Customer Retention: Detecting Churners in Telecoms Industry using Data Mining Techniques

Mahmoud Ewieda[1]
Business Information Systems
Helwan University
Helwan, Egypt

Essam M Shaaban[2]
Business Technology Department
Canadian International College School of Business Technology Department
Cairo, Egypt

Mohamed Roushdy[3]
Faculty of Computers and Information Technology, Future University in Egypt
New Cairo, Cairo, Egypt

***Abstract*—Customers are more concerned with the quality of services that companies can provide. Customer churn is the percentage of service for subscribers, who stop their subscriptions or the proportion of customers, who discontinue using the product of the firm or service within a given time frame. Services by various service providers or sellers are not very distinct that raise rivalry between firms to maintain the quality of their services and upgrade them. This paper aims at manifesting the service quality effect on customer satisfaction and churn prediction to reveal customers who have meant to leave a service. Predictive models can give the extent of the service quality effect on customer satisfaction for the correct determination of possible churners shortly for the provision of a retention solution. This paper analyses the impact of service quality and prediction models that depend on data mining (DM) techniques. The present model contains five steps: data-pre-processing, feature selection, sampling of data, training our classifier, testing for prediction, and output of prediction. A data set with 17 attributes and 5000 records used - which consist of 75% training the model and 25% testing- are randomly selected. The DM techniques applied in this paper are Boruta algorithm, C5.0, Neural Network, Support Vector Machine, and random forest via open-source software R and WEKA.**

Keywords*—Quality of service; churn prediction; classification; data mining; prediction model; customer retention

I. Introduction

In the world of works, customers are the source of gain and revenue for the service of organizations and improvements in QoS that lead to customer loyalty. Organizations become increasingly customer-focused and meet their requirements. QoS is a very effective factor as it becomes equal difficult to please and preserve customers. Research indicates that both QoS and customer satisfaction are distinctive structures but effectively related. This is especially true for companies' service where a higher level of customer satisfaction leads to higher profits [1]. Those facts eventually focus on predicting customer churn as a necessary part of the Communication firms' procedures and decisions, which are the major goal of customer relationship management (CRM) also. The increasing of importance of this tool has led to the enhancement of many predicting tools that reinforce some pivotal tasks in the predictive modeling and operations of classification [2]. Identifying a predicted churn is a beneficial tool to predict a customer in danger of churn. In general, service providers are blamed for poor quality, but the real problem is the design of the service system. Predicting the level of service, good service planning, and knowing customer behavior and desires are way to improve the QoS [1]. The purpose of churn management is to decrease the loss of subscribers generally since subscribers raise profits by handing over a stable and profitable customer basis. [3]. Most Data Mining (DM) techniques play a very substantial role in telecom firms to enhance their marketing efforts, identify a scam, manage their telecom networks, demographic data, behavioral data, and many disciplines [4]. DM techniques are set in telecoms for CRM due to the fast growth of the huge quantity of data, high speed in the market rivalry, and rise in the churn rate. DM techniques can be used in the classification, clustering of customer data to predict churners. And have affected Genetic Algorithms (GA), Fuzzy Logic (FL), Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), and Neural Networks (NN) [5].

This paper is organized as follows: Section 2 demonstrates the definition of the impact of service quality on customer satisfaction. This section presents, also, the concept of churn prediction, and shows the types of churners. Section 3 gives an overview of DM techniques and used algorithms. While Section 4 describes the proposed model for churn prediction; in addition to the results proceed of a case study. Finally; the conclusion and future work.

II. Impact of Service Quality on Customer Satisfaction

1) Quality of services and Customer Satisfaction: QoS can be explained as superiority, measuring the ability of the service that reaches the customer and corresponds to his expectations. Also, it means that the organization's designs deliver the service correctly. Moreover, it enjoys competitive advantages. Perfect QoS leads to the retention of current customers, reduced costs, in addition to reinforcing customers 'loyalty' [6]. Customer satisfaction is a measure of the range of products and services that a company provides to meet customer expectations. Noticed, noticeable emergence of the term customer satisfaction, especially the increasing numbers of business organizations operating in the same sector with the expansion of local and global markets [7], [1].

2) Churn prediction and customer retention: Churn's prediction is used to identify potential churners, before they leave the firm. This helps CRM to prevent potential customers from leaving the company, in the future, via retention policies. Thus, the loss possibility of the firm can be averting [8]. In telecoms-based industries, companies supply customers with rewards to tempt them into switching to their services. Toward off this, the firm must grasp the reasons why the customer decides to depart to another telecom firm [9].

Customer retention is the primary goal of the CRM. It is taken to ensure customer loyalty and reduce churn to move to serve supporters for higher quality, better offers, or more advantages. For this target, churn prediction is an essential part of a proactive scheme to retain a customer [8], [9].

3) Type of churners there are two types of churners: voluntary and involuntary churners. Voluntary churn happens as a result of a customer's resolution to moving into another service provider or another company. Involuntary churn happens because of conditions, as a customer moving to longtime care, death, or moving to a remote place. Involuntary churners are considered the easiest type of churners to determine. The company can decide to exclude them from the subscribers' list. This denomination includes people, who are churned for the scam, not paying, not using the phone [3]. Voluntary churn can be split into two primary classes: "incidental" and "deliberate" churn. Incidental churn happens not for the customers' intention but the actual reason lies in something that has happened in their lifetimes. For example, Change in the financial situation, location. Deliberate change occurring technological reasons; customers wanting new or better technology, economy, price fluctuation, dissatisfaction with the QoS factors, too high prices, no rewarding for customer loyalty, bad support, Social or psychological elements, and/or amenities. Deliberate churn is the issue that the churn department tries to get for solutions (Fig. 1) [3]-[10].

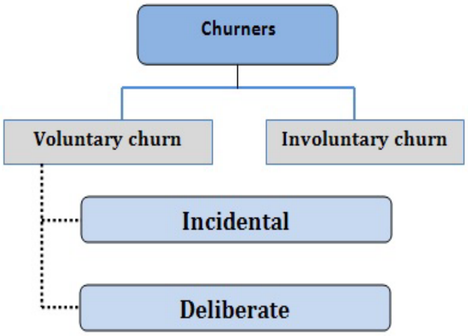


Fig. 1. Churn Classification.

III. Data Mining Techniques

Data mining looks for hidden and unexpected information, such as the optimum classification for customers. DM looks, also, unimportant trends and beneficial patterns in large data sets. DM relates to discovering unexpected or previously-unknown relationships between data. It is a multidisciplinary skill that uses machine learning, statistics, artificial intelligence technology, and database. Numerous main data mining technologies are enhanced and used in recent DM projects, such as association, classification, clustering, prediction, sequential patterns, and decision tree [10], [11].

