# CUSTOMER CHURN PREDICTION USING LOGISTIC REGRESSION AND DECISION TREE (CART) TECHNIQUES

**NAN EI PHYO HTET**

**M.C.Sc.** **DECEMBER 2022**

# CUSTOMER CHURN PREDICTION USING LOGISTIC REGRESSION AND DECISION TREE (CART) TECHNIQUES

**By**

**NAN EI PHYO HTET**

**B.C.Sc.**

**A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of**

**Master of Computer Science**

**(M.C.Sc.)**

**University of Computer Studies, Yangon**

**DECEMBER 2022**

# ACKNOWLEDGEMENTS

I would like to take this opportunity to express my sincere thanks to those who helped me with various aspects of conducting research and writing this thesis. To complete this thesis, many things are needed like my hard work as well as the supporting of many people.

First and foremost, I would like to special thanks to **Dr. Mie Mie Khin**, the Rector of the University of Computer Studies, Yangon who gave me the opportunity to develop this thesis for her general guidance during the period of study.

Secondly, I would like to express my appreciation to **Dr. Si Si Mar Win** and **Dr. Tin Zar Thaw**, Professors, Faculty of Computer Science for their superior suggestions, administrative supports and encouragement during my academic study at the University of Computer Studies, Yangon.

I would like to express my gratitude and appreciation to my thesis supervisor, **Dr. Sandi Winn Aye**, Lecturer, Department of Information Technology Support and Maintenance at the University of Computer Studies in Yangon, for her insightful counsel, thorough guidance, support, and comments throughout the thesis-writing process.

Additionally, I would like to express my deep appreciation to **Dr. Tin Tin Htar**, Associate Professor, for her kind direction, inspiration, and practical advice throughout the completion of this thesis as well as for her numerous supportive recommendations and good ideas.

I would also like to thank the thesis board examiners for their insightful remarks and edits, which helped me produce a quality product. I would like to thank **Daw Hnin Yee Aung**, Lecturer, Department of English at the University of Computer Studies in Yangon, for all of her help and for proofreading my thesis from a linguistic standpoint.

Last but not least, I want to express my gratitude to my family and parents for their unwavering support and encouragement throughout my life. Finally, I want to express my sincere appreciation to all of my professors, coworkers, and friends for your general advice and important assistance.

# STATEMENT OF ORIGINALITY

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

……………………… ………………………………

Date Nan Ei Phyo Htet

# ABSTRACT

The term *"customer churn"* is used to indicate those customers who are about to leave for a competitor or end their subscription. Customer churn or customer attrition has become an important issue for organizations particularly in subscription-based businesses, where customers have a contractual relationship which must be ended. According to numerous studies, acquiring new clients is significantly more expensive than keeping the ones that already have. As a result, businesses are concentrating on creating precise and trustworthy predictive models to pinpoint potential clients who will churn in the near future. This model uses data from telecom companies on a range of aspects, including customers who left within the last month, services that each client has signed up for, demographic data about customers, and customer account information. The model is presented using machine learning techniques, particularly Logistic Regression (LR) and Decision Tree (DT), to forecast churn for telecoms companies. Comparisons are made to determine the algorithm's efficacy using the provided dataset. The results from a strategy based on Logistic Regression (LR) can predict the telecom market better than Decision Tree (DT) techniques.

# TABLE OF CONTENTS Pages

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF EQUATIONS

# CHAPTER 1

# INTRODUCTION

Customers typically have a wide range of options when selecting a supplier of telecommunications services [5]. They are free to select any service provider they want and to leave their existing one. Customers who choose a different provider than their present one cause the existing provider to lose business and money. Churn is the term for the percentage of customers who leave and stop using the service. Any firm must have a consistent consumer base in order to succeed. Businesses strive to keep customers happy and keep them around for a long time. However, in the actual world, the telecom sector's customer churn rate might reach 25% yearly [1]. Additionally, getting a new customer is 10 times more expensive than keeping an existing one. For business owners, this presents a significant difficulty.

Customers with high levels of loyalty can help businesses increase their core competitiveness and perform better. As a result, many businesses invest a significant amount of money to attract new clients because losing clients is expensive for any company. According to research by Reichheld et al., a company's profits from its current clients increase with the length of their commercial relationship. Customers' net present value will rise by 25% to 95% for every 5% improvement in customer retention rates in the business environment. According to Jones and Sasser's [4] research, an organization's average profit rate will climb by 25% to 85% when its customer turnover rate drops by 5%.

Moreover, recruiting new customers is five to six times more expensive than maintaining existing ones, this is a costly issue. Because of this, the practical significance of the customer churn prediction is that it will help businesses financially. Therefore, identifying the churners can aid businesses in keeping customers, and maintaining relationships with current clients is more crucial [10]. Customer Churn generally means that the customers who are about to move their usage of service to a competing service provider. Different Churn prediction methods gives the prediction about customers who likely to churn in the near future and churn management help to identify such churners and give them some positive offers in order to reduce churn effect.

These customers can be identified using their behavior, customer account information, services that the customer sign up for and demographic details etc. Data can be processed and analyzed using data mining techniques to spot trends and behavioral patterns as well as to improve and optimize business operations to cut down on high-churn consumers [7]. Numerous studies have demonstrated the effectiveness of machine learning algorithms in predicting churning and non-churning events by learning from historical corporate data [5]. All consumer data collected over time is included in the data used in this. In this experiment, Decision Tree (CART) algorithm and Logistic Regression is mainly concentrated on efficient prediction model for customer turnover. The experiment focuses on tree-based and regression-based machine learning methods and algorithms for prediction of churn in telecom industries.

## 1.1 Objectives of the Thesis

The main objectives of the thesis are

- To analyze and compare the performance of Logistic Regression and Decision Tree (CART) Techniques in Telecom Customer Churn Prediction
- To maintain high-value customers, minimize acquisition costs and increase marketing efficiency
- To classify telecommunication customer's data by using data mining and matching learning methods
- To know how Logistic Regression and Decision Tree (CART) algorithms are applied to the mining of the real-world database

## 1.2 Related Works

This section gives an analysis on the various works that have been proposed in the area of churn prediction, stating both their merits and demerits.

In this paper [14], the authors suggested a Random Forest algorithm-based approach for predicting client attrition. They made predictions using data from South Asian telecom companies, and the results were very accurate. However, because they employed the Random Forest algorithm, it takes a lot of time. In addition, a lot of features have been used.

In this paper [6], Logistic Regression and Logit Boost were utilized as two machine-learning algorithms for forecasting customer attrition. An actual database from the American company Orange and the WEKA Machine-learning program were used in the experiment. The outcomes were displayed using various evaluation metrics.

In this paper [3], a useful tool is used for predicting customer churn. This paper adopted two main strategies: the first was to determine the key elements that influence customer churn, and the second was to identify consumers with a high propensity to leave by examining social media.

In the first method, a dataset is created using real-world surveys, and it is then examined utilizing matched learning techniques like Deep Learning, Logistic Regression, and Naive Bayes algorithms. The second strategy involves predicting customer attrition by looking at user-generated content (UGC), such as comments, postings, messages, and product or service evaluations. They used sentiment analysis to determine the text's polarity (positive/negative) when assessing the user-generated content. The findings demonstrate that while the algorithms were equally accurate, they differed in how they arranged the qualities according to their relative importance in the conclusion.

### 1.3 Scope of Thesis

This thesis is divided into five chapters. Chapter 1 of the thesis provides the introduction, goals, and overall framework.

Chapter 2 presents the background theory and literature review for data mining and matching learning approaches.

Chapter 3 describes design of the proposed system. This chapter also discusses system overview, details explanation of the proposed system and methodology of the proposed system.

The system architecture, implementation, experimental results, and system discussions are all discussed in great detail in Chapter 4.

The conclusion, the advantages, the disadvantages, and the potential developments of this system are all covered in Chapter 5 of this thesis.

# CHAPTER 2

# BACKGROUND THEORY

The related background theory with the research work is presented in this chapter. All organizations should strive to predict customer churn because it can learn more about the clients and project future revenue. Additionally, it can assist the company in identifying and strengthening its weak points in customer service. There has been a lot of work done on this and there is still plenty to learn about the customer data in many businesses. Results for data from various industries vary.

## 2.1 Customer Churn Prediction

Customer attrition is the process through which a customer switches from one business service to another. The use of customer churn prediction allows businesses to foresee potential customers who may quit. This stage aids the business in developing the necessary retention policies to draw in the potential churners and then keep them, reducing the financial loss of the business [15].

Many sectors worry about customer churn, but it's particularly bad in those with fierce competition. Losing clients results in financial loss due to lower sales, which increases the necessity for acquiring new consumers [4].

Since obtaining new clients is frequently more expensive than maintaining existing ones, customer retention is essential for a variety of enterprises. Predicting whether a customer will leave a firm is a difficult undertaking because of how unpredictable customers can be. Due to the lack of data in the financial sector compared to other industries, it is even more difficult to identify customer churn. In order to predict churn, lengthier inquiry durations are necessary.

It is commonly acknowledged that client retention has economic worth [8]:

(1) When clients are kept, businesses may concentrate more on meeting the requirements of their current clientele rather than looking for riskier new ones.

(2) Long-term clients would be more advantageous and, if pleased, might recommend new clients.

(3) Long-term clients have a propensity to be less sensitive to market competition.

(4) As a result of the bank's understanding, serving long-term customers costs less.

(5) Sales are dropped when clients leave, and they are increased to bring in new ones.

Customer churn has become a significant issue in all businesses, including the banking sector. As a result, banks have always made an effort to monitor customer interactions in order to identify those clients who are most likely to leave the institution. In order to prevent churn, customer churn modeling primarily focuses on the customers who are most likely to depart [14].

More and more businesses are realizing that their current customer base and their data are their most valuable assets in today's competitive environment. Churn incidence predictions are focused on as part of customer relationship management (CRM). The issue of churn management is crucial for keeping loyal consumers.

Organizations in the business world, including telecom sector, banks, insurance providers, and other service providers, are training their staff to be more customer- and service-oriented and developing plans to guarantee customer retention [12]. Retaining current clients and reducing customer churn is the best fundamental marketing approach for the future [8].

Reactive and proactive targeted techniques to controlling client turnover are known to exist, according to prior research. In a reactive strategy, the business waits until the client requests to end the service. The business makes an effort to identify clients who are most likely to leave. The business then makes an effort to keep such clients by offering incentives. Customers will leave organizations if projections of churn are incorrect, hence precise estimates of customer turnover are necessary [20].

### 2.1.1 Types of Data Generated in the Telecom Industry

Three categories of data can be found in the telecom sector:

**1. Call Detail Data**: This relates to call-related data that is kept in call detail records. A call detail record is created for each call made on a network and contains details about the call. The average call duration, average call origin, call period, and calls to and from different area codes are the main topics covered by call detail statistics.

**2. Network Data**: Real-time status messages and information about error generation are included in the network data. The volume of network messages generated is massive, and technology and data mining techniques are used to identify network flaws by learning from network data [11]. The network data also comprises data on the intricate configuration of equipment, data on error production, and data required for configuring network management.

3. **Data about the consumer**: The data about the client comprises information about them, such as their name, age, address, phone type, kind of subscription plan, payment history, etc.

## 2.2 Machine Learning

Machine learning is a technique for data analysis that aids in the construction of analytical models. It belongs to the field of artificial intelligence (AI). The machine learning algorithms generate decisions with a minimum of human interaction, learn from the data, and recognize broad patterns in it.

Machine learning is typically applied to complex problems or tasks that require a large amount of data. It provides quicker, more accurate findings and is an excellent solution for more complex data. It aids a company in spotting lucrative prospects or any unforeseen threats [16].

Two categories of learning techniques are primarily used in machine learning:

1. Unsupervised Machine Learning
2. Supervised Machine Learning

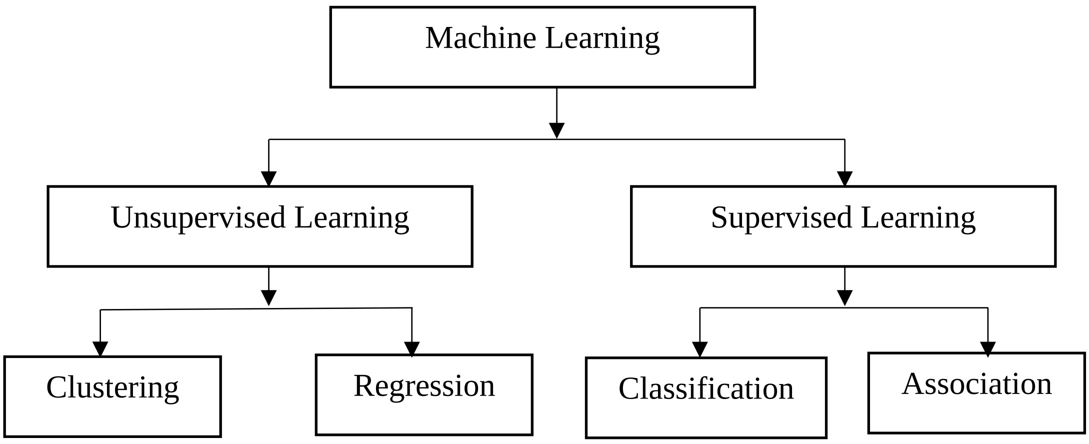


**Figure 2.1 Machine Learning Techniques – Unsupervised and Supervised Learning**

### 2.2.1 Unsupervised Machine Learning

Unsupervised machine learning is a type of machine learning where the users do not have to watch over the model. Instead, it enables the model to operate independently and find previously unnoticed patterns and information. It mostly addresses unlabeled data.

Compared to supervised learning, unsupervised learning algorithms enable users to carry out more complicated processing tasks. Unsupervised learning, however, can be more unpredictable than other types of natural learning. Algorithms for unsupervised learning include neural networks, anomaly detection, and grouping.

Unsupervised learning is the process of using AI algorithms to find patterns in data sets including data points that are neither classed nor labeled. Thus, the algorithms are free to categorize, label, and/or group the data points in the data sets without the need for outside assistance. Unsupervised learning, in other terms, enables the system to recognize patterns on its own in data sets. Unsupervised learning is when a computer program groups unsorted data based on similarities and differences even if no categories are given.

Compared to supervised learning systems, unsupervised learning algorithms are capable of handling more complicated processing tasks. A system can also be tested for AI by undergoing unsupervised learning [13].

Unsupervised learning, however, has the potential to be more unpredictable than supervised learning models. While an unsupervised learning AI system may, for instance, learn how to distinguish between cats and dogs on its own, it may also create unforeseen and undesirable categories to account for uncommon breeds, leading to chaos rather than order.

Although they may also employ a retrieval-based strategy, generative learning models are frequently linked with AI systems that are capable of unsupervised learning (which is most often associated with supervised learning). Robots, self-driving cars, facial recognition software, chatbots, and expert systems are a few examples of systems that may employ both supervised and unsupervised learning techniques [18].

When data sets are fed through training algorithms by machine learning engineers or data scientists, unsupervised learning begins. As it was previously mentioned, the data sets utilized to train such systems do not contain any labels or categories; rather, each piece of information that is fed into the algorithm during training is an unlabeled input object or sample.

With unsupervised learning, the goal is to have the algorithms find patterns in the training data sets and classify the input objects according to the patterns the system discovers. By removing relevant data or attributes, the algorithms examine the underlying structure of the data sets.

Unsupervised machine learning techniques can be used in the following situations:

1. **Clustering**: The dataset can be automatically divided into categories based on similarity using clustering. Cluster analysis, however, frequently exaggerates group similarities and fails to treat individual data points as individuals. Because of this, cluster analysis is not a good option for tasks like consumer segmentation and targeting.
2. **Association mining:** It detects groups of objects in the collection that commonly appear together. Retailers frequently use it for basket analysis since it enables analysts to find products that are frequently bought together and design more efficient marketing and merchandising strategies.

