Computing
https://doi.org/10.1007/s00607-021-00908-y

REGULAR PAPER

Customer churn prediction system: a machine learning approach

Praveen Lalwani[1] · Manas Kumar Mishra[1] · Jasroop Singh Chadha[1] · Pratyush Sethi[1]

Received: 19 June 2020 / Accepted: 12 January 2021

Abstract
The customer churn prediction (CCP) is one of the challenging problems in the telecom industry. With the advancement in the field of machine learning and artificial intelligence, the possibilities to predict customer churn has increased significantly. Our proposed methodology, consists of six phases. In the first two phases, data preprocessing and feature analysis is performed. In the third phase, feature selection is taken into consideration using gravitational search algorithm. Next, the data has been split into two parts train and test set in the ratio of 80% and 20% respectively. In the prediction process, most popular predictive models have been applied, namely, logistic regression, naive bayes, support vector machine, random forest, decision trees, etc. on train set as well as boosting and ensemble techniques are applied to see the effect on accuracy of models. In addition, K-fold cross validation has been used over train set for hyperparameter tuning and to prevent overfitting of models. Finally, the obtained results on test set have been evaluated using confusion matrix and AUC curve. It was found that Adaboost and XGboost Classifier gives the highest accuracy of 81.71% and 80.8% respectively. The highest AUC score of 84%, is achieved by both Adaboost and XGBoost Classifiers which outperforms over others.

Keywords Customer Churn Prediction · Machine Learning · Predictive Modeling · Confusion Matrix · AUC Curve

✉ Praveen Lalwani
praveen.lalwani@vitbhopal.ac.in

Manas Kumar Mishra
manaskumar.mishra@vitbhopal.ac.in

Jasroop Singh Chadha
jasroop.singh2018@vitbhopal.ac.in

Pratyush Sethi
pratyush.sethi2018@vitbhopal.ac.in

1 VIT Bhopal University, Bhopal, India

Published online: 14 February 2021

Mathematics Subject Classification 68T01 · 68T05

1 Introduction

The globalization and advancements of telecommunication industry, exponentially raises the number of operators in the market that escalates the competition [9]. In this competitive era, it has become mandatory to maximize the profits periodically, for that various strategies have been proposed, namely, acquiring new customers, up-selling the existing customers & increasing the retention period of existing customers. Among all the strategies, retention of existing customers is least expensive as compared to others. In order to adopt the third strategy, companies have to reduce the potential customer churn i.e., customer movement form the one service provider to other. The main reason of churn is the dissatisfaction of consumer service and support system. The key to unlock solutions to this problem is by forecasting the customers which are at risk of churning [18,27,34].

One of the main aim of Customer Churn prediction is to help in establishing strategies for customer retention. Along with growing competition in markets for providing services, the risk of customer churn also increases exponentially. Therefore, establishing strategies to keep track of loyal customers (non-churners) has become a necessity. The customer churn models aim to identify early [43] churn signals and try to predict the customers that leave voluntarily. Thus many companies have realized that their existing database is one of their most valuable asset [11] and according to Abbasdimehr, [1] churn prediction is a useful tool to predict customers at risk.

1.1 Problem description

In order to capture the aforementioned problem, company should predict the customer's behaviour correctly. Customer churn management can be done in two ways: (1) Reactive & (2) Proactive. In the reactive approach, company waits for the cancellation request received from the customer, afterwards, company offers the attractive plans to the customer for the retention. In the proactive approach, the possibility of churn is predicted, accordingly the plans are offered to the customers. Its a binary classification problem where churners are separated from the non churners.

In order to tackle this problem, machine learning has proved itself as a highly efficient technique, for forecasting information on the basis of previously captured data [3,42,45], which includes linear regression, support vector machine, naïve bayes, decision tree, random forest, etc.

In machine learning models, after pre-processing feature selection plays a significant role to improve the classification accuracy. A plenty of approaches were developed by researchers for feature selection that are useful to reduce the dimension, computation complexity & overfitting. In churn prediction, those feature are extracted from the given input vector which are useful for the prediction of churn.

In this work, to tackle this problem we have used the following Machine Learning techniques: (1). Logistic Regression, (2) Naive Bayes, (3) Support Vector Machine,

(4) Decision Trees, (5) Random Forest Classifier, (6) Extra Tree Classifier and Boosting Algorithm such as Ada Boost, XGBoost & CatBoost. Furthermore, for better understanding of the data, the data have been pre-processed and important feature vectors have been extracted using gravitational search algorithm (GSA). To use suitable Machine learning methods, the linearity of the data has also been checked and analyzed.

1.2 Author's contribution

Summary of our contribution is as follows:

- We have applied gravitational search algorithm to perform feature selection and to reduce the dimensions of the data-set.
- After, pre-processing of data, we have applied some of the famous machine learning techniques which are used for predictions like logistic regression, SVM, etc. and k-fold cross validation has been performed to prevent overfitting.
- Then we have used the power of ensemble learning in order to optimize algorithms and achieve better results.
- Then we have evaluated the algorithms on test set using confusion matrix and AUC curve, which have been mentioned in form of graphs and tables in order to compare which algorithm performs best for this particular data-set.

1.3 Organization of research article

The rest of the paper is organized as follows. Next , consists of the work carried previously on this complex problem i.e., Customer Churn Prediction. Important preliminaries such as gravitational search algorithm, machine learning models etc. are presented in sect. 3. The proposed terminology to predict Customer Churn is discussed & presented in sect. 4. In sect. 5, confusion matrix and AUC curve of various machine learning models for performance evaluation is presented and discussed. Finally, sect. 6 concludes the paper.

2 Literature review

This presents a short summary of churn prediction in telecom industry as well as related work proposed by renowned researchers [2,7,12,20,21,23,27,28,31,35,38–40].

Adbelrahim et al. [3], author's applied tree based algorithms for the customer churn prediction, namely, decision tree, random forest, GBM tree algorithm, and XGBoost. In comparative analysis, XGBoost performed superior than others in terms of AUC accuracy. However, accuracy can be further improved using the optimization algorithms for the feature selection process.

Praveen et al. [5], provided comparative analysis of machine learning models for customer churn prediction, where, they adopted support vector machine, decision tree, naive bayes, & logistic regression. Thereafter, they also observed the effect of boosting algorithms on the classification accuracy. In the obtained results, SVM-POLY using AdaBoost performed better than others. However, the classification accuracy can be further improved by incorporating feature selection strategies such as uni-variate selection and others.

Horia Beleiu et al. [7], they adopted three machine learning approaches, namely, neural network, support vector machine and bayesian networks for customer churn prediction. In the feature selection process, principle component analysis (PCA) is taken into consideration to reduce the dimensions of the data. But, the feature selection process can be improved using optimization algorithm which increases the classification accuracy. In the performance evaluation, gain measure and ROC curve was used.

J. Burez et al. [8], author's tried to capture the class imbalance problem. They applied logistic regression and random forest with re-sampling technique. In addition, boosting algorithms were also applied. In the performance analysis, AUC and Lift are taken into consideration. They also observed the effect of advanced sampling techniques such as CUBE, but the obtained outcome did not improve the performance. However, still the class imbalance problem can be solved in a better way by using the optimization based sampling techniques.

