CREDIT CARDHOLDERS CHURN ANALYSIS USING DATA MINING METHODS

FATMA ÖZGE ERSÖYLEYEN

MASTER'S THESIS

DEPARTMENT OF MASTER OF BUSINESS ADMINISTRATION

YEDITEPE UNIVERSITY

June, 2017

Approval of the Institute of Social Sciences

Prof. Dr. M. Fazıl GÜLER
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master

Prof. Dr. Dursun ARIKBOĞA
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Doç. Dr. Dilek SAĞLIK ÖZÇAM
Supervisor

Examining Committee Members

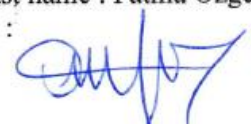
Doç. Dr. Dilek SAĞLIK ÖZÇAM -Yeditepe University

Yrd. Doç. Dr. Yusuf Can ERDEM- Yeditepe University

Prof. Dr. Beril Durmuş - Marmara University

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 : Fatma Özge Ersöyleyen
Signature :



ABSTRACT

Developing information technology, increasing competition and the economic conditions causes big concerns for organizations to manage their customers. In today's competition area between organizational environment the cost of acquiring new customers is much more expensive than keeping existing customers (Roberts, 2000). Therefore, organizations want to know their customers to build long term relationships. To know customers and keeping their existing customers have much more profit than to gain a new customers. The customers who have long term relationships buy more, customer retention may provide new referrals by positive word-of-mouth and a company may become less sensitive to competitor's activities. For this reason, organizations which are especially in service sectors focus on their existing customers and their products to prevent customer churn. (Nie Guangli; Rowe Wei; Zhang Lingling; Tian Yingjie; Shi Yong, 2001).

Forecasts or predictions are absolutely an important and a necessary thing for enterprises to move forward of today's ever-changing and highly interactive business environment (Hanke and Wichern, 2009). Because of that, this study focuses both customer churn and the developing some predictions about customer churn.

In todays, a credit card is most popular product for payments. Besides that the customers use credit card as a necessity product not a luxury product. In order to increase their profit and to have long-term relationship with customers in marketing industry, financial industries also take credit cards seriously.

In this study, credit cardholders who are churners are reviewed. The profile of the churner is stated by using data mining algorithms. Therefore, the credit cardholders who are churners or non-churners are predicted clearly in the next period. The data attributes of credit cardholders are obtained from an anonym private bank. All of these data attributes are masked or transformed to supply data security and data privacy. This churn analysis and forecasts for credit cardholders is made in order to support the marketing process of the anonym private bank.

ÖZET

Gelişen bilgi teknolojisi, artan rekabet ve ekonomik nedenlerden dolayı kurum ve kuruluşların kendi müşterilerini yönetebilme konusunda büyük endişeleri var. Günümüz rekabet alanında; kurum ve kuruluşların yeni müşteriler edinme maliyeti mevcut müşterilerini korumaktan çok daha pahalıdır (Roberts, 2000). Bu nedenle, kurum ve kuruluşlar müşterileri ile uzun ilişkiler kurabilmek ve müşterilerini tanıyabilmek istiyorlar. Müşterileri tanımak ve mevcut müşteri ilişkilerini korumak, yeni bir müşteri kazanmaktan çok daha karlıdır. Uzun süreli ilişkileri olan müşteriler daha fazla ürün ya da hizmet satın alırlarsa, müşteri sadakati ile yeni yönlendirmeler sağlayabilir ve müşteriler pazardaki diğer firmaların faaliyetlerine daha az duyarlı hale gelebilir.

Bu nedenle, özellikle servis sektöründeki kuruluşlar, müşteri değişimini önlemek için mevcut müşterilere ve ürünlerine odaklanmaktadır. (Nie Guangli; Rowe Wei; Zhang Lingling; Tian Yingjie; Shi Yong, 2001).

Tahminler ve öngörüler, günümüz değişen teknolojisinde şirketlerin sektörlerinde ilerlemeleri için kesinlikle önemli ve gereklidir. (Hanke and Wichern, 2009). Bu yüzden, bu çalışma müşteri sadakatı ve müşteri sadakatını arttırmak için öngörüler üzerinde odaklanmaktadır.

Günümüz teknolojisinde en popüler ödeme şekli kredi kartlarıdır. Bunun yanı sıra, müşteriler kredi kartını lüks bir ürün değil bir zorunluluk ürünü olarak görmektedirler. Finansal endüstriler de karlarını arttırabilmek ve pazarlama endüstrisindeki müşterilerle uzun süreli bir ilişki kurmak ya da müşteri sadakatını kazanabilmek için kredi kartlarını ciddiye almaktadır.

Bu çalışmada, kredi kartlarını kullanmayı bırakan müşteriler incelenmektedir. Kredi kartını kullanmayı bırakan müşteriler veri madenciliği algoritmaları ile incelenmiştir. Böylece, ileri ki dönemlerde kredi kartlarını kullanmayı bırakabilecek veya kredi kartlarını kullanmaya devam edecek olan müşteriler açıkça ön görülebilmektedir. Kredi kartı sahiplerinin veri nitelikleri özel bir bankadan elde edilmektedir. Kullanılan tüm bu veri nitelikleri, müşteri veri güvenliği ve veri gizliliği sağlamak için maskelenmiştir. Kredi kartı

sahiplerine yönelik sadık müşteri analizi ve tahminleri, özel bankanın pazarlama sürecini desteklemek üzere yapılmıştır.

ACKNOWLEDGEMENTS

I would like to express my special thanks of gratitude to my advisor Doç.Dr. Dilek Sağlık Özçam who helped me to do this wonderful project on the topic "Credit Card Churn Analysis using Data Mining Methods" which also helped me in doing a lot of research and i came to know about so many new things.

I am extremely thankful to Kübra Canel, who is my friend, for her invaluable comments, support, advice and guidance through all my studies. Her support and advices are more valuable during this time.

I would like to thank my friends who helped me a lot in finalizing this project within the limited time frame.

Last but not the least, I would like to thank my family; my parents and to my two precious sisters for supporting me spiritually throughout writing this thesis and my life in general. Thank them for accepting nothing less than excellence from me. This study is heartily dedicated to my sisters.

Table of Contents

ABBREVIATIONS

JCB	Japon Credit Bureau International
DWH	Data warehouse
BI	Business Intelligence
KDD	Knowledge Data Discovery
CRISP-DM	Cross Industry Standard Process for Data Mining
DM	Data Mining
GLM	Generalized Linear Model
SVM	Support Vector Machine
AC	Accuracy
TPR	True Positive Rate
PPV	Positive Predictive Rate
FNR	False Negative Rate
FPR	False Positive Rate
TNR	True Negative Rate
ROC	Receiver Operating Characteristics
ROI	Return on Investment

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

1. INTRODUCTION

Keeping existing customers is one of the most critical challenges for all service industries. Developing information technology, increasing competition and the economic conditions causes big concerns to managing customers and this becomes much more serious problem for marketing industry. In today's competition area between organizational environment the cost of acquiring new customers is much more expensive than keeping existing customers (Roberts, 2000). Therefore, organizations know to build long term relationships have much more profit than to follow a customer driven vision. For this reason, organizations which are especially in service sectors focus on their existing customers and their products to prevent customer churn. (Nie Guangli; Rowe Wei; Zhang Lingling; Tian Yingjie; Shi Yong, 2001).

A credit card is most popular product for payments. The credit card is taken by a financial company. A credit cardholder who has a credit card is able to option to borrow funds and pay a merchant for goods and services without any cash payment. The financial company accepts that the cardholder promise to repay this loan to the financial company (O'Sullivan, Arthur; Steven M. Sheffrin, 2003). In other words, credit card is a financial tool for both banks and customers. The cardholder could pay for goods and services using credit card and promise to pay for loan of his credit card. The issuer of the credit card which is a kind of bank creates a revolving account and grants a line of credit to the cardholder. In this way, the cardholder can borrow money for payment to a merchant or a cash advance. Credit cards are one of the most profitable and effective factor for banks. Because of that extensive usage of the credit cards, the total amount of it, regular payments, debts of it and especially cancellation are most important criteria for banking sectors.

Churn is used like attrition synonymously (Singh Hergovind; Samalia Harsh Vardhan, 2014). The customers who do not use company's service or product called a churner. Churn analysis explains that customers that are most likely discontinued using a kind of service or product. Service industries use churn modeling to make predictions to prevent customer churn or product churn. Churn analysis using one or combination of many algorithms or

methods, enables the business to address answer of the valuable questions for market and customer churn.

Data science is used by enterprises to transform raw data into useful information. Making sense of data could be extracted from large databases, warehouses. Data science approaches could predict future trends and behaviours based on previous examples. This allows enterprises to make proactive, knowledge-based decisions.

In this work, a credit card churn approach is modeled to pre-predict customers that are possible to leave. Meanwhile, several churn forecasting models are built by the following methods: decision tree, naive bayes method, generalized linear model, support vector machine. Also, training data is obtained from an anonymous private bank's credit card customers. Thus, banks could offer new campaigns or develop new marketing strategies to retain customers.

1.1. The Purpose of the Study

The main purpose of this study, to focus on application of the churn prediction and understand why credit card holders cancel their credit cards. First of all, the related data which are mainly included customer's demographic information, credit card information, credit card transactions, credit risk information of credit cardholders, deposit accounts of credit cardholders, product activity of credit cardholders, complaints data are collected from an anonymous private bank for the descriptive credit card churn and predicting credit card churn in this study. The data are collected, arranged and analyzed with high sensitivity. According to the data structure, analyzed data is enriched and modelled using Data Mining Techniques.

1.2. The Importance of the Study

This paper's most important contribution is that, in traditional software there is a programmer, and the programmer implements all business rules and creates complex decision branches. Decision could fail when the rules are getting more complex. In machine

learning approach, there is data and the software learns all of these business rules from previous data.

In this work, it is hypothesized that credit card churns would be detected by machine learning approaches. By the way, the churn analysis and churn prediction improvement by well-known algorithm. It is also hypothesized that how social media information is able to influence the credit cardholder churn and how compliance of customers is able to influence the credit cardholder churn. Then, to develop new marketing strategies from the model which has more accuracy and the best result for making interpretation.

This study is organized as follows. Section 2 demonstrates relevant literature review. In Section 3, the modeling approach is described. Section 4 presents the experimental results of model. Section 5 includes limitation of the study and section 6 presents the conclusion of the study.

2. THEORATICAL FRAMEWORK

2.1. Credit Card

Nowadays, card payment systems have many alternative ways to decrease the usage of cash and check by the customers. These process have started with the usage of debit card and has being continued with the credit cards.

Todays, credit cards are more popular, an important product for everyday life, is more modern, practical and easiest way for shopping in all over the world. The usage of credit card is not a luxury consumption, it is a necessity. Obviously, there are many benefits using credit card. People who have credit card does not need to carry cash to supply own financial needs. By the way having a credit card is more safety way for consumer's expenditure. On the other hand, with the developing technology, there is no way to buy something on online shopping. Additionally, in order to use the credit card as a payment tool, a cardholder is able to take cash up to cardholder's limit with a certain amount of interest rate.

According to the reports; in the first six months of 2015 alone, there were some 14.5 billion U.S. general purpose credit card transactions accounting for more than $1.4 trillion in purchase volume. Credit cards are also a major driver of consumer indebtedness. As of the end of the second quarter of 2015, there were some $703 billion in credit card loans outstanding, behind only housing debt, automobile debt, and education debt as a component of overall household liability (New York Federal Reserve Bank, 2015).

The credit card has includes three primary elements:

- A credit card issuer: An issuer can be a bank or a financial institution
- A credit card holder: The customer who uses the credit card
- A merchant: The merchant which has a contract to accept the credit card as a payment tool for its sales.

The credit card operations actualize with the membership agreements. These agreements are signed between the international associations and the banks or any other financial institutions. Card Associations can be described as a network formed by banks that regulate

and make payment transactions. There are more than one type card associations. These are Visa, MasterCard or Europay, American Express, Discover, Diner's Club and Japon Credit Bureau International (JCB). These associations determine the standard rules for each credit card's transactions. In Turkey, Visa, MasterCard and American Express cards can be obtainable. In addition, the group names that banks give out to the cards that deal with various facilities such as Visa, MasterCard, American Express and so on.

In this study, two types of credit cards associations have used are Visa and MasterCard.

2.1.1. The Features of Credit Card

Credit card is magnetic stripe card or with chip which is issued by a financial company such as a bank, store, or service provider. It is a short term financing product. It is the essential part of electronic and internet commerce. The card holder's name is written on this card. There is a security password for each credit card. This password which is determined by card holder makes it more securable financial product. Therefore, the owner has an option to borrow funds to buy a good or service with cash or check up to a certain limit. This certain limit is known as a credit limit that is approved by the financial company. Most of credit cards have interest rates on the card holder's loans. The holder has to repay the funds within thirty days to avoid interest or some penalties (investopedia). To refund all of the total term debt amount are means that the cardholder is use that credit card effectively.

2.1.2. The Components of Credit Card

The basic components of credit cards have been listed below. These components are specific for each credit card holders take in part on the credit card.

- Cardholder Name Surname

For Example: Özge Ersöyleyen

- Card Number

For example: 5127 5400 0000 0002 (16 digit for Visa and MasterCard)
3744 270000 00003 (15 digit for American Express)

The credit card numbers are derived according to the Luhn algorithm.

- Valid Thru = Expire Year & Month

For Example: 07 / 15

- CVC2 = CVV.2 = CID

It is a security code which is settled on the backside of credit card

For example: 936 (3 digits for Visa and MasterCard)

7591 (4 digits for American Express)

- Card Association

It is a type of logo which is symbolized card establishment

For example: Visa, MasterCard, American Express, Discover, Diner's Club or JCB

- Issuer Bank

It is included the logo of the bank and name of the bank

- Card Family

It explains that the card Family.