### 2.2.2 Supervised Machine Learning

The computational job of learning correlations between variables in a training dataset and using this knowledge to build a predictive model that can infer annotations for incoming data is known as supervised machine learning [22]. In supervised machine learning, an algorithm is used to learn the mapping from the input to the output. There are an input variable (X) and an output variable (Y).

$$\mathbf{Y = f(X)} \tag{2.1}$$

The objective is to approximate the mapping function so well that the model can forecast the output variable (Y) given new input data (X). When examples with well-known labels are provided, the learning is referred to as supervised learning. The characteristics may be binary, categorical, or continuous [23].

Regression and classification problems are within the category of supervised learning challenges:

**Classification**: Classification issues arise when the output variable is categorical, such as "red" or "blue" and "yes" or "no."

**Regression**: These issues are categorized as regression problems when the output variable has an actual value.

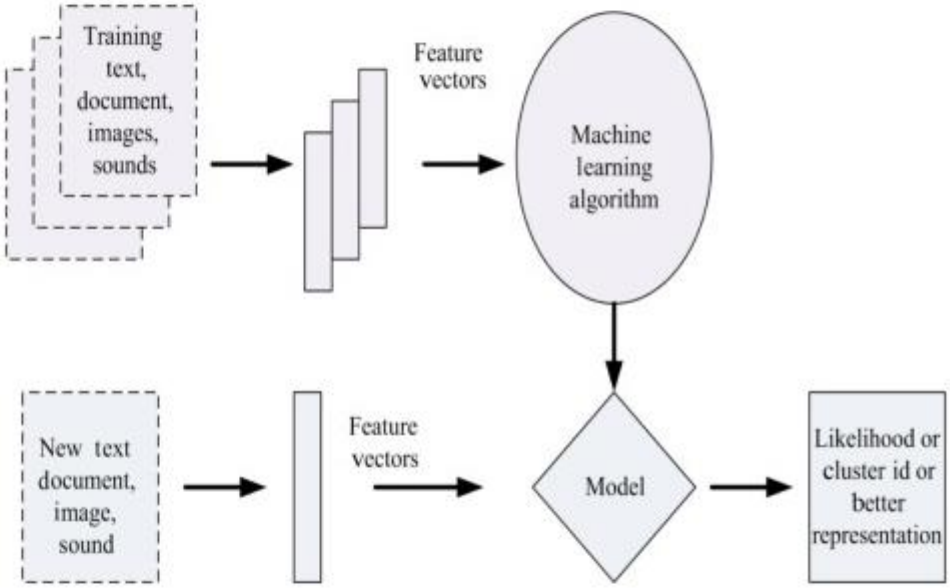


**Figure 2.2 Supervised Machine Learning Model**

## 2.3 Machine Learning Techniques - Classification Algorithms

In the context of data mining and machine learning, classification is the process of locating a model or function that encapsulates a set of data classes in order to use that model to infer the class to which a set of fresh observations belongs. Utilizing training data with a known class label, this is accomplished.

In machine learning, classification is the act of analyzing a new object's features and categorizing it into a set of predetermined classes, according to Berry and Linoff's definition from 2004. Classification may be summed up as a form of supervised learning where models learn from past data and categorize or predict present-day events [21]. Among the most popular classification algorithms are:

### 2.3.1 Random Forest

Random forest and other supervised machine learning algorithms are frequently used in classification and regression problems. It builds decision trees from different samples, using their average for categorization and the majority vote for regression. One of the most important features of the Random Forest Algorithm is its capacity to handle data sets containing both continuous variables, as in regression, and categorical variables, as in classification. It produces better results when it comes to classification problems.

The random forest is created by a large number of decision trees, and the result is the mean prediction for the regression issue and the mode of the class in classification. If the classifier contains a sufficient number of trees, Random Forest will not overfit. It is suitable for categorical variables and can manage missing values. The Random Forest Classification Technique was employed by academics in one of the earlier studies on financial client attrition [9].

### 2.3.2 Support Vector Machine

A supervised machine learning model called the support vector machine can be applied to both classification and regression issues. SVM is frequently employed in classification problems because it may use a hyperplane to distinguish between two classes. Finding a hyperplane that can clearly classify the data is the goal of SVM [4]. Decision boundaries known as hyperplanes assist in categorizing the data points. Support Vectors are data points that are closer to the hyperplane and have an impact on the hyperplane's position and orientation.

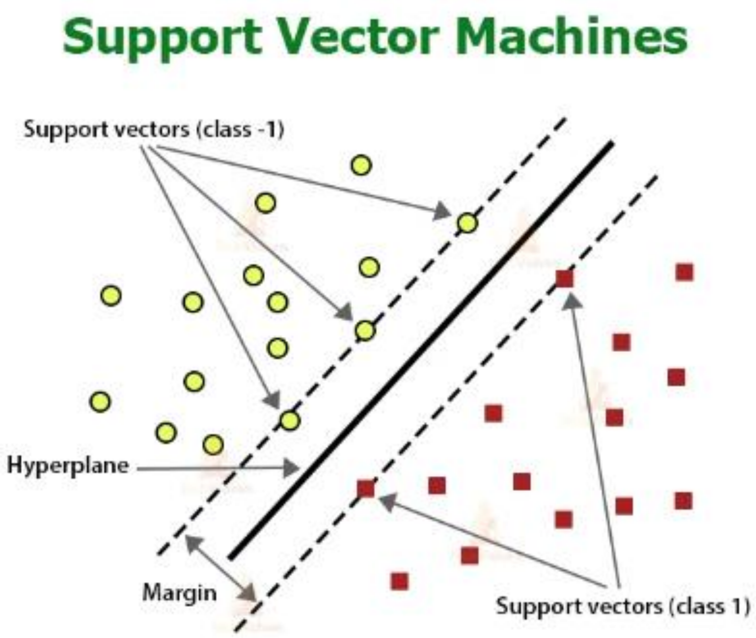


**Figure 2.3 Support Vector Machine**

### 2.3.3 Neural Network

Artificial neural networks, sometimes referred to as "neural networks," are modeled after the biological neural networks that make up the brain. Neural networks are a framework that allows many different machine learning algorithms to cooperate and interpret data inputs, not an algorithm in and of themselves [21].

In order for this to operate, examples must be learned and processed alongside input in order to identify the features of the input and correctly create the output. The neural network can begin processing unknown inputs once the algorithm has gone through a sufficient number of examples, and it will successfully produce the desired outcomes [21].

Understanding neural networks can be improved by looking at an example of how the process of identifying a picture using neural networks operates. The image is divided up into data points and information that a computer may employ using layers of function, according to Deep AI (2018).

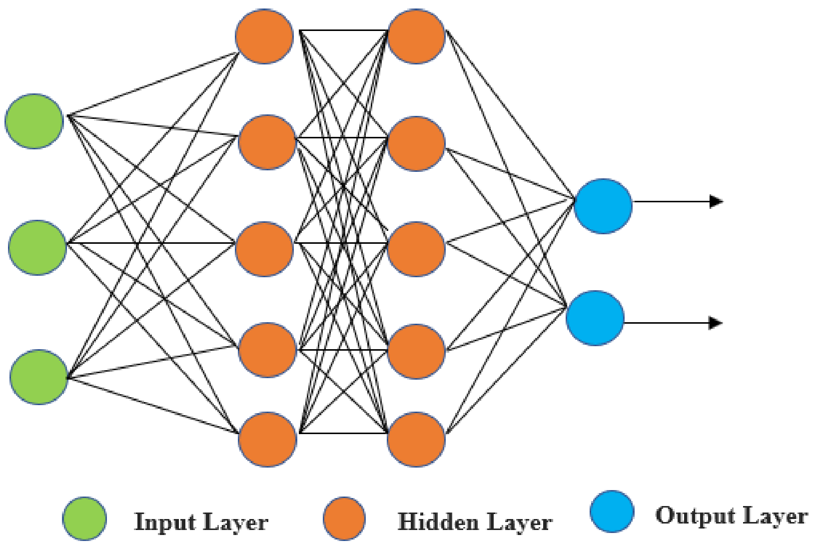


**Figure 2.4 Artificial Neural Networks example**

Following this, "the neural network can start to recognize trends that exist throughout the many, many examples it processes and classify images by their similarity." The algorithm knows what information and components to look for when classifying this particular image again after studying numerous examples of it. Since neural networks learn from experience, it is sometimes said that the more samples the algorithm encounters, the more accurate the results become.

Berry and Linoff (2004) express a similar viewpoint on neural networks and point out that they can learn from examples in a similar way to how people may learn through experience.

According to McDonald (2017), the input layer is the model layer that displays the data in its raw form. Typically, non-linear, complicated interactions between input and output variables are modeled using artificial neural networks. The activation process takes place in a concealed layer. According to Kotu and Deshpande, the activation function utilized in the output function permits a linear transformation for a specific range of values and a non-linear transformation for the remaining values (2014).

### 2.3.4 K-Nearest Neighbors

Lazy learning, a branch of machine learning where generalization of the training data is postponed until a system query, is used in this classification approach (Webb, 2011). A majority vote of the algorithm's neighbors is classified based on the shortest distance between the query and the training set. The training data is set in a multidimensional feature space with a class label, and k is an arbitrary constant that the user defines.

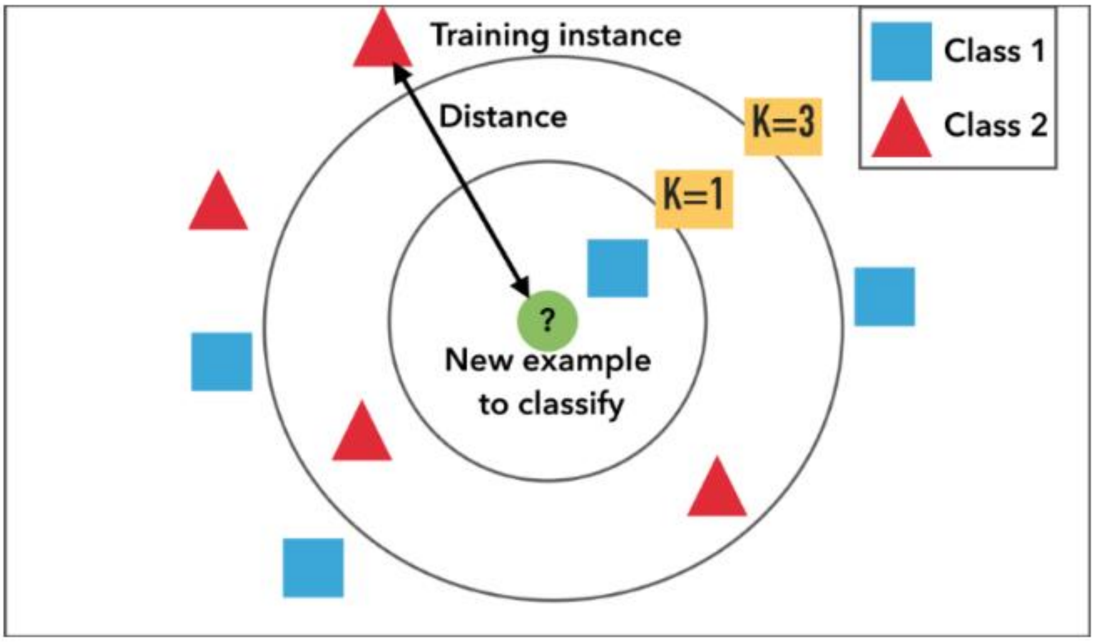


**Figure 2.5: k-nearest neighbor algorithm**

A majority vote of an object's neighbors determines its classification. The most commonly used distance metric is the Euclidean distance. The choice of the parameter k is one of the algorithm's difficulties. According to research carried out by Everitt et al. in 2011, it has been said in research studies that the selection of k should be made on the basis of data and that a large value of k lowers the effect of noise on the data. K-Nearest Neighbor is a straightforward method that works well when there is little to no prior knowledge of the distribution data. However, it also has large memory requirements and exorbitant computational costs.

### 2.3.5 Naïve Bayes

Although this technique is based on the Bayes theorem, it is referred to as "naive" due to the strict assumption of feature independence. Given a class variable y and dependent features $x_1$ through $x_n$, the Bayes' theorem states that:

$$\mathbf{P\ (y \mid x_1, \ldots, x_n) = \frac{P(y)P(x_1,\ldots,x_n \mid y)}{P(x_1,\ldots,x_n)}} \tag{2.2}$$

Using the Naïve condition of independence:

$$\mathbf{P\ (x_i \mid y, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = P\ (x_i \mid y)} \tag{2.3}$$

Because P $(x_1 ,..., x_n)$ is constant due to the independence assumption, the classification rule can be applied:

$$\mathbf{P\ (y \mid x_1, \ldots, x_n)\ \alpha\ P\ (y) \prod_{i=1}^{n} P(\ x_i \mid y)} \tag{2.4}$$

$$\hat{\boldsymbol{y}} = \boldsymbol{arg} \max_{\boldsymbol{y}} \boldsymbol{P(y)} \prod_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{P(\ x_i \mid y)} \tag{2.5}$$

The Naive Bayes classifier nonetheless performs well in spite of this independence requirement, and examples of its applications include document classification and spam filtering. The technique is resilient to outliers and missing data, and it may be fed both categorical and numerical properties.

### 2.3.6 Decision Tree

One of the most often used categorization techniques in data mining and machine learning is the decision tree. As the name implies, a tree-shaped structure is employed in this technique to describe a group of connected decisions or options. The most straightforward explanation of a decision tree is provided by Berry and Linoff (2004), who explain how a huge collection of data is separated into smaller groups of data by applying a set of decision rules. They can either be regression trees, where the target variable consists of a collection of continuous values, or classification trees, where the target variable consists of a set of discrete values. The leaf, which is the output variable and sits at the base of the decision tree, is made up of leaves, branches, and nodes [22].

The decision tree technique can be understood and a number of problems can be resolved by using the example decision tree below, which represents the likelihood that passengers on the Titanic will survive.

The likelihood of a passenger surviving is shown in the figure below as 0.73, with the percentage representing the proportion of observations in the leaf. In summary, there was a probability of 0.73, or a 73% likelihood of survival, if the passenger was female. Male passengers older than 9.5 years had a 17% chance of surviving, whereas those older than 9.5 years who had fewer than 2.5 siblings (2, 1, or 0) had an 89% chance of surviving.

This was a straightforward illustration of how a decision tree functions and how it may be applied to resolve a number of queries on a certain issue. A few decisions tree-specific algorithms exist, some of which were covered in the chapter's prior section. These include C4.5, which replaces ID3, ID3 (Iterative Dichotomizer 3), CHAID (Chi-Square Automatic Interaction Detector), and CART (Classification and Regression Tree).

The split criteria is a crucial measure in decision trees. Information gain, gain ratio, Gini impurity, and variance reduction are frequently employed for this, but only the first two are advised as the most effective techniques in the pertinent literature. Information gain is determined as the information prior to the split minus the information following the split using entropy and information content from information theory. The information value for each node of the tree "represents the expected amount of information that would be needed to define whether a new instance should be classified yes or no, given that the example reached that node," according to Witten, Frank, and Hall (2011).

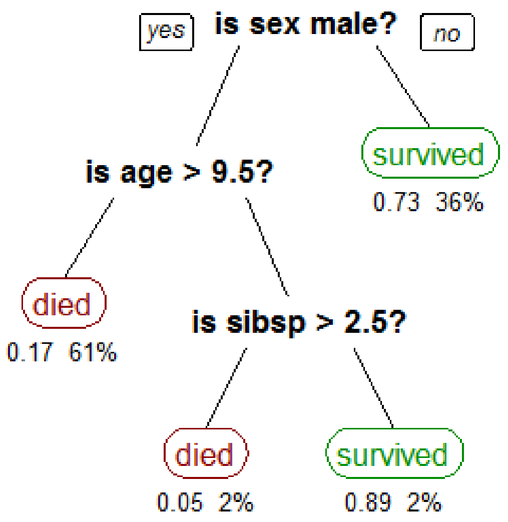


**Figure 2.6 Showing survival probability of passengers on the Titanic ship**

The ratio of information gain to intrinsic information is known as the gain ratio. By taking into account the number of branches when choosing an attribute, it often lessens the bias that information gain has towards multi-valued qualities, according to Witten, Frank, and Hall (2011). A decision tree's ability to be simply comprehended and interpreted by people who are not professionals in data mining and machine learning is one of its main advantages. Additionally, it can handle both numerical and categorical attributes with ease and does not require any normalization of the data. In contrast to k-nearest neighbor, it is computationally economical and works well with huge datasets.

