Results in Control and Optimization 14 (2024) 100342

Contents lists available at ScienceDirect

Results in Control and Optimization

journal homepage: www.elsevier.com/locate/rico

# Customer churn prediction in telecom sector using machine learning techniques

Sharmila K. Wagh [a,*], Aishwarya A. Andhale [b], Kishor S. Wagh [c], Jayshree R. Pansare [a], Sarita P. Ambadekar [d], S.H. Gawande [e]

[a] Department of Computer Engineering, M.E.S. College of Engineering, S.P. Pune University, Pune, Maharashtra 411001, India
[b] Department of Information Technology, MKSSS's Cummins College of Engineering for Women, Pune 411052, India
[c] Department of Computer Engineering, AISSMS Institute of Information Technology, S.P. Pune University, Pune, Maharashtra 411001, India
[d] Department of Computer Engineering, K. J. Somaiya Institute of Technology, Sion, Mumbai, India
[e] Industrial Tribology Laboratory, Department of Mechanical Engineering, M.E.S. College of Engineering, S.P. Pune University, Pune, Maharashtra, 411001, India

A R T I C L E I N F O

*Keywords:*
Churners
Customer churn prediction
Up-sampling
Classifiers
Survival analysis

A B S T R A C T

In the telecom industry, large-scale of data is generated on daily basis by an enormous amount of customer base. Here, getting a new customer base is costlier than holding the current customers where churn is the process of customers switching from one firm to another in a given stipulated time. Telecom management and analysts are finding the explanations behind customers leaving subscriptions and behavior activities of the holding churn customers' data. This system uses classification techniques to find out the leave subscriptions and collects the reasons behind the leave subscription of customers in the telecom industry. The major goal of this system is to analyze the diversified machine learning algorithms which are required to develop customer churn prediction models and identify churn reasons in order to give them with retention strategies and plans. In this system, leave subscriptions collects customers' data by applying classification algorithms such as Random Forest (RF), machine learning techniques such as KNN and decision tree Classifier. It offers an efficient business model that analyzes customer churn data and gives accurate predictions of churn customers so that business management may take action within the churn period to stop churn as well as loss in profit. System achieves an accuracy of 99 % using the random forest classifier for churn predicts, the classifier matrix has achieved a precision of 99 % with a recall factor of 99 % alongwith received overall accuracy of 99.09 %. Likewise, our research work improves churn prediction, scope other business fields, and provide prediction models to hold their existing customers customer service, and avoid churn effectively.

## 1. Introduction

Role of predictive model is to bring the churned customers to light. The proposed model's purpose is to bring churned customers to light. In a targeted approach industry try to identify which customers are likely to churn. The industry then targets those customers or clients and provides them with special incentives, offerings, and plans except normal customers. This approach can bring a huge loss to industry, if churned measures are inaccurate because the industries are wasting a lot of money to the customers who would have stayed

* Corresponding author.
*E-mail address:* skwagh135@gmail.com (S.K. Wagh).

https://doi.org/10.1016/j.rico.2023.100342
Received 30 March 2023; Received in revised form 24 October 2023; Accepted 11 November 2023
Available online 13 November 2023

anyways, irrespective of short or long distance. Communication has become an important part of today's life. It's being used in every field [1–3]. The telecom industry needs to build the best predictive model for churning customers.

Churn Customer refers to the number of existing customers who may leave the service provider over a given period. These customers can be called as churners. The main aim of churn is to predict the churnable customers at the earliest, to identify the reason for churning. The primary goal of churn analysis is to identify and anticipate churnable consumers as soon as possible. This will help us to rectify the issues of the customer. This will be helpful to satisfy the customer needs and will continue to use that service. This will help to meet the needs of the customers, and they will continue to utilize the service. There are promotional costs known as acquisition costs and retention costs in a telco company. The acquisition cost is the price a company pays to gain new consumers. Retention costs, on the other hand, are the costs of keeping existing clients. It is very difficult to predict which customers would churn and which customers will be maintained due to human limitations. As a result, the allocation of money may be incorrect, resulting in a higher amount of cash being issued.

Furthermore, according to some reports, the acquisition cost is 5 times that of the retention cost. If it is incorrect in projecting a client who will churn, but it turns out that we are correct in anticipating a customer who will be kept, we will have to spend more than we should. This paper attempts to develop a Machine Learning model that can predict customer churn prediction and retention. Retaining cost of the company for the existing customers is far economical as compared to addition of new customers within the network. The customer churn is the direct loss in terms of revenue to the company. If the information of the development of churning of customers is known well in advance, then appropriate steps can be taken, and a better service can be provided to such customers. It is observed that the long-term customers add more revenue to the company as they are not much responsive to slight changes. Today's most challenging and critical problems faced by the telecommunication industry is the management of churning customers. Recent studies show that the main objective is to identify the valuable churn customers using a huge amount of data received from the telecommunication industry. Practically there are so many restrictions in using current models, which faces lots of difficulties & hurdles towards the problem of churning in today's present environment. While modeling development, a lot of information-rich features are neglected in In Feature selection Process. Mostly statistical methods are being used in a diverse domain, which tends to give undesirable results of the present predictive models. The feature selection is another huge problem with the existing models. Every customer may be an individual or a group and have different churning reasons. Classification of a churn customer can be as a churner, irrespective observing reasons and factors for churn. There are different patterns of behavior during the churn process and should not treat all of them in the same manner. Some customers may not churn easily than others. Today's need is for a more realistic prediction model which can predict churn customers in advance. This will be of great help to provide strong retention strategies for the different groups of churners may it be different promotions depending upon the churn factors for the different group of churners. Encouraged by the above-mentioned facts & observations, in this study, a model for prediction of churners with the help of different machine learning algorithms is proposed.

Ullah, et al. [1] provided a customer attrition model for data analytics that is validated using common assessment measures. The results demonstrated that utilizing machine learning techniques improved the performance of their proposed churn model. The F-measure result from Random Forest and J48 was 88 % better. The authors used the dataset to identify the primary churn variables and did cluster profiling based on their churn risk. Finally, the authors presented recommendations for telecom decision-makers on client retention. Ammar, et al. [2] presented a hybridized algorithm technique to predict churners that was quite efficient. When compared between the proposed algorithm and normal firefly algorithm, the accuracy obtained by both is found to be similar. Yet the hybrid firefly algorithm performed better than the normal firefly algorithm in terms of time latency which is very low. The study was carried out of these algorithms with respect to F - Measure, Accuracy, Time, PR and ROC. The prediction model in the telecommunication companies should be as efficient as accurate [3]. The efficiency of the model can be gained by making the best training model with reduced dimension and size. The extraction and feature selection techniques used, helps to get the efficient features along with predicting accurately. Thus, this paper shows that being smaller in size i.e., 232 KB this prediction model performs the similar tasks even with more accurate results i.e., 92 % than the initial model of 303 KB size. In many researches, was represented in different ways and data mining is most usable in churn prediction in telecommunication industry [4]. The concept of multilayer perceptron neural network was used to proposed churn classification model and showed that the logistic boost and logical regression both are useful to build a churn prediction model. Black box models are complicated, but actual work tell only logistic regression and using logit boost training example for weak prediction that play an important role in churn prediction. Effective classifiers like KNN, Decision tree classifier, random forest are studied [5]. Classifiers consist of four stages: data access, data wrangling, training, and acquiring insights that all aspects play an important role to analyze churn prediction data. Here, data access is gives input to the proposed model. Data wrangling is the process of collecting distributed data [6]. The training process is constructed by model and test the data set for processing and acquiring insights is useful to predict final output and analyze the final result of churn prediction.

Machine learning techniques used [7,8] in developing this model for this work. The telecom company can predict churn customers with the help of electronic learning technology with developed model. Industries can provide best services so as to reduce the churn level. Such types of models help telecom services for making them profitable. Random Forest and Decision Tree are used for this model. In [9] the PMM (Predictive Mean Matching) algorithm helps to manage lost values, rather than feature removal or recognition of missing data. The combination of two Ensemble classifiers is embedded within the customer speculation model to manage large databases, time-dependent feature labels, and the distribution of inequitable data in the Telecommunication industry. To evaluate the effectiveness of customer churn uplift models, the author presented a novel profit-driven assessment method dubbed the maximum profit uplift measure in [10]. The proposed MPU measure extends the maximum profit measure for customer churn prediction models, allowing for evaluation of a customer churn uplift model's performance in terms of profit per customer in the customer base earned when a retention campaign targets the optimal proportion of customers with the highest uplift scores. Maximizing the profit generated by the retention campaign determines the best proportion of customers to target, which is proven in this article to be strongly tied to the uplift model's capacity to identify the so-called persuadable i.e., Customers who are about to churn but will be kept if the campaign is targeted. Saran Kumar et al. [11] gives a thorough examination of the strategies used to anticipate client churn. Each of their proposed churn prediction models has a low forecast accuracy. To avoid the problem of client turnover, a solid prediction model is essential. Kraljević and Gotovac [12] given a logical foundation for developing data mining applications by defining an enhanced technique for designing applications based on Data Mining technologies. The methodology proposed has been deployed and proven in the telecom industry using data mining applications to predict prepaid subscriber attrition.

Real-life examples of customers are used who decided to go and learn the qualities and behaviors that precede customer profits [13]. There are results that show a clear height for advanced versions of models compared to plain (non-expanded) versions. The best separator was SVM-POLY using AdaBoost with approximately 97 % accuracy and F-rating over 84 %. Alzubaidi, et al. [14], used CDR attributes to determine the strength of social ties between users for each identified community, and then used a model to propagate churn influences on the call graph to determine the net aggregated influences from the churner node. This effect was employed in the logistic regression method to calculate the churn tendency for individual users. When building the Telecom social network, examine the relationship between users to improve the performance of the churn prediction model. CDR attributes were employed to characterize the relationship strength of social communication between edges in the graph. CDR attributes contain information about phone calls and SMS messages. In this work the quality of prediction measurements for various prediction models is compared. Khalid, et al. [15] looked at different prediction models and compared the quality of prediction models such classification algorithms and decision trees. They discovered that the decision tree's accuracy is higher than the other techniques (3 % higher than the second result and 6 % higher than the lowest-achieving algorithm), showing that the decision tree is an excellent churn prediction technique.

The primary goal of study presented in [16] is to develop a method to predict high-value customer turnover based on existing research and customer attribute characteristics in the telecom industry. This study accomplishes customer churn prediction based on the telecom business based on the analysis of big data in the telecom industry and historical information estimation of customers, combined with logistic regression method. It identifies possible churned customers in the customer library by studying the features of customer churn behavior in the telecom industry, and it assists organizations in taking targeted win back efforts based on the characteristics of the potential churned A review of customer churn prediction in the telecommunications industry is presented in [17]. The study demonstrates a huge number of attributes that are put into practice by a big number of paper reviewers to construct a customer churn prediction model. Lalwani et al. [18] presented a comparative study of customer churn prediction in the telecommunication industry using well-known machine learning techniques like Logistic Regression, Nave Bayes, Support Vector Machines, Decision Trees, Random Forest, XGBoost Classifier, CatBoost Classifier, AdaBoost Classifier, and Extra Tree Classifier in this research paper. The experimental results demonstrates that two ensemble learning algorithms, Adaboost classifier and XGBoost classifier, have the highest accuracy when compared to other models, with an AUC score of 84 percent for the churn prediction problem. They outscored other algorithms across the board in terms of accuracy, precision, F-measure, recall, and AUC score. In [19] an improved churn prediction for credit card system using supervised learning and rough clustering technique is presented. Whereas, Amin et al. [20] implemented churn prediction using Naïve Bayes for customers of telecommunication industry with focus on the challenges to get new customers in telecommunication industry instead of handling old customers. Model to analyze customer behavior was developed along with improvement in accuracy of prediction. In line with this work again Amin et al. [21], highlighted telecommunication sector's churn prediction using Just-in-Time approach. It handled various problems of customers intention that may switch from one service provider to another. In this context, different churn prediction methods are applied by practitioners to resolve issues related to preserving customer retention.

In other work, Amin et al. [22] presented churn prediction for customers of cross-company by using transformation of data techniques. It emphasizes target company with lacking data which may be used to source company to predict customer churn effectively. Even, it is also including impact of these techniques on performance using various classifiers needed in telecommunication sector. Amin et al. [23] used rough set method to predict customer churn of telecommunication sector. Customer behavior was used to differentiate the churn from non-churn customers. The major objective was to desperate the necessity of businesses in retention of existing customers. Moreover, this work uses intelligent rule-based techniques for decision making along with rough set to detect customer churn from non-churn customers. Ahmad et al. [25] focused on multi dataset approach for churn prediction which is applicable for telecommunication industry. In this work, major contribution related to feature engineering phase for creation of custom features with the help of machine learning is presented. Moreover, the big data approach along with comparative results of multiple algorithms of machine learning has been used.

In the telecom sector, a number of customer retention tactics are being used to lower churn and boost loyalty. Customer loyalty in the telecom industry is far from guaranteed. Many companies combined its cellular and wire line customer loyalty programmers as part of a reorganization that aimed to give customers more rewards options. Customers can now take advantage of a variety of benefits for their loyalty, including: hardware enhancements include fiber optic service and internet content. To lower their churn rates, telecoms are investing more and more in cutting-edge technologies. Tools for customer analytics and insight can be used to forecast customer behavior, allowing service providers to decide which strategies will increase retention rates the most effectively. Additionally, cutting-edge technology uses artificial intelligence and machine learning techniques to pinpoint the clients most likely to churn. From this discussion it is observed telecom industry playing a very major role in our daily life and generating an enormous amount of customer base. New customer base generation is costlier than retaining the existing customers. Churning is the process of switching of customers in a given time. Telecom management and analysts find the explanations behind customers leaving subscriptions and behavior activities of the holding churn customers' data. The model proposed uses classification techniques to find the leave subscriptions and collects the reasons behind it in the telecom industry. The aim of this model is to analyze the various machine learning algorithms required to develop customer churn prediction models and identify churn reasons in order to give them with retention strategies and plans. The main motivation behind churn prediction in the telecom sector is to reduce churns and retain the existing customers. Section 2 emphasizes system architecture of proposed system and proposed system model is presented in detail. Sequentially, Section 3 focuses on detailed dataset description based on single dataset including data pre-processing, and feature selection and also highlighted multi dataset approach concisely. In Section 4, experimental analysis using various ways such as decision tree classifier, Random Forest algorithm, survival analysis, Cox proportional hazard model, and retention strategy is presented. Finally, concluding remarks are summarized in Section 5.

## 2. System architecture

In this section, system architecture and proposed system model are discussed subsequently.

### 2.1. System architecture of the proposed system

The implementation for churn will require the latest version of Anaconda with built in features that consist of Jupiter notebook for training and testing data. The latest version of Anaconda with built-in functionality, like Jupiter notebook for training and testing data, will be required for the churn implementation. Churn predictions for the telecom industry have been carried out using literature with various methods that includes machine learning algorithms, data mining techniques and retention strategies. These techniques effectively support many companies for predicting, identifying, and retaining churners which help in CRM (Customer relationship management) and decision making. CRM deals with the data to identify a loyal customer for industry. High revenue generating customers (loyal customers) for a company have no impact on the competitor companies. Such loyal customers help to grow profitability of a company by referring to the other people such as their family members, colleagues, and friends. Hence, the role played by CRM is very important in churn prediction and it also helps to retain the churning customers. Fig. 1 depicts a cycle through which churn prediction can be made. For prediction there are many algorithms such as Support vector machine, K-nearest neighbor, J48, naive Bayes, logistic regression, LWL, Random Forest, Decision tree classifier which are used to resolve classification problems. Random forest and Decision tree classifier are considered relevant with better accuracy and performance.

### 2.2. Proposed system model

Let S is the proposed system.

$$S = \{I,\ O,\ DP,\ FS,\ EF\}$$

Where

1) I (Input): Dataset
2) (Output): CM (Confusion matrix) ={R->Actual False, Actual True} {C->Predicted False, Predicted true}
3) DP (Data Processing)
4) FS (Feature Selection) is measured by Pearson correlation formula (Cr) where set of numerical attributes is taken ‘X’ ranging from $X_1, X_2, X_3$……., Xn consider a set of attributes X, form a subset of two attributes each, $X = \{\{X_1, X_1,\}, \{X_1, X_2\}, \{X_1, X_3\} .... \{Xi, X_j\}… \{X_n, X_n\}\}$

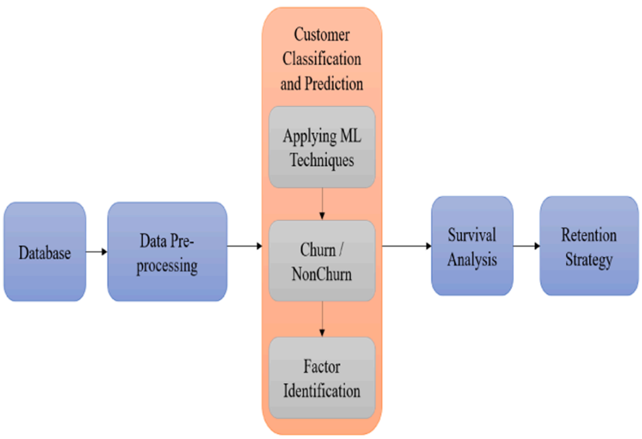


**Fig. 1.** Proposed system.

If each subset = X then evaluates $\sum X_i$, $\sum X_j^2$ & $\sum (X_i * X_j)$
Cr= $\sqrt{\sum} (X_i * X_j) \div \sum X_i^2 * \sum X_j^2$
Where Cr = 1 (+ve correlation), Cr = 0 (no correlation), Cr = −1 (-ve correlation)
Convert string values into numerical (S→N)
Then apply ML algorithm on new Dataset (consider as C)

5) EF – Efficiency of Proposed Model

The proposed system for churn prediction is derived using accuracy, precision, recall, f-measure. Accuracy calculates the accuracy metric.

TP=True positive, TN=True negative, FP=False positive, FN=False negative values, AP=Actual positive, AN=Actual negative.

a) Accuracy= (TP + TN) ÷ (TP + TN + FP + FN)
b) True positive and true negative are values of the confusion matrix after applying classification algorithms. True positive rate is the value that shows us which part of data is classified as correct and false positive classifies incorrect values.
c) (TP) rate = TN ÷ TN +FP
d) (FP) rate= FP ÷ FP+TN
e) Precision = TP ÷ (TP + FP)
f) Recall = TP ÷ (TP + FN)
g) F-Measure = (2 × Precision × Recall) / (Precision + Recall)

*Constraints*: The time is the key for predicted churners. If C (customer) churns after one month, but sometimes dataset shows that it would churn after one week that will lead to incorrect churn. Hence, C depends upon T. Incorrect dataset (ID) with irrelevant features will be other constraints to predict churn model.

*Failure of S*: If customer (C) gets churned out for one month time (T) but then customer rejoins again would lead to loss of company and system (S).

## 3. Dataset description

Here, specifically focus is on Single Dataset presented in sub Section 3.1. Sequentially, Section 3.2 emphasizes multiple datasets to validate proposed system.

### 3.1. Single dataset

The dataset used for experiments in this paper, contains results of Telco-Customer-Churn dataset obtained from Kaggle website (it is also known as IBM Watson dataset which was released in 2015).Each row represents a customer, each column contains attribute described on the column Metadata. It consists of 7043 customer information. Every customer has 21 features and the “Churn” it contains 11 missing values in the Total Charges column. The last attribute contains labelled data with two classes where 26.53 % of total customers are labelled as “T’’ indicating true customers i.e., categorized as churning customers and the remaining 73.46 % customers are labelled as “F’’ indicating false customers i.e., categorized as non-churning customers. The attribute selection depends on the results of techniques of feature selection that find useful, the most similar and effective attributes to predict the churning customers. A total of 5174 are non-churners and 1869 are churners. The dataset contains 16 categorical columns and 5 numeric columns. The dataset helps to figure out customer prophecy and build retention possibilities. Figs. 2 and 3 shows details of database.