For example: Card Family of banks in Turkey; Bonus, Wings, World, Maximum, Axess, CardFinans, Advantage, Paraf, Adios

- Chip

In Turkey, with the implementation of chip and PIN has started to usage on 31 March 2016, usage of the password has been an obligatory for all of type transactions which are done by the credit card since1 June 2007.

2.2. Data Driven Concepts

2.2.1. Data

With the developing of the technology, everywhere and everything is reachable quickly. The technological tools; such as smart phones, ipads, computers can make easier the human's life. Internet has also plays the most important role to supply connectivity and accessibility. Therefore, not only each action, each text, each picture but also each enter process the data. The data volume is increasing exponentially day by day. Todays, many of professors, researchers and scientists are work on data to make it much more meaningful and usable and study to understand the data and make an interpretation by this data.

In literature the word of data has reviewed since 1640s. The word of data is used to refers "transmittable and storable computer information" in 1946. A data is the set of values of qualitative and quantitative variables. The data is commonly related with scientific research and it is collected by the organizations. The data types include geographical data, cultural data, scientific data, financial data, metrological data, natural data, transport data. It can be seen that the data is the measurable variable. In other words, the data is measured, collected and reported, analyzed and it may be visualized with the usage of graphs, images or other analysis tools.

In 1955, the English – American economist and educator, Kenneth Boulding defined that there is a hierarchical arrangement and relationships between data, information, knowledge and wisdom. That hierarchial arrangement is known as DIKW that is obtain from the beginning letter of data, information, knowledge and wisdom (Wallace, Danny P., 2007) , (Boulding, Kenneth, 1955). According to the DKIW pyramid, the definition of the stages are:

- Data: symbols
- Information: data that are processed to be useful; provides answers to "who", "what", "where", and "when" questions
- Knowledge: application of data and information; answers "how" questions
- Wisdom: evaluated understanding.

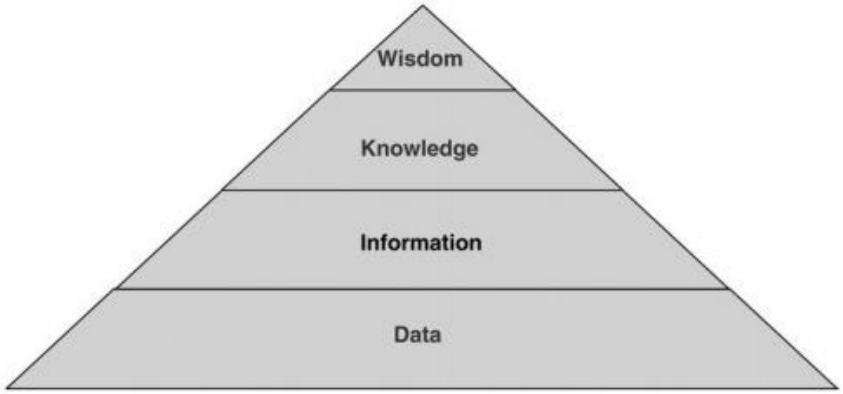


Figure 1. DIKW pyramid (Rowley, 2007)

According to the literature; DIKW pyramid defined like :

"Typically information is defined in terms of data, knowledge in terms of information, and wisdom in terms of knowledge" (Rowley, 2007).

An American organizational theorist Russell Ackoff published the International Society for General Systems Research in1989. According to Ackoff's version of model an understanding steps is found between the parameters of DIKW. The following diagram represents the transitions from data, to information, to knowledge, and to wisdom, each stage composed to the next stage. understanding that support the transition from each stage to the next. Therefore, the content of the DIKW can be classified into five categories with the understanding (Bellinger Gene, Durval Castro, and Anthony Mills, 2004). If necessary to define the understanding stage, it can be like :

- Understanding: appreciation of "why"

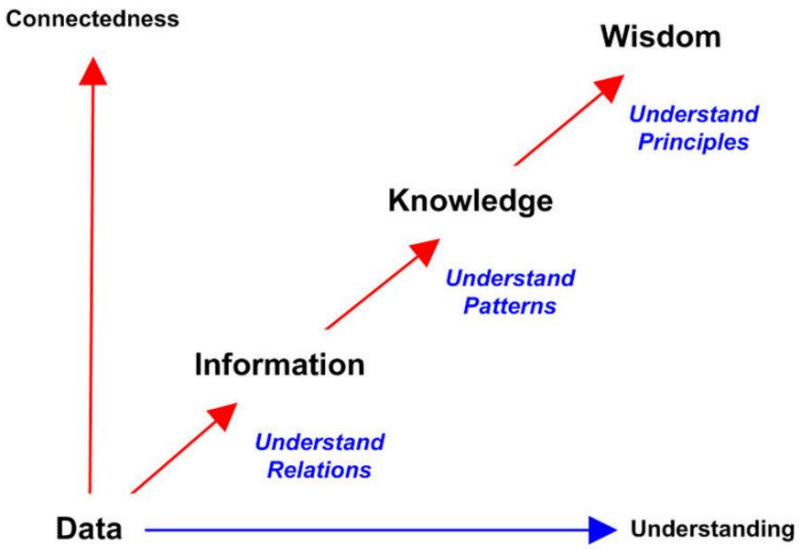


Figure 2. The diagram of Ackoff's DIKW model

(Bellinger Gene, Durval Castro, and Anthony Mills, 2004).

This study aims that the taking data and follow all the stages of Ackoff's DIKW model in order to gain usable and meaningful knowledge from the set of data.

2.2.2. Database

Database is a collection of data or information that store in a kind of server. Databases are used in many applications to supply interactions between customers and the suppliers. The application examples that shows the databases are everywhere such as banking, accounting, insurance, manufacturing, social gaming, personal cloud storage, sports, finances, government, organizations, social media, e-commerce, healthcare, weather and so on. Therefore, databases make easier to access the data or information quickly and efficiently for right time.

A database consists of structured tables. These tables occur from columns and rows (fields and records). It also includes schemas, table queries, reports, objects such as views, indexes, store procedures, functions, packages and so on.

It is possible to separate types of databases due to its characterizations. Some of these different kind of databases are listed like in-memory database, active database, cloud database, Data warehouses, embedded database, federated database system, graph database,

Operational databases, Real-time databases . This study is based on the kind of database is data warehouse.

In this study, there is no need to make any survey in order to obtain qualitative and quantitative data. All of the necessary data have been obtained from an anonym private bank database system and the historical data can be accessible from the data warehouse environment of the bank.

2.2.3. Data Warehouse

Data warehouse (DWH or DW) is also known as enterprise data warehouse (EDW) that was developed by two of IBM researchers Barry Devlin and Paul Murphy in 1980s (Dedić, N. and Stanier C., 2016). Data warehouse is a kind of database and is used for data gathering, data analysis, prediction analysis, reporting and many analytical operations. The data can be stored historically such as daily, weekly or monthly based on data. It is also a simple component of Business Intelligence (BI) that will be explained in next section with details.

The data in data warehouse is uploaded from many real-time databases that are operational databases such as financing, marketing, the organization's sales, salaries, operational data, summaries of data including reports, copies of data, human resource data and so on. To make predictions with a data or to make analyses for customers or product, a historical data is being a key point. Reviewing and reporting historical data in operational database is not an efficient and effective way. At this point, data warehouses are prepared. Data is transferred to the DWH environment by typical software tools, extract-transform-load (ETL) or extract-load-transform (ELT). During the transferring operations, data is cleaned, transformed, aggregated, catalogued, processing some extensions based on DWH that make data is usable and available for manager, business professionals or data scientist for data mining operations, market research, predictive analysis, decision support, data dictionary and so on.

In this study, the historical and transactional data will be used for predictions. Therefore, an anonym private bank's data warehouse environment will be used. The term of data set consists of six months data.

2.3. Business Intelligence

The first usage of the term of Business Intelligence (BI) is in Cyclopædia of Commercial and Business Anecdotes' which is published 1865 by Richard Millar Devens. Some of the BI applications need to DWH to gathered data and to make decision support with applying analytical operations. However, not all data warehouses can be usable for BI applications nor do all BI applications require a DWH platform. The aim of BI applications is to make an interpretation or make inferences with related data and to implement the new decisions or strategies. These decisions can be supply insights to the business against to the competitive market advantage and long term stability (Rud Olivia, 2009).

Business intelligence technologies are used for reporting, online analytical processing, analytics, data mining, process mining, complex event processing, business performance management, benchmarking, text mining, predictive analytics and prescriptive analytics. There are many methodologies for analytics, data mining, and data science project. Knowledge discovery from Data (KDD) and Cross Industry Standard Process for Data Mining (CRISP-DM) are the most popular methodologies.

2.3.1. KDD Methodology

Knowledge discovery from Data (KDD) is an iterative process that is included iterative sequences. While KDD is an artificial intelligence that is related with the extraction of useful and understandable knowledge from raw data, BI is a term that includes data warehouses, KDD and dashboards. With expansion of analytical researches and usage of KDD and BI in real world applications in 1990's, scientists and researchers started to usage

Knowledge Discovery and Business Intelligence (KDBI) together (Cortez Paulo, Santos Manuel Filipe, 2013).

These sequences of KDD process are: Data Selection, Data Preprocessing, Data Transformation, Data Mining, Pattern Evaluation and Knowledge Presentation or Interpretation (Kamber Micheline, Han Jiawei 2006). As it has seen on the diagram below, each step in KDD follow the other step. It is a hierarchical method.

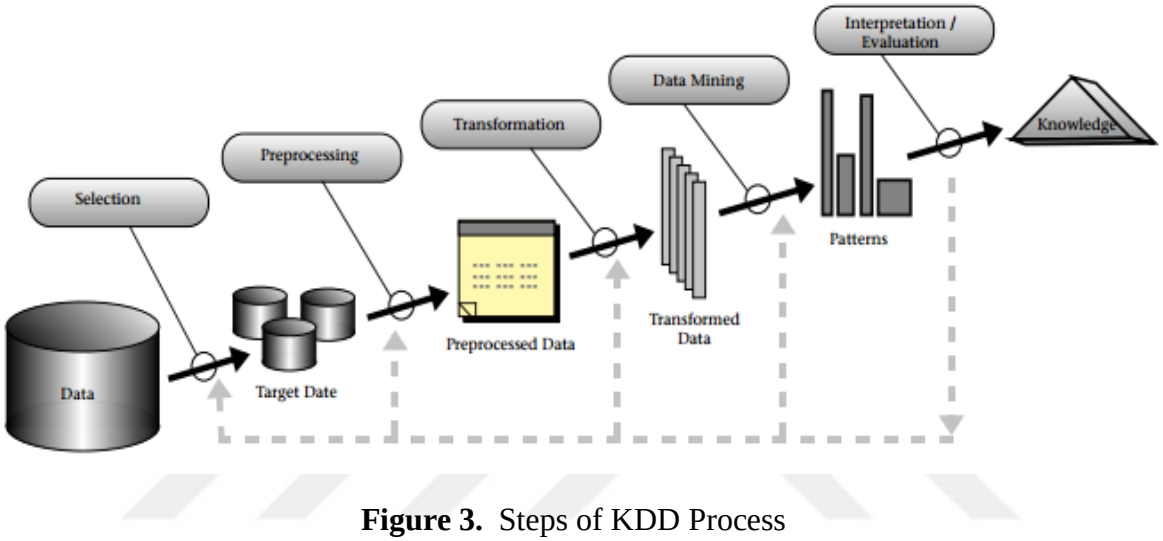


Figure 3. Steps of KDD Process

(Fayyad, Piatetsky-Shapir, Smyth 1996)

Data Selection is the first and the more important stage of KDD. It consists on creating the target data set. The data is collected from any environment like database, data warehouse and son to create the requested data set.

Data Preprocessing is the second step of KDD. Some researchers and in some articles, it is referred as Data Cleaning. By the removing the noise and inconsistent data, the requested data is created. In this way, the data is a cleaned data and the data quality and data reliability and data accuracy is increased.

Data Transformation is the third stage of KDD. Before, starting this stage, the first two stage of KDD have to be completed. In this stage, the data which is obtained and cleaned in stage two is needed. Therefore, transformation of the data can be available. This stage

consists the data which is transformed and consolidated forms to apply data mining algorithms or operating some statistical functions on it.

Data Mining is a significant and considerable stage. There are many intelligent data mining algorithms. These algorithms will have been mentioned in the next section. (2.4 Data Mining). After working with these algorithms, a pattern is extracted that is be usable for descriptive or predictive operations.

Pattern Evaluation and Knowledge Presentation or Interpretation is the latest stage that consists on the interpretation and evaluation of patterns. In this stage, with the interpretation, the pattern can be controllable whether it is meaningful or not.

2.3.2. CRISP-DM Methodology

Cross Industry Standard Process for Data Mining (CRISP-DM) is an iterative and adaptive process model to development a data mining project (Larose, 2005). CRISP-DM was developed by Daimler Chrysler, SPSS, NCR in 1999 (Colin Shearer, 2000). It consists of six phases; business understanding, data understanding, data preparation, modeling, evaluation and deployment. (Chapman et al, 2000; Colin Shearer, 2000)

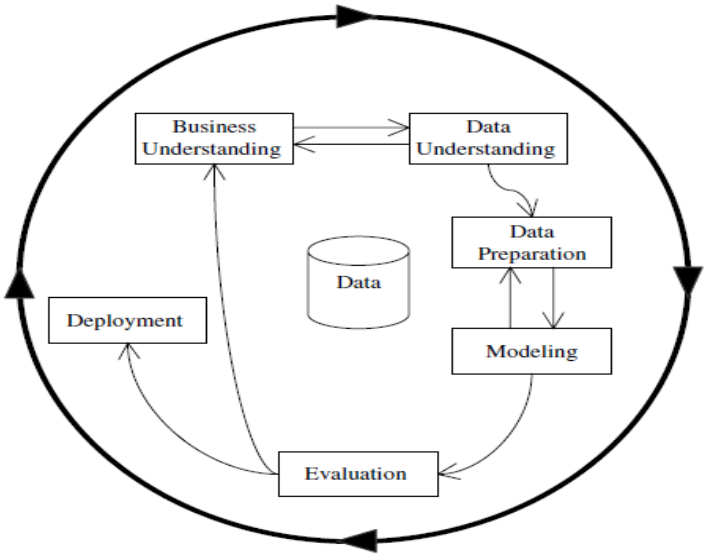


Figure 4. CRISP- DM Life Cycle (Shafique, Qaiser, 2014)

Business understanding is the first phases of the CRISP-DM methodology. The business requirements and definition is made in this phase. The definition of problem is determined. It focuses on understanding the business object and this object turns into DM business problem then it aims to achieve that business object with a plan designing.