K Coussement et al. [11], author's tried to capture the churn prediction problem using support vector machine, logistic regression(LR) and random forest(RF). Initially, performance of SVM was nearly equal to LR and RF, but, when optimal parameter selection was taken into consideration then SVM outperforms over both LR & RF in terms of PCC and AUC.

K. Dahiya et al. [12], researchers applied the two machine learning models, namely, decision tree and logistic regression on churn prediction data-set. In experimentation, WEKA tool was used. However, aforementioned problem can be solved in an efficient way by adopting other machine learning techniques.

Umman et al. [16], author's analyzed the mass data base using logistic regression and decision tree machine learning models, but, obtained accuracy was low. Therefore, further improvement is required for that other machine learning and feature selection techniques can be adopted.

J. Hadden et al. [17], analyze the variables that impact churn in reverence. They also provided the comparative study of three machine learning models such as neural network, regression trees and regression. The obtained results confirm that decision tree is superior than others due to its rule based architecture. The obtained accuracy can be further improved using the existing feature selection techniques.

J. Hadden et al. [18], review of all the machine learning models taken into the consideration as well as they presented deep analysis of existing feature selection techniques. In the prediction models, they found that decision tree performed superior than others. In feature selection, optimization techniques also play a vital role that improves the prediction techniques. After the comparative analysis of existing techniques, author's suggested the path for the future research directions.

Y. Huang et al. [20], author's applied various classifiers on churn prediction dataset, in which the obtained results confirmed that random forest performs superior than others in terms of AUC and PR-AUC analysis. But, accuracy can be further improved using the optimization techniques for the feature extraction.

A. Idris et al. [21], researchers tried the combination of genetic programming(GP) and adaboost machine learning model and then made a comparison with other classification models. The obtained accuracy of GP and adaboost was superior than others. But, accuracy can be further improved using the other optimization techniques such as gravitational search algorithm, bio-geography based optimization and many others.

P. Kisioglu et al. [23], authors applied bayesian belief networks(BBN) for customer churn prediction. In the experimental analysis, correlation analysis and multi-colinearity tests were performed. It was observed that BBN was a good choice for the churn prediction. They also suggested directions for the future research.

2.1 Advantage of proposed technique over the existing

The merits of the proposed algorithm has listed as follows:

- We have applied gravitational search algorithm to perform feature selection and to reduce the dimensions of the data-set, in contrast to existing approaches where prediction accuracy is low due to improper feature selection [8,16,17,20].
- After, pre-processing of data, we have applied some of the famous machine learning techniques which are used for predictions like logistic regression, SVM, etc. and k-fold cross validation has been performed to prevent overfitting, in contrast to recent techniques where overfitting prevention mechanism is not taken into the consideration [20].
- Then we have used the power of ensemble learning in order to optimize algorithms and achieve better results, in contrast to the existing techniques where power of ensemble learning is not taken into consideration, therefore, the obtained accuracy was low [7,11].
- Then we have evaluated the algorithms on test set using confusion matrix and AUC curve, which have been mentioned in the form of graphs and tables in order to compare which algorithm performs best for this particular data-set, in contrast to the existing techniques where obtained results are not properly evaluated [16,18].

3 Preliminaries

In the current , we have tried to describe the notations & abbreviation, techniques we have used for data cleaning and pre-processing in order to make the predictions more robust and machine learning models applied for the classification.

3.1 Notations and abbreviations

In this , description of notations taken into consideration in this article is provided and presented in Table 1.

3.2 Gravitational search algorithm

Various types of optimization techniques can be applied for the different types of segmentation such as particle swarm optimization (PSO), Optics Inspired Optimization

Table 1 Description of notation used in proposed methodology

Notations	Abbreviations
M_{pi}	Mass of the i^{th} passive agent
M_{ai}	Mass of the j^{th} active agent
F_{ij}^{d}	Force between i and j
R_{ij}	Euclidean distance between i and j
$x_i^d(t)$	d^{th} dimension of passive agent i
$x_j^d(t)$	d^{th} dimension of active agent j
$m_i(t)$	i^{th} agent mass at time t
$M_i(t)$	i^{th} agent inertia Mass
$M_i(t)$	i^{th} agent inertia Mass
$A_i(t)$	i^{th} agent acceleration
$V_i(t)$	Velocity of i^{th} agent
h	Its a random number between 0 and 1.

(OIO) [24], and Bio-geography Based Optimization (BBO) [26], and Genetic Algorithm (GA) [6,29]. All evolutionary and swarm intelligence based algorithms needs parameter description before applying to the specific problem, namely, size of population, dimension of individual population member, as well as predefined algorithm dependent parameters. The performance of algorithm for capturing the approximate solutions depends on the fine tuning of algorithm parameters. Rashedi et al. proposed a gravitational search algorithm (GSA) inspired from the law of gravity [25]. It was observed that GSA performs better than well stable optimization techniques such as PSO, GA and SA, when it was tested on various benchmark functions. This is the motivation to apply GSA on image segmentation in the proposed work. Flow of GSA is presented in Fig. 1 and can be described as follows: This algorithm is inspired from the law of gravity. The search agents are modelled as collection of objects which interact with each other based on Newtonian physics. Every mass represents a solution and the algorithm has to adjust between gravitational and inertial mass and the masses will be attracted by the heaviest of them all which will present an optimum solution in the search space. The force acting on the heaviest object drifts it apart from the rest of the population which is basically the optimal solution.

3.2.1 Force estimation:

When agent j acts on agent i, the force is given as:

$$G(t)\frac{(M_{pi}(t) * M_{aj}(t))}{R_{ij} + \epsilon}({x_i}^d(t) - {x_j}^d(t)) \tag{1}$$

The total force acting on iteration t is

$${F_i}^d(t) = \sum_{j \epsilon Kbest, j!=i} rand_j F_i j^d(t) \tag{2}$$

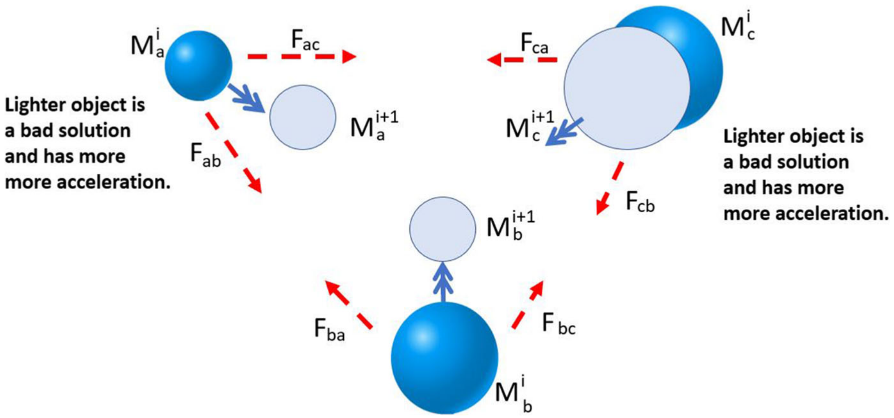


Fig. 1 Gravitational Search Algorithm

3.2.2 Mass estimation using the fitness value:

The inertial mass is estimated with the help of previous equations are as follows:

$$m_i(t) = \frac{fit_i - worst(t)}{best(t) - worst(t)} \tag{3}$$

$$M_i(t) = \frac{m_i(t)}{\sum_{j=1}^{N} m - j(t)} \tag{4}$$

3.2.3 Acceleration:

Finally, the acceleration is calculated as follows:

$$A_i(t) = \frac{F_i(t)}{M_i(t)} \tag{5}$$

3.2.4 Velocity and position update:

In this sub, mathematical equations of velocity and position are shown. Both the equations are applied after generating the acceleration value.

$$V_i^d(t+1) = h * V_i^d(t) + A_i^d(t) \tag{6}$$

$$A_i^d(t+1) = A_i^d(t) + V_i^d(t+1) \tag{7}$$

A sample scenario is illustrated below:

GSA is used to solve the image segmentation problem in the proposed work, which is a non-linear optimization problem.