### 2.3.7 Regression Analysis

Regression is seen to be an effective method for determining and forecasting consumer happiness. Using SPSS, the standard error rate is determined for each variable in a regression model. Then a regression model is built using the factors that have the most significance in relation to linear regressions for churn prediction.

Logistic regression approaches are appropriate since the prediction job in churn prognosis is to determine whether a customer is a churner or not, and the prediction characteristic is therefore simply associated with two values. While logistic regression techniques are appropriate for binary qualities, linear regression models are beneficial for predicting features with continuous values.

### 2.4 Data Pre-processing

Data preprocessing is the process of converting unprocessed data into a form that a machine learning model can use. The first and most crucial step in creating a machine learning model is this one. While working on a machine learning project, the data is not always cleaned and prepared. Additionally, data must always be cleaned and formatted before being used in any activity. Therefore, data preparation work is utilized for this.

It is impossible to directly develop machine learning models using real-world data since it frequently contains noise, missing values, and may be in an unfavorable format. The accuracy and efficiency of a machine learning model are increased by data preprocessing, which is required to clean the data and prepare it for the model. Data preprocessing techniques include finding missing value, label encoding, splitting dataset into training and test set and feature selection.

#### 2.4.1 Handling Missing Data

The handling of missing data in the datasets is the next stage of data preprocessing. The machine learning model may have a very difficult time if the dataset has some missing data. As a result, the dataset contains missing values, which must be handled.

The two primary approaches to dealing with missing data are as follows:

**1. By deleting the particular row**: The first method is typically applied to deal with null data. By doing so, the specific row or column that is empty are merely eliminated. This strategy is useless, though, since deleting data may cause information to be lost, which would lead to erroneous output. By deleting the particular row: The first method is typically applied to deal with null data. In this way, the specific row or column that has no data is eliminated. This strategy is useless, though, since deleting data may cause information to be lost, which would lead to erroneous output.

**2. By figuring out the mean**: By doing so, determine the mean of any missing value-containing column or row and replace it with that value. This method works well for characteristics that contain numerical information, such as age, salary, year, etc.

### 2.4.2 Encoding

Using the technique of encoding, categorical variables can be transformed into numerical values and then quickly fitted to a machine learning model. All input and output variables for machine learning models must be numeric. This means that in order to fit and assess a model, categorical data must first be encoded as integers. An ordinal encoding and a one-hot encoding are the two commonly used methods.

Each distinct category value in ordinal encoding is given an integer value. For instance, "red" equals 1, "green" equals 2, and "blue" equals 3. This is reversible and known as an ordinal encoding or integer encoding. Integer values beginning with zero are frequently utilized. A simple ordinal encoding may be sufficient for some variables. Machine learning algorithms could be able to comprehend and take advantage of the inherent ordered relationship between the integer numbers. It is an organic way to encode ordinal variables. It establishes an ordinal relationship where one may not exist for category variables. This may lead to issues; hence, one-hot encoding may be employed in its place.

The integer encoding may not be sufficient for categorical variables for which there is no ordinal relationship, or it may even be deceptive to the model. Allowing the model to presume a natural ordering between categories instead of forcing an ordinal link through an ordinal encoding could lead to subpar performance or unexpected outcomes (predictions halfway between categories). The ordinal representation in this situation can be encoded using a one-hot method. Here, the integer-encoded variable is dropped, and for each distinct integer value in the variable, a new binary variable is added.

### 2.4.3 Splitting dataset into training and test set

The dataset is split into a training set and test set during the preprocessing stage of machine learning. One of the most important data preprocessing procedures, since it can improve the effectiveness of the machine learning model.

#### 2.4.3.1 Training Data

The training dataset, which is used to train or fit the machine learning model, is the largest (in terms of size) subset of the original dataset. In order for the ML algorithms to learn how to make predictions for the given task, training data must first be supplied to them.

Whether supervised learning or unsupervised learning algorithms are used, the training data changes. In unsupervised learning, inputs are not tagged with the appropriate outputs, hence the training data contains unlabeled data points. In order to produce predictions, models must extract patterns from the provided training datasets. In contrast, labels are included in the training data for supervised learning in order to help the model be trained and predictions made.

The model's accuracy and propensity for prediction heavily depend on the kind of training data that are supplied to it. It implies that the model will perform better the higher the quality of the training data. A typical ML project's training data makes up more than or equal to 60% of the total data.

#### 2.4.3.2 Testing Data

It is time to test the model with the test dataset after it has been trained using the training dataset. This dataset assesses the model's performance and guarantees that it can generalize well to new or unexplored datasets. The test dataset is a different subset of the original data from the training dataset. When the model training is finished, it utilizes it as a benchmark because it has some similar features and a similar class probability distribution. A well-organized dataset called test data provides information for each type of scenario the model might encounter in the actual world. The test dataset for an ML project typically makes up 20–25% of the entire original data.

At this point, it is also possible to verify and compare the model's testing and training accuracy, or how accurate it is when applied to different datasets. The model is considered to have overfitted if its accuracy on training data is higher than its accuracy on testing data. The test results need to be a portion of the initial dataset and it ought to be big enough to make accurate predictions.

### 2.4.4 Feature and variable selection

The process of obtaining the most information from a variety of various variables is called feature and variable selection. It is crucial to only incorporate the most important and helpful variables for the model one is constructing because there are more variables and data available as a result of more sophisticated data collection. Improved predictive performance, quicker and more accurate forecasts, and a better and more accurate understanding of the predictive process are the three main goals of selection. Missing crucial variables causes the model to operate less predictably than adding extraneous variables, which might complicate the model or cause it to become overfit [6]. Filter, wrapper, and embedding techniques are three categories within feature selection.

Utilizing predetermined feature relevance criteria, filtering operates. For instance, it might represent the feature's variance. Each feature's variance is calculated, and a threshold variance is established with a variance that is more significant than the threshold and is incorporated into the model. Another common approach is the ranking method, which is founded on the notion that key characteristics are significant if they can be independent of the input data but not independent of the class labels. "The feature can be deleted if it has no impact on the class labels." Simple filter approaches can fail to account for the interdependence of the features, and ranking methods run the risk of producing an unnecessary subset (2014's Chandrashekar and Sahin).

Utilizing algorithms, the wrapper technique examines potential feature subsets in an effort to improve classification performance. Because the difficulty — also known as an NP-hard problem — rises exponentially as the number of features increases, large feature sets can become computationally highly demanding. It is NP-hard, which categorizes it with other well-known computer science issues. A solution to an NP (nondeterministic polynomial (time)) problem can be efficiently confirmed to be accurate, but it is unknown if an effective algorithm exists to locate the solution. The remaining class P issues can be effectively resolved by an algorithm.

Sequential selection algorithms are more straightforward than more difficult optimized algorithms like genetic or particle swarm optimization. The aforementioned techniques run through the characteristics, adding the top classifier to the subset each time (2014's Chandrashekar and Sahin).

The primary objective of embedded approaches is to shorten the computational time required by reclassifying various subsets and integrating feature selection into the training process. The simplest way to comprehend this approach is to include a penalty variable in the model whenever it introduces new variables or bias (2014, Chandrashekar and Sahin). Principal component analysis (PCA), a linear extraction technique that converts the data into a low-dimensional subspace, is another frequently used approach. The goal is to keep the majority of the data while condensing the features into a smaller vector (2016 Li, Wang, and Chen).

### 2.4.5 Exploratory Data Analysis

An essential part of a data analyst or scientist's daily practice is exploratory data analysis (EDA). It allows for a thorough examination of the dataset, the formulation or disapproval of hypotheses, and the development of prediction models with a solid basis. It uses a number of statistical tools and data manipulation techniques to describe and appreciate the relationship between elements and how they can affect a company.

Exploratory data analysis, or EDA, should be the first step in any project involving data analysis or data science. Exploratory data analysis is an essential method for performing early studies on data in order to detect patterns, spot anomalies, test hypotheses, and triple-check presumptions with the help of summary statistics and graphical representations.

Exploratory data analysis (EDA) is the process of examining a dataset to look for patterns and abnormalities (outliers) and creating hypotheses based on understanding of the dataset studies on data in order to detect patterns, spot anomalies, test hypotheses, and triple-check presumptions with the help of summary statistics and graphical representations. EDA comprises creating summary statistics for the numerical data in the dataset and creating different graphical representations to facilitate understanding of the data.

# CHAPTER 3

# DESIGN OF THE PROPOSED SYSTEM

In this chapter, the design of the proposed system and the methodology will be explained in details. The main goal of this thesis is to analyze and compare the performance of Logistic Regression and Decision Tree (CART) Techniques in Telecom Customer Churn Prediction. The experiments for the study are carried out using the Python programming language.

## 3.1 Overview of the Proposed System

An organization spends five to ten times more to keep a single customer than to acquire a new one. In order to provide a retention solution, predictive algorithms can accurately identify potential churners in the near future. This proposed system presents a novel prediction model based on Data Mining (DM) and Machine Learning approaches.

Data preparation, EDA analysis, feature selection and engineering, model building, and model evaluation are the six phases that make up the proposed model. The used data set has 21 attributes and 7044 instances. 1759 examples are utilized as the testing set, and 5275 instances are used to train the model.

The system flow of the proposed system is depicted in Figure 3.1. The system uses 75% of dataset as training data and 25% of dataset as testing data. As obtaining a new customer is more expensive than keeping an existing one, churn management is crucial for business enterprises. In order to avoid this, the business should be aware of the factors that lead a customer to switch to a different telecom provider [6]. Therefore, adopting a churn prediction model can have a significant impact on the company's revenue and profitability.

The main goal is to assess how well two data mining methods—logistic regression and decision trees—perform in predicting telecom customer churn. Getting the data from the IBM Watson Analytics Community is the first and most important stage. The dataset typically includes a wide variety of errors and noisy data. The pre-processing phase has been completed, and the data have been cleansed to make them usable. The output of the data pre-processing is noise-free data that may be used for further processing stages.

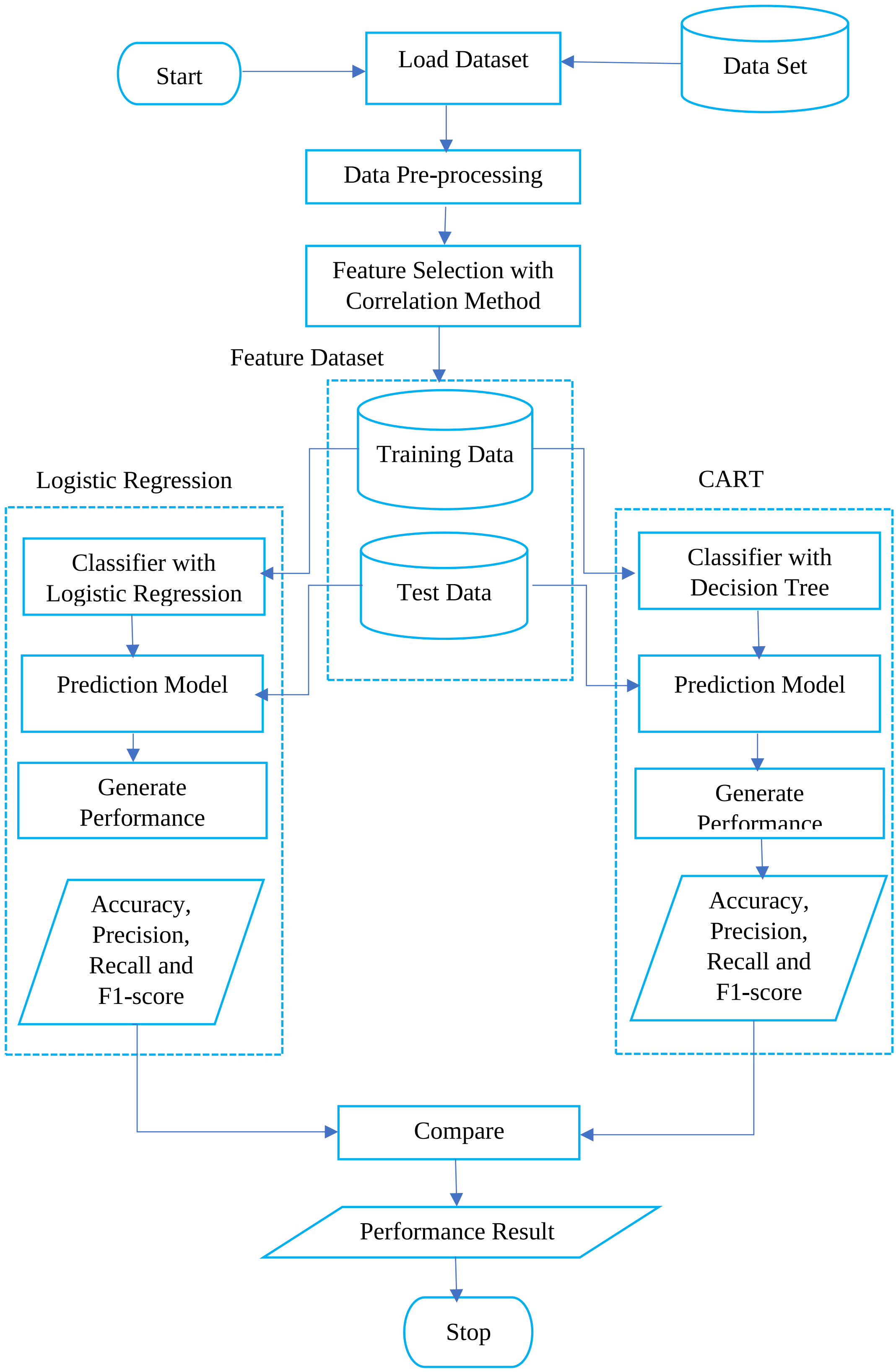


**Figure 3.1Flow Chart of the Proposed System**

### 3.2 Explanation of the Proposed System

The experiment design follows by the EDA Analysis in the proposed system. Python programming is used to carry out the experiments of the proposed system. Data analysis utilizing visual methods is called exploratory data analysis (EDA). With the use of statistical summaries and graphical representations, it is used to identify trends, patterns, or to verify assumptions. The EDA analysis is processed in this proposed system, and each stage is discussed in more details below.

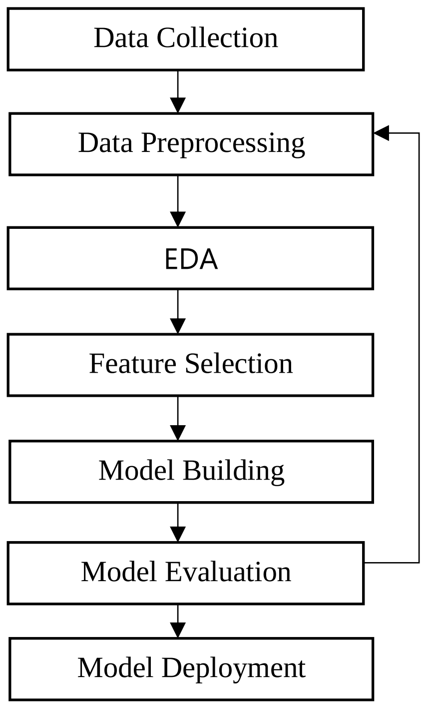


**Figure 3.2 EDA Analysis**

#### 3.2.1 Collection of Data

The proposed system compares the effectiveness of two Machine Learning algorithms, Logistic Regression (LR) and Decision Tree (CART), using data gathered from the IBM Watson Analytics Community. The dataset comprises of 21 columns and 7043 rows of different customer-related data. Using the train-test split technique, 75% of these data will be used for training and 25% for testing.