Data understanding is the second phases of CRISP-DM methodology. It starts with the collection of the data which is needed for the objective. And then it continues with checking data quality and exploring data to get insight from collected data.

Data Preparation is the phase that covers all of the operation to achieve final data set. This final data set accuracy is more important for following the next step. This phase focuses on data transformations, data quality.

Modeling is the fourth phase of CRISP-DM methodology. In this phase, the model is prepared with using the data which is prepared in first three phases. In order to have insight for same business problem, many data mining techniques and algorithms (explained later with all detail) can be selected and applied to build a model.

Evaluation is the decision and interpretation stage for the model and deciding how the knowledge which is obtained by model is able to use.

Deployment is the final stage of CRISP-DM methodology. It includes that determining the usage of knowledge.

In this study, CRISP-DM processes will follow from business understanding to the deployment stage. With CRISP-DM methodology, a business problem which is credit cardholders churn turns into data mining problem. Then data requirement will determine then data collection, data quality operation will do. With the data preparation, the modeling phase will continue. Finally, the business problem will be visible with the output of the model. The anonym private bank has an insight for this churn objective.

2.4. Data Mining

Todays, data mining (DM) is used by the organizations to increase revenues and to reduce costs. Innovative organizations are already using data mining to locate and appeal to higher-value customers, to reconfigure their product offerings to increase sales, and to minimize losses due to error or fraud.

Data mining is an iterative process that produces significant and meaningful data from different sources and different types of data. Fayyad considers that DM is an essential analysis step in the process of analytical methodologies that are explained on previous section. (Fayyad et al., 1996). In other words, data mining is a process that uses a variety of data analysis tools to discover and describe the patterns and relationship between the variables in data that may be used to make valid predictions. In addition, the availability of huge amount data turns into useful information and knowledge. This information and knowledge can be used for database analysis and decision support or any other applications.

The main applications are listed as below:

- Market analysis and management
- For example: target marketing, customer relation management, market basket analysis, cross selling, market segmentation, customer retention, science exploration and so on.
- Risk analysis and management
- For example: Forecasting, customer retention, improved underwriting, quality control, competitive analysis and so on.
- Fraud detection and management
- Text mining (news group, email, documents) and Web analysis
- Intelligent query answering

Data mining aims to work on two important criterias. These are prediction and description. Therefore, data mining activities can be put in two categories which are descriptive data mining and predictive data mining. Descriptive data mining needs an available data set and produces new and undefined information from this available data set. It aims to characterize the general properties of the data which is in the databases. Predictive

data mining is described by the given data set and produces mode. It aims to perform inference from given data set in order to make predictions (Mehmed Kantardzic, 2003). In this study, an anonym private bank's available data is used to make description of credit cardholders who are churner and then with the given data set the predictive data mining is used to make prediction of credit card holders who will be a churner in a period.

There are different data mining functionalities. These are:

- Characterization and discrimination
- The mining frequent patterns, associations and correlations
- Classification and regression
- Clustering analysis
- Outlier analysis

In last decades, observation of substantial contribution of data mining increases the popularity of the discipline. The motivation could be used to control costs as well as contribute to revenue increases. The discipline could be integrated into almost all phases of customer life cycle which includes acquiring new ones, increasing revenues of old ones and most importantly retaining existing ones. Clustering loyal or profit making customers would help an organization to target prospects in same segmentation group. For example, cross-selling is a method looking for customers who have bought a particular product and customers clustered in same segmentation group and haven't bought that product. In similar way, organizations can act to retain customers who are at risk for leaving.

The technique of data mining which is using to make exploration and prediction is called as modeling. Modeling is an application which produces adequate and meaningful results or rules from known situations to use it for unknown situations. There are many modeling data mining technique such as logistic regression, decision tree, neural network, K-Nearest neighbour algorithms, genetic algorithms, fuzzy logic, link analysis and so on.

In this study, decision tree data mining method will be used to supply the retention of the credit cards holder. First of all, the pattern analyses will be drawn to see the churner and predict to customer who can be a churn in the new period.

2.4.1. Decision Tree Method

There are some hidden rules in large data set. The most important idea of machine learning algorithms depend on catching these rules that are in large data sets. These rules are discovered by an algorithm by dividing the large data sets into multiple meaningful subsets. Decision tree is one of the basic machine learning algorithms. Its interpretability makes it more popular against the other classification algorithms. It is a flow chart like tree structure. (Han and Kamber, 2001).

Decision trees are made by nodes that are represented in hierarchical structures. There are three important nodes are called root node, intermediate node and leaf or terminal node in decision tree. The root nodes which are the starting point of decision trees the base level of the tree. The intermediate nodes are referred as a connection node between root node and leaf node. The leaf nodes that are the end of the tree are representing the values of the responsible values. (J. MacLennan, Z.H. Tang, B. Crivăţ,2008).

In classification models, decision trees generate a basic rule set. The rule set is able to explain according to the decision model which is Figure 5, Decision tree algorithms :

In this decision tree, the employees will be split into three different groups which is composed of the employees who shows similarities between each other.

- In the first step, the employees are classified due to their ages
    - If the employee is senior and middle age than the employee can settled into the class C.
    - If the employee is young, than income of the employees are needed to review.
        - The income of the young employee is high then the employee can settle into class A.
        - The income of the young employee is low then the employee can settle into class B.

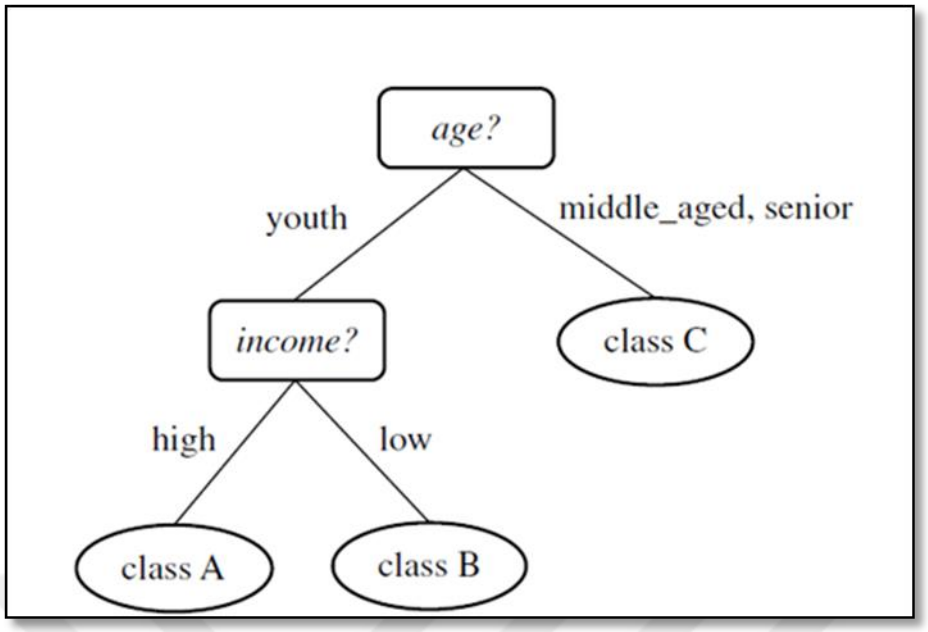


Figure 5. Decision Tree Classification Model (Han and Kamber, 2001).

2.4.2. Naive Bayes Method

Naive Bayes which is a classifier is based on Bayes' theorem. The Bayesian Classification method is a supervised learning method and a statistical method for classification. It is able to solve both diagnostic and predictive problems. (Chai, K.,H. T. Hn, H. L. Chieu, 2002).

Naive Bayes Method is calculating with an algorithm. It is a way of calculating the posterior probability which is represented with P(c|x) below. That is formulated by the following equation:

$$P(c|x) = \left(\frac{P(x|c)P(c)}{P(x)} \right)$$

Equation 1. Naive Bayes Method Formulation

Where ;

- P(c|x) is a posterior probability of class given predictor.
- P(c) is a prior probability of the class.
- P(x|c) is a possibility which is the probability of predictor given class.
- P(x) is a prior probability of predictor.

Naive Bayes is a popular method for text categorization, solving the problem of judging documents as belonging to one category or the other with word frequencies. It finds application in automatic medical diagnosis. (Rish, Irina, 2001).

2.4.3. Generalized Linear Model

Generalized Linear Model which is represented as GLM is a generalization of an ordinary linear regression. GLM is formulated as a way of unifying other statistical models. These statistical models are linear regression, logistic regression and Poisson regression. (Nelder, John; Wedderburn, Robert, 1972).

2.4.4. Support Vector Machine

Support vector machines (SVM) is known also support vector networks are supervised learning models which analyzes is data used for regression analysis and classification.

SVM can be used in many different areas to solve different problems. SVM is a helpful method in text and hypertext categorization, it is able to solve Classification of images. Hand-written problems can be recognized using SVM and so on.

2.5. Definition of Churn

In all over the world, business managers and many organizations take in consideration customer churn as an issue to manage value of the organization effectively and to be a long term market player (Geppert, 2002). Many organizations are confronted with churn predictions. In order to be competitive in the market, organizations have to make clear the

customer satisfactions with their products, services, offers. Organizations have to be able to predict possible churn patterns and take actions to retain loyal customers. Indeed,

Customer churn is that defined that how an organization lost their existing customer to any competitor (Chu,Tsai, and Ho, 2007). In other words, churn explains that why customer leaves from the organizations, why customer cancels the products or services of the organizations. Churners are the customers who do not want use the products or services of the organizations (Glady, Baesens, and Croux, 2009). Customer churn rate is a ratio that shows the customer who leaves from an organizations in a certain time period. This ratio is able to related with the customer satisfaction, even better offers from competitors, following better marketing or sales strategies by competitors or changes customers' life cycle. Customer churn can be reviewed in three different types:

- Voluntary churn: This type churn is caused by the customer dissatisfaction. The customer does not want to take any services from the organization. He / she leaves from the organizations and cancels all of the products or services which are related with the organization.

- Involuntary churn: In this type of churn, the organization does not want to cooperation with the customer and the organization terminates the relationship with the customer. The customer's uncertain behaviors, attitudes can be caused this churn. Fraud, disorder in loan payments, taking non-repayment credits can be served as an example.

- Uninitiated churn: In this type churn, the customer is not settled no longer in the target volume of the organization. The death of the customer can be served as an example.

In order to reduce customer churn, it is important that supply customer satisfaction to the organization's service or product. Many researches demonstrate that the customer retention or customer loyalty is the key point for preventing customer churn and the economic value of an organization related directly with the customer retention. The cost of acquire or obtain new customer is much more expensive than retain existing customers

(Roberts, 2000), (Ennew and Binks, 1996; Dawes and Swailes, 1999). According to the previous studies, the cost of acquiring new customers is higher for five times than making additional sales to the existing customers (Dixon, 1999; Slater & Narver, 2000). Therefore, customers' long term relationship has much more worth than gaining new customers. Long term customers buy more product or service of an organization and because of their loyalty and pleasure they are able to bring new customers to the organization (Reichheld, 1996). Previous studies have proved that how customer retention has much power. Especially, a bank is able to increase its profits by %85 through %5 improvement of customer retention (Reichheld, F. F., & Sasser, W. E. Jr., 1990).

In this study, the research will focus on the credit card churn of an anonym private bank. This application goals that the predict credit card churn in the bank and supply more loyal usage of credit cards of the bank according to the reviewing of the customers which have credit cards.

3. METHODOLOGY

Previous studies for credit card churn analysis and forecasting have focused only on one or two data mining methods. This paper studies will make data analysis on customer's demographic information, credit card information, credit card transactions, credit risk information and deposit accounts of customers who are credit cardholders for applying four data mining methods. As a result, the validity and correctness of these algorithms will be comparable. Then, the new marketing strategies will develop according to the model which has more accuracy than the others.

The organizations want to prevent lost revenue that is caused by customer churn. This study aims preventing credit cardholders churn who customers of an anonym private bank. Building a churn model with a data mining tool will help to understanding why customers cancel their credit cards or they start to work any other financial company or switch their credit cards which are serve any other banks.

3.1. Modeling Approach

This study presents credit cardholders churn modeling and analysis for retail customers in a private bank. The aim of this study to predict credit cardholders who attend to leave from the bank. The data mining methodologies which were used in this study to define the churn model is clarified in previous section.

The key point of data mining projects is that to choose an appropriate and suitable data mining tool. In this study Oracle Data Miner (ODM) which is a data mining and machine learning is used as a data mining software. ODM provides many data mining methods such as decision trees, regression and neural networks.