1) The neural network: The neural network is a group of connected (I /O) input/output units and each communicates with a weight. Through the learning stage, the net learns by tuning weights even predicts correct row classifications for input groups. The NN has a noteworthy ability to extract meaning from complex or inaccurate data and can be used to extract patterns and reveal very complex directions that cannot be observed by humans, or computer technologies e.g., computer training to vocalize English text after reorganizing handwritten letters [9].

2) Random forest: Random forest is considered to be one of the most powerful techniques of machine learning, it almost does not need any data preparation or any modeling experiences, and it is, also, considered as a tool that embodies the power of decision trees in addition to wise randomness and collective learning to produce accurate predictive models, insightful classifications of the values of lost, and new divisions, to help understand the deepest data [11].

3) Support vector machine: Support vector machine is a powerful supervised learning method for regression and classification problems that makes expectations via a linear combination of kernel basis functions. SVM is an implementation of the structural risk minimization (SRM) principles which attempt to minimize an upper bound of the generalization error instead of minimizing the empirical error. Depending on these transformations to find the optimum border between the possible outputs; simply, they perform some very complex data conversions, then they reveal how data is separated based on the labels or outputs that are specified [12].

4) The decision tree: The decision tree is the technique of the most communally DM techniques where its model is very easy to grasp and realize for users. DT is a model that breeds a structure as the shape of a tree that exemplifies a group of decisions. The DT root is an easy question or status that has various answers. All answers lead to a set of questions or situations that help determine the data to be able to make the final decision. The decision trees are created by the C5.0 algorithm work almost well, but are easy to understand and spread [13].

IV. Proposed Model

The suggested model consists of many steps: the first step is to define the problem and data selection, the next step is data pre-processing (data transformation, data cleaning), the next step is feature selection. Feature selection methods are used to eject the iterative and not relevant features that do not contribute significantly to predict performance. The next step sampling stage, represented in the training set, is applied to train our classifier; Moreover, a predicting model builds to be ripe for testing. The last step is the performance and accuracy evaluation, the prediction Outputs and Results. Fig. 2 shows the proposed churn prediction model.

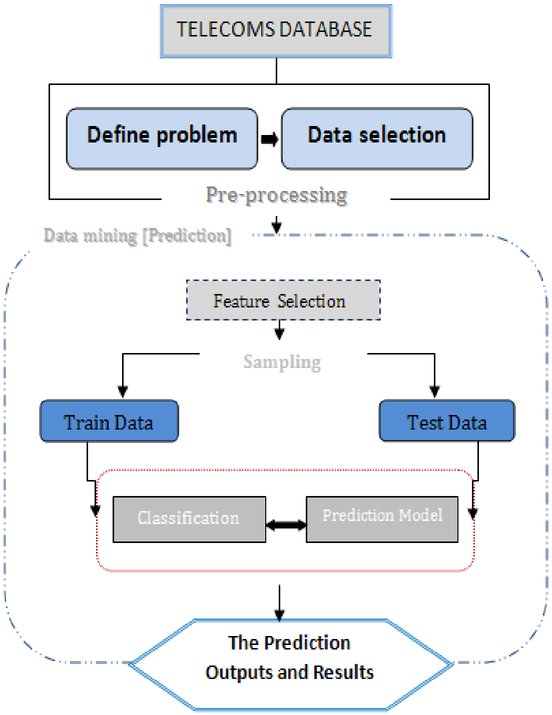


Fig. 2. The Proposed Churn Prediction Model.

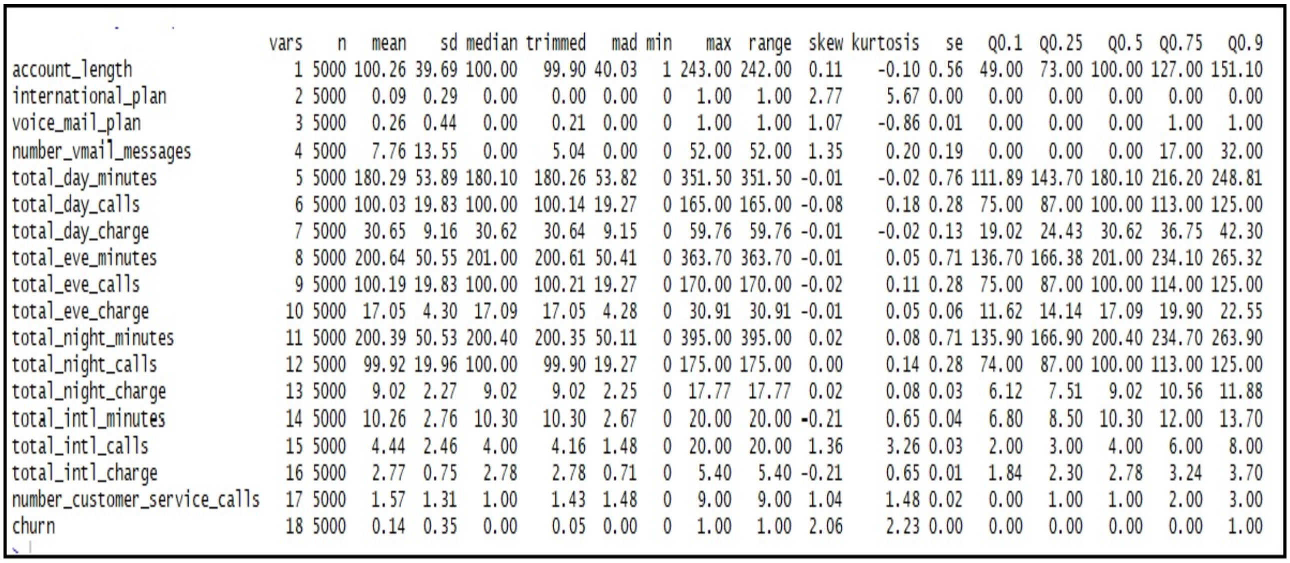
1) The initial stride: Prior to applying specific analytical models to the data, which is previously prepared to be more appropriate for the analysis, by defining the problem, data selection, and analyzing data explicatory?