3.3 Exploratory data analysis (EDA)

It is a way of exploring the hidden features that are present in the rows and columns of data by visualizing, summarizing and interpreting of data. Some of the data visualizations can bee seen in Fig. 2.

Illustration of Fig 2: The distribution of train set attributes over target variable has been shown in Figs 2(a), (b), (c), (d) and (f), whereas, (e) part of Fig. 2 shows that how monthly charges are distributed over total services.

Once EDA is done, meaningful insights are drawn that can be used for supervised and unsupervised machine learning modelling. Some different techniques can also be used to gather more information and insights about customers by following innovative solutions [41]. In our telecommunication data-set we divided the data-set into two parts that is 1st Categorical features and 2nd Numerical features. From 21 features, 16 features were categorical and 5 were numerical as shown in table 2. After preprocessing by dropping null values and replacing keywords graphs were plotted for both categorical features and numerical features.

3.4 Machine learning models

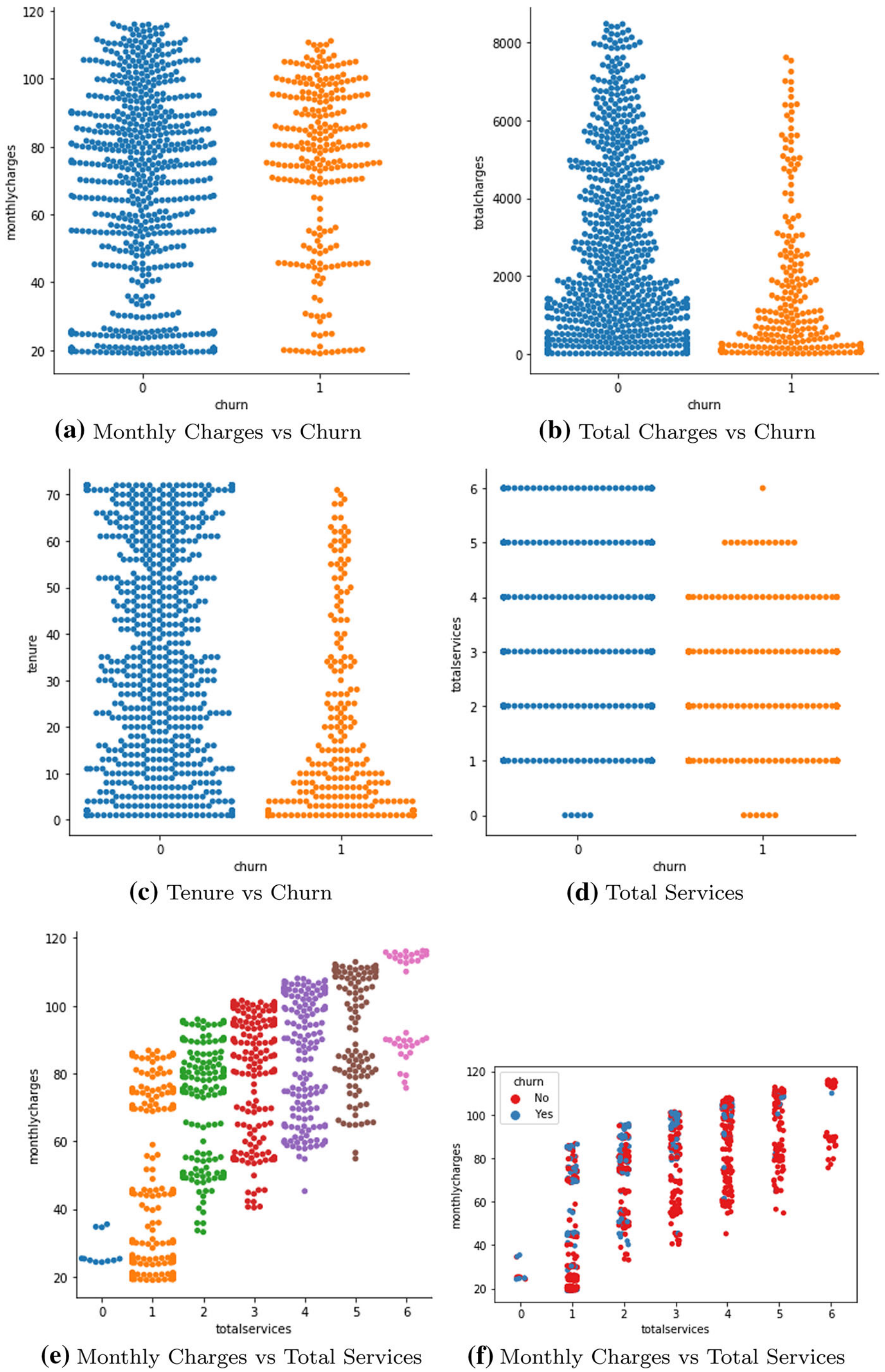
In the following, five well casted and popular techniques used for churn prediction has been presented succinctly, under the canopy of facts considered such as reliability, efficiency, and popularity in the research community [16,17,22,30,33,36].

3.4.1 Regression analysis-logistic regression analysis

Regression is one of the statistical process for estimating how the variables are related to each other. It includes ample amount of techniques for establishing the model and analyzing several variables, when the epicenter of importance is on the bond which is shared between a dependent variable and one or many independent variables. In the light of customer churning, regression analysis is not broadly used because linear regression models are useful for predicting continuous values. But, Logistic Regression or Logit Regression analysis (LR) is a probabilistic statistical classification model. It is also used for binary classification or binary prediction of a categorical value (e.g., house rate prediction, customer churn) which depends upon one or more parameters (e.g., house features, customer features). In addressing the complex problem of customer churn prediction problem, data first has to be casted under proper data transformation from the initial data in order to achieve good performance and sometimes it performs [16] as good as Decision Trees [33].

