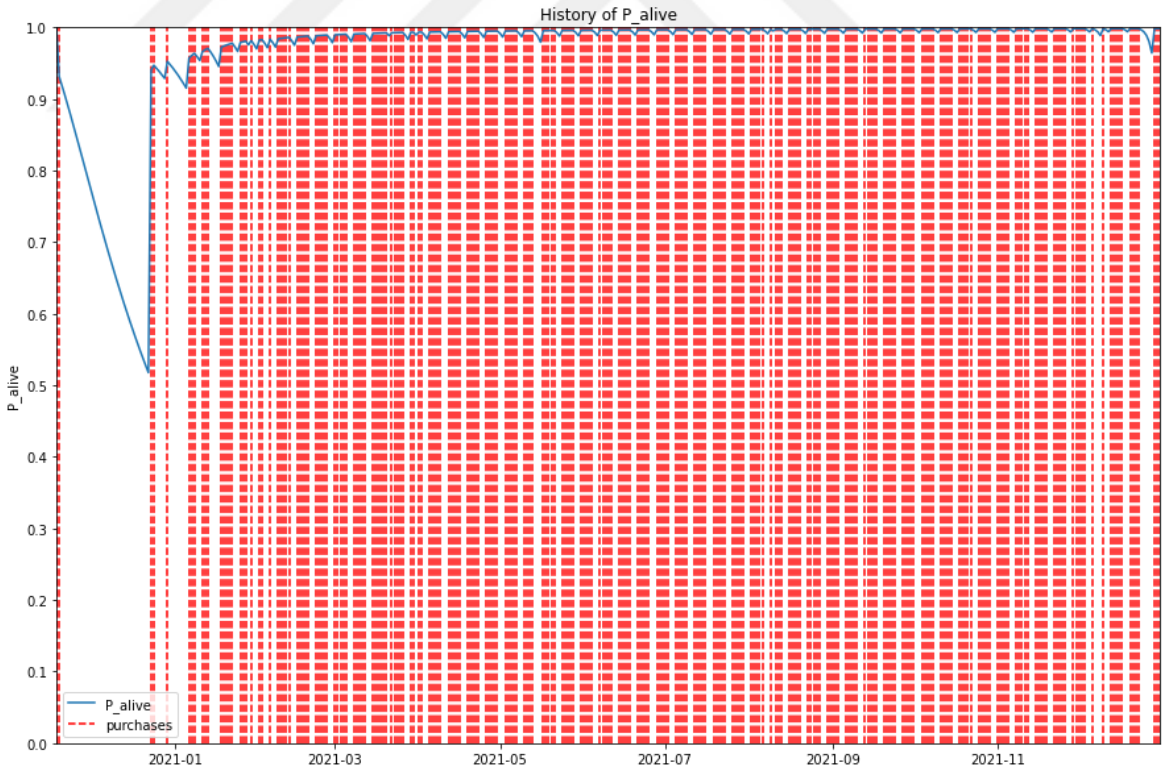

Figure 23. Graph showing probability of alive of a customer in a certain period.

According to this image, the customer continues shopping actively and with vigor. However, at the beginning of 2021, this customer took a break from shopping, so the probability of survival decreased. In addition, an alarm can be created for this interval with the developed automations, and a request can be sent to the relevant teams to prepare a report on the customer's status. In this historical range, macroeconomic data can be compared to the customer's balance sheet. With the data obtained here, a road map can be drawn for the continuation of the customer's relationship with the bank.

4.2.2 The Gamma-Gamma model. To establish the Gamma-Gamma model, which shows the expected conditional average profitability values over a certain period, customers who repeat their shopping at least once must be selected. For this reason, the number of unique customers decreases to 5442.

$$E(M|p,q,\gamma,m_x,x) = \frac{(\gamma + m_x x)p}{px + q - 1} = \left(\frac{q-1}{px+q-1}\right)\frac{\gamma p}{q-1} + \left(\frac{px}{px+q-1}\right)m_x \qquad 0.1)$$

It shows the expected average profitability values. To set up this model, users who repeat their shopping at least once must be selected. As a result, with BG/NBD and Gamma-Gamma probabilistic models, these people are expected to shop soon, so the definitions of the best customers for the company, the examination of whether the customers will be alive or not, the verification of the established model, and the most profitable and lifetime values of all customers can be estimated. And as a result, segmentation can be easily performed according to these values. Gamma-Gamma model fitted with 5442 subjects and parameters with p = 2.04, q = 3.70, and v = 534.86. The 10 customers that are predicted to bring the most profit, which has been applied to the BG/NBD model before, can be listed as follows:

Table 21

The Top 10 Customers that are Predicted to Bring Most Profit

Customer Id	Predicted Monetary
6272	1164035.10
3283	1119255.60
3006	902503.90
783	719588.10
1129	688714.90
4029	661601.90
350	637912.40
4595	621669.50
565	565966.70
5568	515540.00

4.2.3 Calculation of CLTV with BG-NBD and GG model. The top 10 customers with the highest customer lifetime values calculated with BG/NBD and Gamma-Gamma models are listed as follows.

Table 22

The Top 10 Customers that are Predicted to Bring Most Profit

Customer Id	CLV
1198	1925823.70
11	1690495.60
1603	1504263.90
1129	1182440.00
349	967552.80
565	887896.10
308	844211.10
1420	815785.90
187	809445.70
196	768553.00

The CLV value has been calculated for all customers, and these values need to be compressed to values between 1 and 0 to make sense. For this, the min max scaler method was used. Then all customers are divided into 5 different segments according to their CLV values. Customer behaviors will be characterized according to these segments.

Table 23

Customer Behaviors Segmented According to CLV

Customer Segment	Number of Customer
E	1371
D	1367
C	1368
B	1369
A	1369

4.3 Machine Learning Approach Results

Using popular tree-based methods, the preprocessing steps were completed, and the dataset was trained. Performance analyzes were made and compared.

4.3.1 Churn prediction performance analysis. The customers obtained according to the estimation results in the determined models are divided into TP, TN, FP, and FN. However, after this point, another important point for the company is the evaluation metric according to which action will be taken. If it is important to accurately identify the positives, a model with higher precision should be chosen. However, if it is more important to accurately identify negatives, specificity should be chosen as the measurement metric. However, it would be a more accurate approach to choose model success in the first stage. In a much more inclusive view, the AUC value makes it easy to compare the success of two models trained on the same dataset. For

this, a comparison was made according to the ROC curve and the AUC value of the area under this curve to compare the models.