#### 3.1.1. Data pre-processing

This step is required to remove all the irrelevant and dirty data of real world. As the data is congregated from many resources it is important to overcome this issue. Without execution of this step decision makers cannot predict outcomes even if they did it won’t be

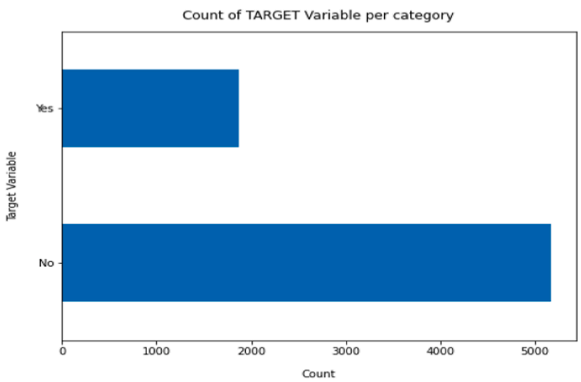


**Fig. 2.** Count of target Variable per category.

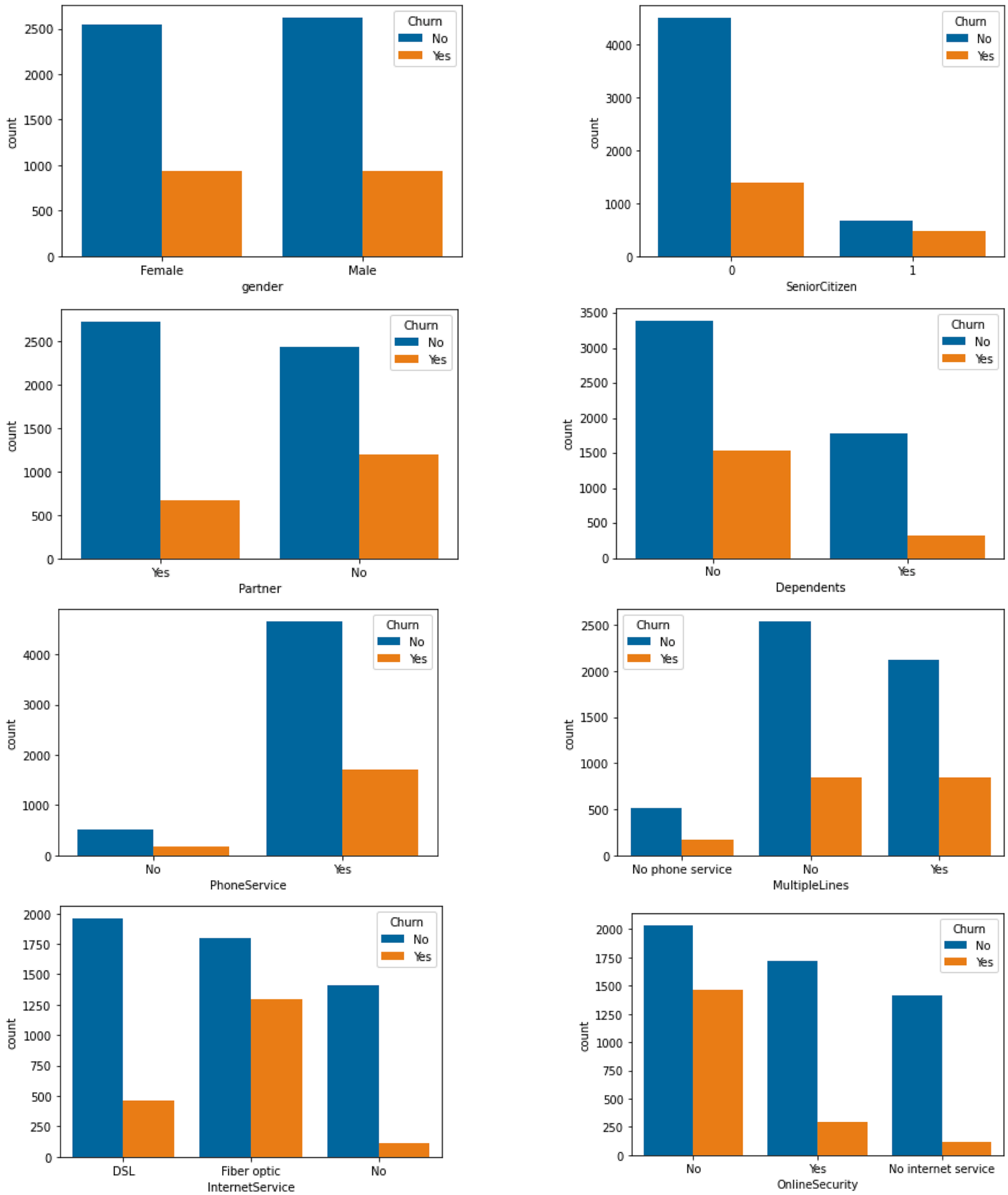


**Fig. 3.** Database attribute details.

correct that leads to no quality data and no quality mining. To solve these issues following methods are considered for cleaning the data. Likewise, there are many such features taken as input to predict model. Classification of data and performing pre-processing cleans the data and makes it easy to use.

Data classification and pre-processing clean the data and make it easier to use. There are three steps in data preparation.

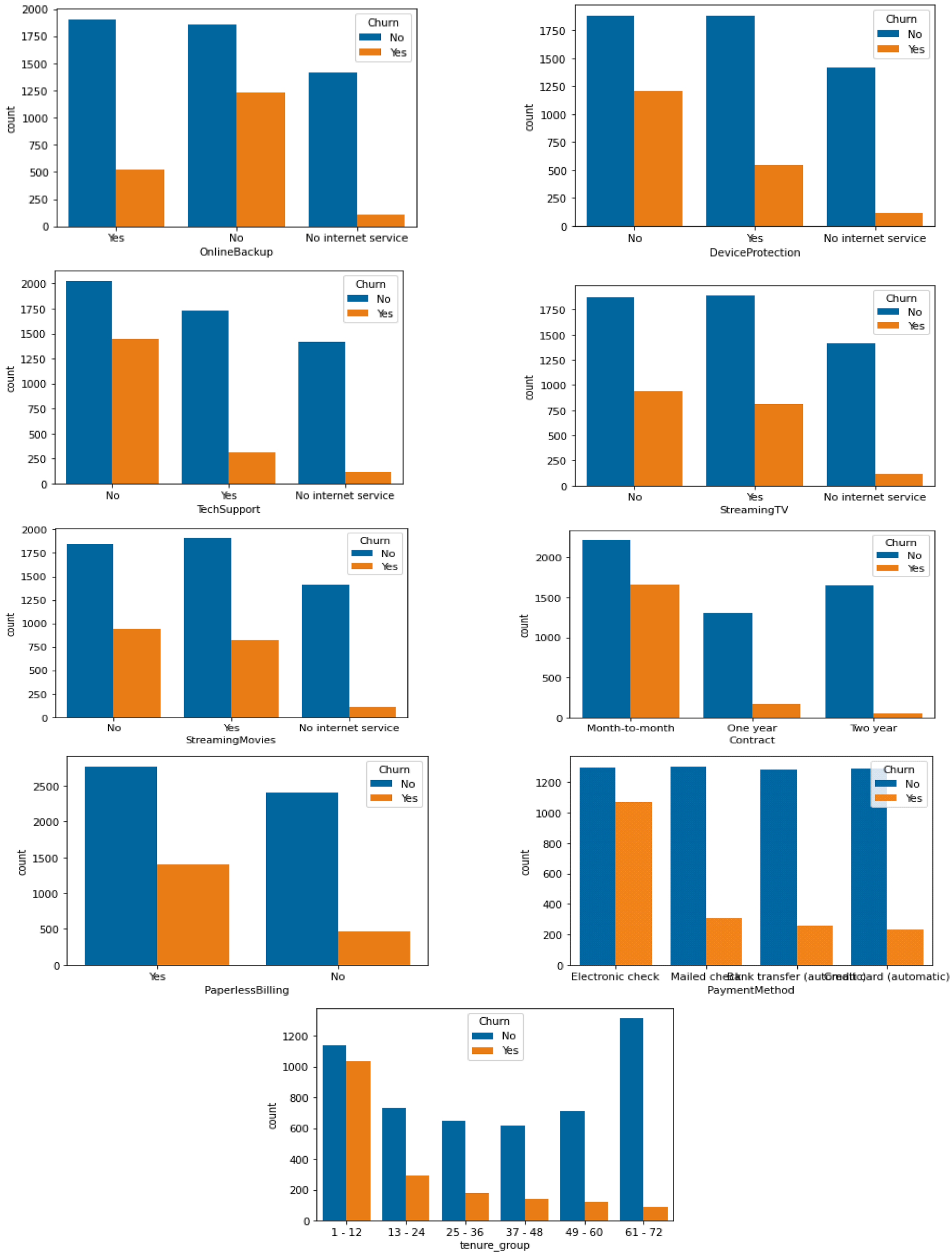


**Fig. 3.** (*continued*).

- A data comparison is carried out in the first stage to identify redundant data. The properties that are repeated are immediately removed.
- Each instance searches for the missing value in step two. If missing values are detected, replacement procedures are used; otherwise, missing values are replaced with the most suited value. When there is no way to replace the instances, they are discarded.
- The final stage entails defining the type of each value in order to eliminate extraneous data. If the information is no longer useful, it is discarded. In other words, noisy data is deleted from the data that has already been pre-processed. Check the data types of each column first.

Because the total charges column's data type is object type, we convert it to numeric and make a duplicate of the underlying data to manipulate and process. This conversion is performed using panda's library imported in the program and using to-numeric function of library. Next, check for duplicate values in dataset, but there are no duplicate values present in dataset. Further continue checking for missing values. So according to general thumb rule the feature with less missing values there to fill means values or simply use regression to predict the missing values based on the particular feature. Similarly, in case of feature with high number of missing values, it would better to drop those columns due to less analysis and insights. Also, the columns with more than 30–40 % are deleted. Now by using is null and sum function there are 11 missing values present in the Total Charges column so drop this missing value. The tenure consists of maximum value of 72 and also dividing the tenure into 6 categories of 1–12, 13–24, 25–36, 37–48, 49–60, 61–72 make easy in visualization of column data of tenure period. Dropping the columns that are not required for processing. It was found the customer ID is not useful and contains unique value which won't affect the prediction results, so drop customer ID and similarly drop the tenure column which is not necessary to evaluate the results.

| | |
|---|---|
| 1–12 | 2175 |
| 61–72 | 1407 |
| 13–24 | 1024 |
| 25–36 | 832 |
| 49–60 | 832 |
| 37–48 | 762 |

Name: tenure_group, dtype: int64

*Univariate analysis* - In data exploration, the distribution of individual predictions by churn will be determined, and the number of churners representing non-seniors will be determined. Here, 1 represents senior citizens, and the plot demonstrates that if a client is a senior citizen, they are more likely to churn. Similarly, when a candidate is single and does not have a partner, the partner churn ratio is high. People with phone service are more likely to churn, similarly payment method, if there is an electronic check appears higher than the case of credit card appears lowest churner because of this they could be having auto debit features, which is one of the important features for churn, similarly remaining people with phone service are more likely to churn, similarly remaining people with phone service are more likely to churn, similarly remaining people with phone service are more likely to churn, remaining figures also shows the churn count respectively above selected features.

*Bivariate analysis* - It is used to find a value prediction for a single variable. Correlations between variables are simple to find. A relationship between two variables is defined as bivariate. There are numerous features in our dataset, and we presented the results. Two variables were examined, and two new data frames for churners and non-churners were generated. A function is created that maintains a data frame that is passed with column, title, and hue information for each feature, similar to how bar graphs for different features may be shown in diagrams. Gender characteristics considers 2500 female and male participants, with a churner/non-churner ratio of around 50 % for each gender. According to gender feature analysis, females are more likely to churn if they have a relationship, but males are more likely to churn if they do not have a partner. The groups are classified into churners and non-churners.

### 3.1.2. Feature selection

This is an important step in achieving our model's goal. Unnecessary data is discovered in datasets while training the model, resulting in a reduction in model accuracy. As a result, feature selection on a dataset is used to solve these issues.

The following are the advantages of feature selection:

- Reduced over fitting means less chance of making conclusions based on noise.
- Accuracy is improved because there is fewer misleading data.
- Training time is reduced providing lesser algorithm complexity with algorithms that train faster.
- Applying Machine Learning Algorithms
  [1] *Decision tree*: Decision tree are used to solve both classification and regression problem in the form of trees that can be incrementally updated by splitting the dataset into smaller dataset, where the result are represented by the leaf node. Each branch represents the possible decision outcome or reaction. It is like a flowchart diagram that shows the various outcomes from a series of decisions. It can be used as a decision-making tool, Decision tree has some series of same craft questions regarding attributes of test data record and it's use to solve classification-based problems, every time it gets solution from it and follows until the final conclusion of class label record. Then visit several decision trees for achieving target value. It can be either true or false. Now pick a majority vote of trees or count the target values provided, then based upon decision trees predict

if customer churn is true or false for research analysis, or for planning strategy. A primary advantage for using a decision tree is that it is easy to follow and understand. Decision tree classifier is simple and adaptive classification technique, this is basically implying a straightforward process to analyze and solve the problem.

[2] *Random Forest tree*: The random forest is a classification algorithm consisting of many decision trees. More numbers of trees in the forest led to more robustness in prediction with higher accuracy. It uses bagging and feature randomness, when building each individual tree to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree. In this algorithm, each tree will have its own output from a dataset provided, such that output generated will be considered from the majority of trees. Decision trees made in this algorithm are of numeric type in which the tree picks any random attribute in the dataset. Advantages of Random Forest that it helps to solve both regression and classification problems.

[3] *Important Features or factors responsible for churning according to Decision Tree model*: Finding the important factors that are responsible for churning makes it possible to find the service that is required to customer to prevent from attrition. This can be done using feature importance. Here the feature is ranked according to their importance. The most important feature at the top of list, while least important are at end of list.

In Fig. 4 it is seen that Contract-Month-To-month ranked first in the list with its importance as 0.517, Total Charges with importance as 0.104, No Internet Service is 0.093, DSL Internet Service as 0.0795, Monthly Charges as 0.0517, Contract of Two years as 0.0464. Contract of one year as 0.0419, tenure group of 1–12 as 0.0397, No Streaming Movies as 0.00696, Fiber Optic Internet Service as0.00643, and last in list is with No Paperless Billing as 0.004727.

[4] *Important Features or factors responsible for churning according to Random Forest Tree model*: In this model from Fig. 5 the features importance ranked from 1st to last as,

- Contract month to month = 0.1245
- Tenure group = 0.10518
- Internet Service Fiber Optic = 0.07693
- Total Charges = 0.07000
- Contract Two year = 0.06123
- Tenure group 61 -72 = 0.04302
- Online Security yes = 0.03977
- No tech support = 0.03943
- Online Backup No Internet Service = 0.0360
- Streaming Movies No Internet Service = 0.03479
- Internet Service DSL = 0.0338
- Monthly Charges = 0.0309
- Online Security No Internet Service = 0.0294
- Contract One year = 0.02819
- No Online Security = 0.02620
- Tech Support Yes = 0.02491
- Tech Support No Internet Service = 0.024313
- Partner Yes = 0.02422
- No Multiple Lines = 0.001289

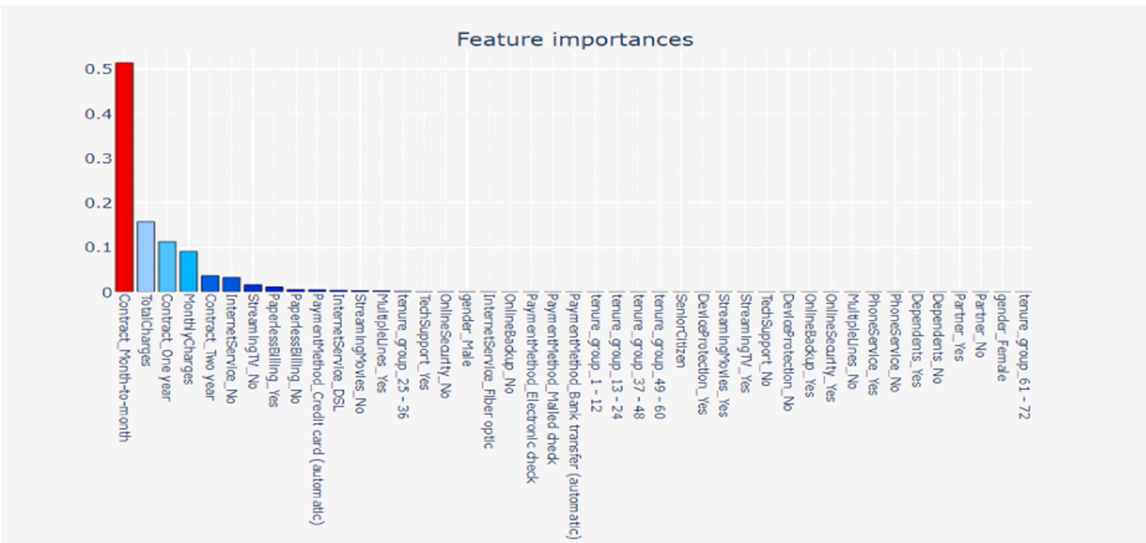


**Fig. 4.** Feature importance for decision tree.

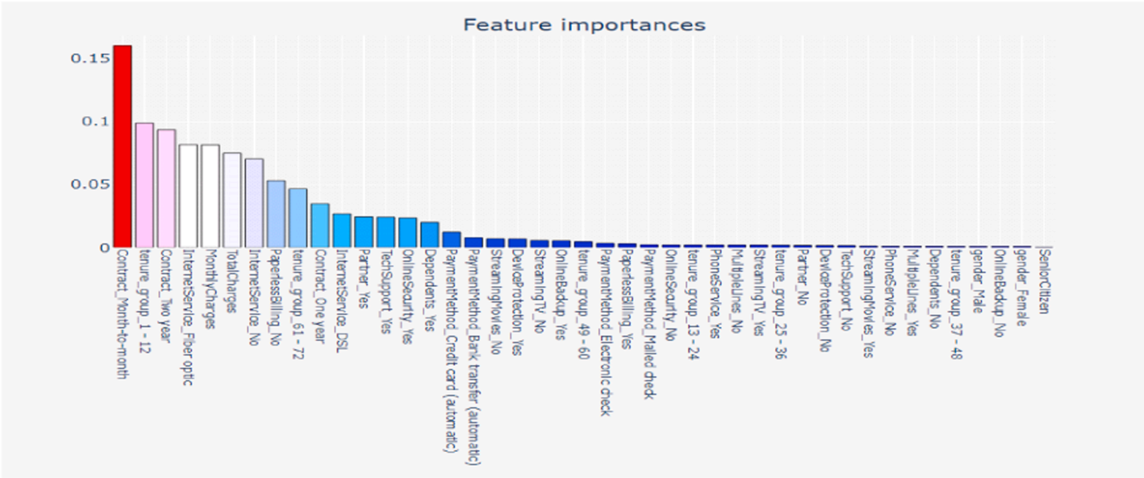


**Fig. 5.** Feature importance for Random Forest classifier.

- Gender Male = 0.0011822
- Multiple Lines No phone service = 0.0011689
- No phone Service = 0.001107
- Gender Female = 0.001101

### *3.2. Multiple dataset*

There exist multiple types of data in SyriaTel [25] as classified below which may be applied to construct our churn prediction model used in telecom sector. It is observed to present dataset structure using spark engine, it is required to include phase-based exploration with suitable pre-preparation for algorithms based on classification.

#### *3.2.1. Towers and complaints database*

In this database, the detailed data of actual location is shown as in the form of digits. The serial digits are mapped with database of towers which offers the actual location of the transaction and providing state, city, area, sub-area, latitude and longitude. Database including various complaints provides all submitted complaints along with inquiries statistics based on coverage, any problem related to the telecom business and issues related to packages and offers.