2) Preparing data: The data, used in this study, is a telecom company database containing a set of statistical data for customers that are 17 illustrative features regarding the use of customers' service daytime, Intl-calls, and customer-service-calls 14.14% of notes have a variable target of "true", and 85.86% of notes have a "false" value. Viewing data set variables of "customer transactions", and their characterization is manifested in Fig. 3, and Table I demonstrates the distribution of every feature and describing a set of data.

Description of the quantiles of a data set is the numbers whose percentiles are the quarter marks of the data set. Specifically, they are the values in the data set that are at 1%, 25%, 50%, 75%., and 90% these are also known as a quartile , mean usually what one means by an average - the sum of dataset, mean usually what one means by an average - the sum of dataset. Standard Deviation (sd) contains percentiles of the measurements, mad = the median absolute deviation (from the median), trimmed = a trimmed mean (by default in this function, this removes the top and bottom 10% from the data, then computes the mean of the remaining values - the middle 80% of the full data set), The range of a quantitative variable is interpreted as the difference between the (maximum (max)and the minimum(min)), a measure of skew, which refers to how much asymmetry is present in the shape of the distribution. a measure of excess kurtosis, which refers to how outlier-prone, or heavy-tailed the shape of the distribution is, as compared to a Normal distribution, se = the standard error of the sample mean, equal to the sample (sd) divided by the square root of the sample size.as show in Fig. 3.

a) Data Transformation: Converting the explicatory variables (Intl-plan, voicemail-plan) to a binary form to be more appropriate (yes = 1 /no = 0) in a specific model.

b) Data clean: This stage involves processing /calculating lost data: Some specified algorithms cannot handle lost data like SVM. For this reason, the lost value can be alternated with average or (0). Yet, it is better option to replace the lost data by the calculated statistical value (computation option), including the data set utilized to the lost values in several variables numerical (total-day-charge, total-evening-minutes, total Int'l-calls, total Int'l-charges, and total nighttime-charges), and nominal variables (Voicemail-plan, Intl-plan). Numerical data is alternated using various RF techniques. Bilateral values are calculated via techniques [13], [14].



```
                            vars    n   mean    sd median trimmed   mad min    max  range  skew kurtosis   se   Q0.1  Q0.25   Q0.5  Q0.75   Q0.9
account_length                 1 5000 100.26 39.69 100.00   99.90 40.03   1 243.00 242.00  0.11    -0.10 0.56  49.00  73.00 100.00 127.00 151.10
international_plan             2 5000   0.09  0.29   0.00    0.00  0.00   0   1.00   1.00  2.77     5.67 0.00   0.00   0.00   0.00   0.00   0.00
voice_mail_plan                3 5000   0.26  0.44   0.00    0.21  0.00   0   1.00   1.00  1.07    -0.86 0.01   0.00   0.00   0.00   1.00   1.00
number_vmail_messages          4 5000   7.76 13.55   0.00    5.04  0.00   0  52.00  52.00  1.35     0.20 0.19   0.00   0.00   0.00  17.00  32.00
total_day_minutes              5 5000 180.29 53.89 180.10  180.26 53.82   0 351.50 351.50 -0.01    -0.02 0.76 111.89 143.70 180.10 216.20 248.81
total_day_calls                6 5000 100.03 19.83 100.00  100.14 19.27   0 165.00 165.00 -0.08     0.18 0.28  75.00  87.00 100.00 113.00 125.00
total_day_charge               7 5000  30.65  9.16  30.62   30.64  9.15   0  59.76  59.76 -0.01    -0.02 0.13  19.02  24.43  30.62  36.75  42.30
total_eve_minutes              8 5000 200.64 50.55 201.00  200.61 50.41   0 363.70 363.70 -0.01     0.05 0.71 136.70 166.38 201.00 234.10 265.32
total_eve_calls                9 5000 100.19 19.83 100.00  100.21 19.27   0 170.00 170.00 -0.02     0.11 0.28  75.00  87.00 100.00 114.00 125.00
total_eve_charge              10 5000  17.05  4.30  17.09   17.05  4.28   0  30.91  30.91 -0.01     0.05 0.06  11.62  14.14  17.09  19.90  22.55
total_night_minutes           11 5000 200.39 50.53 200.40  200.35 50.11   0 395.00 395.00  0.02     0.08 0.71 135.90 166.90 200.40 234.70 263.90
total_night_calls             12 5000  99.92 19.96 100.00   99.90 19.27   0 175.00 175.00  0.00     0.14 0.28  74.00  87.00 100.00 113.00 125.00
total_night_charge            13 5000   9.02  2.27   9.02    9.02  2.25   0  17.77  17.77  0.02     0.08 0.03   6.12   7.51   9.02  10.56  11.88
total_intl_minutes            14 5000  10.26  2.76  10.30   10.30  2.67   0  20.00  20.00 -0.21     0.65 0.04   6.80   8.50  10.30  12.00  13.70
total_intl_calls              15 5000   4.44  2.46   4.00    4.16  1.48   0  20.00  20.00  1.36     3.26 0.03   2.00   3.00   4.00   6.00   8.00
total_intl_charge             16 5000   2.77  0.75   2.78    2.78  0.71   0   5.40   5.40 -0.21     0.65 0.01   1.84   2.30   2.78   3.24   3.70
number_customer_service_calls 17 5000   1.57  1.31   1.00    1.43  1.48   0   9.00   9.00  1.04     1.48 0.02   0.00   1.00   1.00   2.00   3.00
churn                         18 5000   0.14  0.35   0.00    0.05  0.00   0   1.00   1.00  2.06     2.23 0.00   0.00   0.00   0.00   0.00   1.00
```

Fig. 3. Description of the Data Set (Data Set Variables of Customer Transactions).

