VALUE SEGMENTATION OF REMOTELY ACQUIRED CUSTOMERS IN BANKING: A MODEL-BASED APPROACH

RECEP MUMCU

YEDITEPE UNIVERSITY

GRADUATE SCHOOL OF SOCIAL SCIENCES

İSTANBUL, 2023

VALUE SEGMENTATION OF REMOTELY ACQUIRED CUSTOMERS IN BANKING: A MODEL-BASED APPROACH

RECEP MUMCU

SUPERVISOR

ASSOC. PROF. UĞUR T. KAPLANCALI

SUBMITTED TO GRADUATE SCHOOL OF SOCIAL SCIENCES
IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS
IN
DEPARTMENT OF MANAGEMENT INFORMATION SYSTEMS

YEDITEPE UNIVERSITY

GRADUATE SCHOOL OF SOCIAL SCIENCES

İSTANBUL, 2023

PLAGIARISM

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.

Date: 02.08.2023

Name/Surname: Recep MUMCU

ABSTRACT

Innovations in information technologies lead to significant changes in the banking sector. While consumers' adaptation to digital services is accelerating, there are developments in the areas of customer experience and customer expectations. Increasing competition in customer experience field makes digital transformation a necessity for banks. The integration of banks with high technology in order to meet customer expectations ensures that new products and services are emerging in the field of digital banking. Remote customer acquisition has become one of the most important developments in digital banking, enabling banks to acquire new customers by overcoming geographical restrictions and without any physical interaction. The possibility that digital channels will be the most important customer acquisition channels for banks soon makes remotely acquired customers have a strategic importance for banks. It is important for banks to get to know these customers better acquired through digital channels without any physical interaction. Efforts to bring these customers to a value segment that will create more value for the bank are increasing significantly. In this study, it has been tried to emphasize the strategic importance of remote customer acquisition and online account opening process within the scope of digital transformation in the banking sector. Using data obtained from a private bank operating in Turkey, various machine learning models were applied to estimate the value segment of customers opening remote accounts and the results of the models were compared. Random Forest was the best performing machine learning model, which predicted customers' value segment with 76% accuracy.

Keywords: Customer value segmentation, Digital banking, Remote customer acquisition

ÖZET

Bilgi teknolojilerindeki yenilikler bankacılık sektöründe de önemli değişimlere yol açmaktadır. Tüketicilerin dijital hizmetlere adaptasyonu hızlanırken, müşteri deneyimi ve müşteri beklentileri alanlarında da gelişmeler yaşanmaktadır. Müşteri deneyimi alanındaki rekabetin artması bankaları dijital dönüşüme zorlamaktadır. Müşteri beklentilerine karşılık verilmesi için bankaların yüksek teknolojiye entegre olması, dijital bankacılık alanında yeni ürün ve hizmetler ortaya çıkmasını sağlamaktadır. Bankaların coğrafi kısıtları aşarak, herhangi bir fiziksel etkileşim olmaksızın yeni müşteri kazanmalarını sağlayan uzaktan müşteri edinimi, dijital bankacılık alanındaki en önemli gelişmelerden biri haline gelmiştir. Yakın gelecekte bankalar için en önemli müşteri edinme kanallarının dijital kanallar olması ihtimali, bu konsept ile kazanılan müşterileri bankalar için stratejik bir öneme sahip hale getirmektedir. Bankaların herhangi bir fiziki etkileşim kurmadan dijital kanallardan edindikleri bu müşterileri daha iyi tanımaları ve bu müşterileri daha fazla değer üretecek bir değer segmentine yükseltecek çalışmalar yapılması çok önemlidir. Bu çalışmada, bankacılık sektöründeki dijital dönüşüm kapsamında uzaktan müşteri edinimi ve dijital hesap açılış sürecinin stratejik önemini vurgulanmaya çalışılmıştır. Türkiye'de faaliyet gösteren özel bir bankadan temin edilen veriler kullanılarak, uzaktan hesap açan müşterilerin değer segmentini tahminlemek amacıyla çeşitli makine öğrenmesi modelleri uygulanmış ve modellerin sonuçları karşılaştırılmıştır. Müşterilerin değer segmentini %76 doğruluk oranıyla tahminleyen ve en iyi performansı gösteren makine öğrenmesi modeli Random Forest olmuştur.

Anahtar kelimeler: Dijital bankacılık, Müşteri değer segmentasyonu, Uzaktan müşteri edinimi

ACKNOWLEDGMENTS

I would like to express my gratitude to my thesis advisor, Associate Professor Uğur KAPLANCALI, whose valuable experience I benefited from throughout this study, for his advices and constructive feedback.

I also extend my heartfelt thanks to my dear wife Hazal MUMCU, who patiently stood by me and supported me throughout the writing process of my thesis.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

1. INTRODUCTION

Innovations in information technologies lead to significant changes in the banking sector. The inclusion of technology giants like Google, Amazon and Facebook in the financial ecosystem, which they are reshaping customer experience and customer expectations, has made it even more evident that digitalization is inevitable for the future of the banking sector, especially after the pandemic epidemic (Meinert, 2019). While the use of high technology in the banking sector is increasing day by day, there is an increase in digital banking solutions that will improve the customer experience. In this context, the concepts of remote customer acquisition, digital onboarding and branchless banking stand out as the latest developments and trending subjects in digital banking.

The fact that the internet and mobile devices have become more accessible to large masses leads to changes in consumers' behavior and expectations. Increasing consumer adoption of digital services is increasing consumers' demand for remote access to financial services without time and location constraints. (Thu Nguyen et al., 2020). This regard the concept of remote customer acquisition comes forward in terms of facilitating consumers' access to financial services. With this service, consumers can become bank customers without the need for any physical interaction only through mobile devices with internet access.

Remotely acquired customers are a relatively new and strategically important group for banks. These customers even perform their account opening transactions through digital banking channels instead of physical branches. This gives an important clue in terms of the possibility of this group to use digital channels intensively in accessing financial services after becoming a customer. Digital channels are expected to stand out as the most important customer acquisition channel for banks in the coming period. Banks should better understand these customers they gain without any physical interaction. It is also important that they work to raise these customers to segments with high added value that will create more value for the bank.

This study aims to establish a model that can predict customer value segmentation of remotely acquired customers. In this model to be established, the demographic characteristics and financial transactions of remotely acquired customers will be analyzed. In this way, it is aimed to identify the factors that have a significant impact on the promotion of these customers to the higher value segment. In addition to variables such as income level, age, education level of customers, financial transaction information including credit, credit card and debit card usage will be examined and these factors will be tried to be reached.

In addition, marketing suggestions will be presented to increase customers to higher value segments based on the factors determined as a result of this study. Strategies such as providing personalized services, special offers and promotions to strengthen customer relationships will be among the suggestions. In this way, banks will be able to improve the customer experience, increase customer loyalty and gain long-term profitability by taking customers to higher value segments.

In this study, it will be tried to emphasize the strategic importance of the concept of remote customer acquisition and online account opening process within the scope of digital transformation in banking. It will be pointed out that banks should know this customer group better and focus on these customers in order to achieve higher value. It is expected that the results obtained from this thesis will benefit banks in developing their marketing strategies, offering personalized offers to the target audience and increasing their competitive advantages.

2. DIGITAL BANKING

2.1. Digital Banking Concept

Digital banking refers to receiving financial services through a banking system using mobile phones or computers. Financial transactions such as depositing, withdrawing and transferring money are among these services (Ozili, 2018). In addition, the concept of digital banking is a mechanism that creates a distribution channel that enables financial products and services to be offered over virtual environments. This mechanism helps to better recognize customers and anticipate their needs quickly by using high technology. Customers can directly communicate with their banks and perform financial transactions using the internet and mobile phones (Cuesta, 2015).

As the internet and new generation mobile devices become more accessible to wider audiences at competitive prices, changes occur in consumers' habits and preferences. Consumers are increasingly adapting to using digital services. Consumers are now demanding financial services that are compatible with their social lives, regardless of time and location. The traditional products and services offered by banks are no longer able to meet the changes in consumer behavior and expectations. Consumers now demand financial products and services that they can access smoothly and comfortably from their mobile devices or tablets (Thu Nguyen et al., 2020).

The increase in the use of the internet and mobile devices worldwide has led to significant changes in the banking and finance sector and accelerated the development of digital banking. The concept of digital banking can be defined as the use of technology to provide financial services offered by banks more effortlessly and easily (Sardana & Singhania, 2018). The term digital banking is used to express the concepts of electronic banking, internet banking and online banking more generally. The increase in the demand for digital banking and the widespread use of it ensure that the financial services offered to consumers become more efficient due to the increased competition among banks. (Riza, 2019). Developments in digital banking provide benefits for the banking sector to provide traditional banking services at a lower cost and to increase the number of people who demand financial services with ease in terms of accessibility (Choi & Loh, 2023).