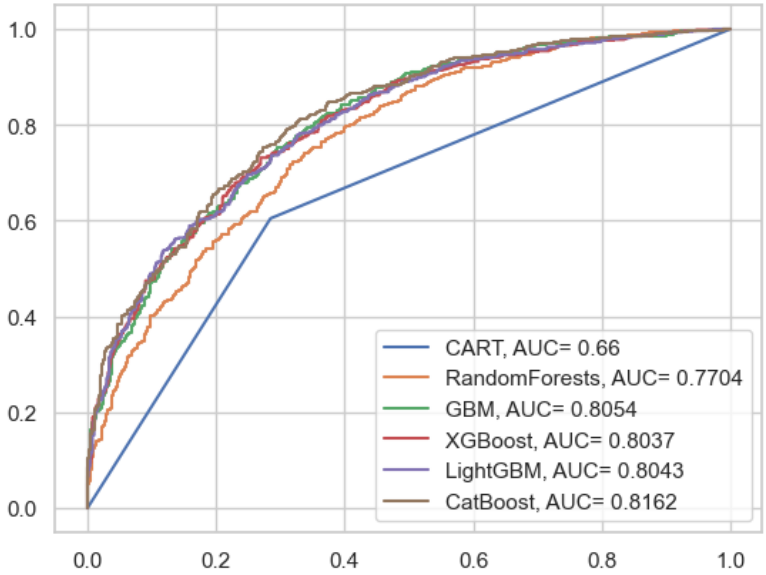

Figure 24. Line graph showing AUC score of data frame eliminated by LOFO

To compare the success of the feature selection methods, in addition to the LOFO, which has the highest accuracy, the same models with the base parameters were established by using the data frame filtered as a result of the SOCP method. The AUC results are as follows.

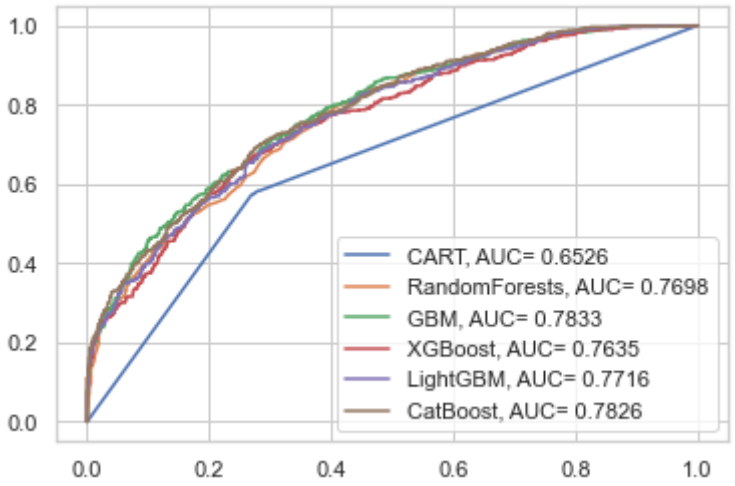

Figure 25. Line graph showing AUC score of data frame eliminated by SOCP

4.3.1.1 Cross validation results with base models. The data set, which was passed through data preprocessing steps, was first trained with various models using default parameters. These models are decision trees (CART), random forests, GBM, XGBoost, LightGBM, and CatBoost. The CatBoost method is the most successful when training is performed with base parameters without any optimization.

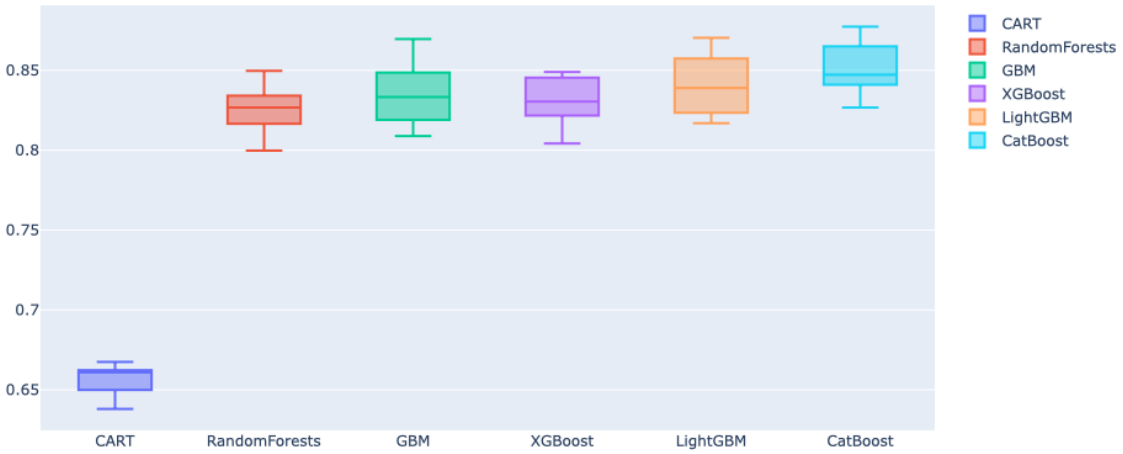

Figure 26. Box plots showing AUC score of data frame eliminated by LOFO

The cross-validation method was also used on the SOCP data frame, and the results were compared. Accordingly, LOFO is more successful than SOCP.

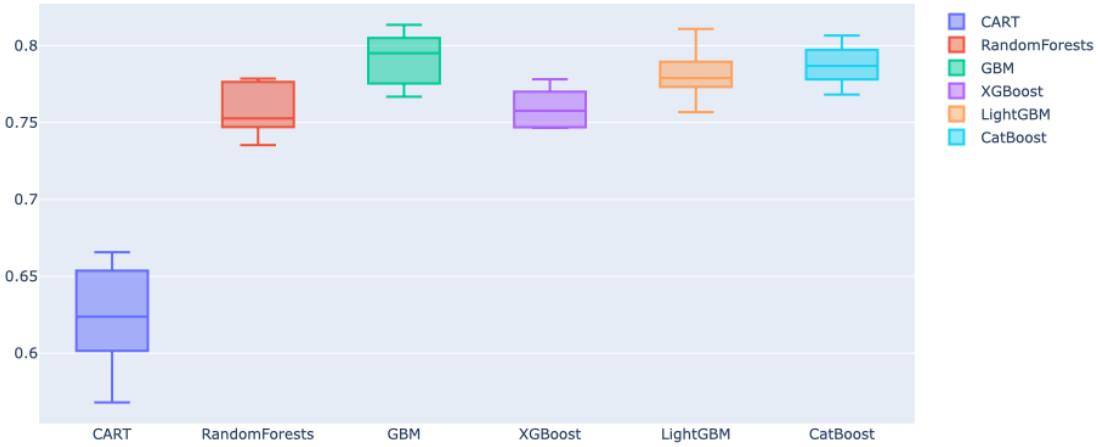

Figure 27. Box plots showing AUC score of data frame eliminated by SOCP

4.3.1.2 Hyperparameter tuning with ROC AUC as a scoring metric. The stratified K-fold method was used for cross validation. The necessary parameters for this were determined as n_splits 8, and by using random_state, consistency in each calculation was ensured. Accordingly, the results in models produced with base parameters were compared with roc_auc scoring.

As a result of grid search, the CART model's 0.761158 AUC score was calculated as the best score, and GINI for the criterion parameter, 5 for the max_depth parameter, 100 for the min_samples_leaf parameter, and 2 for the min_samples_split parameter were determined as the most optimum parameters.

As a result of the grid search, the RandomForests model's 0.828857 AUC score was calculated as the best score, and the values of 15 for the max_depth parameter, 9 for the max_features parameter, 2 for the min_samples_leaf parameter, 15 for the min_samples_split parameter, and 200 for the n_estimators parameter were determined as the most optimal parameters.

As a result of the grid search, the GBM model's 0.845808 AUC score was calculated as the best score, and the values of 0.03 for the learning rate parameter, 6 for the max_depth parameter, 1500 for the n_estimators parameter, and 0.9 for the subsample parameter were determined as the most optimal parameters.