3.4.2 Naïve Bayes

Naive Bayes classifier is a probabilistic approach in which each vector feature is considered as independent of each other. Naive Bayesian classifiers assume that the value of each feature has an independent influence on a given class, and this assumption is called class conditional independence that is used to simplify the computation, and



(a) Monthly Charges vs Churn **(b)** Total Charges vs Churn

(c) Tenure vs Churn **(d)** Total Services

(e) Monthly Charges vs Total Services **(f)** Monthly Charges vs Total Services

Fig. 2 Exploratory Data Analysis ((a)Monthly Charges vs Churn; (b)Total Charges vs Churn; (c) Tenure vs Churn; (d)Monthly Charges vs Total Services; Monthly Charges vs Total Services two plots (e) and (f)

Table 2 Feature vector and their types

Feature vectors	Types
Customer id	alpha numeric
gender	categorical
Senior citizen	numeric
Partner	categorical
Dependents	categorical
tenure	numeric
Phone service	categorical
Multiple lines	categorical
Internet service	categorical
Online security	categorical
Online backup	categorical
Device protection	categorical
Tech support	categorical
Streaming Tv	categorical
Streaming movies	categorical
Contract	categorical
Paperless billing	categorical
Payment method	categorical
Monthly charges	numeric
Total charges	numeric
Churn	categorical

in this sense, we call it “Naive” [13]. In simple terms that this classifier assumes that the presence of feature vector (customer churn) is independent from the other feature vectors that are present in the class. The Naïve Bayes classifier is not regarded as a good classifier for large data-set but as our data-set was only about 7000 instances. It showcased good results.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{8}$$

3.4.3 Support vector machine

In machine learning, Support Vector machine also Known as Support Vector Networks introduced by Boser, Guyon, and Vapnik [5] are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. What support vector machine is trying to do is, it divides the prediction into two parts +1 that is right side of the hyperplane and –1 that is left side of the hyperplane. The hyperplane is of width twice the length of margin. Depending on the type of data i.e. (scattered on the graph) tuning parameter like kernels are used like linear, poly, rbf, callable, pre-calculated [46]. Support Vector machine provides high accuracy than Naïve Bayes and Logistic Regression.

3.4.4 Decision trees

It works on the greedy approach and uses a series of rules for classification. Alternately, this approach elucidates the high categorization accuracy rate it fails to respond to data having noise. The main parameter to decide the root node parameter of decision tree is gain. The decision trees generated by C4.5 can be used for classification and for this reason C4.5 is often referred to as a statistical classifier [37].

3.4.5 Random forest classifier

It works on the divide and conquer approach. It is based on the random subspace method [19]. In this method a number of trees are formed and each decision tree is trained by selecting any random sample of attributes from the predictor attributes set. Each tree matures up to maximum extent based on the attributes or parameters present. The final decision tree is formed for the prediction mainly based on weighted averages. It has the ability to handle thousands of input parameters without deletion. It can also handle the missing values inside the data-set for training the predictive model.

3.4.6 Extra tree classifier

Extra Tree Classifier also called Extreme Randomized Tree Classifier is a type of ensemble learning technique which aggregates the result of multiple de-correlated decision trees collected in a forest to output its classification result. While in comparison with Random Forest Classifier it only differs from it in the manner of construction of the decision trees in the forest. This implements a meta estimator that fits a number of randomized decision trees (extra trees) on various sub-samples of the data-set and uses averaging to improve the predictive accuracy and control over – fitting. In Churn prediction it performed better than all the process and gave good accuracy

3.4.7 Boosting algorithm: adaboost

Ada – boost like Random Forest Classifier is another ensemble classifier. (Ensemble classifier are made up of multiple classifier algorithms and whose output is combined result of output of those classifier algorithms). A single algorithm may perform poorly in classification of the objects. But when combined with boosting ensemble algorithms like Ada-boost and selection of training set at every iteration and assigning right amount of weight in final voting, we can obtain good accuracy score for overall classifier. In short Ada -boost retrains the algorithm iteratively, by choosing the training set based on accuracy of previous training. Ada boost classifier increased the performance, accuracy after combing with Random forest classifier, Decision Trees classifier and Extra Tree Classifier in prediction of the Churn of the telecommunication data-set. Similarly, many boosting techniques or algorithms can be optimized for better performances like [44].

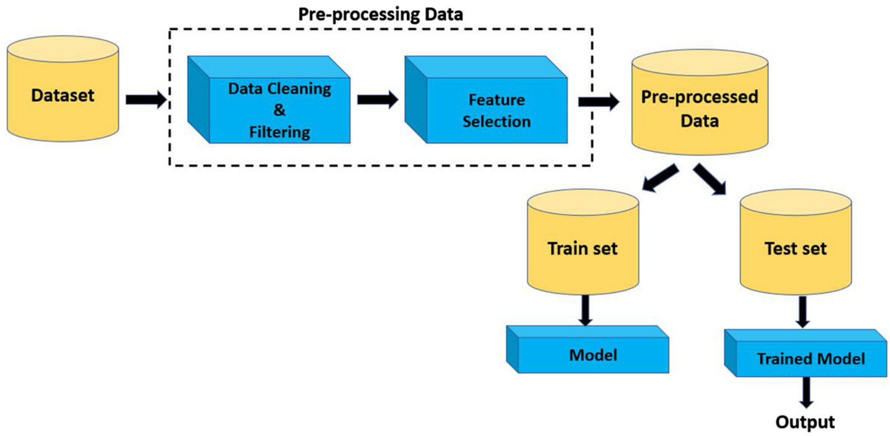


Fig. 3 System Architecture

3.4.8 XGBoost classifier

XGBoost implements decision tree algorithm with gradient boosting. The gradient boosting follows an approach where new models are used to compute the error or residuals of previously applied model and then both are combined to make the final prediction. It also uses gradient descent to locate the minima or reduce the value of loss function.

3.4.9 CatBoost classifier

CatBoost is also a gradient boosting decision tree algorithm but it uses symmetric trees, which in turn decreases the prediction time. After computing the pseudo residuals, it updates the base model in order to produce better results. The major advancement of catboost is that it includes some of most commonly used pre-processing methods like one hot encoding, label encoding,etc. which in turn decreases the pre-processing effort but not completely eliminates the data pre-processing step. It does not include all statistical measures for data pre-processing.

4 Proposed work

This consists of system architecture, algorithm and description of proposed work.

4.1 System architecture

In this sub, pictorial representation of system architecture is shown in Fig. 3 which includes various phases, namely, Data pre-processing and feature selection, Splitting of Pre-processed Data into train and test set, training and testing of models respectively.

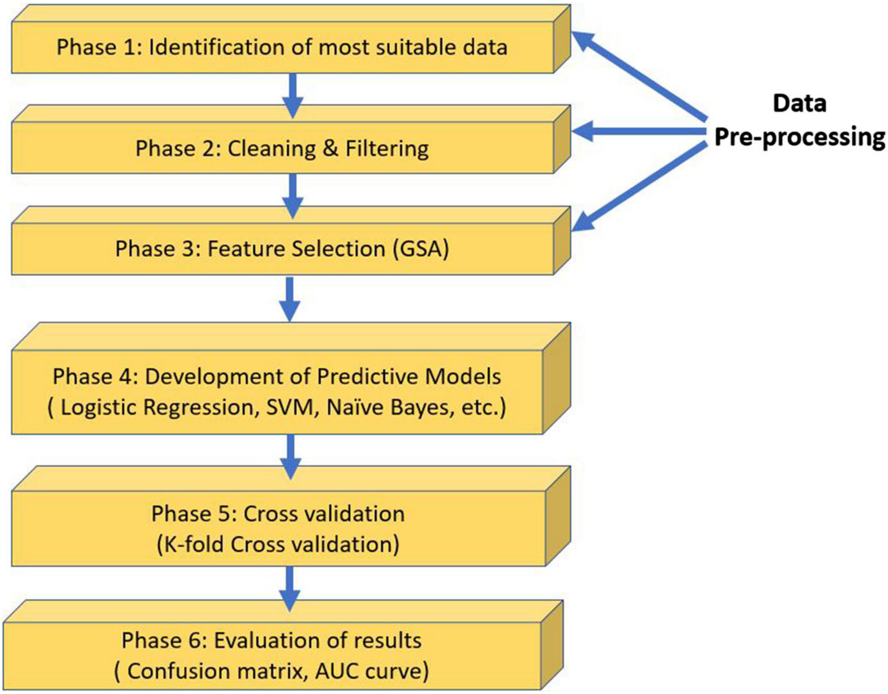


Fig. 4 Multiple phase model for developing a customer churn management framework

4.2 Description of proposed model

This consists of various phases of the proposed model. It consists of five phases, namely, Phase 1: Identification of most suitable data (variance analysis, correlation matrix, outliers removal, etc.), Phase 2: Cleaning & Filtering (handling null and missing values) and Phase 3: Feature Selection (using GSA). Phase 4: Development of predictive models (Logistic Regression, SVM, Naive Bayes, etc.). Phase 5: Cross validation (using k-fold cross validation). Finally, the evaluation of predictive models on test set (using Confusion matrix & AUC curve) has been presented in phase 6.