#### *3.2.2. Customer data*

Customer data comprises the data-based contact information and services of customers. Moreover, various packages, services and offers taken by customer. Also, it contains CRM system including information generated from all customer GSMs such as gender, birth date, the location and type of subscription etc.

#### *3.2.3. Network logs data*

Network logs data includes sessions related to calls, SMS and internet for each transaction used by telecom operator such as required time to initialize a session for call ending and to check internet status. Moreover, it represents whether session is expired or not due to error occurred in the internal network.

#### *3.2.4. Mobile IMEI information*

Mobile IMEI information comprises the model, type, model of the mobile phone and whether mono or dual SIM device. Data may have large size which may require information in detailed. It requires a lot of time for understanding. It also needs to know the original sources along with format for storage. Moreover, related to these records, data must be linked to each other logically using relational databases that actually represent customers detailed information.

#### *3.2.5. Call details records*

Call Details Records (CDRs) includes modifiable data related to MMS, calls, SMS etc. Also, transaction made by customers using internet which is ultimately generated in the form of text files.

## 4. Experimental analysis

After preprocessing on dataset following are the observations:

- Strong correlation exists between tenure and total charges, means as tenure increases so does total charges.
- Strong correlation exists between monthly charges and total charges as well.
- Tenure and Contract duration seems to be strong factors in determining churn.
- Among service types, phone service seems to be most popular.
- CSP should investigate if customers receiving digital invoice have any concern with understanding the bill details.
- Also, they should encourage customers to move to automated payment modes to improve customer experience.
- Gender does not play an important role. However, CSPs should take care of the experience of senior citizens.

From Figs. 6 and 7 it is seen that churn is higher when monthly rates are high. Even with modest overall charges, there is large churns, as shown in the Fig. 7. When all three parameters (Total Charges, Monthly Cost, and Tenure) are combined, higher monthly charges at low tenure result in lower total charges, implying that all of these characteristics are associated to higher churn. Fig. 8 shows used all features.

### 4.1. Experiment analysis using decision tree classifier

So, in our proposed model using a Decision Tree classifier, the obtained accuracy of the model is 78 %, which is very low, and printing the classification report led to the dataset unbalance, which results in less accuracy. Decision Tree classifier took Number of Leaves - 331, Size of the tree - 552, Time taken to build model-1.06 s for execution.

As a result, the accuracy of the Decision tree classifier model before up-sampling & ENN should not be used as a meaningful measure because it leads to unbalanced datasets. As a result, when checking recall, precision, and F1 scores for the minority class, it's clear that the precision, recall, and F1 ratings for Class 1, i.e. churned consumers, are far too low.

For up-sampling training data into a decision tree classifier that differentiates into $x$ train and $y$ train and creates a prediction variable and calls the classification input to process input to produce output with accuracy, and using SMOTE (synthetic minority oversampling technique) by performing oversampling and cleaning using ENN (Edited Nearest Neighbors), the dataset is balanced with dataset values of 493, 40, and 599, 34, and it provides a solution. Tables 1 and 2 shows details results of Decision tree classifier model before and after up-sampling & ENN. The precision of the classifier matrix is 93 %, the recall factor is 93 %, and the F1 score is 93 % and provides 93.85 % accuracy.

### 4.2. Experimental analysis using random forest algorithms

Missing values are handled with carefully, and accuracy is maintained. It's even capable of handling big, multi-dimensional datasets. So, the accuracy of our model using the Random Forest Tree classifier is 98.91 %. Table 3 shows results of Random Forest classifier model before Up-sampling and ENN. There are several techniques used in random forests. Generally bagging technique is used known as ensemble classifier.

For the minority class of churned consumers, employ a classification matrix to increase the model's performance. The imbalance database and its merely oversampled minority class can be addressed by using SMOTE (synthetic minority oversampling technique) and ENN (edited nearest neighbors). In the required minority class, it would include duplicate examples.

The results of the Random Forest classifier model after up-sampling and ENN are shown in Table 4. As a result, the random forest classifier predicts churn with an overall accuracy of 99 %. The classifier matrix has a precision of 99 %, a recall factor of 99 %, and an accuracy of 99.09025616471152 %.

### 4.3. Survival analysis

The survival analysis technique is a valuable statistical technique for predicting how long a client would keep a subscription when

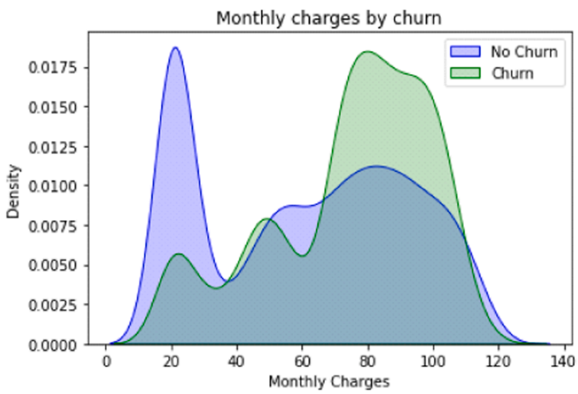


**Fig. 6.** Monthly charges by Churn.

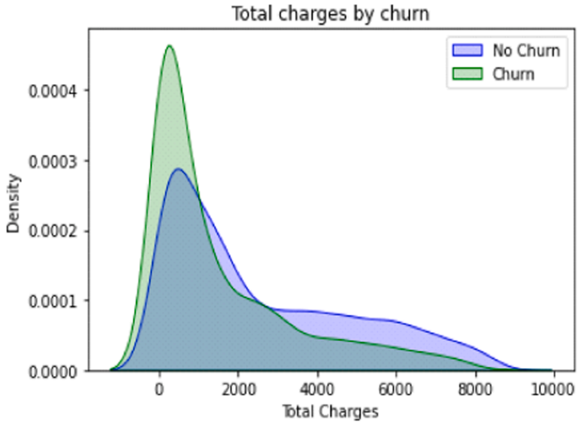


**Fig. 7.** Total charges by Churn.

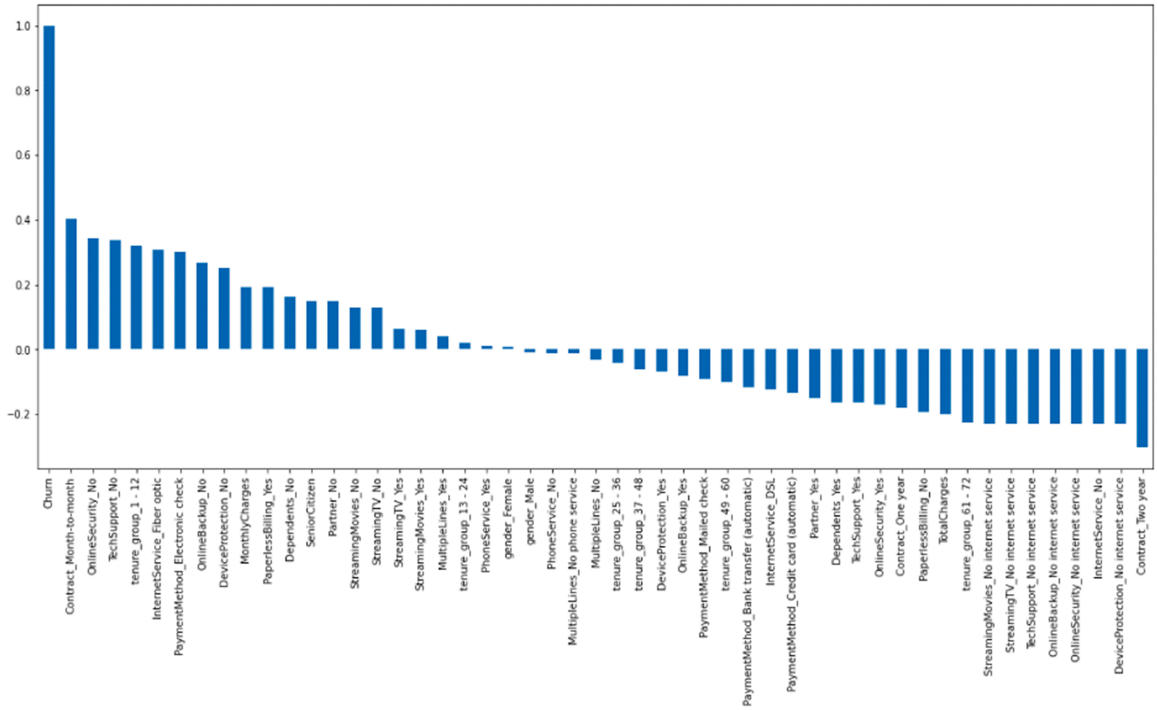


**Fig. 8.** All features in bar graph.

**Table 1**
Results of decision tree classifier model before up-sampling & ENN.

| | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| Class 0 (No) | 0.83 | 0.89 | 0.86 | 1046 |
| Class 1(Yes) | 0.61 | 0.49 | 0.54 | 361 |
| Weighted avg | 0.77 | 0.78 | 0.77 | 1407 |

| | No. of Instances | Percentage |
|---|---|---|
| Correctly Classified Instances | 5493 | 77.9923% |
| Incorrectly Classified Instances | 1550 | 22.0077% |

**Table 2**
Results of decision tree classifier model after up- sampling & ENN.

| | Precision | Recall | F1-score |
|---|---|---|---|
| Class 0 (No) | 0.93 | 0.92 | 0.93 |
| Class 1(Yes) | 0.94 | 0.94 | 0.94 |
| Weighted avg | 0.93 | 0.93 | 0.93 |
| | No. of Instances | Percentage | |
| Correctly Classified Instances | 6615 | 93.85% | |
| Incorrectly Classified Instances | 427 | 6.15 % | |

**Table 3**
Results of Random Forest tree classifier model before Up-sampling & ENN.

| | Precision | Recall | F1-score |
|---|---|---|---|
| Class 0 (No) | 0.99 | 1.00 | 0.99 |
| Class 1(Yes) | 0.99 | 0.96 | 0.97 |
| Weighted avg | 0.99 | 0.99 | 0.99 |
| | No. of Instances | Percentage | |
| Correctly Classified Instances | 6971 | 98.91% | |
| Incorrectly Classified Instances | 77 | 1.09% | |

**Table 4**
Results of Random Forest classifier model after Up-sampling and ENN.

| | Precision | Recall | F1-score |
|---|---|---|---|
| Class 0 (No) | 0.98 | 1.00 | 0.99 |
| Class 1(Yes) | 1.00 | 0.98 | 0.99 |
| Weighted avg | 0.99 | 0.99 | 0.99 |
| | No. of Instances | Percentage | |
| Correctly Classified Instances | 7010 | 99.09% | |
| Incorrectly Classified Instances | 38 | 0.91 % | |

they churn. "Time to event analysis" is another name for survival analysis. Customer retention is heavily influenced by survival analysis. To avoid churn, we concentrate on a large number of consumers with a short survival span. This analysis determines the value of a customer's life time. The event is defined as the precise time when a customer cancels or leaves a subscription, and the time is specified as the time when the consumer joins the service.

Survival function: -

$$\mathrm{S(t)} = P\ \mathrm{r}\ (T > t) = 1 - \mathrm{F(t)} = dx$$

Here $T$ = event time, f(t)= density function

### 4.4. Cox proportional hazard model

'Time-to-event' data is analyzed using the Kaplan-Meier (KM) approach. All-cause mortality is a common outcome in KM analyses. However other outcomes such as the occurrence of a cardiovascular event could also be included. The Cox Proportional Hazard model is useful to predict better survival probability of individuals. In this model, some characteristics include partner, monthly costs, phone service, gender, and remaining variables are covariates, which impute on the survival probability, taking into account each customer's tenure at the time they churned. All variables and survival functions are likewise included in this model.

The log-hazard of an individual is a linear function of their variables and a population-level baseline hazard that changes with time, according to Cox's proportional hazard model. Mathematically:

There are a few things to notice about this model: the baseline hazard, b0, has the only temporal component (t). The partial hazard is a time-invariant scalar element in the preceding equation that solely raises or decreases the baseline hazard. As a result, changes in variables will only affect the baseline hazard.

Fig. 9 show the coefficient in another way. For instance, the coefficient for PhoneService Yes (having a phone service) is around 0.69. In the Cox proportional hazard model, a one-unit increase in PhoneService Yes increases the baseline hazard by a factor of exp (0.69) = 2.00, or nearly 20 %. A greater hazard indicates that the event is more likely to occur. The hazard ratio is defined as exp (0.69) divided by 1.

The key thing to notice here is that although though the (coef) values for covariates MonthlyCharges and gender Male are close to

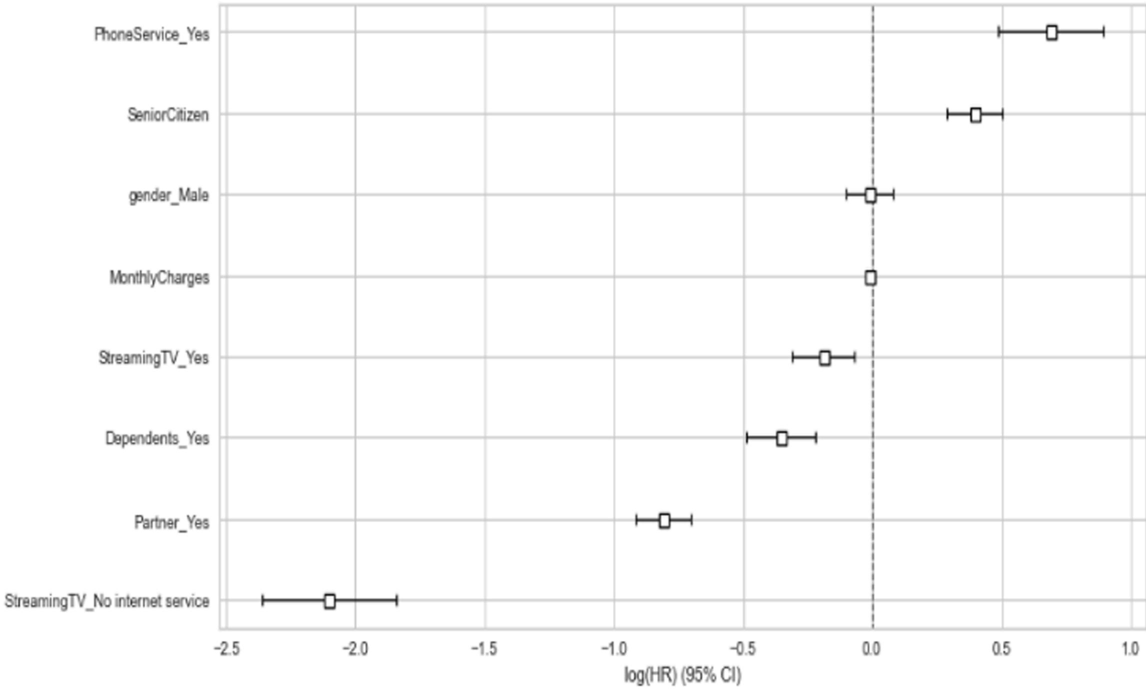


**Fig. 9.** Significance of covariate in predicting churn risk.

zero (−0.01), the former still has a substantial impact in forecasting churn, whereas the latter is inconsequential. The reason for this is because MonthlyCharges is a set of continuous values that can change from one month to the next.

Fig. 10 shows survival curve for the selected customers. So, based on the Fig. 10, it is concluded that customer 2 has the highest chance of churning. Creating survival curves at the customer level allows us to develop a proactive plan for high-value clients for various survival risk segments along the timeline.

### 4.5. Retention strategy

Some high-effort interactions to support customer retention in the telecom industry include Customers Must Be Educated, attractive

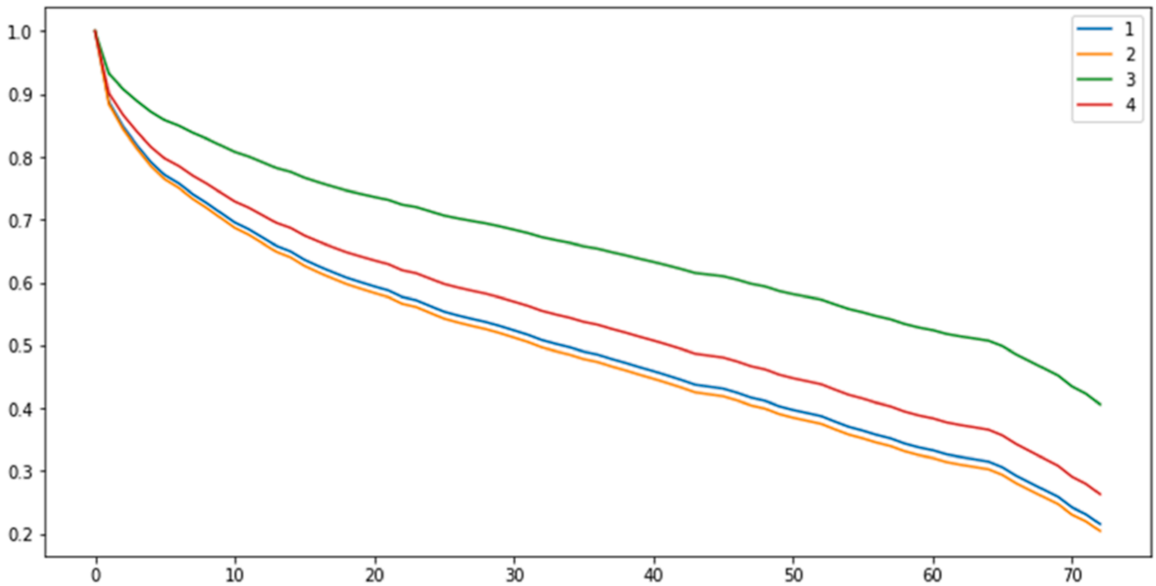


**Fig. 10.** Survival curve for the selected customers (Customer 1, 2, 3 and 4).

Offering Plans, repeat contacts, an emphasis on wasting customers' time, shoddy self-service, avoiding unneeded robotic service and giving complicated instructions.

## 5. Conclusion

Our system works effectively for achieving the major objective to analyze the various machine learning algorithms required to develop customer churn prediction models and in identification of churn reasons in order to give them with retention strategies and plans. Our system focuses on resolving an unavoidable issue Customer churn arise in the telecommunications sector for a variety of reasons. The most perplexing aspect about client turnover is that churn is impossible to manage. Moreover, customer turnover has several causes, some of which are visible and others which are not. On the other hand, operators in the telecommunications industry should be aware that client loss will occur sooner or later and they must be prepared to respond. In this context, our system plays important role on solving the problem of customer churn has become critical for telecom companies' survival.

A customer churn system is based on machine learning methods, a decision tree classifier, and a random forest algorithm. Here, prior to and after up sampling and ENN, both techniques are applied to the system. Decision tree classifier models initially produced poor results on an unbalanced dataset that did not take ultimate accuracy into account when matrices were used to evaluate the model. In comparison to a decision tree classifier, a random forest classifier produces better results. With an overall accuracy of 99 %, the random forest classifier predicts churn. The classifier matrix has a precision of 99 %, a recall factor of 99 %, and an accuracy of 99.09 %. System comprises churn prevention that is based on survival analysis using Cox Proportional Hazard model and retention plans. Survival curve is used for the selected customers plays an important role in customer churn prediction. In the telecom sector, it appears obvious that lowering customer effort is a method to boost customer retention. This churn prevention system may be made more complex and sophisticated in the future to provide better and more precise recommendations. More advanced algorithms such as deep learning recurrent neural networks that help to identify nonlinear complex relationships between data variables which may be applicable in future study to estimate survival likelihood. Moreover, customer churn prediction in the telecommunication sector is possible a rough set approach and data certainty [24] in future.

## Financial and ethical disclosures

This work is not supported fully or partially by any funding organization or agency.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

No data was used for the research described in the article.