As a result of the grid search, the XGBoost model's AUC score of 0.842484 was calculated as the best score, and the values of 1 for the colsample_bytree parameter, 0.1 for the learning_rate parameter, 5 for the max_depth parameter, and 200 for the n_estimators parameter were determined as the most optimal parameters.

As a result of the grid search, the LightGBM model's 0.836782 AUC score was calculated as the best score, and the values of 1 for the colsample_bytree parameter, 0.1 for the learning rate parameter, 5 for the max_depth parameter, and 500 for the n_estimators parameter were determined as the most optimal parameters.

As a result of the grid search, the CatBoost model's AUC score of 0.851853 was calculated as the best score, and a value of 5 for the depth parameter, 800 for the iterations parameter, 3 for the l2_leaf_reg parameter, and 0.03 for the learning rate parameter were determined as the most optimal parameters.

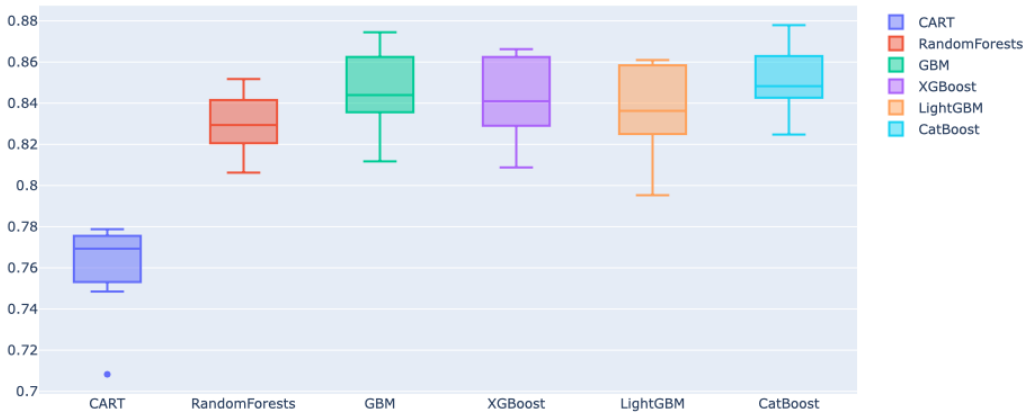

Figure 28. Box plots showing LOFO of television watched per week by age group

Table 24

Roc AUC Score of LOFO Eliminated Models Before and After Hyperparameter - Tuning

Model Name	AUC Score Before Tuning	AUC Score After Tuning
CART	0.6032	0.6599
Random Forests	0.7993	0.8287
Gradient Boosting Machine	0.8031	0.8360
XGBoost	0.8243	0.8438
Light GBM	0.8410	0.8508
CatBoost	0.8521	0.8616

After the hyperparameter optimization was done on the SOCP data frame, AUC comparison was possible.

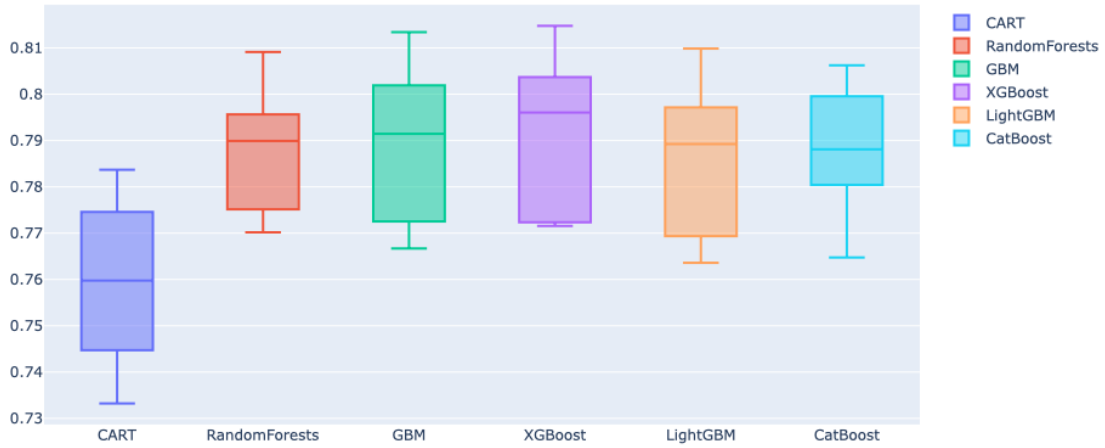

Figure 29. Box plot showing SOCP of television watched per week by age group

Table 25

Roc AUC Score of SOCP Eliminated Models Before and After Hyperparameter - Tuning

Model Name	AUC Score Before Tuning	AUC Score After Tuning
CART	0.6232	0.7597
Random Forests	0.7533	0.7885
Gradient Boosting Machine	0.7901	0.7931
XGBoost	0.7518	0.7995
Light GBM	0.7842	0.7872
CatBoost	0.7826	0.7883

4.3.1.3 Test scores. As a result of hyperparameter optimization, the training step was completed, and the step of measuring the real success of the model with the test data was started. At this stage, the data has never been used before for the model and is completely new. In the CART model, the criterion parameter is GINI, the max_depth parameter is 5, the min_samples_leaf parameter is 100, and the min_samples_split parameter is 2, and the test data is reconstructed for scoring. RandomForests model; it was reconstructed for scoring the test data by assigning a value of 15 for the max_depth parameter, a value of 9 for the max_features parameter, a value of 2 for the min_samples_leaf parameter, a value of 15 for the min_samples_split parameter, and a value of 200 for the n_estimators parameter. GBM was reconstructed model by assigning a value of 0.03 for the learning rate parameter, a value of 6 for the max_depth parameter, a value of 1500 for the n_estimators parameter, and a value of 0.9 for the subsample parameter. XGBoost model; it was reconstructed for scoring the test data by assigning a value of 1 for the colsample_by_tree parameter, a value of 0.1 for the learning_rate parameter, a value of 5 for the max_depth parameter, and a value of 200 for the n_estimators parameter. LightGBM model; it was recreated for scoring with test data by assigning a value of 1 for the colsample_bytree parameter, 0.1 for the learning_rate parameter, 5 for the max_depth parameter, and 500 for the n_estimators parameter. Again, the CatBoost model was reconstructed for scoring the test data by assigning a value of 5 for the depth parameter, a value of 800 for the iterations

parameter, a value of 3 for the l2_leaf_reg parameter, and a value of 0.03 for the learning rate parameter.