4.2.1 Pre-processing of data: phase 1, phase 2, phase 3

Data pre-processing is one of the important techniques of data mining which helps to clean and filter the data. Thus, removing the inconsistencies and converting raw data into a meaningful information which can be managed efficiently. It is important to remove null values or missing values in the data-set and to check the data-set for imbalanced class distributions, which has been one of the emerging problems of data mining [15]. The problem of imbalanced data-set can be solved through re-sampling techniques [32], by enhancing evaluation metrics [8], etc.
Phase 1: Identification of most suitable data: In order to establish a customer churn predictive model, firstly, select the important data or information from raw data in

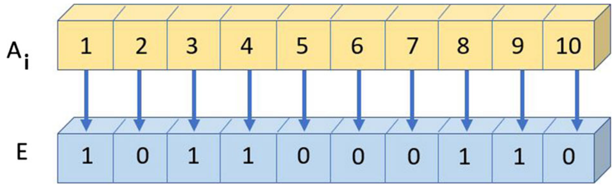


Fig. 5 Agent Representation

order to develop an efficient predictive model. For identification of important data variance analysis has been adopted. Then correlation matrix is used to study the intra-relationship between the attributes. For class balancing dummy rows have been added by using re-sampling techniques. [15,32].

Phase 2: Cleaning & Filtering: This phase consists of data cleaning and filtering by removing missing values, non-relevant parameters, etc. Data cleaning is the key to reduce dimensions of the data-set. As the dimension increases, more time and power of computation is required. In the proposed methodology, data visualization is taken into consideration for understanding or extracting deeper insights from the data [4].

Phase 3: Feature Selection (An Optimized Approach): The main aim of feature selection is to eliminate the non-significant features which remains constant or have no significant dispersion for all instances. In this phase, initially uni-variate selection is applied, afterwards gravitational search algorithm (GSA) is adopted for the feature selection process. In GSA, agent is encoded in binary format, where *1* represents the selected feature, whereas, *0* represents not selected. The dimension of agent A_i is equal to the all available features in the data set.

Derivation of Fitness Function: The objective of feature selection problem is to minimize the error rate, which increases the classification accuracy. In GSA, error rate considered as fitness function which is shown in Eq. 9 and objective is to minimize it.

$$ErrorRate = \frac{FP + FN}{TP + TN + FP + FN} \tag{9}$$

where, false positive, false negative, true positive and true negative represented by *FP*, *FN*, *TP*, and *TN* respectively.

4.2.2 Development of predictive models: phase 4

Phase 4: In this phase predictive models are applied to make predictions. In order to optimize the results obtained from various classifiers, we have applied some existing techniques, namely, ensemble learning (Adaboost, Extra trees, XGBoost, etc.).

Therefore, in the proposed methodology various models are applied, namely, Logistic Regression, Decision trees, Random forest, Naive Bayes, Adaboost Classifier, KNN Classifier, SVM Classifier Linear, Logistic Regression (Adaboost), Adaboost Classifier(Extra tree), Random Forest (Adaboost), SVM Classifier Poly, SVM (Adaboost), XGBoost Classifier and CatBoost Classifier to make the predictions. The obtained

Table 3 k-fold cross validation results for all models

Model	k-fold cross validation (cv=5)%
Logistic regression	79.85
Decision tree	79.56
Adaboost classifier	80.72
Adaboost classifier (Extra Tree)	80.41
KNN classifier	78.51
Random forest	79.28
Random forest (adaboost)	80.39
Naive bayes (gaussian)	75.86
SVM classifier linear	78.65
SVM classifier poly	79.75
SVM (adaboost)	73.48
XGboost classifier	79.5
CatBoost classifier	80.34

results of all the classifiers are mentioned in Sect. 5. Further the models and their respective hyperparameters have been fine tuned using k-fold cross-validation.

4.2.3 K - fold cross validation: phase 5

Phase 5:

It's a re-sampling procedure used to evaluate machine learning models on a limited data sample. The procedure has a single parameter called as *k*, which refers to the number of splitted groups in a given data sample. The *k - Fold Cross Validation* shuffles the data-set randomly, then splits the train set into k groups. From the splitted groups one group is randomly chosen as a test set and remaining as train sets. Thereafter, the model is fitted and the score is validated on unseen data. The results obtained from k-fold cross validation is shown in Table 3 :

It turns out the k- Fold Cross validation has been applied for fine tuning the models and prevent them from overfitting on train set.

4.2.4 Evaluation of results: phase 6

Phase 6 Model evaluation is the key for analysing the performance of the proposed model. For model evaluation confusion matrix and AUC curve are taken into consideration, which has been described in Sect. 5. Then we have compared the results in order to identify the best performing model for the data-set.

4.3 Algorithm of proposed churn prediction model

Algorithm 1: Proposed algorithm for Churn Prediction

Result: Classifier labels for test instances
Input: The train data-set consisting of input features such as x1, x2, x3, x4 and output label y;
Output: Predicted Labels (churn or non-churn);
Procedure;
1. Identification of most suitable data using Variance Analysis, Correlation matrix, etc.;
2. Cleaning & Filtering (handling null and missing values).;
3. Feature Selection using Gravitational Search Algorithm ;
4. Application Predictive Models using Logistic Regression, SVM, etc.;
5. Evaluation of Results using Confusion matrix and AUC curve;

5 Performance analysis

5.1 Confusion matrix

To evaluate the performance of applied models or throughput of Customer Churn Prediction on the test set, different metrics have been used, namely, precision, recall, accuracy and F -measure [39]. It measures the ability of the predictive models for forecasting the churning customers correctly [10]. The aforementioned four measures are calculated from the information captured using confusion matrix and shown in Table 6. The representation of confusion matrix is shown in Table 4. True positive and false positive are denoted as *Tp* and *Fp*, whereas, false negative and true negative as *Fn* and *Tn*.

The four terms to get familiar with for understanding the evaluation criteria are:

- **True Positive (Tp):** The number of customers that are in the churner category and the predictive model has predicted them correctly.
- **True Negative (Tn):** The number of customers that are in the non-churner category and the predictive model has predicted them correctly.
- **False Positive (Fp):** The number of customers who are non-churners but the predictive algorithm has labelled or identified them as churners.
- **False Negative (Fn):** The number of customers who are churners but the predictive model has labelled or identified them as non-churners.