In addition, the concept of digital banking can also be expressed as the digitization of traditional banking products and services that customers could only access through physical bank branches in the past. These services include cash deposits, cash withdrawals, transfers, check transactions, financial product applications, credit transactions and invoice processing (Proctor, 2019). Digital banking technology has affected the essence of banking and has brought the financial services system online, where previously offline and employee-dependent services were at the forefront. Digital banking has been beneficial in terms of increasing the efficiency of financial services provided by using technology and reducing costs. Bank managements have shifted their attention to reducing operational costs and increasing digital channel effectiveness. In addition, the increasing number of digital banking service providers has made it important for banks to take steps to understand how their customers use digital services and to improve customer relations (Son et al., 2016).

Developments in digital banking services have led to a transformation in traditional banking in terms of reducing physical operations in bank branches. With the use of digital banking products and services, customers' obligation to go to physical bank branches for banking transactions has been eliminated. In addition, digital banking has led to a change in the banking sector in terms of providing services to customers abroad, unlike traditional banking, in addition to the benefit it provides to reduce costs (Choi & Loh, 2023).

2.2. Digital Banking Channels

A remarkable trend is observed in the increase of omnichannel services in financial consumption. With the opportunities provided by the developments in information technologies, there is an increase in the consumer offering of omnichannel financial services. In addition to new financial products and services emerging day by day, new channels are emerging where these products and services are offered (Cho et al., 2023).

It has been revealed that the increase in self-service channels offered to customers also brings an increase in the revenues of the banks. Self-service technology gives a positive result for banks as it allows them to better understand their customers' needs and adjust their services accordingly. Also, self-service channels increase the quality of service by providing control of the service offered to consumers. In this way, more benefits can be obtained from service interactions (Giovanis et al., 2019).

Although omnichannel services help increase customer satisfaction with the convenience they provide, these services also have negative aspects. Omnihannels can be challenging and inconvenient, especially for older consumers who struggle with digital complexity. For example, this group, who may have low attention and cognitive abilities in the decision-making process, may find it difficult to adapt to new products or channels. Therefore, age is one of the most important variables when searching for new channels (Cho et al., 2023).

2.2.1. ATM Banking

ATMs are one of the digital channels where many banking transactions can be made and 24/7 service can be obtained without going to bank branches. These systems, called Automated Teller Machine, are abbreviated as ATM.

It is possible to divide the ATMs into two types as in-site (located at branches) and off-site (located independently from branches). In-site ATMs are important in reducing the congestion at the toll booths. For this reason, the number of ATMs in branches varies according to the density of that branch. Off-site ATMs are generally tried to be placed in places far from branches and not easily accessible (Alpergin, 1990).

ATM systems, which were first used in the world in England in 1967, were widely used in Turkey towards the end of the 1980s. While there were 11,397 ATMs in Turkey in 2000, this number reached 51,941 as of 2018. The use of ATMs in Turkey has increased cumulatively every year and the average population per ATM has been 1769 as of 2018. (Egmir & Sagbas, 2021).

2.2.2. Internet Banking

Internet banking draws attention as one of the most preferred sales channels for consumers with its advantages such as convenience, offering many services together, low transaction costs and fast information exchange, and is spreading rapidly around the world (Usta, 2005).

Developments in information technologies have a significant impact on banking services, which are traditionally offered to customers through physical branches.

Thanks to the Internet, customers can perform their banking transactions as long as they have access to the Internet, regardless of time and place. Such services are called "online banking" or "internet banking" (Shao, 2007).

The combination of the convenience provided by the Internet and the demands of consumers has led to changes in the products and services offered by banks. Banks have become more customer-oriented than product-oriented. Banks have realized that in order to gain new customers while increasing customer loyalty, it is necessary to analyze customer expectations well and take actions in line with these expectations. The increase in the number of people using internet banking day by day shows that internet banking offers real value to bank customers (Karjaluoto et al., 2002).

As of December 2022, the number of individual customers registered to use internet banking in Turkey and who have logged into the system at least once has reached 91 million 137 thousand people. The number of individual customers who have logged into internet banking systems at least once in the last year is 25 million 147 thousand. Customers who have made a transaction at least once in the last 3 months (login to the system is sufficient) are classified as 'active' customers, and the number of active retail customers who have logged in to internet banking at least once in the period of October-December 2022 is 10 million 545 thousand people. This amount constitutes 12 percent of the total number of registered retail customers (TBB, 2022).

2.2.3. Mobile Banking

Due to the intense competition in the banking sector, the use of technology becomes important. Thanks to the technological tools used by banks, new products and services that we are not accustomed to in traditional banking are emerging. Today, technological developments have changed the way banking services are purchased and sold. With the developments in communication technologies, the use of mobile devices has increased. One of the newest distribution mechanisms offered to consumers in the banking sector is mobile banking (Suoranta & Mattila, 2004).

With the mobile banking system, it is possible to receive banking services from any location 24/7 from a terminal connected to the internet. The mobile banking channel enables customers to communicate with their banks via mobile devices. With the widespread use of smartphones by consumers, mobile banking applications have started to be used more frequently. Banks have rapidly adopted this system due to its potential benefits and have started to offer many banking services through this channel. Customers receive the most modern mobile banking services through an application they download to their smartphones to perform their banking transactions (Seyrek & Akşahin, 2016).

The most important reason why customers prefer mobile banking is that the system is not limited in terms of time and location (Mallat et al., 2004). According to Engen (2007), mobile banking eliminates the need to open a computer even for a simple transaction such as paying bills. These services are very attractive, especially for lifestyles that have a busy pace and do not have much time.

The total number of customers (individual and corporate) registered in the system to use mobile banking in Turkey and who have logged in at least once is 142 million 629 thousand as of December 2022. Of these, 92 million 90 thousand people (65 percent) are active users who have logged in at least once in the October-December 2022 period. As of October-December 2022, the total number of financial transactions made through mobile banking was 1 billion 816 million, and the total amount of these transactions was 11 trillion 307 billion TL. While 61% of these transactions consist of transfers such as EFT, money order and swift, 27 percent consists of investment transactions (TBB, 2022).

2.3. Branchless Banking

The increasing use of technology in the banking makes important developments. Especially in digital banking the concept of branchless banking emerges as a trending topic.

Skinner (2014) defines the concept of branchless banking as an alternative banking application that allows all banking transactions to be made without going to a physical bank branch. Traditionally, deposit banking works on the basis of physical distribution, while branchless banking works with an electronic distribution model.

Branchless banking has become an important development in terms of offering financial services to wider masses, especially in rural areas due to the presence of fewer physical bank branches. Thanks to the increasing use of this concept in developing countries, it increases the rate of financial inclusion. The advantages of this concept in reaching remote locations without the necessity of opening a physical branch are an important innovation for banks (Palaon et al., 2020).

It is an undeniable fact that the service quality offered by banks to their customers has increased in recent years, with the effect of technological innovations. But not all customers agree on the convenience. For customers unfamiliar with mobile technologies, the closure of physical branches will produce a negative impact. It should be noted that closing physical branches may cause these customers to switch to another bank or lose their motivation to purchase a product or service (Cho et al., 2023).

Currently, branchless banking services are mostly offered to individual and micro-scale companies in Turkey and around the world. Branchless banking services for commercial customers are rarely offered due to the more complex structure of products for the commercial segment.

3. REMOTE CUSTOMER ACQUISITION

The difficult process of becoming a customer can mean a negative experience for potential customers. While technology giants such as Google, Amazon and Facebook are revolutionizing customer experience and customer expectation, it can be difficult for banks to create a successful remote customer acquisition experience that will meet customer expectations with limited technology budgets. The real implementation of digitization in a bank needs to be seen as a cultural shift rather than the successful implementation of digital solutions. Digital transformation must impact the entire organization and onboarding is at the heart of it (Meinert, 2019).

We can refer to the process of acquiring new customers only through mobile devices, without direct physical interaction with customers, as remote customer acquisition. The important developments in technology, the more frequent use of digital platforms and the increase in the level of knowledge of customers in the field of digital literacy have paved the way for banks to reach customers remotely. With this concept, it has become possible for banks to expand their customer base by avoiding the restrictions caused by geographical factors. The concept of remote customer acquisition is not limited to customers opening accounts remotely. This includes using social media to attract and convert potential customers and leveraging many online channels for digital advertising. The convenience, flexibility and cost-effectiveness of remote customer acquisition benefits both banks and customers. In today's world where digital transformation is developing rapidly, remote customer acquisition has gained strategic importance for banks that want to expand their customer base and increase their competitiveness in national markets.

3.1. Digital Onboarding

It can be expressed as digital onboarding when customers register for a new product or service through digital channels and the identity verification processes take place during this process. In this process, distinctive information of customers, such as identity documents, is obtained through digital channels and verified. In this process, advanced technologies and digital tools are used. With digital onboarding, while complying with legal requirements, the need for physical documents is eliminated. Customers' onboarding experience becomes simpler thanks to a faster and more efficient account opening.

Consumers who want to become customers through internet banking and mobile banking must go through the process of recording their information through mobile applications called digital onboarding (Silva et al., 2023).