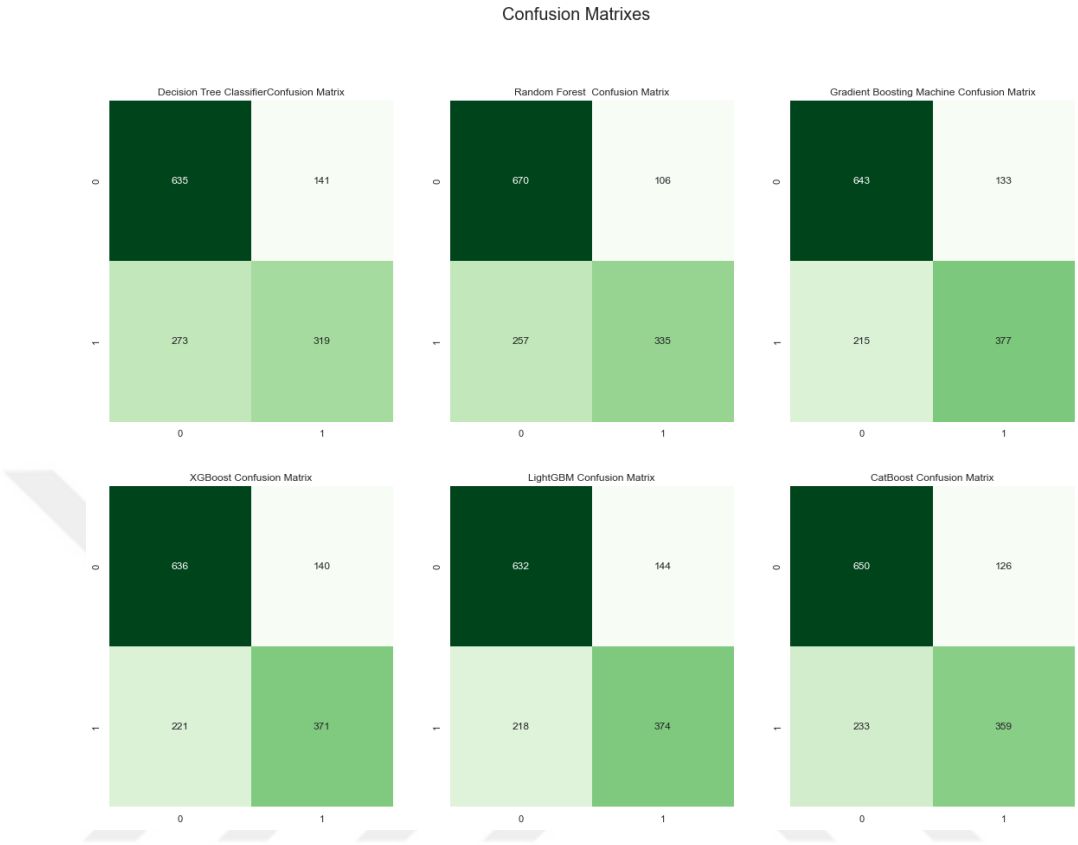

Figure 30. Confusion matrix showing relations between predicted and actual values by all models on LOFO data frame.

Table 26

Evaluation Scores of Each Model with Test Data Derived Using The LOFO Method

Model Name	Accuracy	Precision	Recall	F1 Score	AUC Score	Time(Sec)
CART	0.66	0.59	0.59	0.59	0.66	0.04
RandomForests	0.76	0.74	0.62	0.67	0.81	0.58
GBM	0.75	0.73	0.65	0.68	0.83	0.71
XGBoost	0.76	0.72	0.69	0.70	0.83	1.28
LightGBM	0.76	0.73	0.68	0.71	0.82	0.15
CatBoost	0.78	0.75	0.70	0.72	0.83	3.63

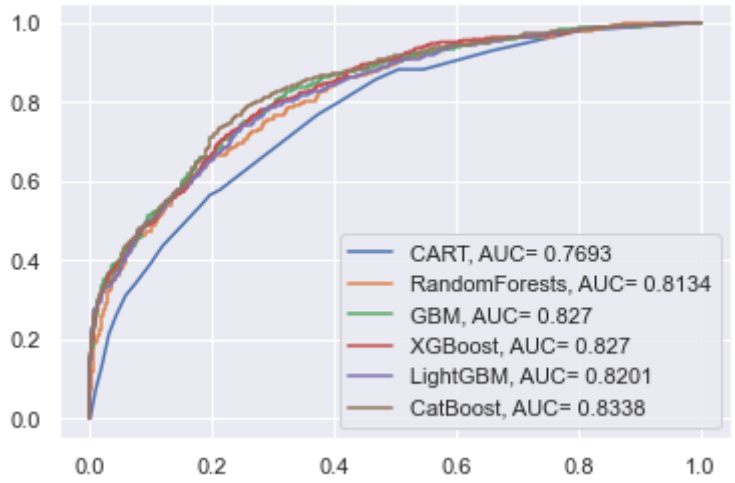

Figure 31. Line graph showing AUC score of test data frame eliminated by LOFO

After the models were finalized, they were presented to the production environment using the best parameters. Success metrics were obtained with the data set of the SOCP data frame reserved for testing, and the results of each model were compared.

Table 27

Evaluation Scores of Each Model With Test Data Derived Using The SOCP Method

Model Name	Accuracy	Precision	Recall	F1 Score	Roc AUC Score	Time(Sec)
CART	0.68	0.70	0.53	0.56	0.76	0.02
RandomForests	0.69	0.69	0.55	0.60	0.78	0.45
GBM	0.70	0.70	0.55	0.62	0.77	0.62
XGBoost	0.71	0.70	0.53	0.62	0.78	1.01
LightGBM	0.71	0.73	0.56	0.58	0.77	0.13
CatBoost	0.71	0.72	0.54	0.61	0.78	2.94