Table 4 Confusion matrix for evaluation of classifier

	Prediction category	
	Churners	con-churners
churn	Tp	Fn
Non-churn	Fp	Tn

5.1.1 Performance indicators

5.1.2 Recall

It is the ratio of real churners (i.e. True Positive), and is calculated under the following:

$$Recall = \frac{T_p}{T_p + F_n} \tag{10}$$

5.1.3 Precision

It is the ratio correct predicted churners, and is calculated under the following:

$$Precision = \frac{T_p}{T_p + F_p} \tag{11}$$

5.1.4 Accuracy

It is ration of number of all correct predictions, and is calculated under the following:

$$Accuracy = \frac{(T_p + T_n)}{(T_p + F_p + Tn + F_n)} \tag{12}$$

5.1.5 F - measure

It is the harmonic average of precision and recall, and it is calculated under the following:

$$F - measure = \frac{(2 \times Precision \times Recall)}{(Precision + Recall)} \tag{13}$$

A better combined precision and recall achieved by the classifier is implied due to a value closer to one [14].

Table 5 k-fold cross validation results for all models

Model	AUC Score %
Logistic regression	82
Logistic regression (Adaboost)	78
Decision tree	83
Adaboost classifier	84
Adaboost classifier (Extra Tree)	72
KNN classifier	80
Random forest	82
Random forest (adaboost)	82
Naive bayes (gaussian)	80
SVM classifier linear	79
SVM classifier poly	80
SVM (adaboost)	80
XGBoost	84
CatBoost	82

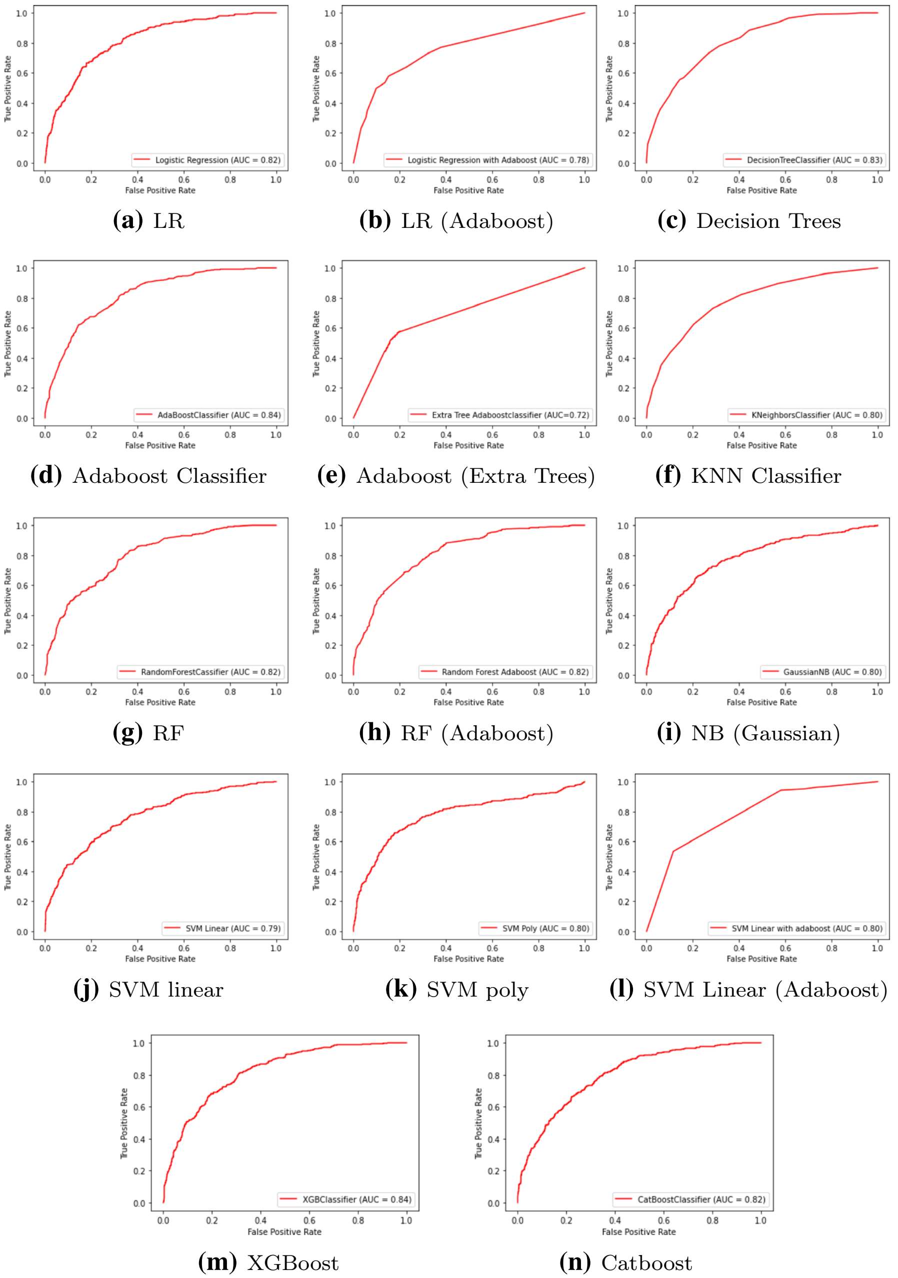
5.2 AUC curve analysis

To quantify the models performance on positive and negative classes of the test set, AUC curve has been used. Higher the value of the AUC score, the better the model performs on both positive and negative classes. The obtained AUC scores of different predictive models which are used to predict the target variable has been represented in Table 5 and Fig. 6. In Fig. 6, (a), (b), (c), (d), (e), (f), (g), (h), (i), (j), (k), (l), (m) & (n) graphically represents the obtained AUC scores of Logistic Regression, Logistic Regression (Adaboost), Decision Trees, Adaboost Classifier, Adaboost Classifier (Extra Trees), KNN Classifier, Random Forest, Random Forest (Adaboost), Naive Bayes (Gaussian), SVM Linear, SVM Poly, SVM Linear (Adaboost), XGBoost Classifier and CatBoost Classifier respectively. In accordance to AUC scores Adaboost classifier and XGBoost Classifier outperforms over other respective algorithms on the test set having an AUC score of 84%.

5.3 Obtained outcome analysis

We tested the final pre-processed data on multiple algorithms such as Logistic Regression, Decision trees, Random forest, Naive Bayes, Adaboost Classifier, KNN Classifier, SVM Classifier Linear, Logistic Regression (Adaboost), Adaboost Classifier(Extra tree), Random Forest (Adaboost), SVM Classifier Poly, SVM (Adaboost), XGBoost Classifier and CatBoost Classifier. The obtained results are mentioned in Table 6. The results are graphically presented in Fig. 7, in which, accuracy, recall, precision and F-measure is represented by Figs 7(a), (b), (c) & (d) respectively.

The LR proved to predict churn with the accuracy of 80.45%, having a good recall of 80.23%, a subtle precision of 79.11%, F – measure of 78.89% and an AUC score of 82%.