In the current financial ecosystem, it is very important to take attention of new customers while increasing the loyalty of old customers. Banks strive to create a personalized, simple and seamless digital customer acquisition and contracting experience. The ability to acquire customers through digital channels simplifies and facilitates a customer's first experience with the bank and significantly reduces the cost of customer acquisition. The pandemic epidemic that has affected the whole world recently has left banks and financial service providers faced with unforeseen challenges. While this period increased the importance of the digital onboarding process, there was an increase in the use of this concept. Digital customer engagement has become a buzzword in the industry as bank managers begin to realize the importance of digitalization (Master, 2022).

According to a survey by banking software company Temenos, only 31% of banks worldwide used digital onboarding in 2016, while in 2019 76% of banks actually embraced the digital onboarding experience. The banking industry has realized that they can no longer ask customers to come to branches for credit or credit cards. This is leading to a "great migration" towards digital onboarding (Meinert, 2019).

The digital onboarding process is the first step in an entire digital journey and has changed the way banks communicate with their customers. Digital onboarding has revolutionized the banking industry by facilitating every step of the customer acquisition process. It enabled an optimized user-friendly experience by removing physical controls from the process, without the need for new customers to deal with lots of paper and forms (Kommu, 2022).

The digital onboarding process is important to the banking industry for the following reasons:

- Enhanced customer experience: User experience is very important for a bank to have a strong reputation in the market. The digital onboarding process helps give customers a positive first impression of what they can expect from a bank, including how simple their banking process is. In today's world where the transition to the digital world is faster than ever, it is important that the services are simple to use and fast-flexible access to these services so that customers can have a seamless experience.

- Increased operational efficiency: Banks can streamline their operations and spend less time and money on paperwork thanks to the digital onboarding process. Digital onboarding enables banks to use their teams for more difficult responsibilities while minimizing reliance on branches and operating costs. Additionally, the digital onboarding process that helps reach customers in different locations can help banks grow their business.

- Reduced fraud: It helps banks to increase system security and reduce fraud with the methods offered by the digital onboarding process to easily verify the identity of the customer. In this process, it is possible to connect with customers, conduct vitality checks, conduct guided interviews and more, using reliable technology solutions.

According to the study carried out by IBSi Research, digital onboarding was chosen as the most important and trending banking product for the third consecutive period. In 2020, approximately 80% of the financial ecosystem has started to use new digital onboarding systems or have gone to upgrade their existent systems (Master, 2022).

3.2. Online Account Opening Process

The digital process in which financial consumers can become customers by opening an account in a bank over the internet or using mobile applications without the need to go to physical bank branches is the online account opening process. This simplified process provides convenience, efficiency and accessibility for users, allowing them to start and complete the account opening process from their mobile devices.

With the relevant regulation that came into force in May 2021, banks operating in Turkey were legally allowed to acquire customers remotely through this process.

In the past years, an increase has been observed in customers' preference for digital channels. This change has been accelerated with the pandemic. Banking sector is working on platforms that digitally help for both their current and prospective consumers (PR Newshire, 2020).

The mobile-first approach is extremely important for the remote account opening process. According to Tracey Dunlap, chief experience officer of Zenmonics (the firm approved by ABA Digital Bank for online account opening solution) and Chief Experience Officer Derek Corcoran of Temenos (banking software firm), the most important steps of the online account opening process can be listed as follows (Meinert, 2019):

- Build for mobile-first: When the online account opening process is carried out on mobile devices, it is important to simplify the process so that the texts can be read easily on a small screen and the account can be opened quickly and efficiently. In addition, banks should keep the Compliance departments on the table from the very beginning of the process in order to ensure the legal obligations of the process. Keeping communication at the forefront while acting within legal limits will help protect the customer experience in the process.

- Prioritize important questions: Customers are often asked for three different sorts of information during the process. Customers' first-hand knowledge, such as their names, addresses, and phone numbers, is the first of these. The second is information that is easy to access but that people frequently forget, like the serial number on an identity card. Customers can have a tougher difficulty locating information like the current mortgage balance, which is the last category. Early on in the process, asking challenging questions will boost subscription. Firstly, asking questions such as the name, e-mail, telephone number that the customers know and then asking the information they can obtain from the identity document will increase the completion rates by making the customers feel like they have made an investment in the process for their subsequent information requests.
- Leverage technological advancements: Thanks to technological developments, the areas where customers enter information manually have greatly reduced. For example, systems that can automatically fill in a form over the image of the relevant form have started to be used.
- Generate and nurture leads: Designing the application flow correctly is not only about reducing subscription but also collecting enough information to allow the

bank to track its customers. Some banks were able to follow up abandoned digital loan applications with phone calls and convert 40 percent of the people they called into borrowers. Designing the application flow properly involves not only reducing the number of drop off application but also gathering sufficient data about customers. Some banks were able to follow up the dropped of some digital application and they got 40% of those customers to complete the process.

- Allow customers to save applications: It is extremely important to save customers' applications and to ensure that those who leave the process unfinished can continue the process later. Most banks send an email as soon as a customer leaves the app with a link they can use to access the app later. People who leave the application unfinished are usually those who do not have the opportunity to deal with it at the moment. So sending an instant email to the person leaving the app might not be of much help. In this context, it would be more appropriate to send an e-mail to customers reminding them that they can complete their applications at regular intervals.
- Identify pain points: It is extremely important to monitor the banks' digital onboarding systems on a daily basis after they are operational. Continuous monitoring of the process may lead to the conclusion that the strategy should be changed if certain patterns are detected in the errors received and customer behavior.

Leveraging technology is important to prioritize important questions in order to adopt an overall mobile-first approach. Identifying and improving the hotspots are key elements to consider when designing a seamless digital engagement experience.

4. CUSTOMER SEGMENTATION

4.1. Customer Segmentation

Customer segmentation is the identification of different customer characteristics and grouping accordingly in order to understand the various characteristics of customers (McDonald, 2012). Companies often segment customers based on their contribution to company revenue, considering factors such as purchase volumes. Thanks to customer segmentation, companies can effectively manage the needs of different customer groups (Batt, 2000).

Customer segmentation is the division of a company's customer base into different groups based on various characteristics such as demographics, financial behavior, and credit risk profiles. With the segmentation concept, significant differences between customers are identified. In this way, it is aimed to group customers with similar needs, preferences and profitability potential. Banks can develop marketing strategies by analyzing customer segments. If the unique demands and expectations of each segment can be met, customer satisfaction, loyalty and profitability can be increased.

Most companies today do not know their customers well enough. For this reason, it is difficult to understand how much profit different customers bring in the short, medium and long term. Few companies know their customers well and can predict which customers are generating significant revenue for their businesses. Although some companies can classify and evaluate their customers according to criteria such as total turnover or gross profit, this practice is not sufficient. Still, it is better to have some degree of classification and evaluation than to have no knowledge of customers at all (Gel, 2004).

With the increase in the amount of data that companies can use and the developments in analytical processes, there have been developments in the field of customer segmentation in recent years. In traditional segmentation applications, segmentation was done based on demographic information. Today, more complex approaches based on advanced analytics and using machine learning algorithms are being used. Thanks to these new techniques used, not only demographic information, but also different sources such as transaction data, social media activities and customer interactions can be brought together. More complex analytical models can be used to use various datasets together. In this way, it is possible to create personalized product offers and marketing strategies. By using data-based customer segmentation methods, companies can use their resources more efficiently. At the same time, customer loyalty can be increased by targeting customers with segmentation studies.

Once you have a deep understanding of customer needs and preferences, it becomes easier to separate customers into different groups. With this information, it is possible to understand what satisfies the customers and even exceed customer expectations. Such insights can be used to improve products and services. Today, services to customers are as important as the product sold, and companies should prioritize increasing the level of service. Ultimately, customer segmentation facilitates decision making by guiding companies on how much importance they should place on different service levels for various customer groups (Buttle, 2009).

4.2. Customer Value Segmentation

Customer value segmentation is a strategic approach that involves dividing a firm's customers into different segments based on their perceived value and preferences. By analyzing the expectations of different customer segments separately, companies can establish marketing strategies to increase customer satisfaction and loyalty. With this approach, companies can identify their high-value customers and offer them personalized offers and exclusive benefits. In this way, it can be ensured that company resources are used more effectively in marketing activities for targeted segments. The concept of customer value segmentation can be useful to improve client relations, increase loyalty and increase long-term profitability.

The process of examining customer value is commonly employed to identify customers who generate profit and create targeted approaches to engage with them effectively (Irwin, 1994).

The use of various classification indexes is important for customer segmentation. Traditionally, these indexes include factors such as the client's age, occupation, gender, and family status. However, understanding customer buying behavior can be difficult if relying solely on these basic indices based on statistical population data. Analytical research on customer values enables applications to be made to identify, retain and develop loyal customers. This approach serves as a crucial strategy for scientifically segmenting customers, providing personalized services, and generating exceptional returns (Reichheld et al., 2000).

According to Kim et al. (2006), segmentation strategies based on customer value can be divided into three different categories.

- With a segmentation management in which only customer value is used, a list containing the values created by customers can be created and segmentation can be made according to the percentages in this list.
- In a segmentation method using customer value and other information, customers are placed in a multidimensional segmentation space. In this space, one axis constitutes the customer value, the other axis information axes.
- In the segmentation method to be applied using the components that make up the customer value, a segmentation field is created in the same size as the number of the components that make up the customer value.

