Churn Prediction using Various Machine Learning Algorithms in Telecom Sector

Vikrant Verma
Student,
Department of Computer Science and Engineering
BIT Raipur, India

Padmavati Shrivastava, PhD
Associate professor,
Department of Computer Science and Engineering
BIT Raipur, India

Dhananjay Sahu
Student
Department of Computer Science and Engineering
BIT Raipur, India

ABSTRACT
The world around us is growing at a very fast rate as a result of this rapid expansion of digital systems and related information technologies, there is a growing trend in the global economy to develop digital CRM systems. One of the pro competitors in this field is the telecommunications sector, where companies are increasingly digitalized. Telecom companies emphasize better and improved CRM's and one of main problem Telecom companies face is Customer churning. Customer churning is the discontinuation of services by customer side. This happens very commonly in telecom sectors like in India people switch from Idea to Jio then from Jio to Airtel, Customer gets service from whoever gives better offer. Due to this, service providers have recognized the significance of retaining current customers. Therefore, providers are compelled to exert additional effort in predicting and preventing churn. Hence From existing churn customers data, business analysts and Customer Relationship Management (CRM) analysts must determine the reasons for churning customers and their behavioural patterns. This study conducts a real-world investigation of customer churn prediction and proposes the use of different classification algorithms and feature selection to create a customer churn prediction model. In this study the researchers have used a fictional dataset which contains real life features on the basis of which customer might churn. They performed an analysis on the WA_Fn-UseC_-Telco-Customer-Churn dataset, which is a telecom customer churn dataset that is available on Kaggle. The purpose of this research was to identify patterns concerning the factors that lead to customer churn. This dataset contains features like Customer ID, gender, tenure, PhoneService, InternetService, PaymentMethod, MonthlyCharges, TotalChargesetc and Churn (Whether the customer churned or not (Yes or No)).

Keywords
CRM, churning

1. INTRODUCTION
In the world we live in today, telecommunications companies are producing an enormous amount of data at an astonishingly quick rate. There is a wide variety of telecom service providers contending with one another in the market in order to increase their percentage of the customer base. Customers now have access to a variety of options, including services that are both better and more affordable. The ultimate objective of businesses in the telecommunications industry is to maximize their profits while simultaneously ensuring their continued viability in a highly competitive market [1]. When a significant proportion of a telecom company's customers feel dissatisfied with the quality of the services they receive, this is known as customer churn. Because of this, customers end up moving their business to competing service providers as a result of the disruption. There are a lot of different reasons to churn. Prepaid customers, as opposed to postpaid customers, are not bound to a particular service provider and are free to switch at any time. Churning has a negative impact not only on a company's overall reputation but also on its brand, which ultimately suffers as a result. It is critical in the telecommunications industry to accurately predict customer churn in order to better serve and retain existing customers through the use of Customer Relationship Management (CRM) software. CRM's toughest task is keeping its current clientele. In today's world average churn rate for the four major carriers in telecom sector is about 1.9% per month, however it can reach 67% for prepaid services annually. Hence fostering customer loyalty is crucial as getting new consumers is 25 times more expensive than retaining them [3]. Customers can choose to use another service provider because the market is so crowded and competitive. As attracting new customers is more expensive than keeping the ones you already have, telecom companies have devised ways to identify and retain their current customers.

The researchers propose a churn prediction model in this study that employs a variety of machine learning algorithms. A classifier's performance is determined by the dataset available. It is validated with a theoretical Telco Churn Dataset. Using information retrieval metrics, the proposed churn prediction model is evaluated. For measuring the accuracy of proposed churn prediction modelthey have used the following parameters: True positive rate (TP rate), False Positive Rate (FP rate), Recall, Precision and F1-measure. This study's goal is to investigate existing machine learning and data mining techniques and use it to create a model for predicting customer churn, identifying their churning factors, and providing retention strategies. This paper starts by giving the introduction about the project followed by Literature review in which researchers of this paper state the findings from various research/reference papers. After that they have explained the main methodology of the project. Thanthey have showcased the observations and results of the compared models and the conclusion drawn from them. Further they highlight the future scope of the project.