TABLE I. ATTRIBUTES OF THE DATA SET

Data Type	**Description**	**Variable**
Account Length	Customer subscription period by weeks	Integer
International plan	participation in the international plan (yes, no)	Categorical
Voice. Mail plan	participation voicemail plan (yes, no)	Categorical
Vmail. messages	No. of voicemails	Integer
Total. Day minutes	Total daily minutes	Integer
Total. Day. calls	Total of calls that was during the day	Integer
Total. Day. charge	Total daily charge	Integer
Total. Eve. minutes	Total eve minutes used	Integer
Total. Eve. calls	Total calls in the evening used	Integer
Total. Eve. charge	Total evening call charges	Integer
Total.night.minutes	Total nighttime minutes used	Integer
Total.night.calls	Total nighttime calls used	Integer
Total.night.charge	Total nightly calls charges	Integer
Total.intl.minutes	total International minutes usage	Integer
Total.intl.calls	Total International calling used	Integer
Total.intl.charge	Total invoice international charges	Integer
Customer.service.calls	No. of calls used serving of customers	Integer
Churn	Customer's case (True = churn, False = no churn)	Categorical

3) Feature selection: Feature selection is one of the most vital elements that can influence model performance. It is also, the operation in which actions are performed automatically or manually, and features that contribute most to the variable or the prediction result. Inappropriate data features can reduce the accuracy of forms and make the model based on irrelevant features. The selection feature helps to give a clearer understanding of the data by identifying the important features of the data and their relationship to each other. In this study, to assess the relationship between each input variable and the target variable, these scores are applied as a basis for selecting (filtering) and arranging the most important variables and reducing dimensions. The RF technique that is used in selection the feature using the average accuracy decrease. Average deficiency means that each feature affects the accuracy of the model. The model allows the values of every feature and estimates the model's accuracy change. Features that have a high effect on accuracy are only important [15], [16]. Technique known as "Boruta" is used for other feature selection. It is amelioration on RF, in which all the features have to be linked to the target variable, while most technologies are following the minimum optimization method interplay among features. Both techniques are used to rank predictors depending on the mean significance of Boruta, and the average decreasing error calculated by RF. The results, indicated in Table II, clarify that the initial three variables correspond to the identical rank (customer-serve-calls, Intl-plan, total-day-minutes). Models agreement is the following ten features of various ranks: (total day-charges, Total-Intl-calls, Total-Int'l-minutes, Total-Int'l-charges, voicemail-plan, total-evening-charges, total-evening-minutes, v-mail-messages, total-nighttime-minutes, and total-nighttime-charges). Models show a low ranking for the remaining variables (total-day-calls, total-nighttime-calls, total-evening-calls, and account-length). Outcomes are indicated in Table II and Fig. 4. Where 13 variables were confirmed and 4 variables rejected.

TABLE II. FEATURES MEAN IMPORTANCE (DESCRIPTION STATISTICS OF NUMERICAL VARIABLES IN THE DATABASE)

Feature	**Mean Import**	**Median Import**	**Min Import**	**Max Import**	**Norm Hits**	**Decision**
customer_service_calls	**79.7030451**	**79.8658717**	**76.297333**	**82.8529157**	**1**	**Confirmed**
international_plan	**65.5254351**	**66.0630689**	**62.212469**	**68.7058141**	**1**	**Confirmed**
total_day_minutes	**39.9455885**	**40.0260729**	**36.90489**	**42.4116863**	**1**	**Confirmed**
total_day_charge	**38.8683454**	**39.0927738**	**36.539083**	**40.806327**	**1**	**Confirmed**
total_intl_calls	**37.9153817**	**37.3873506**	**35.874353**	**41.182217**	**1**	**Confirmed**
total_intl_minutes	**22.8759486**	**22.7362083**	**22.02165**	**24.0693349**	**1**	**Confirmed**
total_intl_charge	**22.7810105**	**22.9715311**	**20.817315**	**24.2876061**	**1**	**Confirmed**
voice_mail_plan	**22.6492168**	**22.8422342**	**20.877226**	**23.6341128**	**1**	**Confirmed**
total_eve_charge	**22.1410863**	**21.8842483**	**20.407409**	**23.6509283**	**1**	**Confirmed**
total_eve_minutes	**21.8988284**	**21.8339512**	**20.498661**	**23.4987201**	**1**	**Confirmed**
number_vmail_messages	**21.8944202**	**21.3996754**	**21.096311**	**24.4659916**	**1**	**Confirmed**
total_night_minutes	**13.4130025**	**13.4427115**	**12.759512**	**14.3821655**	**1**	**Confirmed**
total_night_charge	**13.0082163**	**12.924208**	**12.121582**	**14.3744161**	**1**	**Confirmed**
total_day_calls	**-0.5188858**	**-0.4468013**	**-1.95245**	**0.6344864**	**0**	**Rejected**
total_night_calls	**-0.6262422**	**-0.3491468**	**-2.672444**	**0.874702**	**0**	**Rejected**
total_eve_calls	**-1.3878873**	**-1.1726735**	**-2.787194**	**-0.175634**	**0**	**Rejected**
account length	**-1.0827182**	**-1.1550063**	**-2.631888**	**0.9322869**	**0**	**Rejected**

Features Importance processing of finding and specifying the most beneficial features in a data set. The top essential variables are from the top tier of boruta’s selections and features for RF models. In Fig. 4 show in mean decrease accuracy, it indicates that the right of the blue line is the number of customer service calls, international plan, and total daily minutes.

This means that these are the most important factors in determining customer churn, it is logical the customer, who has to receive many customer service calls to resolve a problem, may become disappointed and leave his business with the company [16] shown in Table II and Fig. 4, is also illustrated cases in Fig. 5.

