ATM DEMAND PREDICTION

WITH MACHINE LEARNING METHODS

(MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE ATM’LERDE TALEP TAHMİNİ)

by

Evrim Gençalp, M.S.

Thesis

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

MASTERS

in

INDUSTRIAL ENGINEERING

in the

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

of

GALATASARAY UNIVERSITY

Supervisor: Assoc. Prof. Dr. Müjde Genevois

Aug 2020

This is to certify that the thesis entitled

ATM DEMAND PREDICTION

WITH MACHINE LEARNING METHODS

Prepared by **Evrim Gençalp** in partial fulfillment of the requirements for the degree of **Master of Science in Industrial Engineering** at the **Galatasaray University** is approved by the

Examining Committee:

Assoc. Prof. Dr. Müjde Genevois (Supervisor)
Department of Industrial Engineering
Galatasaray University -------------------------

Assoc. Prof. Dr. S. Emre Alptekin
Department of Industrial Engineering
Galatasaray University -------------------------

Prof. Dr. İhsan Kaya
Department of Industrial Engineering
Yıldız Technical University -------------------------

Date: -------------------------

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor Assoc. Prof. Dr. Müjde Genevois in all sincerity. She always guided me and prevented me from despair. Without her support and insightfulness, I could not accomplish this study. Being her student is a pleasure.

I would like to express my deep gratitude to Kaan, the one that was always there when I needed. His understanding and continuing support was my motivation on this journey.

I would also like to thank my comrade Sarper for his moral support. Knowing that you are walking the same path with someone makes it easier.

And I am so grateful to my Borusan R&D family for burning to first spark of my graduate study.

June, 2020
Evrim Gençalp

TABLE OF CONTENTS

LIST OF SYMBOLS

ATM : Automated Teller Machine

EDA : Exploratory Data Analysis

MAPE : Mean Absolute Percentage Error

MAE : Mean Absolute Error

KNN : K Nearest Neighbor

SVM : Support Vector Machine

SVR : Support Vector Regressor

DOE : Design of Experiments

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The need for cash is an endless need even if cashless payment methods become popular. The number of ATMs is increasing day by day in Turkey. The most important process to keep an ATM in use is cash replenishment process. The proper ATM replenishment policy is related to forecast the demand of money precisely. Just as lack of money is a problem which decrease customer satisfaction, surplus of money is an extra cost to the bank.

In this study, ATM cash demand forecasting is done by using machine learning algorithms with ATM cash withdrawal data. Python programming language is used to create the model. Machine learning is a subset of artificial intelligence that is developing day by day. There are lots of machine learning algorithm. Each algorithm performs well in a certain data structure. In the study, Support Vector Regressor, K'th Nearest Neighbor Regressor, Random Forest Regressor and Gradient Boosting Regressor algorithms are used to predict cash demand of ATM, which is a time series data, and their results are compared. To increase the performance of the prediction model, Exploratory Data Analysis and Feature Selection is applied to the dataset. Because of multivariate structure of the dataset and dynamic nature of demand makes the prediction process complicated, these data processing methods help prediction algorithm to understand the data well.

Firstly, demand prediction concept and its role in banking sector is explained and machine learning methods and working principles are introduced. Then, methodology of the study is presented, and applications and results are shared.

ÖZET

Günümüzde her ne kadar nakitsiz ödeme yöntemleri popüler olmaya başlasa da nakit ihtiyacı bitmeyecektir. Türkiye'de ATM'lerin günden güne artıyor olması bunun göstergesidir. ATM'lerdeki en önemli süreç nakit ihmal sürecidir. Bu süreç ise nakit talebinin doğru tahminlenmesiyle en iyi şekilde işler. ATM'de ihtiyaçtan eksik para bulunması müşteri memnuniyetsizliğine yol açacağı gibi fazlalık ise bankaya maliyet olarak yansıyacaktır.

Bu çalışmada, makine öğrenmesi algoritmaları kullanılarak ATM nakit talep tahminlemesi yapılmıştır. Çalışmada Python programlama dilinden yararlanılmıştır. Makine öğrenmesi, günden güne gelişmekte olan yapay zekanın bir alt koludur. Birçok makine öğrenmesi algoritmasından her biri belirli bir yapıdaki veri ile iyi sonuç verir. Çalışmada, zaman serileri çeşidindeki nakit talep datasının tahminlenmesi için Destek Vektör Regresörü, En Yakın Komşu Regresörü, Rastgele Orman Regresörü ve Gradyan Arttırıcı Regresör algoritmaları kullanılmış ve bu algoritmaların sonuçları karşılaştırılmıştır. Modelin performansını arttırmak için veri seti üzerinde Keşifsel Veri Analizi ve Öznitelik Seçimi çalışmaları yapılarak sonuçlar tekrar karşılaştırılmıştır. Çok değişkenli yapısı ve nakit talebinin kendi dinamik doğası gereği verinin komplike olması nedeniyle bu çalışmalar modelin veri setini daha iyi anlamasına yardımcı olmuştur.

Öncelikle talep tahminleme konsepti ve bankacılıktaki rolü anlatılmış, makine öğrenmesi yöntemleri detaylı şekilde sunulmuştur. Sonra çalışmanın metodolojisi, uygulaması ve sonuçlar paylaşılmıştır.

1. INTRODUCTION

Turkey is one of the leading countries that use the Alternative Distribution Channels in banking sector (Sayar et al., 2007). Internet banking, POS (Point of safe), ATM banking, mobile and digital banking are Alternative Distribution Channels. ATM banking is one of the pioneers among them in terms of start of use.

In Turkey, first ATM was used in 1987. After starting of ATM banking, usage and number of ATMs increase with an accelerating speed. Over the past 10 years, in Turkey, while total number of banking offices has grown only 6% percent (from 9465 to 10118), The number of automated teller machines (ATMs) has grown 98% percent (from 26.692 to 52.685) according to The Banks Association of Turkey. This increase brings the accessibility and convenience of banking services with it. ATMs are open 24 hours a day, providing more convenience to users than traditional banking offices. With the developing technology, as ATM operations vary, the need for banking offices is decreasing and it is obvious that; in the last decade, demand of banking services switched to ATM’s. Figure 1.1 and Figure 1.2 give total numbers of ATMs and banking offices in Turkey between 2010 and 2020.

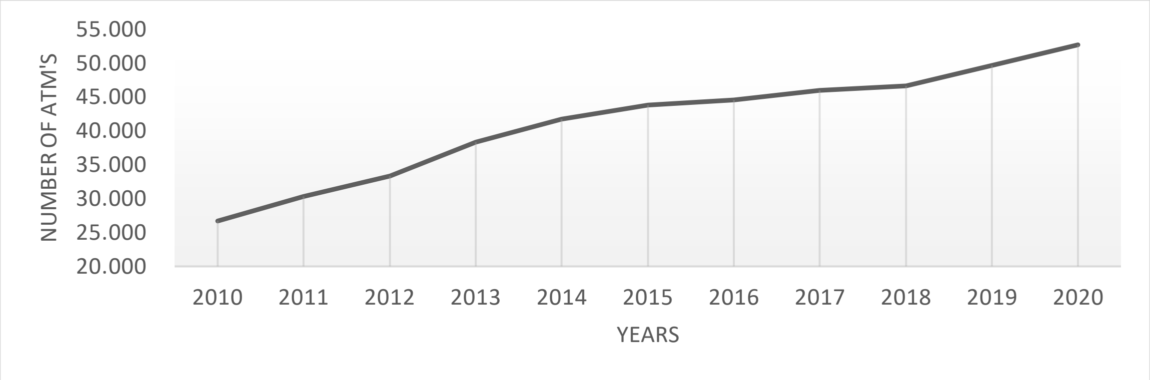


Figure 1.1: Number of ATM's in Turkey

Especially in the last 5 years, number of banking offices have followed an increasing trend. It can be deduced from these graphs that need for banking offices decreased while usage of ATMs increased.

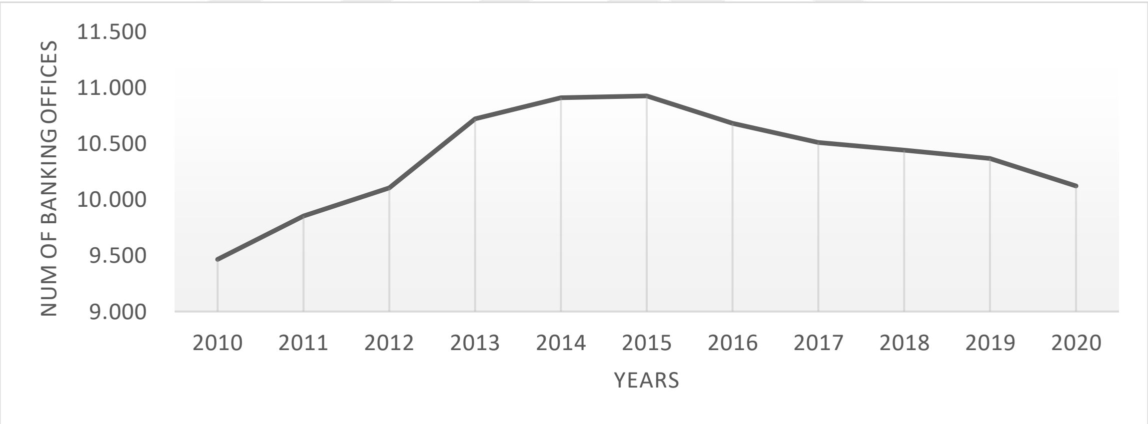


Figure 1.2: Number of Banking Offices in Turkey

Ever since ATMs were first introduced in 1971, they have been a cost reduction method in banking services (Santos et al., 1995). Self-operating system of ATMs eliminates the need of human resource and human-driven mistakes. However, the money stock and the cost of holding money which was not a critical problem with a small number of branches, became a more important issue for banks. Therefore, along with the number and the location of the ATMs, also the amount of money stock in ATMs became an important decision variable. Optimization and demand prediction of these variables have come to the forefront among popular topics in academic research.

Before ATM cash management optimization studies, banks generally used their own business policies and rules came from experiences. However, today, this technique may not be useful in different location or different situations. Nowadays banks that use high technology in operations, prefer scientific methods for cash management by using inhouse or outsource tools.

The main concern in ATM cash management process is to increase customer satisfaction while minimizing cost. According to Dilijonas ve Sakalauskas (2011), ATM service level can be evaluated with some main criteria. These are:

- Customer friendliness,
- Functionality of ATM,
- Proper placement and distribution of ATMs,
- Environmental safety.

Cash management refers to a wide range of finance, including cash collection, handling and usage (Boeschoten, 1998). After collect money from customers with loan repayments and investment accounts, to increase the efficiency, it is important where to use it and how to distribute it.

Demand prediction problem, stock problem and routing problem are parts of cash management problem. Even if these are individual problems in themselves, at the same time these problems are interrelated. As the first step of ATM cash management, cash demand should be determined. Cash transactions of an ATM have trends in certain cycles. These cycles can be weekly, monthly or yearly. Effects of important days like holidays and feasts, pay days, day of week, location of ATM and seasonal fluctuations should be considered in this process. According to calculated cash demand of the ATMs and main costs like delivery cost, holding cost and opportunity cost, optimum amount and replenishment frequency is determined.

Forecasting studies in service industries are ignored areas rather than manufacturing industries (Baltacioglu, 2007). Although the main objectives of these two kinds of industries are similar, like revenue maximization or resource minimization, their environment and the methods widely vary between them.

While inventory cost is one of the main concerns of product-based companies, service-based companies does not have this concern. Essential affairs of a service-based company are quality and availability of the service.

There are fundamental differences between product based and service-based organizations which cause the forecasting methods used for these kinds of organizations differ from each other (Frohlich et al., 2002). Intangibility of service-based products is one of these differences. While this feature gives an advantage to service-based companies by eliminating constraints of logistics processes, decreases measurability of amount of demand of the service. Another difference is need for simultaneity. Demand of a product may be deferrable a fair amount of time, but demand of a service is lost in a case of inability to meet demand. Heterogeneity is one another difference between service based and product-based companies. Services cannot be easily standardized. Every customer experiences a different service each time and quality of a service cannot be measured easily.

Eventually, because of characteristic fluctuations in the demand for services and the fact that services are produced and consumed simultaneously, service-based businesses face the problem of matching capacity and demand continuously (Duk et al., 2002).

Nevertheless, ATM cash management process is a kind of service industry supply chain management problem, it differs in some features of product like tangibility and finiteness of money. Like inventory cost of production industry, holding cost of money is an important parameter addition to opportunity cost of holding money in ATM.

2. DEMAND PREDICTION IN BANKING SECTOR

In this section related literature is reviewed in three steps. Firstly, cash management concept which is the motivation behind the demand prediction is explained in banking perspective. Then demand prediction in manufacturing and service industries are compared and as a combination of them, demand prediction in banking sector is examined.

2.1 Cash Management Concept in Banking Sector

Cash Management is the financial discipline devoted to the management of planned expenditures; therefore, it focuses on operational efficiency and process optimization (Wuttke et al., 2013). However, in banking sector, this term usually refers to a particular business line which is named as transaction services, that sells products and services to assist the bank's cash management activities. Cash management activities can be considered for different operation fields of banking, such as branches, internet banking and ATMs.

Whichever type of operation field it is, cost-effectiveness, risk reduction and efficiency are most important objectives in cash management (Mols et al., 1997). To consider cost effectiveness; holding costs, transportation costs and out of stock costs should be handled together (Castro, 2009). Although these problems are interdependent, because each of these problems is too complex in itself, stocking problem, delivery and routing problem and demand prediction problem are examined individually.

2.2 Demand Prediction in Manufacturing Industries and Service Industries

Demand prediction is an important concept for both manufacturing and service industries. As in the continuation of the business, result of demand prediction is useful for investing a new business.

The term of demand prediction is often used interchangeably with demand planning specially in manufacturing industry. However, demand planning is a wider concept and involves demand planning but not limited with it. In manufacturing, demand planning serves as a starting point for many other processes such as production, inventory holding and delivery. So, the key to increasing efficiency in the entire supply chain is to approach reality as much as possible. The right method to use depends on business type, available resources and main objectives.

