IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 9, 2019, accepted April 30, 2019, date of publication May 6, 2019, date of current version May 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2914999

A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector

IRFAN ULLAH[1], BASIT RAZA[1], AHMAD KAMRAN MALIK[1], MUHAMMAD IMRAN[1], SAIF UL ISLAM[2], AND SUNG WON KIM[3]

[1]Department of Computer Science, COMSATS University Islamabad (CUI), Islamabad 45550, Pakistan
[2]Department of Computer Science, Dr. A. Q. Khan Institute of Computer Sciences and Information Technology, Rawalpindi 47320, Pakistan
[3]Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38542, South Korea

Corresponding authors: Basit Raza (basit.raza@comsats.edu.pk) and Sung Won Kim (swon@yu.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF), through the Brain Korea 21 Plus Program, under Grant 22A20130012814, in part by the Ministry of Science and ICT (MSIT), South Korea, through the Information Technology Research Center (ITRC) support program supervised by the Institute for Information and Communications Technology Planning and Evaluation (IITP), under Grant IITP-2019-2016-0-00313, in part by the National Research Foundation of Korea (NRF), Ministry of Education, through the Basic Science Research Program, under Grant 2018R1D1A1A09082266, and in part by the COMSATS University Islamabad (CUI), Islamabad, Pakistan, through Research Productivity Funds, under Grant CUI/ORIC-PD/19.

ABSTRACT In the telecom sector, a huge volume of data is being generated on a daily basis due to a vast client base. Decision makers and business analysts emphasized that attaining new customers is costlier than retaining the existing ones. Business analysts and customer relationship management (CRM) analyzers need to know the reasons for churn customers, as well as, behavior patterns from the existing churn customers' data. This paper proposes a churn prediction model that uses classification, as well as, clustering techniques to identify the churn customers and provides the factors behind the churning of customers in the telecom sector. Feature selection is performed by using information gain and correlation attribute ranking filter. The proposed model first classifies churn customers data using classification algorithms, in which the Random Forest (RF) algorithm performed well with 88.63% correctly classified instances. Creating effective retention policies is an essential task of the CRM to prevent churners. After classification, the proposed model segments the churning customer's data by categorizing the churn customers in groups using cosine similarity to provide group-based retention offers. This paper also identified churn factors that are essential in determining the root causes of churn. By knowing the significant churn factors from customers' data, CRM can improve productivity, recommend relevant promotions to the group of likely churn customers based on similar behavior patterns, and excessively improve marketing campaigns of the company. The proposed churn prediction model is evaluated using metrics, such as accuracy, precision, recall, f-measure, and receiving operating characteristics (ROC) area. The results reveal that our proposed churn prediction model produced better churn classification using the RF algorithm and customer profiling using *k-means* clustering. Furthermore, it also provides factors behind the churning of churn customers through the rules generated by using the attribute-selected classifier algorithm.

INDEX TERMS Churn prediction, retention, telecom, CRM, machine learning.

I. INTRODUCTION

In the present world, a huge volume of data is being generated by telecom companies at an exceedingly fast rate. There is a range of telecom service providers competing in the market to increase their client share. Customers have multiple options in the form of better and less expensive services. The ultimate goal of telecom companies is to maximize their profit and stay alive in a competitive market place [1]. A customer churn happens when a vast percentage of clients are not satisfied

The associate editor coordinating the review of this manuscript and approving it for publication was Tariq Umer.

2169-3536 © 2019 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

with the services of any telecom company. It results in service migration of customers who start switching to other service providers.

There are many reasons for churning. Unlike postpaid customers, prepaid customers are not bound to a service provider and may churn at any time. Churning also impacts the overall reputation of a company which results in its brand loss. A loyal customer, who generates high revenue for the company, gets rarely affected by the competitor companies. Such customers maximize the profit of a company by referring it to their friends, family members and colleagues. Telecom companies consider policy shift when the number of customers drops below a certain level which may result in a huge loss of revenue [3].

Churn prediction is vital in the telecom sector as telecom operators have to retain their valuable customers and enhance their *Customer Relationship Management (CRM) administration* [5], [6]. The most challenging job for *CRM* is to retain existing customers [7]. Due to the saturated and competitive market, customers have the option to switch to other service providers. Telecom companies have developed procedures to identify and retain their customers as it is less expensive than attracting the new ones [5]. This is due to the cost involved in advertisements, workforce, and concessions which can scale up to almost five to six times than retaining existing customers [3]. Small attention is needed for identifying the existing churn customers, which can help in overturning the situation. The requirement of retaining customers needs to develop an accurate and high-performance model for identifying churn customers. The proposed model should have the capability to identify churn customers and then find the reasons behind churn to avoid loss of customers and provide measures to retain them. In addition, it should employ techniques to predict when such a situation is going to arise in the future.

Due to recent advancements in the field of big data, there exist many data mining and machine learning solutions which can be used to analyze such data. These techniques analyze the data and identify reasons behind customer churning. *CRM* can employ these techniques to maximize their profit [2]. Furthermore, it may be used to design retention strategies to reduce the ratio of customers that are going to churn. The *CRM* can achieve the customer retention objective of a company by identifying accurate customer needs by using data mining techniques. Data mining involves the process of identifying the behavior of churn customers from the patterns extracted from the data. Data mining is known by many different names such as business intelligence, predictive modeling, knowledge discovery, and predictive analytics. Data mining is one of the dimensions of *CRM*, where the CRM analyzer should combine the three main dimensions; data, data mining and decision makers.

Examining the behavior of churn customer helps in measuring the loss incurred due to valuable churning customers and sustaining the current revenue due to massive reforms in telecom sector since the previous decade resulted in massive saturation and enhanced competitors due to deregulation of companies [4]. It is hard to predict the cause of churn and its frequency. Numerous problems with customer development arise primarily because of the quality element including service quality, network coverage, load errors, billing, costs, technologies, etc. These service quality factors allow customers to compare service quality and benefits with another compatible services provider [5]. A telecom sector can do exceptionally well and deal with current customers, irrespective of the possibility that it is not about getting new customers. In general, the prediction rate of the normal customer in the telecom sector is estimated at 2%, which is the total annual loss of approximately 100 billion dollars [8]. Predicting churn customers is 16 times cheaper than attracting new customers and the cost of inviting new customers is 5 to 6 times more than keeping existing customers [8]. Decreasing the churn rate by 5% increases the profit from 25% to 85% [9]. Technological improvements have helped companies to understand other belligerent plans to ensure a high level of churn customers in the business [10]. Researchers are focused on differentiating customers to identify the ones who are likely to churn to another service provider [11]. Due to the deregulation of telecommunication companies, there are many competitors in the market, and customers have more choices to fulfill their needs. So, telecom companies need to better understand customer needs and meet them, taking into account the ultimate goal of escaping from competitor [12]. *CRM* requires the connotation to recognize and understand its business unit and customers. *CRM* also controls improvements in offers and discounts such as which items are offered to which customers and which services and promotions they need.

Existing studies reveal that the primary objective is to use a large volume of telecom data to identify the valuable churn customer. However, there are several limitations in existing models, which put strong obstacles toward this problem in the real-world environment. A large volume of data is being generated in the telecom sector and the data contains missing values, which lead to the poor result of the prediction models. To handle these issues, data preprocessing methods are adapted to remove noise from data, which is effective for a model to correctly classify the data and improve the performance. Feature selection has been used in literature, however, a number of information-rich features are neglected while modeling development [13]. In a diverse domain, mostly statistical methods are used which lead to poor results of the predictive model. In existing studies, models have been validated with benchmark datasets [14], [15] which cannot present the true representation of data and are not valuable for the decision makers. To handle this limitation, multiple algorithms are used on the same dataset and the best classifier is selected for retention. The intelligent mechanism can help in developing prediction models for automated churn prediction and retention. Another major problem in existing models is the feature selection. Every customer or group of customers have different reasons for churn. In literature, a churning customer is simply classified as churner without seeing his/her churning reasons and factors. Churners have different patterns of behavior and all of them should not be treated in the same manner. Some customers are more likely to churn than others. There is a need for such a prediction model that can predict churn customers and provide retention strategies such as different promotions for a different group of churn customers based on their churn factors. Encouraged by the above-mentioned limitations, we used Information Gain and Correlation Attributes Ranking Filter feature selection techniques and selected the top features to form both results.

In this study, we proposed a churn prediction model that uses various machine learning algorithms. The performance of a classifier depends on the available dataset. It is validated by using a real-world dataset of *Call Detail Records (CDR)* of a South Asian company. The proposed churn prediction model is evaluated using information retrieval metrics. The accuracy is calculated for churn prediction model using *TP rate, FP rate, Precision, Recall, F-measure*and*ROC area.* The objective of the study is to investigate the existing techniques in machine learning and data mining and to propose a model for customer churn predictions, to identify churning factors and to provide retention strategies. From the experiments, we observed that our proposed model performed better in term of classification of churners by achieving high accuracy.

Our contribution to this study is to propose a churn prediction model. The important features are selected using feature selection techniques such as information gain and correlation attribute ranking filter. We used a number of machine learning techniques for churn and non-churn classification on two large datasets of the telecom sector. We observed that the Random Forest algorithm produced better accuracy as compared to other machine learning algorithms. We performed customer profiling based on the behavior of customers into three groups Low, Medium and Risky using *k-means* clustering. We identified the factors behind the churning of customers by using the rules generated from Attribute Selected Classifier.

The remaining paper is structured as follows. Section II provides related work. Section III presents the proposed customer churn prediction model. Section IV describes experimental evaluation, and results. Section V provides customer profiling and retention guidelines. Finally, Section VI concludes the discussion and provides future work.

