5-15-2022

Buyer Prediction Through Machine Learning

Rashed Ibrahim Karmostaje
rik7553@rit.edu

RIT

BUYER PREDICTION THROUGH MACHINE LEARNING

By

RASHED IBRAHIM KARMOSTAJE

A Graduate Paper/Capstone Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Professional Studies in Data Analytics

Department of Graduate Programs & Research

Rochester Institute of Technology

RIT Dubai

May 15, 2022

RIT

Master of Science in Professional Studies in

Data Analytics

Graduate Capstone Approval

Student Name: Rashed Ibrahim Karmostaje

Paper/Capstone Title: Buyer Prediction Through Machine Learning

Graduate Capstone Committee:

Name: Dr. Sanjay Modak **Designation: Chair Committee** **Date:**

Name: Dr. Ehsan Warriach **Designation: Member of committee** **Date:**

Acknowledgments

First of all, I am thankful to almighty Allah for all praises and glory to the most compassionate and merciful.

Special thanks and appreciation to my teacher mentor and advisor Dr. Ehsan Warriach. He continuously supports me in this work. His corporation, assistance, and continuous support in this capstone project are appreciable. I want to thank Dr. Sanjay Modak, chair of graduate program and Research at Rochester Institute of Technology RIT Dubai, for his advice, assistance, feedback, and guidance in the research project along with the complete master's program.

I would like to thank my family and parents for supporting me. My parents always supported me throughout my education, especially in the master's program. Additionally, my family, friends, and colleagues also assist me throughout my research.

Abstract

Targeted marketing has grown in popularity in recent years, as well as recognizing when a consumer will desire a commodity may be extremely important to a business. Predicting this demand, however, is a complex procedure. Businesses, promoters/marketers, and sellers are using machine learning approaches to execute buyer prediction. This study focuses on when a customer would buy fast-moving retail merchandise by evaluating a customer's purchase history at partner vendors. The projections should be used to customize special discounts for customers who are about to make a purchase. In addition, buying behavior is a set of consumption habits that can be analyzed to help in predicting the needs of specific target audience. Knowing consumption habits, business is much more likely to formulate sales items tailored to the market. Thus, the chances of success and acceptance of products and services increase. Promotional offers can then be supplied to the most relevant clients (with alerts sent directly to buyers' mobile devices) thus reducing the use of the traditional/general paper-based marketing. More specifically, I will create a machine learning model that predicts potential future buyers based on the supplied market dataset. I will use a data source that gathers clients' consumer history to establish a solid basis for this approach. The study focuses on consumer groupings rather than individual purchasers to forecast purchasing. After analyzing which of these purchase behaviors fits the consumer's decision-making of a product or service, it will be easy to establish appropriate/focused marketing and sales strategies.

Keywords: Targeted Marketing, Forecasting, Machine Learning, Buyer decision making, Prediction buying behavior.

Table of Contents

List of Figures

Chapter 1: Introduction

1.1 Statement of the Problem

Most industrial activities in the automobile sector are still heavily reliant on human judgments. The issue with every manufacturing organization is the number of products produced. If they could not comprehend the product's specifications before manufacture, the corporation may suffer a loss. Our project assists the industry in predicting the proper product demand based on historical client behavior. Using several classifications algorithms on the cycle buyer dataset, the optimum method must be chosen based on accuracy to determine if a person is willing to buy our goods or not.

1.2 Background of the Problem

Many businesses are having to deal with the reality of AI implementation. The advantages of using a predictive system include assisting in determining the actual quantities of potential buyers, which is incredibly strategic and valuable for any firm. Implementing ML-based solutions can result in significant cost savings, better predictability, and increased system availability.

1.3 Project Definition and Goals

Marketing has shifted from a manufacturer to a client strategy. In this information-rich age, client behavior may assist affiliate marketers in selecting the most successful marketing campaign for their clients. If a corporation knows whatever a consumer tries to buy at any particular time, it may sell to the client's demands and achieve a competitive edge. Predicting a client's purchasing behavior allows businesses to promote to a specific person at the right moment. This will have a different influence on a buyer than the usually printed newspaper leaflet.

Customers that use your product leave a trail of actions that suggest how they will behave in the future. We can find predicted patterns in granular consumer behavioural data that may improve the customer experience and earn more income for your organization via automated feature engineering. Predictions are utilized to create customized marketing strategies and service offerings. We will attempt to evaluate and examine the performance of several machine-learning approaches used to forecast churn.

Targeted marketing has gained in popularity in recent years, and as well as recognizing when a consumer will need a product may be pretty beneficial to a company. This, on the other hand, is a challenging assignment to predict. This research describes a study that used machine learning

techniques to anticipate when a client would buy a product. This is accomplished by studying a customer's past purchases at partner retailers. These forecasts would customize special discounts for clients who want to buy an item.

I utilize a vast amount of data on their existing clients, including demographic information and previous purchases. I'm especially interested in examining customer data to see any obvious links between known demographic variables about consumers and the chance of a customer purchasing an item.

My objective is to create a model that can respond to the query asked by the firm's managers. I specifically will use a machine learning model that predicts potential future buyers based on the supplied dataset. I will use a data source that gathers past client data to establish a solid basis for this approach. The study focuses on consumer groupings rather than individual purchasers to forecast purchasing.

1.4 Research Methodology

Correlation metrics, we can see the column set with a high correlation, and we only selected those with high correlation in our features column list. We will use these columns as input features to our model, and the output features will be the status of the purchase pipe so that we will restrict our data into input training and testing and output training and testing. We make a model out of it, train it on the train data set, predict it on the testing data set, and make confusing metrics. So wherever the data is missing, we replace that empty data with a string of no data. So here, we also plot the histogram graph that tells the relation between income and the purchase of bikes, and in the graph, we can see it's a simple graph analysis which shows us that people with more income purchase more bikes. And then what we did was we had some missing data in the marital status column, gender column, and homeowner column. So in our data, wherever we see the word married, written will be replaced by zero, single will be replaced by one, and no Riddle will be replaced by two. So, for the buying analysis, we first imported the necessary libraries and imported the data set. The data set is the link in the initial document of the problem so that I can find the link there. This is because a machine learning model cannot understand strings. It only understands numeric data. So, what we did here we see all the frequencies of those variables and again drop them in the ID column, and so here we can see that there are many dictionaries written an equals to curly braces married single colon one no. Then we've made a correlation metric and plotted the

correlation metrics as a heat map. Some three or four more variants of these plots have different regions with occupation and income.