2. RELATED WORK
Customer Churn prediction is a pretty popular research area, and in recent years there has been a tremendous amount of research done on Telecom customer churn prediction still. However, due to the variations in customer behavior that this problem presents, the implementation of a proper and accurate customer churn prediction system is an extremely rare sight.

In the course of this review of the relevant literature, The researchers analyzed several methods that can be applied in the process of putting into action a binary classification model that estimates the likelihood of customers leaving based on an examination of their patterns of behavior. The purpose of this study is to investigate all of the various possible ways in which machine learning techniques can be implemented to improve the automated Customer Churn prediction system. This is the researchers goal in conducting this study.

Irfan Ullah et al. developed a Churn Prediction Model Using the Random Forest Algorithm in [5]. They used two datasets to study the customer churn prediction problem. Their first dataset which they used was from a South Asian GSM telecom service provider. That dataset has 64,107 instances and 29 features, all of those features are in numerical form. The information is derived from the customer service usage pattern Call Detail Record (CDR). It contains labeled data in two classes, in which 30% of that data contains true customers labeled as "T" which represents represent churners and 70% of that data has false customers labeled as "F" to represent non-churners. They ran that data through 12 different algorithms, with the Random Forest algorithm producing the best results. On that dataset, 88.63 percent accuracy was obtained. They used performance evaluation matrices like Accuracy, False Positive Rate, True Positive rate, *Precision*, Recall, and F-measure.

Hemlata Jain et al. published "Churn Prediction in Telecommunication In that paper, he used Logistic Regression and Logit Boost" [4]. To solve this problem, they used the Logit Boost and Logistic Regression algorithms. To train their model, they used "Orange, an American telecommunications company database." There are 3333 instances in the database. Many performance measures were used to evaluate this model, including Root Relative Squared Error, Mean Absolute Error, Root Mean Square Error, Kappa Statistics, Relative Absolute Error, Case Coverage (0.95), and Mean Rel. Region Size (0.95), and Accuracy. These performance metrics revealed that both techniques outperformed. The results of both techniques were nearly identical. Logistic regression was 85.2385 percent accurate, while Logit Boost was 85.1785 percent accurate.

Boosting is a general and very effective method for improving any learning algorithm. Boosting algorithms uses the concept of iterative learning and adding weak classifiers to make a final strong classifier. Each weak classifier that is added is weighted and trained with reweighted data. Ning Lu et al. [6] used this technique to make an efficient churn prediction model, They collected data from a local telecom database Then they performed variable selection on that dataset and For each customer group, a logistic regression model is trained for predicting the chances that a customer might churn in future. Customers that has a greater predicted likelihood have a greater churn propensity. A churn prediction system's ability to identify churners for marketing purposes should be measured [18], so they used the Receiver operating characteristic also called (ROC) curve and top-quantile-lift prediction model. They compared their results with a logistic regression model without customer separation.

In [7] the researchers have conducted a survey of all ways in which people have tried to solve this problem in their own way. In that paper, they have mentioned that factors like Customer demography and personal details, Customer care service details, Bill and payment details, Customer credit score, Customer usage pattern, Customer’s value-added services, etc should be considered to make a churn prediction model. They even mentioned some datasets like “PAKDD 2006”, “KDD Cup 2009 small”, “Cell2Cell”, and “CrowdAnalytix” which are ideal for doing this research. They also provided a comprehensive table regarding datasets, features, models, and metrics used by various churn prediction systems. The survey paper turned out to be really helpful for this research.

Similarly in [12] Miss.Priyanka Parmar and Mrs. Shilpa Serasiya used XGBoost and Logistic regression algorithm on IBM’s theoretical Telecom Churn dataset and achieved 78.89% accuracy in Logistic Regression and 80.5% accuracy in XGBoost algorithm and By reading this paper the researchers decided to carry out their work on this theoretical dataset and use algorithms like logistic regression ANN , Random Forest , ADA Boost etc and get better accuracy.