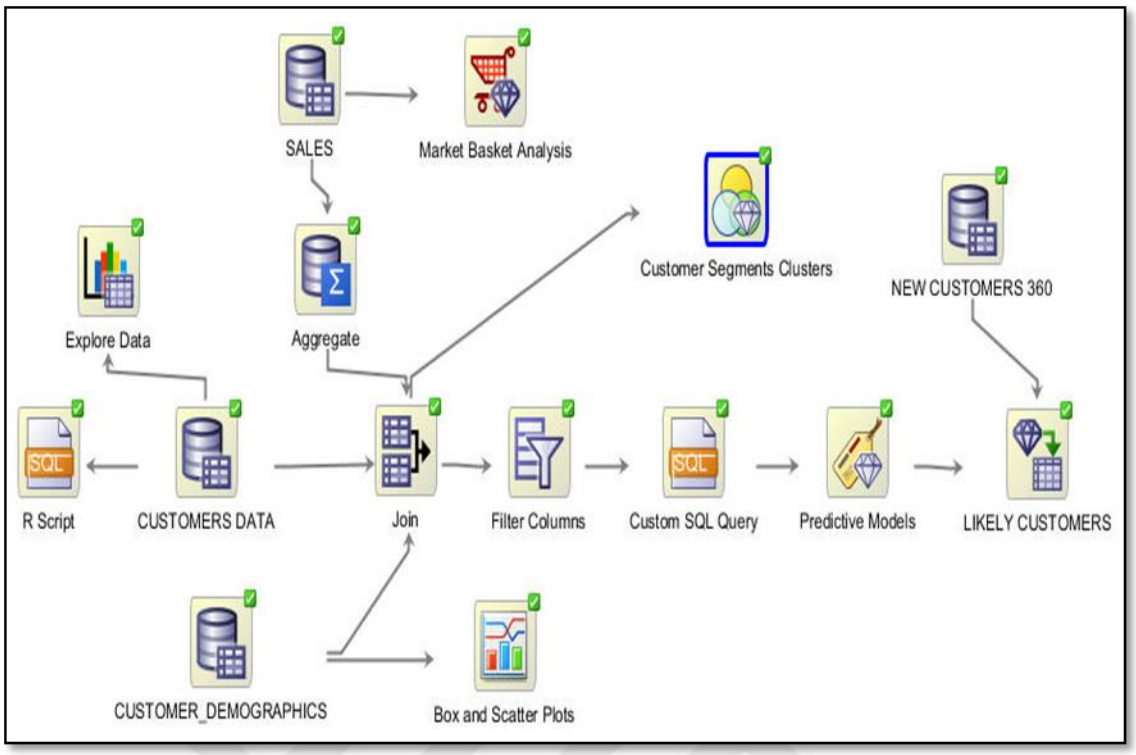


Figure 6. Oracle Data Miner

In order to obtain usable and effective knowledge from the data, business Understanding, data understanding, data preparation - data validation, data cleaning and data quality - modeling and evaluation have to be obtained. These operations will have done according to steps of CRISP-DM which is explained in section 2.3 Business Intelligence, 2.3.2. CRISP-DM Methodology.

3.1.1. Business Understanding

Previous studies demonstrate that the cost retaining the existing customers is lower than the cost of acquiring new customers. (Xu, et al 2009). In other words, retaining existing customers increases the profitability of the bank instead of gaining new customers. For this reason studies about churn analysis has taken an important role to prevent customer loss in revenue for enterprises.

In this study, a churn model will have built to prevent credit cardholder loss in an anonym private bank. The main data of this study consists of credit cardholders. The most

difficult part of this type data mining project to make a churn definition clearly. For this reason, it is necessary to find out in what circumstances the credit cardholder will be a churner.

- The credit cardholders who are closed their credit card accounts in last month have been accepted as a churner. Therefore, their last month activities have been accepted pattern of the credit cardholders churn.

The data is collected from an anonym private bank. The data which will be filtered includes:

- Credit card accounts which are opened closed and open,
- Retail customers who are active customers of the bank
- The credit cardholders who have credit card account of the bank since three months.

3.1.2. Data Description

The three months of the historical data is taken from data warehouse and all of the data will have been customer and credit card based data. There are 30000 unique credit card accounts that are analyzed in the main data set. Different variables which are derived from customers demographic information, credit card information, credit card transaction information, deposit accounts, products activity, complaints data and social media data. The eight variable categories will have been examined in data preparation. These customer based categories are:

- Customer demografic information of credit cardholders,
- Credit card information,
- Credit risk information of credit cardholders,
- Credit card transaction activity information,
- Deposit accounts of credit cardholders,
- Product activity of credit cardholders,
- Complaints data,
- Social media data.

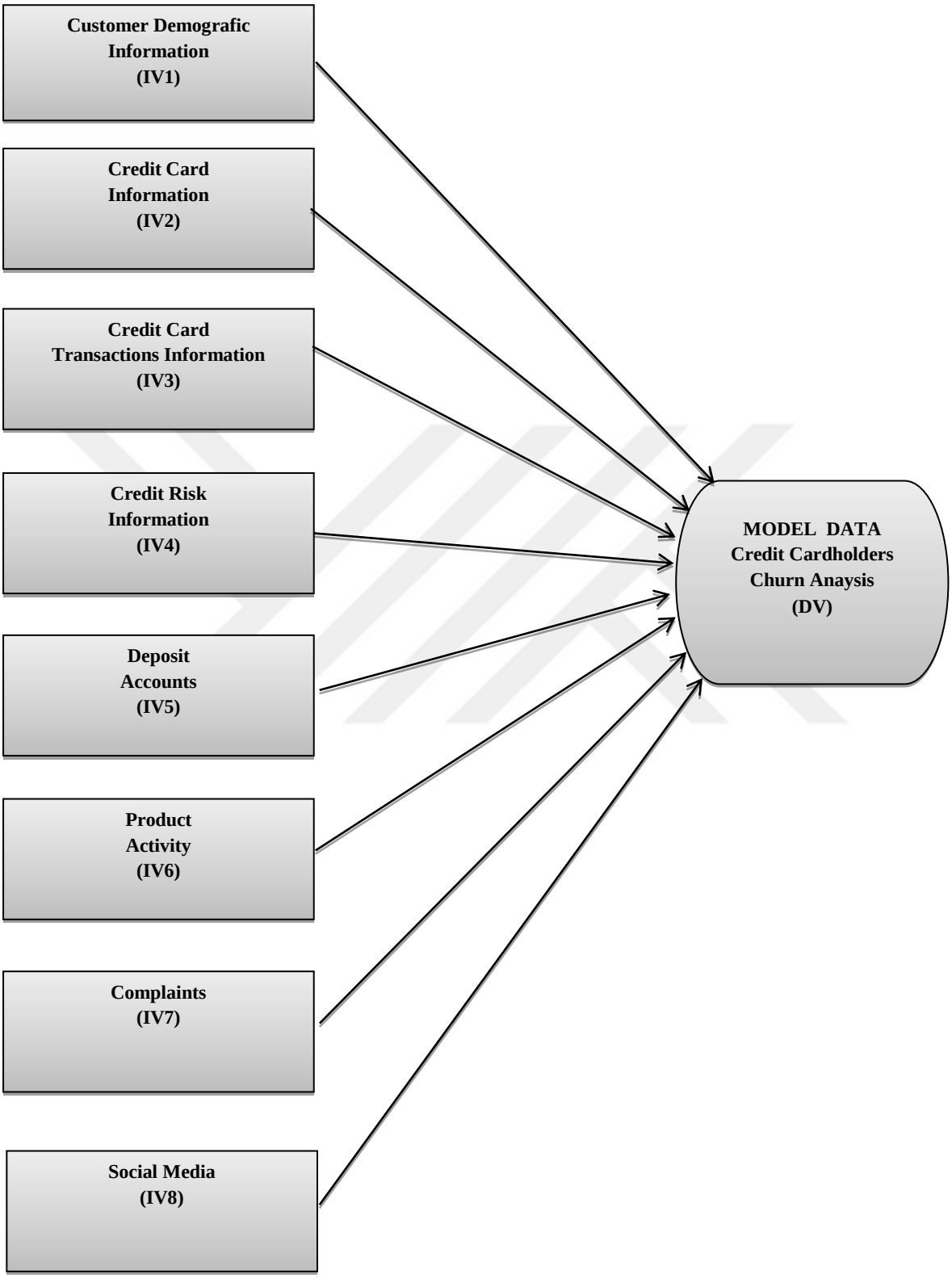


Figure 7. The tables are using in model

The conceptual framework is built on customer-based "Model Data" which is dependent variable (DV). Customer demographic information (IV1), credit card information (IV2), credit card transactions (IV3), credit card risk information (IV4), deposit accounts (IV5), product activity (IV6), complaints (IV7) and social media information (IV8) which are independent variables (IV) are going to explain certain amount of the variance model data (DV).

Customer demographic information (IV1) refers to the basic information of cardholders like age, sex, foreign customers, staff of bank, marriage status, occupation, income, etc. Fifteen variables related to demographic information.

Credit card information (IV2) refers to basic information that is related to credit card like association of credit card, type of credit card, card open date, card end date, card limit and so on.

Credit card transaction information (IV3) refers to active credit card information of credit card holders like transaction date, transaction amount, merchant number and so on.

Credit risk information (IV4) is related with credit card bureau information.

Deposit account (IV5) refers to basic information of deposit open date, deposit end date, deposit amount, balance date and so on.

Product activity (IV6) refers products that have been already active such as count of credit cards, counts of accounts and so on.

Complaints (IV7) refers that the complaints which are related with the credit cards of the bank.

Social media information (IV8) refers that count of pages which are liked by the credit cardholders in Facebook. These pages are related with the bank's social media account.

To make a clear definition of credit cardholders churn through the collected data considered relational research as the search for explanatory and correlational relations between these eight independent variables mentioned above and credit cardholders churn in the research design. According to the light of the below questions, these correlational hypotheses are predicted:

Hypothesis 1:

Social media information influences credit cardholder churn is positively.

Hypothesis 2:

Compliance information influences credit cardholder churn is positively.

Dealing with time is more critical thing in churn analysis. The time zone of this study will be cross sectional. The goal of this churn analysis is to predict what will happen to the credit cardholders in next periods. Therefore, the time will split two phases for observation or training period and performance or target period. In observation period, the independent variables are observed and then the prediction will check whether the credit cardholder became a churner or not during the performance or target period. The independent variables will be calculated from the data that is obtained during observation period while dependent variable is calculated from the performance period. In this study, the data which is collected from an anonym bank is included three months period data from the beginning of the November 2016 to the end of February 2017.

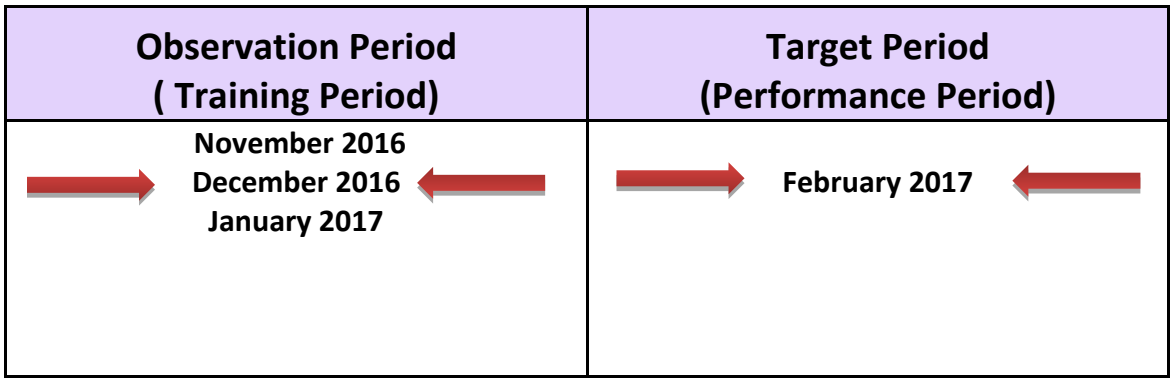


Figure 8. Timeline data set

3.1.3. Data Preparation

The data which is used as an input data in data mining model should be prepared clearly according to the needing. After the data is identified, selected and prepared the next step includes much more variables which are derived from the input data sets. Therefore, the phase of the study includes some aggregations, integration or formatting of the data. New variables are derived from the existing variables to make enrichment of the variables. These derivation operations will have a chance to know customer characteristics behavior. Besides that, enrichment is also important to increase the performance of the model and to increase the accuracy and reliability of the model outputs.

In this study, there are eight independent variables. These necessary data comes from these different eight tables. All of these tables are customer based tables. In this type of project, merging is an assembly operation. Merging refers to collect data for the same key variable. In this study, the key variable is customer number and the credit card account number and all of these eight tables is merged by customer number and credit card account number. In order to supply data security and to protect customer's privacy, all of these keys are masked in this project. These variables in the project do not include the actual data.

There is an another important variable CHURN_FLAG which is demonstrated the churn information of the credit cardholders. CHURN_FLAG has two value that are determined to the rules mentioned in section 3.1.1. Business Understanding. 'T' defines that the credit cardholders is a churner and 'F' defines that the credit cardholders is not a churner.

There is another critical point in data mining project is data cleaning. Data cleaning means that cleaning the noisy, incomplete and inconsistent data which is found in data set. In this step, the missing data is filled and obtain inconsistent data.

In this study, missing values are filled with zero and all rows which is included zero for a customer that are dropped from the model and a row which has no key variable has dropped from the model.

Data transformation is an another critical point to process input data set for data mining model. The data is transferred and consolidated for obtaining appropriate variables or for to reach more accurate data for data mining model.

In this study, birthdate information is transferred in the age and the tenure of the customers is transferred into age for each customer is given as example for data transformation.

To prevent the divergence, some of the variables are calculated due to the normal distribution curve and their quartiles. These variables are calculated for three months period.

The variables of these data set are shown in the Table 1 below. These variables are used as an input data for the data mining models. All of the values of the variables are masked or transformed to not show the real value of the variables.