## References

[1] Ullah I, Raza B, Malik AK, Imran M, Islam SU, Kim SW. A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector. IEEE Access 2019;7:60134–49. https://doi.org/10.1109/ACCESS.2019.2914999.
[2] Ahmed AAQ, Maheswari D. Churn prediction on huge telecom data using hybrid firefly based classification. Egypt Inform J 2017;18(3):215–20. https://doi.org/10.1016/j.eij.2017.02.002.
[3] E V, Ravikumar P, S C, M SK. An efficient technique for feature selection to predict customer churn in telecom industry. In: Proceedings of the 1st International Conference on Advances in Information Technology (ICAIT); 2019. p. 174–9. https://doi.org/10.1109/ICAIT47043.2019.8987317.
[4] Jain H, Khunteta A, Srivastava S. Churn prediction in telecommunication using logistic regression and logit boost. Procedia Comput Sci 2020;167:101–12. https://doi.org/10.1016/j.procs.2020.03.187.
[5] Pamina J, Beschi R, SathyaBama S, Soundarya S, Sruthi MS, Kiruthika S, Aiswaryadevi VJ, Priyanka G. An effective classifier for predicting churn in telecommunication. J Adv Res Dyn Control Syst 2019;11(01–Special Issue). https://ssrn.com/abstract=3399937.
[6] Berger P, Kompan M. User modeling for churn prediction in e-commerce. IEEE Intell Syst 2019;34(2):44–52. https://doi.org/10.1109/MIS.2019.2895788. March-April.
[7] Gaur A, Dubey R. Predicting customer churn prediction in telecom sector using various machine learning techniques. In: Proceedings of the international conference on advanced computation and telecommunication (ICACAT); 2018. p. 1–5. https://doi.org/10.1109/ICACAT.2018.8933783.
[8] Borja B, Bernardino C, Alex C, Ricard G, David MM. The architecture of a churn prediction system based on stream mining. Front Artif Intell Appl 2013:157–66. https://doi.org/10.3233/978-1-61499-320-9-157.
[9] Yildiz M, Varlı S. Customer churn prediction in telecommunication. In: Proceedings of the 23rd signal processing and communications applications conference, SIU 2015 - proceedings; 2015. p. 256–9. https://doi.org/10.1109/SIU.2015.7129808.
[10] Devriendt F, Berrevoets J, Verbeke W. Why you should stop predicting customer churn and start using uplift models. Inf Sci 2021;548:497–515. https://doi.org/10.1016/j.ins.2019.12.075 (Ny).
[11] Saran Kumar A, Chandrakala D. A survey on customer churn prediction using machine learning techniques. Int J Comput Appl 2016;154:13–6. https://doi.org/10.5120/ijca2016912237.
[12] Kraljević G, Gotovac S. Modeling data mining applications for prediction of prepaid churn in telecommunication services. Automatika 2010;51(3):275–83. https://doi.org/10.1080/00051144.2010.11828381.
[13] Hung SY, Yen DC, Wang HY. Applying data mining to telecom churn management. Expert Syst Appl 2006;31(3):515–24. https://doi.org/10.1016/j.eswa.2005.09.080.

[14] Alzubaidi AMN, Al-Shamery E. Predicting customer churn in telecom sector based on penalization techniques and ensemble machine learning. Int J Eng Technol 2018;7:657–64. https://doi.org/10.14419/ijet.v7i4.19.27977.
[15] Khalid LF, Mohsin Abdulazeez A, Zeebaree DQ, Ahmed FYH, Zebari DA. Customer churn prediction in telecommunications industry based on data mining. In: Proceedings of the IEEE symposium on industrial electronics & applications (ISIEA); 2021. p. 1–6. https://doi.org/10.1109/ISIEA51897.2021.9509988.
[16] Zhao M, Zeng Q, Chang M, Tong Q, Su J. A prediction model of customer churn considering customer value: an empirical research of telecom industry in China. Discrete Dyn Nat Soc 2021;2021:12. https://doi.org/10.1155/2021/7160527. Article ID 7160527.
[17] Nadeem AN, Umar S, Shahzad MS. A review on customer churn prediction data mining modeling techniques. Indian J Sci Technol 2018;11(27):1–7. https://doi.org/10.17485/ijst/2018/v11i27/121478.
[18] Lalwani P, Mishra MK, Chadha JS. Customer churn prediction system: a machine learning approach. Computing 2022;104:271–94. https://doi.org/10.1007/s00607-021-00908-y.
[19] Rajamohamed R, Manokaran J. Improved credit card churn prediction based on rough clustering and supervised learning techniques. Cluster Comput 2018;21(1):65–77. https://doi.org/10.1007/s10586-017-0933-1.
[20] Amin A, Adnan A, Anwar S. An adaptive learning approach for customer churn prediction in the telecommunication industry using evolutionary computation and Naïve Bayes. Appl Soft Comput 2023;137. https://doi.org/10.1016/j.asoc.2023.110103.
[21] Amin A, Al-Obeidat F, Shah B, Al Tae M, Khan C, Ur Rehman Durrani H, Anwar S. Just-in-time customer churn prediction in the telecommunication sector. J Supercomput 2020;76:3924–48. https://doi.org/10.1007/s11227-017-2149-9.
[22] Amin A, Shah B, Khattak AM, Lopes Moreiraet FJ, Ali G, Rocha A, Anwar S. Cross-company customer churn prediction in telecommunication: a comparison of data transformation methods. Int J Inf Manag 2019;46:304–19. https://doi.org/10.1016/j.ijinfomgt.2018.08.015.
[23] Amin A, Anwar S, Adnan A, Nawaz M, Alawfi K, Hussain A, Haung K. Customer churn prediction in the telecommunication sector using a rough set approach. Neurocomputing 2017;237:242–54. https://doi.org/10.1016/j.neucom.2016.12.009.
[24] Amin A, Shah B, Khattak AM, Lopes Moreiraet FJ, Ali G, Rocha A, Loo J, Anwar S. Customer churn prediction in telecommunication industry using data certainty. J Bus Res 2019;94:290–301. https://doi.org/10.1016/j.jbusres.2018.03.003.
[25] Ahmad A, Jafar A, Aljoumaa K. Customer churn prediction in telecommunication using machine learning in big data platform. J Big Data 2019;6:1–24. https://doi.org/10.1186/s40537-019-0191-6.

# Implementing machine learning techniques for customer retention and churn prediction in telecommunications

**Article** · August 2024

DOI: 10.51594/csitrj.v5i8.1489



**5 authors**, including:

Angela Omozele Abhulimen
Northumbria University

OPEN ACCESS
Computer Science & IT Research Journal
P-ISSN: 2709-0043, E-ISSN: 2709-0051
Volume 5, Issue 8, P.2011-2025, August 2024
DOI: 10.51594/csitrj.v5i8.1489
Fair East Publishers
Journal Homepage: www.fepbl.com/index.php/csitrj

# Implementing machine learning techniques for customer retention and churn prediction in telecommunications

Ibrahim Adedeji Adeniran[1], Christianah Pelumi Efunniyi[2], Olajide Soji Osundare[3], & Angela Omozele Abhulimen[4]

[1]International Association of Computer Analysts and Researchers, Abuja, Nigeria
[2]OneAdvanced, UK
[3]Nigeria Inter-Bank Settlement System Plc (NIBSS), Nigeria
[4]Independent Researcher, UK

*Corresponding Author: Ibrahim Adedeji Adeniran
Corresponding Author Email: obragado1989@gmail.com
**Article Received:** 05-04-24 **Accepted:** 29-06-24 **Published:** 31-08-24



## ABSTRACT

This review paper explores the application of machine learning techniques in predicting customer churn and enhancing customer retention within the telecommunications industry. The paper begins by discussing the significance of customer churn, its causes, and the limitations of traditional churn prediction methods. It then delves into machine learning algorithms, including decision trees, support vector machines, and ensemble methods. It highlights their effectiveness in handling large and complex datasets typical of the telecom sector. The discussion extends to the challenges faced in data quality, model selection, implementation, and ethical considerations in using customer data for predictive analytics. The paper also compares machine learning models with traditional methods, emphasizing the advantages of scalability, accuracy, and real-time processing. Furthermore, it identifies potential innovations, such as improved data integration, interpretable models, and personalized retention strategies. Finally, the paper reflects on future trends, predicting the growing role of AI and machine learning in telecommunications, particularly in customer service automation and network optimization. The review underscores the importance of adopting machine learning to reduce churn and improve customer retention while considering the field's ethical implications and future opportunities.

**Keywords**: Customer Churn Prediction, Machine Learning, Telecommunications, Customer Retention, Predictive Analytics, AI in Telecom.

## INTRODUCTION

### Overview of Customer Retention and Churn in Telecommunications

Customer retention is critical to business strategy in the telecommunications industry as companies strive to maintain a stable and loyal customer base. Customer churn, the process by which customers stop using a company's services and switch to a competitor, poses a significant threat to profitability and market share (Rane, Achari, & Choudhary, 2023). High churn rates can lead to increased costs associated with acquiring new customers and a loss of revenue from existing customers. In such a competitive environment, where the cost of acquiring new customers often exceeds the cost of retaining existing ones, telecom companies must prioritize strategies that minimize churn and enhance customer loyalty (Hammah, 2020).

Customer churn can occur due to various factors, including dissatisfaction with service quality, pricing, customer service, and the availability of better offers from competitors. The impact of churn is not limited to lost revenue; it also affects brand reputation and market positioning. Consequently, understanding and predicting customer churn has become a top priority for telecom companies, allowing them to retain customers and improve overall business performance proactively (Bhattacharyya & Dash, 2021; Lappeman, Franco, Warner, & Sierra-Rubia, 2022).

Addressing customer churn is crucial for telecom companies for several reasons. First, the telecommunications market is highly saturated, with numerous service providers offering similar products and services. This saturation makes it easy for customers to switch providers if unsatisfied with their current service, increasing the risk of churn. Second, the cost of acquiring a new customer is often much higher than the cost of retaining an existing one. This is due to the significant marketing, promotional, and operational expenses of attracting new customers. As a result, retaining customers can lead to substantial cost savings and higher profitability (Saleh & Saha, 2023).

High customer churn rates can also negatively impact a company's financial stability and growth prospects. When customers leave, their revenue is lost, disrupting cash flow and financial planning. This loss can be particularly damaging in the telecommunications industry, where companies often operate on thin margins. Furthermore, high churn rates can erode customer trust and brand loyalty, making it even more challenging to retain and attract new customers (Melian, Dumitrache, Stancu, & Nastu, 2022). Moreover, regulatory pressures and increased competition in the telecom sector have forced companies to focus more on customer-centric strategies. Governments and regulatory bodies often impose strict guidelines on service quality and pricing, making it imperative for telecom companies to meet customer expectations consistently. Failure to do so can result in penalties, legal challenges, and further loss of customers to competitors. In this context, effective customer retention strategies are essential for maintaining a competitive edge and ensuring long-term business success (Quach, Thaichon, & Hewege, 2020).

**The Role of Machine Learning in Predicting Churn and Enhancing Customer Retention**

Machine learning has emerged as a powerful tool in the fight against customer churn in the telecommunications industry. By leveraging large volumes of customer data, machine learning algorithms can identify patterns and trends that indicate the likelihood of a customer churning. These predictive models can analyze factors such as customer behavior, service usage, payment history, and interactions with customer support to determine which customers are at risk of leaving (Chigwende, 2021).

One of the key advantages of using machine learning for churn prediction is its ability to process and analyze vast amounts of data in real-time. Traditional statistical methods often struggle to handle the complexity and volume of data generated by telecom companies, leading to less accurate predictions. Machine learning, on the other hand, can continuously learn from new data, improving its predictive accuracy over time. This enables telecom companies to make data-driven decisions and implement targeted interventions to retain high-risk customers before they leave (Singh et al., 2024).

Furthermore, machine learning can help telecom companies personalize their customer retention strategies. By segmenting customers based on their likelihood of churning, companies can tailor their retention efforts to address different customer groups' specific needs and preferences. For example, a customer at risk of leaving due to dissatisfaction with service quality may respond positively to an offer of improved service or a discount. In contrast, customers considering switching to a competitor may be retained through a loyalty program or special promotion (Rane et al., 2023).

**Purpose and Objectives of the Paper**

The primary purpose of this paper is to explore the implementation of machine learning techniques for customer retention and churn prediction in the telecommunications industry. By examining the challenges and opportunities associated with using machine learning for churn prediction, the paper aims to comprehensively understand how these advanced technologies can enhance customer retention strategies in telecom companies.

The objectives of this paper are threefold. First, it seeks to provide an overview of the current landscape of customer retention and churn in the telecommunications industry, highlighting the importance of addressing churn for business success. Second, the paper aims to analyze the role of machine learning in predicting churn, discussing the advantages and limitations of various machine learning models and techniques. Finally, the paper will offer insights into the future trends and developments in the application of machine learning for customer retention, exploring potential innovations that could further enhance the effectiveness of churn prediction and prevention strategies.

In conclusion, customer retention and churn prediction are critical areas of focus for telecom companies, given the competitive nature of the industry and the significant impact of churn on profitability and growth. Machine learning offers a promising solution to the challenges of churn prediction, enabling telecom companies to make data-driven decisions and implement targeted retention strategies. By exploring the implementation of machine learning techniques in this context, this paper aims to contribute to the ongoing efforts to improve customer retention and ensure the long-term success of telecom companies.

## BACKGROUND AND LITERATURE REVIEW

### Overview of Customer Churn

Customer churn, or customer attrition, refers to the phenomenon where customers discontinue using a company's products or services over time. In the telecommunications industry, churn is a critical metric that directly impacts a company's revenue, market share, and long-term sustainability. Churn can be voluntary, where a customer actively chooses to leave due to dissatisfaction or better offers from competitors, or involuntary, where customers are lost due to factors like payment failures or relocation to areas outside the service coverage (Bhattacharyya & Dash, 2021). The causes of customer churn in telecommunications are varied and complex. One of the primary causes is dissatisfaction with service quality, including frequent call drops, poor network coverage, and slow internet speeds. Additionally, pricing plays a significant role, with customers often seeking more cost-effective plans or switching to providers offering better value. Customer service quality is another critical factor; customers who experience poor customer support are more likely to leave for a competitor who offers more responsive and helpful service (Rane et al., 2023; Saleh & Saha, 2023).

The implications of customer churn extend beyond immediate revenue loss. High churn rates can indicate underlying issues with service quality, pricing strategies, or customer engagement, signaling deeper operational inefficiencies (Uner, Guven, & Cavusgil, 2020). Moreover, churn can lead to increased customer acquisition costs, as companies must spend more on marketing and promotional efforts to replace lost customers. Over time, this can erode profit margins and negatively impact a company's brand reputation. Furthermore, high churn rates can create a negative feedback loop, where the loss of customers leads to reduced investment in service improvements, further driving up churn (Bhattacharyya & Dash, 2021; Borah, Prakhya, & Sharma, 2020).

### Review of Traditional Methods for Churn Prediction and Customer Retention

Before the advent of advanced machine learning techniques, telecommunications companies relied on traditional statistical methods and business rules to predict customer churn and develop retention strategies. One common approach involved analyzing historical data to identify trends and patterns indicating a customer's likelihood of churning. Logistic regression, decision trees, and survival analysis were commonly used to model customer behavior and predict churn.

Logistic regression, for instance, is a statistical method that estimates the probability of a binary outcome, such as whether a customer will churn or stay. It uses various customer attributes, such as age, contract length, and usage patterns, as predictors. Decision trees, on the other hand, split the data into subsets based on the value of input variables, making it easier to identify segments of customers with a higher likelihood of churn. Survival analysis, often used in medical research, was adapted to estimate the time until a customer would churn, allowing companies to intervene before the critical moment (Routh, Roy, & Meyer, 2021).

While these traditional methods provided valuable insights, they were often limited in scope and accuracy. The models were typically linear and could not capture the complex, non-linear relationships between variables often present in customer behavior data. Additionally, these methods required extensive domain knowledge to define the rules and thresholds for churn prediction, making them less adaptable to changes in customer behavior or market conditions. The manual effort involved in data preprocessing and model building also limited these traditional approaches' scalability and real-time applicability. Customer retention strategies in the traditional framework were largely reactive, focusing on incentives like discounts, loyalty programs, and personalized offers to prevent churn. These strategies were often implemented after a customer showed dissatisfaction, making them less effective. Moreover, the one-size-fits-all approach to retention failed to address different customer segments' diverse needs and preferences, leading to suboptimal results (Capponi, Corrocher, & Zirulia, 2021; Mitchell, 2020).

**Overview of Machine Learning Techniques**

Machine learning represents a significant advancement over traditional methods for churn prediction and customer retention in telecommunications. Machine learning involves training algorithms on large datasets to identify patterns, make predictions, and optimize decisions without explicit programming for every possible scenario. Unlike traditional models, which often rely on predefined rules and linear relationships, machine learning models can learn from data and adapt to new information, making them more flexible and accurate (Sikri, Jameel, Idrees, & Kaur, 2024). Machine learning techniques such as supervised learning, unsupervised learning, and ensemble methods have been widely adopted in the context of churn prediction. Supervised learning algorithms, including support vector machines, random forests, and gradient boosting, are trained on labeled datasets where the outcome (churn or no churn) is known. These models learn to predict churn by finding patterns in customer attributes such as demographics, usage behavior, and interaction history. For example, a random forest model, which uses multiple decision trees to improve prediction accuracy, can identify complex interactions between variables indicative of churn risk (Gattermann-Itschert & Thonemann, 2022).

Unsupervised learning techniques, such as clustering and principal component analysis (PCA), segment customers into groups with similar characteristics or behaviors. These techniques do not require labeled data and are useful for discovering hidden patterns or customer segments at risk of churning. Clustering, for instance, can group customers based on their usage patterns, enabling telecom companies to target specific segments with tailored retention strategies (Sharaf Addin, Admodisastro, Mohd Ashri, Kamaruddin, & Chong, 2022). Ensemble methods, which combine multiple models to improve prediction performance, have also gained popularity in churn prediction. Techniques like bagging and boosting aggregate the predictions of several base models to reduce variance and bias, leading to more robust and accurate predictions. These models are particularly useful in noisy data scenarios or highly complex relationships between variables (Tavassoli & Koosha, 2022). The application of machine learning in telecommunications extends beyond churn prediction to other areas, such as fraud detection, network optimization, and personalized marketing. By leveraging machine learning, telecom companies can analyze vast amounts of data in real-time, automate decision-making processes, and implement proactive strategies that enhance customer satisfaction and retention (Bharadiya, 2023; Hassan & Mhmood, 2021).

### Summary of Recent Research and Trends in Machine Learning for Churn Prediction

Recent research in machine learning for churn prediction has focused on improving model accuracy, interpretability, and scalability. One key trend is the integration of deep learning techniques, such as neural networks, which can model complex, non-linear relationships and process unstructured data such as text and images. For example, recurrent neural networks (RNNs) have been used to analyze sequential customer data, such as call logs or browsing history, to predict churn more accurately (Wu et al., 2022).

Another trend is hybrid models, which combine traditional statistical methods with machine learning techniques. These models leverage the strengths of both approaches, using machine learning to capture non-linear patterns while retaining the interpretability and simplicity of traditional methods. For instance, a hybrid model might use logistic regression for its ease of interpretation and a neural network to capture complex interactions between variables (Sansana et al., 2021).

Explainable AI (XAI) has also become a recent research focus, addressing the challenge of interpretability in machine learning models. Telecom companies often require transparent models that explain why a particular customer will likely churn. Techniques such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) have been developed to make machine learning models more interpretable, enabling businesses to understand the key drivers of churn and make informed decisions (Bramhall, Horn, Tieu, & Lohia, 2020; Wagh et al., 2024). Finally, there is growing interest in real-time churn prediction and intervention. Advances in streaming data processing and real-time analytics have enabled telecom companies to monitor customer behavior continuously and predict churn as soon as risk factors emerge. This allows for timely interventions, such as targeted offers or personalized communication, to retain customers before they decide to leave (Hassan & Mhmood, 2021).