Chapter 2: Literature Review

According to Baderiya and Chawan (2018), the cost of maintaining a client is less than the cost of acquiring new ones due to the marketing expenditures associated with attracting new clients. As a result, with the increase in competition, it has become critical to keep the present consumer base. Customers often leave gradually rather than abruptly. This suggests that by examining consumers' past purchasing habits, one may take a proactive approach to forecasting turnover. In the field of academia, there have been a number of different machine-learning methods analyzed in the past decade to predict profitability and customer retention, and practitioners have been known to have used some of these. In majority of the cases, these methods are mostly based on extracting the customers' latent characteristics from their previous purchase behavior and characteristics. The estimate of what buyers might purchase, particularly in the rapidly personal items industry indicates that customer cross- and up-selling research comprises market basket analysis, forecasting clients' shopping lists, and estimating future purchases.

In a paper exploring the prediction of customer shopping lists from point-of-sale data, Cumby, Fano, Ghani, and Krema (2005) establish that market basket analysis is done to understand which things customers buy to improve marketing targets. It focuses on point-of-sale transaction records to increase sales and control stocks. Retailers also use it to plan the layout of their stores and guarantee that goods commonly purchased together are close to each other. Consumer purchase behavior influenced by several factors for the purchase of different products. Local customs interfere with the way the customer buys. Each region has different habits, unique experiences, different religions, rituals, norms, traditions, musical preferences and so on. Understanding the culture of the area can help in the development of products and services. Finally, the important thing is to be aware of what may embarrass, annoy or even offend potential customers as well as, of course, the qualities of your products that can bring moral benefits and a sense of social responsibility to buyers.

In another study, Els (2019) conducted research that involved the creation of a tailored discount offer system. When consumers visit the store, the system suggests a customized discount offer. The promotions are only available to a single client for a limited time. Consumers were subjected to cross- and up-selling under this system, which opened up the possibility of new revenue streams. The researcher also sought to forecast when a client would buy particular things using analytical approaches. In order to accurately predict the characteristics of the purchasing behavior of the

customers who purchased scooters in the past, the technology related to machine-learning was brought into use to analyze and then predict different characteristics of customer purchasing behavior.

In study on consumer behavior during Covid-19 pandemic, Loxton et al. (2020) established that income level has direct influence on consumer purchase behavior. In this regard, it has been found that there were significant Changes in habits and behavior patterns caused by the coronavirus. The consumer reduced consumption during the period of social isolation due to the impact on income and fear of unemployment. For the post-pandemic period, mostly individual (percentages between 50 and 72%) specified they plan to maintain the level of consumption adopted during social isolation, indicating that consumers in the future will hardly purchase goods at the levels they did before the pandemic.

Xu, Wang, Peng, and Wu (2019) demonstrate that Conscious Consumption showed that 38% of people are concerned about the impacts of production on the environment. The higher the level of education, the greater the willingness to check the origin of a product they purchase is environmentally correct. As around 57% of individual with up to the 4th grade of elementary school never check it and this percentage drops as the degree increases. Education is reaching 29% among people with higher education. Research on conscious consumption has also shown that income also influences the purchase of an environmentally correct product.

In addition, according to Linoff and Berry (2011), empirical research confirms the existence of children's influence on family decisions, but little is known about this, and, according to the author, analysis on the influence of children's consumers around luxury goods or where to go during holidays, has being little explored with regard to consumer choice. In this way, it is evident the need for more research that promotes the influence of the child consumer in other countries, in addition to focusing on specific areas, such as food. In view of this, this study consists of identifying the determining factors of the influence of the child consumer in the purchase decision process and consumption of food in the family.

Cumby et al. (2004) emphasize that traditional norms and parental decision rules in the family have decreased, making communication in the family more open and democratic, making children achieve more power to influence purchasing decisions especially about food. The development of child psychology and the segmentation of increasingly specific niches created an opportune environment for the study of child consumer behavior, mainly focusing on their ability to influence

family decisions, which brings relevant contributions to marketing, enabling the establishment of more targeted marketing strategies for this niche. Many studies on the subject were carried out, however, mainly in the United States and European countries, with little expression in other countries.

According to Khan et al. (2020), the purchasing power of the consumers influences their decision making. By identifying the social class, it is possible to define quality, price the products, establish payment methods and develop appropriate strategies. Products intended for lower classes, for example, need popular appeal. The upper classes, on the other hand, prefer products that make their premium character clear, as they aim at exclusivity. This question concerns the way consumers spend their time, daily habits, and plans for the future. Depending on the lifestyle, the consumer can spend a lot or be economical. Modern people, for example, can invest more in technological innovations and pay attention to the design of objects. Whatever the situation, having this information is critical for your business to reach the right audience.

Jiawei, Micheline, and Jian (2016) note that influencers have great contribution in purchase decision. These people have more power over consumer buying behavior than you might think. With the popularity of the internet, digital influencers are setting the trend and triggering acquisition needs. So, to understand potential customer, business need to know who they look up to and who they look up to. All it takes is a famous blogger to show off the makeup brands she uses, for example, for the products to disappear from store shelves. To understand customer's customs, business also need to know the people they interact with. The circles of friends and family interfere a lot in consumer decisions.

According to Uddin et al.(2019), people do online research before purchasing a product. Opinions exposed on the internet can influence the decision. However, the recommendation of close acquaintances carries even greater weight. All behavior is dictated by the personality of the individual: interests, choice of profession, hobbies, and friendships, among others. Therefore, studying this topic is important to understand the buyer's actions and needs. A shy person, for example, usually opts for homemade programs to have fun, such as watching movies and series. Based on the personality information, you can understand what she is willing to buy.

Neves, Leander, González, and Karoumi (2019) argue that identifying what stage of life the consumer is in is essential to understand the impact that a product or service will have in the market. Over time, human beings can change the way they think, dress, and even replace priorities.

Young buyers, for example, may favor volume of purchases; the more experienced ones tend to opt for the quality of the product or service and the service. Thus, the approach can never be the same for both audiences. Personal motivations are linked to the consumer's life goals or even with the purposes of the moment. Personal motivation can be related to the acquisition of something a car, motorcycle, property, setting up a business, taking a trip, among others. Knowing what led the customer to choose a particular product is important to understand consumer behavior.