Variable Name	Variable Explanation
ID	Derivative id to supply uniquely.
CUSTOMER_NO	Masked customer number
CARD_NO	Masked card number
STATUS_CODE	Shows the code customers are active or not
STATUS	Shows the definition customers are active or not
ENGAGEMENT_WITH_BANK	Shows the customers' engagement with the bank
BIRTH_DATE	Shows the customers' birthday date
GENDER_ID	Shows the code of customers' gender
GENDER	Shows the definition of customers' gender
MARITAL_STATUS_ID	Shows the code of customers' marital status
MARITAL_STATUS	Shows the definition of customers' marital status

EXSERVICE_ID	Shows the code of customer' exservice
EXSERVICE	Shows the definition of customers' exservice
EDUCATION_ID	Shows the code of customers' education
EDUCATION	Shows the definition of customers' education
OCCUPATION_ID	Shows the code of customers' occupation
OCCUPATION	Shows the definition of customers' occupation
SEGMENT_CODE	Shows the code of customers' segmentation
SEGMENT	Shows the definition of customers' education
MONTHLY_INCOME	Shows the monthly income of customers
FIRST_USAGE_DATE_DT	Shows the first usage date of credit card by customers
LAST_TXN_DATE_DT	Shows the last transaction date of credit card by customers
LAST_TXN_AMOUNT_M1	Shows the last transaction amount in November
LAST_TXN_AMOUNT_M2	Shows the last transaction amount in December
LAST_TXN_AMOUNT_M3	Shows the last transaction amount in January
LAST_TXN_AMOUNT_3MO	Shows the total last transaction amount since last three months
AUTOMATIC_PAYMENT_M1	Shows that the customers whether have automatic payment or not in November
AUTOMATIC_PAYMENT _M2	Shows that the customers whether have automatic payment or not in December
AUTOMATIC_PAYMENT _M3	Shows that the customers whether have automatic payment or not in January
OTOMATIK_ODEME_TALIMATI_3MO	Shows that the customers whether have automatic payment or not in last three months
CARD_LIMIT_M1	Shows the credit card limit in November
CARD_LIMIT_M2	Shows the credit card limit in December
CARD_LIMIT_M3	Shows the credit card limit in January
CARD_LIMIT_3MO	Shows the total credit card limit amount since last three months
CUSTOMER_LIMIT_M1	Shows the customer limit in November
CUSTOMER _LIMIT_M2	Shows the customer limit in December
CUSTOMER _LIMIT_M3	Shows the customer limit in January
CUSTOMER _LIMIT_3MO	Shows the total customer limit amount since last three months
CASH_LIMIT_M1	Shows the customers' cash limit in November
CASH_LIMIT_M2	Shows the customers' cash limit in December
CASH_LIMIT_M3	Shows the customers' cash limit in January
CASH_LIMIT_3MO	Shows the total cash limit of customers since last three months
MINIMUN_PAYMENT_M1	Shows the customers' minimum payment for a credit card in November

MINIMUN_PAYMENT _M2	Shows the customers' minimum payment for a credit card in December
MINIMUN_PAYMENT _M3	Shows the customers' minimum payment for a credit card in January
MINIMUN_PAYMENT _3MO	Shows the customers' minimum payment for a credit card since last three months
TERM_DEBT_M1	Shows the term debt of credit cardholders in November
TERM_DEBT_M2	Shows the term debt of credit cardholders in December
TERM_DEBT _M3	Shows the term debt of credit cardholders in January
TERM_DEBT _3MO	Shows the term debt of credit cardholders since last three months
MCC_M1	Shows the masked merchant code that has maximum usage in November
MCC_M2	Shows the masked merchant code that has maximum usage in December
MCC_M3	Shows the masked merchant code that has maximum usage in January
MCC_3MO	Shows the masked merchant code that has maximum usage since last three months
ACIKLAMA_M1	Shows the definition merchant code that has maximum usage in November
ACIKLAMA_M2	Shows the definition merchant code that has maximum usage in December
ACIKLAMA_M3	Shows the definition merchant code that has maximum usage in January
ACIKLAMA_3MO	Shows the definition merchant code that has maximum usage since last three months
TRX_TYPE_M1	Shows the transaction type of credit card that has maximum usage in November
TRX_TYPE _M2	Shows the transaction type of credit card that has maximum usage in December
TRX_TYPE _M3	Shows the transaction type of credit card that has maximum usage in January
TRX_TYPE _3MO	Shows the transaction type of credit card that has maximum usage since last three months
TRX_AMOUNT_M1	Shows the transaction amount of credit card that has maximum usage in November
TRX_AMOUNT _M2	Shows the transaction amount of credit card that has maximum usage in December
TRX_AMOUNT _M3	Shows the transaction amount of credit card that has maximum usage in January

TRX_AMOUNT _3MO	Shows the transaction amount of credit card that has maximum usage since last three months
TRX_COUNT_M1	Shows the transaction count of credit card that has maximum usage in November
TRX_COUNT_M2	Shows the transaction count of credit card that has maximum usage in December
TRX_COUNT_M3	Shows the transaction count of credit card that has maximum usage in January
TRX_COUNT_3MO	Shows the transaction count of credit card that has maximum usage since last three months
DEP_OPEN_DATE	Shows the opened date of deposit by customer
DEP_END_DATE	Shows the closed date of deposit by customer
DEMAND_DEP_BALANCE_TL_M1	Shows the demand TL deposits balance of customers in November
TERM_DEP_BALANCE_TL_M1	Shows the term TL deposits balance of customers in November
TERM_DEP_BALANCE_YP_M1	Shows the term foreign exchange deposits balance of customers in November
DEMANF_DEP_BALANCE_YP_M1	Shows the demand foreign exchange deposits balance of customers in November
DEPOSIT_BALANCE_TL_1MO	Shows the total deposit balance of customer in November
DEMAND_DEP_BALANCE_TL_M2	Shows the demand TL deposits balance of customers in December
TERM_DEP_BALANCE_TL_M2	Shows the term TL deposits balance of customers in December
TERM_DEP_BALANCE_YP_M2	Shows the term foreign exchange deposits balance of customers in December
DEMAND_DEP_BALANCE_YP_M2	Shows the demand foreign exchange deposits balance of customers in December
DEPOSIT_BALANCE_TL_2MO	Shows the total deposit balance of customer in December
DEMAND_DEP_BALANCE_TL_M3	Shows the demand TL deposits balance of customers in January
TERM_DEP_BALANCE_TL_M3	Shows the term TL deposits balance of customers in January
TERM_DEP_BALANCE_YP_M3	Shows the term foreign exchange deposits balance of customers in January
DEMAND_DEP_BALANCE_YP_M3	Shows the demand foreign exchange deposits balance of customers in January
DEPOSIT_BALANCE_TL_3MO	Shows the total deposit balance of customer in January
TOTAL_DEPOSIT_BALANCE_3MO	Shows the total deposit balance of customer since last three months

OPENED_ACCOUNT_COUNT_M1	Shows the count of customer's opened account in November
CREDIT_CARD_COUNT_M1	Shows the count of credit card of customers in November
CREDIT_COUNT_M1	Shows the credit count of customers in November
ACCOUNT_COUNT_M1	Shows the count of account of customers in November
OPENED_ACCOUNT_COUNT_M2	Shows the count of customer's opened account in December
CREDIT_CARD_COUNT_M2	Shows the count of credit card of customers in December
CREDIT_COUNT _M2	Shows the credit count of customers in December
ACCOUNT_COUNT _M2	Shows the count of account of customers in December
OPENED_ACCOUNT_COUNT_M3	Shows the count of customer's opened account in January
CREDIT_CARD_COUNT_M3	Shows the count of credit card of customers in January
CREDIT_COUNT _M3	Shows the credit count of customers in January
ACCOUNT_COUNT _M3	Shows the count of account of customers in January
COMP_CATEGORY	Shows the compliance category that is made by credit cardholders
COMP_PRODUCT	Shows the compliance product that is made by credit cardholders
TOTAL_COMP_COUNT	Shows the total compliance count that is made by credit cardholders
FACE_COUNT_LIKE	Shows the total count of facebook like which are related with page of bank.
FACE_COUNT_OTHER_LIKE	Shows the total count of facebook like which are not related with page of bank.
CHURN_FLAG	Shows the information of customer is a churner or non-churner

Table 1. Used Variables in Data Mining Models

3.1.4. Data Quality

There are many different functions of data mining are mentioned previous sections. These functionalities are classification, clustering, future selection and association rule mining. In this study, classification analysis was used on the data set. The classification algorithms learn from the training data set and then the model is prepared to classify the new

objects or make predictions for the new data set. These classification algorithms are applied on decision tree method, naive bayes method, generalized linear model and support vector machine.

The variables which are listed in Table 2 below. These variables are main variables that affect the credit cardholder churn model directly. In other words, this list showed a clear correlation between listed variables and target variable, CHURN_FLAG.

Name	Rank	Importance
CUSTOMER_LIMIT_3MO	1	0,457
TRX_AMOUNT _3MO	2	0,372
TRX_COUNT_3MO	3	0,372
TRX_TYPE_M3	4	0,371
TRX_COUNT_M3	4	0,3454
TRX_AMOUNT_M3	5	0,345
MCC_M3	6	0,344
TRX_COUNT_M2	7	0,259
TRX_AMOUNT_M2	8	0,2588
MCC_M2	9	0,2586
TRX_TYPE_M2	10	0,2585

Table 2. The variables that have correlation directly with the target variable

Correlation is a statistical technique which is able to whether and how strongly pairs of variables are related with each other. In other words, correlation measures the strength and direction of the relationship between at least two or more variables. It can be seen clearly, there is a strong correlation between these variables in the table above and the target variable. For example, TRX_AMOUNT _3MO that refers to the transaction amount of credit card that has maximum usage since last three months is going to influence the certain amount of CHURN_FLAG. That means that, if the credit cardholder is making any payment in three months, the credit cardholder will not cancel the credit card. Because of this reason the variables which have strong correlations with the CHURN_FLAG do not settled in churn model.

3.1.5. Data Sampling

Sampling is an efficient way for predictive analytics to produces accurate results from the models and run these models large data set. It allows to study with a small data set in order to build an analytical more quickly.

- The balance of CHURN_FLAG is without any sampling in raw data is shown in Figure 9 and Table 2.

Stratify Name	Stratify Value	Source Count	Source Distribution
CHURN_FLAG	T	15000	0,5
CHURN_FLAG	F	15000	0,5

Table 3. Raw data target ratio

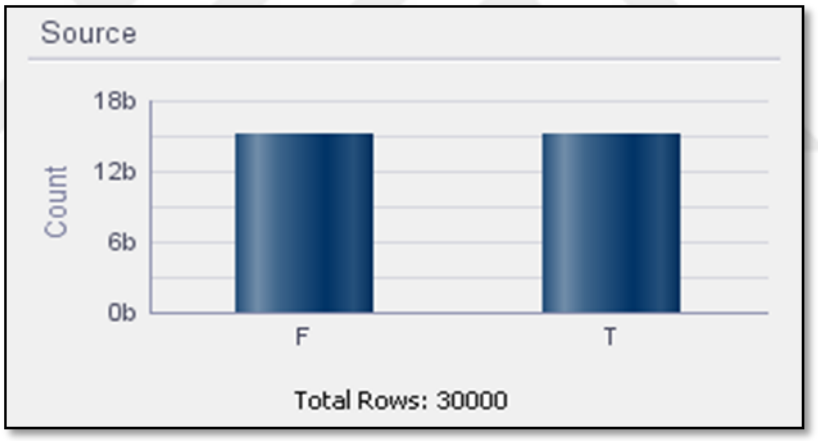


Figure 9. Raw data target ratio in a graph

In this study, sample data is produced in order to save the time and decrease the manpower. The sample data encountered the data which are settled in observation period. These sample data which has the same characteristic features with the raw data is produced with ODM. The data is sampled by three times to obtain best prediction model from the given data set. First of all, the data set is separated %60 -%40 rates due to the CHURN_FLAG demonstrated in the tables below. %60 of the raw data set was used for training and %40 of the data set was used for testing steps. Secondly, the data set is

separated %80 -%20 rates due to the CHURN_FLAG demonstrated in the tables below. %80 of the raw data set was used for training and %20 of the data set was used for testing steps. Thirdly, the data set is separated %90 -%10 rates due to the CHURN_FLAG demonstrated in the tables below. %90 of the raw data set was used for training and %10 of the data set was used for testing steps. The methods, decision tree, naïve bayes, generalized linear model and support vector machine, which are taken part in classification algorithms is used to make predictions. The performance of these algorithms was compared to choose the best sampled data and to determine which algorithms can be applicable to define the credit cardholders churn by comparing of the performance values.

In classification algorithms, to obtain and asses more accuracy result, the performance of a model need to be reviewed. The performance of the model consists of the measurements are percent of prediction confidence, percent of overall accuracy, percent of average accuracy and the cost of the model.

Prediction confidence demonstrate that the estimation of the model accuracy. These value is a number that is between 1 and 0. Prediction confidence shows that how much better predictions made by the tested model.

- Predictive confidence is 0 means that the prediction of the model is not a s better than the prediction of the naïve model,
- Predictive confidence is 1 means that the prediction of the model is perfect,
- Predictive confidence is 0.5 means that the prediction of the model reduce the error of the naïve model by 50 percent.

Average accuracy demonstrates that the percentage of the correct prediction of the prediction model when compared with the actual values of the naïve model. That is formulated by the following equation:

$$Average\ Accuracy = \left(\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} + \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FN}} \right) / \ Number\ of\ classes * 100$$

Equation 2. Formulation of Average Accuracy

While TP indicates true posivite values, FP indicates false positive values. While TN indicates true negative values, FP indicate false negative values. The details of these measurements have more details in the next section in Confusion Matrix.

Overall accuracy demonstrates that the percentage of the correct prediction of the prediction model when compared with the actual values of the naïve model. That is formulated by the following equation:

$$Overall\ Accuracy = \left(\frac{TP + TN}{TP + FP + TN + FN}\right) * 100$$

Equation 3. Formulation of Overall Accuracy

Similarly, While TP indicates true posivite values, FP indicates false positive values. While TN indicates true negative values, FP indicate false negative values. The details of these measurements have more details in the next section in Confusion Matrix.