In the first segmentation method, customers are sorted in descending order based on their customer value, and segmentation is performed using percentiles derived from this ordered list. The second method involves dividing customers into an n-dimensional segment space, where one axis represents customer value, and the remaining (n-1) dimensions represent other relevant information. In the last method, customers are divided into an n-dimensional segment space, where the n-dimensional space is defined by the components of customer value.

Foundation of segmentation lies in the understanding that not all consumers who engage with a company's products and services hold equal value. It is essential for companies to recognize the varying significance of their customers in order to sustain their market presence. Companies must strategically allocate their attention and resources by shifting focus from non-profitable consumers to those who generate higher profits. To ensure continued profitability, companies should prioritize customers who frequently engage with their products or services or make larger purchases. This approach allows for the creation of fruitful customer groups that contribute significantly to a company's success (Kim et al., 2006).

Customer value is a frequently talked about concept. From a company's perspective, customer value can be expressed as profitability and lifetime value from a customer. Customer value can be define as the formula below:

Customer value = Perceived customer benefit – Cost incurred on the customer

To increase customer value, you must analyze current consumer experience, experiment with pricing, analyze customer data, identify loyal customers, and segment customers by value and loyalty. Segmentation helps your marketing efforts with targeted communication, focusing on the best channels, deepening relationships, experimenting with price options, and increasing your revenue (Jackson, 2021).

5. MATERIALS AND METHODS

This section contains general information about the data set and machine learning methods used in the study. Information about data sources and properties of data is given. The descriptive statistics are summarized by explaining the features in the dataset. The data preprocessing and feature engineering steps performed in the study are explained. In addition, the feature selection stage and machine learning algorithms applied in the study are discussed. Brief explanations were made about the selected algorithms and their suitability for research purposes. Lastly, the measurements used in the evaluation of the results will also be included.

5.1. Data Description

The data set used in the study was taken from a private bank operating in Turkey. The data set includes the demographic information of people who open accounts remotely and the financial transaction information of these customers in a certain time period. The data set contains the value segment information that the bank follows related customers and this value is used as the dependent variable in the model. After using customers' demographic information and financial transaction data as independent variables, it is aimed to predict the value segments of these customers.

5.2. Data Properties

The data set obtained includes the demographic information of 73,548 customers who opened accounts remotely in the first half of 2022 and the financial transaction data of these customers for the second half of 2022. The customer Value Segment, which is used as a dependent variable in the modeling phase, shows in which value segment the relevant customer is followed by the Bank. The table below provides a comprehensive overview of all the variables included in the dataset with their respective properties.

Table 1

List of Variables in the Dataset

Variable Name	Description	Variable Type	Data Type
Value Segment	The value segment of the customer can be categorized as A, B, C, or P. Segment A represents the highest value segment, while segment P indicates the passive value segment.	Categorical	String
Average Deposit Balance	The average monthly deposit balance of the customer	Continuous	Integer
Average Debit Card Expenditure	The average monthly spending amount made by the customer using their debit card.	Continuous	Integer
Average Credit Card Expenditure	The average monthly spending amount made by the customer using their credit card.	Continuous	Integer
Credit Card Limit	The credit limit assigned to the customer's credit card.	Continuous	Integer
Credit Risk	The total amount of the customer's credit risk in the bank, including interest	Continuous	Integer
Credit Limit	The total credit limit of the customer at the bank	Continuous	Integer
Productivity Amount	The average monthly productivity amount obtained by the bank from the customer.	Continuous	Integer
ADC Transaction Count	The number of digital banking transactions conducted within a specific time period.	Discrete	Integer
Age	The age of the customer	Discrete	Integer
Gender	Gender of the customer, assigned as 1 for males and 0 for females	Binary	Integer
Marital Status	Marital status of the customer, assigned as 1 for married individuals and 0 for single individuals.	Binary	Integer
Years of Schooling	The number of years of education completed by the customer	Discrete	Integer
Income	Monthly income of the customer	Continuous	Integer

5.3. Data Pre-Processing and Feature Engineering

Before performing data analysis, a set of operations known as preprocessing are conducted to the raw data. This section's major goal is to make sure the data is accurate, dependable, consistent, and understandable. Because of this, measures for feature engineering and data preparation were carried out before building the models. The steps will be thoroughly explained in this section.

5.3.1. Feature Engineering

The columns in the dataset that contain category variables have a string data type. It was chosen to change these categorical variables into dummy variables to guarantee that all values in the dataset are numeric. Through the use of dummy variables, the string data had to be recoded and converted into a numeric format.

The following steps were taken in this regard:

First of all, the string type variables in the table below are converted to numeric values as 1 or 0.

Table 2

List of Variables with String Value

Variable Name	Original Values	New Values
Gender	Male, Female	1, 0
Marital Status	Married, Single	1, 0

Secondly, dummy variables were created to transform certain categorical variables into numerical variables. For instance, a dummy variable named Years of Schooling was created for the variable Education Level. It represents the school year corresponding to the education level required to complete the customer's education.

Table 3

List of Variables for which Dummy Variables were Generated

Variable Name	New Values
Education Level	Primary School, Junior High School, High School, Associate Degree, Bachelor's Degree, Degree, Doctoral Level

5.3.2. Feature Reduction

Data reduction has been applied to the dataset so that only the required features are included. Since there were new dummy variables created for them instead of the categorical features specified in Table 3, they became unnecessary and were excluded from the data set.

Model performance may be adversely affected by multicollinearity, which may result from a high degree of feature correlation. To avoid this situation, it is important to evaluate the correlation coefficients between features and remove highly correlated features. By looking at the correlation coefficients between the features, the features with more than 70% correlation between them were extracted.

- Credit Card Limit
- Credit Limit

It is aimed to improve model's performance with eliminating these highly correlated features and prevent the multicollinearity problem.

The Figure 1 below presents the correlation matrix, illustrating the relationships and correlation coefficients between the variables.

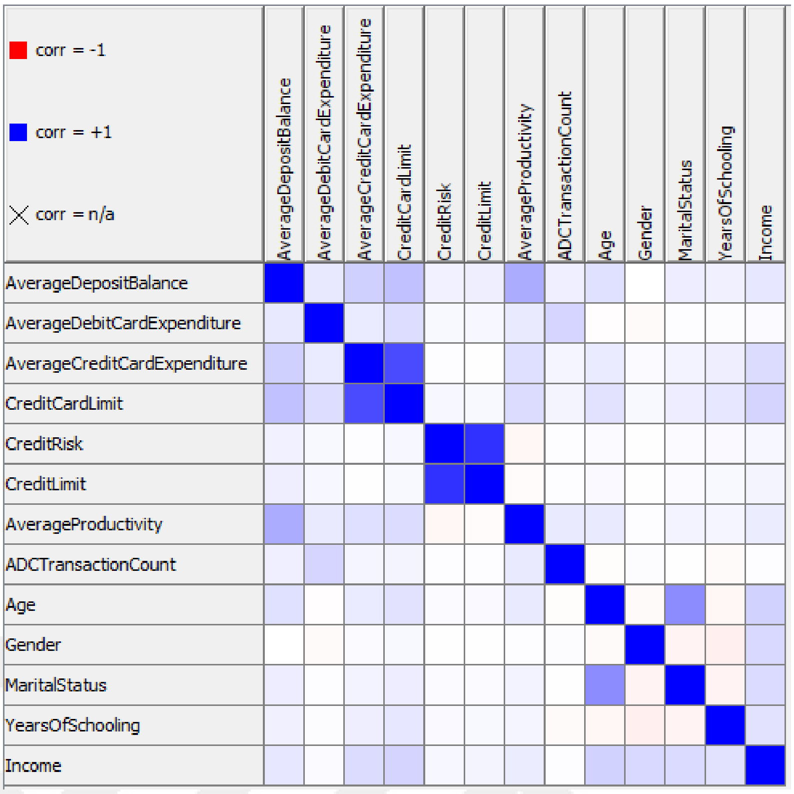

Figure 1. The correlation matrix.

Variables that were determined to have no significant effect on the performance of the model and were of low importance were eliminated using the feature selection method. Feature selection, used as a statistical approach, aims to reduce the complexity of the model and focus only on variables that have a significant impact on the model. An attribute selection cycle was used in this process. By selecting variables and removing those with low impact on the model, the feature selection cycle allowed more variables to be removed iteratively. This iterative process resulted in a more streamlined and optimized structure for the model. It was determined that the eliminated variables did not provide statistically significant contributions and did not have any effect on increasing the performance of the model. Thus, the feature selection method facilitated a more reliable and focused operation of the model.

The table below presents the number of features used in the model and the accuracy of the model, as demonstrated by the feature selection loop.

Table 4

Result Table of Feature Selection Loop

Number of Features	Accuracy of Model
11	0.769
9	0.769
10	0.766
9	0.762
10	0.762
10	0.760
8	0.759
7	0.758
8	0.754

The following features were identified as having low significance in the model after applying the feature selection loop and were subsequently removed from the model.