## CHALLENGES IN CUSTOMER RETENTION AND CHURN PREDICTION

### Data-Related Challenges

One of the most significant challenges in customer retention and churn prediction in the telecommunications industry is data availability and quality. Machine learning models rely heavily on large, high-quality datasets for accurate predictions. However, obtaining such data can be a complex and resource-intensive process. Data may be scattered across different systems, stored in various formats, or incomplete, making it difficult to gather a comprehensive and accurate dataset for churn prediction (Joy, Hoque, Uddin, Chowdhury, & Park, 2024).

Data availability is often a hurdle because telecom companies may not have access to all the necessary data points required for robust churn prediction. While companies typically collect data on customer interactions, usage patterns, billing information, and service complaints, other critical data, such as customer sentiment from social media or third-party data about competitor offers, may be unavailable or difficult to integrate. Additionally, certain customer behaviors that are strong indicators of churn, such as the intention to switch providers, may not be directly observable in the available data (Wagh et al., 2024).

Data quality is equally critical, as poor-quality data can lead to inaccurate predictions and ineffective retention strategies. Issues such as missing data, duplicate records, and inconsistencies in data entry can introduce noise and bias into the model, reducing its accuracy and reliability. Data preprocessing, therefore, becomes a vital step in the churn prediction process. It involves cleaning the data, handling missing values, normalizing variables, and transforming categorical data into numerical formats that machine learning algorithms can process. However, data preprocessing is a time-consuming and complex task that requires significant expertise and resources. Errors in this stage can lead to faulty models, ultimately affecting the effectiveness of customer retention efforts (Tamuka & Sibanda, 2020). Moreover, telecom companies often face challenges obtaining real-time data, which is crucial for timely churn prediction and intervention. Traditional data collection methods may not be fast enough to capture the most recent customer interactions or behaviors, leading to delays in detecting churn risk. This lag can result in missed opportunities to retain customers, especially in a competitive market where customers can easily switch to another provider.

**Model-Related Challenges**

Selecting the appropriate machine learning algorithm for churn prediction is another significant challenge. The telecommunications industry generates vast amounts of data, including structured data, like customer demographics and billing information, and unstructured data, such as customer support transcripts and social media posts. Different algorithms are better suited to different data types, and choosing the wrong model can lead to suboptimal results. For instance, while decision trees are easy to interpret and work well with structured data, they may not perform as well as deep learning models when handling complex, unstructured data (Geiler, Affeldt, & Nadif, 2022).

Overfitting is a common problem in machine learning models, particularly in churn prediction. Overfitting occurs when a model learns the details and noise in the training data to the extent that it negatively impacts its performance on new, unseen data. In churn prediction, this can happen when a model becomes too complex, capturing not only the genuine patterns that indicate churn but also the irrelevant random fluctuations in the data. An overfitted model may perform exceptionally well on historical data but fail to generalize to new data, leading to inaccurate predictions and ineffective retention strategies (Khoh, Pang, Ooi, Wang, & Poh, 2023).

Addressing overfitting requires careful consideration during model development. Techniques such as cross-validation, regularization, and pruning of decision trees can help prevent overfitting by ensuring the model remains general enough to apply to new data. However, striking the right balance between model complexity and generalization is often challenging, especially when dealing with high-dimensional datasets common in telecommunications.

Interpretability is another critical challenge in using machine learning models for churn prediction. While advanced models like neural networks and ensemble methods can offer high accuracy, they often operate as "black boxes," making it difficult to understand how they arrive at their predictions. This lack of transparency can be problematic for telecom decision-makers who need to understand the factors driving churn to develop targeted retention strategies. Without clear insights into why a customer is likely to churn, it becomes challenging to implement effective interventions. Furthermore, regulatory requirements may demand that companies provide explanations for automated decisions, adding another layer of complexity to using opaque machine learning models (Guliyev & Yerdelen Tatoğlu, 2021; Tékouabou, Gherghina, Toulni, Mata, & Martins, 2022).

**Implementation Challenges**

Integrating machine learning models into existing telecommunications systems poses several implementation challenges. Telecom companies often operate legacy systems that were not designed to handle the demands of modern data analytics. These systems may lack the computational power, data storage capacity, or flexibility needed to support machine learning algorithms. As a result, integrating new models with existing infrastructure can require significant investment in hardware, software, and technical expertise.

Cost considerations are a major challenge in implementing machine learning for churn prediction. Developing, deploying, and maintaining machine learning models can be expensive, particularly for smaller telecom companies with limited budgets. The costs associated with data collection, preprocessing, model development, and system integration can quickly add up. Additionally, ongoing expenses such as cloud computing resources, data storage, and the need for continuous model updates to maintain accuracy further increase the financial burden. Companies must carefully evaluate their churn prediction initiatives' potential return on investment (ROI) to justify these costs (Morozov, Mezentseva, Kolomiiets, & Proskurin, 2022).

Scalability is another critical concern, especially for large telecom companies with millions of customers generating massive amounts of data daily. Machine learning models must be scalable to handle this data volume and complexity without sacrificing performance. However, scaling machine learning models is not a straightforward task. It often requires optimizing algorithms for parallel processing, distributing workloads across multiple servers, and ensuring the system can process data in real-time. Failure to address scalability issues can lead to slow processing times, delayed predictions, and missed opportunities to retain customers (Amajuoyi, Nwobodo, & Adegbola, 2024).

Furthermore, deploying machine learning models in a production environment introduces additional challenges. These models must be continuously monitored and updated to remain accurate as customer behaviors and market conditions change. This requires a robust infrastructure for model versioning, testing, and deployment and a team of data scientists and engineers to manage the process. Without proper maintenance, even the most accurate models can quickly become outdated, leading to a decline in prediction accuracy and effectiveness (Paleyes, Urma, & Lawrence, 2022).

**Ethical and Privacy Concerns in Using Customer Data for Predictive Analytics**

Using customer data for predictive analytics in churn prediction raises several ethical and privacy concerns. Telecom companies collect vast amounts of personal data, including sensitive information such as call logs, browsing history, location data, and payment details. While this data is invaluable for building accurate machine-learning models, it raises questions about customer consent, data security, and potential misuse. One of the primary ethical concerns is the issue of informed consent. Customers may not always be aware of the extent to which their data is being collected, analyzed, and used for predictive purposes. Even when consent is obtained, it is often through lengthy and complex terms and conditions that customers may not fully understand. This lack of transparency can lead to a breach of trust, as customers may feel that their privacy has been violated. Telecom companies must ensure that they obtain clear, informed consent from customers and communicate how their data will be used in a way that is easy to understand (Godinho de Matos & Adjerid, 2022).

Data security is another critical concern, as the collection and storage of vast amounts of customer data make telecom companies attractive targets for cyberattacks. A data breach can have severe consequences, including financial losses, legal penalties, and company reputation damage. Protecting customer data requires robust security measures, including encryption, access controls, and regular security audits. However, implementing these measures can be costly and complex, especially for smaller companies with limited resources (Djenna, Harous, & Saidouni, 2021).

There is also the risk of bias and discrimination in machine learning models for churn prediction. Suppose the data used to train the models is biased. In that case, the predictions and subsequent retention strategies may also be biased, leading to unfair treatment of certain customer groups. For example, a model trained on historical data that reflects past discriminatory practices may perpetuate those biases, resulting in some customers being unfairly targeted for retention efforts or ignored altogether. Ensuring fairness and transparency in machine learning models requires careful attention to data quality, model development, and testing processes.

## DISCUSSION AND ANALYSIS

### Analysis of the Effectiveness of Different Machine Learning Techniques in Churn Prediction

The effectiveness of machine learning techniques in churn prediction has been extensively studied, with many approaches demonstrating significant advantages over traditional methods. Among the most effective techniques are supervised learning algorithms, such as logistic regression, decision trees, support vector machines (SVM), and ensemble methods like random forests and gradient boosting machines (GBM). These models are particularly well-suited for churn prediction because they can process large datasets with multiple features and identify complex patterns that indicate a customer's likelihood of churning.

While simple and interpretable, logistic regression often serves as a baseline model in churn prediction studies. It is particularly effective when the relationship between the predictors and the outcome is linear. However, its performance can be limited when dealing with more complex data where non-linear interactions between variables are present. In contrast, decision trees provide a more nuanced approach by capturing non-linear relationships, making them more effective in modeling customer behavior. However, decision trees can be prone to overfitting, especially when the model is complex, which can limit their generalizability.

Support vector machines (SVM) have shown effectiveness in churn prediction by maximizing the margin between different classes (churn vs. no churn). SVMs are particularly powerful in handling high-dimensional data, where many features are used to predict churn. However, they can be computationally intensive, especially with large datasets typical in the telecommunications industry.

Ensemble methods like random forests and gradient boosting machines (GBM) are among the most effective for churn prediction. These models combine multiple decision trees to improve prediction accuracy and reduce overfitting. Random forests, for example, create multiple decision trees using different subsets of the data and average the predictions, which helps stabilize the results and make the model more robust. GBMs, on the other hand, build trees sequentially, where each new tree focuses on correcting the errors made by the previous ones. This approach often leads to higher accuracy but at the cost of increased complexity and computational resources.

Deep learning techniques, particularly neural networks, have also been explored for churn prediction, especially in scenarios involving unstructured data such as text from customer support interactions or social media activity. Recurrent neural networks (RNNs) and long short-term memory (LSTM) networks effectively model sequential data, which can be crucial in understanding customer behavior over time. However, these models are often viewed as black boxes due to their lack of interpretability, making them less desirable in contexts where understanding the factors driving churn is essential.

**Comparison of Machine Learning Models with Traditional Methods**

When comparing machine learning models with traditional methods like logistic regression and decision trees, it is evident that machine learning offers several advantages, particularly in accuracy and scalability. While useful, traditional methods often fall short of capturing the complex, non-linear relationships in customer data. For instance, despite being easy to interpret, logistic regression is limited to linear relationships, which can be overly simplistic for churn prediction.

Decision trees, a traditional method, provide more flexibility by capturing non-linear relationships. However, they are prone to overfitting, especially when dealing with noisy data. In contrast, ensemble methods such as random forests and gradient boosting machines significantly improve upon decision trees by reducing the likelihood of overfitting and increasing prediction accuracy through the aggregation of multiple models.

Machine learning models also excel in handling large and complex datasets. In telecommunications, where companies manage vast amounts of customer data, including usage patterns, billing history, and customer support interactions, the ability to process and analyze this data efficiently is crucial. Traditional models often struggle with scalability, as they may require extensive manual preprocessing and are not designed to handle large volumes of data in real time.

Furthermore, machine learning models can automate the feature selection process, where the algorithm identifies the most relevant features for prediction. This is a significant improvement over traditional methods, which often rely on domain experts to manually select features, which can be time-consuming and prone to human error. Machine learning models can also adapt to changing customer behaviors, providing more accurate and up-to-date predictions. However, one of the key challenges in adopting machine learning models over traditional methods is the trade-off between accuracy and interpretability. Traditional models like logistic regression and decision trees are more interpretable, allowing business stakeholders to understand the factors driving churn easily. This interpretability is crucial for developing targeted retention strategies. In contrast, while offering higher accuracy, more complex machine learning models often lack transparency, making it difficult to derive actionable insights.

**Potential Improvements and Innovations in the Application of Machine Learning**

There are several potential areas for improvement and innovation in the application of machine learning for customer retention in the telecommunications industry. One area is the enhancement of data integration capabilities. Telecom companies collect data from various sources, including customer interactions, network usage, and external data like social media activity. Integrating these diverse data sources into a unified model could provide a more comprehensive view of customer behavior and improve the accuracy of churn predictions.

Another area for improvement is the development of more interpretable machine learning models. As mentioned earlier, the lack of transparency in complex models like deep learning networks is a significant barrier to their adoption in business settings. Innovations in explainable AI (XAI) are promising in this regard. Techniques such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) can help make the predictions of complex models more understandable, providing insights into which features are driving churn and how retention strategies can be optimized.

Personalization is another area ripe for innovation. Machine learning models can be used to develop more personalized retention strategies that consider individual customer preferences and behaviors. For example, telecom companies could use machine learning to predict whether a customer will churn and which specific incentives, such as discounts or personalized offers, would be most effective in retaining that customer. This level of personalization could significantly improve customer satisfaction and loyalty.

Real-time analytics and prediction are also critical areas for future innovation. As customer behavior evolves rapidly, the ability to predict churn in real time and intervene immediately could provide a significant competitive advantage. Advances in streaming data processing and real-time machine learning models could enable telecom companies to monitor customer interactions continuously and take proactive measures before a customer decides to leave.

**Future Trends and the Evolving Role of AI and Machine Learning in Telecommunications**

The role of AI and machine learning in telecommunications is expected to expand significantly, driven by ongoing advancements in technology and increasing competition in the industry. One of the key future trends is the integration of AI-driven automation in customer service. Chatbots and virtual assistants powered by machine learning algorithms are becoming increasingly sophisticated, capable of handling complex customer inquiries and providing personalized service. These AI-driven systems can help telecom companies improve customer satisfaction and reduce churn by providing timely and accurate support.

Another emerging trend is the use of machine learning for predictive maintenance of telecom infrastructure. By analyzing data from network equipment and customer usage patterns, machine learning models can predict when equipment will likely fail and schedule maintenance before disruptions occur. This proactive approach can improve service reliability, a critical factor in customer retention.

AI and machine learning are also expected to play a significant role in optimizing network resources. As telecom companies deploy 5G networks, the complexity of managing network resources will increase. Machine learning algorithms can help optimize network traffic, allocate resources more efficiently, and ensure a high quality of service for customers. This, in turn, can reduce customer churn by providing a more reliable and consistent user experience. Finally, the ethical use of AI and machine learning in telecommunications will become increasingly important. As telecom companies collect and analyze more customer data, they must ensure that their practices comply with privacy regulations and ethical standards. This includes being transparent about how customer data is used, obtaining informed consent, and ensuring that AI-driven decisions do not perpetuate bias or discrimination.

## CONCLUSION

This paper has explored the crucial role of machine learning in predicting customer churn and enhancing customer retention within the telecommunications industry. As highlighted, customer churn poses a significant challenge for telecom companies, leading to substantial revenue losses and increased customer acquisition costs. While useful, traditional methods of churn prediction often fall short in dealing with the complexity and scale of modern telecommunications data. Machine learning techniques, on the other hand, offer powerful tools to analyze large datasets, identify patterns, and predict churn with greater accuracy.

Key findings from the paper include the effectiveness of various machine learning algorithms, such as decision trees, support vector machines (SVM), and ensemble methods, like random forests and gradient boosting machines (GBM), in improving the accuracy of churn prediction. These models are particularly adept at handling the large volumes of data generated by telecom companies, capturing non-linear relationships between variables, and adapting to changing customer behaviors. The paper also discussed the challenges associated with data quality, model selection, and the implementation of machine learning models, emphasizing the importance of addressing these issues to maximize the benefits of churn prediction.

Furthermore, the paper analyzed the limitations of traditional methods compared to machine learning models, particularly regarding scalability and accuracy. It also highlighted potential improvements and innovations in the application of machine learning for customer retention, including the integration of diverse data sources, the development of more interpretable models, and the personalization of retention strategies. The discussion extended to future trends, focusing on the evolving role of AI and machine learning in the telecommunications industry, particularly in areas such as customer service automation, predictive maintenance, and network optimization.

The importance of using machine learning for churn prediction and customer retention cannot be overstated. In an industry as competitive as telecommunications, retaining existing customers is critical to maintaining profitability and market share. Machine learning offers a significant advantage over traditional methods by enabling telecom companies to predict churn more accurately and efficiently. With the ability to analyze vast amounts of data, machine learning models can uncover hidden patterns and provide insights that are impossible with manual analysis or traditional statistical methods.

Machine learning also allows for more personalized customer retention strategies. By understanding the specific behaviors and preferences of individual customers, telecom companies can tailor their retention efforts to address each customer's unique needs, thereby improving satisfaction and loyalty. This level of personalization is essential in today's market, where customers have more choices than ever and expect tailored experiences from their service providers. Moreover, machine learning's ability to process real-time data is crucial for proactive churn prevention. By continuously monitoring customer behavior, machine learning models can identify signs of dissatisfaction or intent to churn before the customer takes action. This allows telecom companies to intervene promptly with targeted offers or support, potentially saving the customer relationship. In this way, machine learning predicts churn. It provides the tools to prevent it, making it an invaluable asset for telecom companies.

## References


Amajuoyi, C. P., Nwobodo, L. K., & Adegbola, M. D. (2024). Transforming business scalability and operational flexibility with advanced cloud computing technologies. *Computer Science & IT Research Journal, 5*(6), 1469-1487.

Bharadiya, J. P. (2023). Machine learning and AI in business intelligence: Trends and opportunities. *International Journal of Computer (IJC), 48*(1), 123-134.

Bhattacharyya, J., & Dash, M. K. (2021). Investigation of customer churn insights and intelligence from social media: A netnographic research. *Online Information Review, 45*(1), 174-206.

Borah, S. B., Prakhya, S., & Sharma, A. (2020). Leveraging service recovery strategies to reduce customer churn in an emerging market. *Journal of the Academy of Marketing Science*, *48*, 848-868.

Bramhall, S., Horn, H., Tieu, M., & Lohia, N. (2020). Qlime-a quadratic local interpretable model-agnostic explanation approach. *SMU Data Science Review, 3*(1), 4.

Capponi, G., Corrocher, N., & Zirulia, L. (2021). Personalized pricing for customer retention: Theory and evidence from mobile communication. *Telecommunications Policy, 45*(1), 102069.

Chigwende, S. (2021). *The Impact of Corporate Brand Image on Customer Satisfaction, Loyalty and Switching Behavior: A Case Study of Mobile Telecommunications Customers in Zimbabwe.*

Djenna, A., Harous, S., & Saidouni, D. E. (2021). Internet of things meet internet of threats: New concern cyber security issues of critical cyber infrastructure. *Applied Sciences, 11*(10), 4580.

Gattermann-Itschert, T., & Thonemann, U. W. (2022). Proactive customer retention management in a non-contractual B2B setting based on churn prediction with random forests. *Industrial Marketing Management*, *107*, 134-147.

Geiler, L., Affeldt, S., & Nadif, M. (2022). A survey on machine learning methods for churn prediction. *International Journal of Data Science and Analytics*, *14*(3), 217-242.

Godinho de Matos, M., & Adjerid, I. (2022). Consumer consent and firm targeting after GDPR: The case of a large telecom provider. *Management Science, 68*(5), 3330-3378.

Guliyev, H., & Yerdelen Tatoğlu, F. (2021). Customer churn analysis in banking sector: Evidence from explainable machine learning model. *Journal of Applied Microeconometrics, 1*(2).

Hammah, C. A. (2020). A customer retention strategy for Phoenix Insurance Company.

Hassan, A., & Mhmood, A. H. (2021). Optimizing network performance, automation, and intelligent decision-making through real-time big data analytics. *International Journal of Responsible Artificial Intelligence, 11*(8), 12-22.

Joy, U. G., Hoque, K. E., Uddin, M. N., Chowdhury, L., & Park, S.-B. (2024). A Big data-driven hybrid model for enhancing streaming service customer retention through churn prediction integrated with explainable AI. *IEEE access*.

Khoh, W. H., Pang, Y. H., Ooi, S. Y., Wang, L.-Y.-K., & Poh, Q. W. (2023). Predictive churn modeling for sustainable business in the telecommunication industry: optimized weighted ensemble machine learning. *Sustainability, 15*(11), 8631.

Lappeman, J., Franco, M., Warner, V., & Sierra-Rubia, L. (2022). What social media sentiment tells us about why customers churn. *Journal of Consumer Marketing, 39*(5), 385-403.

Melian, D. M., Dumitrache, A., Stancu, S., & Nastu, A. (2022). Customer churn prediction in telecommunication industry. A data analysis techniques approach. *Postmodern Openings, 13*(1 Sup1), 78-104.

Mitchell, W. D. (2020). *Proactive Predictive Analytics Within the Customer Lifecycle to Prevent Customer Churn*: Northcentral University.