(a) LR **(b)** LR (Adaboost) **(c)** Decision Trees

(d) Adaboost Classifier **(e)** Adaboost (Extra Trees) **(f)** KNN Classifier

(g) RF **(h)** RF (Adaboost) **(i)** NB (Gaussian)

(j) SVM linear **(k)** SVM poly **(l)** SVM Linear (Adaboost)

(m) XGBoost **(n)** Catboost

Fig. 6 Models AUC curve (a)Logistic Regression(LR) (b)Logistic Regression (Adaboost) (c) Decision Trees (d) Adaboost Classifier (e) Adaboost (Extra Trees) (f) K-Nearest Neighbor (g) Random Forest (h) Random Forest (Adaboost) (i) Naive Bayes (Gaussian) (j), (k) and (l) represents Support Vector Machines (m) XGBoost Classifier (n) CatBoost Classifier

Table 6 Comparison of machine learning models

Model	Accuracy(%)	Recall(%)	Precision(%)	F-Measure(%)	AUC Score %
Logistic Regression	80.45	80.23	79.11	78.89	82
Logistic Regression (Adaboost)	76.57	75.57	56.61	64.71	78
Decision Tree	80.14	80.1	78.81	78.89	83
Adaboost Classifier	81.71	81.21	80.14	80.28	84
Adaboost Classifier (Extra Tree)	81.14	81.64	80.57	80.60	72
KNN Classifier	79.64	79.71	78.38	77.00	80
Random Forest	78.04	78.68	77.54	77.91	82
Random Forest (Adaboost)	81.21	81.28	80.19	80.29	82
Naive Bayes (Gaussian)	77.07	77.12	77.60	77.31	80
SVM Classifier Linear	79.14	79.89	78.67	78.86	79
SVM Classifier Poly	80.21	80.64	79.66	78.11	80
SVM (Adaboost)	74.07	74.43	54.91	63.17	80
XGBoost	80.8	80.7	80.3	78.7	84
CatBoost	81.8	82.2	81.2	79.6	82

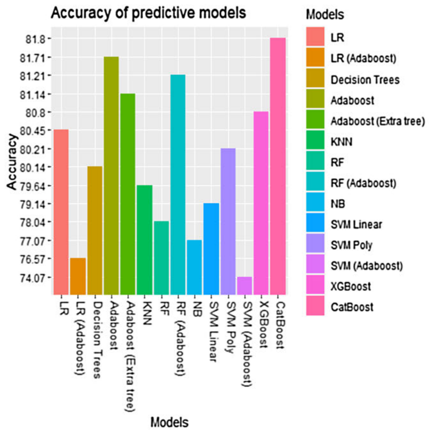
Another model which came out to prove its ability is DT model. It forecasted Customer Churn with accuracy of 80.14%, precision of 78.81%, F – measure of 78.89%, recall of 80.1% and an AUC score of 83%.

Among the tested algorithms, some of them also came out to give significant results like SVM-POLY, SVM-LINEAR, Naïve Bayes, Random Forest and KNN Classifier.

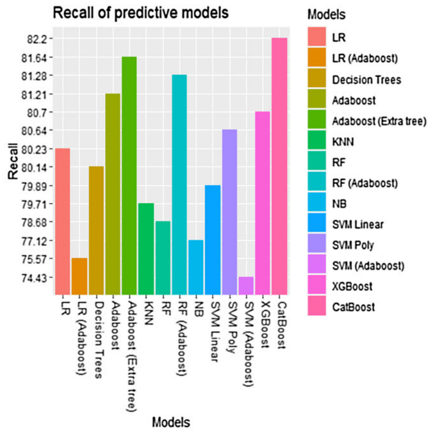
The most prominent predictive model without boosting came out to be LR on our data-set, but DT and SVM-POLY came out to be pretty close and thus, LR came out to be the most significant, having slightly more accuracy then others.

The XGBoost and CatBoost Classifier also gave significant results having good precision, recall, accuracy and F-measure as shown in Table 6. XGBoost performed better than other respective algorithms having an AUC score of 84%.

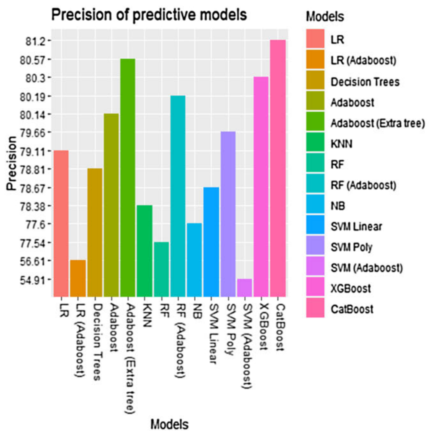
But, with the power of ensemble learning AdaBoost Classifier also gave the highest accuracy with respect to others i.e., 81.71% also having a high recall of 80.21% with good precision and F-measure, along with an AUC score of 84%. Hence, Adaboost Classifier and XGBoost Classifier gives the most significant results.



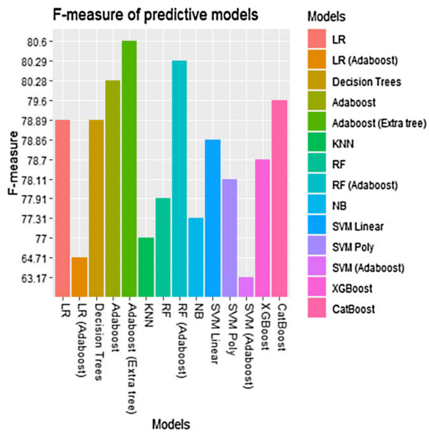
(a) Accuracy of Models



(b) Recall of Models



(c) Precision of Models



(d) F-measure of Models

Fig. 7 Evaluation of Models on Performance Indicators ((a) Accuracy; (b) Recall;(c) Precision; (d) F-measure)

6 Conclusion and future findings

In the 21st century the trend of growth has been proving the most drastic boom ever. With advancement of technology, there comes an increase in services and it is hard for a company to predict the customers who are likely to leave their services. In telecom industry, churn prediction is a problem which has gathered attraction by various researchers in the recent years. Through this research paper we provide a comparative study of Customer Churn prediction in Telecommunication Industry using famous machine learning techniques such as Logistic Regression, Naïve Bayes, Support Vector Machines, Decision Trees, Random Forest, XGBoost Classifier, CatBoost Classifier, AdaBoost Classifier and Extra tree Classifier. The experimental results show that two ensemble learning techniques that is Adaboost classifier and XGBoost classifier gives maximum accuracy with respect to others with an AUC score of 84% for the churn prediction problem with respect to other models. They outperformed other algorithms in terms of all the performance measures such as accuracy, precision, F-measure, recall and AUC score. Churn prediction for a company tends to be a very tedious task and as of many upcoming company's and startups there is a tough competition in the market to retain the customers by providing services that are beneficial to both sides. It is very difficult to predict genuine customers of the company. In future, with the upcoming concepts and frameworks in the field of reinforcement learning and deep learning sector, machine learning is proving to be one of the most efficient way to address problems like churn prediction with better accuracy and precision.