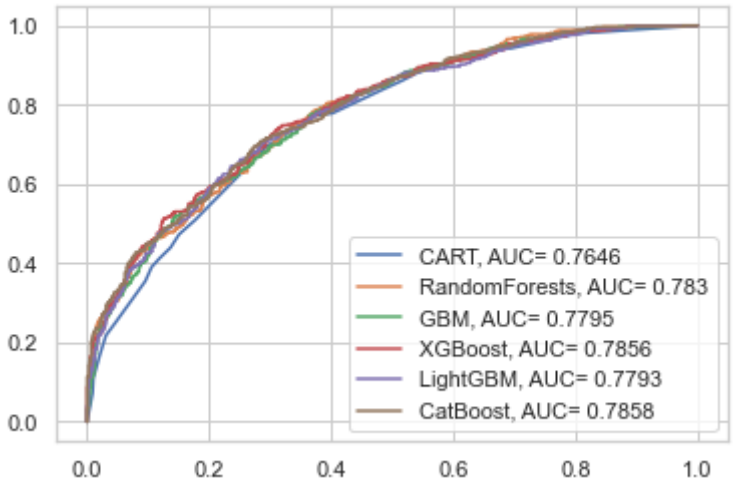

Figure 32. Bar graph showing AUC score of test data frame eliminated by SOCP

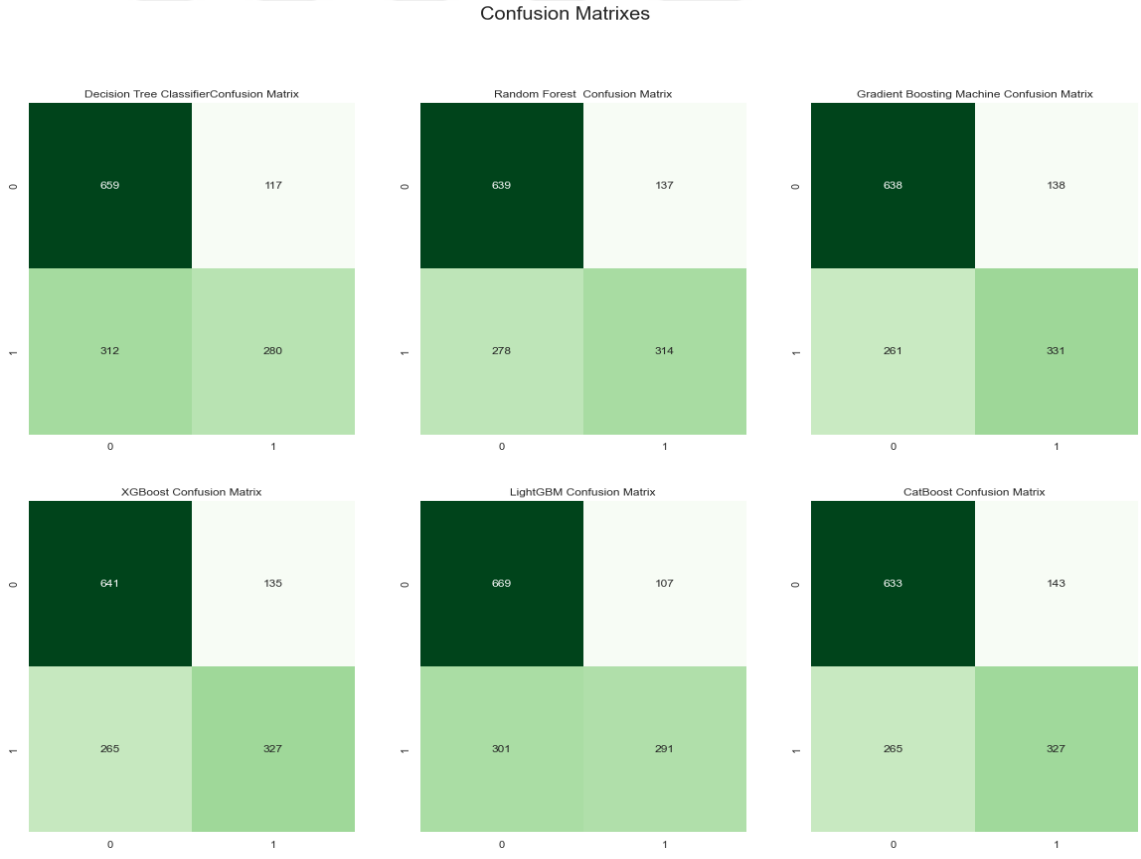

Figure 33. Confusion matrix showing relations between predicted and actual values by all models on SOCP data frame.

4.3.2 Feature analysis and importance. As a result of the data being trained with tree-based popular methods, the way each model makes use of the features in the data set differs according to its own working principle. Accordingly, feature importance rankings can be obtained for each method separately. Accordingly, the first 5 features in terms of importance in the CART method are listed as “Last Year Treasury Efficiency”, “USD/TL”, “Last 1 Month Amount”, “Last 1 Month Count”, and “Last 4 Month Amount” while, in the random forest method, it is “Last Year Treasury Efficiency”, “USD/TL”, “Last Year Efficiency”, “TotalFXAmount”, and “USD Amount”, respectively. On the other hand, the 5 most important features for the GBM algorithm, which is a boosting method, were “Last Year Treasury Efficiency”, “USD/TL”, “Last 1 Month Amount”, “Total FX Amount”, and “USD Amount”. The top 5 features determined by today's popular and widely used XGBoost model are “Last 6 Month Count”, “USD/TL”, “Last 6 Month Amount”, “Last 1 Month Count”, “Last Year Treasury Efficiency”, and “Open Treasury Products Count”. For Light GBM, another popular model that is XGBoost's biggest competitor, the most important 5 features are listed as “Last Year Treasury Efficiency”, “Last 1 Month Amount”, “USD Amount”, “Total FX Amount”, and “Last Year Efficiency”. Finally, the 5 features that have the highest impact on the model trained with CatBoost, which is less popular but at least as successful as other models, were observed as “Last 6 Month Count”, “Last 1 Month Amount”, “USD/TL”, “Last Year Treasury Efficiency”, and “Open Treasury Products Count”.

Chapter 5

Discussion and Conclusions

5.1 Discussion of Findings for Research Questions

As a result of the loss analysis, it will be possible to ask the questions of "who" and "how do they behave" to recognize the customers who are preparing to leave. Identifying these customers will provide a concrete expression of the potential effects on profitability and sales if they leave. The position that the company will take to retain these customers will be clearly demonstrated and will ensure that the company uses its resources efficiently and reflects the value in its customer profile in the best way possible. In addition to individual measures, customer segmentation is critical in developing an overall policy to prevent churn among customers who are difficult to predict and have the potential to leave. According to Keramati et al. (2016) it is expected that, with a better understanding of the features of churners, managers of the banks should consider some strategies to prevent churn. These strategies should be used for customers whose features are becoming more like those of the churner groups identified. These strategies can include providing required facilities, growing consumer reactivity, identifying the needs of different groups, and improving the quality of services.

Table 28

Probability of Churn Derived from Different Models for Some Customers Obtained Probabilistically

Customer Id	Churn	CART	RF	GBM	LGBM	XGboost	CatBoost
1366	True	0.835	0.920	0.999	0.999	0.998	0.993
3662	True	0.421	0.613	0.998	0.999	0.989	0.829
147	False	0.413	0.417	0.207	0.102	0.207	0.245
475	True	0.649	0.589	0.952	0.980	0.952	0.843
1582	False	0.808	0.470	0.191	0.060	0.191	0.256
5341	True	0.162	0.439	0.713	0.918	0.713	0.520
509	True	0.413	0.519	0.619	0.471	0.619	0.507
841	False	0.710	0.003	0.006	0.001	0.006	0.034

5.2 Conclusions

If a company is creating customer strategies to maximize profits, it needs to do a detailed and end-to-end study on churn analysis. At this point, it is impossible for only machine learning companies to implement churn strategies. According to the estimation results, in addition to the operation of different actions in marketing and IT organizations, a good observation of these processes will lead to a healthy functioning. In this study, using the data of a financial institution in the interest-free finance system, churn classification was made with the highest performance classification algorithms available, as well as RFM and CLV analysis to predict the churn behavior on the mobile application. Algorithms were applied over two years of data for 2020 and 2021 obtained from the interest-free financial institution. As a result of the test, it was revealed that the size of the studied data is suitable for such an analysis. In such studies, explanatory variables are of great importance.

Many different methods and strategies are used by companies in customer churn analysis studies. In this study, 3 different approaches were mentioned, and their results were analyzed.

First, customer classification was applied with the RFM method. As a result, customers are grouped in a 2-dimensional plane. In this way, the first step of creating a sales strategy for each segment was taken. With this method, CRM departments can achieve effective results and plan campaign management by spending minimum effort. However, the disadvantage of RFM is that the generated scores do not belong to each customer. The obtained scores are generalized for all customers in the segment. In addition, the data obtained for the classes are discrete on a scale between 1 and 5. It offers a low-sensitivity insight. In addition, it does not use any information about the characteristics of customers. Behavioral analysis of customers as well as demographic data of customers are also stored in banks. These data can carry vital data for churn analysis. For this reason, the Machine Learning approach comes to the fore in customer churn analysis.

In BG/NBD, another method, a statistical path was followed. In addition to the R, F, and M values, the T value, which can be expressed as the age of the customer, is also used. When this method, which has a stronger mathematical background, is used, a specific survival rate can be calculated for each customer, considering their behavior separately. In this respect, it provides the opportunity to create a strategy based on the customer. However, when calculating with this method, many existing information of the customers is not used. Additionally, it may be a new concept for many data scientists and analysts. It is a difficult method to use and evaluate for the technical capacity of CRM departments. Therefore, it will require additional personnel support for its interpretation.