II. RELATED WORK

Churn prediction has been performed in the literature using various techniques including machine learning, data mining, and hybrid techniques. These techniques support companies to identify, predict and retain churning customers, help in decision making and *CRM*. The *decision trees* are the most commonly recognized methods used for prediction of problems associated with the customer churn [18]. There is a constraint in the decision tree that it is not appropriate for complex nonlinear connections between attributes but perform better for linear data in which the attributes depend on each other. However, the study shows that pruning improves the accuracy of the *decision tree* [19]. There are many advantages of decision tree algorithms: they can be easily visualized and understood, can process categorical and numerical data, and use a nonparametric method that does not need prior assumptions [42]. The data used in this analysis is linear and we intend to identify rules and hidden pattern through the decision tree. A *neural network* based methodology for the prediction of churn customers in the telecom sector is provided in [14]. In literature, churn prediction is also performed using data certainty [16] and particle swarm optimization [30]. Another study provides a comparison of churn prediction between *ANN*and *decision trees* which results revealed that the accuracy of the decision tree based approach is better than the *neural network* based approach [20]. This work was further extended by a study which aimed at finding answers to customer loyalty results in prepaid mobile phone organizations [21]. In this work, a two-step approach is used for prediction. In the first step, RFM related features are divided into four clusters and in the second step the churn data, which is extracted in the first step, is tested on different algorithms using *Decision Tree (C5.0), Decision (CART), Neural Networks,*and*Decision Tree (CHAID)*. It shows that the hybrid approach resulted in better performance as compared to a single algorithm. The study proposed by [22] is a hybrid approach for churn prediction and results showed better performance using existing tree induction algorithm with genetic programming to derive classification rules based on customer behavior. Predictive models for churn customers regarding prepaid mobile phone companies are described in [21], [23]. In another study, authors use *Support Vector Machine (SVM)*, *Neural net*, *Naïve Bayes*, *K-nearest neighbors* and *Minimum-Redundancy Maximum-Relevancy (MRMR)* features selection technique [9].

In Statistical approaches, hybrid techniques are used for processing large amounts of customer data including regression-based techniques that produced good results in predicting and estimating churn [17]. Data mining algorithms are often used in customer history analysis and prediction. The techniques of regression trees were discussed with other commonly used data mining methods such as *decision trees*, rules-based learning, and neural networks [17], [18]. *Naive Bayes* is a guided learning module that predicts invisible data based on the position of Bayesian, is used to predict churn customer [20]. Churn problem for wireless-based customer data is discussed in [24].

There is a range of hybrid techniques proposed in the literature for churn prediction. One such technique, named *KNN-LR*, is a hybrid approach using *Logistic Regression* (*LR*) and *K-Nearest Neighbor* (*KNN*) algorithm is used in the study [25]. They conducted a comparison between *KNN-LR*, *logistic regression*, *C 4.5* and *Radial Basis Function* (*RBF*) network and found that *KNN-LR* is superior in performance to all the other approaches. The novel model presented in [13] shows a hybrid approach linking the adapted *k-means* clustering algorithm with the classical rule inductive

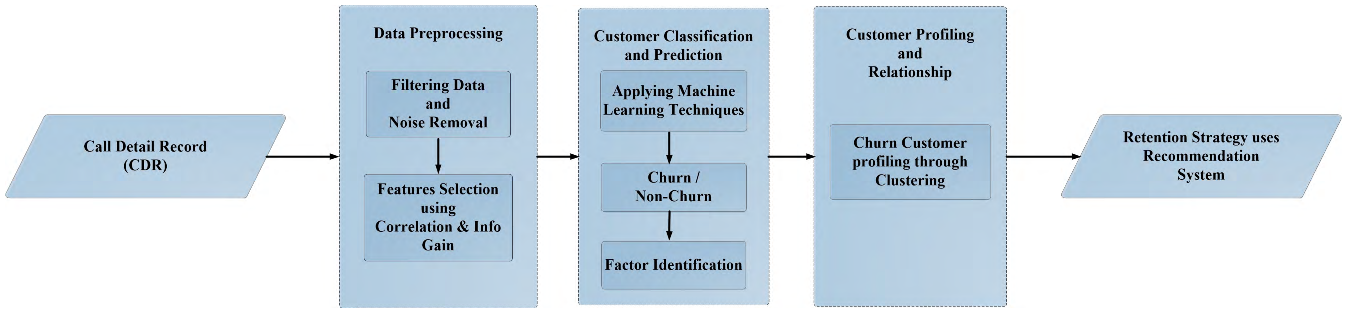


FIGURE 1. **Proposed model for customer churn prediction.**

technique (*FOIL*) to predict churn customer behavior. The control of a large volume of data in today's world provides an opportunity to improve the quality of service to the users. This data includes information about customers behavior, usage pattern and network operations. The study [10] proposed a model for both online and offline distributed framework based on data mining techniques to predict and identify churn customers. The model is appropriate for telecom to improve the CRM and its quality of service in different aspects. *Particle Swarm Optimization (PSO)* techniques are used for features selection as its preprocessing mechanism.

The telecommunications service sector has undergone a major change over the past decade due to new services, state-of-the-art upgrades [31]–[36] and intensified competition due to deregulation [4]. There is a need to secure important customers, strengthen connection management of CRM and improve the profitability [5], [11]. CRM needs that a company knows and understands its business units and customers. CRM controls improvements in offers and discounts. CRM also controls which services (including media and promotions etc.) are offered to which customers. In churn prediction, the distance factor approach is used for classification based on certainty estimation [16]. The study [37] used a likely maximum profit standard which is one of the core performance measures to provide insight factors from a cost-effective point of view. The prediction problem is also addressed using *EPMC* and *ProfLogit* approaches [38]. Deep learning is applied for churn prediction through *Convolutional Neural Network (CNN)* [39]. Gaining operational efficiency by acquiring new features and reusable techniques are used to determine the best features by applying Pareto multi-criteria optimization [40]. In this paper, in the first phase, customers are segmented using decision rules and in the second phase, a model is developed for every leaf of the tree. This hybrid approach is compared with logistic regression, decision trees, random forests, and logistic model trees [41] for churn prediction.

III. PROPOSED MODEL FOR CUSTOMER CHURN PREDICTION

This section presents the proposed customer churn prediction model. Fig. 1 shows the proposed churn prediction model and describes its steps. In the first step, data preprocessing is performed which includes data filtering for noise removal, removal of imbalanced data features and normalization of the data. Important features are extracted from data using information gain attributes ranking filter and correlation attributes ranking filter. In the second step, different classification algorithms are applied for categorizing the customers into the churn and non-churn customers. The classification algorithms include *Random Tree (RT), J48, Random Forest (RF), Decision Stump, AdaboostM1 + Decision Stump, Bagging + Random Tree, Naïve Bayes (NB), Multilayer Perceptron (MLP), Logistic Regression (LR), IBK*and*LWL*. This step also identifies factors which are used in the next step for applying clustering algorithms. In the third step, customer profiling is performed using *k-means* clustering techniques. Cluster analysis is based on the patterns of customer transactional behavior captured from the data. In the final step, the model recommends retention strategies for each category of churn customers.

A. DATA PREPROCESSING

1) NOISE REMOVAL

It is very important for making the data useful because noisy data can lead to poor results. In telecom dataset, there are a lot of missing values, incorrect values like "Null" and imbalance attributes in the dataset. In our dataset, the number of features is 29. We analyzed the dataset for filtering and reduced the number of features so that it contains only useful features. A number of features are filtered using the delimiter function in Python. TABLE 1 shows the 29 features which are available in the dataset.

2) FEATURES SELECTION

Feature selection is a crucial step for selecting the relevant features from a dataset based on domain knowledge. A number of techniques exist in the literature for feature selection [26], [27] in the context of churn predictions. In this study, we used Information Gain and Correlation Attributes Ranking Filter techniques for feature selection using WEKA toolkit [28]. In churn dataset, we selected only the top 17 features out of the total 29 features, having high ranking values in the results of both techniques. The dataset used in this study contains 29 attributes. In such a high dimensional dataset,

TABLE 1. Number of filtered features.

S#	Features	S#	Features
1	TOTAL_CALLS	16	IDD_CALLS
2	TOTAL_MINS	17	IDD_MINS
3	TOTAL_CALLS_REV	18	IDD_REV
4	TOTAL_INCOMING_MINUTES	19	CHRGD_CALLS
5	TOTAL_INCOMING_REV	20	CHRGD_MINS
6	TOTAL_OUTGOING_MINUTES	21	CHRGD_REV
7	TOTAL_OUTGOING-REV	22	FREE_CALLS
8	ONNET_CALLS	23	FREE_MINS
9	ONNET_MINS	24	TOTAL_SMS
10	ONNET_REV	25	CHRGD_SMS
11	OFFNET-CALLS	26	FREE_SMS
12	OFFNET_MINS	27	REVENUE_SMS
13	OFFNET_REV	28	RECHRG_TOTAL_LOAD
14	INCOMING_INC_REV	29	TOTAL_VAS_REV
15	INCOMING_TOTAL_CALLS		

some attributes improve performance measure and are useful for decision-making process while others are less important attributes. The performance of classification increases if the dataset contains highly predictive and valuable variables. Therefore, focusing on selecting significant features and decreasing the number of irrelevant attributes increases classification performance. Two machine learning methods are used for attribute selection to acquire the most relevant attributes. Information Gain (IG) entropy is used in decreasing order. And Correlation Attributes Ranking Filter techniques is used for selecting a subset of relevant features. From these techniques, the ranking of the most significant subsets of attributes is selected having low computational cost and avoiding the dimensionality problems. The attributes ranking is employed to identify the factors and hidden pattern in data that are the main reasons of churning. The ranking values of Information Gain and Correlation Attributes Ranking Filter are shown in TABLE 2.