According to Armstrong (2001), not all the techniques are applicable for each product type. The correct application of the right technique with the right parts is crucial for accurate demand prediction. For example, in food industry, because of the seasonality and perishability factors are limitations to storage process, choosing an accurate forecast method is vital. Barbosa et al. (2015) claim that time series analysis is meaningful for food and beverage producers because one of the biggest challenges is to coordinate production and stock to minimize product loss that is caused by of short durability. Production strategy is another criterion to decide the demand prediction method. In a make-to-order environment, manufacturers often use model option logic in production planning which consists of producing specific numbers of each model and option of the product according to orders, while those in make-to-stock environment should rely heavily on forecasts.

No matter which business type is it, there are key points to make right prediction. According to Whitfield et al. (2013), these key points are identifying the various types of statistical data available, deciding on the specific degree of the accuracy required and considering horizon of the forecast. Helms (2000) emphasize that whichever technique is selected to forecast, time horizon, technological change, barriers to entry, dissemination of information and elasticity of demand are key principles that need to be considered for accuracy of forecast. Generally, for mid/long-term planning, established products and stable demand traditional methods can be best fit while machine learning methods can give better results for short/mid-term planning, new products and volatile demand scenarios.

Contrary to the fact that there is a lot of study about demand prediction in manufacturing industries for decades, there is not much in the service industries. Although the objective of market research for service-based and product-based companies is similar, which is maximizing return with limited resources; there are also significant differences. The main difference is cost of inventory. Product based companies should estimate the demand to minimize inventory costs while ensuring their customers access to their products. On the other hand, inventory is not as important in service-based companies as trained workers. Service-based companies must focus more on workforce qualification (Lay et al., 2009). They must invest strongly on the training and certification of their employees. Employing and training qualified staff is both expensive and time consuming; this requires service-based companies to plan further than product-based companies.

Service-based companies generally have a wider range of variables on which demand is based. While a product-based company target its research on a particular type of product, service-based companies must consider whole industry and use wider range of data (Lee et al., 2014). So, while for product-based companies traditional prediction methods can be more useful, they may not be sufficient for service-based companies. When it comes to the service industries, there are multiple variables to be considered such as workforce efficiency, training requirements and keeping the long-term goals in mind. These evidences reveal the fact that service-based companies should choose more advanced methods than traditional prediction methods.

Another human factor in service industry except than employee is, customer. Unlike the product customer, behavior of the service customer is more ambiguous and varying by time. Koshiba et al, (2013) conduct a user-centered study which claims that required service is extremely reliant on factors such as customer interest and environment in service organizations like restaurants and retailers. For example, if customers in a retail shop have time and need advisory, employees should take care of the customers closely. Contrarily, if customers do not need help, caring of staff can make a negative impression. So, right decision can be change according to situation and while making a service-based market research customer behavior should be studied.

2.3 Demand Prediction in Banking Sector

For cash management problem in banking sector, first, Baumol (1952) developed a model called as “an Inventory Theoretic Approach” and defined the problem as an inventory problem. In his model, unused cash money is compared to idle raw materials in stocks and Economic Order Quantity model is used to determine cash management strategy.

Johnson and Montgomery (1976) associated cash management problem to single period stochastic inventory problem.

Constantinides and Richard (1978) modeled the problem with Wiener Process which is a continuous time stochastic process.

Boeschoten (1992) studied on parameters which affects cash demand and classify cash demand for ATMs according to time, environment and customer features.

Simutis (2008) proposed an approach which focuses on customer behavior and aimed to predict customer behavior by using Artificial Neural Networks. Simutis used Flexible Artificial Neural Networks Method and Support Vector Regression Algorithms to predict daily demand of ATMs.

In ATM cash replenishment, banks want to use less resources, like cash kept in ATMs, trucks for transport money, while meeting customer demand. According to Simutis, generally banks keep %40 more money than they need while %15 is efficient as safety stock (2008). As well as surplus is an undesirable and costly situation (Rajwani et al., 2017), stock-out also an unwelcomed case and generally its cost is an inestimable cost (Paul, 2010).

In recent studies in ATM cash demand prediction main methods to forecast the demand can be grouped in three: statistical methods, simulation techniques and artificial intelligence techniques. In Figure 2.1, an overview of forecasting methods is presented.

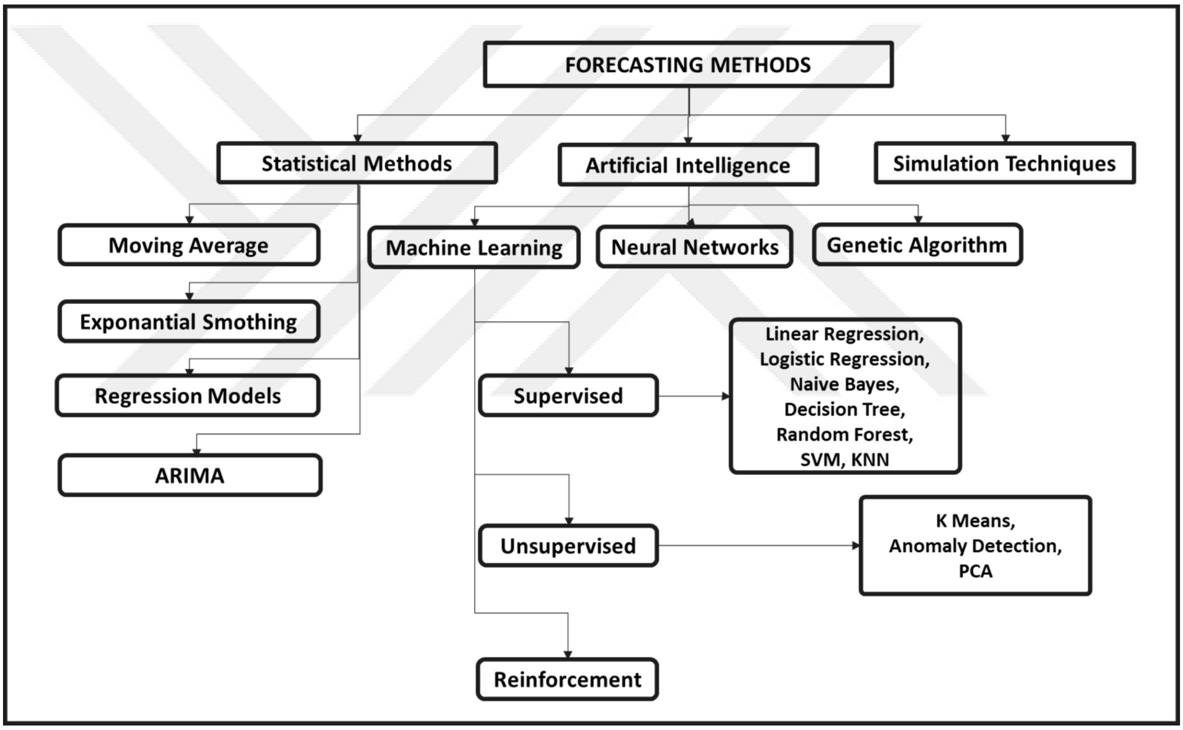


Figure 2.1: Forecasting Methods Overview

Statistical methods are widely used and traditional forecasting methods in ATM cash demand prediction. (Laitinen et al., 1998), (Petrovic et al., 2006), (Abraham et al., 2009), (Gubar et al., 2011), (Huynh et al., 2014). Exponential smoothing and weighted moving average methods are applied to daily cash withdraws for individual ATMs (Cui et al., 2014). These methods use only cash withdrawal data and try to catch a trend or seasonality in the historic data (Brentnall et al., 2008). Another statistical model of predicting ATM cash demand is ARIMA (Simutis et al., 2007). ARIMA models are the most generic and basic time series forecasting models. The fact that these models are easy to implement and can be solved with shorter run times, increases their preferability. One disadvantage of ARIMA models is that these models predict future demand by using only demand data and does not use features to correlate sudden changes and fluctuations with them (Liu et al., 2013).

Simulation is another method used in cash demand prediction. Even if they are effective to model a problem, simulation methods are rarely used techniques because of long run times and modelling difficulty in complex systems. (Sepúlveda et al., 2004), (Scholtz et al., 2012), (Stahl, 2019).

Although the methods described so far are useful methods, they are insufficient at some point. They only use historical data of subject that is forecasted. For example, if weather condition is forecasted, the only input is daily weather condition data for a certain time. With this way, the other factors which affect weather are neglected like atmospheric pressure, emission level of harmful gases.

Artificial intelligence is more advanced technique to make time series forecasting by ability to use many variables related to variable that is forecasted. As a part of artificial intelligence, machine learning methods are commonly used in recent times with their easiness in implementation and precision in results. (Atsalaki et al., 2011), (Cardona et al., 2012), (Venkatesh et al., 2014), (Arabani et al., 2019), (Leonov et al., 2020). This ability increases the precision of the models (Kotsiantis et al., 2006). To use exploratory data analysis is another distinctive feature of machine learning. It is a critical process of investigating data with help of statistics and visualization tools (Hartwig et al., 2002).

Another part of the artificial intelligence is neural networks. Concept of neural networks is an ever-growing concept for years. As examples of recurrent neural networks, long short term memory networks (Wei et al., 2017), (Hua et al., 2019) and time delay neural networks (Jha et al., 2014), (Sitte et al., 2000) are good performance algorithms for time series data. These algorithms do not need data handling processes unlike machine learning algorithms. However neural networks need big datasets to give a good performance.

So far in the study, it is seen that machine learning methods are very effective and accurate methods in ATM cash demand prediction. So, the rest of the study focuses on the machine learning algorithms.

In the next section, machine learning algorithms are explained in detail and use cases are given. Then, methodology of the study is introduced, and flow chart of the method is provided. After that, application of the method and results are presented. Finally, in the conclusion section, an overall brief of the study is given, and further works are discussed.

3. MACHINE LEARNING ALGORITHMS

Machine learning is a subset of artificial intelligence, whose goal is to make machines learn from past data without programming explicitly and obtain the accurate results (Michalski et al., 1983). Artificial intelligence and machine learning are two concepts that are often confused. Artificial intelligence is the theory of simulating human intelligence processes by machines, especially by computer systems. Although the name “artificial intelligence” was first used in 1950s, simple decision trees can be shown as an early examples of human-like decision system studies (Nilsson, 2010). On the other hand, machine learning is a specific subset of an artificial intelligence systems, which is more related with statistics. Despite their similarity, there are some differences that distinguish machine learning from statistics.

The most important feature that distinguishes machine learning from statistics is the purpose of their usage. While machine learning methods are trained to obtain predictions as accurate as possible, statistical methods are used for revealing the relationships between the variables (Paluszek et al., 2016). So, while a machine learning model is trained with much more data than the statistics model and used to predict future data by using historical data, a statistical model is used to detect relationship between variables and expecting a future prediction remains just an assumption.

Besides, relevance between increase in the frequency of usage of machine learning algorithms and the development computer processor technology is a proof of this situation (Arel et. Al., 2010). The huge amount of data obtained from the digital systems of today and the increasing complexity of the problems made impossible to establish explicitly programmed statistical models for each problem and lead professionals to use machine learning algorithms.

Another advantage of machine learning is effectiveness in time series prediction due to easiness in introducing time effect to model (Längkvist et al., 2014). Time series data have complicating agents like seasonality and machine learning algorithms are strong tools to detect and measure them.

Machine Learning concept was born in 1940s based on a model of brain cell interaction. As the first case, neurophysiologist Warren McCulloch and mathematician Walter Pitts wrote a paper about neurons and their working principles (1947). They decided to create a model of this. At the same years, Hebb presented his theories on neuron excitement and communication between neurons in a book titled “The Organization of Behavior: A Neuropsychological Theory” (1949).

In 1950, Alan Turing introduced Turing Test which is a method to test a computer whether it is able to convince a human that it is a human, not a computer.

Arthur Samuel developed a computer program for playing checkers (1952). This program could learn and decide its next move by using the positions of the pieces on the board.

Frank Rosenblatt designed the first artificial neural network, called “Perceptron” for pattern and shape recognition (1958). Perceptron can be called as ancestor of machine learning algorithms.

3.1 Data Requirements and Data Handling

Machine learning algorithms consume and process large amounts of data to learn complex patterns. However more data is not always better data. High accuracy for prediction is not possible without the right data that is captured and processed the right way in machine learning (Domingos, 2012). Exploratory Data Analysis and Feature Selection are powerful data handling methods which are widely used in machine learning applications (Friedman, 1974).

Exploratory Data Analysis (EDA) is an approach for summarizing, visualizing, and becoming intimately familiar with the important characteristics of a dataset (Behrens, 1997). It helps model to discover patterns, to spot anomalies, to test hypothesis and to check assumptions. Even though it is a time-consuming process, skipping EDA can cause generating inaccurate models, generating accurate models on the wrong data or choosing the wrong variables for the model (Martinez et al., 2010). In order to understand data better and make it easier to apply the model, the first steps are to detect missing data points and outliers. After that, balance of data is investigated by using graphical tools and wrong and misleading parts of data is eliminated. Most used graphical tools used in EDA are correlation matrix, histogram, pareto chart and scatter plot (Behrens, 2012). Detailed application of EDA is presented in Methodology section.

Feature engineering involves the careful selection and manipulation of dataset features. The target is to use most necessary part of input to make accurate predictions and does not have to deal with any extra noise that comes from the rest of the data (Xue et al., 2015). By applying principal component analysis (PCA), features have extremely low correlation with the output can be found and extracted from the dataset (Abdi et al. 2010). Some features can be redundant, such as the ID or recording date and do not contribute to the prediction at all.

3.2 Machine Learning Forecasting Use Cases

3.2.1 New Product Introduction

New product introduction is an area where machine learning is extraordinarily strong. Traditional forecasting needs two to five years of sales data to guarantee an acceptable level of accuracy (Ramanathan, 2012). With new items, although there is no sales history, predicting demand cannot be neglected as it drives multiple important processes, from procurement to logistics management. Apart from market research and collecting expert opinions, the most common approach to forecasting launches is to identify clusters of predecessors with similar properties and product life cycle curves (Christensen et al., 2017). Machine learning algorithms can be employed to extract specific patterns from huge volumes of unstructured data, find similarities and develop predictions.