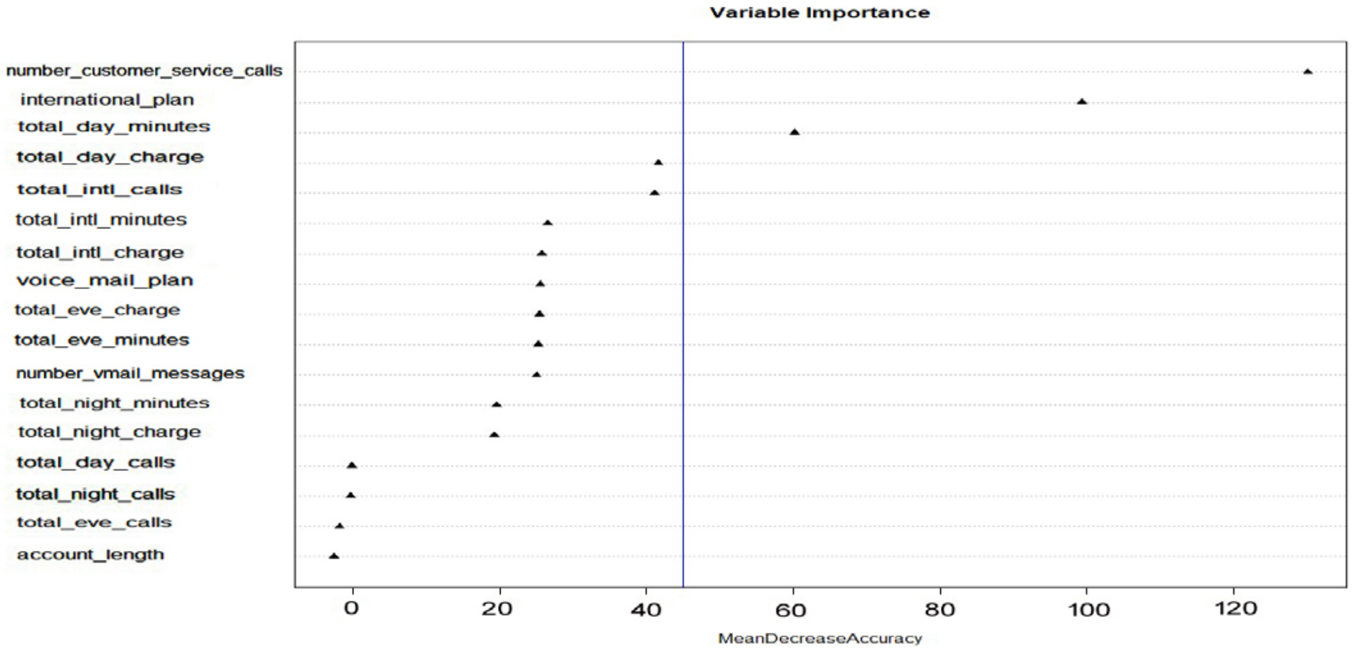
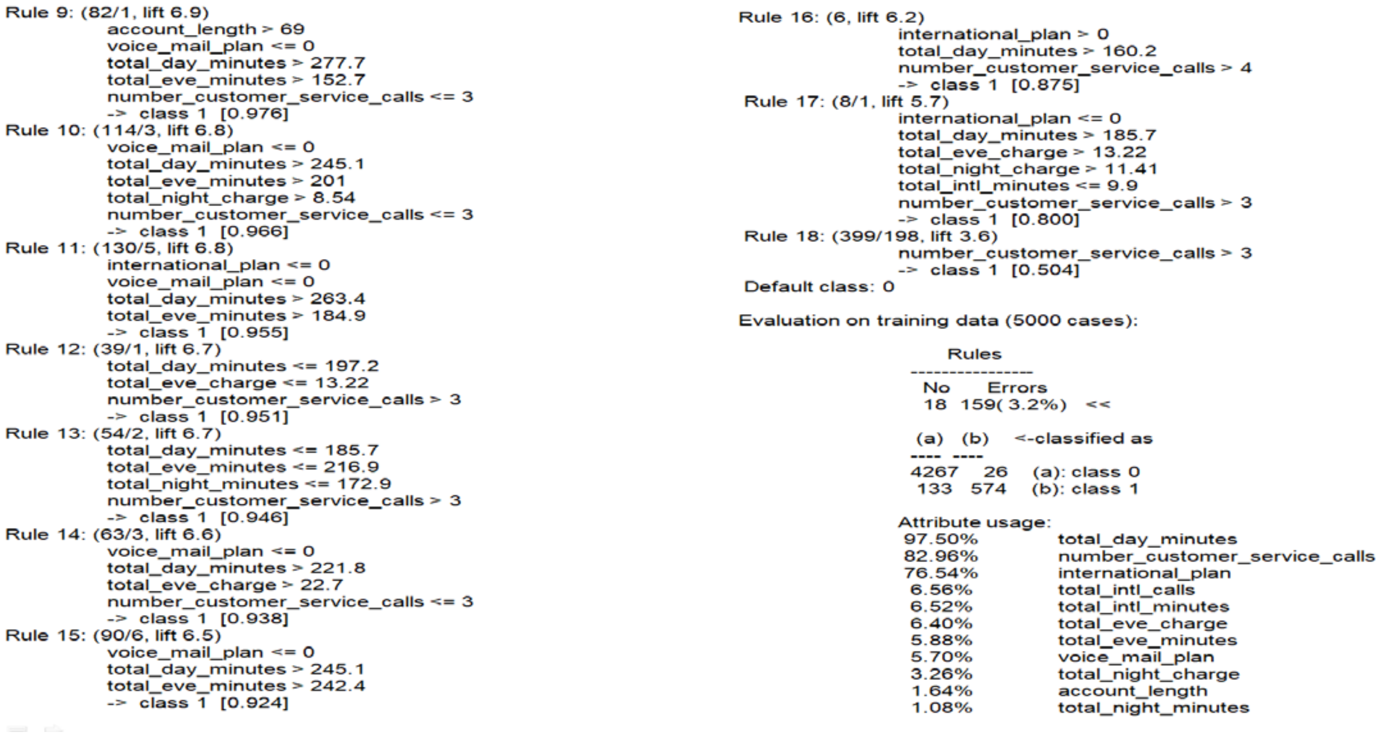


Fig. 4. Features mean Importance (Mean Decrease Accuracy).



```
Rule 9: (82/1, lift 6.9)
                account_length > 69
                voice_mail_plan <= 0
                total_day_minutes > 277.7
                total_eve_minutes > 152.7
                number_customer_service_calls <= 3
                -> class 1 [0.976]
Rule 10: (114/3, lift 6.8)
                voice_mail_plan <= 0
                total_day_minutes > 245.1
                total_eve_minutes > 201
                total_night_charge > 8.54
                number_customer_service_calls <= 3
                -> class 1 [0.966]
Rule 11: (130/5, lift 6.8)
                international_plan <= 0
                voice_mail_plan <= 0
                total_day_minutes > 263.4
                total_eve_minutes > 184.9
                -> class 1 [0.955]
Rule 12: (39/1, lift 6.7)
                total_day_minutes <= 197.2
                total_eve_charge <= 13.22
                number_customer_service_calls > 3
                -> class 1 [0.951]
Rule 13: (54/2, lift 6.7)
                total_day_minutes <= 185.7
                total_eve_minutes <= 216.9
                total_night_minutes <= 172.9
                number_customer_service_calls > 3
                -> class 1 [0.946]
Rule 14: (63/3, lift 6.6)
                voice_mail_plan <= 0
                total_day_minutes > 221.8
                total_eve_charge > 22.7
                number_customer_service_calls <= 3
                -> class 1 [0.938]
Rule 15: (90/6, lift 6.5)
                voice_mail_plan <= 0
                total_day_minutes > 245.1
                total_eve_minutes > 242.4
                -> class 1 [0.924]
```

```
Rule 16: (6, lift 6.2)
                international_plan > 0
                total_day_minutes > 160.2
                number_customer_service_calls > 4
                -> class 1 [0.875]
 Rule 17: (8/1, lift 5.7)
                international_plan <= 0
                total_day_minutes > 185.7
                total_eve_charge > 13.22
                total_night_charge > 11.41
                total_intl_minutes <= 9.9
                number_customer_service_calls > 3
                -> class 1 [0.800]
 Rule 18: (399/198, lift 3.6)
                number_customer_service_calls > 3
                -> class 1 [0.504]
 Default class: 0

Evaluation on training data (5000 cases):

                      Rules
                 ----------------
                   No      Errors
                   18  159( 3.2%)   <<

                  (a)   (b)    <-classified as
                 ----  ----
                 4267    26    (a): class 0
                  133   574    (b): class 1

                Attribute usage:
                 97.50%          total_day_minutes
                 82.96%          number_customer_service_calls
                 76.54%          international_plan
                  6.56%          total_intl_calls
                  6.52%          total_intl_minutes
                  6.40%          total_eve_charge
                  5.88%          total_eve_minutes
                  5.70%          voice_mail_plan
                  3.26%          total_night_charge
                  1.64%          account_length
                  1.08%          total_night_minutes
```