- Average Debit Card Spend
- Gender
- Years of Schooling
- Income

5.3.3. Data Descriptive

In this chapter, some general information about the dataset will be given. It was checked whether there were missing variables in the data set, and no missing variables were found in the data set.

The Table 5 below shows the statistical summaries of the dataset.

Table 5

Descriptive Statistics of Some Features in the Dataset

Feature Name	Min.	Max.	Mean	St. Dev.
Average Deposit Balance	0	9,847,934	61,480	257,204
Average Credit Card Expenditure	0	143,446	233.3	1,620
Credit Risk	0	2,999,999	741.7	16,458
Average Productivity	-112,383	97,951	224.7	1,671
ADC Transaction Count	0	4,461	22.1	63.9
Age	18	86	30.5	9.5
Marital Status	0	1	0.47	0.5

5.4. Selection of the Model

While selecting the model, the model that finds the patterns in the data most accurately will be determined. This step is very important in terms of making accurate predictions and providing reliable predictions for the model to be used at the end of the study. The models used in the study were evaluated and the one with the best performance and interpretability were selected.

The most important indicator of a model's predictive power is how well the model captures patterns in historical data and can accurately predict future values. Therefore, when evaluating the performance of models, it is important to use not only historical data, but also unseen data. For this reason, train-test splitting was performed on the data. To help the model understand patterns and correlations within the data, approximately 80% of the data was randomly selected to be used as a training set. The remaining 20% test set was used as a completely independent data set to evaluate the predictive capacity of the model.

A solid framework for model selection was tried to be established by making a train-test distinction. With this method, it is aimed to facilitate the selection of the most suitable model for the study. An objective evaluation has been made about the prediction performance of the model on unseen data.

Machine learning (ML) models are used to solve many problems from simple to complex and outperform traditional techniques in terms of speed and accuracy. Machine learning enables predictions, classifications, and data-driven decision making. This enables companies to leverage historical data to understand consumers, optimize product strategies, and plan future activities more efficiently (Miklosik et al., 2019). The machine learning algorithms used during the modeling process will be briefly discussed in this section.

Decision tree methods are generally used for classifications. It provides top-down hierarchical mapping of organizational and operational links among given features and distributed judgments (Wang et al., 2020). Machine learning algorithms used for classification include Random Forest (FR), Decision Tree (DT) and Multilayer Perceptron (MLP) (Koranga et al., 2022).

5.4.1. Random Forest Algorithm

“Random Forest is a machine learning algorithm that uses multiple decision trees to make predictions”. Random forests consist of multiple decision trees that are aggregated together. While individual decision trees are easily interpretable, this interpretability is sacrificed in random forests due to the ensemble approach. However, this trade-off comes with the advantage that random forests generally outperform individual decision trees in prediction tasks (Breiman, 2001). One notable improvement of random forests over decision trees is their ability to estimate the error rate more accurately. In fact, it has been mathematically proven that as the trees in the random forest multiply, the rate of inaccuracy converges consistently. This convergence provides confidence in the predictive performance of random forests and highlights their effectiveness in handling complex datasets.

When a machine learning model is overly complex, it will do well on training data, but cannot generalize to new and untested data. When this happens, the model underperforms on new data as it memorizes the training data instead of understanding the underlying patterns. "overfitting" is the term for this problem (Dietterich, 1995). To create each tree in a random forest, subgroups of the data and characteristics are chosen at chance. As a result, overfitting is decreased.

Linear regressions assume a linear relationship between variables. But random forests can better handle nonlinear situations and generally produce better predictions. The limitation of linear regression that arises when the number of variables exceeds the number of observations is overcome, especially with random forests, which can produce more effective results for medium to large datasets.

While individual decision trees are easy to interpret, this interpretability is compromised in random forests due to the aggregation of multiple trees. However, this trade-off allows random forests to excel in prediction tasks. In fact, compared to decision trees, random forests provide more accurate estimates of the error rate. Furthermore, it has been mathematically demonstrated that the error rate consistently converges as the total of trees in the random forest increases (Breiman, 2001).

Another important feature of Random Forests is that it can process high-dimensional data and calculate the relative value of each feature. This makes it an effective tool for many different tasks, including feature selection, regression, and classification.

The main structure of the random forest algorithm represented in Figure 2.

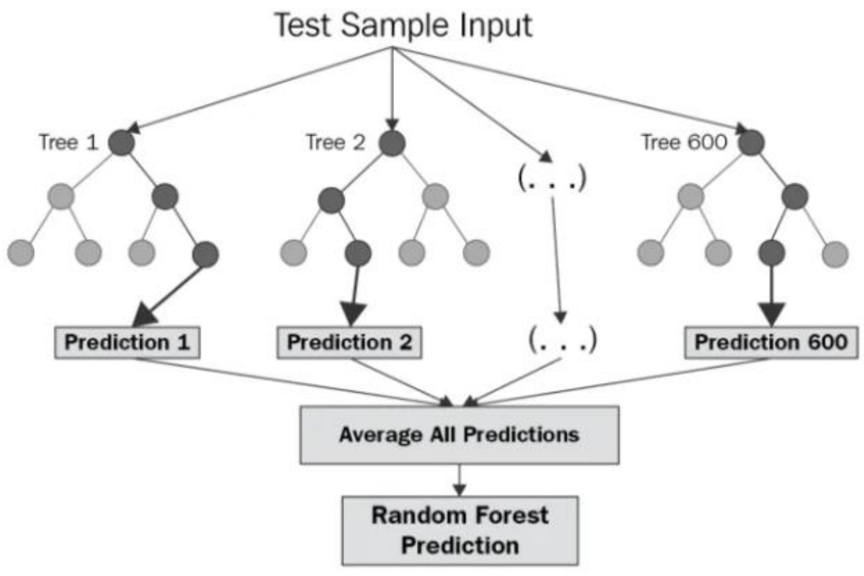

Figure 2. Typical representation of a random forest regressor.

5.4.2. Decision Tree Algorithm

The decision tree algorithm is a tree-based algorithm that is widely used in both regression and classification tasks.

A widely used supervised learning technique for constructing classification systems or developing prediction algorithms for the target variable is decision trees. Using internal nodes, leaf nodes, and root nodes, this technique branches a population. Each edge is a test result, each leaf node represents a class label, and each inner node represents a test on a feature. The algorithm can handle very large and complex datasets well without the need for a complex parametric framework. When the number of examples are high enough, the study's data can be separated into training and testing datasets. A decision tree model is created using the training dataset, and the testing dataset is utilized to establish the ideal tree size for the best ending model. (Song & Lu, 2015).

Some key features of decision trees are as follows; (Gepp et al., 2018)

- Flexibility: It can be used for classification and regression tasks, also different types of variables can be used depending on the data type.
- Easy to interpret: The tests performed at each node are classified due to its hierarchical structure. In this way, model results and decision trees can be easily interpreted.
- Non-parametric property: When creating data-based decision rules without requiring complex parametric structures, simple decision rules are used instead of data-specific parameters.
- Data-based predictions: Decision tree working with data-applied tests and decision rules can make accurate predictions based on what is learned from the training data set.
- Optimization of Dataset: Data management optimization is achieved by dividing the dataset into segments and dividing the decision trees into smaller parts.

Figure 3 illustrates how the decision tree approach is represented.

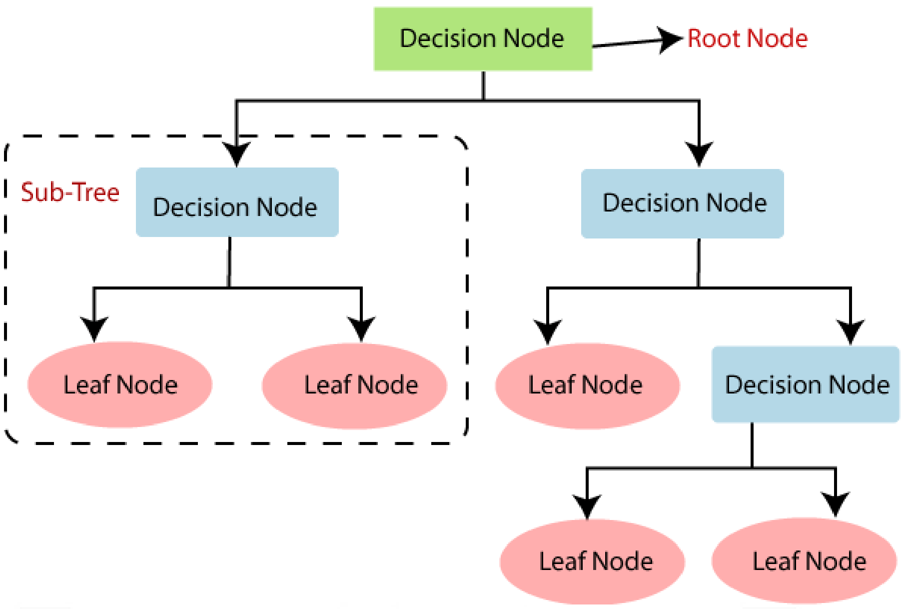

Figure 3. Simplified structure of the decision tree algorithm.