3.2.2 Marketing Campaigns

Forecasting is often used to adjust ads and marketing campaigns and can affect the number of sales. Advanced machine learning prediction models can also take into account marketing data. Companies run thousands of consumer promotions that are supposed to drive sales. According to a report of Gartner Research Company, 59% of trade promotions don't break even, or, in other words, result in additional expense not profit and these companies that still use spreadsheets to plan promotions and forecast their impact (2017). Obviously, without improved technologies, companies can hardly generate reliable predictions for costly marketing campaigns. Outcomes of promotions depend on numerous factors with complex relationships, hidden in large batches of raw data. Machine learning can cope with this challenging task.

3.2.3 Predictive Sales Analytics

One of the most common applications of machine learning is demand forecasting solution which is being able to model the future. Predictive analytics allows the organization to estimate the demand for their products or services, and also understand what will drive the sales and how consumer behavior is likely to change under certain conditions like trend alerts or seasonal changes (Koch, 2017). A demand forecasting solution using machine learning aggregates historical and new data from different sources. This includes data from ERP (enterprise resource planning system), CRM (customer relationship management system), POSs (point of sales), customer demand studies, marketing surveys, social media engagement and more.

3.2.4 Demand Sensing

Demand sensing is another machine learning application that focuses on capturing real-time fluctuations in market demand and consumer purchasing behavior (Folinas et al., 2012). With these demand planning solutions, organizations can better optimize their forecasts to match anticipated needs on the go. This is a critical demand forecasting method in industries that have a fast-changing market. In demand sensing, the data is aggregated from point-of-sale systems and warehouses and is backed by marketing data that highlights customer engagement, trending content and more. Basically, a demand sensing solution captures demand trends on a short-term basis and the factors that cause fluctuations in it, enabling the organization to build a data-driven supply chain.

3.3 Types of Machine Learning Algorithms

Types of machine learning algorithms are basically divided into three main branches according to the task or problem as supervised, unsupervised and reinforcement algorithms.

3.3.1 Supervised Learning

As the name suggests, the learning method that is designed for dataset with target values are called as supervised learning (Caruana et al., 2006). In this type of machine learning models, the training is made on the basis of labels of the target values. Supervised learning problems can be grouped in two sections as “classification” and “regression”. When the target variable is a category, like “Ok” or “Not Ok”, “Positive” or “Negative”, the supervised learning problem is a classification problem. If the target variable is a dynamic value like the currency or amount of item sold, the problem is a regression problem. In Figure 3.1, graphical representation of classification is given. A and B classes split according to distances between samples.

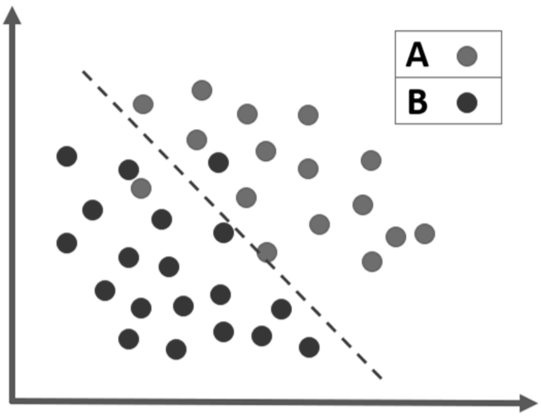


Figure 3.1: Supervised Learning

3.3.2 Unsupervised Learning

Unsupervised learning problems are the problems those data do not include target values (Ghahramani, 2003). In this kind of problems, dataset only includes feature variables and the unsupervised learning problem needs a model that will expose the structure or distribution in the data.

The most common unsupervised learning problems are “clustering” problems. Clustering problem is kind of a problem that, its solutions should include groupings in the data such as purchasing behavior of customers. Figure 3.2 represents clustering in graph.

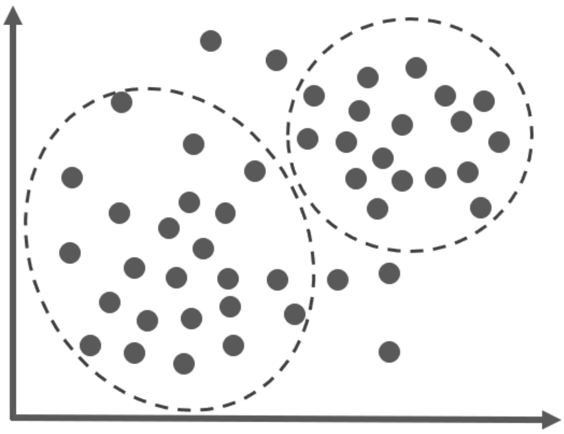


Figure 3.2: Unsupervised Learning - Clustering

Dimensionality reduction and anomaly detection are other unsupervised learning problems.

Dimensionality reduction is the process of reducing the number of random variables under consideration by obtaining a set of principal variables (Kumar et al., 2017). Dimensionality reduction techniques are used for feature elimination process.

Anomaly detection techniques detect outliers in the unlabeled data under an assumption that most of the data examples are normal by observing the instances that fit the remainder of the dataset (Eskinet al., 2002).

3.3.3 Reinforcement Learning

Before the explanation of reinforcement learning, it will be useful to explain what software agent is. Software agent is the section of an algorithm that does the given task within defined rules and borders. Reinforcement learning arises from the idea that a software agent can be conditioned to a particular movement or sequence of motion by giving a penalty or reward at the end of each movement (Kaelbling et al., 1996). Initially a random utility value is assigned to every move that can be made in any situation.

Starting from a random starting state, the software agent starts moving. In a situation, when a move is made, if the award or penalty is received, the utility value of that move is increased or decreased according to a formula. Figure 3.3 gives a representation of reinforcement learning principles.

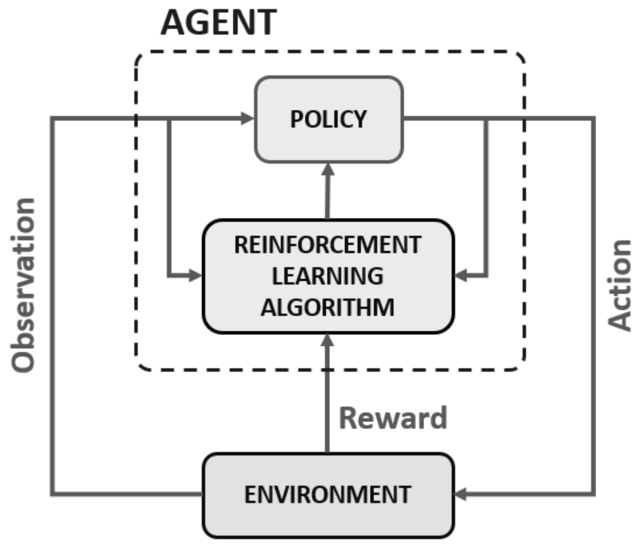


Figure 3.3: Reinforcement learning principles

3.4 Machine Learning Models

In this section, most commonly used machine learning models will be explained. Because of their easiness to use, suitability of different datasets and speed in run times they are the most frequent models in literature (Matsunaga et al., 2010).

3.4.1 Linear Regression

Linear regression is a linear model that assumes a linear relationship, which can be called $y = A + B*x$, between the input variable (x) and the target variable (y) (Seber et al., 2012). In this method, initial A and B values are assigned to model and the feedback is examined in terms of R, RMSE (Root Mean Squared Error), R2 etc. According to the error of the model, A and B values are being updated and the best accuracy is being tried to get in optimum iterations. Figure 3.2 represents linear regression in graph.

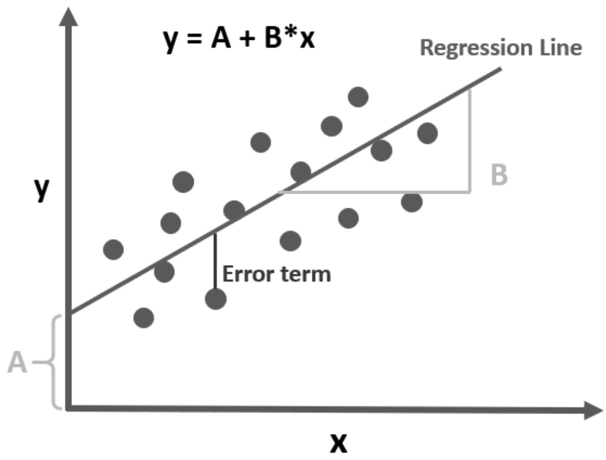


Figure 3.4: Linear Regression

3.4.2 Logistic Regression

Logistic regression is the method which is suitable for making predictions when the target variable is divided in two categories (Mohammed et al., 2016). Instead of fitting a line, logistic regression fits an “S” shaped logistic function to the data which goes from 0 to 1. The data points are evaluated depending on which side they are closer, and the predictions are made within this perspective. In Figure 3.5, graphical representation of logistic regression is given.

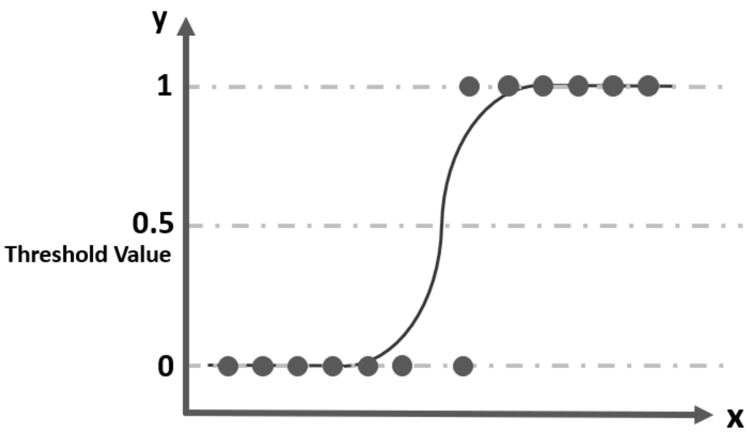


Figure 3.5: Logistic Regression

3.4.3 Decision Tree

A decision tree uses a tree-like graph in order to represent all possible solutions graphically (Freund et al., 1999). Decision trees are frequently used not only in machine learning applications, but also operations research and decision analysis. In this method, the initial rules form the leaves of the tree and by following rules respectively the decision is made. After getting feedback from the current error of decisions, iterations are done and the updates on the rules are made. In Figure 3.6, representation of a sample decision tree is given.

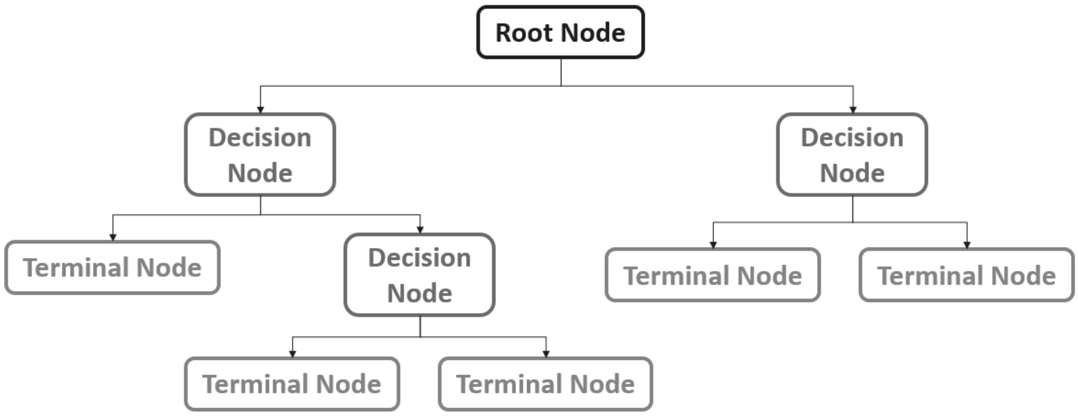


Figure 3.6: Decision Tree

3.4.4 K Nearest Neighbors Algorithm

K Nearest Neighbors algorithm, as known as KNN, is a type of supervised machine learning algorithms which can be used in both regression and classification problems (Kuang et al., 2009). In KNN method, at first, the value k is assigned as an integer. This k-value is used to determine the class of the unknown datapoint with checking the k nearest neighbors' values. The unknown data point is predicted as the most commonly seen target value class among k neighbors. Since the value of k can change the prediction, it must be fine-tuned on the test data. Figure 3.7 represents k nearest neighbors algorithm graphically with different k-values.

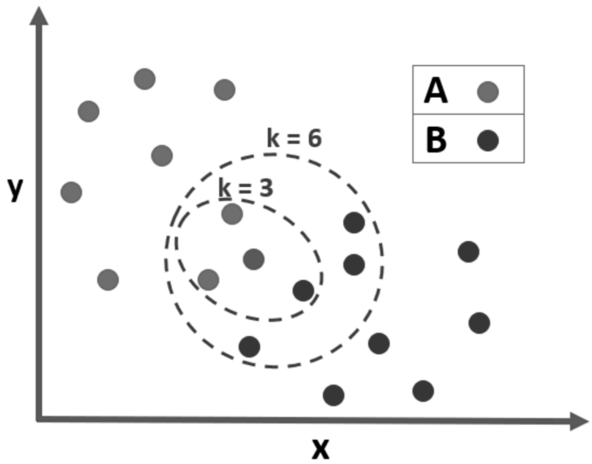


Figure 3.7: K Nearest Neighbors Algorithm

3.4.5 Support Vector Machine Algorithm

The aim of Support Vector Machine algorithm, as known as SVM, is to find a plane in N-dimensional space to classify the data point (Kotsiantis, 2007). Support Vector Machine method has applications for both supervised and unsupervised learning type problems. To make it clearer, Figure 3.8 can be examined as an application for 2-dimensional unsupervised learning problem.

As it can be seen in Figure 3.8.a, in the first step of this problem, a random plane, in this situation a line, is assigned on the dataset. This plane divides data points in two categories. This data groups' center of gravity is marked with x for both of them and a new plane is assigned to be equidistant from these two gravity centers, which can be seen in Figure 3.8.b. A new plane will divide the datapoints into two new groups and the process is repeated for this new situation and this leads the datapoints and the plane to the circumstance in Figure 3.8.c. These iterations can be continued until the displacement amount of gravity centers become less than a determined value.