Further they did more research on this same “Telco Customer churn” dataset by IBM and found out that “NyashadzasheTamuka” and “KhulumaniSibanda”[13] used the same dataset and trained their model using Logistic regression, Random forest and Decision tree model and got really decent accuracy 97% in logistic regression, 79% in Random forest and 78.3% in Decision Tree. Here the literature review was concluded and work on the project was initiated on the chosen dataset “Telco Customer churn dataset”.

3. DATASET USED

Telco customer churn (WA_Fn-UseC_-Telco-Customer-Churn.csv) is an open source dataset available on Kaggle. The file name is WA Fn-UseC -Telco-Customer-Churn.csv. The Telco customer churn data includes information about a made-up telco company that, during the third quarter of the year, offered home phone and Internet services to 7043 customers in the state of California. It shows which customers have left, which customers have stayed, and which customers have signed up for their service. Each customer's Satisfaction Score, Churn Score, and Customer Lifetime Value (CLTV) index are included, in addition to a number of important demographics that are included for the customer.

Each row in the table represents a different customer, and each column in the table contains information about some aspect of that customer. The raw data has a total of 7043 rows, which represent the customers, and 21 columns (features). The "Churn" column will serve as the objective.

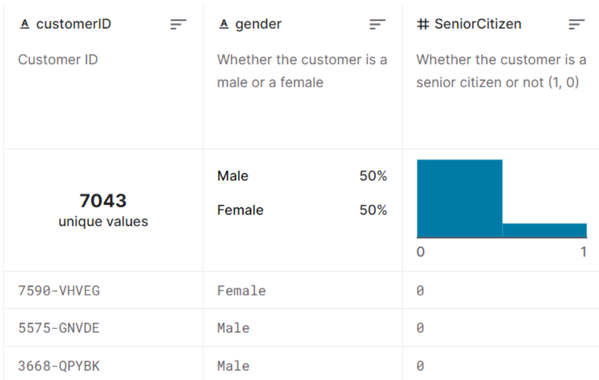

Fig 1: Snapshot of Telco customer churn dataset.

The dataset contains following features :

Table 1. Table captions should be placed above the table

S.no.	Feature	Description
1	Customer ID	Id of customer.
2	Gender	This property tells if the customer is male or female.
3	Senior citizen	This property tells if the customer is a senior citizen or not (1, 0).
4	Partner	This property tells if the customer has any partner or not (Yes or No).
5	Dependents	This property tells if the customer has any dependents or not (Yes or No).
6	Tenure	The number of months that the consumer has been a loyal customer to the business.
7	Phone service	This property tells if the customer has a phone service or not (Yes or No).
8	Multiple lines	This property tells if the customer has multiple lines or not (Yes, No or No phone service).
9	Internet service	This property tells the Customer's internet service type (DSL, Fiber optic or No).
10	Online becurity	This property tells if the customer has online security or not (Yes, No or No internet service).
11	Online backup	This property tells if the customer has online backup or not (Yes, No or No internet service).
12	Device protection	This property tells if the customer has device protection or not (Yes, No or No internet service).
13	Tech support	This property tells if the customer has tech support or not (Yes, No or No internet service).
14	Streaming TV	This property tells if the customer has streaming TV or not (Yes, No or No internet service).
15	Streaming movies	This property tells if the customer has streaming movies or not (Yes, No or No internet service).
16	Contract	This property tells about the contract term of the customer (Month-to-month or One year or Two year).
17	Paperless billing	This property tells if the customer has paperless billing or not (Yes or No).
18	Payment method	This property tells the payment method used by the customer like (Mailed check, Electronic check, Credit card (automatic) or Bank transfer (automatic)).
19	Monthly charges	The amount charged to the customer on monthly basis.
20	Total charges	The total amount charged to the customer.
21	Churn	This property tells if the customer has churned or not (Yes or No).