Fig. 5. The Result of Churners.

4) Calculation of variable importance: This step for sampling training and testing through this stage predicts uses the dataset containing 5000 records. A data set with 17 attributes 75% is used to train the model and 25% are applied as testing. Training is used to explore data and build models, while testing will be used to measure model performance. For significance grades generated from varImp.train, the drawing method can be used to visualize results. Variable importance is only 5 of the 17 features are utilized by "rpart". The tree structure shown here provides a neat and easy-to-follow description of the problem under consideration and its solution. Note that the implementation of R for the CART algorithm is called RPART (Recursive Partitioning and Regression Trees). Dataset and constructing an algorithm, used in the training, can be used to calculate the variable importance (varImp), model. The varImp is, then, used to value the different importance, which is plotted. It shows that the number of customer-service-calls, international-plan, total-day-minutes, total-day-charger, and voicemail-plan attributes are the top 5 most essential attributes. Ranks of features variable importance is shown in Fig. 6 and 7.

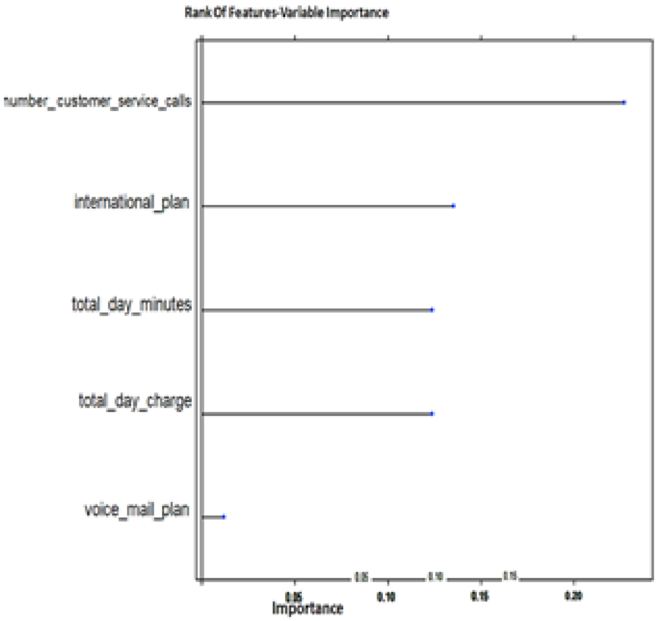


Fig. 6. The Rank of Features Variable Importance.

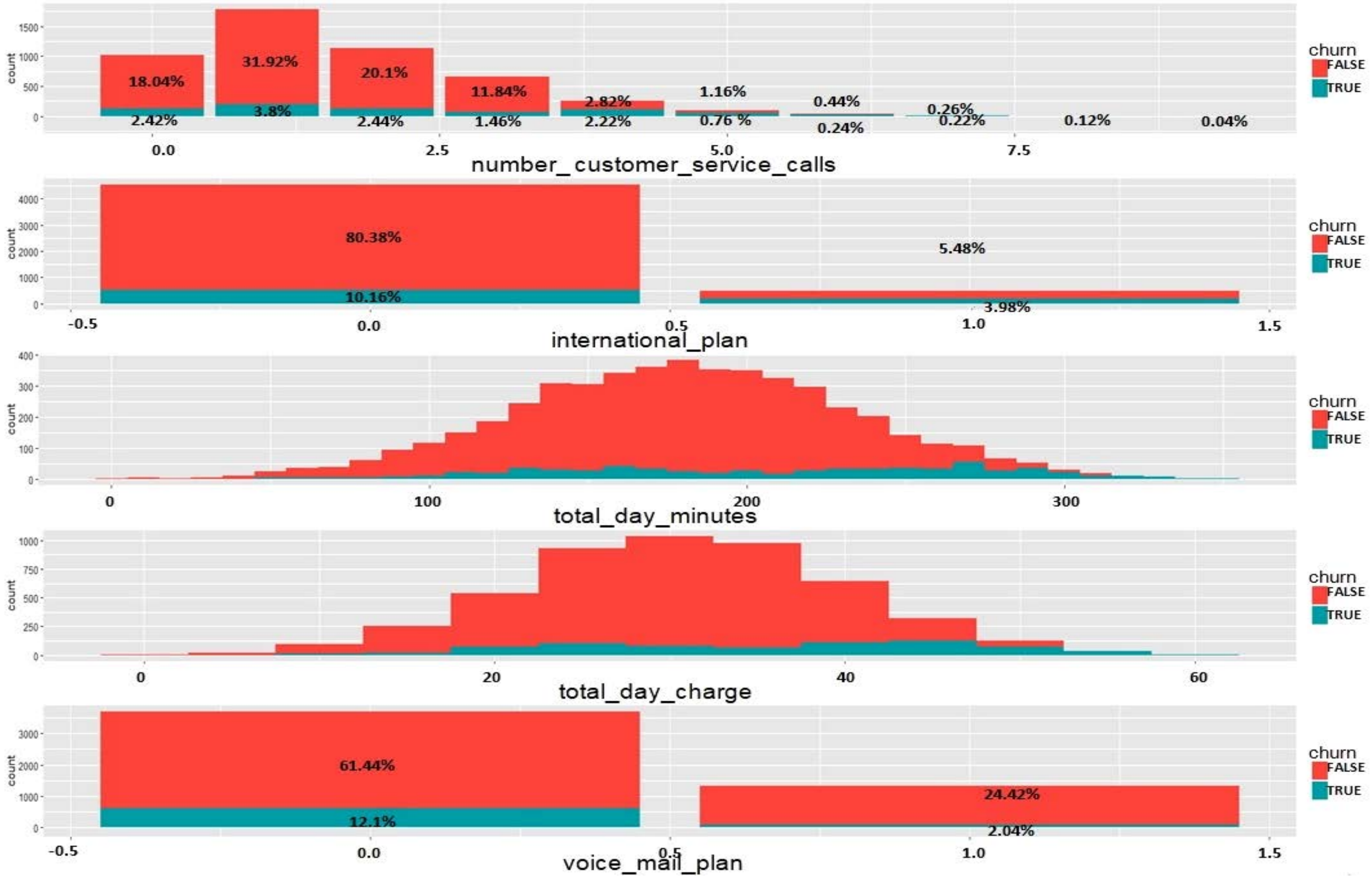


Fig. 7. Variables Importance Distribution (Ranks of Features Variables Distribution).