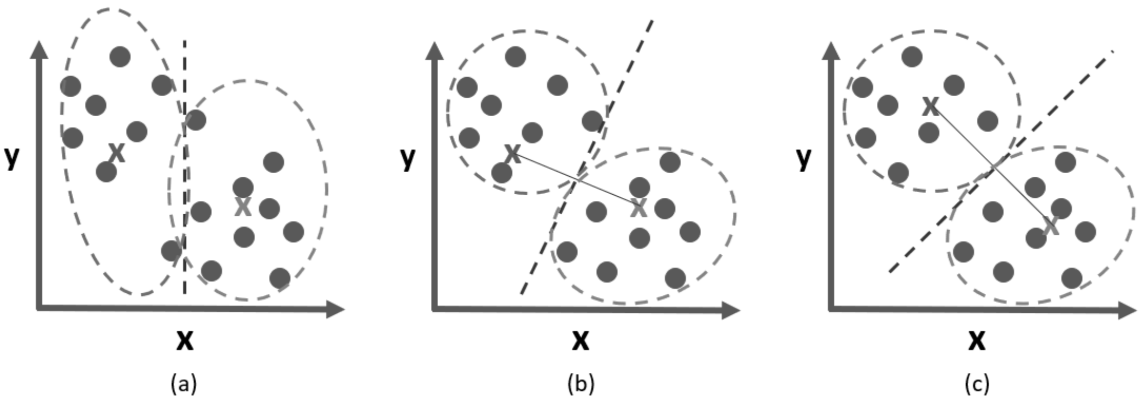


Figure 3.8: Support Vector Machine

In the next section, methodology of this study and dataset which is used in the application are introduced. Among all the mentioned machine learning models and types, suitable ones for dataset are compared and results are shared.

4. METHODOLOGY

Demand for money is the main concern of banking sector. Because of the fact that there is no production process; cost in banking sector comes from holding and transferring money. Both excess and insufficiency in supply of money in ATM's cause the company making loss. While excess of money supply eliminates the other opportunities like investment or giving loan, insufficient money supply leads up to customer satisfaction and distrust.

With the excessive importance of money demand, the value of accurate prediction increases. The problem is money demand includes lots of factors and these factors and their weights can be change with time and location dramatically. So, demand quantity and trend are not same for two different cities or neighborhood even if they have similar population densities. Fundamental factors that quantity and trend of demand for money to an ATM is affected from population density, demographic structure of location, location of ATM and its proximity to other ATMs, special days, holidays, feasts, pay days of government and company employees, surrounding malls and shopping facilities so on. Being affected by that many of factors necessitates using a developed prediction model to investigate different variables and their significance on data. As stated before, rather than statistical methods or other traditional techniques, machine learning works better in complicated and multivariate data.

In this study, machine learning algorithms are used in order to predict cash demand of an ATM. First of all, machine learning algorithms are run with raw dataset. Then, exploratory data analysis and feature selection are applied to the dataset and the algorithm given the best result with raw data run with this processed dataset. By comparing these two results it is aimed to emphasize the importance of exploratory data analysis and data processing in prediction with machine learning algorithms. Flow chart of the methodology can be seen in Figure 4.1.

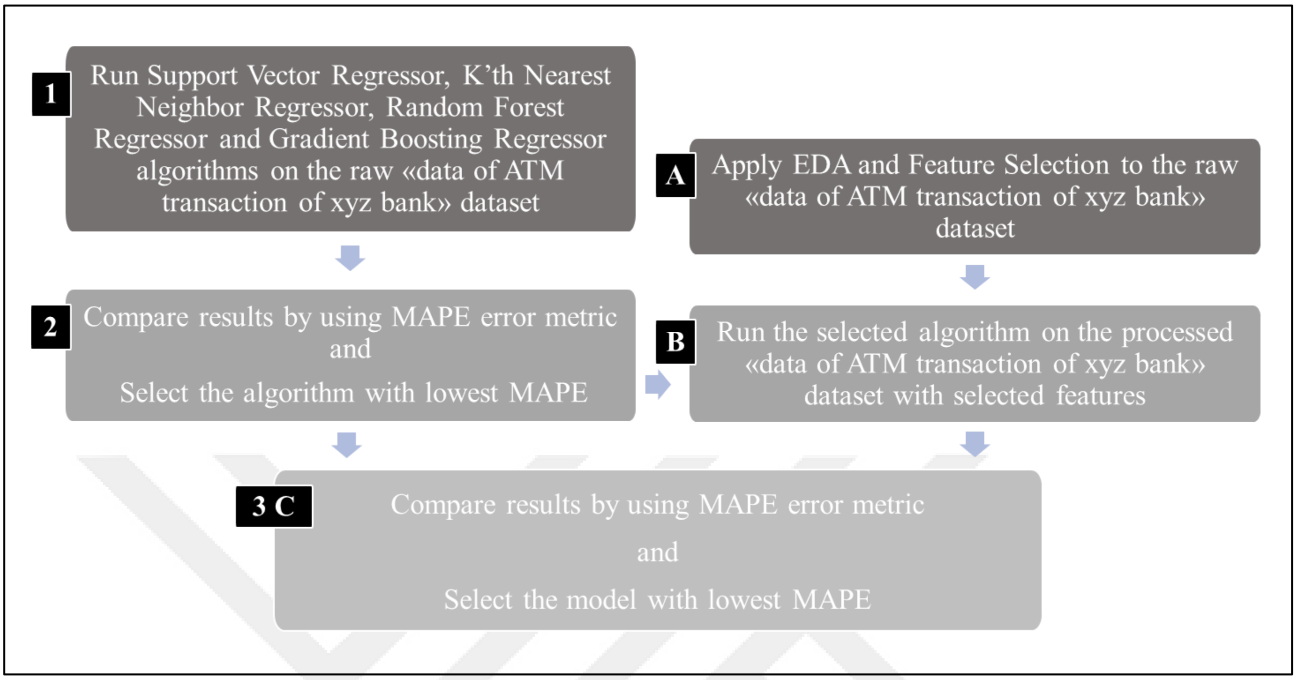


Figure 4.1: Flow Chart of Methodology

4.1 Algorithms and Error Metric

In the previous sections, it is mentioned that all type of machine learning models gives better accuracies in different data structures and problems. In time series prediction the most accurate and widely used machine learning algorithms are Support Vector Regressor, K'th Nearest Neighbor Regressor, Random Forest Regressor and Gradient Boosting Regressor (Sapankevych et al., 2009), (Ak et al., 2015), (Weigend, 2018).

To measure the accuracy, choosing the right error metric is extremely important (Davydenko, 2013). In time series prediction, Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) are the most common encountered error metrics (Saigal et al., 2012), (Franses, 2016), (Khair et al. ,2017). R-squared is also a widely used error metric in machine learning.

In MAE metric shown below, large errors only have a proportional effect on the overall metric:

$$MAE = \frac{1}{N} * \sum_{i=1}^{N} |x_i - \tilde{x}_i| \tag{1}$$

MAPE measures the size of the error in percentage. It is calculated as the average of the unsigned percentage error, as shown below:

$$MAPE = \frac{100}{N} * \sum_{i=1}^{N} \left| \frac{x_i - \tilde{x}_i}{x_i} \right| \tag{2}$$

Where, x_i is the actual observation and $\tilde{x}_i$ is the estimated value, with a N sample.

R-squared, which is also called as coefficient of determination is the proportion of the variance in the dependent variable that is predictable from the independent variable(s).

The most general definition of the coefficient of determination is given below:

$$R^2 = 1 - \frac{Unexplained\ Variation}{Total\ Variation} \tag{3}$$

R-squared is always between 0 and 1. 0 indicates that the model explains none of the variability of the response data around its mean.1 indicates that the model explains all the variability of the response data around its mean.

In this study, MAPE metric is used in order to avoid misguidance of magnitudes of the results. MAE and R-squared metrics are also considered in order to control accuracy of the model.

4.2 Introduction of Dataset

The data used in this study is taken from the online platform www.kaggle.com. Kaggle is a data science platform which data scientists and data analysts can share datasets. By using these datasets, users can open notebooks, analyze data, create prediction models, share them and compare with another's work. This dataset is a simulated data with name "data of ATM transaction of xyz bank", consisted of 11590 rows. Dataset is uploaded to platform in January 2019. Because of there is no notebook for this dataset, a comparison with another work cannot be done. A sample of application data is given in the Table 4.1.

Table 4.1: The Sample Dataset

	ATM name	# of withdrawals	Total Amount	Weekday	Festival Religion	Working Day
1	Big Street	50	123800	Saturday	H	H
2	Mount Road	253	767900	Saturday	C	H
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
11589	KK Nagar	76	76	Friday	H	H
11590	Christ College	143	143	Friday	H	H

The dataset consists of cash withdrawal data of 6 ATMs. In this study, cash withdrawal data of the ATM named Airport is used. The prediction algorithm is trained with 2 years of data and 1 year of data is predicted.

4.3 Exploratory Data Analysis and Feature Selection

After choosing the right algorithm and error metric, exploratory data analysis is applied to dataset to investigate data and prepare it to prediction.

First of all, data types of dataset's columns are checked. Prediction models does not work with string and categorical data types directly. They should be mapped to integer values (Myatt, 2007). In order not to misleading the model, these integer values should not be large numbers. Table 4.2 gives the columns and their data types of the dataset. Weekday, Festival Religion and Working Day columns should be converted to numeric values.

Table 4.2: Columns of Dataset

Column Name	Type
Transaction Date	datetime
No Of Withdrawals	integer
No Of XYZ Card Withdrawals	integer
No Of Other Card Withdrawals	integer
Total amount Withdrawn	integer
Amount withdrawn XYZ Card	integer
Amount withdrawn Other Card	integer
Weekday	category
Festival Religion	category
Working Day	category

In the other step of exploratory data analysis, statistical descriptions and distributions of dataset's columns should be analyzed. Statistical analysis is used to detect the extreme points and examine consistency of data (Van den Broeck et al., 2005). Thus, categorical data types do not require this process. Consistency of a data can be deduced by examining whether the normal distribution can be fit to data (Chatfieldet al., 2019). If data has a normal distribution, prediction methods are more powerful, and results are well understood. Statistical description consists of mean, standard deviation, minimum value, maximum value, lower, middle and upper quartiles of data. In a consistent data, which fits to the normal distribution, mean and middle quartile values should be close to each other. Also, the difference between mean and maximum / minimum values should not be extremely higher than three times standard deviation. This is called as three-sigma method. According to three sigma method, an event is considered to be practically impossible if it lies outside of three standard deviations from the mean (Pukelsheim, 1994). With the three-sigma method, outliers should be extracted from the data for the accuracy of model results.

Feature scaling is an important process to prepare data to model. In multivariate data, features can have varying degrees of magnitudes. Machine learning algorithms, especially distance-based ones like KNN, K-means and SVM, are affected by the range of features excessively (Li et al., 2015). Because they use distances between instances to determine their similarity. If feature scaling is not applied higher weightage is given to features with higher magnitudes. Performance of the algorithm is biased towards features with higher magnitude.

For feature scaling, normalization and standardization techniques can be used. Normalization is rescaling values of data in a range between 0 and 1. It gives better results when distribution of the data does not follow a normal distribution. In standardization values are centered around the mean with a unit standard deviation. Standardization can be helpful when the data follows a normal distribution.

Feature selection is a performance enhancing process for machine learning algorithms. Machine learning models are advanced algorithms which can work with multivariate data (Tsai, 2009). However, there can be correlation between some features. This fact can decrease accuracy of the model. Generally, a correlation degree greater than 0.7 is considered a strong correlation and less than 0.4 is considered a weak or no correlation (Jin, 2006). In the dataset, features which have high correlation between each other should be detected and they should not be used together in the model.

4.4 Design of Experiments

After choosing the best algorithm and applying data processing tools to analyze the data and improve prediction results, specific adjustments should be made to optimize the performance of the model. The design of experiments called DOE is a systematic approach to understanding how process parameters affect response variables (Antony, 2014). DOE is a mathematical tool used to define the importance of specific processing and/or product variables, and how to control them to optimize the system performance while maximizing features. DOE uses statistical methodology to analyze data and predict product feature performance under all possible conditions within the limits chosen for experimental design. In addition to understanding how a variable affects product performance, interactions between different process and product variables are also defined.

For each machine learning algorithm there are different parameters to apply DOE. For KNN algorithm, the main parameter is k value. This parameter is explained in Application and Results section in detail and DOE is applied.

In the next section, machine learning algorithms are run with the raw data and results are compared. The best algorithm is run with the processed data and results are compared. With the application of the methodology, the effect of exploratory data analysis in prediction is proven.

5. APPLICATION AND RESULTS

Accuracy of the prediction model is based on to select the right algorithm and to investigate the data and understand it. These two processes are interdependent. There are specific algorithms that work well with specific kind of data structures. Also, each machine learning algorithm needs particular data structures and data preparation processes.

5.1 Application of Algorithm Selection (Step1 & Step2)

As the first step of application, in order to choose the right model, alternative models which is introduced in Methodology section are tried with the raw data. According to results which are given in Table 5.1 and Figure 5.1, the most accurate model for this dataset is KNN with 27% MAPE. MAE value of KNN is 92311 which is the lower than the other models. R2 metric shows that all models except SVR can establish strong relationship between independent and dependent variables.