Morozov, V., Mezentseva, O., Kolomiiets, A., & Proskurin, M. (2022). *Predicting customer churn using machine learning in IT startups.* Paper presented at the Lecture Notes in Computational Intelligence and Decision Making: 2021 International Scientific Conference" Intellectual Systems of Decision-making and Problems of Computational Intelligence", Proceedings.

Paleyes, A., Urma, R.-G., & Lawrence, N. D. (2022). Challenges in deploying machine learning: a survey of case studies. *ACM Computing Surveys, 55*(6), 1-29.

Quach, S., Thaichon, P., & Hewege, C. (2020). Triadic relationship between customers, service providers and government in a highly regulated industry. *Journal of Retailing and Consumer Services, 55*, 102148.

Rane, N. L., Achari, A., & Choudhary, S. P. (2023). Enhancing customer loyalty through quality of service: Effective strategies to improve customer satisfaction, experience, relationship, and engagement. *International Research Journal of Modernization in Engineering Technology and Science, 5*(5), 427-452.

Routh, P., Roy, A., & Meyer, J. (2021). Estimating customer churn under competing risks. *Journal of the Operational Research Society, 72*(5), 1138-1155.

Saleh, S., & Saha, S. (2023). Customer retention and churn prediction in the telecommunication industry: a case study on a Danish university. *SN Applied Sciences, 5*(7), 173.

Sansana, J., Joswiak, M. N., Castillo, I., Wang, Z., Rendall, R., Chiang, L. H., & Reis, M. S. (2021). Recent trends on hybrid modeling for Industry 4.0. *Computers & Chemical Engineering, 151*, 107365.

Sharaf Addin, E. H., Admodisastro, N., Mohd Ashri, S. N. S., Kamaruddin, A., & Chong, Y. C. (2022). Customer mobile behavioral segmentation and analysis in telecom using machine learning. *Applied Artificial Intelligence, 36*(1), 2009223.

Sikri, A., Jameel, R., Idrees, S. M., & Kaur, H. (2024). Enhancing customer retention in telecom industry with machine learning driven churn prediction. *Scientific Reports, 14*(1), 13097.

Singh, P. P., Anik, F. I., Senapati, R., Sinha, A., Sakib, N., & Hossain, E. (2024). Investigating customer churn in banking: A machine learning approach and visualization app for data science and management. *Data Science and Management, 7*(1), 7-16.

Tamuka, N., & Sibanda, K. (2020). *Real time customer churn scoring model for the telecommunications industry*. Paper presented at the 2020 2nd International Multidisciplinary Information Technology and Engineering Conference (IMITEC).

Tavassoli, S., & Koosha, H. (2022). Hybrid ensemble learning approaches to customer churn prediction. *Kybernetes, 51*(3), 1062-1088.

Tékouabou, S. C., Gherghina, Ș. C., Toulni, H., Mata, P. N., & Martins, J. M. (2022). Towards explainable machine learning for bank churn prediction using data balancing and ensemble-based methods. *Mathematics, 10*(14), 2379.

Uner, M. M., Guven, F., & Cavusgil, S. T. (2020). Churn and loyalty behavior of Turkish digital natives: Empirical insights and managerial implications. *Telecommunications Policy, 44*(4), 101901.

Wagh, S. K., Andhale, A. A., Wagh, K. S., Pansare, J. R., Ambadekar, S. P., & Gawande, S. (2024). Customer churn prediction in telecom sector using machine learning techniques. *Results in Control and Optimization, 14*, 100342.

Wu, X., Li, P., Zhao, M., Liu, Y., Crespo, R. G., & Herrera-Viedma, E. (2022). Customer churn prediction for web browsers. *Expert Systems with Applications, 209*, 118177.

Machine Learning with Applications 17 (2024) 100567

Contents lists available at ScienceDirect

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

# Explaining customer churn prediction in telecom industry using tabular machine learning models

Sumana Sharma Poudel [a], Suresh Pokharel [b], Mohan Timilsina [c,*]

[a] Pokhara University, Nepal college of Information Technology, Nepal
[b] Westcliff University, Presidential Graduate School, Nepal
[c] Data Science Institute, University of Galway, Ireland

ARTICLE INFO

*Keywords:*
Customer churn
Explainable model
Global explainable
Local explainable
Telecommunication

ABSTRACT

The study addresses customer churn, a major issue in service-oriented sectors like telecommunications, where it refers to the discontinuation of subscriptions. The research emphasizes the importance of recognizing customer satisfaction for retaining clients, focusing specifically on early churn prediction as a key strategy. Previous approaches mainly used generalized classification techniques for churn prediction but often neglected the aspect of interpretability, vital for decision-making. This study introduces explainer models to address this gap, providing both local and global explanations of churn predictions. Various classification models, including the standout Gradient Boosting Machine (GBM), were used alongside visualization techniques like Shapley Additive Explanations plots and scatter plots for enhanced interpretability. The GBM model demonstrated superior performance with an 81% accuracy rate. A Wilcoxon signed rank test confirmed GBM's effectiveness over other models, with the *p*-value indicating significant performance differences. The study concludes that GBM is notably better for churn prediction, and the employed visualization techniques effectively elucidate key churn factors in the telecommunications sector.

## 1. Introduction

The service-oriented industries, such as telecommunications, face considerable challenges due to customer churn, where valuable customers are lost to competitors. As the world rapidly embraces digitization, the telecommunications sector serves as a crucial backbone. Notably, it represents a significant contributor to national income, particularly in developing countries, where it plays a substantial role in generating revenue (Liao & Lien, 2012). With its substantial business volume, telecommunications is recognized as a key industry, evident in ongoing technical advancements and a growing number of operators. Consequently, fierce competition among service providers persists (Gerpott, Rams, & Schindler, 2001), leading to the introduction of new technologies, services and strategies aimed at attracting new customer and retaining existing customers. The churn rate in this sector is approximately 2.6% monthly (Hawley, 2003). Comparing the return on investment between acquiring a new customer and retaining an existing one reveals that the latter is less expensive (Reinartz & Kumar, 2003; Yang & Peterson, 2004) and generally easier than upselling (Ascarza, Iyengar, & Schleicher, 2016). Therefore, customer retention is recognized as the most profitable strategy (Qureshi, Rehman, Qamar, Kamal, & Rehman, 2013; Wei & Chiu, 2002) and can positively influence the company's reputation, reducing marketing costs for new customer acquisition (Bolton & Bronkhorst, 1995; Reichheld & Sasser, 1990). So, it is desirable to have thorough research on customer churn and taking proactive measures in response by decision maker can provide a competitive edge to stay ahead in this competition.

The primary goal of the churn prediction is to support the creation of client retention plans in a market that is highly competitive. Churn models are made to predict which customers are likely to quit on their own will and to spot early signs of churn (Wei & Chiu, 2002). For this, companies must leverage their databases as valuable assets to comprehend customer churn behavior (Coussement & Van den Poel, 2008). Fundamentally, these databases contain information on customer service usage, billing details, and satisfaction levels. In addition to predicting customers likely to switch, companies seek to understand churn causes, which aids in profiling prone customers and devising effective retention campaigns (Leung, Pazdor, & Souza, 2021). Effective churn modeling has two important components: (i) predicting whether a specific customer will churn, (ii) discovering the reasons behind their churn, either at a local or global level. While

* Corresponding author.
*E-mail addresses:* sumana.192978@ncit.edu.np (S.S. Poudel), suresh.pokharel@presidential.edu.np (S. Pokharel), mohan.timilsina@insight-centre.org (M. Timilsina).

https://doi.org/10.1016/j.mlwa.2024.100567
Received 9 February 2024; Received in revised form 28 March 2024; Accepted 19 June 2024
Available online 24 June 2024

much of the existing research predominantly focuses on the first aspect. They are treating churn prediction as a binary classification task and employing various machine learning techniques around it such as feature extraction (Zhao, Gao, Dong, Dong, & Dong, 2017), feature selection (Umayaparvathi & Iyakutti, 2017), treatment of imbalanced datasets (Fujo et al., 2022), and utilizing classifiers like SVC (Cortes & Vapnik, 1995), Logistic Regression (Hosmer, Lemeshow, & Sturdivant, 2013), Random Forest (Breiman, 2001), XGBoost (Friedman, 2001), and Neural networks (Goodfellow, Bengio, & Courville, 2016). However, this alone may not suffice to fully grasp customer behavior and these ignore the second important component. These model cannot explain the reason behind the churning.

This study aims to close the research gap in the field of churning prediction by focusing not only on forecasting whether a certain customer would churn or not, but also on the reason why. For the reasoning we adapt SHapley Additive exPlanations (SHAP) to explain machine learning predictions by identifying influential customers from the training set (Lundberg & Lee, 2017a). The specific research questions (RQs) investigated are:

- **What are the best available off-the-shelf machine learning algorithms for predicting customer churn?**
  1. Which classification algorithm performs best for churn prediction in terms of different evaluation metrics?
  2. Is there a significant difference in the predictions made by these classifiers?
- **How can we explain the factors responsible for customer churn?**
  1. What are the most important predictors, and how do they influence prediction performance?
  2. Is there any interaction between the churn predictors?

**Contributions:** Our contributions are summarized as follows:

With this, this research's contributions are summarized as follows:

- We rigorously compared state-of-the-art supervised machine learning
  algorithms for churn prediction.
- We performed statistical tests to find the most significant model for churn prediction.
- We provide explanations for each predictor corresponding to customer churn, highlighting both positive and negative contributions to churn prediction.

To the best of our knowledge, our approach is the first to generate global and/or local explanations for churn prediction. We conducted rigorous experiments to evaluate tabular machine learning algorithms using different evaluation metrics and to choose the most significant model.

The remainder of the paper is organized as follows: related work, problem definition, method description, experiments, and conclusion.

## 2. Related work

Recently, data mining techniques have emerged to tackle the challenging problems of customer churn in telecommunication service field (Au, Chan, & Yao, 2003; Hadden, Tiwari, Roy, & Ruta, 2007). As one of the important measures to retain customers, churn prediction has been a concern in the telecommunication industry and research (Bin, Peiji, & Juan, 2007). Majority of the research focused on churn prediction were dedicated in voice services available over mobile and fixed-line networks. In most of the cases, the features used for churn prediction in mobile telecommunication industry includes customer demographics, contractual data, customer service logs, call details, complaint data, bill and payment information (Bin et al., 2007; Hadden et al., 2007). However, the information for land-line services providers are different than mobile services (Bin et al., 2007). Some of this data is missing, less reliable or incomplete in land-line communication service providers. For instances, customer ages and complaint data, fault reports are unavailable and only the call details of a few months are available. Due to business confidentiality and privacy, there are no public datasets for churn prediction (Huang, Kechadi, & Buckley, 2012).

Customer churn prediction models have demonstrated significant value beyond telecommunications, notably within industries like digital marketing, e-commerce, and banking, where understanding and mitigating churn is equally critical. In digital marketing, the application of churn models facilitates the optimization of customer engagement and retention strategies. For instance, Ascarza (2018) in their work delve into how digital marketing efforts can be tailored to retain customers showing signs of churn, offering insights into the effectiveness of targeted interventions. In the banking sector, Miguéis, Van den Poel, Camanho, and e Cunha (2012) apply churn prediction to understand and predict customer churn concerning specific banking products and services. These references collectively highlight the broad applicability of churn prediction models across various industries, emphasizing their potential to inform and refine customer retention strategies in diverse business contexts.

Churn analysis and prediction task is also tackled from statistical modeling perspective. A very popular approach to model churn is time to event prediction (Bhattacharya, 1998; Van den Poel & Lariviere, 2004). In the context of customer attrition, the time to failure links to the churn behavior. The potential churner behavior has also been considered using structural equation modeling (Nguyen & LeBlanc, 1998; Varki & Colgate, 2001). Such technique can be of great interest for managerial decisions, as it evaluates the effect of suspected influential features on a specific customer decision, such as churn (Geiler, Affeldt, & Nadif, 2022). The variance analysis was also widely used in marketing and business areas to uncover customer behavior (Maxham, 2001; Mittal & Kamakura, 2001; Zeithaml, Berry, & Parasuraman, 1996). Financial and retail services also rely on classical T-test and Chi square statistics to forecast customer behavior and perceptions (Hitt & Frei, 2002; Mittal & Lassar, 1998). The churn prediction problem has one important issue of class imbalance (Kong, Kowalczyk, Menzel, & Bäck, 2020) that might cause biased towards the negative samples which might hinder training the machine learning models (Zhu, Baesens, & vanden Broucke, 2017). Typically, this problem occurs when the classes in a given dataset are unequally distributed between the minority and majority classes that is low number of "churners" than "non churners". Without considering this problem, effective learning process by classification algorithms will be a challenge, since the main goal is the detection of minority classes (Dwiyanti et al., 2016; Sun, Wong, & Kamel, 2009). The popular algorithms like k-nearest neighbors ($k$-NN) is also applied in the churn-like data however studies (Dubey & Pudi, 2013; Tan, 2005) have shown several significant drawbacks. In the context of class imbalance issues in churn prediction problem, Naive Bayes classifier also appeared to be sensitive due to the strong bias in the prior estimation (Bermejo, Gámez, & Puerta, 2011). However, Huang et al. (2012) demonstrated reasonable results using Naive Bayes method.

Earlier studies have provided for various customer churn models they have analyzed the model based on customer behavior data and used different data mining techniques (Moayer & Gardner, 2012; Naz, Shoaib, & Shahzad Sarfraz, 2018; Pushpa, 2012). In these studies, all churn prediction models were analyzed and models with the best results were presented. There are various approaches for that for example Lazarov and Capota (Lazarov & Capota, 2007) showed that a model based on the customer's lifetime value analysis is the best way to predict customer churn. Similarly Naz et al. (2018) and Bandara, Perera, and Alahakoon (2013) analyzed model based on a dataset they used and showed that a big dataset with more features causes model training and evaluation difficult. Hence, this research suggested focusing on feature selection to reduce the number of features. In terms of machine learning models the study showed that for true churn rate and false churn rate, SVM should be used and in case of churn probability, logistic regression should have been used. Similarly Ahmed and Linen (2017) proposed that using hybrid models are useful and accurate for churn prediction.

The user churn prediction is also studied from the network science perspective. Recently the studies (Ahmad, Jafar, & Aljoumaa, 2019; Huang et al., 2015; Mitrović & De Weerdt, 2020; Xu et al., 2021; Zhang, Zeng, Zhao, Jin, & Li, 2022) showed the effect of social influence on user churn. The techniques to approach this problem is categorized from two perspective. The first one is to model the network structure as a surrogate of social influence. For instance, Ahmad et al. (2019) used social network analysis to extract network-based features for machine learning model. Similarly, Yang, Shi, Jie, and Han (2018) extracted network features to cluster users in different communities and predict customer churn with a deep learning model. The second one is to model the sequential order of churn as a diffusion process and use propagation models such as inflection and stopping rule (Ji et al., 2021) and spreading propagation activation (Dasgupta et al., 2008) to simulate the diffusion process and give predictions. However, the main caveat of this method is that these approaches failed to capture the causal nature of social influence. There is also a graph-based semi-supervised effort to predict the customer-churn in telecommunication (Benczúr, Csalogány, Lukács, & Siklósi, 2007). Liu et al. (2018) propose a novel graph-based inductive semi-supervised embedding model that jointly learns the prediction function and the embedding function for user-game interaction to predict the user churn from the games.

Recent studies begin to investigate how to use causal information to build better deep learning models (Bonner & Vasile, 2018; Yoon, Jordon, & Van Der Schaar, 2018). It includes the applications to eliminate the bias between the observed data and the application scenarios and learning the causal effects to give more accurate churn predictions (Johansson, Shalit, & Sontag, 2016). The studies by Umayaparvathi and Iyakutti (2017) demonstrated that deep learning models have similar performance to conventional classifiers such as support vector machine and random forest. The transfer learning which is very popular in image classification has also been employed in the customer churn prediction (Ahmed et al., 2019). Similarly, Seymen, Dogan, and Hiziroglu (2020) proposed a novel deep learning model which is compared to logistic regression and artificial neural network models. In a similar note, Momin, Bohra, and Raut (2020) demonstrated that deep Learning enables multi-stage models to represent the data at multiple abstraction levels which reduces the time and effort of feature selection considerably as it automatically creates useful features for accurate customer churn prediction. In spite of the popularity, the deep learning models can still be considered as a black box because of the complicated architecture and there is a little visibility into its decision rationale (Colbrook, Antun, & Hansen, 2022). Furthermore, it is also ambitious to recognize problems in a machine learning model or otherwise find improvements for it if the model's behavior cannot be understood (Adadi & Berrada, 2018). EXplainable Artificial Intelligence (XAI) (Emmert-Streib, Yli-Harja, & Dehmer, 2020) is a research area that studies how to make models transparent and explainable. In terms of black box models such as random forest and artificial neural network they require the application of XAI techniques to explain the model recommendation (Leung et al., 2021).

From the above listed studies, we observed that customer churn has investigated a wide range of algorithms from white box to black box models. They have good abilities to differentiate between "churn" and "no churn" customers. However, previous studies have not primarily focus on explaining churn prediction model. Therefore, successfully discriminating between these two categories is not only the aspect that is utmost importance. For customer churn prediction, understanding of the model and its outputs is important as well to target incentives to customers who have a high risk of churning and inducing them to stay. Thus in this work, we exploit the power of XAI to uncover **local** and **global** explanation of churn prediction. In particular, these explanations will enable the understanding of machine learning reasoning for the domain expert for customer churn prediction. From global explanation, one can learn about the most important pattern learned by the machine learning model for churn prediction about training population. It helps to understand the interaction between the confounding predictors. From local explanation, it enables the reasoning that the model applied to a particular case to answer the very specific questions such as "Why customer Alex churned?" and "Why has Jane continued to subscribe the plan".

## 3. Solution approach

The overall solution of our approach is illustrated in Fig. 1. The main aim of this study is to assess machine learning classifiers to predict the customer churn and provide local and global explainability for those predictions. In the next section, we explained our methodology of our approach.

## 4. Methods

The figure above depicts the methodology of the proposed model approach for the churn prediction. The step wise working of methodology is described as below:

- **Dataset:** The input to the model is the Telecommunication dataset in any tabular format. The dataset used in the paper is from Kaggle. The dataset consists of missing data which requires cleaning. For this, the dataset is passed to data preparation and preprocessing steps.
- **Selection Criteria:** The Telecommunication dataset consist of data of both churners and non-churners. Some of field might consist of missing values as well. Such data should be handled before the data are fed into the model. Thus, in this steps missing values are drop.
- **Feature Engineering and feature selection:** The raw datasets need to be handled before fetching to the classifiers. The input datasets consists of duplicate columns and unique value columns as well. Such data does not provide any significance in the churn prediction and thus these columns are drop.
- **Encoding:** The Telecommunication dataset consist of both numeric as well as categorical data. However, all of the machine level models do not work with categorical data. Thus, numeric conversion of data need to be done before application of ML models. For handling of such categorical data one-hot encoding technique is implemented in the model. This led to the increment in the column of the dataset.
- **Hyper parameter selection:** the optimization of hyperparameters across diverse machine learning models deployed for predicting customer churn within the telecommunications sector. These models are characterized by a multitude of hyperparameters, each necessitating precise calibration to enhance model efficacy.
- **Training Models:** In our methodology, we have incorporated a suite of state-of-the-art classification algorithms to ensure robust and accurate modeling. This includes the utilization of the SVM (Cortes & Vapnik, 1995), known for its effectiveness in high-dimensional spaces, and LR (Hosmer et al., 2013), a staple for binary classification problems. Additionally, we have leveraged the Random Forest Classifier (Breiman, 2001), which excels in handling large datasets with numerous features. The GBM (Friedman, 2001) has been selected for its prowess in predictive accuracy by combining multiple weak prediction models into a strong one. Lastly, Neural Networks (Goodfellow et al., 2016) have been implemented for their unparalleled capacity to learn from complex data patterns through layers of interconnected nodes, making our approach comprehensive and powerful.