The cost of the predictive model may be important to describe the cost for making in correct decision in classification problem. Cost measurement can be more useful when the different misclassifications vary significantly.

3.1.5.1. Measurements of Sampled Data by %60 -%40

The sample data which are separated %60 -%40 rates due to the CHURN_FLAG demonstrated in the tables below. %60 of the raw data set was used for training and %40 of the data set was used for testing steps.

Name	Prediction Confidence %	Overall Accuracy %	Average Accuracy %
Decision Tree	61,3311	80,6655	80,6655
Generalized Linear Model	42,8644	71,4322	71,4322
Naive Bayes	27,5316	63,7658	63,7658
Support Vector Machine	43,5046	71,7523	71,7523

Table 4. Performance comparison of models with %60 -%40 sample data

3.1.5.2. Measurements of Sampled Data by %80 -%20

The sample data which are separated %80 -%20 rates due to the CHURN_FLAG demonstrated in the tables below. %80 of the raw data set was used for training and %20 of the data set was used for testing steps.

Name	Prediction Confidence %	Overall Accuracy %	Average Accuracy %
Decision Tree	63,1831	81,5915	81,5915
Generalized Linear Model	42,2473	71,1236	71,1236
Naive Bayes	28,8934	64,4467	64,4467
Support Vector Machine	42,7254	71,3627	71,3627

Table 5. Performance comparison of models with %80 -%20 sample data

3.1.5.3. Measurements of Sampled Data by %90 -%10

The sample data which are separated %90 -%10 rates due to the CHURN_FLAG demonstrated in the tables below. %90 of the raw data set was used for training and %10 of the data set was used for testing steps.

Name	Prediction Confidence %	Overall Accuracy %	Average Accuracy %
Decision Tree	62,542	81,271	81,271
Generalized Linear Model	42,2999	71,15	71,15
Naive Bayes	29,0518	64,5259	64,5259
Support Vector Machine	41,0222	70,5111	70,5111

Table 6. Performance comparison of models with %90 -%10 sample data

3.1.5.4. Selection of the Model and the Sampled Data

All of the steps of data mining processes; business understanding, data description, data preparation, data quality; were done, the next step will have been developing the model to obtain the best prediction result. In this study, that is mentioned previous section, the data set is sampled and four different data mining methods are applied the data set. Performance results of the algorithms are used to determine which sampled data gives the best prediction result and to select which data mining algorithm can be applicable to make predictions.

When analyzing the performances of the sampled data results, it is obviously visible that the overall accuracy and the average accuracy have same results approximately. That is the point that this result is a desired result to obtain best prediction models. The values of the average accuracy and overall accuracy have closed values indicates that the data mining model is working. That means that the models do not memorized anything, the models are learning and are able to make reliable predictions on general untrained data.

When making comparison all of the sampled data in itself, decision tree method approach had the best result obviously. Therefore, decision three method can be applicable to make the definition of credit cardholders churn. When making comparison all of the sampled data with each other, the second sampled data which has %80 of the raw data set was used for training and %20 of the data set was used for testing steps has the best prediction and the accuracy values.

In this study, the data set was sampled by %80 for training and %20 for testing steps. The decision tree model was used to make predictions.

3.1.6. Decision Tree Method Application

Decision three is the most basic and common machine learning technique. In classification algorithm, it is the easiest technique to implement and understand the data. Besides these positive features, in this study, the performance results show that the measurement of accuracy results and percentage of predictions are much more reliable results.

In this study, because all of these reasons explained above, decision tree method as a machine learning algorithm was used.

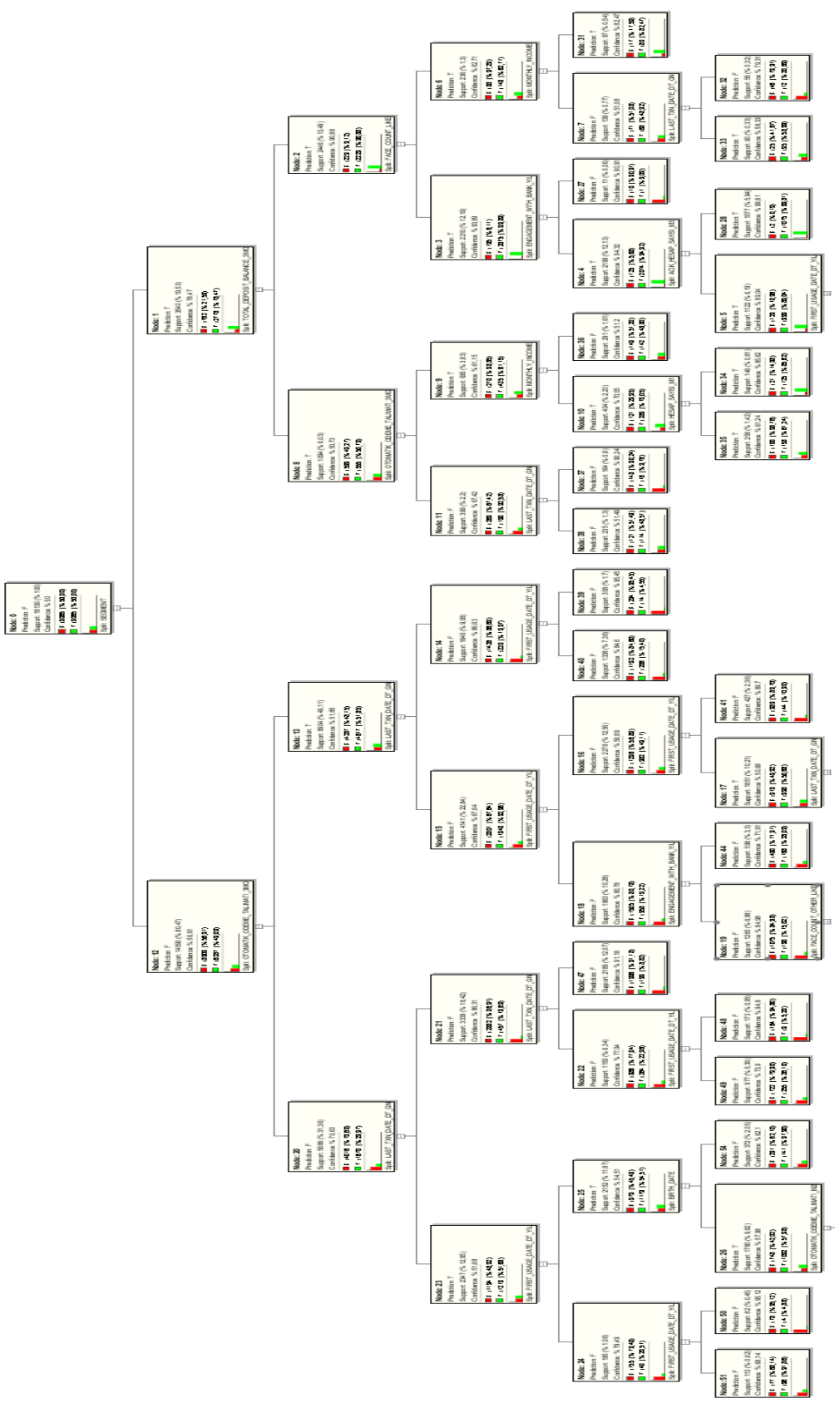


Figure 10. Screenshot of Decision Tree Method

When the decision tree is reviewed, making an interpretation about these variables that affect model significantly:

- Segment
- Total_Dep_Balance_3MO
- Face_Count_Like
- Engagement_with_Bank
- Open_acc_count

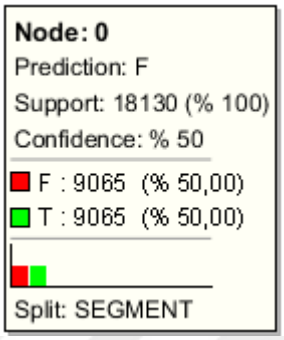


Figure 11. Screenshot of Node 0 of the Decision Tree

First of all, while the credit cardholders churn (T) rate is percent 50 of records in the data set, similarly the non-churner credit cardholders (F) rate is percent 50 of the records in data set. The first split is for segment variable. This points that the values of the segment variable has the most important indicator between churners and non-churners of credit cardholders. The node which is settled left side is shows the percentage of non-churners credit cardholders, the right node of the tree is classified the percentage of the churner credit card holders.

Figure 12 represents the churner credit cardholders below:

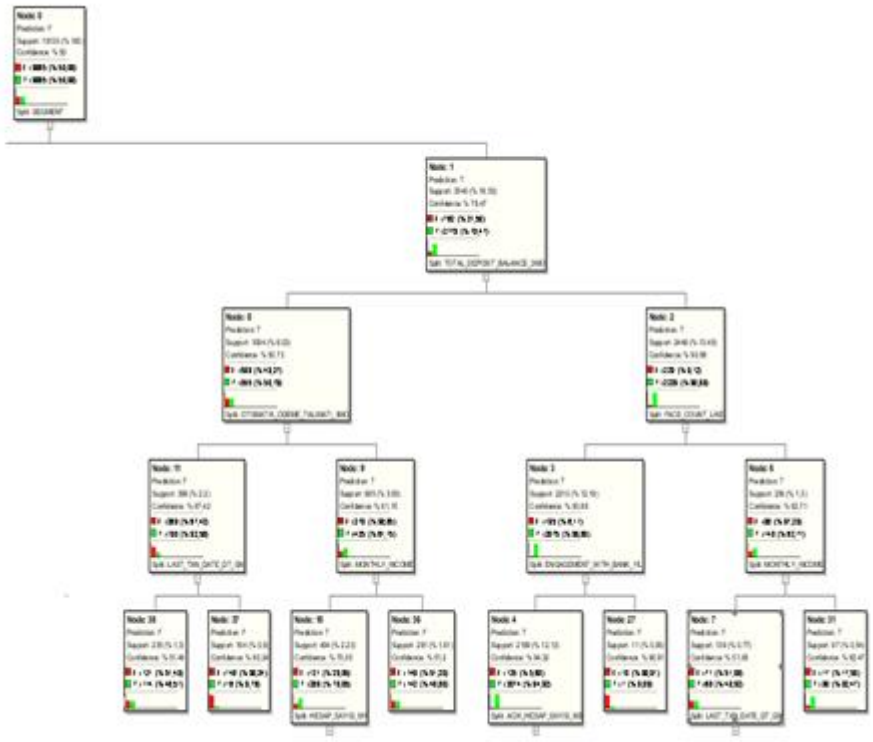


Figure 12. Churners part of the Decision Tree

At the beginning the credit carholder churner's rate in main data is 50 percent of records. The first split is on Segment. This implies that the greatest and most valuable difference between credit cardholder churners and non-churners is Segment. In Node 1, Total_Deposit_Balance_3MO, 78.47 percent of records represent churners, while in Node 2, Face_Count_Like, 90.88 percent do. Clearly, a record in Node 2 is more likely to represent churners than a record in Node 8. For this reason splitting operations continue using this branch. In node 3 is represent Engagement_with_Bank the concentration of churners rises from 90.88 percent to 93.89 percent. In node 4 which is the last leaf of the decision tree is shows Opened_Account_Count rises credit cardholders churners from 93.89 to 94.32.

There is an another important point in this decision tree. This decision tree also supports one of the hypotheses of this study. This hypothesis is "Social media information of customers influences credit cardholder churn positively". In node 2 represents the Face_Count_Like shows the count of like which likes are related with the banking pages. In Node 2, the concentration of the credit cardholders churners rise from 50 percent which is in

first split to 90.88 percent. Therefore social media information influences credit cardholder churn positively.

On the other hand, one of the hypothesis of the study is related with complaints of the customers. These complaints are related with the credit cards. The hypothesis is "Complaint information of customers influences credit cardholder churn positively". As it seen in the decision tree there is no information about complaints data. In this study, complaints do not affect the behavior of the credit cardholders.

Figure 13 represents the non-churner credit cardholders below:

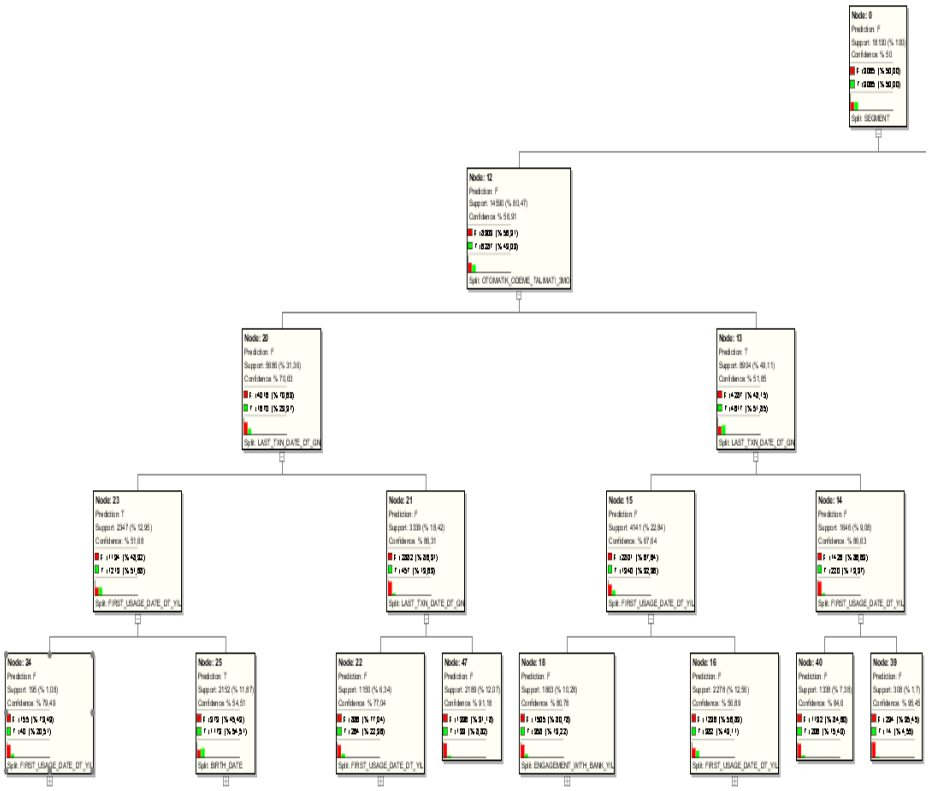


Figure 13. Non – churners part of the Decision Tree

At the beginning the credit cardholder non-churner's rate in main data is 50 percent of records. In Node 20, 70.63 percent of records which is Last_Trx_Date represent non-

churners credit cardholders. In node 21, the concentration of churners rises from 70.63 percent to 86.31 percent. And at the last node the concentration of churners increases to 91.18 percent.