Table 5.1: Prediction Results with Raw Data

KNN	Random Forest	GBR	SVR
MAPE: 27 %	MAPE: 29 %	MAPE: 32 %	MAPE: 28 %
MAE: 92311	MAE: 116572	MAE: 127757	MAE: 107589
R2: 0.549	R2: 0.951	R2: 0.965	R2: -0.22

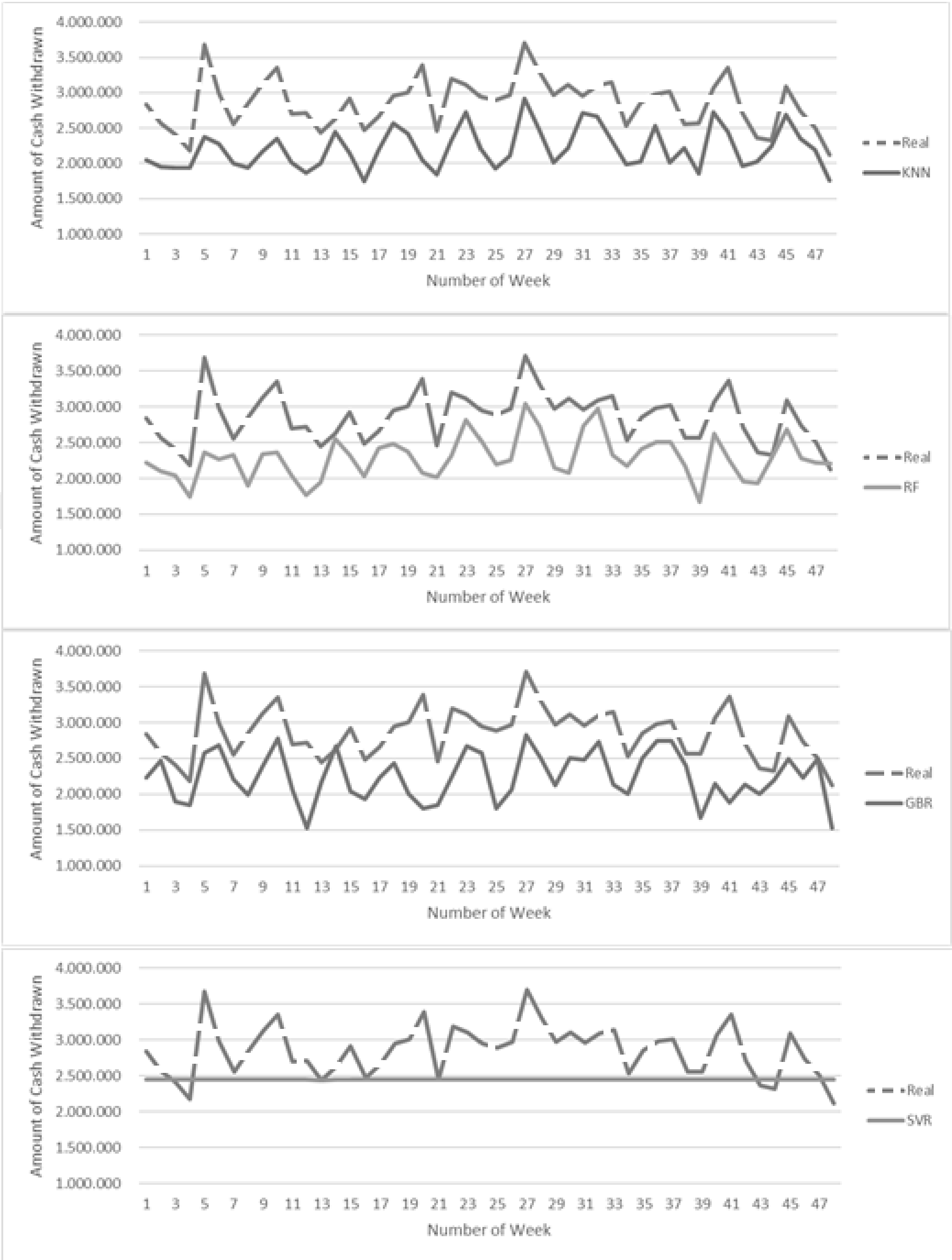


Figure 5.1: Line Charts of Real (Raw Data) vs Predicted Values

KNN firstly was proposed to solve pattern recognition problem as a classification problem (Gweon et al., 2017). Recently, due to its successful and practical implementation, KNN becomes a widely used method in time series prediction (Martínez et al., 2019). As a supervised learning, KNN creates a function from input values of training data according to output values. After learning from training data, desired output values are predicted by using input values of test data.

Classification model of KNN is mentioned in Machine Learning section. The model basically groups data points according to their distance of each other and a determined K value. KNN Regression model works in a similar way. The difference is while in classification variables are categorical, in regression they are continuous (Yin et al., 2016).

K value is number of neighbor data points which is considered together in the function. Because of based of K value the results of model tend to change, to find the optimum value of K is important. As a design of experiments (DOE) application, finding the k value with best results is important in terms of performance of the algorithm. The well-known and widely used technique is “elbow curve” technique (Mehrotra et al., 2006). In this heuristic approach model is run with an extremely low value of k and overfits with a high error rate. As increases as the k value, error rate decreases until a point. In this point error curve reaches a minima. The right value of the k is this value. Figure 5.2 gives the elbow curve of used model and dataset in this application. In Figure 5.3 the differences in results of same model with same data but different K values. In Figure 5.3.a K value is set to 1 and model overfits. When K value is set to 10, prediction is smoothened (Figure 5.3.b).

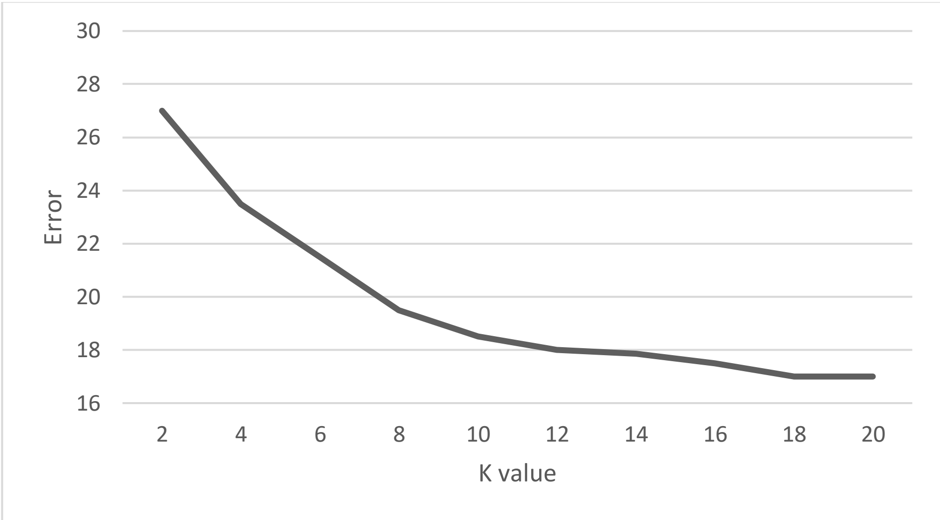


Figure 5.2: KNN Regression errors with different K values in used model

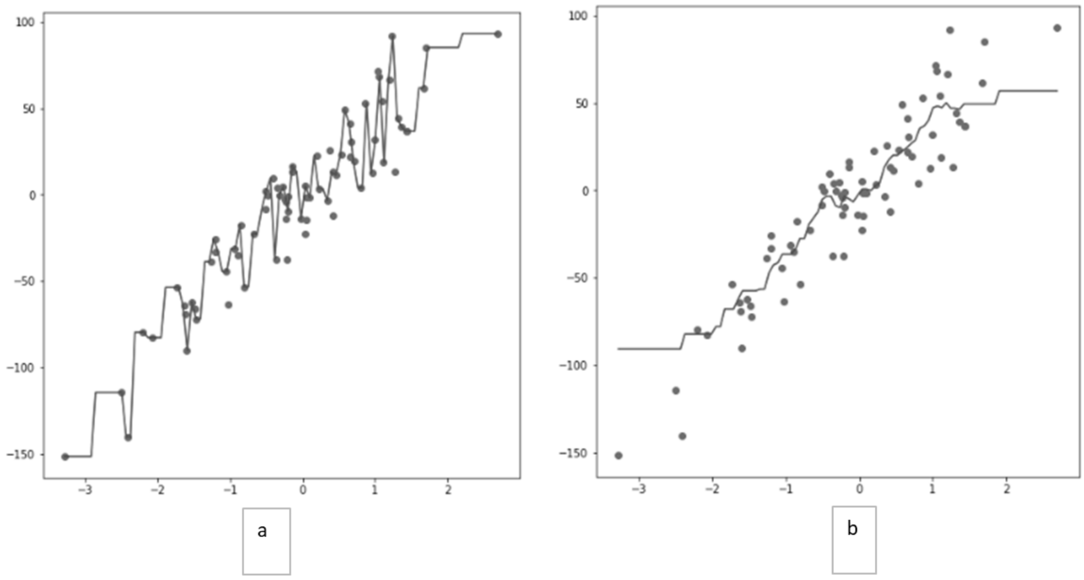


Figure 5.3: Representation of KNN Regression results with different K values

5.2 Application of Exploratory Data Analysis and Feature Selection (StepA & StepB)

After finding the right algorithm as KNN, statistical descriptions and distributions of dataset's columns are examined as the first step of exploratory data analysis as it is mentioned in Methodology section. Table 5.2 gives the statistical descriptions of integer columns of dataset. In the table, means, standard deviations, minimum values, maximum values, lower, middle and upper quartiles of data is given. According to statistics in the table, it can be said that the dataset is close to consistency. However, 'Amount withdrawn other card' column can have a remarkable amount of extreme points. In Figure 5.4, distribution plots of the data support this inference. Although not as much as in this column, it is seen that the other columns have extreme points too. With the three-sigma method, these outliers will be extracted from the data for the accuracy of model results.

Table 5.2: Statistical Descriptions of Columns of Dataset

	No Of Withdrawals	No Of XYZCard Withdrawals	No Of Other Card Withdrawals
count	731.000.000	731.000.000	731.000.000
mean	102.917.921	76.551.300	26.370.725
std	23.449.248	17.435.785	11.672.737
min	26.000.000	17.000.000	0
25%	89.000.000	66.000.000	19.000.000
50%	102.000.000	76.000.000	24.000.000
75%	118.000.000	88.000.000	32.000.000
max	183.000.000	134.000.000	88.000.000
	Total Amount Withdrawn	Amount Withdrawn XYZ Card	Amount Withdrawn Other Card
count	731.000.000	731.000.000	731.000.000
mean	466.283.153.215	383.293.384.405	82.989.768.810
std	142.142.671.132	118.196.409.202	47.835.953.111
min	123.800.000.000	100.200.000.000	0
25%	363.900.000.000	300.250.000.000	47.850.000.000
50%	457.100.000.000	371.400.000.000	75.300.000.000
75%	556.550.000.000	460.600.000.000	109.300.000.000
max	968.700.000.000	719.600.000.000	301.500.000.000

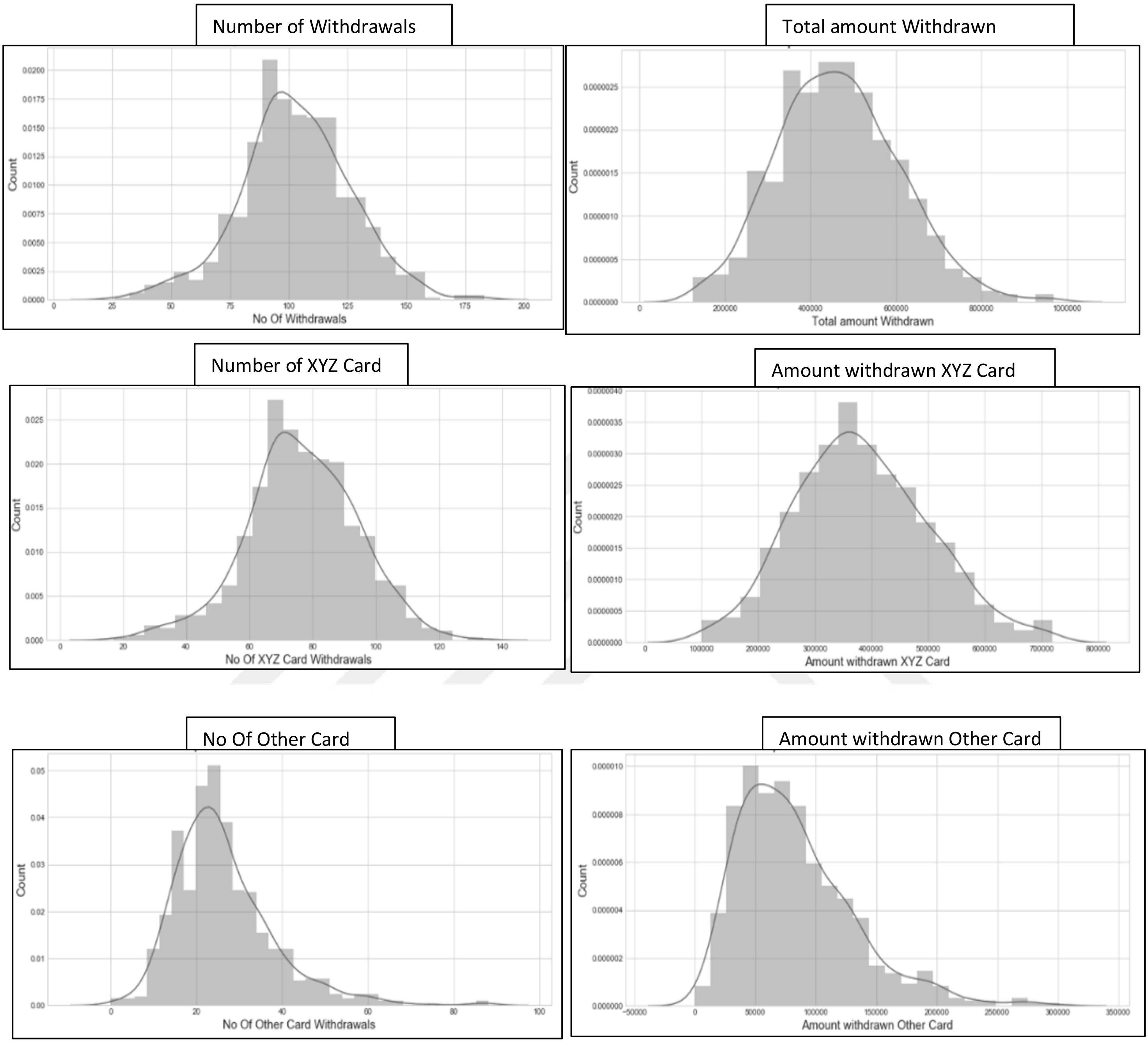


Figure 5.4: Distribution Plots of Data

As it is mentioned in the previous section, to use the right columns of the dataset in the model, it is important to measure correlation between features and not to use correlated features together. In order to detect correlated features, correlation matrix of the dataset is examined. Correlation matrix of columns of the dataset is given in Figure 5.5. Columns

‘No of Withdrawals’, ‘'No of XYZ Card Withdrawals’, ‘No of Other Card Withdrawals’, ‘Total amount Withdrawn’ and ‘Amount withdrawn Other Card’ have strong correlation (higher than 0.7) with each other. Thus, these columns should not be used together in the model. The other columns, “Working Day categorical”, “Festival Religion categorical”, “Weekday categorical”, “year”, “month” and “day”, are suitable to use together with weak correlation values lower than 0.4.