B. CUSTOMER CLASSIFICATION AND PREDICTION

There are two types of customers in the telecom dataset. First, are the non-churn customers; they remain loyal to the company and are rarely affected by the competitor companies. The second type is churn customers. The proposed model targets churn customers and identify the reasons behind their migration. Furthermore, it devises retention strategies to overcome the problem of switching to other companies. In this study, a range of machine learning techniques is used for classifying customers' data using the labeled datasets. It is to assess which of the algorithm best classifies the customers into the churn and non-churn categories. First, the decision tree algorithm is used for classification. It is categorized as an eager learning algorithm where training data is generalized to classify new samples. It has been extensively used in the literature for data analysis and is the modified version of the original *ID3 and C4.5* algorithms [29]. Secondly, we used

TABLE 2. Ranking values of information Gain and Correlation Attributes Evaluator.

Attributes	Information Gain Ranking Values	Correlation Attributes Ranking values
TOTAL_CALLS	0.010614	0.07856
TOTAL_MINS	0.007962	0.0497
TOTAL_CALLS_REV	0.009111	0.07175
ONNET_CALLS	0.008609	0.06123
ONNET_MINS	0.006335	0.04303
ONNET_REV	0.008882	0.06251
OFFNET_CALLS	0.007919	0.06542
OFFNET_MINS	0.006929	0.06139
OFFNET_REV	0.007164	0.0646
INCOMING_TOTAL_CALLS	0.003773	0.04296
CHRGD_CALLS	0.010331	0.0757
CHRGD_MINS	0.008834	0.05974
CHRGD_REV	0.009111	0.07175
FREE_CALLS	0.005597	0.04683
FREE_MINS	0.006043	0.04066
REVENUE_SMS	0.005483	0.04333
RECHRG_TOTAL_LOAD	0.003697	0.05451

Random Forest, *Decision Stump*, *J48* and *Random Tree* with 10-fold cross-validation. Other than analyzing individual algorithms, hybrid algorithms are also selected for experimentation. This includes *AdaboostM1+DecisionStump* and ***Bagging* + *Random Tree*** algorithms. Additionally, the prediction model was also tested using the *Bayes algorithm* which too lies in eager learning class and performs better on larger datasets consisting of millions of records. It could be used for real-time prediction, multi-class prediction, text classification, spam filtering, sentiment analysis, and recommendation system. The classification algorithms *Random Forest (RF)*, *Artificial Neural Networks (ANN)*, *decision tree, C5, Multilayer Perceptron (MLP)* and *Logistic Regression (LR)*are also used in the simulation experiment. The classification process is performed using WEKA 3.8 toolkit [28].

IV. EXPERIMENTS AND RESULTS

We performed a number of experiments on the proposed model using machine learning techniques on two datasets. The subsequent sections present the results obtained using different machine learning techniques. Further, it provides the factors behind customer churn. *Random Forest* is a useful technique for classification and can handle nonlinear data efficiently. Unlike others, it performs better if correlated features exist in the data. *RF*produced better results because it handled very well with our data and produced a better performance as compared to other techniques. *RF* uses multiple decision trees to make a prediction. We need better results of churn customer prediction for further segmentation and we got 88.63% correct classification through *RF*. However, *RF* is not appropriate for rule generation for factor identification as it generates complex forest which is difficult to visualize and rule inference. Therefore, in this study, for factor identification, a comparable classifier such as *Attribute Selected Classifier* is used for rule generation that can be easily visualized.

These rules provide hidden patterns or factors of potential churn customer's usage and behavior which can later be used in customer profiling to specify policies for retention. There exist other methods for rule generation such as Rough Set Theory (RST) [33]. Rough Set Decision based Tree (RDT) performs well, however, in this study we performed customer profiling based on their behavior through *k-means* clustering algorithm for creating retention policies by decision makers.

DATASET DESCRIPTION

In this study, two datasets are used. The first dataset is obtained from South Asia GSM telecom service provider for studying customer churn prediction problem. It has 64,107 instances with 29 features, in which all features are numerical. The data is extracted from the customer service usage pattern Call Detail Record (CDR). It contains labeled data with two classes where 30% data is labeled as "T" (true customers) that represents churner and 70% data is labeled as "F" (false customers) that represent non-churners.

It has three types of' attributes that include call behavior or usage attributes, marketing related attributes, and financial information attributes. Selection of attributes depends upon results of feature selection techniques that allows identifying the most relevant, useful and effective attributes for customer churn prediction. The second dataset is a publicly available churn-bigml dataset "*http://bigml.com/user/francisco/gallery/dataset/5163ad540c0b5e5b22000383*"[1]. The dataset contains 3333 instances and 16 features which is in numerical form and the targeted churn customers class is labeled as "T" which is 14.5% of the total data whereas 85.5% are non-churn customers which are labeled as "F". Table 3 describes both datasets.

PERFORMANCE EVALUATION MATRIX

In this study, the proposed churn prediction model is evaluated using accuracy, precision, recall, f-measure, and ROC area. *Equation 1* calculates the accuracy metric. It identifies a number of instances that were correctly classified.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{1}$$

Here "TN" stands for True Negative, "TP" stands for True Positive, "FN" stands for False Negative and "FP" stands for False Positive. TP Rate is also known as sensitivity. It tells us what portion of the data is correctly classified as positive.

[1] "*http://github.com/caroljmcdonald/mapr-sparkml-churn/tree/master/data*"

TABLE 3. Dataset description.

	Instances	Attributes	Target Class
Dataset 1 Description	64,107 Numerical Data	29	Two class classification T represent Churn Customer F represent Non-churn Customer
Dataset 2 Description	3333	16	Two class classification T represent Churn Customer F represent Non-churn Customer

For any classifier, the TP rate must be high. TP rate is calculated by using *Equation 2*.

$$TP\ Rate = \frac{True\ Positives}{Actual\ Positives} \tag{2}$$

FP Rate tells us which part of the data are incorrectly classified as positive. The result of the FP rate must be low for any classifier. It is calculated by using *Equation*3.

$$FP\ Rate = \frac{False\ Positives}{Actual\ Negatives} \tag{3}$$

Accuracy, also known as Positive Predictive Value (PPV), indicates which part of the prediction data is positive. It is calculated by using Equation 4.

$$Precision = \frac{True\ Positive}{(True\ Positive + False\ Positive)} \tag{4}$$

The recall is another measure for completeness i.e. the true hit of the algorithm. It is the probability that all the relevant instances are selected by the system. The low value of recall means many false negatives. It is calculated by using *Equation 5*.

$$Recall = \frac{(True\ Positive)}{(True\ Positive + False\ Negative)} \tag{5}$$

The F-measure value is a trade-off between correctly classifying all the data points and ensuring that each class contains points of only one class. It is calculated by using *Equation 6*.

$$F - measure = 2 * \left(\frac{Precision * Recall}{Precion + Recall}\right) \tag{6}$$

ROC area denotes the average performance against all possible cost ratios between FP and FN. If the ROC area value is equal to 1.0, this is a perfect prediction. Similarly, the values 0.5, 0.6, 0.7, 0.8 and 0.9 represent random prediction, bad, moderate, good and superior respectively. Values of ROC areas other than these indicate something is wrong.

TABLE 4. Performance measure of various classification algorithms with 10-fold cross-validation on own churn-balance dataset.

Method used	Incorrectly Classified Instances (%)	Correctly Classified Instances (%)	Time for Building Tree (Sec)
Random Forest	11.37	88.63	108.48
Attribute Selected Classifier	11.66	88.34	4.08
J48	11.42	88.58	7.44
Random Tree	15.66	84.34	2.06
Decision Stump	29.02	70.98	0.97
AdaBoostM1 Classifier + Decision Stump	16.05	83.95	9.24
Bagging + Random Tree	11.39	88.61	13.98
Naïve Bayes	52.37	47.63	0.48
Multilayer Perceptron	17.96	82.04	214.18
Logistic Regression	29.02	70.98	1.87
IBK	19.63	80.37	0.02
LWL	18.41	81.59	0.05

TABLE 5. Performance measure of various algorithms with 10-fold cross-validation on the churn-bigml dataset.

Method used	Incorrectly Classified Instances (%)	Correctly Classified Instances (%)	Time for Building Tree (Sec)
Random Forest	10.41	89.59	1.39
Attribute Selected Classifier	8.09	91.91	0.13
J48	8.09	91.91	0.11
Random Tree	17.18	82.82	0.05
Decision Stump	13.13	86.87	0.03
AdaBoostM1 Classifier + Decision Stump	13.13	86.87	0.33
Bagging + Random Tree	11.12	88.88	0.2
Naïve Bayes	11.12	88.88	0.05
Multilayer Perceptron	10.71	89.29	137.41
Logistic Regression	14.15	85.85	35
IBK	14.14	85.86	11
LWL	11.12	88.88	10

A. APPLYING MACHINE LEARNING TECHNIQUES

Different classification techniques are applied to two churn datasets using WEKA toolkit. We obtained the performance (in terms of correctly classified, incorrectly classified and time to build tree) from two datasets as mentioned in Table 4 and TABLE 5 that show the accuracy of all algorithms with 10-fold cross-validation. It is observed that the *Random Forest* and *J48* outperform other techniques for both datasets and correctly classify the data with 88.63% and 88.58% accuracy respectively. *Decision Stump* and *Random Tree* have low accuracy rate with 70.98% and 84.34%, respectively.