5.4.3. Multilayer Perceptron Neural Network

A method for creating artificial neural networks (ANN) and used generally in classification is called a multilayer perceptron (MLP). MLP can map input and output vectors in a nonlinear way. This approach utilizes weights and activation functions along with simply connected neurons. The usage of a learning process, in which the weights are modified by repeatedly presenting the training dataset, is one of the multilayer perceptron's key characteristics. Using error signals, it also learns to approximate the output to the goal output. These characteristics allow the multilayer perceptron to generalize to new input and model complex functions (Gardner and Dorling, 1994).

With the right choice of coupling weights and transfer functions, MLP may estimate whatever measurable, smooth function among input and produced vectors. In order to give the appropriate input-output mapping, MLP, which has the ability to learn using education data, modifies the strengths in the network during training. The MLP learns under supervision and modifies the weights based on the error signals. When correctly trained, MLP can predict outcomes from novel and unheard-of input data (Hornik et al., 1989).

The structure and operation of the brain and its neurons serve as the model for how neural networks work. This method uses sensors to deliver numerical signals in the place of neurons. Input, output, and occasionally a hidden layer that affects the result are all layers that make up artificial neurons. There is no interconnection between neurons in the same layer. In a neural network, learning is a continuous process that involves maximizing the weighted connections between layers to identify the best answer (Stangierski et al., 2019).

The representation of the multilayer perceptron neural network method is shown in figure 4.

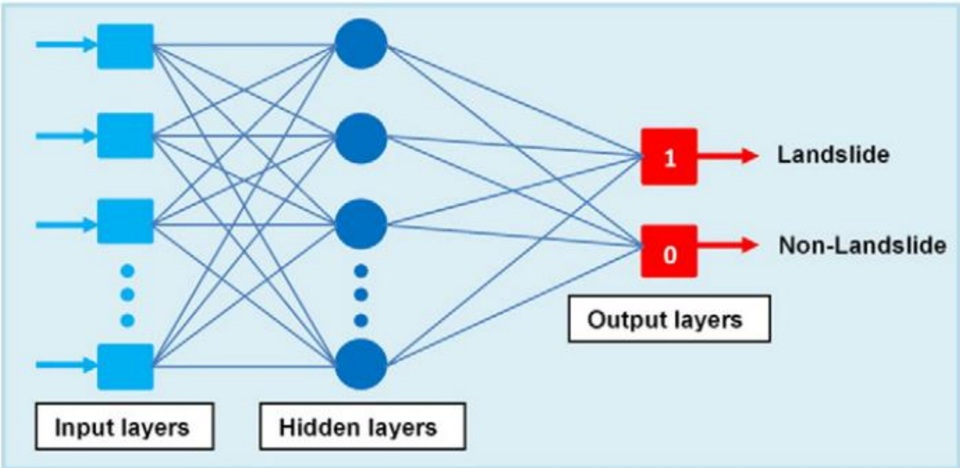

Figure 4. Structure of a Multilayer Perceptron Neural Networks.

5.5. Evaluation Metrics for Assessing Model Performance

In the modeling process, three methodologies were employed. Their outcomes were assessed using specific indicators. The results from the test data are typically used to generate a confusion matrix, which is used to evaluate the effectiveness of the model in classification algorithms.

The definitions of the values in the confusion matrix are shown in the following table:

Table 6

Confusion Matrix

	True Positives (TPs)	False Positives (FPs)	True Negatives (TNs)	False Negatives (FNs)
Actual Class (1)	TPs 1	FPs 1	TNs 1	FNs 1
Actual Class (2)	TPs 2	FPs 2	TNs 2	FNs 2
Actual Class (3)	TPs 3	FPs 3	TNs 3	FNs 3
Actual Class (4)	TPs 4	FPs 4	TNs 4	FNs 4

- True Positives (TPs): The count of positive records the model correctly predicted.
- False Positives (FPs): The count of negative records the model incorrectly predicted as positive.
- True Negatives (TNs): The count of negative records the model correctly predicted.
- False Negatives (FNs): The count of positive records the model incorrectly predicted as negative.

Class statistics like accuracy and precision are calculated based on the results of the confusion matrix. Following is a definition of class statistics used in the model assessment:

- Accuracy: The proportion of correctly classified records (both true positives and true negatives) between the entire count of records.
  i.e. (TP + TN) / (TP + TN + FP + FN)
- Sensitivity: The proportion of correctly identified positive records among all

actual positive records.

i.e. TP / (TP + FN)

- Specificity: The proportion of correctly identified negative records among all actual negative records.

i.e. TN / (TN + FP)

- Precision (positive predictive value): The accuracy of the model's predictions for positive records.

i.e. TP / (TP + FP)

- Recall: The rate at which the model accurately predicts positive records.

i.e. TP / (TP + FN)

- F-measure: Harmonic mean of the precision and recall values.

i.e. 2*(Recall*Precision) / (Recall+Precision)

- Kappa: It accounts for the potential that agreement could come about by chance alone. The range of kappa values is -1 to 1, with values closer to 1 denoting a high level of agreement above chance, values around 0 or negative denoting low agreement or worse than chance agreement, and a value of 0 denoting no agreement above chance.

6. MODEL BUILDING AND EVALUATION

Three different methodologies were used in the study. Among these methodologies, it was aimed to find the model that best predicts the value segment of customers. While various methods were explored, the focus was on determining the model that gave the optimum result. In this section, the models applied in the study and their results are explained. Each model was evaluated using specific performance measures and information on the results was included. Comparison of the models is also included in order to determine the model with the best prediction performance.

The KNIME Analytics Platform, a comprehensive software tool, was used in the modeling process. The platform has various data processing, analysis and machine learning features. These features facilitate the steps of data analysis, feature selection, and application of predictive models. The platform's integrated machine learning tools enabled the implementation and evaluation of different modeling techniques. The steps in the modeling process and the general scheme of the model are shown in the image below.

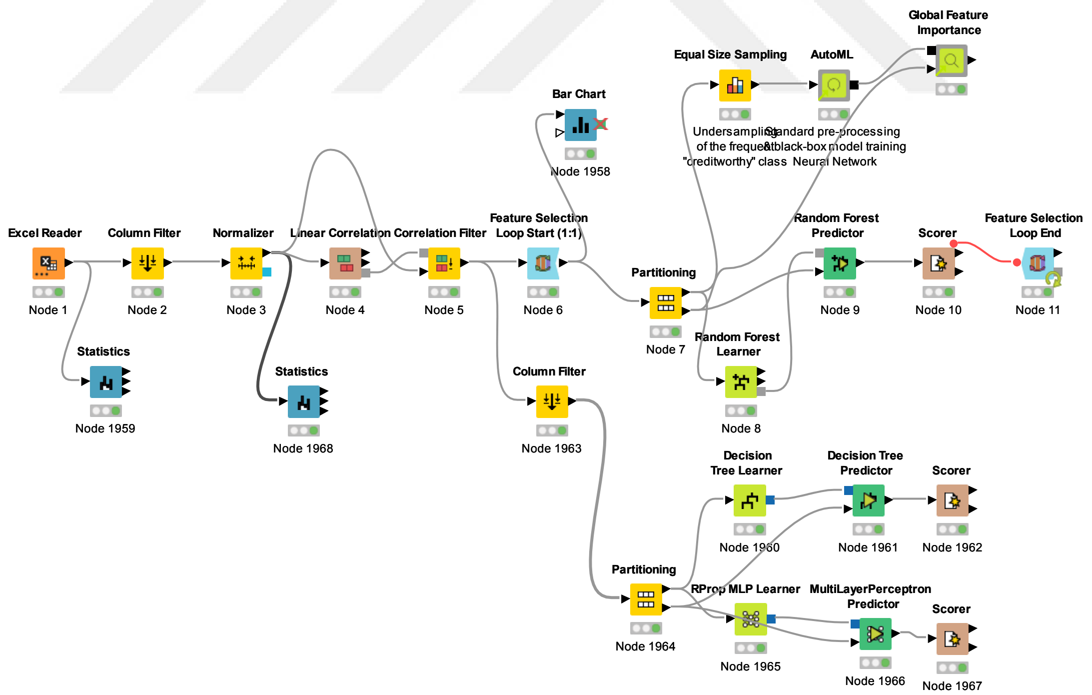

Figure 5. The KNIME-based modeling process.

6.1. Model Results

6.1.1. Random Forest

The global importance values of the features in the random forest model are given in Figure 6.