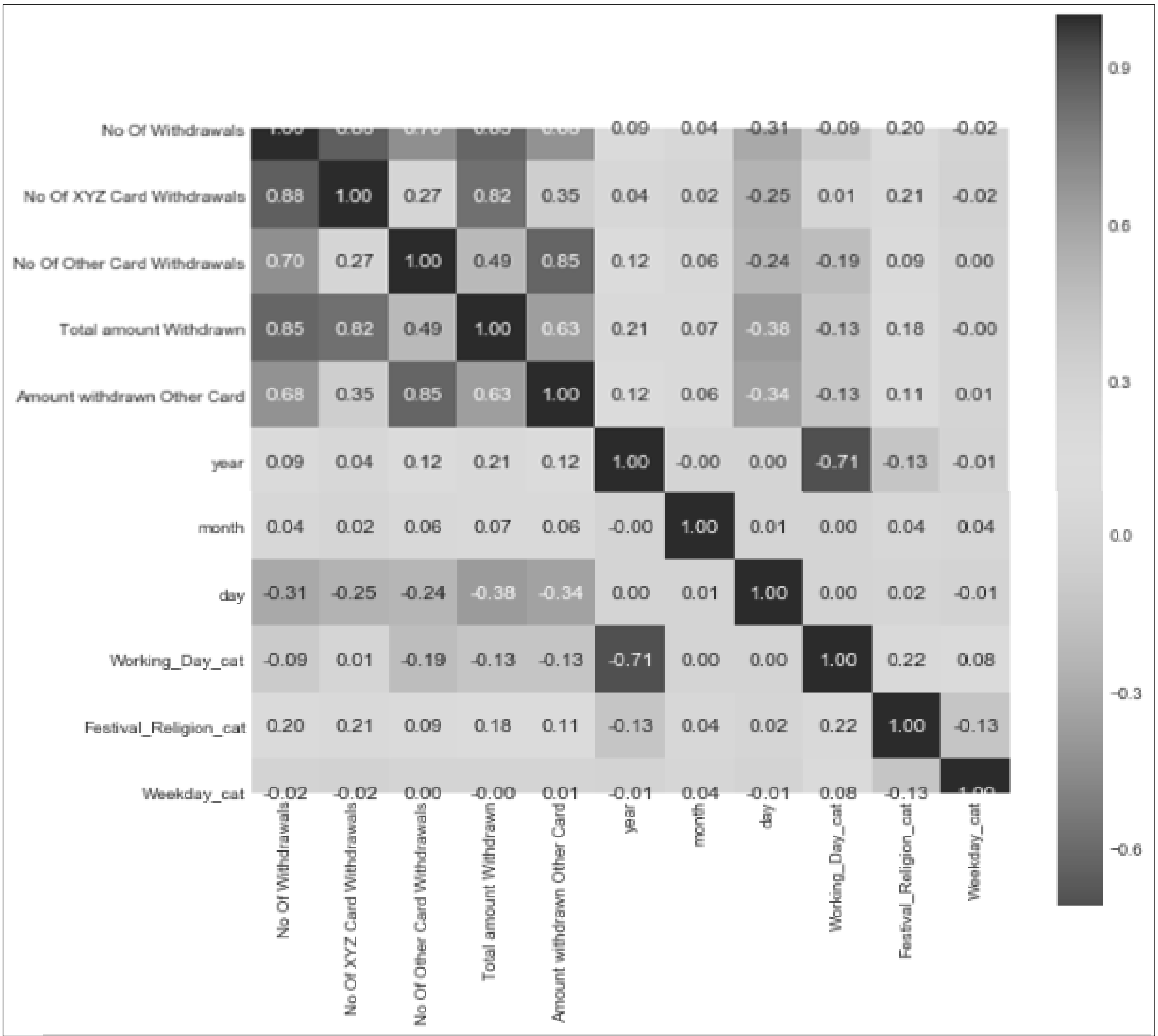


Figure 5.5: Correlation Matrix of Column to Predict

Using the right features is a key factor to make accurate predictions. To predict 'Amount withdrawn XYZ Card' column in the dataset, it is important to catch the trend of the data, seasonality and time dependent rises and falls. Thus, together with features of the dataset, year, month and day columns should be added to model to improve prediction model. These features are extracted from 'Transaction Date' column.

Data handling processes are mentioned previous parts of this section. After data handling, columns of the dataset turned from Table 5.3 to Table 5.4. As the next step of the model, in the feature selection process correlated features are determined. Normally, one of the correlated five columns should be chosen and used in the model. However, because of the fact that these features are unknown information for the future prediction time period, in this study, all of them have been excluded from the model. Finalize version of columns of the dataset is given in the Table 5.5. The dataset is used in the prediction model as such.

Table 5.3: Columns of the Raw Data

Column Name
Transaction Date
No Of Withdrawals
No Of XYZ Card Withdrawals
No Of Other Card Withdrawals
Total amount Withdrawn
Amount withdrawn XYZ Card
Amount withdrawn Other Card
Weekday
Festival Religion
Working Day

Table 5.4: Columns of Processed Data

Column Name
Transaction Date
No Of Withdrawals
No Of XYZ Card Withdrawals
No Of Other Card Withdrawals
Total amount Withdrawn
Amount withdrawn XYZ Card
Amount withdrawn Other Card
Weekday - converted
Festival Religion - converted
Working Day - converted
year
month
day

Table 5.5: Columns Used in Prediction

Column Name
Transaction Date
Amount withdrawn XYZ Card
Weekday - converted
Festival Religion - converted
Working Day - converted
year
month
day

5.3 Results of the Models with Processed Data (Step 3 C)

After running the prediction model with processed data, it is seen that exploratory data analysis and data handling are as important processes as model selection. With processed data, the error metric of KNN algorithm decreases from 27% MAPE to 9,4% MAPE. Figure 5.6 shows the comparison of KNN model results with the raw data and the processed data. After processing, model can follow trend and peak points of the data better.

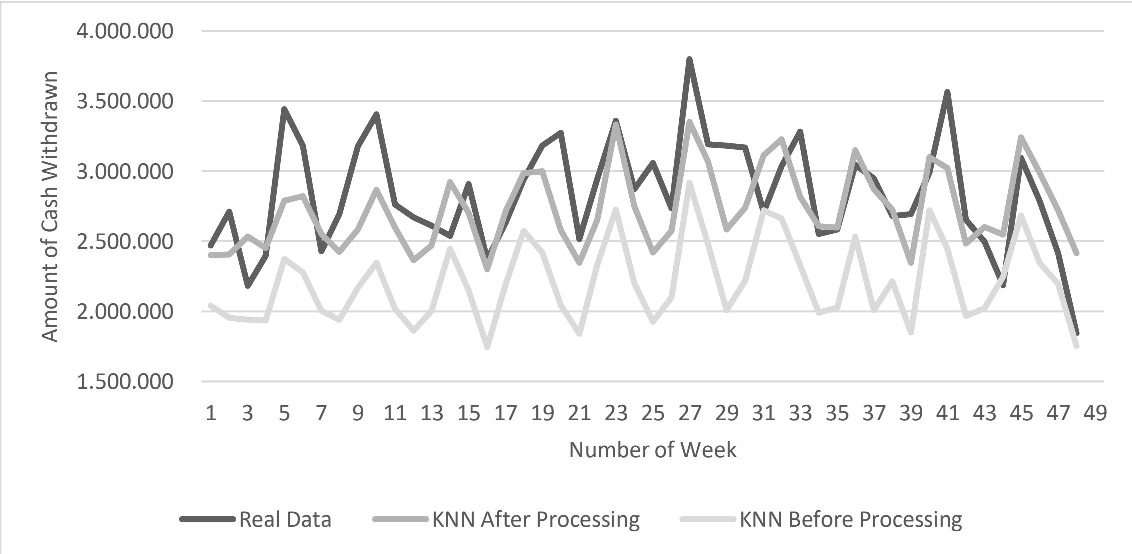


Figure 5.6: Comparison of Prediction Results Before and After EDA

In the methodology section, four machine learning algorithms are run with the dataset and results are presented. The algorithm with second best result is SVR with %28 MAPE by using the raw data. After processing the data, the error of the results decreases to 15%. Even though error of the model seems quite low, when examining the results, it is seen that the results create a horizontal line and does not follow the descents and ascents of the real data. As the Figure 5.7 shows, SVR is not a proper method to model a peaky demand data. However, it can be suitable for a make to stock production system (Smola et al., 2004).

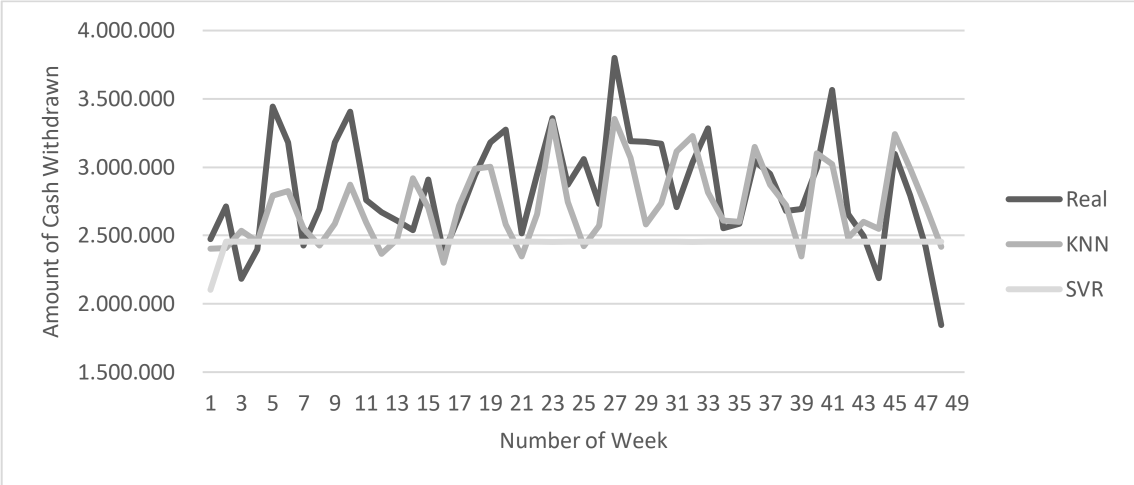


Figure 5.7: Comparison of SVR and KNN Prediction Results

In the next section, conclusion of the study is presented, a summary is given, and the possible further works are introduced.

6. CONCLUSION

In this thesis, ATM cash demand prediction problem is taken into consideration in banking sector. To determine the right amount of money to have in stock for an ATM is extremely important for avoiding excessive costs which are delivery cost, holding cost and opportunity cost. Therefore, to predict cash demand is a critical issue. Cash demand data has a multivariate and fluctuant structure and to construct a model for this kind of data is very difficult. An advanced forecasting method is necessary for accurate results. Machine learning algorithms are strong tools to understand data and predict future demand.

To provide real cash demand data is very difficult for information security reasons of banks. Thus, in the study, a cash withdrawal dataset is used which is provided from online data science platform Kaggle.com. The dataset has features like transaction date, weekday, working day and number of transactions. These features help algorithms to be trained better. Python programming language is used to create prediction model. Python codes are given in Appendices section.

Four machine learning algorithms; Support Vector Regressor, K'th Nearest Neighbor Regressor, Random Forest Regressor and Gradient Boosting Regressor, are chosen by considering their suitability to structure of the dataset that is a kind of time series data. Their performance on the dataset shows the effect of the working principle of algorithm on performance of prediction. While KNN algorithm can understand the structure of the data well and can predict the fluctuations and peak points of it, SVR algorithm creates a straight line even if its error is low in average. Mean Absolute Percentage Error (MAPE) is chosen as the error metric to measure performances of algorithms. Because of the large magnitudes of data elements, a percentage type error is more suitable to compare results.

In order to ensure accuracy of the results, MAE and R-squared error metrics are taken into consideration too. With the raw data, KNN algorithm gives best result with 27% MAPE and 92311 MAE. According to R-squared metric, the model can understand the variables of the data with 0.549 coefficient of determination value. A value which is near or below to 0 means that model cannot understand the data well. The second best algorithm is SVR according to MAPE (28%) and MAE (107589) values. However, SVR algorithm gives a negative R-squared value and shows that the algorithm cannot explain the data very well.

After selecting right algorithm for the model, exploratory data analysis and feature selection is applied to the dataset. Cleaning the data from noise and selecting proper features helps algorithm is trained well. Especially, statistical analysis is a useful tool as a part of EDA to examine the dataset in detail. Although multivariate data is beneficial to train machine learning algorithm better, selecting right variables for model is a complex decision. Correlation degree is a practical indicator to select features to use together. Correlation matrix is a useful tool to visualize correlation degrees of features to each other. On the other hand, algorithm-specific design of experiment application is also an important process to increase performance of the algorithm. For KNN algorithm to determine the optimum value of k-value is an application of DOE.

Comparison of models with raw data and processing data shows the importance of EDA and feature selection. MAPE of the model decreases from 27% to 9.4% with the same algorithm.

In the end of the study, 1 year of cash demand of an ATM is forecasted with a 9.4% error which is a quite low value for a demand prediction problem.

6.1 Limitations and Future Work

In this study, a dataset from the online platform kaggle.com is used because of to provide transaction data from banks is a difficult issue. The number of features of the dataset is limited and they tried to be used effectively. However, a dataset which has more features can be beneficial to increase the accuracy of the model. Features like location of the ATM, information of nearby shopping centers and corporate buildings and information of pay days of public and private sector employees can help the model to be trained and predict the demand better. In the future of this study, comprehensive dataset can be used to understand the behavior of the transaction data well.

As another further work of this problem, neural networks can be used with a bigger real-world dataset. As it is mentioned in Demand Prediction in Banking Sector section, time delay neural networks and long-short term memory networks as examples of recurrent neural networks are good performing algorithms for time series data like cash demand data. Comparison of these neural network algorithms with machine learning algorithms which are used in this study can be beneficial to develop the study.

REFERENCES

Sayar, C. and Wolfe, S. (2007). Internet banking market performance: Turkey versus the UK, *International Journal of Bank Marketing*, Vol. 25 No. 3, pp. 122-141.

The Banks Association of Turkey, Statistical Reports, **URL:** https://www.tbb.org.tr/tr/bankacilik/banka-ve-sektor-bilgileri/istatistiki-raporlar/59

Santos, B and Peffers, K. (1995). Rewards to Investors in Innovative Information Technology Applications: First Movers and Early Followers in ATMs, *Organization Science*, Vol. 6 No. 3, pp. 241-335.