According to Kim, Chae, and Olson (2012), a bad experience with a brand, product or service can influence all purchase decisions in the future. Therefore, it is important to study the customer's history, including interaction with the competition. This analysis serves to understand which attitudes are well received by the customer and what he does not tolerate. When you find out that he stopped going to an establishment because of poor service, it means that you need to offer excellent assistance to win that customer. Mapping and understanding consumer buying behavior is essential for attaining business goals. Several factors can interfere in the customer's purchase journey, and knowing these characteristics is essential to understand what customer needs are and wants.

Hosseini and Shabani (2015) argue that consumers study the market to make decisions on critical purchase interests. For example, buying a home is a complex operation in several aspects, such as legal, bureaucratic, and financial. It can be even more complicated when it involves a mortgage loan. This is because the approval and granting of real estate credit is based on the client's risk assessment. The bank, in general, makes several requirements and analyzes the buyer's profile and ability to pay for high and long-term debt, which can reach 30, 35 years, depending on the financial institution. Therefore, among the conditions observed by the bank is the age of the borrower. However, the truth is that the property belongs to the bank until you finish paying for it. If individual cannot pay the installments, the property will be taken. Another important point: the interest on the financing can make the property cost up to twice as much. Even though the interest is low, it is necessary to evaluate because it is applied to a high amount and for a long period of time. When buying a physical property, the investor will have expenses such as certificates, registrations and, depending on the property, the value will be high. When investing in Real Estate Investment Fund, one can purchase just one share, making it much more affordable. In addition, selling a fund share is easier and faster than selling a property. Another advantage is that in an

Real Estate Investment Fund, all administration is done by the manager and administrator, unlike a property, where the owner has to bear possible works and rent payment delays.

Lo-Ciganic et al. (2019), in the analysis of customers' behavior for detecting customer behavior through artificial intelligence, found that homeowners have a number of cars according to their region. In several sectors, the hiring of professionals with good salary averages has grown in the beginning of 2018. The scenario is a good indicator to evaluate the economic recovery of the country after some years of retraction. The IT development area is among those that pay the best. Part of this is due to the digitization of numerous processes and activities in the most diverse segments of the economy, from industry to services.

Brei (2020) notes that the accounting field has sought increasingly qualified professionals, attentive to the new rules. If individual need to dispose of their property, for example, this will depend on the market and a buyer, which does not always appear quickly. Another important factor is vacancy. An investor who owns a property runs the risk of the tenant leaving and the property being vacant. Therefore, he will have to bear all the costs involved. If a small and cheaper house meets their needs and that of their family, take the opportunity to save and make other types of investments, both financial and personal.

Raorane, Kulkarni, and Jitkar (2012) note that in times of crisis, differentiating products and services is the best strategy to conquer a prominent space in the market. In this regard, experts are unanimous in saying that investing in a property too early, through financing, can be a choice with a not very positive impact on the young person's life. That's because these individuals are still at the beginning of his career and, in most cases, with little money to make a good down payment. Thus, the young person pays high installments, with interest and for a long period. Meanwhile, individual gives up on investing in himself and his professional career. In other words, he is totally without financial conditions to do a postgraduate course, miss activities that provide unique and enriching experiences, such as shows, tours and trips. Many will say the opposite, but rent can be an ally of flexibility. Even most of the customers are by opting for a smaller, cheaper place and saving the rest to invest in the best way.

Huang et al. (2019) argue that the biggest transformation in the society is the increase in the status of children in the household. Based on new dynamics that have been established in the family, creating a new power relationship, in which decisions are not made in isolation by the parents, can establish a process developed in a bidirectional way (anchored on the influence between parents

and children). Drastic changes in society and family structure, such as urbanization, the use of contraceptive methods and the greater presence of women in the labor market, have made motherhood a choice, not a chance. In this way, families are smaller and more planned and the children in these families are more sovereign, influencing and deciding what to eat, wear, among others, even for the adults in the house.

Breiman (2001) indicates lifetime has a direct influence on the value of the installments and even on the period in which the buyer can dilute a debt. In a financing, the installments are composed of interest, fees and also mandatory insurance, such as Death or Permanent Disability. This insurance tends to increase in value the older the buyer is, because it is understood that over the years, he presents more risks of getting sick, becoming disabled or dying. In this way, individual could not afford the debt and the insurance would have to be activated. Therefore, policies are more expensive according to age. A buyer over 50 years of age can pay 18% more than a younger buyer in their 20s on a 30-year loan. The important thing is that individual have at least 25% of the value of the property saved to give the 20% down payment and pay the fees. The younger one can get to that level, the lower the interest they pay. So, the age ideal to buy is a balance between youth and this minimum of 25% of the property. He also states that in a simulation one found that insurance became 600% more expensive for the 50-year-old buyer compared to the 20-year-old.

According to Syam and Sharma (2018), another limiting factor is the rule that requires the financing to be paid up to the maximum age of 80 years and six months of the person who buys the property, precisely because of the MIP. It is also a way for banks (creditors) to reduce the risk that this borrower may not be able to pay off everything until the end of his life or until he loses the ability to pay his bills. The age limit to finance, in fact, is related to the settlement date, which has to happen until the proponent turns 80 years and 6 months old. Therefore, if a person is 60 years old, he will only be able to finance the property for a period of 20 years and six months, so that the transaction is completed when he is 80 years and six months old. A longer term is not available. If a person wants to take out a 30-year loan, he or she must be at most 50 years and six months old at the time of applying for credit from the financial institution. Otherwise, the term will have to be shortened, which ends up making the installments even more expensive, because the debt dilution will be lower.

Finally, Martínez et al. (2020) argue that purchase of tangible assets requires higher level of income. Even for taking loans, financial institutions want a guarantee that the contractor will fulfill the contract signed. Therefore, he must prove his ability to pay for future installments. That is why the older insured person has to prove greater earnings, even in the case of equal properties compared to a younger person, since housing insurance impacts the value of the installments. In addition, as the installment term is reduced, the installment may be higher, requiring a higher minimum income. The conditions and formats for analyzing and approving mortgage loans vary from bank to bank. According to analysts, financial institutions observe the characteristics of the proponents in order to reduce the risk of default or default, even more so when it comes to long-term credit. Therefore, in addition to age, there are other factors that are observed such as profession, employment relationship and the so-called customer score in the market that is, whether he is a good payer or is in debt. Even with variations, the objective is to check the buyer's ability to pay. A lot of people think that the criteria are fixed for everyone. Income commitments and other financial commitments are checked, such as alimony, car consortium, loans and debts, credit card expenses and payment history.