#### 3.2.2 Dataset Information

The information about the attribute of a telecom subscriber is provided details in the following Table 3.1.

**Table 3.1 Telecom Customer's Attributes and Its Information**

| No | Attribute and Its Information | | |
|---|---|---|---|
| | **Attribute** | **Information** | **Value** |
| 1 | Customer ID | A unique ID that identifies each customer | Customer ID |
| 2 | Gender | Binary attribute: customer gender | Male<br>(or)<br>Female |
| 3 | Senior Citizen | Indicates if the customer is 65 years or not | 1 (or) 0 |
| 4 | Partner | Married or not | Yes (or) No |
| 5 | Dependents | Live with any dependents. Dependents could be children, parents, grand-parents, etc. | Yes (or) No |
| 6 | tenure | Total amount of months that the customer has been with the company | Numeric value |
| 7 | PhoneService | Has home phone service | Yes (or) No |
| 8 | MultipleLines | Has multiple telephone phone service | Yes (or) No |
| 9 | InternetService | Customer's internet service type | DSL, Fiber optic (or) No |
| 10 | OnlineSecurity | Has online security or not | Yes, No, No internet service |
| 11 | OnlineBackup | Has online backup or not | Yes, No, No internet service |
| 12 | DeviceProtection | Has device protection of not for their Internet equipment | Yes, No, No internet service |
| 13 | TechSupport | Has technical support plan to reduce wait | Yes, No, No internet service |
| 14 | StreamingTV | Has Streaming TV or not | Yes, No, No internet service |
| 15 | StreamingMovies | Has Streaming movies or not | Yes, No, No internet service |

| | | | |
|---|---|---|---|
| 16 | Contract | Customer’s current contract type | Month-to-Month, One Year, Two Year |
| 17 | PaperlessBilling | Ha paperless billing or not | Yes, No |
| 18 | PaymentMethod | Customer’s payment method | Electronic check, Mailed check, Bank transfer, Credit card |
| 19 | MonthlyCharges | The monthly charges accrued by the customer | Numeric value |
| 20 | TotalCharges | Total charges accumulated by the customer | Numeric value |
| 21 | Churn | Whether the customer churned or not | Yes (or) No |

### 3.2.3 Pre-processing and Exploratory Data Analysis

In machine learning, data preprocessing is used to transform the dataset's raw data into clean data. Data preprocessing entails filling in missing data with median values, deleting unnecessary aspects from the dataset (such as Customer ID), and identifying the key features needed to train a high-performing model. In the proposed system, data cleaning, data transformation, and data normalization are performed in this stage.

After that, the exploratory data analysis (EDA) step is continued to process. Exploratory Data Analysis (EDA) is a technique for analyzing the dataset in order to draw conclusions from its key properties. The dataset includes a predictive feature (Churn or No Churn), numerical features, and categorical features. It is really difficult to glance at a large spreadsheet and identify key details about the data. Therefore, EDA is crucial in certain situations. The purpose of conducting EDA is to make sense of the data and help to be better understand each feature before feeding them to machine learning models, thereby, making the modelling more efficient.

At the beginning of EDA, “pandas.DataFrame.info” method is used to get information about the data. This method outputs a brief summary of the data frame that includes the names of the columns and their data types, the number of non-null values, and the memory consumption.

```
------------------------------
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 21 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   customerID        7043 non-null   object
 1   gender            7043 non-null   object
 2   SeniorCitizen     7043 non-null   int64
 3   Partner           7043 non-null   object
 4   Dependents        7043 non-null   object
 5   tenure            7043 non-null   int64
 6   PhoneService      7043 non-null   object
 7   MultipleLines     7043 non-null   object
 8   InternetService   7043 non-null   object
 9   OnlineSecurity    7043 non-null   object
 10  OnlineBackup      7043 non-null   object
 11  DeviceProtection  7043 non-null   object
 12  TechSupport       7043 non-null   object
 13  StreamingTV       7043 non-null   object
 14  StreamingMovies   7043 non-null   object
 15  Contract          7043 non-null   object
 16  PaperlessBilling  7043 non-null   object
 17  PaymentMethod     7043 non-null   object
 18  MonthlyCharges    7043 non-null   float64
 19  TotalCharges      7043 non-null   object
 20  Churn             7043 non-null   object
dtypes: float64(1), int64(2), object(18)
memory usage: 1.1+ MB
None
```

**Figure 3.3 Summary of the Data Frame**

The data set has 21 columns and 7043 observations, as seen above. The data collection does not appear to include any null values, yet the column. The customer's unique identity is found in the column "Customer ID". To determine whether or not a customer will churn, the column "Customer ID" is worthless.

Therefore, this column is removed from the data set although it is not required for the prediction. “Total Charges” column is incorrectly identified as an object. This column is a numeric variable since it shows the total cost incurred by the customer. This column must be converted into a numeric data type for further analysis.

The “pd.to” numeric function can be used to accomplish this numeric conversion. This function throws an exception by default if it encounters non-numeric data. Nevertheless, it can avoid those cases with the errors = "coerce" parameter, and substitute a NaN instead. The “Total Charges” column has 11 missing values as shown in the following figure 3.4.

```
------- Check for null entries ----------
gender              0
SeniorCitizen       0
Partner             0
Dependents          0
tenure              0
PhoneService        0
MultipleLines       0
InternetService     0
OnlineSecurity      0
OnlineBackup        0
DeviceProtection    0
TechSupport         0
StreamingTV         0
StreamingMovies     0
Contract            0
PaperlessBilling    0
PaymentMethod       0
MonthlyCharges      0
TotalCharges       11
Churn               0
dtype: int64
```

**Figure 3. 4Checking Missing Values**

In spite of the fact that "Monthly Charges" column is not null for these entries, these observations likewise have a tenure of 0. These observations are eliminated from the dataset because they seemed to be conflicting. Then EDA Analysis is carried out to get the domain knowledge and to understand the data in telecom dataset.

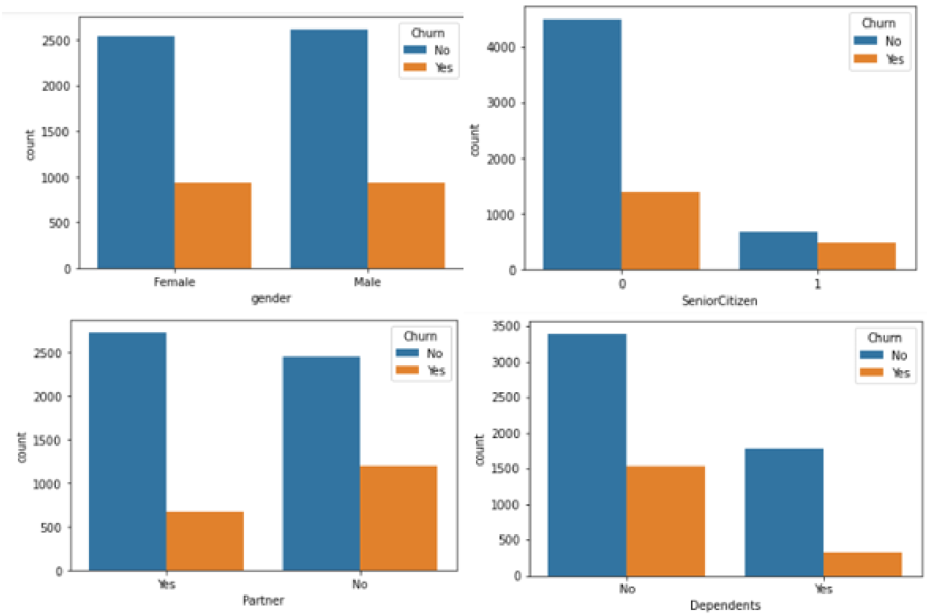


**Figure 3.5 Demographic Features of Telecom Dataset**

Senior citizens' turnover is approximately twice as high as that of young people. Gender is not expected to significantly influence predictions. Both male and female customers exhibit the same amount of churn. Customers who have a partner are to churn less frequently than those who don't. Younger customers (Senior Citizen = No), customers without partners, and customers without dependents all have greater rates of attrition. The data's demographic section singles out senior citizens without partners or dependents as a particular client group that is more likely to leave.

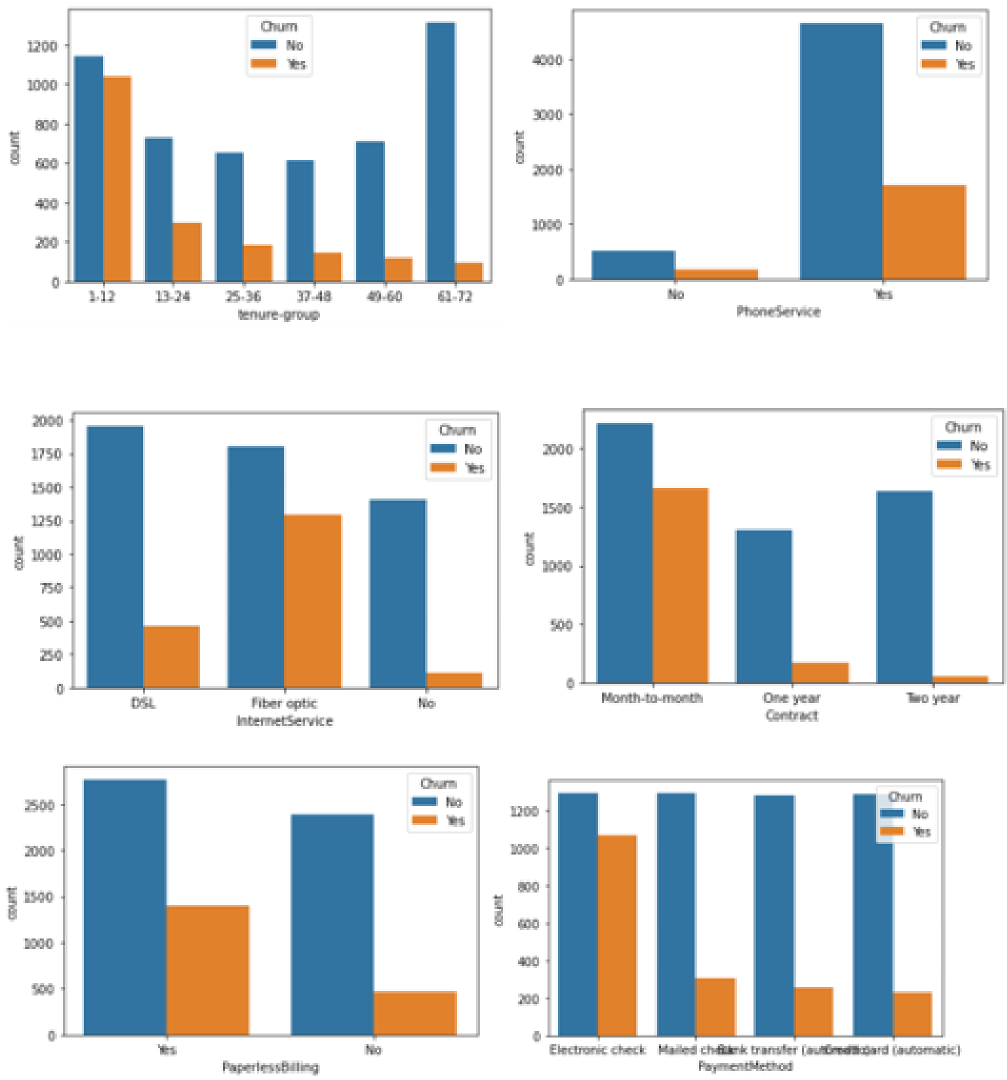


**Figure 3.6 Customer Service and Numerical Features of Telecom Dataset**

The tenure is appropriately demonstrating that the majority of consumers have only recently joined the telecom provider (0–9 months). Additionally, the first few months have the highest turnover rates (0–9 months). Within the first 30 months, 75% of clients who decide to leave the telco firm do so. Nearly 90.3% of consumers use phone services, and their rate of turnover is greater. Customers with fiber-optic internet are more likely to leave the company. High costs, competition, poor customer service, and a host of other factors could all contribute to this. One of the possible causes of customer attrition is the fact that fiber-optic service is significantly more expensive than DSL.

The shorter the contract, the higher the churn rate. Those with more extended plans face additional barriers when canceling early. This clearly explains the motivation for companies to have long-term relationships with their customers. The churn rate is higher for customers who opted for paperless billing. About 59.2% of customers use paperless billing. Customers who pay with electronic checks are more likely to churn, and this kind of payment is more common than other payment types.

The monthly charges reveal that customers with higher monthly rates have a higher rate of churn. This implies that offers such as discounts and specials may persuade customers to stay. In comparison to clients with yearly contracts, customers with month-to-month contracts see greater rates of customer attrition.

The distribution of tenure and monthly charges by churn are displayed in the above charts. The distributions of the two classes (No and Yes) are different for all quantitative attributes, indicating that all of the attributes will be helpful in determining whether or not a client churns.

### 3.3 Correlation matrix for Feature Selection

Given that the dataset contains a substantial number of dependent variables, feature selection is applied to it. Finding pertinent features for the model's construction is done through feature selection.

**Table 3.2 Size of Correlation Matrix**

| Size of Correlation | Interpretation |
|---|---|
| .90 to 1.00 (-.90 to -1) | Very high (positive/negative) Correlation |
| .70 to .90 (-.70 to -90) | High (positive/negative) Correlation |
| .50 to .70 (-.50 to -.70) | Moderate (positive/negative) Correlation |
| .30 to .50 (-.30 to -.50) | Low (positive/negative) Correlation |
| .00 to .30 (-0 to -.30) | Negligible Correlation |

A measure of correlation is the degree to which one variable is influenced by another. The variables will not be taken into account if the correlation exceeds the threshold of correlation bigger than 0.5 since it will have an impact on the model's accuracy (Mukaka, 2012). In Table 3.2 shows the degree of association between interpretation and size.

In this proposed system the feature selection is carried out using the Correlation Matrix with Heatmap approach. The accuracy of the model is enhanced by feature selection. The model is trained more quickly, and its complexity has decreased. The dependent and independent factors, as well as the dependent variables, are shown to be correlated.

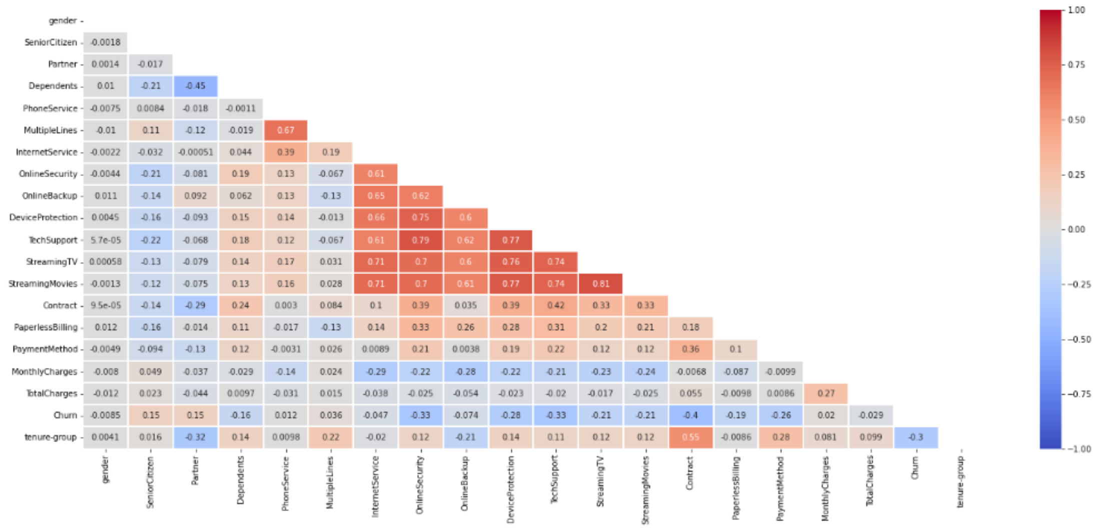


**Figure 3. 7 Feature Selection with Correlation Matrix**

This plot shows that variables like “Multiple Lines” is highly correlated with “Phone Services”. Moreover, “Online Security”, “Online Backup”, “Device Protection”, “Tech Support”, “Streaming TV” and “Streaming Movies” are also strongly correlated with “Internet Service”.