3.1.7. Evaluation and Scoring

A threshold model is any model where a threshold value in mathematical or statistical modelling. This value is used to distinguish ranges of values where the behaviour predicted by the model varies in some important way. (Dodge Y, 2003)

	Minimum Probability Value	Maximum Probability Value
T	0,82	1,00
F	0,05	0,69

Table 7. Threshold value of Churn

4. FINDING AND RESULTS

4.1. Lift Curve

In direct marketing lift curve is a very popular technique. This is a technique to measure the effectiveness of the data mining model. The lift curve ratio is measured by between the results that are obtained with and without a predictive model. The lift charts shows that how rapidly a model finds the actual positive target values.

While the x-axis of the graph is divided by quantiles, the y-axis shows the cumulative positive results. In this study, the quantile or threshold value is selected at 10 that is indicated by blue line vertically. While the red line is indicated the random guess line , the light green line is indicated the ideal line for the decision tree model. The dark green is the line that is represented the decision tree model.

From these result, the ideal line of the lift chart peaks at around 50 percent that means that if the model is a perfect model, it could be reachable 100 percent of the targeted customers who are credit cardholder churners by the using of the decision tree model result to only 50% of the total population. The actual lift for the decision tree model when 50 percent of the population is taken as a targeted customers is at around %80. In other words, it means that %80 of the credit cardholders churner can be predictable with using decision three model.

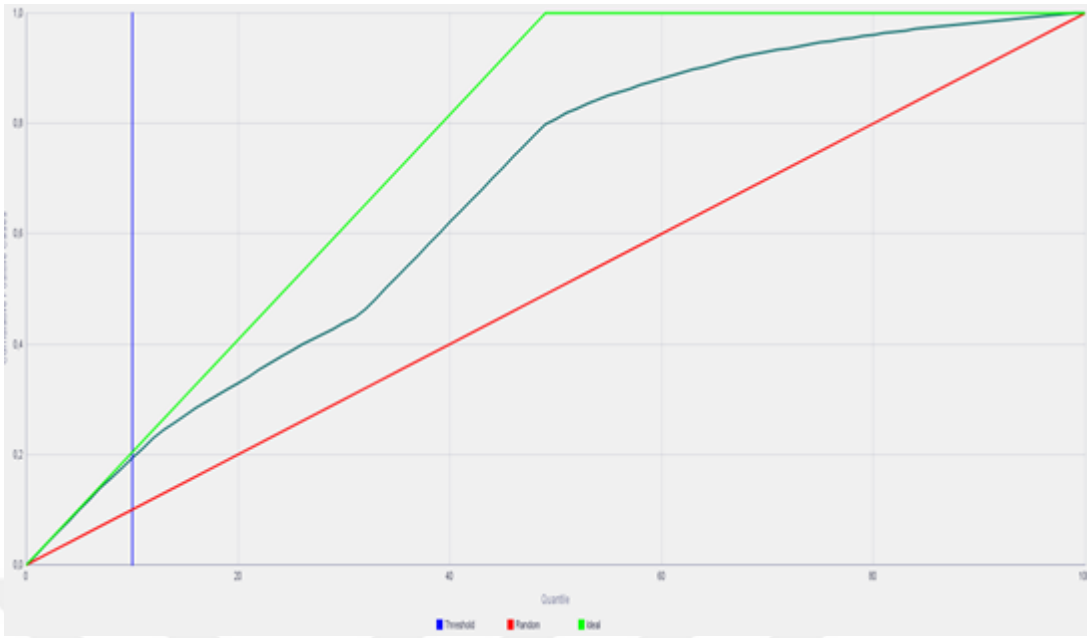


Figure 14. Lift Curve of Decision Tree Model

4.2. Confusion Matrix

A confusion matrix, also known as a performance matrix, contains information about actual and predicted classifications done by classification system. (Santra and Christy, 2012). The table 7 shows the entries of the confusion matrix.

	Predictive Negative	Predictive Positive
Actual Negative	TN	FP
Actual Positive	FN	TP

Table 8. Confusion Matrix Entries

While the rows represented the actual values, columns represented the predicted values.

- TN symbolized the true negative value that is the number of the correct predictions of the negative examples. In other words, negative examples classified as a negative example correctly.
- FP symbolized the false positive value that is the number of the incorrect predictions of the negative examples. In other words, negative examples classified as a positive example incorrectly.
- FN symbolized the false negative value that is the number of the incorrect predictions of the positive examples. In other words, positive examples classified as a negative example incorrectly.
- TP symbolized the true positive value that is the number of the correct predictions of the positive examples. In other words, positive examples classified as a positive example correctly.

The accuracy is represented as AC is calculated by the summation of the correct predictions is divided by the summation of the all prediction values. (Santra and Christy 2012). It can be formulated by the following equation:

$$Accuracy\ (AC) = \left(\frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \right)$$

Equation 4. Formulation of Accuracy

Recall, also known as sensitivity or true positive rate (TPR), is calculated by the true positives value divided by the summation of the false negative and true positive values. (Brennan, 2012). Recall can be defined as a measure of the completeness or quantity. (Hong Hu, Jiuyong Li, Ashley Plank, 2006) (Milan Kumari, Sunila Godara, 2011). It can be formulated by the following equation:

$$Recall\ (TPR) = \left(\frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \right)$$

Equation 5. Formulation of Recall

Precision is also known as a true positive predictive rate (PPV) , is calculated by the true positives value divided by the summation of the false positive and true positive values.

(Brennan, 2012). Recall can be defined as a measure of the exactness or quality. (Hong Hu, Jiuyong Li, Ashley Plank, 2006) (Milan Kumari, Sunila Godara, 2011). It can be formulated by the following equation:

$$Precision\ (PPR) = \left(\frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}\right)$$

Equation 6. Formulation of Precision

Miss Rate is also known as False Negative Rate (FNR). False negative rate refers the proportion of positives cases. These positive cases were incorrectly classified as negative. It can be formulated by the following equation:

$$\text{Miss Rate (FNR)} = \left(\frac{\mathrm{FN}}{\mathrm{FN} + \mathrm{TP}}\right)$$

Equation 7. Formulation of Miss Rate

Fall-out is also known as False Positive Rate (FPR). False positive rate refers that the proportion of negative cases. These cases were classified incorrectly as positive. It can be formulated by the following equation:

$$Fall - out\ (FPR) = \left(\frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}\right)$$

Equation 8. Formulation of Fall-out

Specificity is also known as True Negative Rate (TNR). True negative rate refers that the proportion of negatives cases. These negative cases were classified correctly. It can be formulated by the following equation:

$$Specificity\ (TNR) = \left(\frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}\right)$$

Equation 9. Formulation of Specificity

In this study, the decision three method is used and the confusion matrix of the model is showed in Table 8 below. Whereas the rows represented the actual values, columns represented the predicted values.

	F(Non-churner)	T(Churner)
F (Non-churner)	4.844	1.091
T (Churner)	1.204	4.731
Total	6.048	5.822
Correct %	80,0926	81,2607
Cost	2.408	2.182

Table 9. Confusion Matrix of Decision Tree

For above confusion matrix, the first result cell, which contains the value 4.844, indicates the number of true negative for the value F. Because F indicates that the customer who are non-churner credit cardholders, this statistic tells that model predicted the correct value for non-churner credit cardholders in 4.844 cases. The cell which contains the value 1.901, shows the number of false positive, or how many times the model predicted that someone would be a churner when actually they did not.

The cell that contains the value 4.731 indicates the number of true positives for the value T. Because T means that the credit cardholder is a churner. In other words, In 4.731 cases the model correctly predicted that someone would be a churner. The cell that contains the value 1.204 indicates the number of false negative for the value T. This statistic tells that in 1.204 cases, the model predicted someone would not be non-churner when in fact they were churner.

While 6.408 represent the prediction of the total prediction count that is predicted incorrectly, 5.822 represent the prediction of the total prediction count that is predicted correctly.

Diagonal elements of matrix 4.844 + 4.731 = 9575 represent the correct prediction count is classified and the other elements of the matrix 1.091 + 1.206 = 2297 represent incorrect prediction count.

Precision for churner credit cardholders' count is predicted is 4.731 / (1.091 + 4.731) = 0.713. That means that 0.713 times the model predicted the positive cases correctly.

Recall is 4.731 / (4.731 + 1.204) = 0.797. That means that 0.797 times the model predicted the actual positive cases correctly.

Miss Rate is 1.204 / (4.731 + 1.204) = 0.181. That means that 0.181 times the model predicted the actual negative cases correctly.

Fall out is 1.204 / (1.204 + 4.844) = 0.314. That means that 0.314 times the model predicted the negative cases that were incorrectly as positive.

Specificity is 4.844 / (4.844 + 1.901) = 0.718. That means that 0.718 times the model predicted negative cases as negative cases correctly.

Accuracy of the model is (4.731 + 4.844) / (1.091 + 4.731 + 4.844 + 1.2014) =. 0.755. This is the proportion of the churners or non-churners' count of the predictions that is predicted by the model correctly.

The summary of these performance indicators' calculation is represented in Table 9 below:

		Predictive (Model)			
		Negative	Positive		
Actual	Negative	4.844	1.091	*Fall Out or FPR*	0.314
				Miss Rate or FNR	0.203
	Positive	1.204	4.731	*Recall or Sensitivity or TPR*	0.797
		Specificity or TNR	*Precision or PPV*	Accuracy = 0.755	
		0.718	0.713		

Table 10. Performance Indicators of Decision Tree

4.3. ROC Curve

The ROC is an abbreviation of Receiver Operating Characteristics. The ROC supports what-if analysis. The ROC curve is a graphical plot. The curve is plotting with the values of true positive rate against the false positive rate at different threshold. While the true positive rate is also known as sensitivity, recall, false positive rate is also known as fall out or probability of false alarm in machine learning. As with any classification process, the methods deal with establishing a balance between precision (ability to eliminate false positives) and sensitivity (ability to detect correct positives). Since the positive and negative samples in the data set do not distribute equally, ROC curve was used to evaluate the

balance between precision and sensitivity. The area under the ROC curve can be defined as the ROC score. The ROC curve shows the number of correct positives according to the changing classification thresholds. If the ROC score 1, that means that positive values are separated from negative values clearly. On the other hand, if the ROC score is 0 that means that there is no positive values in the model. (Powers, David M W, 2011)

According to Thomas (2010) which is the one of the ways to find model accuracy is measured by the area under the ROC curve. An area of 1 represents a perfect test; an area of 0.5 represents a worthless test. A rough guide for classifying the accuracy of a diagnostic test is the traditional academic point system:

- 0.90-1 = excellent (A)
- 0.80-.90 = good (B)
- 0.70-.80 = fair (C)
- 0.60-.70 = poor (D)
- 0.50-.60 = fail (F)

The Figure 15 represent the ROC curve of the decision three model for credit cardholder churner. According the plot, the area under the ROC curve is 0.8313 . Due to point system of Thomas the accuracy of the diagnostic test is grading good (B). In other words, the model can be usable for decision making in the marketing strategies.

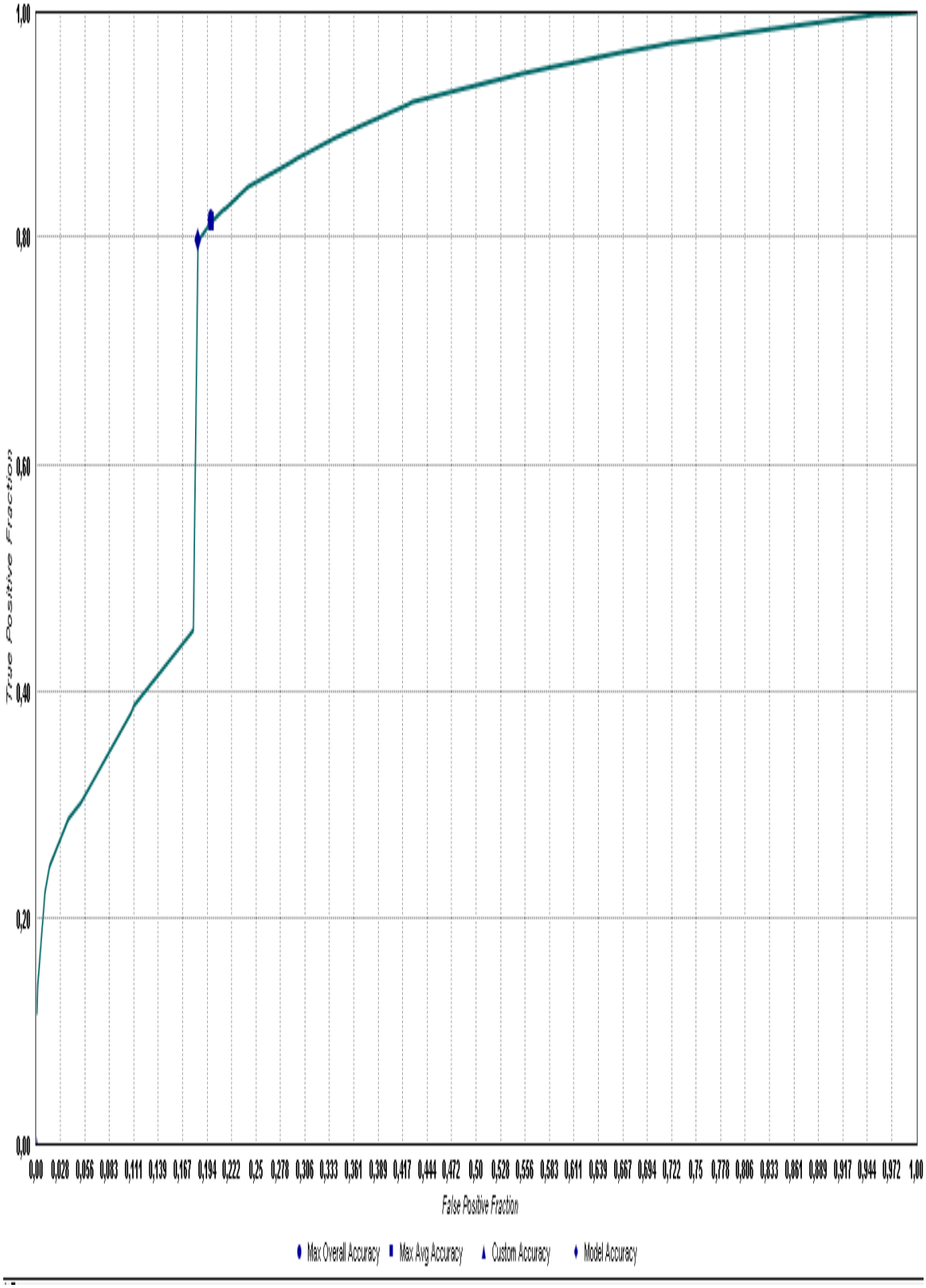


Figure 15. ROC curve of Decision Tree Method

5. LIMITATION

In this section of the study, the limitations of the study will be discussed.