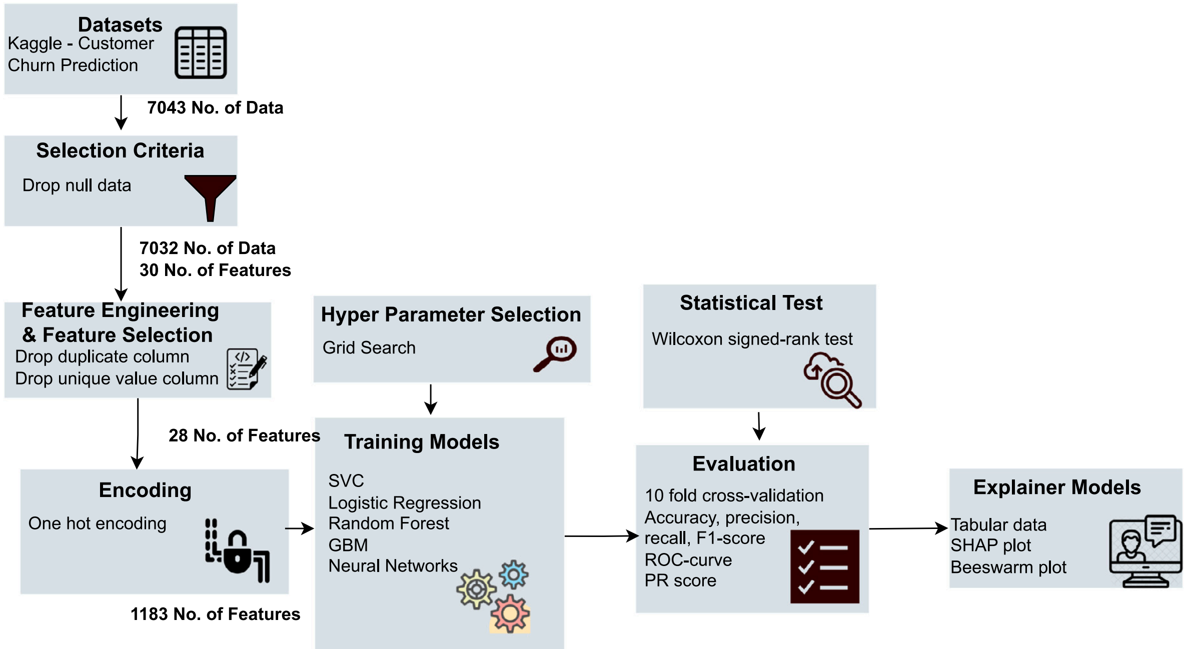


**Fig. 1.** An illustration of the data processing, model training, evaluation and explainer models on the customer churn data.

- **Statistical Test:** For a fair assessment of whether any model's predictive ability is statistically better or not in comparison to others Wilcoxon signed-rank test is implemented in the model. This provide a rigorous validation for model selection.
- **Evaluation Strategy:** To rigorously evaluate the performance of the classification models, we used 10-fold cross-validation strategy. This technique involves partitioning the original dataset into 10 equal-sized subsets. In each fold of the validation process, nine subsets are used to train the model, and the remaining one subset is used to test the model. This cycle is repeated 10 times, with each of the 10 subsets serving as the test set exactly once. By employing 10-fold cross-validation, we aim to achieve a more accurate and generalized understanding of the model's predictive power and ensure that classifier's performance is not dependent on a particular random split of the data. This is considered a robust method for assessing the generalizability of the model to an independent dataset.
- **Explainer models:** To elucidate the complexities of the telecommunication dataset within our study, we have integrated explainer models that substantially improve data visualization. Our approach incorporates SHAP (Lundberg & Lee, 2017b) plots for a macro-level analysis, providing global explanations of feature influences on the predictive model, alongside scatter plots for micro-level insights into individual customer behaviors. This bifurcated visualization strategy enables a comprehensive understanding of the dataset, facilitating the identification of systemic and case-specific factors influencing customer churn. Consequently, these methods enhance the interpretability of our model and strengthen the predictive accuracy regarding the key determinants of churn in the telecommunications domain.

**Table 1**
Summary of Attributes used in the Dataset.

| S.N | Attributes | Data type |
|---|---|---|
| 1 | CustomerID | object |
| 2 | Count | int64 |
| 3 | Country | object |
| 4 | State | object |
| 5 | City | object |
| 6 | Zip_Code | int64 |
| 7 | Lat_Long | object |
| 8 | Latitude | float64 |
| 9 | Longitude | float64 |
| 10 | Gender | object |
| 11 | Senior_Citizen | object |
| 12 | Partner | object |
| 13 | Dependents | object |
| 14 | Tenure_Months | int64 |
| 15 | Phone_Service | object |
| 16 | Multiple_Lines | object |
| 17 | Internet_Service | object |
| 18 | Online_Security | object |
| 19 | Online_Backup | object |
| 20 | Device_Protection | object |
| 21 | Tech_Support | object |
| 22 | Streaming TV | object |
| 23 | Streaming Movies | object |
| 24 | Contract | object |
| 25 | Paperless_Billing | object |
| 26 | Payment Method | object |
| 27 | Monthly Charges | float64 |
| 28 | Total_Charges | object |
| 29 | CLTV | int64 |
| 30 | Churn_Label | object |

## 5. Feature engineering and data processing

Table 1, demonstrates the feature we used in our study. We prepared the dataset for analysis, ensuring a robust foundation for our predictive models. The dataset comprises an array of features spanning customer demographic information, account details, and service subscriptions, each playing a crucial role in understanding customer behavior and predicting churn.

In our churn prediction model, we placed significant emphasis on feature engineering to enhance the model's ability to predict customer churn accurately. Among the additional features we created, the engagement score. This composite score is derived from Tenure_Months, Monthly Charges, and Total Charges, offering customer engagement and loyalty over time. By integrating these elements, the engagement score provides a multifaceted understanding of how deeply and satisfactorily customers are connected to the services offered. Another critical feature we introduced is service utilization, which quantifies the

**Table 2**
Model hyperparameters.

| Model | Hyperparameter tuning range | Hyperparameter |
|---|---|---|
| SVC | 0.001641949 – 464.0812108 | C |
| Logistic Regression | 5.15E−05 - 4534347.358 | C |
| Random Forest | 9 – 20<br>14–20 | max-depth<br>n-estimators |
| GBM | 5–29<br>5–10<br>auto<br>3–7 | max-depth<br>min-samples-leaf<br>max-features<br>max-leaf-nodes |
| Neural networks | 5–9<br>relu<br>adam | hidden-layer-sizes<br>activation<br>solver |
| AdaBoost | 50 – 500<br>0.01 – 1.0 | n-estimators<br>learning-rate |
| XGBoost | 100 - 1000<br>0.01 – 0.3<br>3 – 10 | n-estimators<br>learning-rate<br>max-depth |

total number of services a customer utilizes, including Phone_Service, Multiple_Lines, Internet_Service, among others. This feature reflects the depth of product penetration and serves as an indicator of potential customer satisfaction. A higher service utilization often suggests that customers find value in a wider range of services, potentially increasing their loyalty and decreasing their likelihood of churn.

## 6. Results

### 6.1. Model hyperparameter tuning

Table 2 illustrates the optimization of hyperparameters across diverse machine learning models deployed for predicting customer churn within the telecommunications sector. These models are characterized by a multitude of hyperparameters, each necessitating precise calibration to enhance model efficacy. Detailed in the table are the ranges of hyperparameter tuning, alongside the specific hyperparameters selected for each model, underscoring their pivotal role in refining model performance.

### 6.2. Experiments

Table 3 presents the specifications of the dataset employed in our studies. The use of detailed telecommunication data poses substantial challenges, primarily due to rigorous privacy regulations and proprietary limitations, which significantly hinder external analytical endeavors and innovative developments. Kaggle[1] enhances these constraints by providing anonymized datasets, thereby ensuring adherence to privacy standards while simultaneously facilitating the extraction of valuable analytical insights. The platform's dynamic community further promotes a culture of collaboration and knowledge exchange, catalyzing the development of novel solutions for intricate sector-specific issues such as churn prediction. Consequently, our study leverages this publicly accessible data to train our models and derive predictive insights from this dataset.

To assess the performance of the state-of-the-art classifier model, we utilized a comprehensive set of evaluation metrics, including Accuracy, Precision, Recall, F1-score, Receiver Operating Characteristic (ROC) curve, and Precision–Recall (PR) score. Table 4 summarizes the performance metrics of various machine learning models used for churn prediction tasks. Each evaluation metric is accompanied by a mean value and a standard deviation ±, indicating the variability of the

**Table 3**
Summary of the Dataset.

| Description | Dataset |
|---|---|
| Number of samples | 7043 |
| Number of features | 30 |
| % of positive samples (Churn) | 26.54% |
| %of negative samples (Non-Churn) | 73.46% |
| Data source | Kaggle |

model's performance. The models are ranked by their Accuracy, with GBM showing the highest Accuracy of 0.81 ± 0.02 and Neural Networks the lowest at 0.74 ± 0.06. The ROC-score follows a similar trend, with GBM having the highest score. The PR-score is also highest for GBM, suggesting its superior performance across various aspects of churn prediction tasks in this evaluation. The data presented in the table reveals that the GBM model exhibits superior performance compared to other models.

We have used the Wilcoxon signed-rank test is used to determine if there is a significant difference in the predictive power of GBM compared to each of the other models when applied to the same churn prediction task. This allows for a fair assessment of whether GBM's predictive ability is statistically better or not, providing a rigorous validation for model selection.

The test results showcased in the Table 5 demonstrate that GBM significantly outperform several other supervised machine learning models in the context of churn prediction for this specific dataset. AdaBoost, with a $p$-value of 0.05, indicates that its difference in performance compared to GBM is on the threshold of statistical significance, suggesting a competitive but slightly less effective model than GBM in this context. XGBoost's $p$-value of 0.07, slightly above the conventional threshold for statistical significance, suggests that while it may offer strong predictive capabilities, it does not statistically outperform GBM to a significant degree in this dataset. Both Neural Networks and Logistic Regression, with p-values well below the 0.05 threshold, demonstrate a statistically significant difference in performance compared to GBM, indicating GBM's superior capabilities in churn prediction. The SVC's performance, with a $p$-value marginally above the threshold, and Random Forest, with a higher $p$-value, suggest a less significant difference compared to GBM, underscoring GBM's robustness and effectiveness as a churn prediction tool. This comprehensive comparison underscores the importance of selecting the right model based on the dataset's specific characteristics and the predictive task at hand. While GBM shows strong performance, the nuanced differences between models highlight the potential benefits of model ensemble approaches or further hyperparameter tuning to optimize predictive accuracy.

To further understand the effectiveness of the GBM, we utilized a confusion matrix to examine its predictive accuracy and identify the areas where the model may be making errors.

Table 6 presents the confusion matrix for the GBM model, a key tool in our churn prediction analysis. The matrix indicates that the model is highly effective at identifying customers who will remain with the service, as evidenced by the 466 true negatives. However, it also points to a notable challenge in the form of 84 false negatives, which represent customers who were predicted to stay but actually churned. While the model successfully identified 103 actual churners (true positives), it incorrectly flagged 51 loyal customers as likely to churn (false positives), suggesting a need for refinement. The GBM model's strong suit is its ability to recognize stable customers, a vital aspect of preserving a customer base and avoiding the costs associated with unwarranted retention incentives. Yet, its tendency to overlook some churners could lead to substantial customer loss if not addressed. Improving the model's sensitivity, to capture more true churn cases, and its precision, to reduce the mistaken identification of loyal customers as churners, emerges as a critical focus for advancing its utility in practical

[1] https://www.kaggle.com/

**Table 4**
Results of the 10 Fold cross validation of supervised machine learning classification model for churn prediction. The figure behind ± is the standard deviation.

| Models | Accuracy | Precision | Recall | F1-score | ROC-score | PR-score |
|---|---|---|---|---|---|---|
| Neural networks | 0.74 ± 0.06 | 0.58 ± 0.26 | 0.43 ± 0.31 | 0.41 ± 0.21 | 0.83 ± 0.02 | 0.64 ± 0.03 |
| SVC | 0.78 ± 0.01 | 0.68 ± 0.03 | 0.34 ± 0.02 | 0.45 ± 0.02 | 0.77 ± 0.02 | 0.57 ± 0.04 |
| Logistic Regression | 0.79 ± 0.02 | 0.64 ± 0.04 | 0.47 ± 0.06 | 0.54 ± 0.05 | 0.81 ± 0.03 | 0.61 ± 0.04 |
| AdaBoost | 0.79 ± 0.01 | 0.65 ± 0.02 | 0.50 ± 0.06 | 0.57 ± 0.03 | 0.82 ± 0.01 | 0.63 ± 0.02 |
| XGBoost | 0.80 ± 0.03 | 0.68 ± 0.01 | 0.55 ± 0.02 | 0.61 ± 0.03 | 0.85 ± 0.02 | 0.67 ± 0.02 |
| Random Forest | 0.80 ± 0.02 | **0.71 ± 0.04** | 0.43 ± 0.08 | 0.53 ± 0.07 | 0.84 ± 0.01 | 0.64 ± 0.03 |
| **GBM** | **0.81 ± 0.02** | 0.67 ± 0.04 | **0.55 ± 0.03** | **0.60 ± 0.02** | **0.86 ± 0.01** | **0.68 ± 0.03** |

**Table 5**
Wilcoxon signed rank test.

| Models | Statistics | pvalue |
|---|---|---|
| Neural Networks | 0.0 | 0.03125 |
| SVC | 1.0 | 0.0625 |
| Logistic Regression | 0.0 | 0.03125 |
| AdaBoost | 2.0 | 0.05 |
| XGBoost | 1.5 | 0.07 |
| Random Forest | 3.0 | 0.15625 |

**Table 6**
Confusion Matrix analysis for GBM.

| | | Prediction outcome | |
|---|---|---|---|
| | | Non-churners | Churners |
| Actual value | Non-churners | 466 | 51 |
| | Churners | 84 | 103 |

business scenarios. These enhancements are imperative for tailoring customer retention strategies more effectively and securing a healthier churn rate, thereby improving the business's financial performance and customer satisfaction.

Qualitative Benchmark with Other State-Of-The-Art Models: In Table 7, we introduce an innovative approach to customer churn prediction, leveraging Gradient Boosting Machines (GBM) to analyze the Kaggle customer churn prediction dataset. Our methodology achieved a ROC-Score of 0.86, positioning it competitively among state-of-the-art methods in churn prediction for the telecommunications industry. Notably, Ebrah et al.'s use of SVM on both the IBM Watson dataset and the cell2cell dataset resulted in ROC-Scores of 0.83 and 0.99, respectively, indicating a high benchmark for model performance in varied contexts (Ebrah & Elnasir, 2019). Similarly, Shrestha et al. demonstrated the efficacy of XGBoost in achieving a ROC-Score of 0.98 with data from a Telecom service provider in Nepal (Shrestha & Shakya, 2022), while Saha et al. utilized CNN and ANN models to reach a ROC-Score of 0.99 across datasets from both Southeast Asian and American telecom markets (Saha et al., 2023). These findings underscore the significant advancements in churn prediction methodologies, with SVM, XGBoost, CNN, and ANN models setting high standards for accuracy and reliability. Our GBM-based approach contributes to this evolving landscape by not only achieving a commendable ROC-Score but also by emphasizing the adaptability and effectiveness of GBM models in handling the complexities of customer churn prediction. This comparative analysis highlights our model's potential in bridging the gap between traditional machine learning techniques and the demands of modern-day churn prediction challenges.

### *6.2.1. Selection of most important predictors*

Fig. 2 presents a beeswarm plot generated using SHAP values, which delineates the influence of various features on the GBM model's churn predictions. The plot reveals that the 'Contract_Month-to-month', 'Tenure_Months', and 'Monthly Charges' features exert the most substantial impact on the model's output, with the 'Contract_Month-to-month' feature, in particular, strongly pushing predictions towards churn. A gradation from blue to red denotes the range of feature values, with red signifying higher values. The horizontal spread of the dots reflects the magnitude of each feature's SHAP value; points to the right of the central vertical line indicate a feature's propensity to increase the likelihood of churn, while points to the left suggest a decrease. Notably, features such as 'Internet_Service_Fiber optic' and 'Payment_Method_Electronic check' predominantly contribute positively to churn predictions, whereas features like 'Online_Security_No', 'Dependents_Yes', and 'Tech_Support_No' display a mixture of positive and negative effects on the model's predictions. In the next section, we have demonstrated the top two ranked features 'Contract_Month-to-month', 'Tenure_Months' by the GBM and its interaction with the other features in the data.

### *6.2.2. Interaction between the churn predictors*

Fig. 3 visualizes the relationship between month-to-month contracts and the provision of fiber optic internet service in the context of customer churn. The red dots represent customers who have churned (discontinued their service), and the blue dots represent those who have not churned (continued their service). The $x$-axis differentiates customers based on their contract type, with a particular focus on month-to-month contracts. The $y$-axis measures some standardized metric related to churn, possibly a probability or a churn score. From the plot, we can observe a higher density of red dots at the higher end of the month-to-month contract axis, indicating that customers with month-to-month contracts and fiber optic internet service are more likely to churn. Conversely, there are more blue dots concentrated towards the lower end of the axis, suggesting that customers without fiber optic service or with longer contract terms are less likely to churn. This implies an interaction where the likelihood of churn is amplified for customers who have fiber optic service on a month-to-month basis compared to those without such service or with more extended contracts.

Fig. 4 illustrates the relationship between customer tenure, measured in months on the $x$-axis, and the amount they are charged monthly, represented by the color intensity of the dots, with magenta indicating higher charges and blue indicating lower charges. The $y$-axis shows a standardized value metric, which might represent customer satisfaction or likelihood of churn. The pattern suggests that customers with shorter tenure and higher monthly charges (magenta dots) experience a more substantial negative impact on the standardized value metric, which could indicate lower satisfaction or higher churn risk. As tenure increases, the density of magenta dots diminishes, particularly beyond the 20-month mark, suggesting that customers with higher monthly charges either improve in their standardized value metric or possibly churn out of the service, leaving behind those more satisfied or less sensitive to the charge amount. The convergence of magenta and blue dots as tenure increases indicates that the impact of monthly charges on the standardized metric decreases over time. Customers with longer tenure, irrespective of their monthly charges, show similar values of the standardized metric, which could imply that the initial sensitivity to pricing diminishes, or that the remaining customer base has adapted to or accepted the monthly charges.