Dilijonas, D. and Sakalauskas, V. (2011). Self-service Systems Performance Evaluation and Improvement Model, *IFIP Advances in Information and Communication Technology*, Vol. 353, pp. 87–98.

Boeschoten, W.C. (1998). Cash Management, Payment Patterns and the Demand for Money, *De Economist*, Vol. 146, pp.117–142.

Baltacioglu, T., Ada, E., Demirbag M., Yurt, O. (2007). A New Framework for Service Supply Chains, *Service Industries Journal*, Vol. 27, pp. 105-124.

Frohlich, M. and Westbrook, R. (2002). Demand chain management in manufacturing and services: web-based integration, drivers and performance, *Journal of Operations Management*, Vol.20, pp. 729-745.

Duk, B. and Seon, K. (2002). Forecasting telecommunication service subscribers in substitutive and competitive environments, *International Journal of Forecasting*, Vol. 18, pp. 561-581.

Wuttke, D. A., Blome, C., Henke, M. (2013). Focusing the financial flow of supply chains: An empirical investigation of financial supply chain management, *International Journal of Production Economics*, Vol. 145, pp. 773-789.

Castro, J. (2009). A stochastic programming approach to cash management in banking, *European Journal of Operational Research*, Vol. 192, No. 3, pp. 963-974.

Mols, N.P., Nikolaj D. Bukh, P. and Blenker, P. (1997). European corporate customers' choice of domestic cash management banks, *International Journal of Bank Marketing*, Vol. 15 No. 7, pp. 255-263.

Armstrong, J. (2001). Principles of Forecasting: A Handbook for Researchers and Practitioners. Kluwer Academic Publishers.

Barbosa, N., Christo, E.S., Alonso Costa, K. (2015). Demand forecasting for production planning in a food company, *ARPN Journal of Engineering and Applied Sciences*, Vol. 10, No. 1610. 7137-7141.

Whitfield, R. and Duffy, A. (2013). Extended revenue forecasting within a service industry, *International Journal of Production Economics*, Vol. 141, pp. 505–518.

Helms, M.M., Ettkin, L.P. and Chapman, S. (2000). Supply chain forecasting – Collaborative forecasting supports supply chain management, *Business Process Management Journal*, Vol. 6 No. 5, pp. 392-407.

Lay, G., Schroeter, M., Biege, S. (2009). Service-based business concepts: A typology for business-to-business markets, *European Management Journal*, Vol. 27, pp. 442-455.

Lee, J., Kao, H.-A., Yang, S. (2014). Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment, *Procedia CIRP*, Vol. 16, pp. 3–8.

Koshiba, H., Takenaka, T., Motomura, Y. (2013). A Service Demand Forecasting Method using a Customer Classification Model. *The Philosophers Stone for Sustainability*, pp. 281–285.

Baumol, W. J. (1952). The Transactions Demand for Cash: An Inventory Theoretic Approach. *The Quarterly Journal of Economics*, Vol. 66 No. 4, pp. 545.

Montgomery, D.C. and L.A. Johnson (1976). Forecasting and Time Series Analysis, McGraw-Hill, New York.

Constantinides, G. M. and Richard, S. F. (1978). Existence of Optimal Simple Policies for Discounted-Cost Inventory and Cash Management in Continuous Time. *Operations Research*, Vol 26(4), pp. 620–636.

Boeschoten, W. (1992). Currency use and payment patterns, *Financial and Monetary Policy Studies,* Vol. 23, pp. 244.

Simutis, R., Dilijonas, D., Bastina L. (2008). Cash Demand Forecasting for Atm Using Neural Networks and Support Vector Regression Algorithms, *EurOPT-2008*, pp. 416–421.

Rajwani, A. & Syed, T. (2017). Regression analysis for ATM cash flow prediction, *2017 International Conference on Frontiers of Information Technology*, pp. 212-217.

Paul, J. (2010). ATMs and Cash Demand Forecasting: A Study of Two Commercial Banks, Journal of Regional Development.

Cui, W. & Wang, J. & Ning, C. (2014). Time Series Prediction Method of Bank Cash Flow and Simulation Comparison, *Algorithms*, Vol. 7(4), pp. 650-662.

Brentnall, A., Crowder, M., Hand, D. (2008). A statistical model for the temporal pattern of individual automated teller machine withdrawals, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Vol. 57(1), pp. 43-59.

Simutis, R., Dilijonas, D., Bastina, L. (2007). A Flexible Neural Network for ATM Cash Demand Forecasting, *6. WSEAS international conference on computational intelligence, man-machine systems and cybernetics (CIMMACS 07)*, Vol. 162165.

Liu, N., Ren, S., Choi, T., Hui, C. (2013). Sales Forecasting for Fashion Retailing Service Industry: A Review, *Mathematical Problems in Engineering* 2013.

Kotsiantis, S.B., Zaharakis, I.D., Pintelas, P.E. (2006). Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review*, Vol. 26, pp. 159–190.

Hartwig, F., & Dearing, B. E. (2002). Exploratory data analysis. Newbury Park, CA: Sage Publ.

Michalski, R. S., Carbonell, J. G., Mitchell, T. M. (1983). Machine learning: an artificial intelligence approach. Palo Alto, CA: Tioga Pub. Co.

Nilsson, N. J. (2010). The quest for artificial intelligence: a history of ideas and achievement. Cambridge: Cambridge University Press.

Paluszek, M., & Thomas, S. (2016). An Overview of Machine Learning. *MATLAB Machine Learning*, Vol. 1, pp. 3–15.

Arel, I., Rose, D. C., Karnowski, T. P. (2010). Deep Machine Learning - A New Frontier in Artificial Intelligence Research, *IEEE Computational Intelligence Magazine*, Vol. 5 No. 4, pp. 13-18.

Laitinen, E.K., Laitinen, T. (1998), Cash Management Behavior and Failure Prediction, *Journal of Business Finance & Accounting*, Vol. 25, pp. 893-919.

Petrovic, D., Xie, Y., Burnham, K. (2006). Fuzzy decision support system for demand forecasting with a learning mechanism, *Fuzzy Sets and Systems*, Vol. 157(12), pp. 1713-1725.

Abraham, B., Ledolter, J. (2009). Statistical methods for forecasting (Vol. 234). John Wiley & Sons.

Gubar, E., Zubareva, M., Merzljakova, J. (2011). Cash Flow Optimization in ATM Network Model, *Contributions to Game Theory and Management*, Vol. 4, pp. 213–222.

Huynh, K. P., Schmidt-Dengler, P., Stix, H. (2014). Whenever and Wherever: The Role of Card Acceptance in the Transaction Demand for Money, *SFB/TR* 15, pp. 472.

Acuna, G., Currilem, M. (2012). Comparing NARX and NARMAX models using ANN and SVM for cash demand forecasting for ATM. *WCCI 2012 IEEE World Congress on Computational Intelligence.*

Andrawis, R., Atiya, A. (2011). Forecast combinations of computational intelligence and linear models for the NN5 time series forecasting competition, *International Journal of Forecasting*.

Sepúlveda, J. A., Akin, H. M. (2004). Modeling a garment manufacturer's cash flow using object-oriented simulation, *2004 IEEE Winter Simulation Conference*, Vol. 2, pp. 1176-1183).

Scholtz, E., Bekker, J., Du Toit, D. (2012). Multi-objective optimisation with stochastic discrete-event simulation in retail banking: A case study, *ORiON*, Vol. 28(2), pp. 117-135.

Stahl, I. (2019). Topics in discrete events simulation for business students, *2019 WSC Winter Simulation Conference*, pp. 3332-3343.

Atsalaki, I., Atsalakis, G., Zopounidis, C. (2011). Cash withdrawals forecasting by neural networks, *Journal of Computational Optimization in Economics and Finance*, Vol. 3, pp. 133.

Cardona, L., Moreno, L. (2012). Cash management cost reduction using data mining to forecast cash demand and LP to optimize resources, *Memetic Computing*. Vol. 4, pp. 127-134.

Venkatesh, K., Ravi, V., Prinzie, A., Van den Poel, D. (2014). Cash demand forecasting in ATMs by clustering and neural networks, *European Journal of Operational Research*, Vol. 232(2), pp. 383-392.

Arabani, S. P., Komleh, H. E. (2019). The Improvement of Forecasting ATMs Cash Demand of Iran Banking Network Using Convolutional Neural Network, *Arabian Journal for Science and Engineering,* Vol. 44(4), pp. 3733-3743.

Leonov, P., Sviridenko, A., Leonova, E., Epifanov, M., Nikiforova, E. (2020). The use of artificial intelligence technology in the process of creating an ATM service model, *Procedia Computer Science,* Vol. 169, pp. 203-208.

Längkvist, M., Karlsson, L., Loutfi, A. (2014). A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters,* Vol. 42, pp. 11-24.

McCulloch, W. S., Pitts, W. (1947). How we know universals: the perception of auditory and visual forms, *Bulletin of mathematical Biophysics,* Vol. 9(3), pp. 127-147.

Hebb, D. O. (1949). The organization of behavior: a neuropsychological theory. J. Wiley; Chapman & Hall.

Turing, A. (1950). Computing machinery and intelligence-AM Turing, *Mind,* Vol. 59(236), pp. 433.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers, *IBM Journal of research and development,* Vol. 3(3), pp. 210-229.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain, *Psychological review*, Vol. 65(6), pp. 386.

Domingos, P. (2012). A few useful things to know about machine learning, *Communications of the ACM,* Vol. 55(10), pp. 78-87.

Friedman, J. H., Tukey, J. W. (1974). A projection pursuit algorithm for exploratory data analysis, *IEEE Transactions on computers*, Vol. 100(9), pp. 881-890.

Behrens, J. T. (1997). Principles and procedures of exploratory data analysis, *Psychological Methods*, Vol. 2(2), pp. 131.

Martinez, W. L., Martinez, A. R., Solka, J., Martinez, A. (2010). Exploratory data analysis with MATLAB. Crc Press.

Behrens, J. T., DiCerbo, K. E., Yel, N., Levy, R. (2012). Exploratory data analysis. Handbook of Psychology, Second Edition, 2.

Xue, B., Zhang, M., Browne, W. N., Yao, X. (2015). A survey on evolutionary computation approaches to feature selection, *IEEE Transactions on Evolutionary Computation*, Vol. 20(4), pp. 606-626.

Abdi, H., Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, Vol. 2(4), pp. 433-459.

Ramanathan, U. (2012). Supply chain collaboration for improved forecast accuracy of promotional sales. *International Journal of Operations & Production Management.*

Christensen, K., Nørskov, S., Frederiksen, L., & Scholderer, J. (2017). In search of new product ideas: Identifying ideas in online communities by machine learning and text mining. *Creativity and Innovation Management*, Vol 26(1), pp. 17-30.

Gartner Research, Market Guide for Trade Promotion Management and Optimization for the Consumer Goods Industry,
URL: https://www.gartner.com/en/documents/3587241

Rod Koch CMA, P. M. P. (2015). From business intelligence to predictive analytics. *Strategic Finance*, Vol. 96(7), pp. 56.

Folinas, D., & Rabi, S. (2012). Estimating benefits of Demand Sensing for consumer goods organisations. *Journal of Database Marketing & Customer Strategy Management*, Vol. 19(4), pp. 245-261.

Caruana, R., Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms, *23rd International Conference on Machine learning*, pp. 161-168.

Ghahramani, Z. (2003). Unsupervised learning, *Summer School on Machine Learning, Springer*, pp. 72-112.

Kumar, V., Kalitin, D., Tiwari, P. (2017). Unsupervised learning dimensionality reduction algorithm PCA for face recognition, *2017 International Conference on Computing, Communication and Automation (ICCCA)*, pp. 32-37.

Eskin, E., Arnold, A., Prerau, M., Portnoy, L., Stolfo, S. (2002). A geometric framework for unsupervised anomaly detection, *Applications of data mining in computer security*, pp. 77-101.

Kaelbling, L. P., Littman, M. L., Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of artificial intelligence research*, Vol. 4, pp. 237-285.

Matsunaga, A., & Fortes, J. A. (2010). On the use of machine learning to predict the time and resources consumed by applications, *2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing*, pp. 495-504.

Seber, G. A., & Lee, A. J. (2012). Linear regression analysis, Vol. 329. John Wiley & Sons.

Mohammed, M., Bashier, E. (2016). Machine Learning: Algorithms and Applications. Crc Press.

Freund, Y., Mason, L. (1999). The alternating decision tree learning algorithm, *ICML*, Vol. 99, pp. 124-133.

Kuang, Q., Zhao, L. (2009). A practical GPU based kNN algorithm, *2009 International Symposium on Computer Science and Computational Technology (ISCSCI 2009)*, pp. 151.

Kotsiantis, S. (2007). Supervised Machine Learning: A Review of Classification Techniques, *Emerging artificial intelligence applications in computer engineering*, Vol. 160, pp. 3-24.

Weigend, A. S. (2018). Time series prediction: forecasting the future and understanding the past, Routledge.

Ak, R., Fink, O., Zio, E. (2015). Two machine learning approaches for short-term wind speed time-series prediction, *IEEE transactions on neural networks and learning systems*, Vol. 27(8), pp. 1734-1747.

Sapankevych, N. I., Sankar, R. (2009). Time series prediction using support vector machines: a survey, *IEEE Computational Intelligence Magazine*, Vol. 4(2), pp. 24-38.

Davydenko, A., Fildes, R. (2013). Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts, *International Journal of Forecasting*, Vol. 29(3), pp. 510-522.

Saigal, S., Mehrotra, D. (2012). Performance comparison of time series data using predictive data mining techniques, *Advances in Information Mining*, Vol. 4(1), pp. 57-66.

Franses, P. H. (2016). A note on the mean absolute scaled error, *International Journal of Forecasting*, Vol. 32(1), pp. 20-22.

Khair, U., Fahmi, H., Al Hakim, S., Rahim, R. (2017). Forecasting error calculation with mean absolute deviation and mean absolute percentage error, *Journal of Physics: Conference Series*, Vol. 930 No. 1.

Myatt, G. J. (2007). Making sense of data: a practical guide to exploratory data analysis and data mining, John Wiley & Sons.