b) Data visualization of descriptive statistics: After taking various steps including calculating variable importance, descriptive statistics for data are examined under some simple plots, which define several important predictive variables for the model and which represent the five variables shown in Fig. 6. Fig. 7 shows illustrates the distribution of this feature concerning the churn class, the churn rate via the categorical predictors (red = false) and (blue = true). While the percentage of change for customers, who do not use voicemail plans, seems higher, the rate of change in the international plan is 4 times the average, and there is a higher increase in the rate of calls with the increase of customer calls exceeding 4 customer calls. Many churning customers use more "minutes of the time of day" while charges are high. It also shows (increases and decreases) between churners' factors, find an increase in the number of calls to customer of service and a rise in using international plan, increases charge process daily, daily contact ,and Less use on voicemail.

c) Neural Networks: Neural networks are built via 17 inputs, 2 outputs, and 1 hidden layer with 6 neurons. The algorithm that is used in neural net by default is based on the resilient back-propagation without weight backtracking, to perform a classification (linear. Output = FALSE) in this case.

d) Support Vector Machine: Support vector machine builds and specifies a type for c-classification, in addition to using two extra parameters: gamma and cost. Gamma is the argument used by the function kernel= radial "the default". Test with various gamma and cost values to find better classification accuracy. Whereas, actual support offer1018 detects the model support vectors dispenser in the categories (false for 585 and 433 for true).

e) Random forest: Forests of 500 DT are built by using RF 100 trees; they do not decrease in the error. Besides, the parameter is metric, which refers to the number of predictors that are sampled to divide by every node. Optimum performance at metric Mtry= 5.

f) Results and discussion: A dataset of 5,000 customers dissected, included 707 churnings. The results reveal that there are five correlations between customers' behavior and variables that affect the QoS, which affects customer loyalty. A dataset was applied to an available dataset of publicly, obtained from the KDD library. The data is acquired from a wireless communications operator. For a complete description of the dataset, one can refer to D.Larose (2014) [17].

1) Number customer service calls: About 50% of customers who contact the company more than 3 times are classified as "Churn" while only 5 % of "non-churners" Called for service more than three times, this reveals that 27 % of churners have done so. The company must monitor the QoS in its call centers in an attempt to solve technical problems in less than three calls. Besides, customers who call more than three times should be noteworthy. As shown in Fig. 3 and 7.

2) International plan: It reveals that customers who have an international plan quadruple their odds to be churn. The company should examine its Intl plan and its suitability for customer needs. For every unit of increase in international fees, there is an 18% increase in the possibility of high rate (leaving the company) 10% of those without an international plan are classified "Churn" and 8% of those who own an international plan is churn. A high rate of churn among customers has international plans. The count of customer service calls, the international plan, and the total of international calls. This means that it is the most important factor that determines churn. The customer who has to receive many customer service calls to solve a problem is likely to become frustrated. And leave his dealings with the company. This is shown in Fig. 3 and 7.

3) Total day minutes and total day charge: Total day minutes & Total day charge: Apparent that many churn customers use more "daytime minutes", and as an outcome for paying high invoices. They are "the significant use". These are the company's most values customers because of the high fees they pay. Higher charges may push these customers to find a cheaper plan or other providers. Targeting these higher in-use customers by promoting the best prices may be an incentive for them to remain as customers. This is shown in Fig. 3 and 7.

4) Total day minutes and total eve minutes: If the whole minutes per day exceed 277.7, and the total evening minutes are 152.7, the customer is probable to depart the company. This is shown in rule 9 in Fig. 5. High churn rate among intense time users today, it is probably customers who pay high bills and some of these customers seek to find cheaper options. Show in Fig. 5.

5) Voice- mail- plan: customers who have a voicemail plan are less probable to "churn" than customers without a voicemail plan. This is shown in Fig. 3 and 7.

The proposed model categorizes churn customer data by using classification algorithms to identify the root causes of amplification [18]. By knowing the important churners factors from customer the data. CRM can improve productivity; recommend related promotions to set of potential churners [19], [20] Results have been discussed based on the performance of three classifiers; NN, SVM, and RF were evaluated as confusion matrix mentioned in the basic abbreviations used in confusion matrix (Table III).

Accuracy defines the percentage of rightly classified cases from test assigned by the classifier [11] - [13].

$$\text{Accuracy} = \frac{\text{tp+tn}}{\text{tp+tn+fp+fn}} \quad (1)$$

*(TP)*Are statuses where predicted churners by correctly “churn"

(FP) Are statuses where the predicted churners by incorrectly “churn"

(TN) Are statuses where the predicted churners by correctly "non-churn"

(FN) Are statuses where the predicted churners by incorrectly “non-churn"

The results performed by the tool R, considering the NN, SVM, and RF techniques, are shown in Table IV. The results performed by the WEKA tool considering the NN, SVM, and RF techniques are shown in table 5. The summary of the total results of tools and techniques are shown in Table VI.

Table VI compares the accuracy of the R and WEKA tools considering the NN, SVM, and RF techniques. For the NN technique, the accuracy of the R tool is 96.9%, which shows better results than the WEKA tool that is 95.92%. Considering the SVM technique, the accuracies of the WEKA and R tools are 94.8% and 92.1%, respectively. That means the WEKA tool outperforms the R tool in such a case. For the RF technique, the accuracy of the R tool is 96.4%, which shows better results than that of the WEKA tool that is 95.44%.