One of the limitation of this study is the content of the data. The related data are collected from an anonymous private bank for the descriptive credit card churn and predicting credit card churn. The related are mainly included customer's demographic information, credit card information, credit card transactions, credit risk information of credit cardholders, deposit accounts of credit cardholders, product activity of credit cardholders, complaints data. The related of the data can be expanded with any other information of the customer. These information is able to expanded with any other adding variables such as investment of the customer or borrowing of the customer to the other bank.

Second limitation of this study is the sample size. The related dataset is including 30.000 credit cardholders which is composed of 15.000 credit cardholder churners, the other half is composed of credit cardholders who are non-churners. This means that, in order to take much more reliable results, the study should be enlarged by the count of the customers.

Third limitations of this study is about time series of the data. Dealing with time is more critical thing in churn analysis. In this study, the time will split two phases for observation or training period and performance or target period. For this study, the data which is collected from an anonym bank is included three months period data from the beginning of the November 2016 to the end of February 2017. For observation or training period, the period of the data set can be expanded to at least two years.

Fourth limitation of this study is about social media information which is related with one of the hypothesis. Social media information is limited to the Facebook data. These data set can be expanded with any other social media accounts.

Last limitation of this study is data mining techniques. That means that there is only one data mining technique was used in this study that was classification. This study can be

reviewed with any other data mining technique such as clustering. So that, different views can be compared.

6. CONCLUSION

In this study, the data was collected from an anonym private bank. The data is based on credit cardholders of this bank. The churn definition is characterized according to the data set of this bank. At the end of this study, there are some proposed a process of churn prediction of credit cardholders in an anonym private bank.

The main data set was consisted of credit cardholders of the anonym private bank. The data set merged with the tables that included customer demographic information, credit card information, credit card transactions, credit card risk information, deposit accounts, product activity, complaints, and social media information. The data set was reorganized as a customer based data in basic tables. Finally, the tables were merged by credit card numbers to process the input data set for data mining model.

In order to supply data security and to protect customer's privacy, all of key variables were masked in this project. These variables in the project did not include the actual data.

For this study, 30.000 unique customers who had credit card in the bank were reviewed. These 30.000 unique customers are based on customers that 15.000 unique customers who cancelled their credit cards in the last month, the 15.000 unique customers who were the active credit cardholders in the last month. The data mining process, the models were built with 85 chosen variables. The data was sampled as a training and test data test. After all, studying on data the next month values were used to predict the credit cardholders will be a churner or not in next period.

Four data mining methods were applied to the input data set to review which model is much more applicable for credit cardholders churn analysis. These methods were decision tree method, naïve bayes method, generalized linear model and support vector machine method. The methods were helped to make a classification for credit card holders who were churner or non-churners. In addition to classification, these methods demonstrated the correlations of the variables in data set. The variables which have more correlations between the churn flag were eliminate from the data set. After the data quality was supplied, the data set was used as an input for the model processes.

The model was prepared ant the four method which are described were applied to the data set. Performance results of the algorithms are used to determine which sampled data gives the best prediction result and to select which data mining algorithm can be applicable to make predictions. When making comparison all of the sampled data in itself, decision tree method approach had the best result obviously. Therefore, decision three method can be applicable to make the definition of credit cardholders churn. When making comparison all of the sampled data with each other, the second sampled data which has %80 of the raw data set was used for training and %20 of the data set was used for testing steps has the best prediction and the accuracy values.

This study demonstrates importance of social media information of customers on credit cardholders. That means that the study sports one of the hypothesis. This hypothesis is "Social media information of customers influences credit cardholder churn positively". On the other hand, the study is related with complaints of the customers. These complaints are related with the credit cards. The hypothesis is "Complaint information of customers influences credit cardholder churn positively". As it seen in the decision tree there is no information about complaints data. In this study, complaints do not affect the behavior of the credit cardholders.

Previous studies show that the keeping existing customers is one of the most critical challenges for all service industries. Today, organizations know to build long term relationships have much more profit than to follow a customer driven vision. For this reason, organizations which are especially in service sectors focus on their existing customers and their products to prevent customer churn. In this point, data mining has a critical and an important role to help organization's decision making. This study can assist to the organizations for their their marketing strategies. This bank is able to take an account to know their credit cardholders' attitudes according to the result of the data mining methods.

The model shows that the social media information has much more influence on credit cardholder churn positively. This information shows that the credit cardholders who has a facebook account and who like the social media pages are related with this anonym private bank could not cancel their credit cards. Therefore, at the end of this study, the anonym private bank knows that it's customers are interacting with brands through social media. If

the bank is having a strong social media marketing plan and presence on the web is the key to tap into their interest. If implemented correctly, marketing with social media is able to bring remarkable success to your business. According to the result of this situation, the bank can improve their marketing strategies with the social media. In addition to the importance of the social media, the study shows that the complaints about credit card has not much more influences on credit cardholders churn directly. The bank knows that the focusing on social media accounts is much more valuable than focusing on customer complaints.

According to the this study, the bank can create new campaigns or improve their campaigns for its credit cardholders to use social media accounts much more effective. The customers' like on facebook shows that whether a credit cardholder is a churner or non-churner. If a credit cardholder starts to like other banks pages, the customers can lead to be a churner, and the bank can follow the customer easily then has a chance to prevent this churn. With these developing marketing strategies, the bank increase the loyalty and the life cycle of the customers and make more profit in the market.

For future process, it is recommended to apply recent and much more powerful machine learning algorithms such as Artificial Neural Network (ANN), clustering algorithms to build an alternative churn model. Then, additional parameter, which can be any other social media accounts like twitter, instagram etc…, can be added to the training data set that can affect to the result of the churn statements. Also, the time period can be more keep more wide to obtain much more accurate results. The time period can be expanded with the historical data which should be at least two years historical data to obtain much more accurate results.

7. REFERENCES

Roberts, J. H. (2000). Developing new rules for new markets. Journal of the Academy of Marketing Science, 28(1), 31–44.

Nie Guangli; Rowe Wei; Zhang Lingling; Tian Yingjie; Shi Yong (2001). Credit card churn forecasting by logistic regression and decision tree. Journal of the Expert Systems with Applications 38 (2011) 15273‑15285

O'Sullivan, Arthur; Steven M. Sheffrin (2003). Economics: P55rinciples in action. Upper Saddle River, New Jersey 07458: Pearson Prentice Hall.ISBN 0-13-063085-3.

Singh Hergovind, Samalia Harsh Vardhan (2014). A Business Intelligence Perspective for Churn Management. Journal of the Science Direct. Procedia - Social and Behavioral Sciences 109 (2014) 51 – 56

Blanchflower, David G., & Evans David S., & Oswald Andrew J. (1988). Credit Cards and Consumers. Dartmouth College, June 1998, National Economic Research Associates, Cambridge, MA.

Wallace, Danny P. (2007). Knowledge Management: Historical and Cross-Disciplinary Themes. Libraries Unlimited. pp. 1–14. ISBN 978-1-59158-502-2.

Boulding, Kenneth (1955). "Notes on the Information Concept". Exploration. Toronto. 6: 103–112. CP IV, pp. 21–32.

Rowley, Jennifer (2007). "The wisdom hierarchy: representations of the DIKW hierarchy". Journal of Information and Communication Science. 33 (2): 163–180. doi:10.1177/0165551506070706.

Bellinger Gene, Durval Castro and Anthony Mills (2004). "Data, Information, Knowledge, and Wisdom.

Dedić, N. and Stanier C. (2016). "An Evaluation of the Challenges of Multilingualism in Data Warehouse Development" in 18th International Conference on Enterprise Information Systems - ICEIS 2016, p. 196.

Fayyad Usama, Piatetsky-Shapiro Gregory, and Smyth Padhraic (1996). “From Data Mining to Knowledge Discovery in Databases”. Al Magazine, 17(3): 37-54

Shafique Umair, Qaiser Haseeb (2014). “A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA)” . International Journal of Innovation and Scientific Research

Rud, Olivia (2009). Business Intelligence Success Factors: Tools for Aligning Your Business in the Global Economy. Hoboken, N.J: Wiley & Sons. ISBN 978-0-470-39240-9.)

Mehmed Kantardzic, (2003). Data Mining: Concepts, Models, Methods, and Algorithms. Wiley-IEEE Press:Totowa, NJ, 2003, ISBN 0-471-22852-4 $59.95

New York Federal Reserve Bank (August, 2015). Quarterly Report on Household Debt and Credit

http://www.investopedia.com/terms/c/creditcard.asp

Kamber Micheline, Han Jiawei (2006). Data Mining: Concepts and Techniques

Cortez Paulo, Santos Manuel Filipe (2013). Recent Advances on Knowledge Discovery and Business Intelligence. Journal of the Expert Systems

Chu, B. H., Tsai, M. S., & Ho, C. S. (2007). Toward a hybrid data mining model for customer retention. Knowledge-Based Systems , 20, 703-718.

Glady, N., Baesens, B., & Croux, C. (2009). Modeling churn using customer lifetime value. European Journal of Operational Research , 197, 402-411.

Han, J. and Kamber, M., (2006). Data Mining: Concepts and Techniques, Second Edition, Elsevier Inc.

J. MacLennan, Z.H. Tang, B. Crivăţ .(2008). Data Mining with Microsoft SQL Server 2008. Wiley Publishing

Geppert C., (2002). Customer Churn Management, KPMG International, p.7.

Ennew, C. T., & Binks, M. R. (1996). The impact of service quality and service characteristics on customer retention: Small businesses and their banks in UK. British Journal of Management , 7 (3), 219-230.

Dixon, M. (1999). 39 Experts predict the future. America's Community Banker, 8(7),20‑31.

Slater, S. F., & Narver, J. C. (2000). Intelligence generation and superior customer value. Journal of the Academy of Marketing Science, 28(1), 120‑127.

Reichheld, F. F. (1996). Learning from customer defections. Harward Business Review ,74, 56-69.

Reichheld, F. F., & Sasser, W. E. Jr., (1990). Zero defections: Quality comes to service. Harvard Business Review, 68(5), 105‑111.

Larose,D., (2005). Discovering Knowledge in Data An Introduction To Data Mining, John Wiley & Sons, Inc., Publication.

Colin Shearer, (2000). "The CRISP-DM Model: The New Blueprint for Data Mining", Journal of Data Warehousing, Volume 5, Number 4, page. 13-22.

Chapman, P. et al, (2000). CRISP-DM 1.0 - Step-by-step data mining guide

Huang Ming, Niu Wenying, Liang Xu., (2009). An Improved Decision Tree Classification Algorithm Based On ID3 and The Application in Score Analysis, Chinese Control and Decision Conference, 1876-1879.

Christy, J. and Santra, K.. (2012). Genetic Algorithm and Confusion Matrix for Document Clustering, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 2.

Hong Hu, Jiuyong Li, Ashley Plank, " A Comparative Study of Classification Methods for Microarray Data Analysis", published in CRPIT, Vol. 61, 2006.

Milan Kumari, Sunila Godara, "Comparative Study of Data Mining Classification Methods in Cardiovascular Disease Predictions", IJCST, Vol. 2, Issue 2,2011.

Powers, David M W, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation" .Journal of Machine Learning Technologies. 2011

Thomas G. Tape, MD, University of Nebraska Medical Center, http://Gim.Unmc.Edu/Dxtests/Roc3.Htm, 2010

Chai, K.,H. T. Hn, H. L. Chieu, “Bayesian Online Classifiers for Text Classification and Filtering”, Proceedings of the 25th annual international ACM SIGIR conference on Research and Development in Information Retrieval, August 2002

Rish, Irina, “An empirical study of the naive Bayes classifier” . IJCAI Workshop on Empirical Methods in AI.,2001.

Nelder, John, Wedderburn, Robert, "Generalized Linear Models". Journal of the Royal Statistical Society. Series A (General). Blackwell Publishing, 1972

Dodge, Y. “The Oxford Dictionary of Statistical Terms”, (2003)