As you see the dataset contains all kinds of features. It contains demographic features like CustomerID, Gender, SeniorCitizen , Dependents , Partner. Service based features like TotalCharges, MonthlyCharges, PaymentMethod, Contract , StreamingMovies and TV etc.

All the values on the rows that describe the customer's status are either in String or Numeric form which will be further converted to fit it in a machine learning model.

4. PROPOSED EXPERIMENTATION

In this section, various methods have been presented that has been experimented with and implemented that has yielded some result, The researchers used the Telco churn dataset and feed it into different Binary classification algorithms and compared their accuracies. The methodology is divided into four parts :

- Data Exploration and analysis
- Data pre-processing
- Model training
- Model evaluation

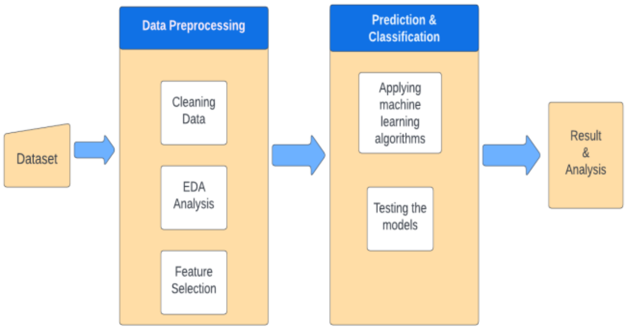

Fig 2: Methodology flowchart.

4.1 Data Exploration and analysis

In this step exploration on the dataset has been done researchers looked for null values , removed and plotted informatory graphs , They plotted graphs to view the relation between different categorical features and churning of customer on the basis of those features.

Have a look at the below figures given in the next page get a good understanding.