Therefore, “Multiple Lines”, “Online Security”, “Online Backup”, “Device Protection”, “Tech Support”, “Streaming TV” and “Streaming Movies” columns are dropped before building the model. No other variables are correlated with each other but dropped genders column because it has almost equal number of churners. Therefore, there is no significance in prediction.

### 3.3.1 Correlation between Churn and Selected Features

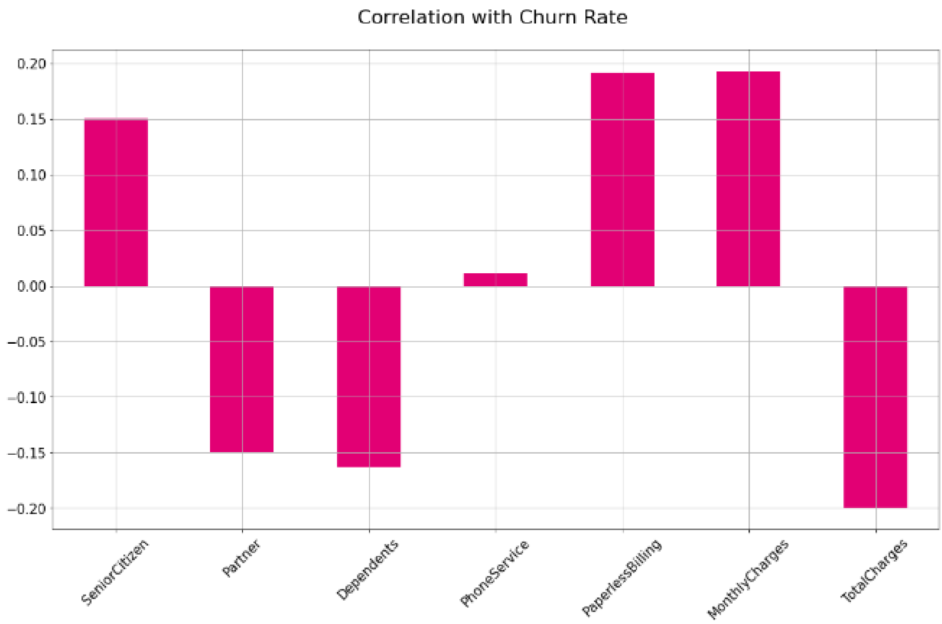


**Figure 3. 8 Correlation between churn and selected Boolean and numeric variables**

The majority of older citizens turnover, which is correlated positively with customer age Longer tenure may logically also indicate greater loyalty and a decreased likelihood of churn.

Additionally, it makes sense that higher monthly fees could raise the risk of turnover. The fact that overall charges have a negative link to churn, however, is intriguing. The reason might be that final costs also depend on how much time a client has spent with a business (tenure has a negative correlation).

It is debatable whether "total charges" is a sufficient variable to comprehend consumer behavior and whether the customer tracks it. There is a possible link between paperless billing and turnover that deserves further investigation (the underlying causes of that behavior are unknown).

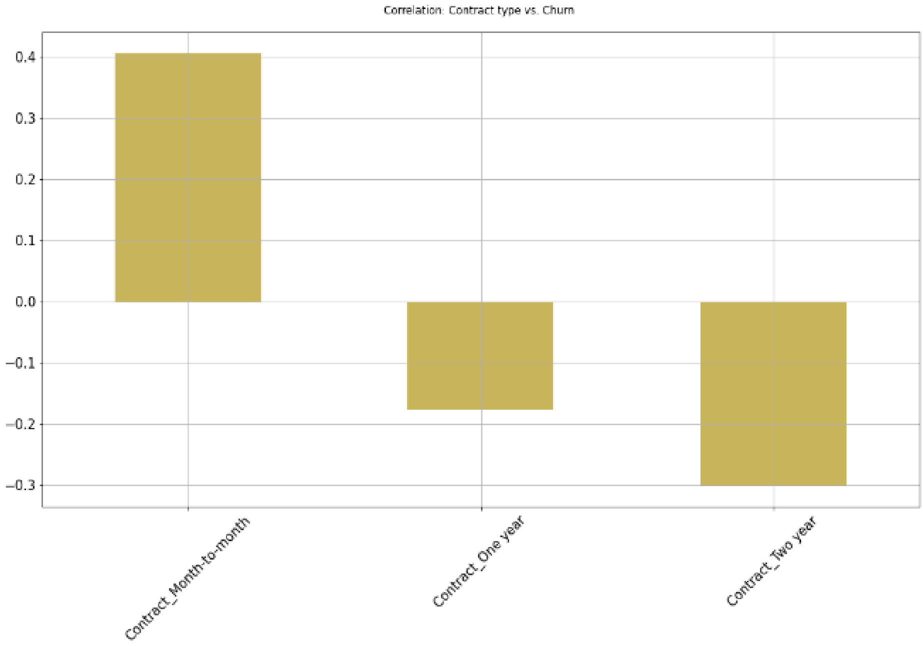


**Figure 3. 9 Correlation between Churn and Contract types**

The month-to-month type of subscription is most exposed to churn risk. A longer contract duration is a good churn prevention mechanism.

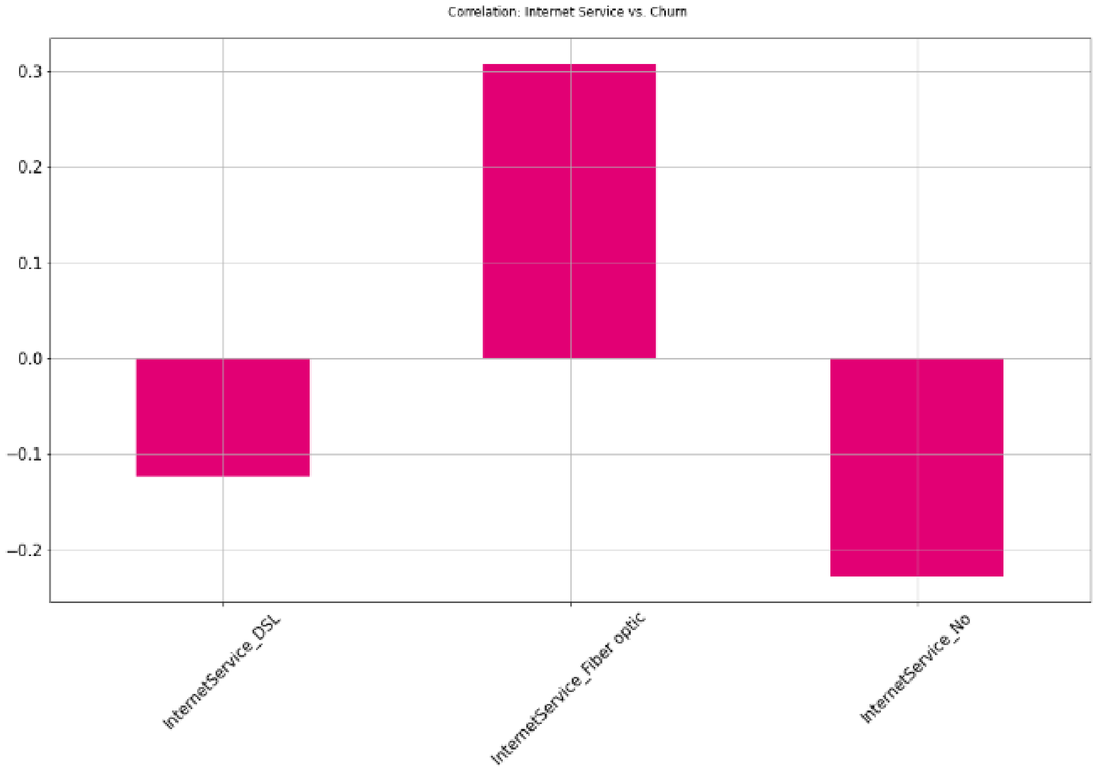


**Figure 3.10 Correlation between Churn and Internet Service types**

The" internet service_ Fiber optic" is highly correlate with 'Churn' variable than other two types of services.

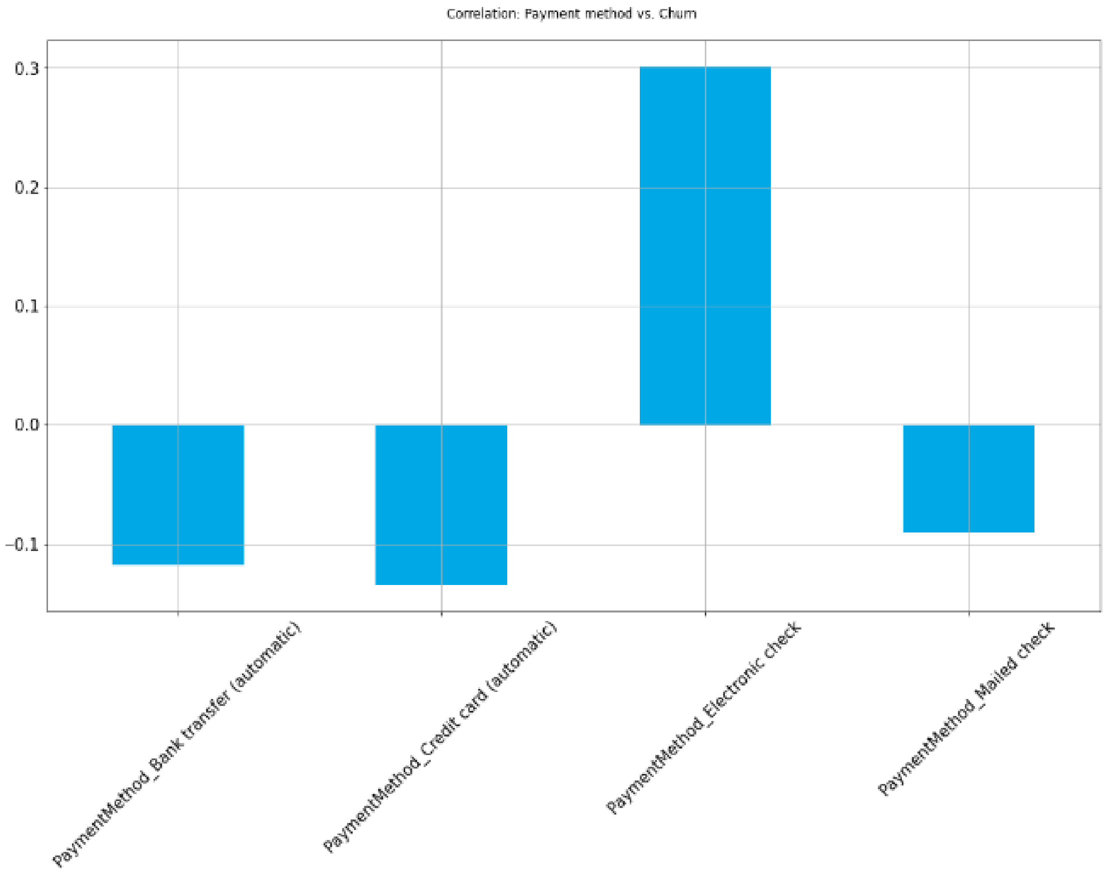


**Figure 3. 11 Correlation between Churn and Payment Method**

It is necessary to investigate the reasons for a positive correlation between electronic check as a payment method and churn.

### 3.3.2 Multicollinearity

Multicollinearity is a term used in data science to describe a situation caused by a high correlation between two or more independent variables. In other words, a variable that is independent can predict a variable that is independent. High multicollinearity variables are redundant. Therefore, this model occurs complicated interpretation, and the risk of overfitting is occurred.

A fantastic technique for evaluating multicollinearity is VIF (Variable Inflation Factors). VIF assesses how a variable strongly correlates with a set of additional independent variables in a dataset. The VIF scales from 1 to 10, and a value greater than 10 indicates a significant degree of multicollinearity between the independent variables.

| | variables | Variable Inflation Factors |
|---|---|---|
| 0 | SeniorCitizen | 1.327866 |
| 1 | Partner | 2.813692 |
| 2 | Dependents | 1.916758 |
| 3 | tenure | 10.451417 |
| 4 | PhoneService | 7.881414 |
| 5 | PaperlessBilling | 2.815119 |
| 6 | MonthlyCharges | 13.840300 |
| 7 | TotalCharges | 12.451150 |

**Figure 3.12 Checking Multicollinearity between Variables (VIF 1)**

A high VIF value can be found in the features "Monthly Charges" and "Total Charges." The numbers for "Monthly Charges" and "Total Charges" are shown on a scatter graph to see how they relate to one another.

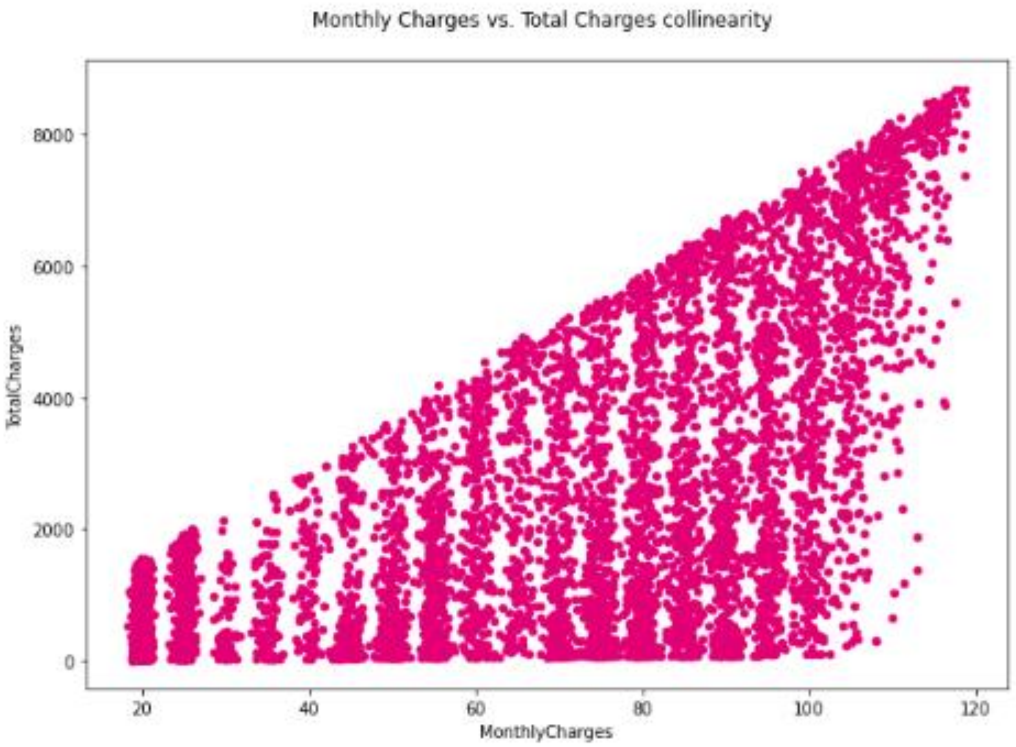


**Figure 3.13 Multicollinearity between Monthly Charges and Total Charges**

The scatter graph demonstrates the collinearity between the variables (features) "Total Charges" and "Monthly Charges." The multicollinearity between correlated features will be decreased by eliminating one of those features. The "Total Charges" feature will be removed because "Monthly Charges" variable is strongly positive correlation with Churn.

| | variables | Variable Inflation Factors |
|---|---|---|
| 0 | SeniorCitizen | 1.322783 |
| 1 | Partner | 2.812721 |
| 2 | Dependents | 1.899884 |
| 3 | tenure | 3.286508 |
| 4 | PhoneService | 5.596376 |
| 5 | PaperlessBilling | 2.741617 |
| 6 | MonthlyCharges | 7.448061 |

**Figure 3.14 Checking Multicollinearity between Variables (VIF 2)**

The test dataset's multicollinearity between correlated features is decreased by removing the "Total Charges" variable (including "tenure").