However, when it is used in ensemble algorithms with AdaboostM1 and *Bagging*, it performs better with 83.95% and 88.61% accuracy in both datasets respectively. Furthermore, *Random Forest, J48, AdaboostM1* and *Bagging* ensemble with *Decision Stump* and *Random Tree* have a minimum incorrect classification. The correct classification of instances of both datasets proves that these four techniques performed better than the others. It is important to note that after the classification of data with cross-validation, Random Forest outperforms other algorithms in terms of correct classification. Table 4 and Table 5 show that *Random Forest* takes much time to build the prediction model, 108.48 seconds, however, it has a maximum classification accuracy of 88.63% among all algorithms. *Multilayer Perceptron* is not good with 82.04 % accuracy and 214.18 seconds run time. *J48* performs better in terms of building model time but its accuracy is not good like *Random Forest*. Moreover, the *Random Tree* and *Decision Stump* algorithms have taken minimum time for building a prediction model, however, their classification is not up to the mark. *Random Tree* result is not stable due to a random selection of attributes in making a decision tree but its performance increases and performs equivalently to *J48* when it is used in ensemble algorithms with *AdaboostM1* + *Decision Stump*and *Bagging* + *Random Tree*. *Naïve Bayes, Multilayer Perceptron*, and *Logistic Regression* have low performance and they also take much time in the model building. The computational cost of Naïve Bayes is low, however, it is not good for accurate measurement. Lazy learning algorithms perform better in terms of accuracy and time as compared to *Neural Network* and *Bayes* algorithms, whereas, they have low performance as compared to the *Decision tree* and ensemble algorithms. The overall result of the *Random Forest* algorithm in terms of accuracy is better than other algorithms for prediction problem because in both datasets its performance is high.

Random Forest uses a divide and conquers approach. It makes the numeral type of decision tree and every Decision Tree is trained by picking any random attribute from the whole predictive set of attributes. Every tree grows up to maximum level based upon features subset. After this, a final Decision Tree is constructed for prediction of the test dataset. Random forest well performs on a large dataset and handle missing variable without variable deletion. Random Forest handles missing values inside the dataset for training the model. Table 4 and Table 5 show the accuracy and building prediction model time for both datasets.

To further validate our findings, the performance of algorithms in TABLE 6 and Table 7 show that TP and FP rate is higher for *Random Forest* classifier as compared to others. The Area under Curve (AUC) is a selective performance measure which is used by many researchers in the prediction model for measuring the accuracy. TP and FP rate

TABLE 6. Accuracy of various algorithms on own churn-balance dataset.

Method used	TP Rate	FP Rate	Precisi on-on	Recall	F-measu re	ROC area
Random Forest	0.888	0.236	0.893	0.888	0.882	0.947
Attribute Selected Classifier	0.888	0.258	0.902	0.888	0.880	0.913
J48	0.887	0.243	0.893	0.887	0.880	0.940
Random Tree	0.843	0.215	0.844	0.843	0.844	0.814
Decision Stump	0.700	0.700	0.490	0.700	0.577	0.645
AdaBoost M1 Classifier + Decision Stump	0.835	0.339	0.839	0.835	0.822	0.788
Bagging + Random Tree	0. 881	0.233	0.883	0.881	0.876	0.930
Naïve Bayes	0.473	0.284	0.715	0.473	0.456	0.629
Multilayer Perceptron	0.822	0.346	0.821	0.822	0.810	0.804
Logistic Regression	0.700	0.700	0.490	0.700	0.577	0.496
IBK	0.810	0.358	0.805	0.810	0.797	0.828
LWL	0.812	0.302	0.806	0.812	0.807	0.847

TABLE 7. Accuracy of various algorithms on a churn-bigml dataset.

Method used	TP Rate	FP Rate	Precision	Recall	F-measure	ROC area
Random Forest	0.896	0.565	0.891	0.896	0.876	0.835
Attribute Selected Classifier	0.914	0.419	0.909	0.914	0.905	0.799
J48	0.912	0.419	0.906	0.912	0.904	0.798
Random Tree	0.825	0.551	0.821	0.825	0.823	0.646
Decision Stump	0.867	0.636	0.843	0.867	0.844	0.598
AdaBoostM1 Classifier + Decision Stump	0.868	0.610	0.846	0.868	0.848	0.814
Bagging + Random Tree	0.883	0.574	0.868	0.883	0.864	0.801
Naïve Bayes	0.881	0.527	0.866	0.881	0.868	0.792
Multilayer Perceptron	0.895	0.429	0.886	0.895	0.888	0.797
Logistic Regression	0.854	0.778	0.810	0.854	0.809	0.742
IBK	0.856	0.852	0.877	0.856	0.790	0.660
LWL	0.833	0.512	0.869	0.883	0.871	0.782

of *Random Forest* is high, however, its FP rate is like *J48* and *AdaboostM1* ensemble techniques. Whereas, *Naïve Bayes* performance is worst as compared to all another classifier. *J48*and *Random Forest* have high precision to predict churners correctly by having the highest TP rate in both datasets as shown in TABLE 6 and Table 7. Furthermore, ensemble algorithms *AdaboostM1 + DecisionStump*and *Bagging + Random Tree* also have good prediction sensitivity. Lazy learning methods *IBK* and *LWL*have least TP rate as compared to *Random Forest* and *J48*, however, they are better than *MLP*, *Logistic Regression*, and *Naïve Bayes,* which have very low sensitivity TP rate. *Random Forest, J48*and *AdaboostM1* have minimum FP rate, which indicates that they are good classifiers for prediction. It is observed that the *MLP, Logistic Regression* and *Naïve Bayes* perform badly in terms of *TP, FP, specificity, and sensitivity*. *Random Forest* has a maximum value of recall. It means that this algorithm found the maximum number of true positives in the dataset and it can correctly identify the churn customers. The precision of *J48* and *Random Forest* is the highest which indicate that these algorithms outperform other algorithms in the prediction of real positive values. The precision rate of Random Forest and J48 is 0.893 which is better as compare to other algorithms. *Random Forest* and J48 algorithms have performed very well as compared to all other algorithms by having 0.959 ROC area under the curve. *Random Forest* is an outstanding classifier for prediction of instances. According to the ROC value scale discussed earlier, J48 and ensemble algorithms also performed better.

B. FACTORS IDENTIFICATION OF CHURN CUSTOMERS

The factors of churn customers are discussed in this subsection which is classified by the *Attribute Selected Classifier* algorithm. This method provides rules for churn prediction and provides churning user behavior and patterns. These key rules are very valuable for the decision makers for the retention of churning customers. The *Attribute Selected Classifier* algorithm identifies many reasons of churn and provides features which depend on each other. The churn related rules provided by the *Attribute Selected Classifier* algorithm are described below.

Factor 1:

OFFNET_CALLS, OFFNET_MINS, ONNET_CALLS, and TOTAL_CALLS are highly dependent features of churning because when the calls rate to other networks are higher than the On-Net calls of this customer fall in churner class, see rule 1 in Fig.4.

Factor 2:

TOTAL_CALLS_REV is in relation to TOTAL_CALLS when the call rate is high then it will increase revenue but rule 2 describes that when revenue decrease from 607 then chances of churn also increase. Rule 2 describes it in detail as shown in Fig. 4.

Factor 3:

TOTAL_CALLS, REVENUE_SMS, TOTAL_MINS, RECHRG_TOTAL_LOAD, and FREE_MINS are dependent on each other. If customers have a greater number of total calls, SMS revenue, total minutes and also greater RECHRG_TOTAL_LOAD but he gets free minutes less

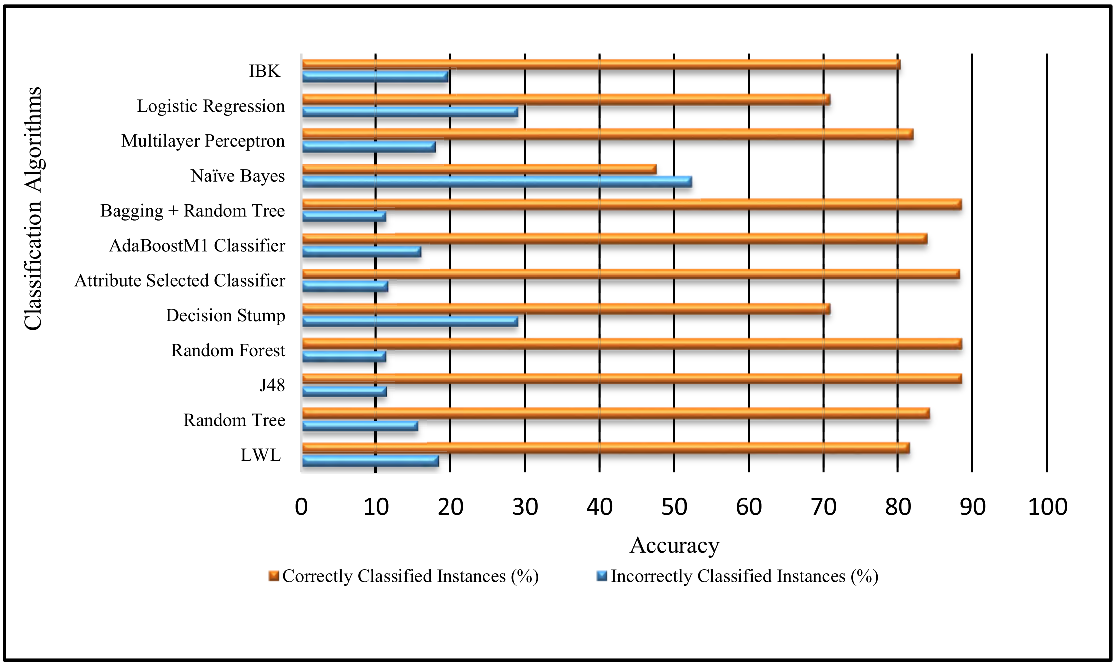


FIGURE 2. Accuracy performance of classification algorithms on the churn-balance dataset.