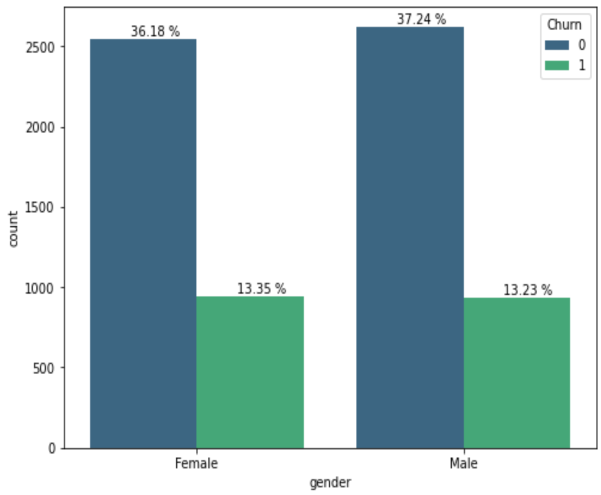

Fig 3: churning customers based on Gender distribution

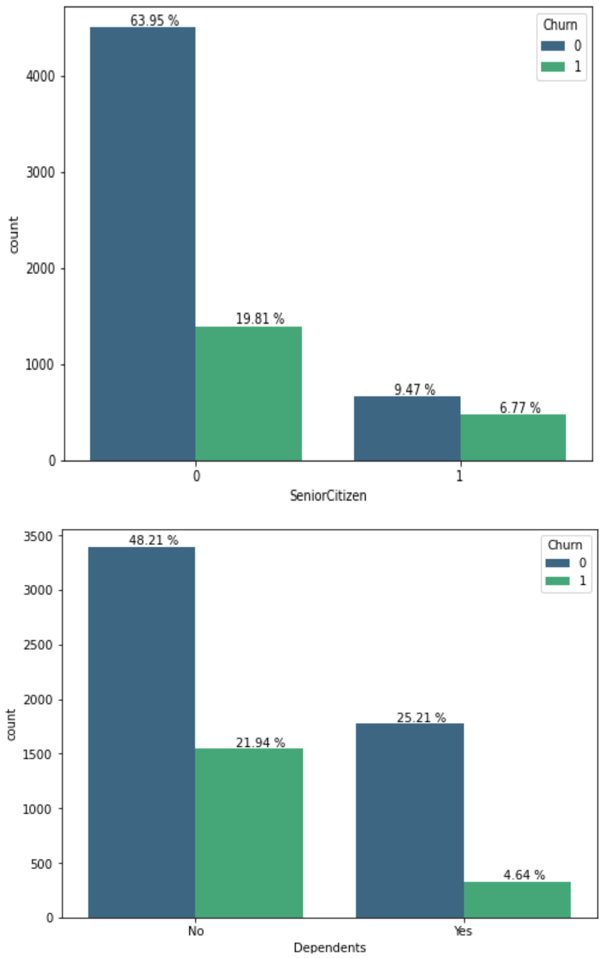

Fig 4: churning customers based on Senior Citizen and dependents distribution.

As you can see in the above figures that based on gender 13.3% Female and 13.23% Male out of total 7014 customers have churned similarly 6.77% senior citizens and 4.64% people with dependents have discontinued their service. In the below figure you can see distribution based on contract which says that most of the churner take service on month-month basis.

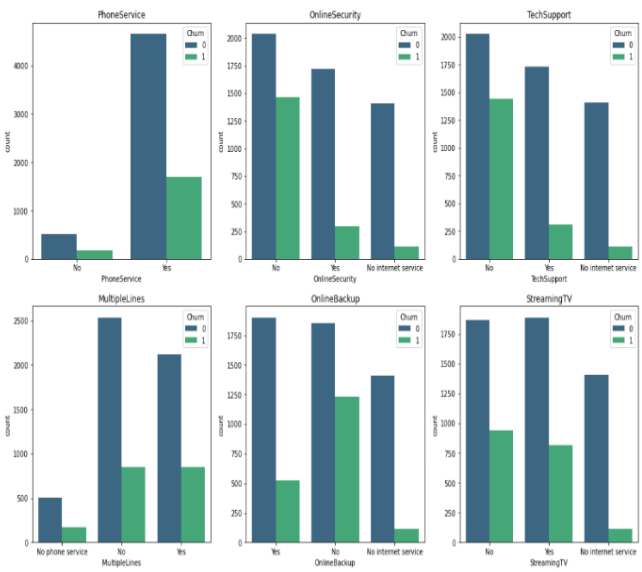

Fig 5: churning customers based on services.

4.2 Data pre-processing

A dataset is a collection of data points that a computer can analyse and forecast as a single entity. This implies that the gathered data must be made standardised and machine-readable, since machines do not see data in the same manner as people. In order to do this, it is necessary to preprocess the data by cleaning and finishing it, and to annotate the data by adding machine-readable tags that contain useful information.

For preprocessing of data identification and removal of null values and false values has been done. Then all the yes and no values of categorical features are transformed into 1 and 0 and for those categorical features which has more than yes or no as value(e.g.multipleLines, etc)these features were further divided and split into more categories providing them with binary values. The use ofMinMaxScalar function was also done to scale down the values which are greater than 1.

4.3 Model Training

For model training thestandard procedure was followed the use of scikit-learn library was done for implementing Logistic regression, Random Forest, Ada boost, Support Vector Machine and Naive Byaes. Keras library was used to implement Neural network andxgboost library to make XGBoost classifier.

The procedure was pretty standard :

- Import the classifier and make the model object.
- Feed the training data and train the model
- evaluate the model and note the evaluation report
- make a confusion matrix to get a better view of results on the test dataset.

The same steps were taken to make the classifier model using all algorithms except for the neural network. The neural network was built by importing the Sequential() object from TensorFlow's Keras library Then addition of 3 dense layers was done followed by the addition of 3 dropout layers with a dropout rate of 0.5 and one output layer was done the resultant model performs the classification. Two neural networks were made and only the units were different in the two neural networks. The first neural network units were 28 x 14 x 14 and after doing feature selection it was reduced to 18 x 9 x 3 (layer 1 x layer 2 x layer 3).