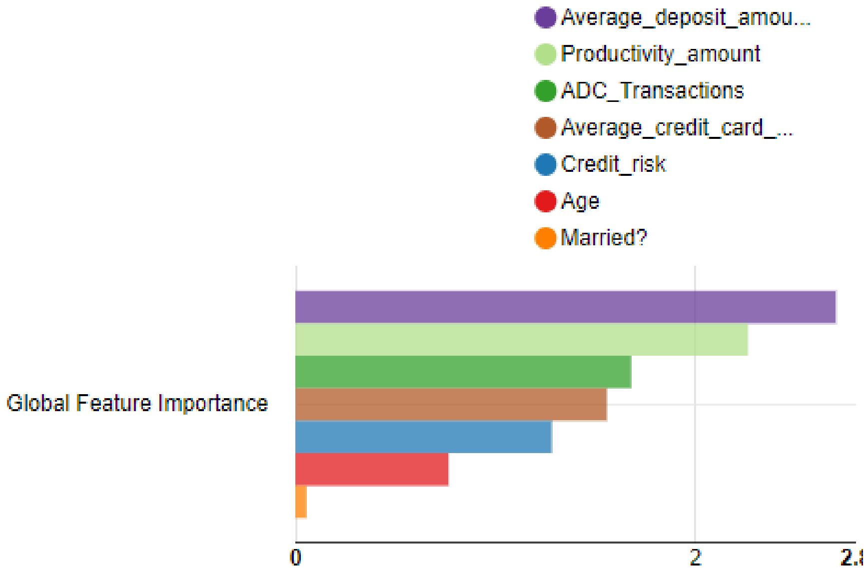

Figure 6. The global importance values of the random forest model.

The global importance values of the features obtained as a result of the random forest model give an idea about the importance levels of the variables in the data set in terms of the model. The contribution of each variable to the model can be measured over these values. The visual in Figure 6 visually represents the significance levels of the variables. Variables with high significance in the graph have a greater impact on the model. Likewise, variables with low significance have a smaller effect on the model. With this graph, it can be understood which variables contribute more to the performance of the model. It is an important indicator for the variables to be selected to increase the accuracy of the model. According to the results of the random forest model, the three features that contribute the most to the accuracy of the model are the *Average Balance of Deposit*, *Efficiency Amount* and *ADC Transactions* variables, respectively. These variables are the most effective factors in estimating the target variable. The significance levels of these variables are higher in the model. This shows that they play a very important role in determining the accuracy of the model's predictions. By focusing on these features, valuable information can be gained and more informed decisions can be made regarding the target variable.

The confusion matrix, shown in the table below, presents the results of the random forest model:

Table 7

Confusion Matrix of the Random Forest Model

Value Segments	True Positives	False Positives	True Negatives	False Negatives
Class A Actual	985	249	13,140	336
Class B Actual	1,740	939	11,030	1,021
Class C Actual	2,704	1,540	9,348	1,118
Class P Actual	5,742	831	7,053	1,084

Based on the data provided in the table, the value segment of 11,151 (TPs) out of 14,710 customers was accurately predicted. The overall accuracy of the model is 75.8%.

6.1.2. Decision Tree

The following table presents the results of the decision tree model used in the study:

Table 8

Confusion Matrix of the Decision Tree Model

Value Segments	True Positives	False Positives	True Negatives	False Negatives
Class A Actual	865	320	13,086	439
Class B Actual	1,464	1,241	10,695	1,310
Class C Actual	2,121	1,545	9,437	1,607
Class P Actual	5,752	1,402	6,404	1,152

According to the data presented in the table, the value segment of 10,202 (TPs) out of 14,710 customers was correctly predicted. The model achieves an overall accuracy of 69.4%.

6.1.3. Multilayer Perceptron

The following table presents the results of the multilayer perceptron model employed in the study:

Table 9

Confusion Matrix of the Multilayer Perceptron Model

Value Segments	True Positives	False Positives	True Negatives	False Negatives
Class A Actual	944	249	13,157	360
Class B Actual	1,630	849	11,087	1,144
Class C Actual	2,538	1,547	9,435	1,190
Class P Actual	5,864	1,089	6,717	1,040

According to the data presented in the table, the value segment of 10,976 (TPs) out of 14,710 customers was correctly predicted. The model achieves an overall accuracy of 74.6%.

6.2. Model Comparison

The accuracy statistics of each model used in the study, in which the value segment of the customers is used as the target variable, are calculated, and given in the tables below.

Table 10

Random Forest Model Results in terms of Class Statistics

Model Results	Specificity	Sensitivity	Precision	F-measure	Accuracy	Cohen's Kappa
Class A	**0.981**	**0.746**	**0.798**	**0.771**	**0.758**	**0.642**
Class B	**0.922**	**0.628**	**0.647**	**0.637**		
Class C	**0.859**	**0.707**	**0.637**	**0.670**		
Class P	**0.895**	**0.841**	**0.874**	**0.857**		

Table 11

Decision Tree Model Results in terms of Class Statistics

Model Results	Specificity	Sensitivity	Precision	F-measure	Accuracy	Cohen's Kappa
Class A	**0.976**	**0.663**	**0.730**	**0.695**	**0.694**	**0.540**
Class B	**0.896**	**0.528**	**0.541**	**0.534**		
Class C	**0.859**	**0.569**	**0.579**	**0.574**		
Class P	**0.820**	**0.833**	**0.804**	**0.818**		

Table 12

Multilayer Perceptron Model Results in terms of Class Statistics

Model Results	Specificity	Sensitivity	Precision	F-measure	Accuracy	Cohen's Kappa
Class A	**0.981**	**0.724**	**0.791**	**0.756**	**0.746**	**0.620**
Class B	**0.929**	**0.588**	**0.658**	**0.621**		
Class C	**0.859**	**0.681**	**0.621**	**0.650**		
Class P	**0.860**	**0.849**	**0.843**	**0.846**		

As seen in the tables above, the random forest model has the highest accuracy values in the class statistics of the models. Accuracy statistics calculated for the A value segment were especially examined, since it is more important to correctly estimate the A value segment for the purpose of the study. The overall model accuracy was 75.8% and the sensitivity value for the A value segment was calculated as 74.6% and the random forest model produced better performance compared to other models. Other metrics such as f-measure and Cohen's kappa are also relatively higher in the random forest model than in other models.

7. CONCLUSION

In this thesis, by focusing on the strategic importance of remote customer acquisition in the banking sector, a model is presented that can help the customer group that remotely acquired to reach a more valuable segment. In this context, a model is presented to estimate the value segment of customers who open remote accounts. During the modeling phase, the demographic characteristics and financial transactions of customers who open remote accounts were analyzed and factors that would move customers to a more valuable segment were identified.

In the modeling phase, in which three different machine learning algorithms were applied, the results of the models were close to each other. However, the random forest model showed the best performance. Although the number of customers in different value segments was not close to each other in the data set to which the model was applied, the final accuracy of the model was 76%. In line with the purpose of the study, since suggestions will be presented to help banks take their customers to the higher value segment, the model results of the A segment, which is the highest value segment, were also paid attention. The sensitivity value of the model was realized as 75% for this segment. In other words, with this model, the value segment of the customers can be predicted correctly at a rate of 76%.

According to the results of this study, the most important features in the random forest model are as follows:

- *The Average Deposit Balance*: represents the average deposit balance in customers' accounts. Deposit Balance is an important factor in promoting customers to a more valuable segment. This variable indicates the financial strength and level of financial accumulation of the client. Customers with high deposit balances are candidates for a more valuable customer segment for the bank.
- *The Productivity Amount*: represents the total revenue generated by a customer to the bank through their financial transactions. Another important variable that has a high impact on the customer's value segment. Customers with a high amount of efficiency bring more revenue to the bank and are potential candidates for inclusion in the higher value segment.
- *The ADC Transaction Count*: represents the number of financial transactions

performed by customers through digital channels. It is a variable that reflects the customer's digital activities. Customers who perform a large number of transactions through digital banking channels are considered active users and show great interest in digital banking. These customers are more likely to belong to a higher value segment due to their inclination and active use of digital banking services.

These above-mentioned variables appear as factors that have a high impact on the value segment of customers. Banks can use these variables to improve their marketing strategies. For example, special campaigns can be offered to customers with low deposit balances so that they can reach higher deposit balances. Thus, an increase in the value segments of the customers in this group can be expected. On the other hand, customers with high deposit balances can be offered special services or investment opportunities to increase their loyalty to the bank. Similarly, banks can offer special campaigns or discounts to their customers with high productivity levels. In this way, it is possible to work on improving the relations of the customers in this group with the bank and for these customers to perform more transactions. The high number of transactions made by customers through digital channels also carries these customers to the higher value-added segment. For this reason, investments can be made to improve the customer experience in digital channels in order to increase the number of transactions performed by customers through digital channels.

In conclusion, this study aims to present a model and insight into the strategic steps that banks can take for remote customer acquisition and value segmentation concepts. Banks can work on target marketing strategies to leverage customers' value segments by focusing on the above-mentioned variables that have a significant impact on customers' value segments. These strategies can help banks improve the customer experience, strengthen customer loyalty and achieve long-term profitability.

REFERENCES

Alpergin, P., (1990). *Bireysel Bankacılık*, Türkiye Bankalar Birliği Yayınları, Yayın no:160, Ankara.

Batt, R. (2000). Strategic segmentation in front-line services: Matching customers, employees and human resource systems. *International Journal of Human Resource Management*, 11(3), 540-561. https://doi.org/10.1080/095851900339756

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

Buttle, F. (2009). *Customer Relationship Management: Concepts and Technologies* (2nd ed.). Elsevier Ltd.