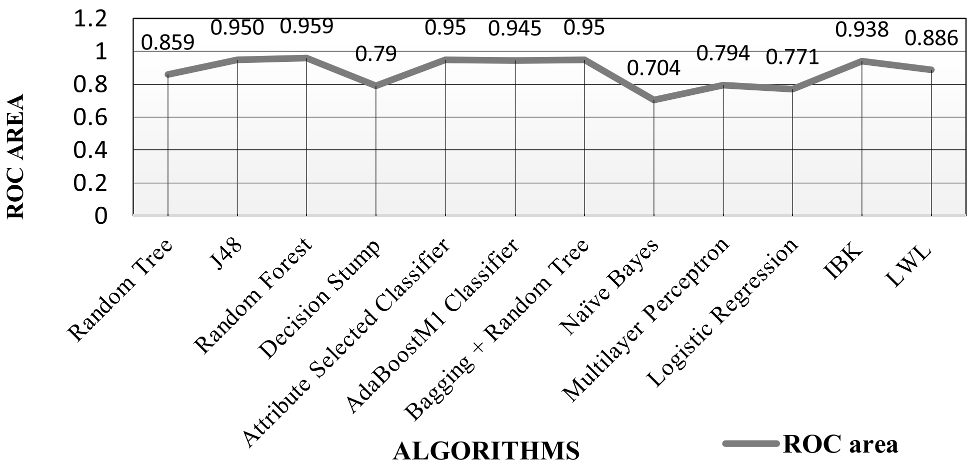


FIGURE 3. ROC area curve.

than 17 then the customers will fall in churn class. See rule 4 as shown in Fig. 4.

Factor 4:

If the customer's call revenue is greater but On-Net minutes are less than his free minutes and if the customer also gets less free minutes, then check total minutes. Also, if call revenue is less than the customer will churn because the customer is charged more as compared to call revenue and gets no benefit in return. If a customer On-Net revenue is greater but gets less free minutes, then this customer also falls in churn class. The call revenue, free minutes, On-Net revenue, recharge total load and total minutes are described in detail in rule 5 as shown in Fig. 4.

V. CUSTOMER PROFILING AND RETENTION

Customer cluster is used to partition the complete customers' data into groups based on their behavior information and their relationship. A number of clustering algorithms can be applied for hierarchical, fuzzy and partition clustering. We used *k-means* technique that is the best for partition clustering which can segment the data into different groups as the given problem involves very complex, heterogeneous and very large dataset [13]. The *k*-means is a well-known iterative approach to partition the data. In this technique, the data is segmented as belonging to one of the *k*-groups. We consider real-valued data in which the arithmetic mean value is the representative of the cluster. *k-means* is useful to find a relationship and hidden pattern in data which belong to one class. In this study, the *k-means* algorithm segments the data into three group due to the nature of the data. The three groups represent Low, Medium and Risky customers. Figure 5 shows the threshold value and number of customers in each segment according to the distance to the nearest cluster. This analysis verifies that setting the value of k in *k-means* to 3 can lead to better segmentation results. *k-means* can represent the relationship and pattern on the basis of which decision maker can retain the customers and provide specific policies to a specific class of customers. We describe the customer profiling in subsection A and provide retention guidelines in subsection B.

A. CUSTOMER PROFILING

A number of steps are involved in cluster analysis which include *Data Compilation* and *Cluster Analysis*. Data compilation is the activity that contains a collection of data from different dimensions whereas cluster analysis is used to create clusters using *k-means* clustering algorithm [19] which runs in a repetitive mode. Through every repetition, data points are assigned to a cluster based on the minimum Euclidean distance from the k-cluster centroids. The *k-means* clustering algorithm aims to split the set of n observations $(x_1, x_2, \ldots x_n)$, into $K(\leq n)$ disjoint sets $S = \{S_1, S_2, \ldots, S_k\}$ so as to minimize the sum of squares within the cluster. Mathematically, this goal can be achieved by using Equation 7.

$$j = \min \sum_{i=1}^{k} \sum\nolimits_{x \epsilon S1} ||x - \mu i||^2 \quad (7)$$

where μ_1 is the mean of point S_i.

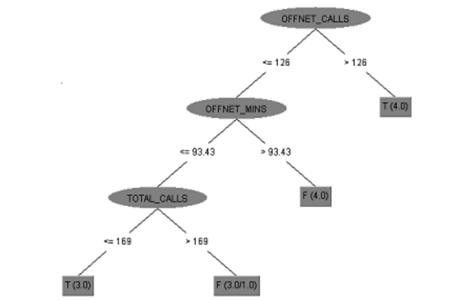
Algorithm 1 categorizes the customers based on their behavior. In our case, three clusters are developed according to three different customer behaviors.

In existing studies, the researchers analyzed the behavior of both classes churner and non-churner [12]. However, our approach is different, as we are analyzing the behavior of churner only and find the behavior of similar customers because the decision makers are interested to find the behavior of churner and make appropriate policy for retention to maximize companies profit. We focused on churn class data only to assess the behavior of similar customers and compare their prediction. The clustering dataset contains 19213 instances of churn customers only which is correctly classified by RF algorithm. We take the results of the RF algorithm because of its better performance and less error rate. It contains only 17 attributes during the classification process with high ranking value to construct the model and generate a valuable pattern of behavior for churn. Based on similar behavior of a group of customers appropriate policy is developed for churner only. Using *k-means* clustering algorithm the data is partitioned into three segments which include Low, Medium, and Risky customers for policy making to retain the churn customers. Fig. 5 summarizes the segmentation of

Algorithm 1 *k-Means* Clustering for User Behavior

```
Input: Churn class data
Cluster the data using Means Number of clusters = 3
  Import
    K_Means,
    Find cosine similarity
    Identify user behavior
    Generate_Segmant0, Generate_Segmant1,
    Generate_Segmant2,
Print (Cluster) // mapping cluster data with columns

// Find cosine similarity in percentage for each cluster
    Val = find cosine similarity (item ['cluster'],
        user data)
          Return max similar value
// Generate rules within clusters of data for each defined
attribute & identify user behavior based on rules.
Headers = ['Clusters', 'total_calls', 'onnet_calls',
'offnet_calls'......n]
  If i == 0: // i denote each attribute in dataset
      Writing headers (headers, 'Rule1.CSV')
      With open ('Rules/Rule1.csv', 'a') as f:
      Writer = CSV.DictWriter (f, fieldnames= headers)
      //each cluster is saved in separate CSV file
  Try:
      writer.writerow ({
      'Clusters': i,
      'total_calls': data ['total_calls'],
      'onnet_calls': data ['onnet_calls'],
      'offnet_calls': data ['offnet_calls'],.
      . // for every attribute the rules are define.
      })
Calculate percentage of each segment of cluster churn
customers:
    Result = float (data1) / float (total)
    Result = round (result * 100)
Return result
// Grouping churn customers in three segments (Risky,
Medium, Low) by rules using threshold value.
If offnet_calls >= 64 & onnet_calls <= 34:
    Print ("Risky")
    risky_count + = 1
If offnet_calls >= 55 & offnet_calls <= 64 & onnet_calls
>= 44 & onnet_calls < 83:
    Print ("Medium")
    medium_count + = 1
If offnet_calls <= 16 & onnet_calls >= 83:
    Print ("Low")
    low_count + = 1
    .
    .
    Return
Print headers (Risky, Medium, and Low)
```

<table>
<tr>
<td>
</td>
<td>Rule 1:

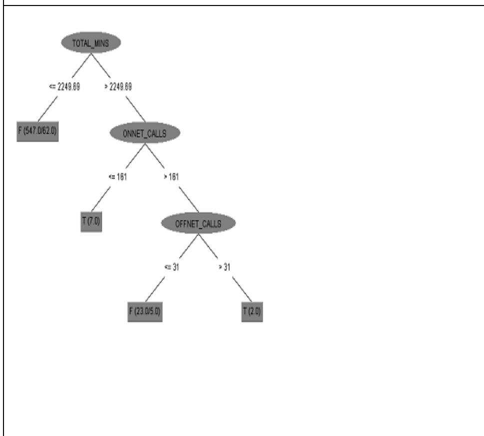
If (OFFNET_CALLS > 126) then
Output True

Else If (OFFNET_CALLS <= 126 & OFFNET_MINS > 93) then
Output False

Else If (OFFNET_CALLS <= 126 & OFFNET_MINS <= 93 & TOTAL_CALLS <= 169) then
Output True
Else False</td>
</tr>
<tr>
<td>
</td>
<td>Rule 2:

If (TOTAL_CALLS_REV > 607.58) then
Output False

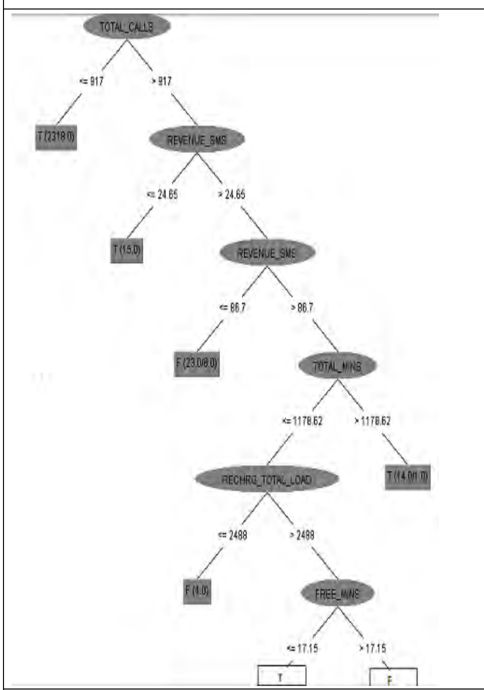
Else If (TOTAL_CALLS_REV <= 607.58 & TOTAL_CALLS <= 209 or TOTAL_CALLS <= 149) then
Output True

Else If (TOTAL_CALLS_REV <= 607.58 & TOTAL_CALLS > 209) then
Output False</td>
</tr>
<tr>
<td>
</td>
<td>Rule 3:

If (TOTAL_MINS <= 2249) then
Output False

Else If (TOTAL_MINS > 2249 & ONNET_CALLS <= 161) then
Output True

Else If (TOTAL_MINS > 2249 & ONNET_CALLS > 161 & OFFNET_CALLS <= 31) then
Output False

Else If (TOTAL_MINS > 2249 & ONNET_CALLS > 161 & OFFNET_CALLS > 31) then
Output True</td>
</tr>
<tr>
<td>
</td>
<td>Rule 4:

If (TOTAL_CALLS <=917) then
Output True

Else If (TOTAL_CALLS > 917 & REVENUE_SMS <= 24) then
Output True

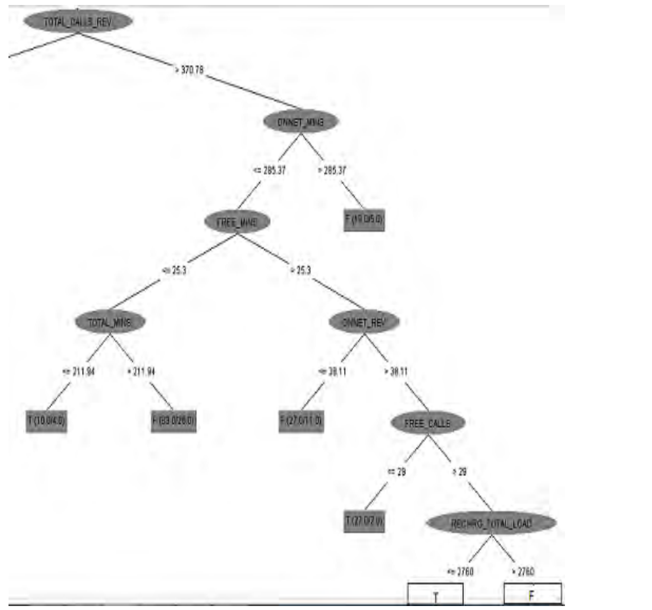
Else If (TOTAL_CALLS > 917 & REVENUE_SMS > 24 & REVENUE_SMS <= 86) then
Output False

Else If (TOTAL_CALLS > 917 & REVENUE_SMS > 86 & TOTAL_MINS > 1178) then
Output True

Else If (TOTAL_CALLS > 917 & REVENUE_SMS > 86 & TOTAL_MINS > 1178 & RECHRG_TOTAL_LOAD <= 2488) then
Output False

Else If (TOTAL_CALLS > 917 & REVENUE_SMS > 86 & TOTAL_MINS > 1178 & RECHRG_TOTAL_LOAD > 2488 & FREE_MINS <= 17) then
Output True</td>
</tr>
</table>

FIGURE 4. Sub-trees from attribute selected classifier generated tree.



Rule 5:

If (TOTAL_CALLS_REV > 370 & ONNET_MINS <= 285 & FREE_MINS <= 25 & TOTAL_MINS <=211) then
Output True

Else If (TOTAL_CALLS_REV > 370 & ONNET_MINS <= 285 & FREE_MINS > 25 & ONNET_REV > 38 & FREE_CALLS <= 29) then
Output True

Else If (TOTAL_CALLS_REV > 370 & ONNET_MINS <= 285 & FREE_MINS > 25 & ONNET_REV > 38 & FREE_CALLS > 29 & RECHRD_TOTAL_LOAD <= 2760) then
Output True

FIGURE 4. ***(Continued.)*** **Sub-trees from attribute selected classifier generated tree.**

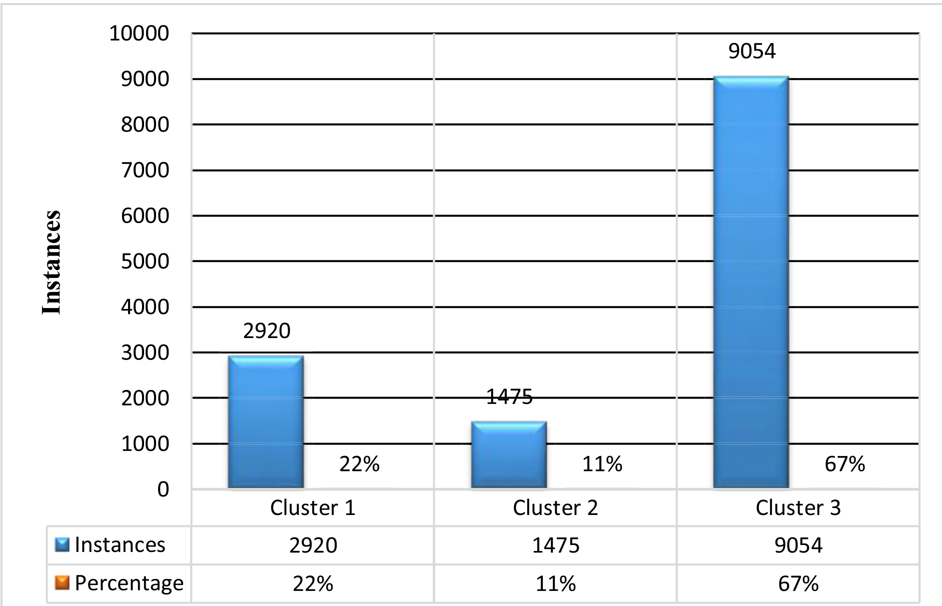


FIGURE 5. **Segmentation of churn customers.**

churner and the decision makers can easily understand the behavior of a group of customers that are more valuable and need a serious policy to improve the retention mechanism which is profitable for the organization.

Fig. 5, shows that the cluster 1 and cluster 3 have maximum similar churn customers 22% and 67% respectively. These two clusters are more valuable for the company to maximize the profit by retaining them as compared to cluster 2 with only 11% churners. Based on segmentation, we can easily find a similar pattern and factors of similar churn customers. From this pattern and behavior, we make rules for the recommendation of only similar customers in the future. Fig. 6, Fig. 7, Fig. 8 and Fig. 9 show the behavior of churner having different attributes. Each segment in the cluster represents the different pattern and behavior of churn customers. Three clusters represent a unique collection of behavior and these characteristics can extend the scope of decision making for churner. In order to differentiate the outcome of three clusters, meaningful justification can be analyzed from attributes. We identify rules and factors from

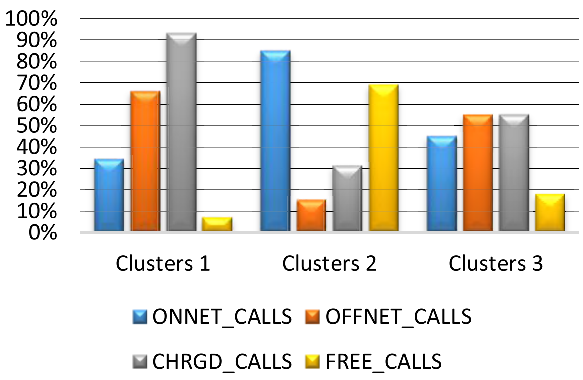


FIGURE 6. TOTAL_CALLS behavior in each cluster.

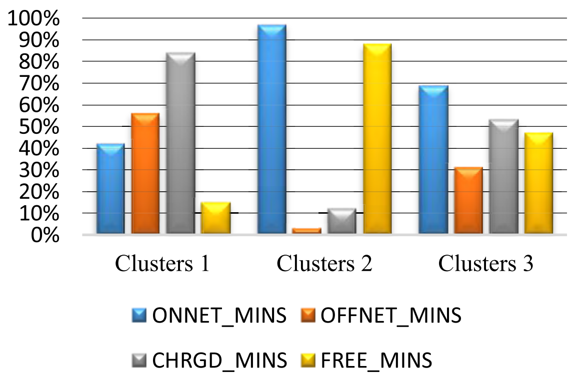


FIGURE 7. TOTAL_MINS behavior in each cluster.

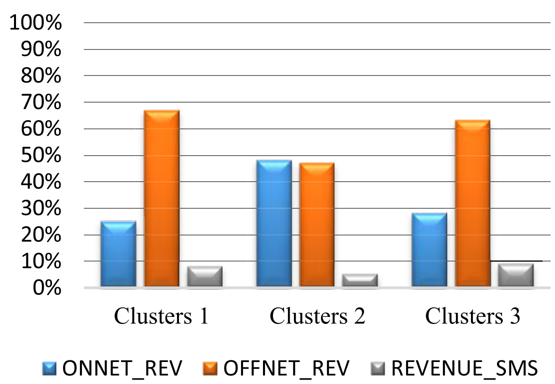


FIGURE 8. TOTAL_REVENUE behavior in each cluster.

the attributes in each cluster which can further allocate each customer in appropriate class in which it falls by using a recommender system. By targeting the appropriate class for retention, win back campaigns can be planned for each class based on its categorization and offer a loyalty program, a special offer, or a package as a reward to ensure customer retention. Using categorization, decision-makers can easily generate various approaches and personalized actions.

For finalizing the retention strategies, there is a need for better targeting churn customers and controlling them through the marketing process. Top-down approach has a deficiency in terms of targeting similar customers, Bottom-up and customized approach are not good in terms of the

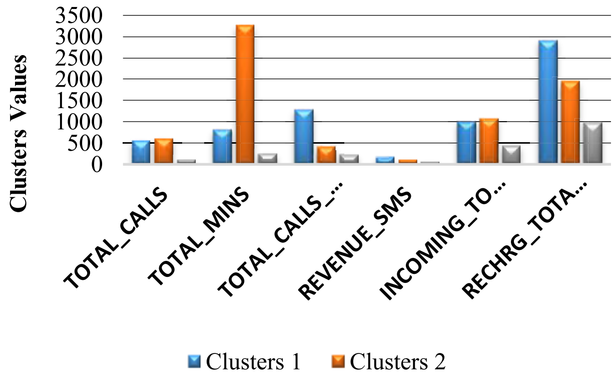


FIGURE 9. Behavior of different attributes in each cluster.

TABLE 8. Threshold values of churn customers.