4.4 Model Evaluation

Evaluating anything is putting it through its paces with data that you haven't seen before and determining whether or not it can achieve its goals. accuracy, precision, recall, f-measure, and the confusion matrix are the metrics that are utilized in this investigation for the purpose of assessing the proposed churn prediction model.

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

Here "TN" represents true negative, "TP" represents true positive, "FN" represents false negative and "FP" represents false positive. The true positive rate (TP rate) is also known as sensitivity. It indicates how much of the data is correctly classified as positive. The TP rate for any classifier must behigh.TP rate is calculated by using Equation 2. $\boldsymbol{TP\ Rate} = \frac{\boldsymbol{True\ Positives}}{\boldsymbol{Actual\ Positives}}$ (2)

The FP Rate indicates which parts of the data have been incorrectly classified as positive. For any classifier, the FP rate must yield a low result. It is calculated by using Equation 3.

$$\textbf{FP Rate} = \frac{\textbf{False Positives}}{\textbf{Actual Negatives}} \quad (3)$$

Accuracy, which can also be referred to as Positive Predictive Value (PPV), determines which aspect of the prediction data contains positive information. Equation 4 is used to do the calculations needed to determine it.

$$\textbf{Precision} = \frac{\textbf{True Positives}}{\textbf{True Positives} + \textbf{False Positives}} \quad (4)$$

The recall is yet another measure for completeness, and represents the algorithm's actual success rate. It is the likelihood that the algorithm will choose all of the cases that are relevant to the question. The low value of recall results in a large number of false negatives. It is determined by employing Equation 5 which you can see below.

$$\textbf{Recall} = \frac{\textbf{True Positives}}{\textbf{True Positives} + \textbf{False Negatives}} \quad (5)$$

The value of the F-measure is a compromise that must be reached between correctly classifying all of the data points and ensuring that each class contains points that belong to only that class. Equation 6 is used in the computation of this value.

$$\boldsymbol{F - measure} = \mathbf{2} \times \frac{\textbf{Precision} \times \textbf{Recall}}{\textbf{Precision} + \textbf{Recall}} \quad (6)$$

Confusion matrix is the compilation of TP, FP , TN, FN in a matrix form. All these metrics were used to evaluate the models.

5. RESULT & DISCUSSION

These are the results of the churn prediction model before doing any feature selection. The below table represents the score of prediction of customers who have churned.

Table 2. Accuracy of churn prediction of customer churning.

S. No.	Method used	TP Rate	FP Rate	Precision	Recall	f1-score	Accuracy
1.	Logistic regression	0.824	0.3596	0.85	0.91	0.88	0.80758
2.	Random Forest	0.838	0.27	0.84	0.94	0.88	0.81876
3.	Support Vector Machine (SVM)	0.853	0.306	0.85	0.91	0.88	0.82018
4.	ADA Boost	0.845	0.306	0.85	0.91	0.88	0.8159
5.	XG Boost	0.850	0.339	0.85	0.90	0.88	0.8095
6.	Naive Bayes	0.92	0.544	0.92	0.65	0.76	0.6979
7.	ANN	0.852	0.298	0.85	0.92	0.88	0.81947

In the above table you can see the classification report of all the models trained using different classification algorithms. Almost every algorithm has same accuracy but support vector machine and neural network are showing the best result. However Naïve bayes algorithm shows the best precision and TP Rate (True Positive rate).

Table 3. Accuracy of churn prediction of customer churning after feature selection

S. No.	Method used	TP Rate	FP Rate	Precision	Recall	f1-score	Accuracy
1.	Logistic regression	0.850	0.298	0.85	0.92	0.88	0.82089
2.	Random Forest	0.842	0.2911	0.84	0.93	0.88	0.81805
3.	Support Vector Machine (SVM)	0.8428	0.305	0.84	0.92	0.88	0.81449
4.	ADA Boost	0.850	0.311	0.85	0.91	0.88	0.8173
5.	XG Boost	0.8468	0.3839	0.85	0.88	0.86	0.79388
6.	Naive Bayes	0.9134	0.4822	0.91	0.74	0.82	0.7547
7.	ANN	0.8118	0.2356	0.86	0.96	0.88	0.8359

In the above table, you can see the classification report of all the models trained using different classification algorithms after applying feature selection. For features selection, some of the fewer correlation features were replaced. As you can see in the table for some algorithms feature selection improved the performance but for some, it decreased performance i.e Accuracy of the Naïve Bayes algorithm, Logistic regression, and ANN increased whereas the accuracy of XG boost and Random forest decreased.