Cho, S., Lee, Z., Hwang, S., & Kim, J. (2023). Determinants of bank closures: What ensures sustainable profitability in mobile banking? *Electronics (Basel)*, 12(5), 1196. https://doi.org/10.3390/electronics12051196

Choi, H. S., & Loh, R. (2023). Physical frictions and digital banking adoption. *Management Science*, 1-53. [Research Collection Lee Kong Chian School of Business].

Cuesta, C. M. R. (2015). The digital transformation of the banking industry. *Digital Economy*, 1-10.

Dietterich, T. (1995). Overfitting and undercomputing in machine learning. *ACM Computing Surveys*, 27(3), 326-327. https://doi.org/10.1145/212094.212114

Egmir, R. T., & Sagbas, İ. (2021). Atm Bankacilik Sistemi Ve Regülasyonlar. *Gümrük ve Ticaret Dergisi*, 8(24), 24-36. Retrieved from https://dergipark.org.tr/tr/pub/gumrukticaretdergisi/issue/63502/841987 (Accessed: February 25, 2023).

Engen, J. (2007). Mobile banking's second act. *U.S. Banker*, 117(6), 35.

Gardner, M. W., & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences.

Atmospheric Environment (1994), 32(14), 2627-2636. https://doi.org/10.1016/S1352-2310(97)00447-0

Gel, C. O. (2004). Operasyonel Yönetimde Finansal Farkındalık (2nd edition). Istanbul: Sistem Yayıncılık.

Gepp, A., Linnenluecke, M. K., O'Neill, T. J., & Smith, T. (2018). Big data techniques in auditing research and practice: Current trends and future opportunities. *Journal of Accounting Literature*, 40(1), 102-115. https://doi.org/10.1016/j.acclit.2017.05.003

Giovanis, A., Assimakopoulos, C., & Sarmaniotis, C. (2019). Adoption of mobile self-service retail banking technologies: The role of technology, social, channel and personal factors. *International Journal of Retail & Distribution Management*, 47(9), 894-914. https://doi.org/10.1108/IJRDM-05-2018-0089

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359-366. https://doi.org/10.1016/0893-6080(89)90020-8

Irvin, S. (1994). Using lifetime value analysis for selecting new customers: Moving beyond traditional modeling applications. *The Credit World*, 82(3), 37.

Jackson, J. (2021). How to measure customer value and start using it for segmentation. Retrieved [30.05.2021], from [https://www.yaagneshwaran.com/blog/customer-value-based-segmentation/].

Kim, S., Jung, T., Suh, E., & Hwang, H. (2006). Customer segmentation and strategy development based on customer lifetime value: A case study. Expert Systems with Applications, 31(1), 101-107. https://doi.org/10.1016/j.eswa.2005.09.004

Kommu, K. (2022). Significance of digital customer onboarding in the banking space. *Dataquest*, Retrieved from http://lproxy.yeditepe.edu.tr/login?url=https://www.proquest.com/trade-journals/significance-digital-customer-onboarding-banking/docview/2708052785/se-2

Koranga, M., Pant, P., Kumar, T., Pant, D., Bhatt, A. K., & Pant, R. P. (2022). Efficient water quality prediction models based on machine learning algorithms for nainital lake, uttarakhand. *Materials Today : Proceedings*, 57, 1706-1712. https://doi.org/10.1016/j.matpr.2021.12.334

Mallat, N., Rossi, M., & Tuunainen, V. (2004). Mobile banking services. *Communications of the ACM*, 47(5), 42-46. https://doi.org/10.1145/986213.986236

Master, W. (2022, Apr 04). Why Digital Onboarding is Imperative for Banking Today. *Addis Fortune*, Retrieved from http://lproxy.yeditepe.edu.tr/login?url=https://www.proquest.com/magazines/why-digital-onboarding-is-imperative-banking/docview/2647053352/se-2

Karjaluoto, H., Mattila, M. and Pento, T. (2002) Factors Underlying Attitude Formation towards Online Banking in Finland. *International Journal of Bank Marketing*, 20, 261-272.

McDonald, M. (2012). Market segmentation: How to do it and how to profit from it (1. Aufl.;4; ed.). Wiley.

Meinert, M. C. (2019). Practical tips from two chief experience officers on designing a user-friendly digital onboarding process. *ABA Banking Journal*, 111(5), 30.

Miklosik, A., Kuchta, M., Evans, N., & Zak, S. (2019). Towards the adoption of machine learning-based analytical tools in digital marketing. *IEEE Access*, *7*, 85705-85718. https://doi.org/10.1109/ACCESS.2019.2924425

ODX launches digital account opening to support contactless, online banking: Digital account opening enables services ranging from online checking to lending for financial institutions of all sizes (2020). PR Newswire Association LLC.

Ozili, P. K. (2018). Impact of digital finance on financial inclusion and stability. *Borsa Istanbul Review*, *18*(4), 329–340. https://doi.org/10.1016/j.bir.2017.12.003

Palaon, H., Wiryono, S. K., & Faturohman, T. (2020). Branchless banking agents: Business satisfaction, continuity, and viability. *Cogent Business & Management*, 7(1), 1823585. https://doi.org/10.1080/23311975.2020.1823585

Proctor, D. (2023, June 3). What is Digital Banking? Temenos. Retrieved from https://www.temenos.com/news/2019/12/19/what-is-digital-banking/

Reichheld, F. F., Markey, R. G. J., & Hopton, C. (2000). The loyalty effect -- the relationship between loyalty and profits. *European Business Journal*, 12(3), 134.

Riza, A.F. (2019). Customer Acceptance of Digital Banking In Islamic Bank: Study On Millennial Generation. Conference of Islamic Management Accounting & Economic, 66-74.

Sardana, V., & Singhania, S. (2018). Digital technology in the realm of banking: A review of literature. *International Journal of Research in Finance and Management*, *1*(2), 28–32. https://www.allfinancejournal.com/archives/2018.v1.i2.12

Seyrek, İ. H. & Akşahin, A. (2016). Mobil bankacilik uygulamalari kalite faktörlerinin analitik hiyerarşi prosesi ile karşilaştirilmasi . *International Review of Economics and Management*, 4 (3), 47-61. Retrieved from https://dergipark.org.tr/tr/pub/iremjournal/issue/15725/285226

Shao, G. (2007). The diffusion of online banking: research trends from 1998 to 2006. *The Journal of Internet Banking and Commerce*, 12(2), 1–13. https://www.icommercecentral.com/open-access/the-diffusion-of-online-banking-research-trends-from-to-.pdf

Silva, M. C. d., Tavares, G. M., Gritti, M. C., Ceravolo, P., & Barbon Junior, S. (2023). Using process mining to reduce fraud in digital onboarding. Fintech, 2(1), 120-137. https://doi.org/10.3390/fintech2010009

Skinner, C. (2014). *Digital bank: Strategies to launch or become a digital bank.* Marshall Cavendish International (Asia) Private Limited.

Son, Y., Kwon, H. E., Tayi, G. K., & Oh, W. (2020). Impact of customers' digital banking adoption on hidden defection: A combined analytical–empirical approach. *Journal of Operations Management*, 66(4), 418-440. https://doi.org/10.1002/joom.1066

Song, Y., & Lu, Y. (2015). Decision tree methods： applications for classification and prediction. *Shanghai Jingshen Yixue*, 27(2), 130-135. https://doi.org/10.11919/j.issn.1002-0829.215044

Stangierski, J., Weiss, D., & Kaczmarek, A. (2019). Multiple regression models and artificial neural network (ANN) as prediction tools of changes in overall quality during the storage of spreadable processed gouda cheese. *European Food Research & Technology*, 245(11), 2539-2547. https://doi.org/10.1007/s00217-019-03369-y

Suoranta, M., & Mattila, M. (2004). Mobile banking and consumer behaviour: New insights into the diffusion pattern. *Journal of Financial Services Marketing*, 8(4), 354-366. https://doi.org/10.1057/palgrave.fsm.4770132

TBB (2022). TBB- Dijital, İnternet ve Mobil Bankacılık İstatistikleri. Retrieved from https://www.tbb.org.tr/tr/bankacilik/banka-ve-sektor-bilgileri/istatistiki-raporlar/59

Thu Nguyen, T., Thi Nguyen, H., Thi Mai, H., & Thi Minh Tran, T. (2020). Determinants of digital banking services in vietnam: Applying UTAUT2 model. *Asian Economic and Financial Review*, 10(6), 680-697. https://doi.org/10.18488/journal.aefr.2020.106.680.697

Usta, R. (2005). Tüketicilerin internet bankaciligini kullanmama nedenleri üzerine bir arastirma. *Doğuş Üniversitesi Dergisi*, 6 (2), 279-290. Retrieved from https://dergipark.org.tr/tr/pub/doujournal/issue/66653/1042894 (Accessed: March 4, 2023).

Wang, Y., Wu, S., Yan, D., Li, F., Chengcheng, W., Min, C., & Wenyu, S. (2020). Determining and mapping the spatial mismatch between soil and rice cadmium (cd) pollution based on a decision tree model. *Environmental Pollution* (1987), 265, 115029-115029. https://doi.org/10.1016/j.envpol.2020.115029