The biggest benefit of tree-based machine learning models is that all factors that can cause customer churn are evaluated together, and these can be easily explained with tree-based methods. The aim for interest-free finance companies is not only to predict which customers will churn but also to understand the reasons for customer losses, for this reason, the effect of the parameters that may cause the loss should be included in the analysis. Organizational improvements should be made based on the results of these models. Thus, it becomes easier to make systemic changes to reduce customer losses and to take appropriate measures for the customer. From this point of view, it is thought that some of the explanatory variables that come to the fore in the study will also contribute to the sector's achievement of these goals. As mentioned before, when the algorithms run with the most optimal parameters are compared, it has been observed that the CatBoost algorithm model reaches the most successful result. Considering all models, the most important variables to detect customer churn are "Last 1 Month Amount," "Total FX Amount," "Last Year Treasury Efficiency," "USD/TL," and "Open Treasury Products Count." When both the number and transaction volume of customers' USD/TL transactions are analyzed monthly, it provides many insights about the customer's churn probability. The temporal graph created by the months before the last transaction can clearly show the churn process of a customer to companies. It has been observed that customers with a high open treasury products count do not leave the system and continue, while customers with a low total FX and precious metal amount value can leave the system more easily. Considering that participation banks are generally preferred for the low-cost

transaction opportunity they provide in precious metal purchases and sales, it is quite reasonable. It is obvious that this behavior of the customers will push the bank to make different campaigns or promotions in this regard. Therefore, it will be a risk-reducing factor for the bank to score customers based on the number and volume of FX and PM transactions in the last six months and take precautions. Another factor that has as much impact as FX and PM transactions is efficiency. If a customer's profitability has started to decline in the bank, it can be considered a good signal for churn. Considering the economic conditions in Turkey, the volatility of the exchange rate between the dollar and TL increases the probability of customers making transactions due to the increase in transactions.

In the feature selection step, in addition to the wrapper methods, the LOFO and SOCP methods were also used. LOFO and SOCP came to the fore as the two most successful methods. Although LOFO is much more successful, the processing time is too long. For this reason, choosing it for real-life problems will cause a lot of time loss for companies. On the other hand, although SOCP produced less successful results than LOFO, it worked much faster.

In summary, as a result of the study, it is seen that the CatBoost Classification algorithm is quite successful in predicting the churn behavior in the interest-free finance system. In addition to the CatBoost algorithm, LGBM and XGBoost algorithms have also produced successful results. In addition, although LGBM and XGBoost do not produce the best results, it can be preferred for businesses because it works faster than CatBoost. According to the studies and reviews, churn behavior cannot be fully explained unless it is supported by more information about other companies in the sector. This study was carried out using information from a single interest-free finance system. Information about other companies can be developed by including them in the studies. Another area for improvement is to focus on problems arising from the development of customer churn over time. Since the models that are the subject of this study are models that ignore the growth in the time axis, they cannot explain the evolution of the behavior of churn over time. Subsequent studies and research should be developed with this in mind as well.

5.3 Recommendations

Companies can devise innovative tactics to combat churn. These techniques must be implemented for the targeted consumers who display churn-like behavior. Offering a new plan, providing better quality service, determining the needs of different segments of customers, and providing tailored offers for different customer groups can all be included in the strategies. (Shen & Shibghatullah, 2022). Among the reasons for leaving, the low number of USD/TL transactions is also an important indicator. Realization of the orders at the desired margin and exchange rate is one of the factors that should be considered in terms of customer satisfaction and the continuity of this satisfaction. Despite the conditions being met, the failure of the orders to be fulfilled will have an important role in removing the customers from the system. Customers, known as traders, follow the market and enter their buy or sell orders into the system by using their own technical skills. The problems experienced by these customers will return as customer dissatisfaction. The biggest reason why customers prefer a bank for foreign exchange buying and selling transactions is the margins applied during buying and selling transactions. Considering this situation, CRM departments can prevent profitable and churning customers from becoming churned by providing low-margin transactions. Meetings with customers are another important factor for the system. Interviews and customer visits are very important in explaining the system to the customer and arousing curiosity. The occupational group to which the interest-free finance system appeals is critical when developing a new campaign or planning an advertising campaign. When analyzed by occupational groups, it was observed that private sector employees preferred the system more than other occupational groups, and customers in the private sector group left the system more.

Because of the privacy issues and regulations that applied to the bank, real time datasets and some of the most important features are not allowed to be used in this research.

REFERENCES

Ahn, J. H., Han, S. P., & Lee, Y. S. (2006). Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry. Telecommunications policy, 30(10-11), 552-568.

Aleksandrova, Y. (2018, July). Application of machine learning for churn prediction based on transactional data (RFM analysis). In 18 International Multidisciplinary Scientific Geoconference SGEM 2018: Conference Proceedings (Vol. 18, No. 2.1, pp. 125-132).

Alexandropoulos, S. A. N., Kotsiantis, S. B., & Vrahatis, M. N. (2019). Data preprocessing in predictive data mining. The Knowledge Engineering Review, 34.

Al-Tashi, Q., Abdulkadir, S. J., Rais, H. M., Mirjalili, S., & Alhussian, H. (2020). Approaches to multi-objective feature selection: A systematic literature review. IEEE Access, 8, 125076-125096.

Anitha, P., & Patil, M. M. (2019). RFM model for customer purchase behavior using K-Means algorithm. Journal of King Saud University-Computer and Information Sciences.

Aslan, D., & Tolon, M. (2018). Comparing customer segmentation with CLV using data mining and statistics: A case study. İşletme Araştırmaları Dergisi, 10(4), 887-900.

Atici, G. (2018). Islamic (Participation) banking and economic growth: empirical focus on Turkey. Asian Economic and Financial Review, 8(11), 1354-1364.

BAŞKOL, M. (2020). DETERMINING CUSTOMER SEGMENTATION BY USING RFM AND CORRESPONDENCE ANALYSIS. Business & Management Studies: An International Journal, 8(4), 902.

Batista, G. E., & Monard, M. C. (2002). A study of K-nearest neighbour as an imputation method. His, 87(251-260), 48.

Behnke, M., Briner, N., Cullen, D., Schwerdtfeger, K., Warren, J., Basnet, R., & Doleck, T. (2021). Feature engineering and machine learning model comparison for malicious activity detection in the dns-over-https protocol. IEEE Access, 9, 129902-129916.

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. Artificial Intelligence Review, 54(3), 1937-1967.

Binh, T. V., Thy, N. G., & Phuong, H. T. N. (2021). Measure of CLV toward market segmentation approach in the telecommunication sector (Vietnam). SAGE Open, 11(2), 21582440211021584.

Birant, D. (2011). Data mining using RFM analysis. In Knowledge-oriented applications in data mining. IntechOpen.

Borg, L. (2021). Investigating training set and learning model selection for churn prediction in online gaming (Master's thesis, University of Malta).

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Breiman, L., & Cutler, A. (2005). Random Forests. Berkeley.

Brownlee, J. (2019). How to choose a feature selection method for machine learning. Machine Learning Mastery, 10.

Cai, J., Luo, J., Wang, S., & Yang, S. (2018). Feature selection in machine learning: A new perspective. Neurocomputing, 300, 70-79.