Softwares (2014) highlights that by consuming tailored products and services, the customer will feel understood and satisfied with the quality of delivery, which will meet their expectations and solve problems that impact their daily lives. It has been found from the literature that people often have specific needs but aren't sure what will help them. Knowing what negatively impacts a potential client's life can be a golden opportunity to develop creative solutions. Moreover, any company can be more likely to grow and be highly profitable if it is attentive to purchasing behavior, and this advantage does not stop there. Financial resources can also be directed to what will really bring results to the business, which eliminates monetary, operational, and strategic waste. One of the important things to remember when learning more about consumer behavior is that your customer's buying habits change. They will change if a pandemic occurs, for example. After all, this difficult time changes practically everything that is happening in the world. This concern is real, obviously thinking about the future of the planet. But beyond that, for companies, it's something that consumers expect.

Chapter 3: Project Description

3.1 Data Preprocessing

Data preprocessing is a method of transforming original data into a presentable data collection. In other words, if data is collected from numerous sources, it is gathered in original form, making analysis impossible.

3.2 Data Cleaning

Data cleaning or cleansing methods aim to fill in incomplete data, smooth out distortion while detecting anomalies, and fix discrepancies. The approach I used was to ignore the tuples that were unnecessary for our model and remove the duplicating values.

3.3 Data Visualization

Data visualization is a crucial step in analyzing and creating a model. Visuals and graphs assist us to grasp the data so that we can readily choose features.

3.4 Data Modeling

Data modelling is a collection of tools and strategies used to understand and assess how a company should gather, update, and retain data. It is necessary for the business analyst responsible for detecting, evaluating, and defining changes to how software systems generate and preserve information. Data categorization is an essential aspect of data modelling.

3.5 Data Classification

The process of classifying and categorizing data into numerous types, forms, or any other separate class is known as data classification. Data classification allows for the separation and variety of data based on data set needs for various corporate or personal goals. It is primarily a data management procedure. For data classification, several techniques and classifiers are available.

3.6 Splitting Data

We divided our data into two portions to train and test our model while picking features. It may be 70 per cent and 30 per cent, or 80 per cent and 20%. The smaller one is for testing, while the larger one is for training.

3.7 Features Selection

There are several columns in the data. Some of those columns are extra to our model and forecast. To achieve high accuracy and efficiency in our model, we choose some characteristics and columns that will make a difference and boost accuracy and precision in our model.

3.8 Data Sources

We will use the internet as an external data source. We need to collect a considerable amount of data about existing customers, such as demographic information and information about purchases they have made. Furthermore, the study should anticipate a client's typical monthly spend based on known consumer attributes. As a result, we will evaluate any of the following data sources:

- https://www.kaggle.com/heeraldedhia/bike-buyers
- https://www.kaggle.com/rahulsah06/bike-buying-prediction-for-adventure-works-cycles
- https://www.statista.com/statistics/1222828/england-electric-bicycle-buying-intention/
- https://www.statista.com/statistics/1196079/italy-likelihood-to-buy-an-e-bike-by-frequency-of-use/

3.9 Data Processing

3.9.1 Overview

The many machine learning approaches are used in this research to predict whether or not a consumer would buy a product. This technique will boost the chances of purchasing things based on accuracy and considering a dataset that is classified so that the dataset falls within the classification model. On the same dataset, apply all classification algorithms and compare accuracy values from all algorithms to identify the best-suited approach for predicting potential customers that provide the most outstanding performance.

Some of the classification algorithms that will be used in our project are as follows:

- Decision trees.
- Random Forest Classifier.
- Support Vector Machine.

3.9.2 Random Forest Classifier

Tree models are recognized to have a significant variance and a low bias. They are prone to overfitting the training data in the following step. If we summarize how we do not prune, this is unforgettable. There are no more attributes on which to spit the dataset because of this uncertainty, a minor change in the dataset's composition results in a different tree model.

3.9.2.1 Advantages

- Multiple trees reduce the likelihood of encountering a classifier that does not perform well due to the connection between the train and test data.
- Averaging numerous trees reduces the danger of overfitting marginally.

3.9.2.2 Disadvantages

- It takes longer to train samples.

3.9.3 Support Vector Machine

The goals of SVM are to find an ideal hyperplane that divides data into different groups by employing a model known as the kernel. As a result, it may quickly move through while maintaining the most significant feasible gap between itself and these data points.

3.9.3.1 Advantages

- Works well with multidimensional data. Works effectively with unstructured and semi-structured data as well. The kernel technique is the SVM's strength. To solve the complex problem, use the kernel function.
- The SVM model has been implemented in practice; the danger of overfitting is lower in SVM.

3.9.3.2 Disadvantages

- Choosing a decent kernel function is complex, and training massive datasets takes along.
- The final result, variable weights, and effect are challenging to understand and communicate.

3.9.4 Decision Tree Classifier

The decision tree should be used to physically and concisely represent decisions and decision making. It makes use of a decision-tree model, even though it is a widely utilized approach in data mining for building a plan to attain a specific goal.

3.9.4.1 Advantages

- Capable of dealing with category and numerical data.
- Easy to grasp, display, and interpret.

3.9.4.2 Disadvantages

- Complex Decision Trees do not generalize well to data, leading to overfitting.

The cross-industry standard data mining process creates and assesses the suggested model. The steps of this procedure are as follows:

1. Cleaning of data (to eliminate noise and inefficiencies in data)
2. Integration of data (When several data sources may be integrated)
3. Data collection (when data from the database relevant to the analysis job is gathered)
4. Data conversion (where data is processed and condensed into mining-ready formats through brief or aggregation procedures)
5. 5th data mining (an essential method that use an intelligent approach to extract patterns from data).
6. Pattern analysis (based on interestingness criteria, locate the genuinely fascinating ways of expressing knowledge)
7. Knowledge demonstration (where visualization and conceptual modelling methods are utilized to show users with mined information)