Van den Broeck, J., Cunningham, S. A., Eeckels, R., Herbst, K. (2005). Data cleaning: detecting, diagnosing, and editing data abnormalities, *PLoS medicine*, Vol. 2(10).

Chatfield, C., Xing, H. (2019). The analysis of time series: an introduction with R, CRC press.

Pukelsheim, F. (1994). The three sigma rule, *The American Statistician*, Vol. 48(2), pp. 88-91.

Li, D., Zhang, B., Li, C. (2015). A feature-scaling-based k-nearest neighbor algorithm for indoor positioning systems, *IEEE Internet of Things Journal*, Vol. 3(4), pp. 590-597.

Tsai, C. F. (2009). Feature selection in bankruptcy prediction, *Knowledge-Based Systems*, Vol. 22(2), pp. 120-127.

Jin, Y. (2006). Multi-objective machine learning, Berlin: Springer.

Antony, J. (2014). Design of experiments for engineers and scientists. Elsevier.

Gweon, H., Schonlau, M., Steiner, S. H. (2017). Statistical learning approaches to some classification problems, Waterloo, Ontario, Canada: University of Waterloo.

Martínez, F., Frías, M. P., Pérez, M. D., Rivera, A. J. (2019). A methodology for applying k-nearest neighbor to time series forecasting, *Artificial Intelligence Review*, Vol. 52(3).

Yin, Y., Shang, P. (2016). Forecasting traffic time series with multivariate predicting method, *Applied Mathematics and Computation*, Vol. 291, pp. 266-278.

Mehrotra, R., Sharma, A. (2006). Conditional resampling of hydrologic time series using multiple predictor variables: A K-nearest neighbour approach, *Advances in water resources*, Vol. 29(7), pp. 987-999.

Smola, A.J. & Schölkopf, B. (2004). A tutorial on support vector regression, *Statistics and Computing*, Vol. 14(3), pp. 199-222.

Jha, G.K., Sinha, K. (2014). Time-delay neural networks for time series prediction: an application to the monthly wholesale price of oilseeds in India. *Neural Comput & Applic*, Vol. 24, pp. 563–571

APPENDICES

Appendix A. – Python Code

Import Libraries

In [3]:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from datetime import datetime
```

In [4]:

```
#Veri oku
data = pd.DataFrame(pd.read_excel("AggregatedData.xlsx"))

#print(data.columns)
#sns.countplot(data['ATM Name'],label="Count")

#Big Street ATM
#Mount Road ATM
#Airport ATM **
#KK Nagar ATM
#Christ College ATM *


data = data.loc[data['ATM Name'] == 'Airport']
data = data.loc[data['Transaction Date'] < '2013-01-01']

data["year"]= data["Transaction Date"].apply(lambda x: x.year)         #.apply(lambda
 x: x.split('-')[0])
data["month"]= data["Transaction Date"].apply(lambda x: x.month)
data["day"]= data["Transaction Date"].apply(lambda x: x.day)
data.index= data["Transaction Date"]


data['Working Day'] = data['Working Day'].astype('category')
data['Working_Day_cat'] = data['Working Day'].cat.codes
data['Festival Religion'] = data['Festival Religion'].astype('category')
data['Festival_Religion_cat'] = data['Festival Religion'].cat.codes
data['Weekday']=data['Weekday'].apply(lambda x: x.lower())
data['Weekday'] = data['Weekday'].astype('category')
data['Weekday_cat'] = data['Weekday'].cat.codes
```

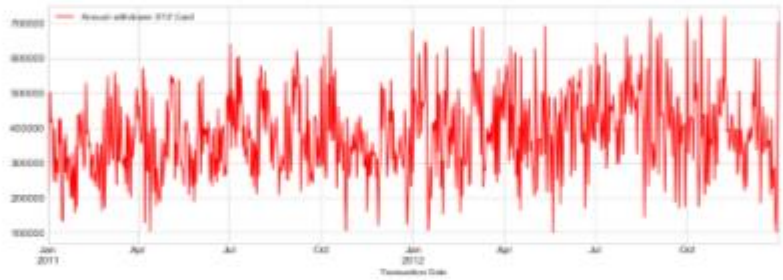
Data Analysis

In [5]:

```
print(data.info())
data.isna().sum()
data = data[pd.isna(data)==False]
print(data.describe())
sns.countplot(data['Amount withdrawn XYZ Card'],label="Count")
```

In [6]:

```
#plot çiz
pred=data["Amount withdrawn XYZ Card"]
plt.rcParams['figure.figsize'] = [15, 5]
#plt.figure(figsize = (36,36))
plt.style.use('seaborn-whitegrid')
pred.plot(fontsize=12, color='red',legend=True)
plt.show()
```



In [7]:

```
# Veri dağılımının incelenmesi
col_names = [ 'No Of Other Card Withdrawals', 'Amount withdrawn Other Card'   ] #  'No
 Of Withdrawals','No Of XYZ Card Withdrawals',  'Total amount Withdrawn', 'Amount withd
rawn XYZ Card']



fig, ax = plt.subplots(len(col_names), figsize=(12,12))

for i, col_val in enumerate(col_names):

    sns.distplot(data[col_val], hist=True, ax=ax[i])
    ax[i].set_title('Freq dist '+col_val, fontsize=20)
    ax[i].set_xlabel(col_val, fontsize=15)
    ax[i].set_ylabel('Count', fontsize=15)

plt.show()
```

Feature Selection

In [8]:

```
#Feature Selection

from sklearn.ensemble import ExtraTreesClassifier

y = data["No Of XYZ Card Withdrawals"]

from sklearn import preprocessing
data_ = data.drop(data.columns[[0,1,6,8,9,10,11]], axis=1)
data_std = pd.DataFrame(preprocessing.scale(data_))
data_std.columns =data_.columns
data_std.index= data["Transaction Date"]
X = data_std


forest = ExtraTreesClassifier(n_estimators=250,
                              random_state=0)

forest.fit(X, y)
importances = forest.feature_importances_
std = np.std([tree.feature_importances_ for tree in forest.estimators_],
             axis=0)
indices = np.argsort(importances)[::-1]

print("Feature ranking:")

for f in range(X.shape[1]):
    print("%d. feature %d (%f)" % (f + 1, indices[f], importances[indices[f]]))

plt.figure()
plt.title("Feature importances")
plt.bar(range(X.shape[1]), importances[indices],
       color="r", yerr=std[indices], align="center")
plt.xticks(range(X.shape[1]), indices)
plt.xlim([-1, X.shape[1]])
plt.show()
```

In [9]:

```
#Correlation matrix
features = data.drop(data.columns[[0,1,6,8,9,10,11]], axis=1)

corr = features.corr()
plt.figure(figsize=(10,10))
sns.heatmap(corr, cbar = True,  square = True, annot=True, fmt= '.2f',annot_kws={'size'
: 12},
           xticklabels= features.columns, yticklabels= features.columns,
           cmap= 'coolwarm')
```

Data Processing

In [10]:

```
#Outlier elimination

minval = np.mean(data["Amount withdrawn XYZ Card"]) - 3*np.std(data["Amount withdrawn X
YZ Card"])
maxval = np.mean(data["Amount withdrawn XYZ Card"]) + 3*np.std(data["Amount withdrawn X
YZ Card"])

data = data[data["Amount withdrawn XYZ Card"] >= minval]
data = data[data["Amount withdrawn XYZ Card"] <= maxval]


#print (data)
#Veri standardizasyonu

from sklearn import preprocessing
dt = data.drop(data.columns[[0,1,6,8,9,10,11]], axis=1)
up_data_std = pd.DataFrame(preprocessing.scale(dt))
up_data_std.columns =dt.columns
up_data_std.index = data.index
up_data_std
```

Prediction Model

In [12]:

```
#Train ve validation

from sklearn.neighbors import KNeighborsRegressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.gaussian_process import GaussianProcessRegressor
from sklearn.svm import SVR
from sklearn.neural_network import MLPRegressor

from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import r2_score

from sklearn import ensemble
import warnings
warnings.filterwarnings('ignore')


from math import sqrt

from sklearn.model_selection import TimeSeriesSplit
from sklearn.model_selection import cross_val_score
```

```
training_end = '2011-12-31'
prediction_start = '2012-01-31'
prediction_end = '2012-12-31'

selected_columns = [ 'year' , 'month' ,'day',
                     'Working_Day_cat','Festival_Religion_cat', 'Weekday_cat'] # 'No Of
 Other Card Withdrawals', 'Total amount Withdrawn','Amount withdrawn Other Card','No Of
Withdrawals','No Of XYZ Card Withdrawals',


 #'year' , 'month' ,'day', 'Working_Day_cat','Festival_Religion_cat', 'Weekday_cat'

x_train = up_data_std[selected_columns][:training_end]
y_train = data["Amount withdrawn XYZ Card"][:training_end]
x_test = up_data_std[selected_columns][prediction_start:prediction_end]
y_test= data["Amount withdrawn XYZ Card"][prediction_start:prediction_end]

best_model_list =[]
result=dict()


#KNN Modeli
print('KNN')
model = KNeighborsRegressor(n_neighbors=50).fit(x_train, y_train)
y_predict = model.predict(x_test)
y_predict = pd.DataFrame(y_predict,index=y_test.index, columns=['prediction'])
#print(y_predict)
y_true, y_pred = np.array(y_test), np.array(y_predict)

tscv = TimeSeriesSplit(n_splits=4)
scores = cross_val_score(model, x_train, y_train, cv=tscv, scoring = 'neg_mean_absolute
_error')
print("Loss: {0:.3f} (+/- {1:.3f})".format(scores.mean(), scores.std()))
print(scores)
perc_error =np.mean(np.abs((y_true - y_pred) / y_pred))*100
```

```
print("mean absolute percentage error: {} %".format(perc_error))

mae = mean_absolute_error(y_test, y_predict)
print("MAE: {} ".format(mae))
result["model"]="KNN"
result["error"]=perc_error
result["sonuclar"]=y_predict
best_model_list.append(result)
result=dict()
print("\n")


#Random Forest Modeli
print('Random Forest')
model = RandomForestRegressor(n_estimators=10, random_state=10).fit(x_train, y_train)
y_predict = model.predict(x_test)
y_predict = pd.DataFrame(y_predict,index=y_test.index, columns=['prediction'])
y_true, y_pred = np.array(y_test), np.array(y_predict)

tscv = TimeSeriesSplit(n_splits=4)
scores = cross_val_score(model, x_train, y_train, cv=tscv, scoring='neg_mean_absolute_e
rror')
print("Loss: {0:.3f} (+/- {1:.3f})".format(scores.mean(), scores.std()))
print(scores)
perc_error =np.mean(np.abs((y_true - y_pred) / y_pred))*100
print("mean absolute percentage error: {} %".format(perc_error))

mae = mean_absolute_error(y_test, y_predict)
print("MAE: {} ".format(mae))
result["model"]="RF"
result["error"]=perc_error
result["sonuclar"]=y_predict
best_model_list.append(result)
result=dict()
print("\n")
```

```
#Boosting Modeli
print('GBR')
params = {'n_estimators': 1000, 'max_depth': 3, 'min_samples_split': 2,
          'learning_rate': 0.5, 'loss': 'ls'}
model = ensemble.GradientBoostingRegressor(**params)
model.fit(x_train, y_train)
y_predict = model.predict(x_test)
y_predict = pd.DataFrame(y_predict,index=y_test.index, columns=['prediction'])
y_true, y_pred = np.array(y_test), np.array(y_predict)

tscv = TimeSeriesSplit(n_splits=4)
scores = cross_val_score(model, x_train, y_train, cv=tscv, scoring='neg_mean_absolute_e
rror')
print("Loss: {0:.3f} (+/- {1:.3f})".format(scores.mean(), scores.std()))
print(scores)
perc_error =np.mean(np.abs((y_true - y_pred) / y_pred))*100
print("mean absolute percentage error: {} %".format(perc_error))

mae = mean_absolute_error(y_test, y_predict)
print("MAE: {} ".format(mae))
result["model"]="GBR"
result["error"]=perc_error
result["sonuclar"]=y_predict
best_model_list.append(result)
```

```
result=dict()
print("\n")


#SVR Modeli
print('SVR')
model = SVR().fit(x_train, y_train)
y_predict = model.predict(x_test)
y_predict = pd.DataFrame(y_predict,index=y_test.index, columns=['prediction'])
y_true, y_pred = np.array(y_test), np.array(y_predict)
#print(y_predict)

tscv = TimeSeriesSplit(n_splits=4)
scores = cross_val_score(model, x_train, y_train, cv=tscv, scoring='neg_mean_absolute_e
rror')
print("Loss: {0:.3f} (+/- {1:.3f})".format(scores.mean(), scores.std()))
print(scores)
perc_error =np.mean(np.abs((y_true - y_pred) / y_pred))*100
print("mean absolute percentage error: {} %".format(perc_error))

mae = mean_absolute_error(y_test, y_predict)
print("MAE: {} ".format(mae))
result["model"]="SVR"
result["error"]=perc_error
result["sonuclar"]=y_predict
best_model_list.append(result)
result=dict()
print("\n")


best_model_sel= [(e["error"])for e in best_model_list]
best_model_ind = np.argmin(best_model_sel)
best_model=best_model_list[best_model_ind]["model"]
best_sonuclar=best_model_list[best_model_ind]["sonuclar"]

print("best model: {} ".format(best_model))
print(best_sonuclar)
print("\n")
```

```
y_predict_ = y_predict.resample('W').agg(['sum']).astype(int)
print("aylık tahmin")
print(y_predict_)
print("\n")
y_test_ = y_test.resample('W').agg(['sum'])
print("aylık gerçek")
print(y_test_)
```

BIOGRAPHICAL SKETCH

Evrim Gençalp was born in Edirne on February 16, 1991. She studied at Çanakkale Science High School where she was graduated in 2009. He attended the undergraduate program of Industrial Engineering in Middle East Technical University. She received her B.S. degree in the Industrial Engineering in 2014. Since graduation she has been working in the technology development in the private sector. Also, she is working towards Master of Science degree in Industrial Engineering under the supervision of Assoc. Prof. Dr. Müjde Genevois the Institute of Science and Engineering, Galatasaray University.

PUBLICATIONS

Gencalp, E., Genevois, M. E., Cedolin M. (2019). Atm Replenishment Forecasting with Support Vector Machine – A Case Study, *International Journal of Industrial Electronics and Electrical Engineering*, Vol. 7(10).