Cenggoro, T. W., Wirastari, R. A., Rudianto, E., Mohadi, M. I., Ratj, D., & Pardamean, B. (2021). Deep learning as a vector embedding model for customer churn. Procedia Computer Science, 179, 624-631.

Christy, A. J., Umamakeswari, A., Priyatharsini, L., & Neyaa, A. (2021). RFM ranking–An effective approach to customer segmentation. Journal of King Saud University-Computer and Information Sciences, 33(10), 1251-1257.

Çelik, O., & Osmanoglu, U. O. (2019). Comparing to techniques used in customer churn analysis. Journal of Multidisciplinary Developments, 4(1), 30-38.

Dachyar, M., Esperanca, F. M., & Nurcahyo, R. (2019, August). Loyalty improvement of Indonesian local brand fashion customer based on customer lifetime value (CLV) segmentation. In IOP Conference Series: Materials Science and Engineering (Vol. 598, No. 1, p. 012116). IOP Publishing.

De Caigny, A., Coussement, K., De Bock, K. W., & Lessmann, S. (2020). Incorporating textual information in customer churn prediction models based on a convolutional neural network. International Journal of Forecasting, 36(4), 1563-1578.

de Lima Lemos, R. A., Silva, T. C., & Tabak, B. M. (2022). Propension to customer churn in a financial institution: A machine learning approach. Neural Computing and Applications, 1-18.

Díaz Méndez, G. A. (2021). Data mining guided process for churn prediction in retail: from descriptive to predictive analytics (Doctoral dissertation).

Do Ruibin, K., & Vintilescu Borglöv, T. (2018). Predicting customer lifetime value: understanding its accuracy and drivers from a frequent flyer program perspective.

Dogan, O., Ayçin, E., & Bulut, Z. A. (2018). Customer segmentation by using RFM model and clustering methods: a case study in retail industry. International Journal of Contemporary Economics and Administrative Sciences, 8(1), 1-19.

Domingos, E., Ojeme, B., & Daramola, O. (2021). Experimental analysis of hyperparameters for deep learning-based churn prediction in the banking sector. Computation, 9(3), 34.

Ernawati, E., Baharin, S. S. K., & Kasmin, F. (2021, April). A review of data mining methods in RFM-based customer segmentation. In Journal of Physics: Conference Series (Vol. 1869, No. 1, p. 012085). IOP Publishing.

Fader, P. S., Hardie, B. G., & Huang, C. Y. (2020). Incorporating Time-Varying Covariates in the BG/NBD Model.

Fader, P. S., Hardie, B. G., & Lee, K. L. (2005). RFM and CLV: Using iso-value curves for customer base analysis. Journal of marketing research, 42(4), 415-430.

Fagerholm, F. (2022). Predicting customer churn in the financial industry.

Fan, C., Chen, M., Wang, X., Wang, J., & Huang, B. (2021). A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data. Frontiers in Energy Research, 9, 652801.

Farzanfar, E., & Delafrooz, N. (2016). Determining the customer lifetime value based on the benefit clustering in the insurance industry. Indian Journal of science and Technology, 9(1), 1-8.

Flach, P. A. (2016). ROC analysis. In Encyclopedia of machine learning and data mining (pp. 1-8). Springer.

García, S., Luengo, J., Herrera, F. (2015). Introduction. In: Data Preprocessing in Data Mining. Intelligent Systems Reference Library, vol 72. Springer, Cham. https://doi.org/10.1007/978-3-319-10247-4_1

Grange, S. K., Carslaw, D. C., Lewis, A. C., Boleti, E., & Hueglin, C. (2018). Random forest meteorological normalisation models for Swiss PM 10 trend analysis. Atmospheric Chemistry and Physics, 18(9), 6223-6239.

Guldogus, B. C., Abdullah, A. N., Ali, M. A., & Ozogur-Akyuz, S. (2023). Autoselection of the Ensemble of Convolutional Neural Networks with Second-Order Cone Programming. arXiv preprint arXiv:2302.05950.

Guldogus, B., Ozogur-Akyuz, S., & Ali, M. (07 2022). Using Second-Order Conic Programming as an Embedded Feature Selection Algorithm (Abstract), EURO 2022, Espoo, Finland. ISBN 978-951-95254-1-9.

Guliyev, H., & Tatoğlu, F. Y. (2021). Customer churn analysis in banking sector: Evidence from explainable machine learning models. Journal Of Applied Microeconometrics, 1(2), 85-99.

Hadden, J., Tiwari, A., Roy, R., & Ruta, D. (2008). Churn prediction: Does technology matter?. International Journal of Industrial and Manufacturing Engineering, 2(4), 524-536.

Hajipour, B., & Esfahani, M. (2019). Delta model application for developing customer lifetime value. Marketing Intelligence & Planning.

Hamzehei, A., Fathian, M., Gholamian, M., & Farvaresh, H. (2011). A new methodology to study customer electrocardiogram using RFM analysis and clustering. Management Science Letters, 1(2), 139-148.

Hardy, L. (2012). The evolution of participation banking in Turkey. Al Nakhlah Online Journal of Soutwest Asia and Islamic Civilization.

Ho, T. K., & Baird, H. S. (1995). Evaluation of ocr accuracy using synthetic data. In Proceedings of the 4th Annual Symposium on Document Analysis and Information Retrieval.

Huang, B., Kechadi, M. T., & Buckley, B. (2012). Customer churn prediction in telecommunications. Expert Systems with Applications, 39(1), 1414-1425.

Hwang, H., Jung, T., & Suh, E. (2004). An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. Expert systems with applications, 26(2), 181-188.

Iranmanesh, S. H., Hamid, M., Bastan, M., Hamed Shakouri, G., & Nasiri, M. M. (2019, July). Customer churn prediction using artificial neural network: An analytical CRM application. In Proceedings of the International Conference on Industrial Engineering and Operations Management, Pilsen, Czech Republic (pp. 23-26).

Jaiswal, J. K., & Samikannu, R. (2017, February). Application of random forest algorithm on feature subset selection and classification and regression. In

2017 world congress on computing and communication technologies (WCCCT) (pp. 65-68). IEEE.

Jasek, P., Vrana, L., Sperkova, L., Smutny, Z., & Kobulsky, M. (2019). Comparative analysis of selected probabilistic customer lifetime value models in online shopping. Journal of Business Economics and Management, 20(3), 398-423.

Jasek, P., Vrana, L., Sperkova, L., Smutny, Z., & Kobulsky, M. (2019). Comparative analysis of selected probabilistic customer lifetime value models in online shopping. Journal of Business Economics and Management, 20(3), 398-423.

Jasek, P., Vrana, L., Sperkova, L., Smutny, Z., & Kobulsky, M. (2019). Comparative analysis of selected probabilistic customer lifetime value models in online shopping. Journal of Business Economics and Management, 20(3), 398-423.

Javed, U., Ijaz, K., Jawad, M., Ansari, E. A., Shabbir, N., Kütt, L., & Husev, O. (2021). Exploratory Data Analysis Based Short-Term Electrical Load Forecasting: A Comprehensive Analysis. Energies, 14(17), 5510.

Jung, H., & Lee, B. G. (2020). The impact of transforming unstructured data into structured data on a churn prediction model for loan customers. KSII Transactions on Internet and Information Systems (TIIS), 14(12), 4706-4724.

Kabasakal, İ. (2020). Customer segmentation based on recency frequency monetary model: A case study in E-retailing. Bilişim Teknolojileri Dergisi, 13(1), 47-56.