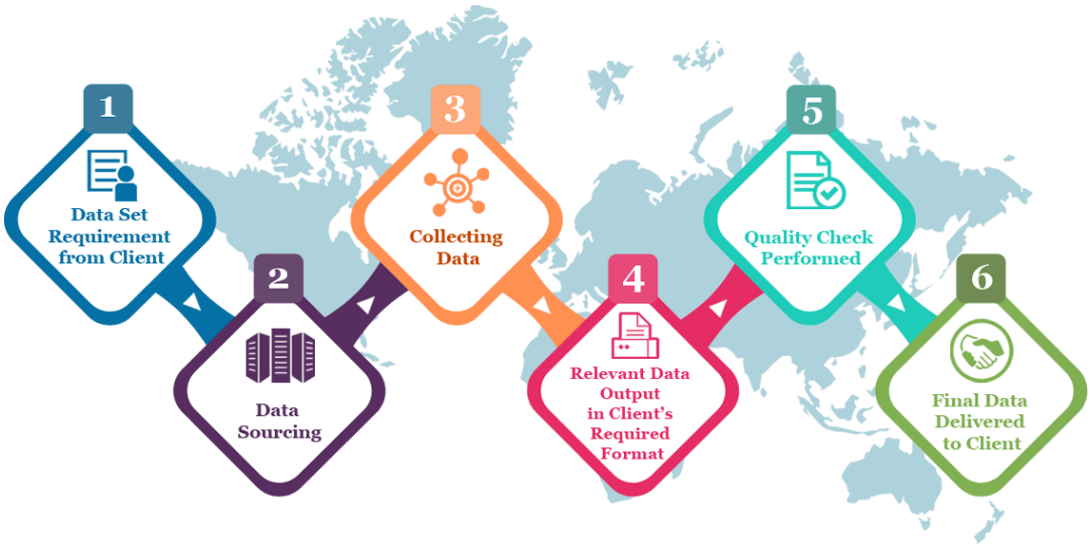

Figure 1: Data Mining Diagram

3.9.5 Algorithm Used

Algorithm	
Algo 1 – TfIdf	This is used for retrieval of information for identifying short term frequency. In an inverse document frequency this represents numerical statistics that is used for reflecting importance of word for documentation in corpus collection.

Algo 2 - Neural network	This include a series of algorithm that used for identifying the relationship in data set through specific process which represents mimics about working of human brain.
Accuracy	This is used for problem classification by presenting accurate information.

3.9.6 Accuracy

Accuracy is used for classifying the issues in order to identify accurate percentage for predicting behavior. In this regard, the research calculates accuracy by dividing correct prediction with total number of predictions in a data.

3.9.7 Confusion matrix

This matrix is used for identifying performance measurement in classification of machine learning. In this aspect, the researcher used this to predict buying behavior. This table support for identifying performance of classification model on specific set of data in which true values in data set are known.

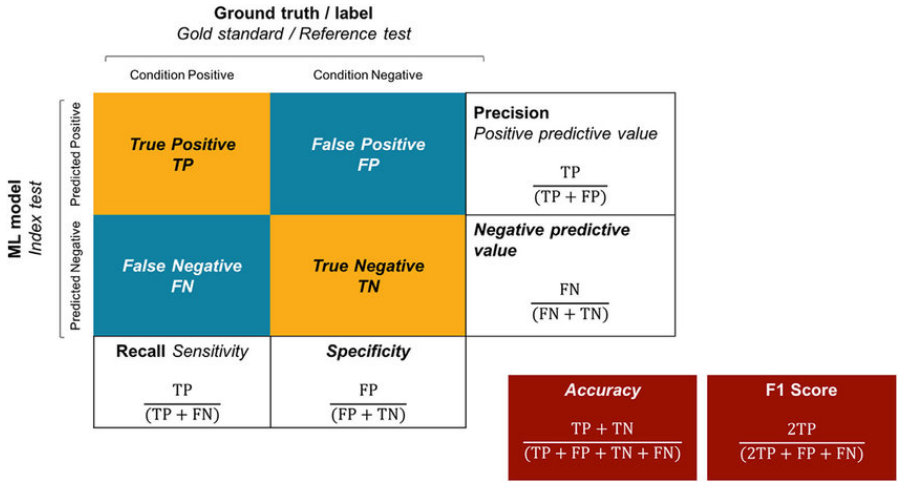

Figure 2: Confusion Matrix

Chapter 4: Data Analysis and Results

The data established that consumer purchase behavior dictates how customer chooses, buys, uses and discards products. In other words, it established the habits and customs that determine consumption needs and desires. Understanding this path is critical to strategizing the sales process and influencing buyer decisions. It is necessary to evaluate a series of factors and variables that can intervene in the customer's purchase process. It may seem like difficult task, but the analysis of the items described can help to identify consumer buying behavior. There are several ways to do this: in this research secondary data analysis has been carried out for analysis of purchase data. Regardless of the size of business, all companies can benefit from this information.

It is evident that creating a predictive system for potential buyers require customized ML approaches to predict whether or not a consumer would buy a commodity. The technique improves purchasing chances based on accuracy and considering a dataset that is classified so that the dataset falls within the classification model. On the same dataset, classification algorithms and compare accuracy values from all algorithms to identify the best-suited approach for predicting potential customers that provide the most outstanding performance. The applicable models in this context can be Decision trees, Random Forest Classifiers, or Support Vector Machines, where tree models are recognized to have a significant variance and a low bias. In concurrence with some information obtained in the literature review, there are no more attributes on which to spit the dataset because of uncertainty, a minor change in the dataset's composition can result in a different tree model. Although the Random Forest Classifiers takes longer to train samples, its multiple trees minimize the likelihood of encountering a classifier that does not perform well due to the connection between the train and test data. It also reduces the danger of overfitting marginally. On the other hand, the Support Vector Machine finds ideal hyperplane that divides data into different groups by employing a model known as the kernel. Resultantly, it quickly moves through and maintains the most significant feasible gap between itself and these data points.

Data has established that ML can improve the relationship with consumers. Artificial Intelligence, Machine Learning, and Data Science to maximize their operations in many ways. Artificial Intelligence (AI) technology allows computers to make decisions and interpret

data automatically, based on algorithms. It is not necessary to program them to perform specific actions. An algorithm is a finite sequence of actions and rules aimed at solving a problem. Each of them triggers a different type of operation when making contact with the data the computer receives. The result of all operations is what makes ML possible. In this way, machines improve the tasks performed by processing data such as images and numbers. That's why machine learning depends on Big Data to be effective. Big Data is the strategic gathering and analysis of a large volume of data. In this article, you can delve deeper into the concept and understand how Big Data works. The computer can learn in different ways, depending on the data it receives which is helpful for predicting consumer behavior.

To accurately predict the purchasing behavior of the customers who purchased scooters in the past, the technology related to machine learning analyzed and predicted different characteristics of customer purchasing behavior. Therefore, the root node seems to have no parental node, whereas another node seems to have one parental node. While logistic regression, decision tree, neural network, support vector machine, Bayesian network & other methods are some methods in machine learning. The process is based on defining the classification of the data set that allows training for future observations. In academia, there have been several different machine-learning methods analyzed in the past decade to predict profitability and customer retention, which practitioners have been known to have used. Internal nodes defined differences between the attributes or related features where the leaf nodes present different categories in the data. It is based on testing the data set separated from the definite training set. The training sessions define the data set as consisting of the pairs within the input object. The consideration of the decision tree is based on the node and defining a directed edge.