### 3.4 Data Normalization

The process of extracting features from the data and converting them into a format appropriate for the machine learning model is known as "data normalization". Both numerical and categorical variables must be transformed in this proposed system. Before training the model, all categorical features in the dataset are converted into numerical labels because the majority of machine learning techniques demand them.

#### 3.4.1 Label Encoding

Numerical numbers are used to substitute category values in label encoding. Each category is replaced by a numerical label using this encoding. The following binary variables are encoded using labels in this project: Gender, Senior Citizen, Partner, Dependents, Paperless Billing, Phone Service, and Churn are the first seven factors.

**Table 3.3 Label Encoding**

| Gender | SeniorCitizen | Partner | Dependents | PhoneService | PaperlessBilling | Churn |
|---|---|---|---|---|---|---|
| Female | 1 | Yes | Yes | Yes | Yes | Yes |
| Male | 0 | No | No | No | No | No |

### 3.4.2 One-Hot Encoding

For each level of the categorical variable, a new binary column is produced through one-hot encoding. The new column has zeros and ones in order to indicate whether the value of categorical variable or not. One-hot encoding is used in this proposed system to categorize the following variables: “Internet Services”, “Contract” and “Payment Method” are the first three items on the list.

## 3.5 Splitting the data in training and testing sets

Splitting the data into two groups, known as the training and testing sets, is the first stage in creating a model. The machine learning algorithm creates the model using the training set. The test set is used to assess the performance of the model and contains samples that are not included in the learning process. To ensure an impartial assessment, it is crucial to evaluate the model's quality using unobserved data.

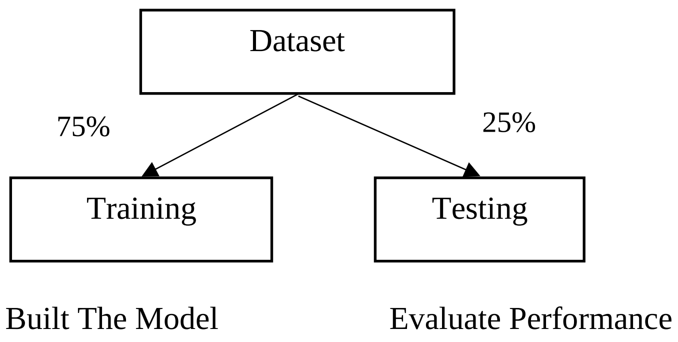


**Figure 3.15 Splitting the dataset into training and test set**

## 3.6 Methodology of the Proposed System

In this stage, the preprocessed data is utilized to create the machine learning model to forecast customer turnover. For building accurate and comprehensible churn prediction models, the methods used in this proposed system are Logistic Regression and Decision Tree (CART). The goal of these method applied in this system is to evaluate and contrast the effectiveness of Logistic Regression and Decision Tree (CART) Techniques in Predicting Telecom Customer Churn. This proposed system uses a dataset as input and builds prediction model using it. The input dataset for a classification task is typically split into train and test datasets. With the training dataset is used to develop the prediction model and the test dataset is used to evaluate the model's performance using evaluation metrics.

### 3.6.1 Logistic Regression

At first, the logistic regression model is created. It is the most frequently used algorithm to model binary dependent variables. It is a particular kind of statistical probability classification model that is mostly applied to classification issues [9]. This method can be applied to a variety of variable combinations and aid in more accurate customer churn prediction. It is possible to calculate the variables' prediction ability.

The curve is fitted to the dataset in a statistical model. When the target variable is binary, this approach is helpful. This algorithm for predictive analysis is built on the idea of probability. The sigmoid function, also known as the "logistic function," is a complex cost function that is used in logistic regression. The predictions are converted to probabilities using the sigmoid function. The output is constrained, and a probability score between 0 and 1 was given. One of the widely used algorithms for classification problems is logistic regression.

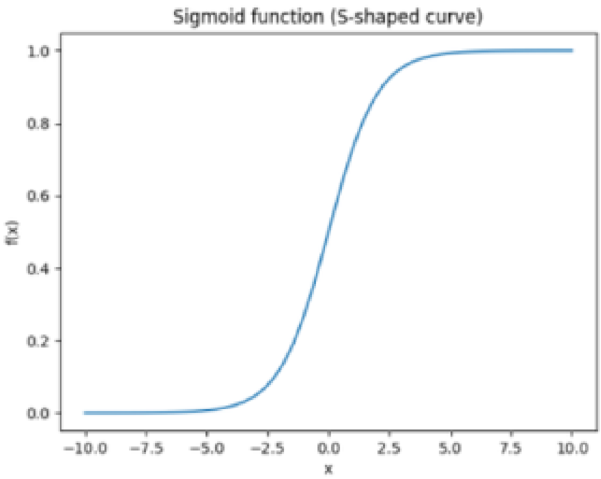


**Figure 3.16 Logistic Regression**

Logistic regression is a data mining technique used to predict occurrence probability of customer churn [2]. A statistically based method for analyzing how one variable affects another is called logistic regression. By creating a set of equations linking the input values (i.e., factors influencing customer turnover) with the output field, predictions are created (probability of churn).

The equation describes the mathematical formulas for a logistic regression model.

$$\text{Logit (Y)} = \log\left[\frac{Y}{1-Y}\right] = Y = \frac{e^{\beta_0+\beta_1 x_1+\cdots+\beta_n x_n}}{1+e^{\beta_0+\beta_1 x_1+\cdots+\beta_n x_n}} \quad \textbf{3. 1}$$

where,

- $\beta_0$ = Constant Coefficient
- $\beta_n$= Coefficient of $x_n$
- $x_n$ = independent variable (where n= 1,2,…,n)
- P(Y)= Probability that Y equals 1

### 3.6.2 Classification and Regression Tree (CART)

A prediction technique used in machine learning is called a Classification and Regression Tree (CART). It illustrates how the values of a target variable can be predicted from those of other variables. It is a decision tree with predictions for the target variable at each node's end for each fork in the predictor variables [6].

A crucial decision tree algorithm that forms the basis of machine learning is the CART algorithm. Additionally, it serves as the foundation for other potent machine learning techniques, including enhanced decision trees, random forest, and bagged decision trees.

One of the earliest and most essential algorithms is the classification and regression tree (CART) approach. Based on specific predictor factors, it is used to forecast outcomes. They require extremely little data pre-processing, making them ideal for data mining jobs. When compared to other analytical models, decision tree models have the distinct advantage of being simple to comprehend and use.

Leo Breiman, Jerome Friedman, Richard Olshen, and Charles Stone introduced the Classification and Regression Tree methodology, or CART, in 1984. It is organized like a tree, with the root node at the top.

In this system, the Gini index is the splitting criterion Gini index. The Gini index G(S) at a node S in a CART tree, is defined as:

$$\textbf{Gini=1-}\sum_{i=1}^{c}(P_i)^2 \qquad \textbf{3.2}$$

**CART Algorithm**

Input: Data partition, D, which is a set of training tuples and their related class labels;

Attribute_list: Attribute_selection_method, to determine the splitting criterion that "best" partitions.

Output: A decision tree.

Algorithm:

**Start**

**Step 1- CREATING A ROOT NODE**

1. Create a root node N

2. If tuples in D are all of the similar class, C then

3. Return N as a leaf node label with the class C;

4. If attribute list is empty then

5. Return N as a leaf node label with the majority class in D

**Step 2- ATTRIBUTE SELECTION**

6. Apply attribute_selection_method(D, attribute_list) to discover the "best " splitting _criterion attribute;

7. Label node N with splitting _criterion;

8. Update the attribute_list

**Step 3- SPLIT THE TREE**

9. for each outcome j of splitting_criterion //partition the tuples and produce subtrees for each partition

10.Based on splitting_criterion attribute Split the tree into two part

12. attach a leaf labeled with the majority class D in node N:

13. else attach the node returned by Generate_decision_tree(Dj:attribute_list)to node N: end for

14. return N

**End**

# CHAPTER 4

# IMPLEMENTATION OF THE PROPOSED SYSTEM

This chapter represents a step-by-step comparison of the telecom churn prediction system. The effectiveness of two models is assessed by comparison in three different contexts (without feature selection, with feature selection and proposed system with multicollinearity). The supervised machine learning models are evaluated for their prediction abilities in this chapter. The accuracy, precision, recall, and F1-score of each model are evaluated in these comparisons. The analysis results of these comparisons are shown by figures, and these comparisons indicate that how the performance vary.

## 4.1 Performance Evaluation

A critical stage in the churn prediction process is evaluating the performance of churn prediction models to determine how well the model generalizes. Accuracy [15], precision and recall, and F-Measure are some of the popular evaluation metrics used by many researchers for churn prediction evaluation.

In this proposed system, several measures are used to evaluate the models' performance. Accuracy, precision, recall, and specificity are calculated once each model's confusion matrix is built [23]. It is determined how many samples are true positive (TP), true negative (TN), false positive (FP), and false negative (FN). The confusion matrix is a common technique used to assess the effectiveness of a classification model. Using a confusion matrix, a classifier's actual and anticipated classifications are shown.

**True positives (TP):** the number of customers that are actually churners and the classification model has identified them correctly as churners.

**True negatives (TN):** the number of customers that actual are non-churners and the classification model has identified them correctly as non-churners.

**False positives (FP):** the number of customers who are non-churners but the classification model incorrectly determined them as churners.

**False negatives (FN):** the number of customers who are churners but the classification model incorrectly determined them as non-churner.

**Table 4. 1 A confusion matrix for a binary classifier**

| | Predicted Classes | | |
|---|---|---|---|
| **Actual Classes** | Class=Yes /+/ Churn | Class=Yes /+/ Churn TP (true positive) | Class=No/-/No-Churn FN (false negative) |
| | Class=No/-/No-Churn | FP (false positive) | TN (true negative) |

**Accuracy:** the portion of the total number of correctly predicted cases, is calculated as follows:

$$\textbf{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{4.1}$$

**Precision**: the fraction of predicted churners that do churn, is calculated as follow:

$$\textbf{Precision} = \frac{TP}{TP + FP} \tag{4.2}$$

**Recall**: the fraction of real churners which are correctly determined as churners is calculated as follow:

$$\textbf{Recall} = \frac{TP}{TP + FN} \tag{4.3}$$

**F-Measure**: can be considered a weighted average of precision and recall, with the best value being 1 and the poorest being 0. Precision and memory make equal relative contributions to the F1 score. It is calculated as follow:

$$\textbf{F-measure} = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4.4}$$

### 4.2 Experimental Setup

This chapter's objective is to represent implementation of the proposed system, design, and performance assessment. High-value customers who are expected to leave soon can be kept by using the telecom churn prediction system. Python is the programming language used to implement this proposed system.

### 4.3 Implementation of the System

The “Welcome” page of the proposed system is shown in Figure 4.1. Firstly, “Start” button is clicked.

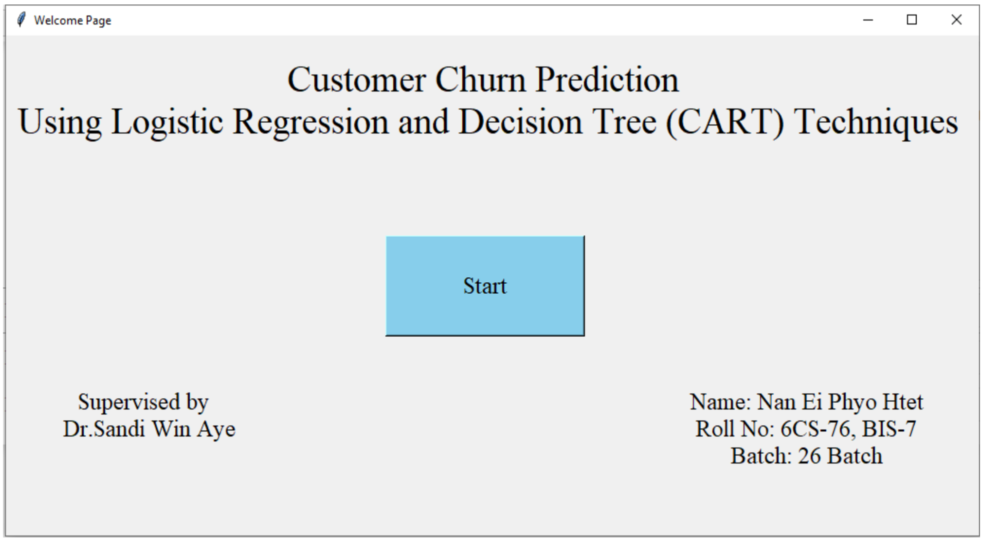


**Figure 4.1 Welcome Page of the Proposed System**

When the "Start" button is clicked, “Home” page of this system, as depicted in Figure 4.2, is displayed. This page includes “Comparative Results” button of the proposed system, “Load Feature Data Set”, “Training Feature Data Set”, and “Calculation” buttons. Moreover, “Decision trees (CART)” and “Logistic Regression” buttons to show accuracy of each method are included in this page.

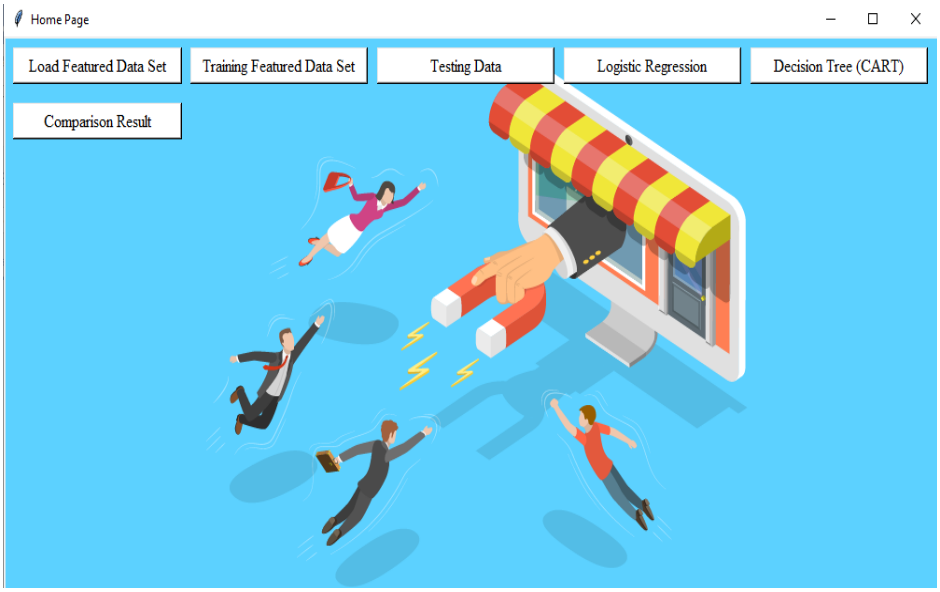


**Figure 4.2 Home Page of the Proposed System**

The selected feature dataset that is used in this proposed system can be viewed by clicking the "Load Feature Dataset" button. The result is shown in Figure 4.3.