Low (%)	Medium (%)	Risky (%)
ONNET_CALLS = 85	ONNET_CALLS = 45	ONNET_CALLS = 34
OFFNET_CALLS = 15	OFFNET_CALLS = 55	OFFNET_CALLS = 66
CHRGD_CALLS = 31	CHRGD_CALLS = 82	CHRGD_CALLS = 93
FREE_CALLS = 69	FREE_CALLS = 18	FREE_CALLS = 7
ONNET_MINS = 97	ONNET_MINS = 69	ONNET_MINS = 42
OFFNET_MINS = 3	OFFNET_MINS = 31	OFFNET_MINS = 56
CHRGD_MINS = 12	CHRGD_MINS = 53	CHRGD_MINS = 84
FREE_MINS = 88	FREE_MINS = 47	FREE_MINS = 15
ONNET_REV = 48	ONNET_REV = 28	ONNET_REV = 25
OFFNET_REV = 47	OFFNET_REV = 63	OFFNET_REV = 67
RECHRG_TOTAL_LOAD = 982.7691	RECHRG_TOTAL_LOAD = 1945.7423	RECHRG_TOTAL_LOAD = 2916.2544

marketing process. We choose the similarity-based approach to create personalized retention activities. This approach is the most well-known example of a recommender system. It can help in recommending appropriate policies or some personalized set of offers to each churn customers on the basis of its preferences and past behavior that is analyzed from clusters and from decision tree in rules as shown in Fig. 4. Recommender system, using a similarity measure for targeting similar customers, depends upon the following approaches.

- *Content-based*: It is based on past behavior and recommends similar preference and items to the same customer.
- *Collaborative*: It is based on past behavior and preference only for those customers who have a similar preference and similar behavior.
- *Hybrid*: It is the combination of both content and collaborative approaches.

This model is based on the collaborative approach for customers who have similar past behavior and similar preferences. Each customer is categorized in the Low, Medium and High categories to carry out modified retention strategies in the future. This model using cosine similarity measure of different attributes like ONNNET_CALLS, OFFNET_CALLS, ONNET_REV, FREE_CALLS, etc, in each cluster. The cosine similarity measure determines the churning customer. The obtained value indicates similar customers based on categorization. The neighborhood of each similar customer is identified, and a personalized retention offer can be proposed.

The categorization of groups can be defined as a set of customers by using threshold value for each group and make rules for it. TABLE 6 shows the threshold value extracted from clustering for each attribute and some of the rules are represented which categorized the customers into groups.

Rule 1:
IF $OFFNET\_CALLS >= 66$ *and* $ONNET\_CALLS <= 34$ *THEN*
Risky
ESLE IF $OFFNET\_CALLS >= 55$ *and* $< OFFNET\_CALLS = 66$ *and* $ONNET\_CALLS >= 45$ *and* $ONNET\_CALLS < 85$ *THEN*
Medium
ELSE IF $OFFNET\_CALLS <= 15$ *and* $ONNET\_CALLS >= 85$ *THEN*
Low

Rule 2:
IF $OFFNET\_MINS >= 56$ *and* $ONNET\_MINS <= 42$ *THEN*
Risky
ELSEIFOFFNET_MINS $<= 56$ *and* $OFFNET\_MINS >= 31$ *and* $ONNET\_MINS >= 42$ *and* $ONNET\_MINS <= 69$ *THEN*
Medium
ELSE IF $OFFNET\_MINS <= 3$ *and* $ONNET\_MINS >= 97$ *THEN*
Low

Rule 3:
IF $CHRGD\_MINS >= 84$ *and* $FREE\_MINS <= 15$ *THEN*
Risky
ELSE IF $CHRGD\_MINS < 84$ *and* $CHRGD\_MINS \geq 53$ *and* $FREE\_MINS \geq 47$ *and* $FREE\_MINS <= 15$ *THEN*
Medium
ELSE IF $<= 12$ *and* $FREE\_MINS >= 88$
THEN
Low

Consequently, the customers are categorized into groups by using rules and a threshold value which is extracted from clusters as shown in TABLE 6. The core concept of this recommended system is to identify similar churn customers in the context of similar behavior and pattern. The churners are grouped into three categories so that personalized retention offers are recommended to a specific group with similar behavior. At the end of the grouping, the churn customers are offered retention packages according to the specified groups by decision makers.

B. CUSTOMER RETENTION

The impact of retention is positive on a company profile that ultimately enhances the retention and marketing performance of the companies. Through this strategy, the company can understand the behavior of customers and retain the customers by offering suitable packages to a specific group of customers.

This research work indicates the important factors of customers, their behavior and addresses some important issues which are valuable for companies. These issues are described below.

- The research identified and used marketing factors through data mining from all dimensions of customers such as; ONNET_CALLS, OFFNET_CALLS, CHRGD_CALLS, FREE_CALLS, ONNET_MINS, OFFNET_MINS, CHRGD_MINS, FREE_MINS, ONNET_REV and OFFNET_REV, and RECHRG_TOTAL_LOAD.
- This research extracted the behavior of customers and categorized them into groups by applying the data mining techniques and focusing on the CRM, to retain customers.
- The analysis shows the relationship between the customers and improves the productivity of the company to achieve effective marketing campaigns.
- The segmentation model helped to identify the behavior of customers in the form of segments to retain them in categories - Risky, Medium and Low on a long-term basis.
- From results and customer behaviors, we can also achieve "one-to-one marketing" which emphasizes marketing with individual customers.
- To fulfill the needs of customers, their unique behavior and preferences are identified to offer them a variety of packages and services.
- CRM needs to build a specific relationship with customers in light of their identified needs and behavior to build a strong relationship with them.
- CRM system should be integrated between customers and operational front-line system to effectively manage customer's needs.

For retaining customers, introducing an integrated system could be effective across the channel for offering different services such as productivity and efficiency services, customized services, and price control services. In this context, CRM is concerned with personalizing the relationship with the customers and organization response to the customer's satisfaction such as developing a collaborative communication system, extending front offices to incorporate all employees, partners, and suppliers to interact with customers through email, telephone, web pages, contacts, text, etc. Through segmentation, the customer's behavior is monitored and tracked to identify the patterns and usage. A direct marketing method is a retaining process through which customers are retained and a variety of services are offered through many channels. Direct methods are more effective which can decrease the cost and increase productivity by offering different services and personalized offers directly. In the direct interaction with customer's, the organization gets a better response and can immediately take a decision. From the customer's response, the organization can also develop a model that can estimate likelihood of churn, identify similar groups of customers based on their responses and is more beneficial for decision makers.

VI. CONCLUSION

In the present competitive market of telecom domain, churn prediction is a significant issue of the CRM to retain valuable customers by identifying a similar groups of customers and providing competitive offers/services to the respective groups. Therefore, in this domain, the researchers have been looking at the key factors of churn to retain customers and solve the problems of CRM and decision maker of a company. In this study, a customer churn model is provided for data analytics and validated through standard evaluation metrics. The obtained results show that our proposed churn model performed better by using machine learning techniques. Random Forest and J48 produced better F-measure result that is 88%. We identified the main churn factors from the dataset and performed cluster profiling according to their risk of churning. Finally, we provided guidelines on customer retention for decision-makers of the telecom companies.

In future, we will further investigate eager leaning and lazy learning approaches for better churn prediction. The study can be further extended to explore the changing behavior patterns of churn customers by applying Artificial Intelligence techniques for predictions and trend analysis.

REFERENCES

[1] S. Babu, D. N. Ananthanarayanan, and V. Ramesh, "A survey on factors impacting churn in telecommunication using datamininig techniques," *Int. J. Eng. Res. Technol.*, vol. 3, no. 3, pp. 1745–1748, Mar. 2014.

[2] C. Geppert, "Customer churn management: Retaining high-margin customers with customer relationship management techniques," *KPMG & Associates Yarhands Dissou Arthur/Kwaku Ahenkrah/David Asamoah*, 2002.

[3] W. Verbeke, D. Martens, C. Mues, and B. Baesens, "Building comprehensible customer churn prediction models with advanced rule induction techniques," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2354–2364, Mar. 2011.

[4] Y. Huang, B. Huang, and M.-T. Kechadi, "A rule-based method for customer churn prediction in telecommunication services," in *Proc. Pacific–Asia Conf. Knowl. Discovery Data Mining*. Berlin, Germany: Springer, 2011, pp. 411–422.

[5] A. Idris and A. Khan, "Customer churn prediction for telecommunication: Employing various various features selection techniques and tree based ensemble classifiers," in *Proc. 15th Int. Multitopic Conf.*, Dec. 2012, pp. 23–27.

[6] M. Kaur, K. Singh, and N. Sharma, "Data mining as a tool to predict the churn behaviour among Indian bank customers," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 1, no. 9, pp. 720–725, Sep. 2013.

[7] V. L. Miguéis, D. van den Poel, A. S. Camanho, and J. F. e Cunha, "Modeling partial customer churn: On the value of first product-category purchase sequences," *Expert Syst. Appl.*, vol. 12, no. 12, pp. 11250–11256, Sep. 2012.

[8] D. Manzano-Machob, "The architecture of a churn prediction system based on stream mining," in *Proc. Artif. Intell. Res. Develop., 16th Int. Conf. Catalan Assoc. Artif. Intell.*, vol. 256, Oct. 2013, p. 157.

[9] P. T. Kotler, *Marketing Management: Analysis, Planning, Implementation and Control*. London, U.K.: Prentice-Hall, 1994.

[10] F. F. Reichheld and W. E. Sasser, Jr., "Zero defections: Quality comes to services," *Harvard Bus. Rev.*, vol. 68, no. 5, pp. 105–111, 1990.

[11] J. Hadden, A. Tiwari, R. Roy, and D. Ruta, "Computer assisted customer churn management: State-of-the-art and future trends," *Comput. Oper. Res.*, vol. 34, no. 10, pp. 2902–2917, Oct. 2007.

[12] H.-S. Kim and C.-H. Yoon, "Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market," *Telecommun. Policy*, vol. 28, nos. 9–10, pp. 751–765, Nov. 2004.