Prediction Algorithms use decision trees and acyclic tree are tricky to manage due to unbalanced dataset considering Random Forest methods in training. It helped provide trained sessions to the selected actions and manage the random subset for the attribute of the whole set of the predictor attribute. Therefore, the process helps provide the average or weighted average method for implementing the final Decision Tree. Customer cross- and up-selling research comprises market basket analysis, forecasting clients' shopping lists, and estimating future purchases. Time series decomposition, exponential smoothing, linear regression, moving average, gray theory, and regression are some of the data mining tools. The conducted research involved

the creation of a tailored discount offer system. The differentiation of the attributes and features is seen in the leaf nodes considering different classifications. The integration defines the accuracy of the classifier where the values are used for defining possible classification errors. Market basket analysis's central purpose is to understand which things customers buy at the same time to improve marketing. It can also be handled with the missing values and considering the dataset for the training sessions of the model. The objective defines efficient learning towards the classification.

From the study, it is evident that decomposition, exponential smoothing, linear regression, moving average, gray theory and a regression are some of the data mining tools. While logistic regression, decision tree, neural network, support vector machine, Bayesian network & other methods are some of the methods in machine-learning. Data mining mines the previously unknown, practical and effective information, making it a complete process. In this regard, it is evident that algorithms use decision trees that are definite and known as the acyclic tree. The structures are based on classifying the instances. The consideration of the decision tree is based on the node and defining a directed edge. The node is associated with the internal and the leaf nodes. Internal nodes define differences between the attributes or related features where the leaf nodes present different categories in the data. The differentiation of the attributes and the features are seen in the leaf nodes as considering different classifications. The data indicates that the root node seems as having no parental node where another nodes seems having one parental node.

Finally, the ML algorithm that worked the best is the Random Forest Classifier because it can be divided and conquered into specific data sets. It also provides trained sessions to the selected actions and managing the random subset for the attribute of the whole set of the predictor attribute. In addition, the tree increases with maximum implementation of the data sets. It is based on defining the attribute while presenting in the given subset. Therefore, the Random Forest Classifier ML helps in providing the weighted average method for implementing the final Decision Tree. It will help in providing the constructed methods for the prediction within the testing of dataset. Random forest can also be efficiently used in defining the larger dataset. It can be handled easily for thousands of input variables without managing the variable deletion. It can also be handled with the missing values and considering the dataset for the training sessions of the model.

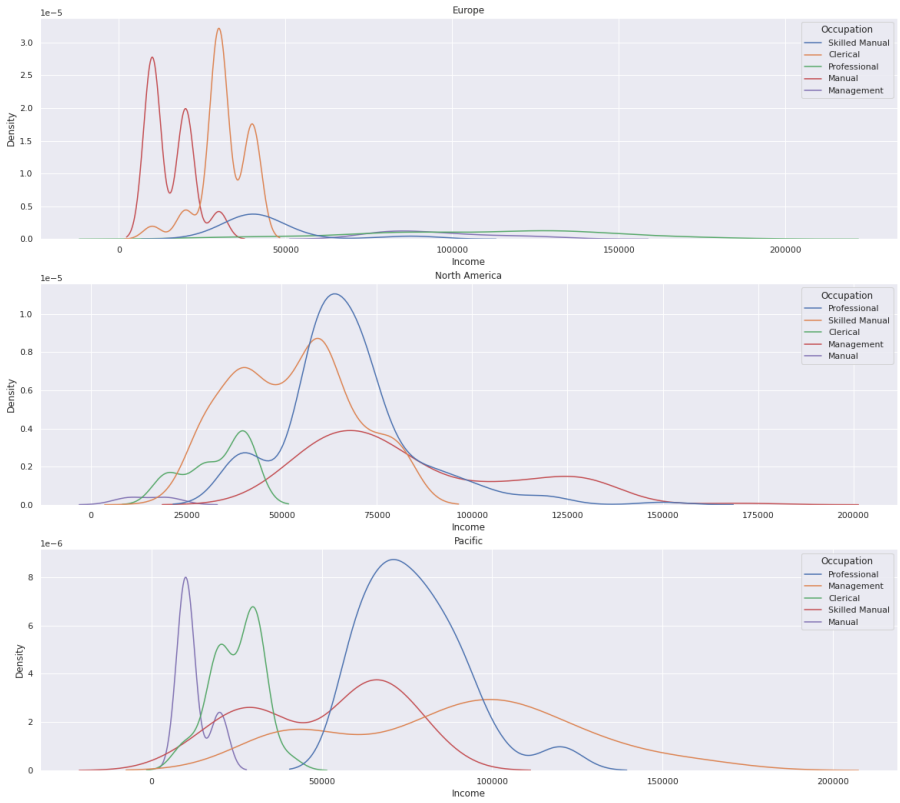

Figure 3: Income level in different regions

The income level has direct influence on consumer purchase behavior. above figure is presenting density plot of profession against their income with respect to different regions. It has been found that there were significant Changes in habits and behavior patterns caused by the coronavirus. It has been observed that consumer reduced consumption during the period of social isolation due to the impact on income and fear of unemployment. For the post-pandemic period, mostly individual (percentages between 50 and 72%) specified they plan to maintain the level of consumption adopted during social isolation, indicating that consumer behavior in the future is that they will not be willing to resume the same level of purchases before the pandemic.

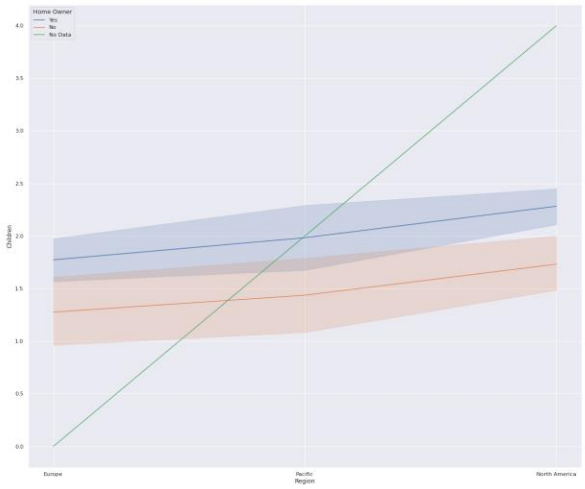

Figure 4: Homeowners and number of children

The above graph is presenting homeowners having number of children according to their region. There is today a new family model, the biggest transformation of which is the increase in the status of children in the household, based on new dynamics that have been established in the family, creating a new power relationship, in which decisions are not made in isolation by the parents, the process being developed in a bidirectional relationship, with mutual influence between parents and children.