**Table 7**
Performance comparison of various models on telecom customer churn prediction, highlighting our GBM approach.

| Reference | Dataset | Evaluation metric | Model |
|---|---|---|---|
| Yabas, Cankaya, and Ince (2012) | Orange Telecom | ROC-Score (0.653) | Random Forest |
| Ebrah and Elnasir (2019) | IBM Watson dataset | ROC-Score (0.83) | SVM |
| Ebrah and Elnasir (2019) | cell2cell | ROC-Score (0.99) | SVM |
| Shrestha and Shakya (2022) | Telecom service providerof Nepal | ROC-Score (0.98) | XGBoost |
| Saha et al. (2023) | Southeast Asian telecom industry, and American telecom market. | ROC-Score (0.99) in bothdataset | CNN and ANN |
| Our approach | Kaggle customer churn prediction | ROC-Score (0.86) | GBM |

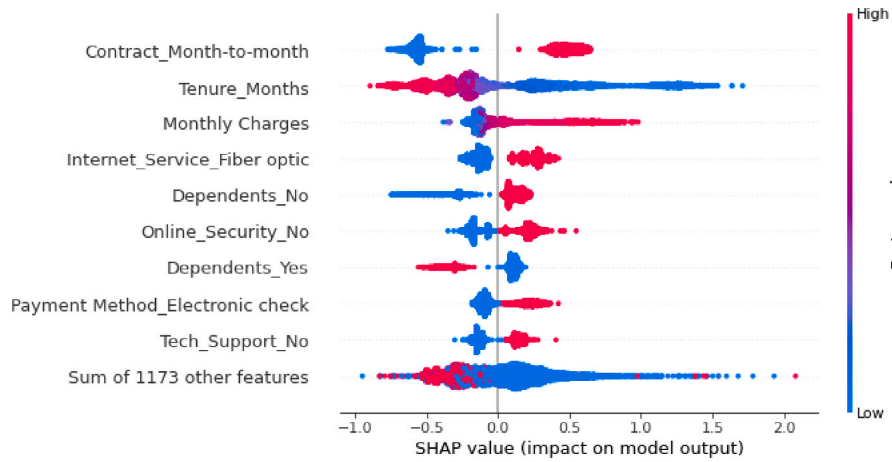


**Fig. 2.** Beeswarm plot for GBM model.

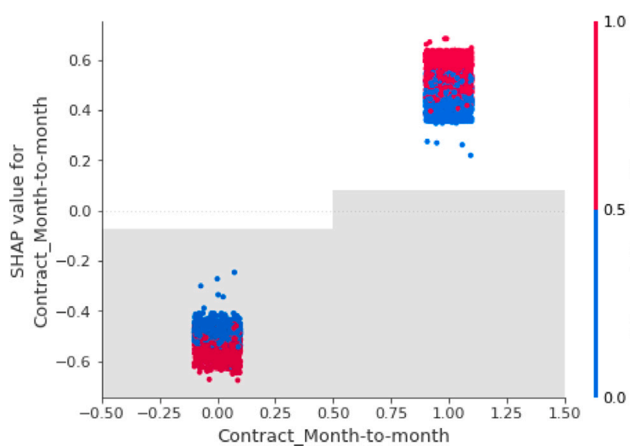


**Fig. 3.** Feature Interaction of Contract-Month-to-month.

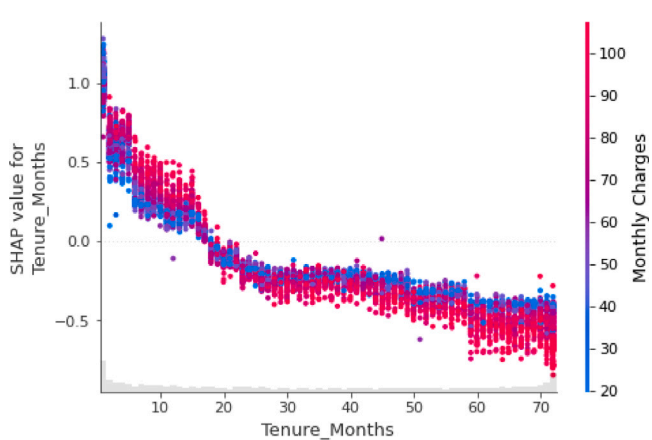


**Fig. 4.** Feature Interaction of Tenure-Months.

## 7. Discussion

The telecommunications industry is at the forefront of customer-centric strategies, where understanding and mitigating churn is not just beneficial but essential for sustaining growth and profitability. In our study, we sought to identify a machine learning model that not only excels in churn prediction but also offers clear insights into the reasons behind customer turnover. GBM model emerged as the front-runner in our analyses, substantiated by a rigorous statistical comparison using the Wilcoxon signed-rank test. The test revealed that GBM significantly outperforms Neural Networks and Logistic Regression in predicting churn, with p-values indicating the improbability of such results being due to chance.

What sets our approach apart is the incorporation of SHAP, which provided both global and local interpretability of the GBM model's predictions. Globally, SHAP values allowed us to rank features by their importance and to understand the overall direction and strength of each feature's impact on churn prediction. For instance, features like month-to-month contracts, tenure, and monthly charges were identified as key drivers of churn. Customers with short-term contracts or higher monthly charges were predisposed to churn, implying that long-term contracts and competitive pricing could be effective retention strategies.

Locally, SHAP offered insights into individual predictions, explaining why specific customers were likely to churn according to the model. This level of detail is crucial for customer relationship management, as it allows for personalized intervention strategies. For example, a customer predicted to churn due to high monthly charges could be offered a discount or a bundle package as an incentive to stay. The GBM model's ability to reveal complex interactions between features was another advantage. Through SHAP interaction values, we observed how the impact of one feature on churn could change in the presence of another feature. For example, the negative effect of a month-to-month contract on customer retention was exacerbated when combined with fiber optic internet service, suggesting that customers with this combination of services were particularly churn-prone.

The combination of GBM and SHAP explanations thus provided a powerful tool for telecom operators. Not only could they accurately predict which customers were at risk of churning, but they could also understand the underlying factors contributing to these predictions. This understanding facilitates the development of targeted strategies to retain specific customer segments, enhancing the efficiency of marketing efforts and potentially improving customer satisfaction. Incorporating these insights into business operations could lead to more nuanced customer segmentation and more effective churn prevention initiatives. For instance, identifying at-risk customers based on their usage patterns and service preferences enables the deployment of tailored communication strategies and personalized offers, thereby fostering customer engagement and loyalty.

Our work's core contribution lies in enhancing the interpretability of machine learning (ML) models for customer churn prediction, particularly through the use of SHapley Additive exPlanations (SHAP) values. The creation of unique features before data classification indeed presents a valuable avenue for research; however, it poses substantial challenges, including the need for deep domain expertise, limitations posed by data availability and quality, the balance between model complexity and interpretability, and the risk of overfitting. Our study focuses on leveraging existing, well-understood features and enriching the analysis with detailed interpretability to provide actionable insights. This approach not only aids telecom providers in identifying and addressing churn risks but also maintains the model's generalizability and robustness, carefully navigating the complexities inherent in feature engineering.

## 8. Conclusion

In the telecom sector, accurately predicting which customers are likely to leave the service is crucial. The ability to identify at-risk customers early on allows companies to intervene with targeted retention strategies. Machine learning models, particularly those that handle tabular data, are key to making these predictions. These models analyze customer data and can effectively forecast who might churn. This predictive power is essential for reducing churn rates, which is a persistent problem for telecom providers. Our research found that the GBM model was especially effective in this data. To confirm GBM's performance, we compared it with other advanced models using the Wilcoxon signed-rank test. The test results showed that GBM was significantly better at predicting churn. The $p$-value from the test helped us understand the strength of this evidence. A lower $p$-value indicates a more definitive difference between the models, and in our case, GBM's lower $p$-value confirmed its superior predictive ability. similarly, we leveraged the SHAP (SHapley Additive exPlanations) values to gain insights into the importance of different features in our predictive model. This information is invaluable for telecom companies looking to pinpoint the factors that most influence customer churn. By utilizing SHAP values, we were able to identify which specific customer attributes, such as call duration, plan type, or contract length, had the most significant impact on the churn prediction. These insights helped telecom providers tailor their retention efforts towards addressing the key factors driving customer attrition. SHAP values provided a transparent and interpretable way to analyze the model's decision-making process, making it a valuable tool for optimizing customer retention strategies in the telecommunications sector.

## Funding

This work receive no funding.

## Ethical approval

All data used in this work is freely available online. No other aspect of this work cause ethical issues.

## CRediT authorship contribution statement

**Sumana Sharma Poudel:** Conducted experiments, Analysed the results, Prepared the original draft. **Suresh Pokharel:** Revised the original draft. **Mohan Timilsina:** Provided the guidance, Revised the manuscript.

## Declaration of competing interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

## Data availability

Data will be made available on request.

## Acknowledgments

We would like to thank Data Science Institute, Insight Center for Data Analytics, at University of Galway Ireland for providing us constructive feedback and improvement of the manuscript.

## References

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access*, *6*, 52138–52160.

Ahmad, A. K., Jafar, A., & Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, *6*(1), 1–24.

Ahmed, U., Khan, A., Khan, S. H., Basit, A., Haq, I. U., & Lee, Y. S. (2019). Transfer learning and meta classification based deep churn prediction system for telecom industry. arXiv preprint arXiv:1901.06091.

Ahmed, A., & Linen, D. M. (2017). A review and analysis of churn prediction methods for customer retention in telecom industries. In *2017 4th international conference on advanced computing and communication systems* (pp. 1–7). IEEE.

Ascarza, E. (2018). Retention futility: Targeting high-risk customers might be ineffective. *Journal of Marketing Research*, *55*(1), 80–98.

Ascarza, E., Iyengar, R., & Schleicher, M. (2016). The perils of proactive churn prevention using plan recommendations: Evidence from a field experiment. *Journal of Marketing Research*, *53*(1), 46–60.

Au, W.-H., Chan, K. C., & Yao, X. (2003). A novel evolutionary data mining algorithm with applications to churn prediction. *IEEE Transactions on Evolutionary Computation*, *7*(6), 532–545.

Bandara, W., Perera, A., & Alahakoon, D. (2013). Churn prediction methodologies in the telecommunications sector: A survey. In *2013 international conference on advances in ICT for emerging regions* (pp. 172–176). IEEE.

Benczúr, A. A., Csalogány, K., Lukács, L., & Siklósi, D. (2007). Semi-supervised learning: A comparative study for web spam and telephone user churn. In *In graph labeling workshop in conjunction with ECML/pKDD*. Citeseer.

Bermejo, P., Gámez, J. A., & Puerta, J. M. (2011). Improving the performance of Naive Bayes multinomial in e-mail foldering by introducing distribution-based balance of datasets. *Expert Systems with Applications*, *38*(3), 2072–2080.

Bhattacharya, C. (1998). When customers are members: Customer retention in paid membership contexts. *Journal of the Academy of Marketing Science*, *26*(1), 31–44.

Bin, L., Peiji, S., & Juan, L. (2007). Customer churn prediction based on the decision tree in personal handyphone system service. In *2007 international conference on service systems and service management* (pp. 1–5). IEEE.

Bolton, R. N., & Bronkhorst, T. M. (1995). The relationship between customer complaints to the firm and subsequent exit behavior. *ACR North American Advances*.

Bonner, S., & Vasile, F. (2018). Causal embeddings for recommendation. In *Proceedings of the 12th ACM conference on recommender systems* (pp. 104–112).

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Colbrook, M. J., Antun, V., & Hansen, A. C. (2022). The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem. *Proceedings of the National Academy of Sciences*, *119*(12), Article e2107151119.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*, 273–297.

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.

Dasgupta, K., Singh, R., Viswanathan, B., Chakraborty, D., Mukherjea, S., Nanavati, A. A., et al. (2008). Social ties and their relevance to churn in mobile telecom networks. In *Proceedings of the 11th international conference on extending database technology: advances in database technology* (pp. 668–677).

Dubey, H., & Pudi, V. (2013). Class based weighted k-nearest neighbor over imbalance dataset. In *Pacific-Asia conference on knowledge discovery and data mining* (pp. 305–316). Springer.

Dwiyanti, E., Ardiyanti, A., et al. (2016). Handling imbalanced data in churn prediction using rusboost and feature selection (case study: Pt. telekomunikasi Indonesia regional 7). In *International conference on soft computing and data mining* (pp. 376–385). Springer.

Ebrah, K., & Elnasir, S. (2019). Churn prediction using machine learning and recommendations plans for telecoms. *Journal of Computer and Communications*, *7*(11), 3. http://dx.doi.org/10.4236/jcc.2019.711003.

Emmert-Streib, F., Yli-Harja, O., & Dehmer, M. (2020). Explainable artificial intelligence and machine learning: A reality rooted perspective. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(6), Article e1368.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189–1232.

Fujo, S. W., Subramanian, S., Khder, M. A., et al. (2022). Customer churn prediction in telecommunication industry using deep learning. *Information Sciences Letters*, *11*(1), 24.

Geiler, L., Affeldt, S., & Nadif, M. (2022). A survey on machine learning methods for churn prediction. *International Journal of Data Science and Analytics*, 1–26.

Gerpott, T. J., Rams, W., & Schindler, A. (2001). Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market. *Telecommunications Policy*, *25*(4), 249–269.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

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.

Hawley, D. (2003). International wireless churn management: research and recommendations. Yankee Group report, (June), URL http://www.ams.com/cme/pdfs/yankeechurnstudy.pdf. (Accessed January 2006).

Hitt, L. M., & Frei, F. X. (2002). Do better customers utilize electronic distribution channels? The case of PC banking. *Management Science*, *48*(6), 732–748.

Hosmer, D. W., Jr., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression: vol. 398*, John Wiley & Sons.

Huang, B., Kechadi, M. T., & Buckley, B. (2012). Customer churn prediction in telecommunications. *Expert Systems with Applications*, *39*(1), 1414–1425.

Huang, Y., Zhu, F., Yuan, M., Deng, K., Li, Y., Ni, B., et al. (2015). Telco churn prediction with big data. In *Proceedings of the 2015 ACM SIGMOD international conference on management of data* (pp. 607–618).

Ji, H., Zhu, J., Wang, X., Shi, C., Wang, B., Tan, X., et al. (2021). Who you would like to share with? a study of share recommendation in social e-commerce. In *Proceedings of the AAAI conference on artificial intelligence, vol. 35, no. 1* (pp. 232–239).

Johansson, F., Shalit, U., & Sontag, D. (2016). Learning representations for counterfactual inference. In *International conference on machine learning* (pp. 3020–3029). PMLR.

Kong, J., Kowalczyk, W., Menzel, S., & Bäck, T. (2020). Improving imbalanced classification by anomaly detection. In *International conference on parallel problem solving from nature* (pp. 512–523). Springer.

Lazarov, V., & Capota, M. (2007). Churn prediction. *Business Analysis Course. TUM Computer Science*, *33*, 34.

Leung, C. K., Pazdor, A. G., & Souza, J. (2021). Explainable artificial intelligence for data science on customer churn. In *2021 IEEE 8th international conference on data science and advanced analytics* (pp. 1–10). IEEE.

Liao, C.-H., & Lien, C.-Y. (2012). Measuring the technology gap of APEC integrated telecommunications operators. *Telecommunications Policy*, *36*(10–11), 989–996.

Liu, X., Xie, M., Wen, X., Chen, R., Ge, Y., Duffield, N., et al. (2018). A semi-supervised and inductive embedding model for churn prediction of large-scale mobile games. In *2018 ieee international conference on data mining* (pp. 277–286). IEEE.

Lundberg, S. M., & Lee, S.-I. (2017a). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*.

Lundberg, S. M., & Lee, S.-I. (2017b). A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems 30* (pp. 4765–4774). Curran Associates, Inc..

Maxham, J. G., III (2001). Service recovery's influence on consumer satisfaction, positive word-of-mouth, and purchase intentions. *Journal of Business Research*, *54*(1), 11–24.

Miguéis, V. L., Van den Poel, D., Camanho, A. S., & e Cunha, J. F. (2012). Modeling partial customer churn: On the value of first product-category purchase sequences. *Expert Systems with Applications*, *39*(12), 11250–11256.

Mitrović, S., & De Weerdt, J. (2020). Churn modeling with probabilistic meta paths-based representation learning. *Information Processing & Management*, *57*(2), Article 102052.

Mittal, V., & Kamakura, W. A. (2001). Satisfaction, repurchase intent, and repurchase behavior: Investigating the moderating effect of customer characteristics. *Journal of Marketing Research*, *38*(1), 131–142.

Mittal, B., & Lassar, W. M. (1998). Why do customers switch? The dynamics of satisfaction versus loyalty. *Journal of Services Marketing*, *12*(3), 177–194.

Moayer, S., & Gardner, S. (2012). Integration of data mining within a strategic knowledge management framework. *International Journal of Advanced Computer Science and Applications*, *3*(8).

Momin, S., Bohra, T., & Raut, P. (2020). Prediction of customer churn using machine learning. In *EAI international conference on big data innovation for sustainable cognitive computing* (pp. 203–212). Springer.

Naz, N. A., Shoaib, U., & Shahzad Sarfraz, M. (2018). A review on customer churn prediction data mining modeling techniques. *Indian Journal of Science and Technology*, *11*(27), 1–27.

Nguyen, N., & LeBlanc, G. (1998). The mediating role of corporate image on customers' retention decisions: an investigation in financial services. *International Journal of Bank Marketing*.

Pushpa, S. (2012). An efficient method of building the telecom social network for churn prediction. *International Journal of Data Mining & Knowled Management Process*, *2*(3), 31–39.

Qureshi, S. A., Rehman, A. S., Qamar, A. M., Kamal, A., & Rehman, A. (2013). Telecommunication subscribers' churn prediction model using machine learning. In *Eighth international conference on digital information management* (pp. 131–136). IEEE.

Reichheld, F. F., & Sasser, W. E. (1990). Zero defeofions: Quoliiy comes to services. *Harvard Business Review*, *68*(5), 105–111.

Reinartz, W. J., & Kumar, V. (2003). The impact of customer relationship characteristics on profitable lifetime duration. *Journal of Marketing*, *67*(1), 77–99.

Saha, L., et al. (2023). Deep churn prediction method for telecommunication industry. *Sustainability*, *15*(5), 4543.

Seymen, O. F., Dogan, O., & Hiziroglu, A. (2020). Customer churn prediction using deep learning. In *International conference on soft computing and pattern recognition* (pp. 520–529). Springer.

Shrestha, S. M., & Shakya, A. (2022). A customer churn prediction model using XGBoost for the telecommunication industry in Nepal. *Procedia Computer Science*, *215*, 652–661.

Sun, Y., Wong, A. K., & Kamel, M. S. (2009). Classification of imbalanced data: A review. *International Journal of Pattern Recognition and Artificial Intelligence*, *23*(04), 687–719.

Tan, S. (2005). Neighbor-weighted k-nearest neighbor for unbalanced text corpus. *Expert Systems with Applications*, *28*(4), 667–671.

Umayaparvathi, V., & Iyakutti, K. (2017). Automated feature selection and churn prediction using deep learning models. *International Research Journal of Engineering and Technology (IRJET)*, *4*(3), 1846–1854.

Van den Poel, D., & Lariviere, B. (2004). Customer attrition analysis for financial services using proportional hazard models. *European Journal of Operational Research*, *157*(1), 196–217.

Varki, S., & Colgate, M. (2001). The role of price perceptions in an integrated model of behavioral intentions. *Journal of Service Research*, *3*(3), 232–240.

Wei, C.-P., & Chiu, I.-T. (2002). Turning telecommunications call details to churn prediction: a data mining approach. *Expert Systems with Applications*, *23*(2), 103–112.

Xu, F., Zhang, G., Yuan, Y., Huang, H., Yang, D., Jin, D., et al. (2021). Understanding the invitation acceptance in agent-initiated social e-commerce. In *Proceedings of the international AAAI conference on web and social media, vol. 15* (pp. 820–829).

Yabas, U., Cankaya, H. C., & Ince, T. (2012). Customer churn prediction for telecom services. In *2012 IEEE 36th annual computer software and applications conference* (pp. 358–359). http://dx.doi.org/10.1109/COMPSAC.2012.54.

Yang, Z., & Peterson, R. T. (2004). Customer perceived value, satisfaction, and loyalty: The role of switching costs. *Psychology & Marketing*, *21*(10), 799–822.

Yang, C., Shi, X., Jie, L., & Han, J. (2018). I know you'll be back: Interpretable new user clustering and churn prediction on a mobile social application. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 914–922).

Yoon, J., Jordon, J., & Van Der Schaar, M. (2018). GANITE: Estimation of individualized treatment effects using generative adversarial nets. In *International conference on learning representations*.

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1996). The behavioral consequences of service quality. *Journal of Marketing*, *60*(2), 31–46.

Zhang, G., Zeng, J., Zhao, Z., Jin, D., & Li, Y. (2022). A counterfactual modeling framework for churn prediction. In *Proceedings of the fifteenth ACM international conference on web search and data mining* (pp. 1424–1432).

Zhao, L., Gao, Q., Dong, X., Dong, A., & Dong, X. (2017). K-local maximum margin feature extraction algorithm for churn prediction in telecom. *Cluster Computing*, *20*, 1401–1409.

Zhu, B., Baesens, B., & vanden Broucke, S. K. (2017). An empirical comparison of techniques for the class imbalance problem in churn prediction. *Information Sciences*, *408*, 84–99.