The best accuracy which is achieved so far is in the ANN model which gave an "83.59%" accuracy, with a TP rate of 0.8118, FP Rate of 0.2356, precision 0f 0.81, and recall of 0.96, and F1- Score of 0.88.

The confusion matrix of ANN is :

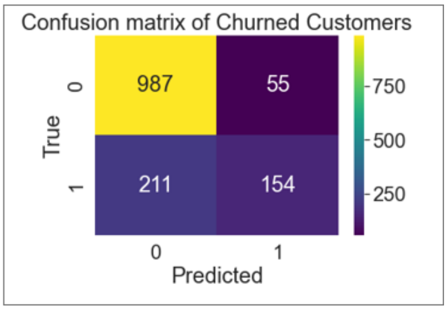

Table 4. Confusion matrix churn prediction of customer churning

S.No.	Method used	TP	FP	FN	TN
1.	Logistic regression	675	367	58	307
2.	Random Forest	975	67	188	177
3.	Support Vector Machine (SVM)	953	89	164	201
4.	ADA Boost	958	84	175	190
5.	XG Boost	939	103	165	200
6.	Naive Bayes	675	367	58	307
7.	ANN	959	83	171	194

Table 5. Confusion matrix churn prediction of customer churning after feature selection

S.No.	Method used	TP	FP	FN	TN
1.	Logistic regression	675	367	58	307
2.	Random Forest	975	67	188	177
3.	Support Vector Machine (SVM)	953	89	164	201
4.	ADA Boost	953	89	168	197
5.	XG Boost	918	124	166	199
6.	Naive Bayes	770	272	73	292
7.	ANN	959	83	171	194

6. CONCLUSION AND FUTURE SCOPE

6.1 Conclusion

An empirical investigation of the prediction of customer churn based on theoretical data sets is carried out as part of this research. In contrast to the majority of the research that has been done on churn prediction, The researchers trained their model using a variety of binary classification algorithms after studying and comparing them. All of the algorithms produced results that were quite comparable to one another, but the neural network architecture produced the best outcomes. Further implementation of feature selection and removal of non-necessary features, resulted in better classification, and an accuracy of 83 percent was achieved in this study.

6.2 Future Scope

In the future, this study plans to conduct additional research into eager learning and lazy learning strategies to improve churn prediction capabilities. Applying techniques from artificial intelligence for trend analysis and prediction can be used to further investigate the shifting behavior patterns of churned customers. This can be done by expanding the scope of the current study. In addition to this, one of the main goals of this study is to collect some data from actual life situations based on the features described above and use that data to remake the models and check their accuracy. Aside from this, The researchers would also like to work on additional different datasets for exploratory purposes, such as the Cell2Cell Dataset, PAKDD 2006, and so on, and compare the features of this dataset to those of the other datasets to see if the features in the other datasets are superior.

7. ACKNOWLEDGMENTS

Our thanks to our project guide, Dr. Padmavati Shrivastava, who guided us throughout this project. We would also like to thank our friends and classmates who supported us and offered deep insight into the study.

8. REFERENCES

[1] Babu, S., Ananthanarayanan, D. N., & Ramesh, V. (2014). A survey on factors impacting churn in telecommunication using datamininig techniques. International Journal of Engineering Research & Technology (IJERT), 3(3).

[2] Verbeke, W., Martens, D., Mues, C., &Baesens, B. (2011). Building comprehensible customer churn prediction models with advanced rule induction techniques. Expert systems with applications, 38(3), 2354-2364.