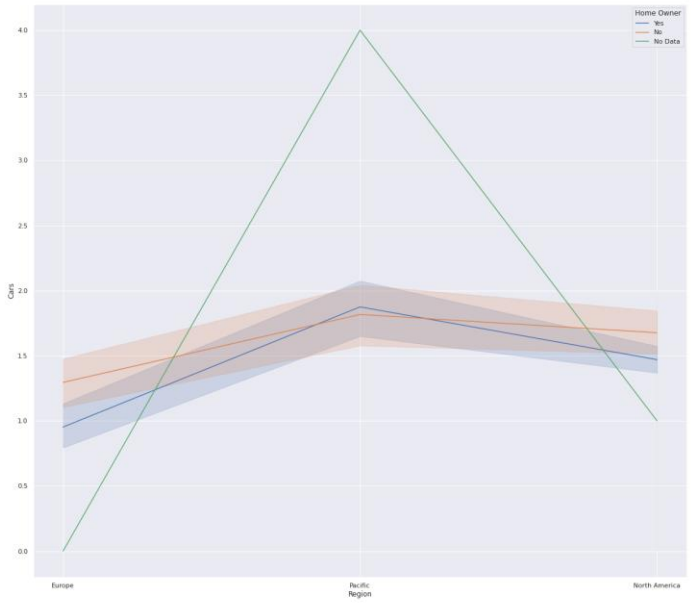

Figure 5: Homeowners and Number of cars

From the analysis of customers' behavior for detecting customer behavior through artificial intelligence it has been found that line plot describes homeowners having number of cars according to their region. In several sectors, the hiring of professionals with good salary averages has grown in the beginning of 2018. The scenario is a good indicator to evaluate the economic recovery of the country after some years of retraction.

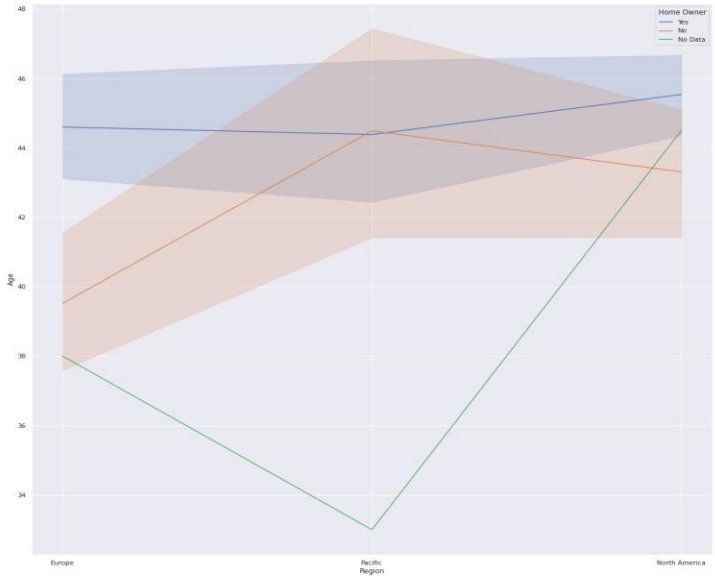

Figure 6: Ages of Homeowners

In time of crisis, differentiating products and services is the best strategy to conquer a prominent space in the market. The above figure describes homeowners' ages according to their region. Experts are unanimous in saying that investing in a property too early, through financing, can be a choice with a not very good impact on the young person's life. That's because these individuals are still at the beginning of their career and, in most cases, with not enough money to make a good down payment. Thus, the young individual pays high installments, with interest and for a long period.

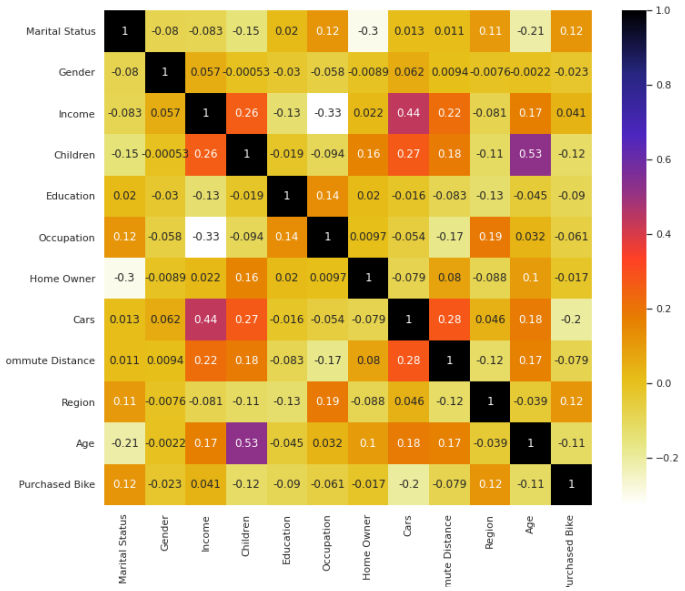

Figure 7: Correlation of features

Correlation heatmap showing all the features and their association, as the consumer purchase behavior influenced by several factors for the purchase of different products. Local customs interfere with the way the customer buys. Each region has different habits, unique experiences, different religions, rituals, norms, traditions, musical preferences and so on. Understanding the culture of the area can help in the development of products and services.

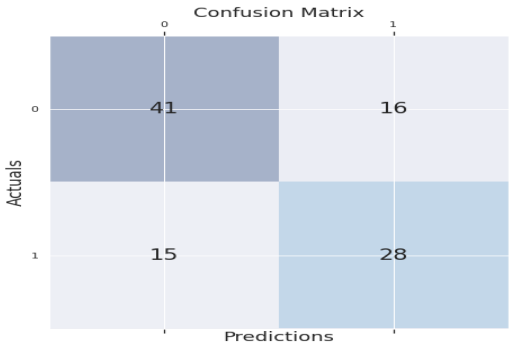

Figure 8: Confusion matrix showing the output predicted results

Here we can find the confusion matrix. The true positive was 41 and the true negative was 28. While the false positive was 16 and the false negative was 15. According to those numbers, the accuracy is 69%.