Keramati, A., Ghaneei, H., & Mirmohammadi, S. M. (2016). Developing a prediction model for customer churn from electronic banking services using data mining. Financial Innovation, 2(1), 1-13.

Khan, Y., Shafiq, S., Naeem, A., Ahmed, S., Safwan, N., & Hussain, S. (2019). Customers churn prediction using artificial neural networks (ANN) in telecom industry. International Journal of Advanced Computer Science and Applications, 10(9).

Koike, K., Miyazawa, N., Machida, K., Kawamura, M., Takenaka, K., Sagawa, D., & Tanaka, K. (2022). Product Portfolio Optimization for LTV Maximization. In 人工知能学会全国大会論文集 第 36 回 (2022) (pp. 2S4IS2b04-2S4IS2b04). 一般社団法人 人工知能学会

Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2022). Customer churn prediction system: a machine learning approach. Computing, 104(2), 271-294.

Lee, E., Kim, B., Kang, S., Kang, B., Jang, Y., & Kim, H. K. (2018). Profit optimizing churn prediction for long-term loyal customers in online games. IEEE Transactions on Games, 12(1), 41-53.

Li, W., & Zhou, C. (2020, March). Customer churn prediction in telecom using big data analytics. In IOP Conference Series: Materials Science and Engineering (Vol. 768, No. 5, p. 052070). IOP Publishing.

Li, Y. (2022). Discrete Hyperparameter Optimization Model Based on Skewed Distribution. Mathematical Problems in Engineering, 2022.

Libai, B., Bart, Y., Gensler, S., Hofacker, C. F., Kaplan, A., Kötterheinrich, K., & Kroll, E. B. (2020). Brave new world? On AI and the management of customer relationships. Journal of Interactive Marketing, 51, 44-56.

Melian, D. M., Dumitrache, A., Stancu, S., & Nastu, A. (2022). Customer Churn Prediction in Telecommunication Industry. A Data Analysis Techniques Approach. Postmodern Openings, 13(1 Sup1), 78-104.

Mzoughia, M. B., & Limam, M. (2014). An improved BG/NBD approach for modeling purchasing behavior using COM-Poisson distribution. International Journal of Modeling and Optimization, 4(2), 141.

Naz, N. A., Shoaib, U., & Shahzad Sarfraz, M. (2018). A review on customer churn prediction data mining modeling techniques. Indian Journal of Science and Technology, 11(27), 1-27.

Perišić, A., & Pahor, M. (2020). Extended RFM logit model for churn prediction in the mobile gaming market. Croatian Operational Research Review, 249-261.

Prathima, J., Vaishnavi, M., Perumalraja, R., & Kamalesh, S. Finding the Most Profitable customer using CLV.

Praveena, M., & Jaiganesh, V. (2017). A literature review on supervised machine learning algorithms and boosting process. International Journal of Computer Applications, 169(8), 32-35.

Raeisi, S., & Sajedi, H. (2020, October). E-Commerce Customer Churn Prediction By Gradient Boosted Trees. In 2020 10th International Conference on Computer and Knowledge Engineering (ICCKE) (pp. 055-059). IEEE.

Rijnen, M., Hazen, B., & Vletter, C. (2018). Predicting Churn using Hybrid Supervised-Unsupervised Models (Doctoral dissertation, Master's thesis. Leiden University).

Rothmeier, K., Pflanzl, N., Hüllmann, J. A., & Preuss, M. (2020). Prediction of player churn and disengagement based on user activity data of a freemium online strategy game. IEEE Transactions on Games, 13(1), 78-88.

Sabbeh, S. F. (2018). Machine-learning techniques for customer retention: A comparative study. International Journal of Advanced Computer Science and Applications, 9(2).

Sabuncu, İ., Türkan, E., & Polat, H. (2020). Customer segmentation and profiling with RFM analysis. Turkish Journal of Marketing, 5(1), 22-36.

Sahoo, K., Samal, A. K., Pramanik, J., & Pani, S. K. (2019). Exploratory data analysis using Python. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8(12), 2019.

Santharam, A., & Krishnan, S. B. (2018). Survey on customer churn prediction techniques. International Research Journal of Engineering and Technology, 5(11), 3.

Shen, T. J., & Shibghatullah, A. S. B. (2022). Developing Machine Learning and Deep Learning Models for Customer Churn Prediction in Telecommunication Industry.

Singh, S. S., & Jain, D. C. (2017). Measuring customer lifetime value: Models and analysis. In Review of marketing research(pp. 37-62). Routledge.

Singh, S. S., Borle, S., & Jain, D. C. (2009). A generalized framework for estimating customer lifetime value when customer lifetimes are not observed. Qme, 7(2), 181-205.

Singh, Y., Pandit, Y., Joshi, N., & Avhad, V. PREDICTION OF CUSTOMER CHURN USING MACHINE LEARNING.

Sofaer, H. R., Hoeting, J. A., & Jarnevich, C. S. (2019). The area under the precision-recall curve as a performance metric for rare binary events. Methods in Ecology and Evolution, 10(4), 565-577.

Tiwari, A., Hadden, J., & Turner, C. (2010, March). A new neural network based customer profiling methodology for churn prediction. In International Conference on Computational Science and Its Applications (pp. 358-369). Springer, Berlin, Heidelberg.

Tukey, J. W. (1962). The future of data analysis. The annals of mathematical statistics, 33(1), 1-67.

Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., & Kim, S. W. (2019). A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector. IEEE access, 7, 60134-60149.

Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. Simulation Modelling Practice and Theory, 55, 1-9.

Vogelsang, A., & Borg, M. (2019, September). Requirements engineering for machine learning: Perspectives from data scientists. In 2019 IEEE 27th International

Requirements Engineering Conference Workshops (REW) (pp. 245-251). IEEE.

Wu, J., Shi, L., Lin, W. P., Tsai, S. B., Li, Y., Yang, L., & Xu, G. (2020). An empirical study on customer segmentation by purchase behaviors using a RFM model and K-means algorithm. Mathematical Problems in Engineering, 2020.

Xia, G., & He, Q. (2018, March). The research of online shopping customer churn prediction based on integrated learning. In 2018 International Conference on Mechanical, Electronic, Control and Automation Engineering (MECAE 2018) (pp. 259-267). Atlantis Press.

Xu, T., Ma, Y., & Kim, K. (2021). Telecom churn prediction system based on ensemble learning using feature grouping. Applied Sciences, 11(11), 4742

Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing, 415, 295-316.

Yoseph, F., & AlMalaily, M. (2019). New Market Segmentation Methods Using Enhanced (RFM), CLV, Modified Regression and Clustering Methods. International Journal of Computer Science & Information Technology (IJCSIT) Vol, 11.

Yu, T., & Zhu, H. (2020). Hyper-parameter optimization: A review of algorithms and applications. arXiv preprint arXiv:2003.05689.

Yulianti, Y., & Saifudin, A. (2020, July). Sequential feature selection in customer churn prediction based on Naive Bayes. In IOP Conference Series: Materials Science and Engineering (Vol. 879, No. 1, p. 012090). IOP Publishing.

Zainuri, N. A., Jemain, A. A., & Muda, N. (2015). A comparison of various imputation methods for missing values in air quality data. Sains Malaysiana, 44(3), 449-456.