[3] Strategies for Reducing Churn Rate in the Telecom Industry. (2022). Retrieved 24 June 2022, from https://www.heavy.ai/blog/strategies-for-reducing-churn-rate-in-the-telecom-industry#:~:text=The%20average%20churn%20rate%20in,fostering%20customer%20loyalty%20is%20key

[4] Jain, H., Khunteta, A., & Srivastava, S. (2020). Churn prediction in telecommunication using logistic regression and logit boost. Procedia Computer Science, 167, 101-112.

[5] Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., & Kim, S. W. (2019). A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector. IEEE access, 7, 60134-60149.

[6] Lu, N., Lin, H., Lu, J., & Zhang, G. (2012). A customer churn prediction model in telecom industry using boosting. IEEE Transactions on Industrial Informatics, 10(2), 1659-1665.

[7] Umayaparvathi, V., &Iyakutti, K. (2016). A survey on customer churn prediction in telecom industry: Datasets, methods and metrics. International Research Journal of Engineering and Technology (IRJET), 3(04).

[8] Dalvi, P. K., Khandge, S. K., Deomore, A., Bankar, A., &Kanade, V. A. (2016, March). Analysis of customer churn prediction in telecom industry using decision trees and logistic regression. In 2016 symposium on colossal data analysis and networking (CDAN) (pp. 1-4). IEEE.

[9] Canbek, G., Sagiroglu, S., Temizel, T. T., & Baykal, N. (2017, October). Binary classification performance measures/metrics: A comprehensive visualized roadmap to gain new insights. In 2017 International Conference on Computer Science and Engineering (UBMK) (pp. 821-826). IEEE.

[10] Amin, A., Shehzad, S., Khan, C., Ali, I., & Anwar, S. (2015). Churn prediction in the telecommunication industry using a rough set approach. In New trends in computational collective intelligence (pp. 83-95). Springer, Cham.

[11] Prashanth, R., Deepak, K., &Meher, A. K. (2017, July). High accuracy predictive modeling for customer churn prediction in the telecom industry. In the International Conference on Machine Learning and Data Mining in Pattern Recognition (pp. 391-402). Springer, Cham.

[12] Parmar, M. P., &Serasiya, M. S. (2021). Telecom Churn Prediction Model using XgBoost Classifier and Logistic Regression Algorithm.

[13] Tamuka, N., & Sibanda, K. (2020, November). Real time customer churn scoring model for the telecommunications industry. In 2020 2nd International Multidisciplinary Information Technology and Engineering Conference (IMITEC) (pp. 1-9). IEEE.

[14] De Caigny, A., Coussement, K., & De Bock, K. W. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. European Journal of Operational Research, 269(2), 760-772.

[15] Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2022). Customer churn prediction system: a machine learning approach. Computing, 104(2), 271-294.

[16] Lazarov, V., &Capota, M. (2007). Churn prediction. Bus. Anal. Course. TUM Comput. Sci, 33, 34.

[17] Xie, Y., Li, X., Ngai, E. W. T., & Ying, W. (2009). Customer churn prediction using improved balanced random forests. Expert Systems with Applications, 36(3), 5445-5449.

[18] Amaresan, S. (2022). What Is Customer Churn? [Definition]. Retrieved 24 June 2022, from https://blog.hubspot.com/service/what-is-customer-churn

[19] Logistic Regression in Machine Learning - Javatpoint. (2022). Retrieved 24 June 2022, from https://www.javatpoint.com/logistic-regression-in-machine-learning#:~:text=Logistic%20regression%20is%20one%20of,of%20a%20categorical%20dependent%20variable.

[20] Random Forest Algorithm: A Complete Guide. (2022). Retrieved 24 June 2022, from https://builtin.com/data-science/random-forest-algorithm

[21] Support Vector Machine (SVM) Algorithm - Javatpoint. (2022). Retrieved 24 June 2022, from