[13] Y. Huang and T. Kechadi, "An effective hybrid learning system for telecommunication churn prediction," *Expert Syst. Appl.*, vol. 40, no. 14, pp. 5635–5647, Oct. 2013.

[14] A. Sharma and P. K. Kumar. (Sep. 2013). "A neural network based approach for predicting customer churn in cellular network services." [Online]. Available: https://arxiv.org/abs/1309.3945

[15] Ö. G. Ali and U. Aritürk, "Dynamic churn prediction framework with more effective use of rare event data: The case of private banking," *Expert Syst. Appl.*, vol. 41, no. 17, pp. 7889–7903, Dec. 2014.

[16] A. Amin, F. Al-Obeidat, B. Shah, A. Adnan, J. Loo, and S. Anwar, "Customer churn prediction in telecommunication industry using data certainty," *J. Bus. Res.*, vol. 94, pp. 290–301, Jan. 2019.

[17] S. A. Qureshi, A. S. Rehman, A. M. Qamar, A. Kamal, and A. Rehman, "Telecommunication subscribers' churn prediction model using machine learning," in *Proc. 8th Int. Conf. Digit. Inf. Manage.*, Sep. 2013, pp. 131–136.

[18] V. Lazarov and M. Capota, "Churn prediction," Bus. Anal. Course, TUM Comput. Sci, Technische Univ. München, Tech. Rep., 2007. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.462.7201&rep=rep1&type=pdf

[19] R. Vadakattu, B. Panda, S. Narayan, and H. Godhia, "Enterprise subscription churn prediction," in *Proc. IEEE Int. Conf. Big Data*, Nov. 2015, pp. 1317–1321.

[20] V. Umayaparvathi and K. Iyakutti, "Applications of data mining techniques in telecom churn prediction," *Int. J. Comput. Appl.*, vol. 42, no. 20, pp. 5–9, Mar. 2012.

[21] A. T. Jahromi, M. Moeini, I. Akbari, and A. Akbarzadeh, "A dual-step multi-algorithm approach for churn prediction in pre-paid telecommunications service providers," *J. Innov. Sustainab.*, vol. 1, no. 2, pp. 2179–3565, 2010.

[22] V. Yeshwanth, V. V. Raj, and M. Saravanan, "Evolutionary churn prediction in mobile networks using hybrid learning," in *Proc. 25th Int. FLAIRS Conf.*, Mar. 2011, pp. 471–476.

[23] G. Nie, W. Rowe, L. Zhang, Y. Tian, and Y. Shi, "Credit card churn forecasting by logistic regression and decision tree," *Expert Syst. Appl.*, vol. 38, no. 12, pp. 15273–15285, Nov./Dec. 2011.

[24] S. V. Nath and R. S. Behara, "Customer churn analysis in the wireless industry: A data mining approach," in *Proc. Annu. Meeting Decis. Sci. Inst.*, vol. 561, Nov. 2003, pp. 505–510.

[25] Y. Zhang, J. Qi, H. Shu, and J. Cao, "A hybrid KNN-LR classifier and its application in customer churn prediction," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Oct. 2007, pp. 3265–3269.

[26] H. Yu *et al.*, "Feature engineering and classifier ensemble for KDD cup 2010," Dept. Comput. Sci. Inf. Eng., National Taiwan Univ., Taipei, Taiwan, Tech. Rep., 2010, vol. 1, pp. 1–16.

[27] L. Zhao, Q. Gao, X. Dong, A. Dong, and X. Dong, "K-local maximum margin feature extraction algorithm for churn prediction in telecom," *Cluster Comput.*, vol. 20, no. 2, pp. 1401–1409, Jun. 2017.

[28] G. Holmes, A. Donkin, and I. H. Witten, "WEKA: A machine learning workbench," in *Proc. Austral. New Zealnd Intell. Inf. Syst. Conf.*, Dec. 1994, pp. 357–361.

[29] F. S. Gharehchopogh and S. R. Khaze. (2013). "Data mining application for cyber space users tendency in blog writing: A case study." [Online]. Available: https://arxiv.org/abs/1307.7432

[30] J. Vijaya and E. Sivasankar, "An efficient system for customer churn prediction through particle swarm optimization based feature selection model with simulated annealing," *Cluster Comput.*, pp. 1–12, Sep. 2017. doi: 10.1007/s10586-017-1172-1.

[31] A. Amin *et al.*, "Cross-company customer churn prediction in telecommunication: A comparison of data transformation methods," *Int. J. Inf. Manage.*, vol. 46, pp. 304–319, Jun. 2019.

[32] A. Amin, B. Shah, A. M. Khattak, T. Baker, H. ur Rahman Durani, and S. Anwar, "Just-in-time customer churn prediction: With and without data transformation," in *Proc. IEEE Congr. Evol. Comput.*, Jul. 2018, pp. 1–6.

[33] A. Amin *et al.*, "Customer churn prediction in the telecommunication sector using a rough set approach," *Neurocomputing*, vol. 237, pp. 242–254, May 2017.

[34] A. Amin *et al.*, "Comparing oversampling techniques to handle the class imbalance problem: A customer churn prediction case study," *IEEE Access*, vol. 4, pp. 7940–7957, 2016.

[35] M. Ahmed, H. Afzal, A. Majeed, and B. Khan, "A survey of evolution in predictive models and impacting factors in customer churn," *Adv. Data Sci. Adapt. Anal.*, vol. 9, no. 3, Jul. 2017, Art. no. 1750007.

[36] R. Rajamohamed and J. Manokaran, "Improved credit card churn prediction based on rough clustering and supervised learning techniques," *Cluster Comput.*, vol. 21, no. 1, pp. 65–77, Mar. 2018.

[37] B. Zhu, B. Baesens, and S. K. van den Broucke, "An empirical comparison of techniques for the class imbalance problem in churn prediction," *Inf. Sci.*, vol. 408, pp. 84–99, Oct. 2017.

[38] E. Stripling, S. van den Broucke, K. Antonio, B. Baesens, and M. Snoeck, "Profit maximizing logistic model for customer churn prediction using genetic algorithms," *Swarm Evol. Comput.*, vol. 40, pp. 116–130, Jun. 2018.

[39] A. Mishra and U. S. Reddy, "A novel approach for churn prediction using deep learning," in *Proc. IEEE Int. Conf. Comput. Intell. Comput. Res.*, Dec. 2017, pp. 1–4.

[40] S. Mitrović, B. Baesens, W. Lemahieu, and J. D. Weerdt, "On the operational efficiency of different feature types for telco churn prediction," *Eur. J. Oper. Res.*, vol. 267, no. 3, pp. 1141–1155, Jun. 2018.

[41] A. D. Caigny, K. Coussement, and K. W. D. Bock, "A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees," *Eur. J. Oper. Res.*, vol. 269, no. 2, pp. 760–772, Sep. 2018.

[42] M. Hassouna, A. Tarhini, T. Elyas, and M. S. AbouTrab. (Jan. 2016). "Customer churn in mobile markets a comparison of techniques." [Online]. Available: https://arxiv.org/abs/1607.07792

IRFAN ULLAH received the bachelor's degree from the Institute of Management Sciences (IMSciences), Peshawar, Pakistan. He is currently pursuing the master's degree in computer science with COMSATS University Islamabad (CUI), Pakistan. His research interests include big data, data mining, and machine learning.

BASIT RAZA received the Ph.D. degree in computer science, in 2014. He is currently an Assistant Professor with the Department of Computer Science, COMSATS University Islamabad (CUI), Islamabad, Pakistan. He has published several conference and journal papers of internal repute. His research interests include database management systems, security and privacy, data mining, data warehousing, machine learning, and artificial intelligence.

AHMAD KAMRAN MALIK received the Ph.D. degree from the Vienna University of Technology (TU-Wien), Austria. He is currently an Assistant Professor with the Department of Computer Science, COMSATS University Islamabad (CUI), Islamabad, Pakistan. He has published many articles in journals of international repute. His current research interests include data science, social network analysis, and information security.

MUHAMMAD IMRAN received the B.Sc. degree in software engineering from the University of Engineering and Technology, Taxila, Pakistan, in 2006, and the master's degree in software engineering and the Ph.D. degree in computer science from the University of Southampton, U.K., in 2009 and 2015, respectively. From 2007 to 2008, he was a Lecturer with COMSATS University Islamabad (CUI), Islamabad, Pakistan, where he is currently an Assistant Professor with the Department of Computer Science. He secured the Faculty Development Scholarship from CIIT. His research interests include social network analysis, artificial intelligence, and semantic web.

SAIF UL ISLAM received the Ph.D. degree in computer science from University Toulouse III Paul Sabatier, France, in 2015. He was an Assistant Professor with COMSATS University Islamabad, Islamabad, Pakistan, for three years. He has been a part of the European Union-funded research projects during his Ph.D. degree. He was a Focal Person of a research team at COMSATS working in O2 project in collaboration with CERN, Switzerland. He is currently an Assistant Professor with the Department of Computer Science, Dr. A. Q. Khan Institute of Computer Sciences and Information Technology, Rawalpindi, Pakistan. His research interests include resource and energy management in large-scale distributed systems [edge/fog, cloud, and content distribution network (CDN)] and the Internet of Things (IoT).

SUNG WON KIM received the B.S. and M.S. degrees from the Department of Control and Instrumentation Engineering, Seoul National University, South Korea, in 1990 and 1992, respectively, and the Ph.D. degree from the School of Electrical Engineering and Computer Sciences, Seoul National University, in 2002. From 1992 to 2001, he was a Researcher with the Research and Development Center, LG Electronics, South Korea. From 2001 to 2003, he was a Researcher with the Research and Development Center, AL Tech, South Korea. From 2003 to 2005, he was a Postdoctoral Researcher with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA. In 2005, he joined the Department of Information and Communication Engineering, Yeungnam University, South Korea, where he is currently a Professor. His research interests include resource management, wireless networks, mobile networks, performance evaluation, and embedded systems.