Chapter 5: Conclusion

5.1 Conclusion

From the above analysis, it has been identified that there are several factors which influence consumer decision-making during purchases. In this regard, ML offers valuable insights to ensure the best customer experience. Implementing ML in customer relationships has several benefits, as it provides greater user convenience; it helps if the customer wants to find the solution to the problem quickly, preferably on their smartphone. The second contact channel they look for is chat, according to the same Aspect report, hence the importance of having a well-built Chabot strategy. Moreover, this provides greater predictability: machine learning combined with big data allows predictive tests to be carried out and, thus, avoid customer attrition situations. Data analysis to predict an overload on the energy and internet network or an increase in demand for a certain product is some examples. In relation to this, this contributed for understanding behavior: the data allows you to know the best offer and the best time to send it to a specific customer. They also help to identify new demands and guide the development of products to meet them. Furthermore, this provides problem Solving autonomy: Aspect's Consumer Index Report 2020 pointed out that, if possible, the customer would like to solve their problems alone, without depending on a human agent. Automation and machine learning will help you to have more autonomy, in an agile and well-oriented way. From the analysis of the customer's purchase history, it is possible to automate the relationship with relevant content and offers. It is conclusive to state that machine learning allows the analysis of consumer behavior when using a certain product. Thus, it is possible to predict what other services may be offered to it.

5.2 Recommendations

To explore improved ways for predicting when a client will purchase a product, modern technologies in the field of predictive analytics should be utilized, with more research done on specific markets or products in the marketplace. In this regard, researchers should use various statistical techniques, such as “data mining, predictive modeling, and machine learning” anchored in Internet of Things (IoT) to model and analyze historical and current data/facts to predict future consumer behaviors. In this regard, retailers and manufacturers will be able to utilize this information to identify specific target markets/customers to obtain a competitive advantage. Marketing managers may decide how to sell to a person if they know when they require specific items. Future researchers should shift all the marketing strategies from a commodity based to a client-focused approach for specific address of customer needs.

References

1 Baderiya, M.S.H. and Chawan, P.M., (2018). Customer buying Prediction Using Machine-Learning

2 Cumby, C., Fano, A., Ghani, R. & Krema, M. (2004). *Predicting customer shopping lists from point-of-sale purchase data.* KDD '04 Seattle, Washington, USA.

3 Cumby, C., Fano, A., Ghani, R., & Krema, M. (2005). Building intelligent shopping assistants using individual consumer models. *Proceedings of the 10th International Conference on Intelligent User Interfaces*. https://doi.org/10.1145/1040830.1040915

4 Els, Z. (2019, April 1). Development of a data analytics-driven information system for instant, temporary personalised discount offers. Retrieved March 15, 2022, from scholar.sun.ac.za website: http://scholar.sun.ac.za/handle/10019.1/106194

5 Hosseini, M., & Shabani, M. (2015). New approach to customer segmentation based on changes in customer value. *Journal of Marketing Analytics*, *3*(3), 110–121. https://doi.org/10.1057/jma.2015.10

6 Huang, C., Wu, X., Zhang, X., Zhang, C., Zhao, J., Yin, D., & Chawla, N. V. (2019). Online Purchase Prediction via Multi-Scale Modeling of Behavior Dynamics. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. https://doi.org/10.1145/3292500.3330790

7 Jiawei, H., Micheline, K., & Jian, P. (2016). Data Mining Concepts and Techniques. *188.242*. https://doi.org/http://hdl.handle.net/123456789/5809

8 Kim, G., Chae, B. K., & Olson, D. L. (2012). A support vector machine (SVM) approach to imbalanced datasets of customer responses: comparison with other customer response models. *Service Business*, *7*(1), 167–182. https://doi.org/10.1007/s11628-012-0147-9

9 Linoff, G. S., & Berry, M. J. A. (2011). Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. In *Google Books*. Retrieved from https://books.google.com/books?hl=en&lr=&id=AyQfVTDJypUC&oi=fnd&pg=PR37&dq=Data+mining+techniques:+For+marketing

10 Martínez, A., Schmuck, C., Pereverzyev, S., Pirker, C., & Haltmeier, M. (2020). A machine learning framework for customer purchase prediction in the non-contractual setting.

European Journal of Operational Research, *281*(3), 588–596. https://doi.org/10.1016/j.ejor.2018.04.034

11 Neves, A. C., Leander, J., Gonzálelz, I., & Karoumi, R. (2019). An approach to decision-making analysis for implementation of structural health monitoring in bridges. *Structural Control and Health Monitoring*, *26*(6), e2352. https://doi.org/10.1002/stc.2352

12 Raorane A.A., Kulkarni, R.V. & Jitkar, B.D. (2012). Association rule - Extracting knowledge using market basket analysis. *Research Journal of Recent Sciences,* 1(2), 19-27.

13 RESEARCH OF CREDIT RISK OF COMMERCIAL BANK PERSONAL LOAN BASED ON ASSOCIATION RULE. (2011). *Proceedings of the 13th International Conference on Enterprise Information Systems*. https://doi.org/10.5220/0003413101290134

14 Softwares, L. (2014, October 22). Data Mining and Its Importance. Retrieved from Loginworks Softwares Pvt. Ltd. website: https://www.loginworks.com/blogs/217-data-mining-and-its-importance/

15 Khan, W., Ghazanfar, M. A., Azam, M. A., Karami, A., Alyoubi, K. H., & Alfakeeh, A. S. (2020). Stock market prediction using machine learning classifiers and social media, news. *Journal of Ambient Intelligence and Humanized Computing*, 1-24.

16 Lo-Ciganic, W. H., Huang, J. L., Zhang, H. H., Weiss, J. C., Wu, Y., Kwoh, C. K., ... & Gellad, W. F. (2019). Evaluation of machine-learning algorithms for predicting opioid overdose risk among medicare beneficiaries with opioid prescriptions. *JAMA network open*, *2*(3), e190968-e190968.

17 Xu, X., Wang, J., Peng, H., & Wu, R. (2019). Prediction of academic performance associated with internet usage behaviors using machine learning algorithms. *Computers in Human Behavior*, *98*, 166-173.

18 Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. *BMC medical informatics and decision making*, *19*(1), 1-16.

19 Syam, N., & Sharma, A. (2018). Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice. *Industrial marketing management*, *69*, 135-146.

20 Brei, V. A. (2020). Machine learning in marketing: Overview, learning strategies, applications, and future developments. *Foundations and Trends® in Marketing*, *14*(3), 173-236.