Import CSV File to Tkinter Table

| SeniorCitizen | Partner | Dependents | Tenure | PhoneService | PaperlessBilling | MonthlyCharges | InternetService_D | InternetService_Fi | InternetService |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 66 | 1 | 1 | 105.65 | 0 | 1 | 0 |
| 1 | 1 | 0 | 4 | 1 | 1 | 74.4 | 0 | 1 | 0 |
| 0 | 1 | 1 | 11 | 0 | 1 | 29.6 | 1 | 0 | 0 |
| 0 | 1 | 1 | 72 | 1 | 1 | 103.2 | 0 | 1 | 0 |
| 0 | 1 | 1 | 24 | 1 | 1 | 84.8 | 1 | 0 | 0 |
| 0 | 0 | 0 | 72 | 1 | 1 | 21.15 | 0 | 0 | 1 |
| 0 | 0 | 0 | 12 | 0 | 0 | 60.65 | 1 | 0 | 0 |
| 0 | 0 | 0 | 19 | 1 | 1 | 78.7 | 0 | 1 | 0 |
| 0 | 0 | 0 | 67 | 1 | 1 | 102.95 | 0 | 1 | 0 |
| 0 | 0 | 0 | 38 | 1 | 1 | 69.5 | 0 | 1 | 0 |
| 1 | 0 | 0 | 1 | 1 | 1 | 75.75 | 0 | 1 | 0 |
| 1 | 1 | 0 | 55 | 1 | 0 | 60.0 | 1 | 0 | 0 |
| 0 | 0 | 0 | 2 | 1 | 1 | 20.05 | 0 | 0 | 1 |
| 1 | 0 | 0 | 6 | 0 | 1 | 44.4 | 1 | 0 | 0 |
| 0 | 1 | 0 | 68 | 1 | 0 | 64.1 | 1 | 0 | 0 |
| 0 | 0 | 0 | 13 | 1 | 0 | 73.35 | 1 | 0 | 0 |
| 0 | 0 | 0 | 9 | 1 | 1 | 44.2 | 1 | 0 | 0 |
| 0 | 0 | 0 | 18 | 1 | 1 | 95.05 | 0 | 1 | 0 |
| 0 | 1 | 0 | 44 | 1 | 1 | 84.8 | 0 | 1 | 0 |
| 1 | 1 | 0 | 63 | 1 | 1 | 103.5 | 0 | 1 | 0 |
| 0 | 0 | 0 | 72 | 1 | 1 | 104.95 | 0 | 1 | 0 |
| 0 | 0 | 0 | 12 | 1 | 1 | 59.8 | 1 | 0 | 0 |

**Figure 4.3 Feature Dataset of Proposed System**

To check whether the customer is churner or not, the necessary data are entered into the data entry form firstly. After entering the data, the user must select their desired model: decision tree (CART) (or) logistic regression. If "Logistic Regression" button is clicked, result of this model is shown in the following figure 4.4.

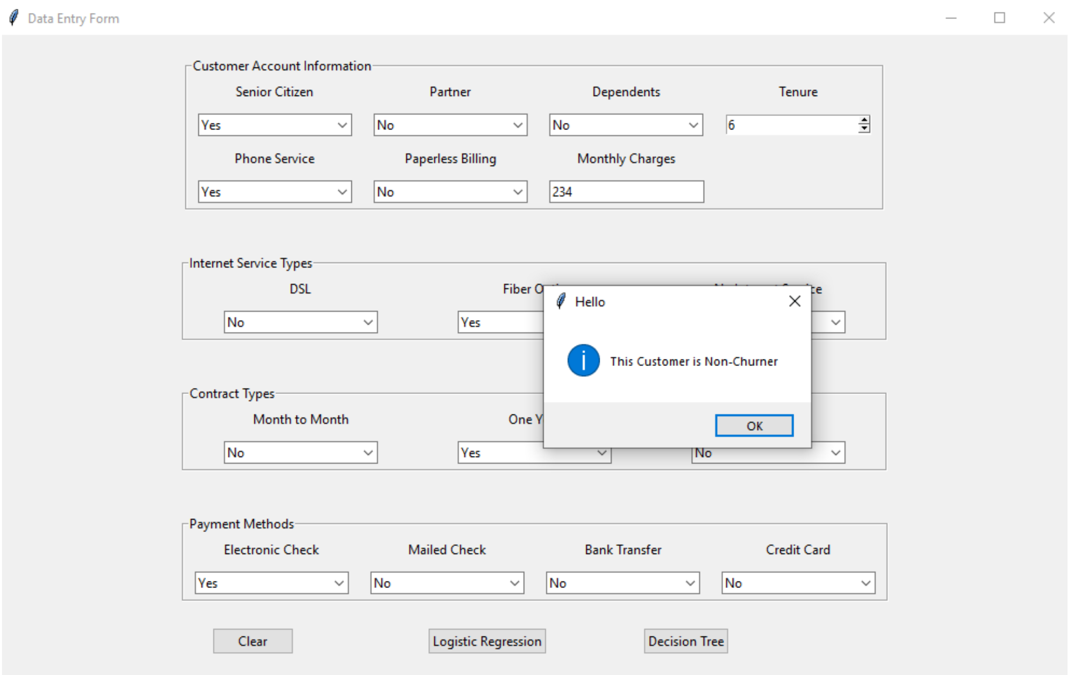


**Figure 4.4 Logistic Regression Result for Non-Churner**

If "Decision Tree (CART)" button is clicked, result of this model is displayed in the following figure 4.5.

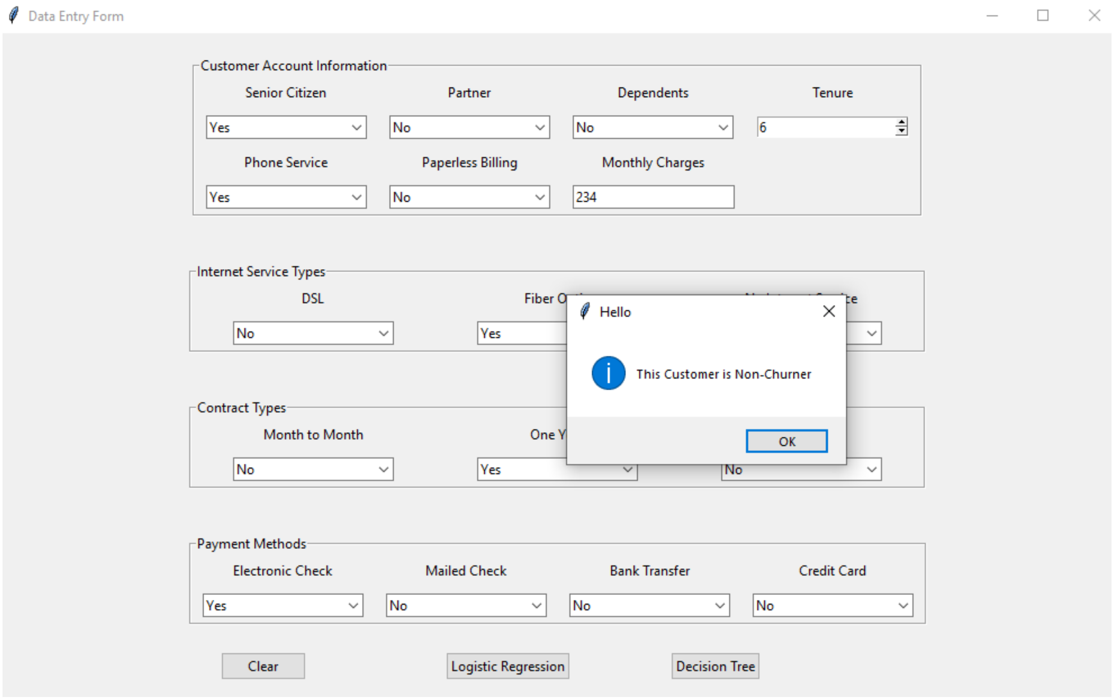


**Figure 4.5 Decision Tree (CART) result for Non-Churner**

## 4.4 Experimental Results

The Telecom Customer Churn dataset undergoes an experiment to determine the most effective supervised machine learning model for forecasting customer churn. The data used in this experiment are the dataset without feature selection, the dataset with feature selection, and the dataset with the proposed system, which includes screening for multicollinearity.

In term of accuracy, precision, recall, and F1-score as evaluation metrics, it has been found that the supervised machine learning algorithms: Logistic Regression and Decision Tree (CART) perform better when feature selection and multicollinearity are processed. The Logistic Regression machine learning algorithm has outperformed the Decision Tree (CART) model in all studies conducted in terms of all assessment measures.

### 4.4.1 Experimental Result without Feature Selection

The following results are obtained from the initial dataset (without feature selection). The results of Logistic Regression and Decision Tree (CART) are shown in the figure 4.6 and 4.7.

```
------------------------------------------------------------
Accuracy of Logistic Regression:  0.7940841865756542
Execution time: 0.13391685 seconds
------------------------------------------------------------
              precision    recall  f1-score   support

           0       0.84      0.89      0.86      1291
           1       0.64      0.53      0.58       467

    accuracy                           0.79      1758
   macro avg       0.74      0.71      0.72      1758
weighted avg       0.78      0.79      0.79      1758
```

**Figure 4.6 Logistic Regression Result of Without Feature Selection**

```
------------------------------------------------------------
Accuracy of Decision Tree:  0.7201365187713311
Execution time: 0.07295632 seconds
------------------------------------------------------------
              precision    recall  f1-score   support

           0       0.81      0.81      0.81      1291
           1       0.47      0.48      0.48       467

    accuracy                           0.72      1758
   macro avg       0.64      0.64      0.64      1758
weighted avg       0.72      0.72      0.72      1758
```

**Figure 4.7 Decision Tree (CART) Result of Without Feature Selection**

The confusion matrix of Logistic Regression and Decision Tree (CART) models are shown in following figure 4.8.

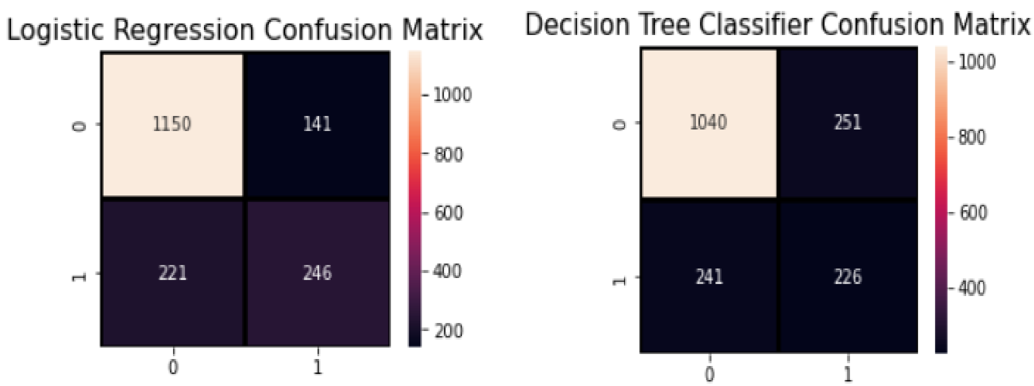


**Figure 4.8 Confusion Matrix of Without Feature Selection**

The performance evaluation of Logistic Regression and Decision Tree (CART) models without feature selection as bar chart is shown in the following figure 4.9.

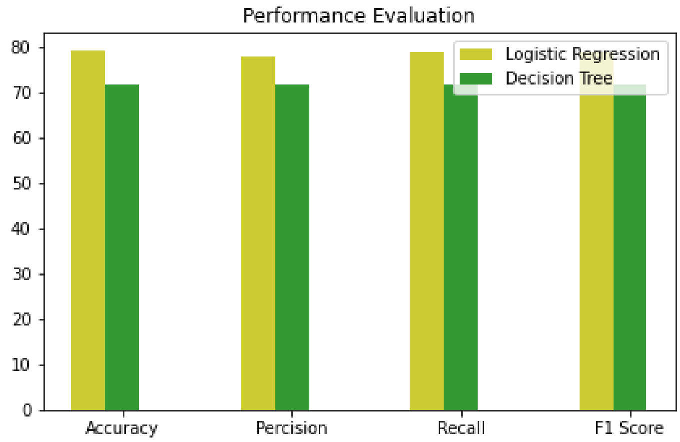


**Figure 4.9 Performance Evaluation of Without Feature Selection**

### 4.4.2 Experimental Result with Feature Selection

In this experiment, the correlation matrix is used for feature selection. Therefore, among 21 features, “Senior Citizen”, “Partner”, “Dependents”, “tenure”, “Phone Service”, “Internet Service”, “Contract”, “Paperless Billing”, “Payment Methods”, “Monthly Charges” and “Total Charges” (11 features) are used for this system. The accuracy of Logistic Regression and Decision Tree (CART) models increases slightly from 79% to 81% and 72% to 73%, respectively. The results are shown in the following figure 4.10 and 4.11 respectively.

```
--------------------------------------------------------
Accuracy of Logistic Regression:  0.8083048919226393
Execution time: 0.26317286 seconds
--------------------------------------------------------
              precision    recall  f1-score   support

           0       0.85      0.90      0.87      1291
           1       0.67      0.55      0.61       467

    accuracy                           0.81      1758
   macro avg       0.76      0.73      0.74      1758
weighted avg       0.80      0.81      0.80      1758
```

**Figure 4.10 Logistic Regression Result of Feature Selection**

```
------------------------------------------------------------
Accuracy of Decision Tree:  0.7286689419795221
Execution time: 0.07132268 seconds
------------------------------------------------------------
              precision    recall  f1-score   support

           0       0.82      0.80      0.81      1291
           1       0.49      0.52      0.51       467

    accuracy                           0.73      1758
   macro avg       0.66      0.66      0.66      1758
weighted avg       0.73      0.73      0.73      1758
```

**Figure 4.11 Decision Tree (CART) Result of Feature Selection**

The confusion matrix for Logistic Regression and Decision Tree (CART) models are shown in following figure 4.12.

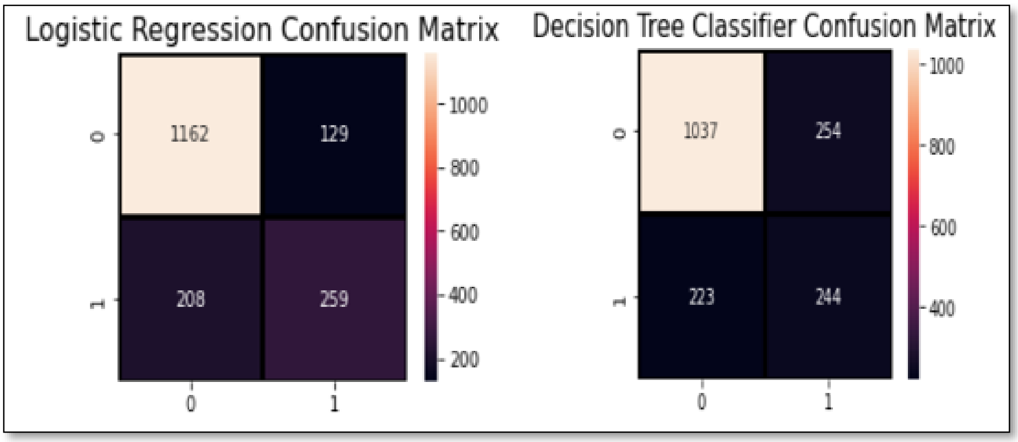


**Figure 4.12 Confusion Matrix of Feature Selection**

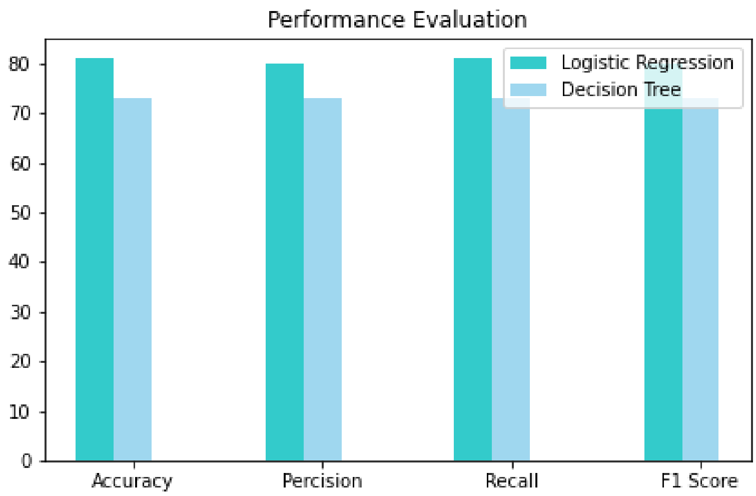


**Figure 4.13 Performance Evaluation of With Feature Selection**

The performance evaluation of Logistic Regression and Decision Tree (CART) models as bar chart is shown in the above figure 4.13.