References

1. Abbasimehr H, Setak M, Tarokh M (2011) A neuro-fuzzy classifier for customer churn prediction. International Journal of Computer Applications 19(8):35–41
2. Adwan O, Faris H, Jaradat K, Harfoushi O, Ghatasheh N (2014) Predicting customer churn in telecom industry using multilayer preceptron neural networks: Modeling and analysis. Life Science Journal 11(3):75–81
3. Ahmad AK, Jafar A, Aljoumaa K (2019) Customer churn prediction in telecom using machine learning in big data platform. Journal of Big Data 6(1):28
4. Archambault, D., Hurley, N., Tu, C.T.: Churnvis: visualizing mobile telecommunications churn on a social network with attributes. In: 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013), pp. 894–901. IEEE (2013)
5. Asthana P (2018) A comparison of machine learning techniques for customer churn prediction. International Journal of Pure and Applied Mathematics 119(10):1149–1169
6. Aziz R, Verma C, Srivastava N (2018) Artificial neural network classification of high dimensional data with novel optimization approach of dimension reduction. Annals of Data Science 5(4):615–635
7. Brânduşoiu, I., Toderean, G., Beleiu, H.: Methods for churn prediction in the pre-paid mobile telecommunications industry. In: 2016 International conference on communications (COMM), pp. 97–100. IEEE (2016)
8. Burez J, Van den Poel D (2009) Handling class imbalance in customer churn prediction. Expert Systems with Applications 36(3):4626–4636
9. Chen, H., Chiang, R.H., Storey, V.C.: Business intelligence and analytics: From big data to big impact. MIS quarterly pp. 1165–1188 (2012)
10. Coussement K, De Bock KW (2013) Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning. Journal of Business Research 66(9):1629–1636
11. Coussement K, Van den Poel D (2008) Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. Expert systems with applications 34(1):313–327
12. Dahiya, K., Bhatia, S.: Customer churn analysis in telecom industry. In: 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), pp. 1–6 (2015)
13. Dong, T., Shang, W., Zhu, H.: Naïve bayesian classifier based on the improved feature weighting algorithm. In: International Conference on Computer Science and Information Engineering, pp. 142–147. Springer (2011)
14. Fawcett T (2006) An introduction to roc analysis. Pattern recognition letters 27(8):861–874
15. García S, Fernández A, Herrera F (2009) Enhancing the effectiveness and interpretability of decision tree and rule induction classifiers with evolutionary training set selection over imbalanced problems. Applied Soft Computing 9(4):1304–1314
16. Gürsoy UŞ (2010) Customer churn analysis in telecommunication sector. İstanbul Üniversitesi İşletme Fakültesi Dergisi 39(1):35–49
17. Hadden J, Tiwari A, Roy R, Ruta D (2006) Churn prediction: Does technology matter. International Journal of Intelligent Technology 1(2):104–110

18. Hadden J, Tiwari A, Roy R, Ruta D (2007) Computer assisted customer churn management: State-of-the-art and future trends. Computers & Operations Research 34(10):2902–2917
19. Han J, Pei J, Kamber M (2011) Data mining: concepts and techniques. Elsevier,
20. Huang, Y., Zhu, F., Yuan, M., Deng, K., Li, Y., Ni, B., Dai, W., Yang, Q., Zeng, J.: Telco churn prediction with big data. In: Proceedings of the 2015 ACM SIGMOD international conference on management of data, pp. 607–618 (2015)
21. Idris, A., Khan, A., Lee, Y.S.: Genetic programming and adaboosting based churn prediction for telecom. In: 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1328–1332. IEEE (2012)
22. Kirui, C., Hong, L., Cheruiyot, W., Kirui, H.: Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining. International Journal of Computer Science Issues (IJCSI) **10**(2 Part 1), 165 (2013)
23. Kisioglu P, Topcu YI (2011) Applying bayesian belief network approach to customer churn analysis: A case study on the telecom industry of turkey. Expert Systems with Applications 38(6):7151–7157
24. Lalwani P, Banka H, Kumar C (2017) Crwo: Clustering and routing in wireless sensor networks using optics inspired optimization. Peer-to-Peer Networking and Applications 10(3):453–471
25. Lalwani, P., Banka, H., Kumar, C.: Gsa-chsr: gravitational search algorithm for cluster head selection and routing in wireless sensor networks. In: Applications of Soft Computing for the Web, pp. 225–252. Springer (2017)
26. Lalwani P, Banka H, Kumar C (2018) Bera: a biogeography-based energy saving routing architecture for wireless sensor networks. Soft Computing 22(5):1651–1667
27. Lejeune MA (2001) Measuring the impact of data mining on churn management. Internet Research
28. Massey AP, Montoya-Weiss MM, Holcom K (2001) Re-engineering the customer relationship: leveraging knowledge assets at ibm. Decision Support Systems 32(2):155–170
29. Musheer RA, Verma C, Srivastava N (2019) Novel machine learning approach for classification of high-dimensional microarray data. Soft Computing 23(24):13409–13421
30. Nath SV, Behara RS (2003) Customer churn analysis in the wireless industry: A data mining approach. Proceedings-annual meeting of the decision sciences institute 561:505–510
31. Petrison LA, Blattberg RC, Wang P (1997) Database marketing: Past, present, and future. Journal of Direct Marketing 11(4):109–125
32. Qureshi, S.A., Rehman, A.S., Qamar, A.M., Kamal, A., Rehman, A.: Telecommunication subscribers' churn prediction model using machine learning. In: Eighth International Conference on Digital Information Management (ICDIM 2013), pp. 131–136. IEEE (2013)
33. Radosavljevik D, van der Putten P, Larsen KK (2010) The impact of experimental setup in prepaid churn prediction for mobile telecommunications: What to predict, for whom and does the customer experience matter? Trans. MLDM 3(2):80–99
34. Rajamohamed R, Manokaran J (2018) Improved credit card churn prediction based on rough clustering and supervised learning techniques. Cluster Computing 21(1):65–77
35. Rodan A, Faris H, Alsakran J, Al-Kadi O (2014) A support vector machine approach for churn prediction in telecom industry. International journal on information 17(8):3961–3970
36. Shaaban E, Helmy Y, Khedr A, Nasr M (2012) A proposed churn prediction model. International Journal of Engineering Research and Applications 2(4):693–697
37. Sharma H, Kumar S (2016) A survey on decision tree algorithms of classification in data mining. International Journal of Science and Research (IJSR) 5(4):2094–2097
38. Simons, R.: Siebel systems: Organizing for the customer (2005)
39. Sokolova, M., Japkowicz, N., Szpakowicz, S.: Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation. In: Australasian joint conference on artificial intelligence, pp. 1015–1021. Springer (2006)
40. Tamaddoni Jahromi, A.: Predicting customer churn in telecommunications service providers (2009)
41. Ultsch A (2002) Emergent self-organising feature maps used for prediction and prevention of churn in mobile phone markets. Journal of Targeting, Measurement and Analysis for Marketing 10(4):314–324
42. Umayaparvathi V, Iyakutti K (2016) A survey on customer churn prediction in telecom industry: Datasets, methods and metrics. International Research Journal of Engineering and Technology (IRJET) 4(4):1065–1070
43. Wei CP, Chiu IT (2002) Turning telecommunications call details to churn prediction: a data mining approach. Expert systems with applications 23(2):103–112

44. Xie Y, Li X, Ngai E, Ying W (2009) Customer churn prediction using improved balanced random forests. Expert Systems with Applications 36(3):5445–5449
45. Yu, W., Jutla, D.N., Sivakumar, S.C.: A churn-strategy alignment model for managers in mobile telecom. In: 3rd Annual Communication Networks and Services Research Conference (CNSR'05), pp. 48–53. IEEE (2005)
46. Zhao, Y., Li, B., Li, X., Liu, W., Ren, S.: Customer churn prediction using improved one-class support vector machine. In: International Conference on Advanced Data Mining and Applications, pp. 300–306. Springer (2005)