TABLE III. CONFUSION MATRIX

Confusion matrix	**Predicted value**	**Predicted value**
Actual	True	False
True	TP	FP
False	FN	TN

TABLE IV. CONFUSION MATRIX R TOOL

Tool and Techniques using		**Actual Class**	**Actual Prediction**	
			Non-churners	*Churners*
R Tool	NN	*Non- churners*	1405	11
		Churners	39	211
	SVM	*Non- churners*	1417	79
		Churners	20	150
	RF	*Non- churners*	1078	35
		Churners	10	127

TABLE V. CONFUSION MATRIX WEKA TOOL

Tool and Techniques using		**Actual Class**	**Actual Prediction**	
			Non-churners	*Churners*
Weka Tool	NN	*Non- churners*	1059	11
		Churners	40	140
	SVM	*Non- churners*	1403	20
		Churners	65	178
	RF	*Non- churners*	1063	7
		Churners	50	130

TABLE VI. ACCURACY, ERROR RATES COMPARISON FOR R TOOL AND WEKA TOOL OF TECHNIQUES

Tool	**Technique**	**Accuracy**	**Error rate**
R	NN	96.9%	3.1 %
	SVM	92.1%	7.9%
	RF	96.4%	3.6%
WEKA	NN	95.92	4.08 %
	SVM	94.8%	5.2%
	RF	95.44	4.56%

V. CONCLUSIONS AND FUTURE WORK

This study manifests an effective methodology to expect fluctuations in industries depending on customer service; the customer is affected by the quality of service, and that the QoS is one of the most important factors affecting the survival and continuation of the customer. Besides, the higher costs, associated with purchasing new customers, highlight the need for telecom operators to identify the churn to reduce costs, increase revenue, and analyze case types of churners. The model depending on DM techniques is offered to aid in the CRM management tracking its customers and their conduct versus disturbances Using 3 various techniques predicting NN, SVM, and RF for classification results hint that the best output for the data set used is the NN technique. In the future, the present methodology can be used to modern data sources such as flowing data in real-time to obtain a prediction of churn in real-time and would be more fit for data-based industries. The idea of churn prediction can be expanded to include other areas such as employee churn, drop-out customers' expectations. Customers can also learn about the best services suitable for them through a detailed stride-by-stride guide without communicating with the service providers.

REFERENCES

[1] Qadeer, Sara, “Service Quality & Customer Satisfaction: A case study in Banking Sector,” pp.1-101, 2014.

[2] Amin, A., Anwar, S., Adnan, A., Nawaz, M., Alawfi, K., Hussain, A., & Huang, K., “Customer churn prediction in the telecommunication sector using a rough set approach,”Neurocomputing, 237, pp. 242-254, 2017.

[3] E. Shaaban, Y. Helmy, A. Khedr, and M. Nasr, “A proposed churn prediction model,” International Journal of Engineering Research and Applications, Vol. 2, No. 4, pp. 693 - 697, 2012.

[4] K. Kaur and S. Vashisht, “Enhanced Boosted Trees Technique for Customer Churn Prediction Model,” IOSR Journal of Engineering (IOSRJEN), Vol. 04, Issue 03, pp 41-45, 2014.

[5] KiranDahiya ,Surbhi Bhatia “Customer Churn Analysis in Telecom Industry,” 4th InternationalConference on Reliability, InfocomTechnologies and Optimization (ICRITO) (Trends and Future Directions), pp. 1-6,2015.

[6] Oghojafor , Benjamin &Bakarea , Rasaki&Omoera, Charles &Adeleke, I.A, “Discriminant Analysis of Factors Affecting Telecoms Customer Churn,” International Journal of Business Administration 3(2),pp.59-67, 2012.

[7] Sidra Ansar, co authorSamreenLodhi, “the impact of service quality on customer satisfaction in telecom sector of Pakistan. An empirical study of Pakistan,” International Journal of Scientific & Engineering Research, Volume 6, Issue 10, pp.1639-1645, 2015.

[8] ManpreetKaur, Dr. PrernaMahajan, “Churn Prediction in Telecom Industry Using R,” International Journal of Engineering and Technical Research (IJETR) ISSN: 2321-0869, Volume-3, Issue-5, pp.46-53, 2015.

[9] Bharati, M. &Ramageri, “data mining technique applications,” Indian Journal of Computer Science and Engineering, Vol. 1 No. 4, pp. 301-305, 2010.

[10] Mamčenko, J. &Gasimov, J., “Customer churn prediction in mobile operator using combined model,” ICEIS 2014 - Proceedings of the 16th International Conference on Enterprise Information Systems. 1, pp. 233-240, 2014.

[11] Jayaswal, Pretam& Prasad, Bakshi&Tomar, Divya&Agarwal, Sonali. , “An Ensemble Approach for Efficient Churn Prediction in Telecom Industry,” International Journal of Database Theory and Application. 9, pp. 211-232, 2016.

[12] Brandusoiu, Ionut&Toderean, G., “Churn Prediction in the Telecommunications Sector Using Support Vector Machines,” Annals of the Oradea University. Fascicle of Management and Technological Engineering. XXII (XII), pp.19-22, 2013.

[13] I. I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam and S. W. Kim, "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector," in *IEEE Access*, vol. 7, pp. 60134-60149, 2019.

[14] Shah AD, Bartlett JW, Carpenter J, Nicholas O, Hemingway H., “Comparison of random forest and parametric imputation models for imputing missing data using MICE: a CALIBER study,” Am J Epidemiol. 15; 179(6):764-74, 2014.

[15] Oralhan, Burcu&Uyar, Kumru& ORALHAN, Zeki, “Customer Satisfaction Using Data Mining Approach,” International Journal of Intelligent Systems and Applications in Engineering. 4. 63-63, 2016.

[16] F., Sahar. “Machine-Learning Techniques for Customer Retention: A Comparative Study,” International Journal of Advanced Computer Science and Applications. Vol. 9, No. 2, pp.273-281, 2018.

[17] Daniel T. Larose and Chantal D. Larose “Discovering Knowledge in Data: An Introduction to Data Mining, Second Edition”,pp.1-366,2014

[18] Ahmad, A.K., Jafar, A. &Aljoumaa, K., “Customer churn prediction in telecom using machine learning in big data platform,” Journal of Big Data, 6(1), pp.1-24, 2019.

[19] Abdulrahman, S. A., Khalifa, W., Roushdy, M., & Salem, A.-B. M., “Comparative study for 8 computational intelligence algorithms for human identification. Computer Science Review,” 36, 100237, pp.1-11, 2020.

[20] HomaMeghyasi and AbasRad, “Customer Churn Prediction in Irancell Company Using Data Mining Methods,” EasyChair Preprint No. 2422, pp.1-6, 2020.