### 4.4.3 Experimental Result of the Proposed System

The proposed system is done by using "Feature Selection and Checking Multicollinearity". In proposed system, "Senior Citizen", "Partner", "Dependents", "tenure", "Phone Service", "Internet Service", "Contract", "Paperless Billing", "Payment Methods" and "Monthly Charges" are the best 10 selected feature for building two models (Logistic Regression and Decision Tree (CART)).

```
---------------------------------------------------------
Accuracy of Logistic Regression:  0.8270762229806599
Execution time: 0.12798548 seconds
---------------------------------------------------------
              precision    recall  f1-score   support

           0       0.86      0.92      0.89      1291
           1       0.72      0.57      0.64       467

    accuracy                           0.83      1758
   macro avg       0.79      0.75      0.76      1758
weighted avg       0.82      0.83      0.82      1758
```

**Figure 4.14 Logistic Regression Result of Proposed System**

```
---------------------------------------------------------
Accuracy of Decision Tree:  0.7394766780432309
Execution time: 0.03999376 seconds
---------------------------------------------------------
              precision    recall  f1-score   support

           0       0.83      0.81      0.82      1291
           1       0.51      0.54      0.52       467

    accuracy                           0.74      1758
   macro avg       0.67      0.68      0.67      1758
weighted avg       0.74      0.74      0.74      1758
```

**Figure 4.15 Decision Tree (CART) Result of Proposed System**

The obtained accuracy of Logistic Regression is 83% while Decision Tree (CART) gets 74% of accuracy. This is the best accuracy among three experiments: Without Feature Selection, With Feature Selection and Proposed System (With Feature Selection and Multicollinearity). These results of proposed system are shown in the above figure 4.14 and 4.15 respectively

In Logistic Regression, the likelihood of a client leaving is determined by the data, which shows that 1187 and 267 are, respectively, true positives and false positives, or correctly identified instances. These instances amount to 1454, or 83% of the total 1758 occurrences that make up the total. Decision Tree Model is correctly classified the data 1043 and 254 respectively and gets 74% of the accuracy. The confusion matrix of this proposed system is shown in the following figure 4.16.

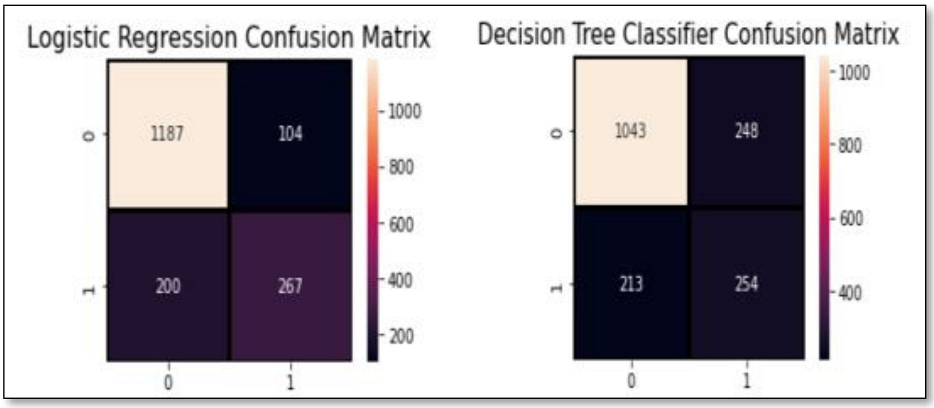


**Figure 4.16 Confusion Matrix of Proposed System**

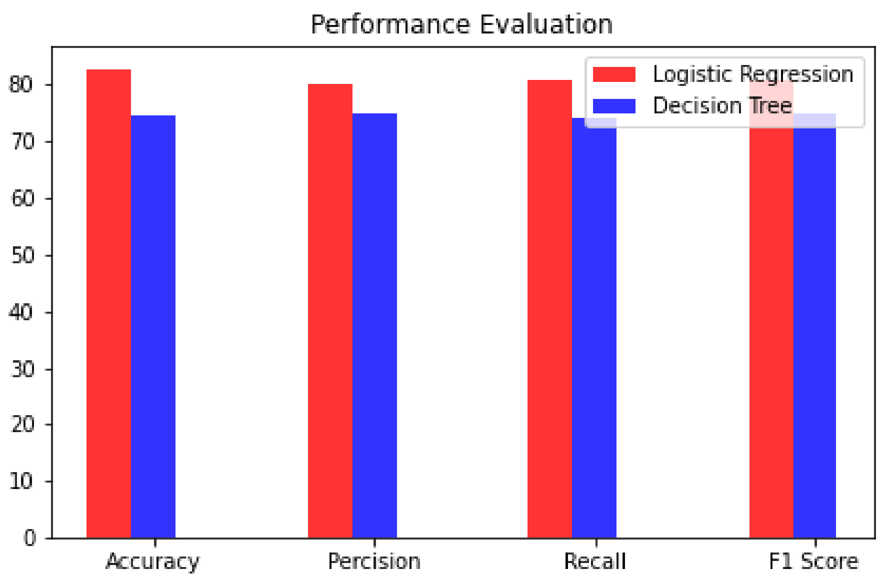


**Figure 4.17 Performance Evaluation of Proposed System**

The performance evaluation of Logistic Regression and Decision Tree (CART) models as bar chart is shown in the following figure 4.17.

### 4.5 Model Comparison

When it comes to the overall performance of how well they forecast the outcome, the accuracy of the initial model (without feature selection), the initial model with feature selection, and the proposed model with multicollinearity are not particularly different from one another.

**Table 4.2 Models Comparison**

| Comparison | Without Feature Selection Model | | Feature Selection Model | | Proposed Model | |
|---|---|---|---|---|---|---|
| | LR | CART | LR | CART | LR | CART |
| **Accuracy** | 79 | 72 | 81 | 73 | 83 | 74 |
| **Execution time** | 0.1339s | 0.0729s | 0.2631s | 0.0713s | 0.1279s | 0.0399s |
| **Precision** | 78 | 72 | 80 | 73 | 82 | 74 |
| **Recall** | 79 | 72 | 81 | 73 | 83 | 74 |
| **F1-score** | 79 | 72 | 80 | 73 | 82 | 74 |

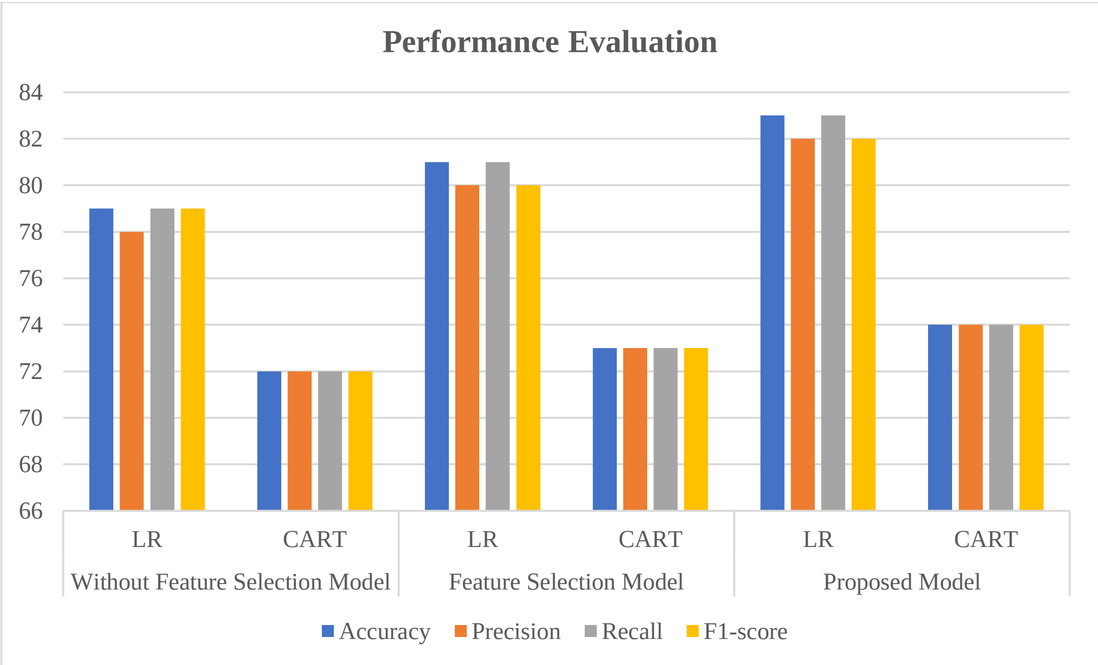


**Figure 4.18 Performance Evaluation of Models Comparison**

Nevertheless, the simplified model is leaner, quicker, and less resource-intensive because there is a significant difference in the number of variables. The performances of the initial model, the initial model with feature selection, and the final models are contrasted as shown in the above Table 4.1.

# CHAPTER 5

# CONCLUSION

In this proposed system, the comparison of two Data Mining and Machine Learning Techniques in telecom industry is focused on particularly in Logistic Regression and Decision Tree (CART). This proposed system presents the algorithm, Logistic Regression and Decision Tree (CART) are used to compare the accuracy, precision, recall and F1-score. The system is implemented using python programming language on the window platform. In this chapter, the summary of the main conclusion and advantages, limitations, and further extensions are suggested.

In this proposed system, the performance of two machine learning techniques: Logistic Regression and Decision Tree (CART) are analyzed. The Telecom dataset from the IBM Watson Analytics Community is used in this proposed system. In this system, three experiments: without feature selection, with feature selection and proposed system (with feature selection and multicollinearity) are carried out by using these two techniques.

In the first experiment, Logistic Regression Model get 79% accuracy score while CART can only have 72%. In the second experiment, the accuracy of Logistic Regression and CART improve from 79% to 81% and from 72% to 73% respectively. The last experiment, the proposed system gives the highest accuracy score: improving from 81% to 83% in Logistic Regression and from 73% to 74% in CART respectively. This proposed system suggests that Logistic Regression outperformed than the Decision Tree (CART) models.

## 5.1 Limitation and Further Extension

The proposed system identifies some potential future work that may be done. Only one branch of the dataset is investigated and examined in this study. Another area of the Telecom Customer Churn dataset may be investigated in the future. The Telecom Customer Churn dataset is utilized in this study to test the two machine learning methods. Additional strategies can also be investigated. It is possible to examine various machine-learning algorithms and analyze data.

# PUBLICATION

[1] Nan Ei Phyo Htet, Sandi Winn Aye, “Customer Churn Prediction Using Logistic Regression and Decision Tree (CART) Techniques”, University of Computer Studies, Yangon, Myanmar, 2022.

## REFERENCES


[1] Abdelrahim Kasem Ahmad, Assef Jafar and Kadan Aljoumaa, “Customer churn prediction in telecom using machine learning in big data platform”, Journal of Big Data, 2019

[2] Chinnu P Johny, Paul P. Mathai, “Customer Churn Prediction:A Survey”, International Journal of Advanced Research in Computer Science, Volume 8, No. 5, May – June 2017

[3] Essam Abou el Kassem, Essam Abou el Kassem, Shereen Ali Hussein, Fahad Kamal Alsheref, “Customer Churn Prediction Model and Identifying Features to Increase Customer Retention based on User Generated Content”, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 5, 2020

[4] Guo-en and Wei_dong, “Model of Customer Churn Prediction on Support Vector Machine”, Systems Engineering - Theory & Practice,Volume 28, Issue 1, January 2008, Pages 71-77

[5] Jones, T.O. and Sasser, W.E. (1995) Why Satisfied Customers Defect. Harvard Business Review, 73, 88-99

[6] H. Abbasimehr, S. Alizadeh, “A Novel Genetic Algorithm Based Method for Building Accurate and Comprehensible Churn Prediction Models”, International Journal of Research in Industrial Engineering, Volume 2, Number 4, 2013

[7] Hemlata Jain, Ajay Khunteta, Sumit Srivastava, “Churn Prediction in Telecommunication using Logistic Regression and Logit Boost”, International Conference on Computational Intelligence and Data Science (ICCIDS 2019)

[8] Kim, M., Park, M. and Jeong, D. (2004) The Effects of Customer Satisfaction and Switching Barrier on Customer Loyalty In Korean Mobile Telecommunication Services, Telecommunication Policy, 28, 145-159

[9] K. Sandhya Rani, Shaik Thaslima, N.G.L. Prasanna3, R.Vindhya4, and P. Srilakshmi, “Analysis of Customer Churn Prediction in Telecom Industry Using Logistic Regression”, International Journal of Innovative Research in Computer Science & Technology (IJIRCST) ISSN: 2347-5552, Volume-9, Issue-4, July 2021

[10] Lingling Yanga, Dongyang Li, Yao Lu, “Prediction Modeling and Analysis for Telecom Customer Churn in Two Months”, School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, China

[11] Mohamad Firman Maulana, Meriska Defriani, “Logistic Model Tree and Decision Tree J48 Algorithms for Predicting the Length of Study Period”, Journal Penelitian Ilmu Komputer, System Embedded & Logic, Vol. 8 (1): 39 – 48 (March 2020)

[12] Nashwan, Saeed Awadh and Hassan, Haslinda , “Impact of customer relationship management (CRM) on customer satisfaction and loyalty: a systematic review”, Journal of Advanced Research in Business and Management Studies, 6. pp. 86-107. ISSN 2462-1935

[13] Oleksandr Dorokhov, Liudmyla Dorokhova, Lyudmyla Malyarets and Iryna Ushakova,” Customer Churn Predictive Modeling By Classification Methods”, Bulletin of the Transilvania University of Bra_sov _ Vol 13(62), No. 1 – 2020

[14] Oyeniyi and Adeyemo, “Customers Churn Prediction in Financial Institution Using Artificial Neural Network”, Conference: International Multi-Conference on ICT, December 2019

[15] Pretam Jayaswal, Bakshi Rohit Prasad, Divya Tomar, and Sonali Agarwal, “An Ensemble Approach for Efficient Churn Prediction in Telecom Industry”, International Journal of Database Theory and Application Vol.9, No.8 (2016), pp.211-232.

[16] Shin-Yuan Hung, David C. Yen, and Hsiu-Yu Wang, “Applying data mining to telecom churn management,” Expert Syst. Appl., vol. 31, pp. 515–524, 2006.

[17] T. O. Jones and W. E. Sasser, “Why satisfied customers defect,” *Harvard Business Review*, vol. 73, no. 6, p. 88, 1995.

[18] Van den Poel, Dirk & Lariviere, Bart, 2004. "Customer attrition analysis for financial services using proportional hazard models," European Journal of Operational Research, Elsevier, vol. 157(1), pages 196-217, August.

[19] Vrushabh Jinde and Prof. Amit Savyanava,” Customer Churn Prediction System using Machine Learning”, International Journal of Advanced Science and Technology Vol. 29, No. 5, (2020), pp. 7957-7964

[20] V. Umayaparvathi, K. Iyakutti, “A Survey on Customer Churn Prediction in Telecom Industry: Datasets, Methods and Metrics”, Computer Science, Business, Published 2016

[21] Yasser Khan, Shahryar Shafiq, Abid Naeem3, Sabir Hussain, Sheeraz Ahmed and Nadeem Safwan, “Customers Churn Prediction using Artificial Neural Networks (ANN) in Telecom Industry”, International Journal of Advanced Computer Science and Applications, Vol. 10, No. 9, 2019.

[22] Yajing Zhang, Guotai Chi, Zhipeng Zhang, "Decision tree for credit scoring and discovery of significant features: an empirical analysis based on Chinese microfinance for farmers", Published by Faculty of Sciences and Mathematics,University of Niˇs, Serbia

[23] Zeynep Uyar Erdem, Banu Çalış Uslu, Seniye Ümit Fırat, "Customer Churn Prediction Analysis In A Telecommunication Company With Machine Learning Algorithms", Journal of Industrial Engineering 32(3), 496-512, 2021