iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IV **Month of publication:** April 2023

DOI: https://doi.org/10.22214/ijraset.2023.50646

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

Customer Churn Prediction Using Machine Learning Algorithms

Pallavi Hangarge[1], Gunjan Jadhav[2], Vaishnavi Janagave[3], Sujal Kadam[4], Prof. P.S. Pise[5]
[1, 2, 3, 4, 5]*Department of Information Technology*

Abstract: ***Customer churn is a serious problem in the telecommunications industry and occurs more often. Customer churn is the percentage of customers that stopped using your company's product or service during a certain time frame. One of the most important problems in predicting customer churn is the imbalanced data, which has been tried to be solved and compared with different methods. The machine learning algorithms will be use in this paper are Decision Tree, Support Vector Machine, Random Forest . Also, the performance of support vector were better than other algorithms.***
Keywords: Customer Churn, Machine Learning.

I. INTRODUCTION

Customer churn is the percentage of customers who give up your services or products and can happen due to competition in price and quality of service with other companies, weakness in building a strong relationship with the customer, competitive marketing for companies. Customer churn happens gradually, over time, and is not a one-time event. In other words, analyzing the customer buying pattern over time can make it possible to predict customer churn. Many companies, including telecom companies, financial service providers, and pay -TV companies, have been analyzing customer data to predict customer churn in order to stay competitive. Customer churn.maximization. Owing to heavy competition from rival companies due to the competitive rates of various providers, customers often tend to switch between them. Also, it is pretty common to hear people express frustration regarding convoluted billing, unwanted marketing emails, hard-tonavigate customer service, high plan prices, etc. All these factors combined today make the necessity of effective customer churn identification in the telecommunications industry of paramount importance.

Churners are customers who will be switching from one telecom service provider to another. Prediction of telecom churners has been an area of interest for researchers and many researchers have worked on various techniques to predict telecom customer churn. Telecom industry has been battling the threat of losing more than 25% of its customers every year, which is believed to result in huge revenue loss. Another known fact is that adding or acquiring a new customer costs between 5 to 10 times more than retaining an old customer with the company. Hence it is believed that the best marketing strategy is to retain the existing subscribers or to avoid customer churn. Over the years, there has been an increasing need to automate the process of identifying customer churn. This process has become such an expensive affair that generally only 15 percent of revenues earned by mobile companies are spent on network infrastructure and IT while 15 to 20 percent of revenues on the acquisition and retention of customers . AI has been really successful in being able to counter this problem by considerably comprehending large amounts of customer data to draw valuable inferences while reducing the workforce this job demands.

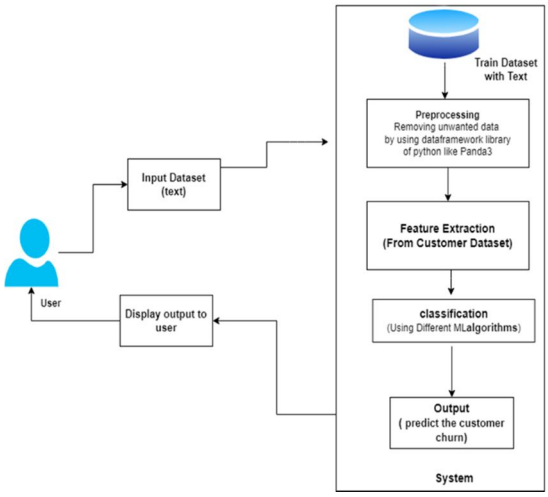
In this paper, a comparative analysis is performed on techniques such as Random Forest, SVM, Extreme Gradient Boosting (XGBoost), Ridge classifier and Neural Networks to predict various customer churn patterns. Through Random Forest, we explore the older but trusted and lightweight divide and conquer methodology making decision trees of numeral type and and using them through a random selection of attributes. Finally, a decision tree is created for classification on test data. Random forest, not only performs well on large datasets but also handles missing variables efficiently. A final decision tree is constructed for prediction of the test dataset. Random forest performs well on a large dataset and handles missing variables without deletion of variables. Support Vector Machine (SVM) models are parametric. Its accuracy and performance is greatly influenced by the initial values of its parameters.

Therefore, while tuning the parameters of SVM, a new combined evaluation metric is applied to maximize its effectiveness for churn management. Another recent prediction algorithm as a part of the ensemble method in many machine learning challenges is Extreme Gradient Boosting (XGBoost), which generally performs well with imbalanced-classes data. The XGBoost algorithm uses the exact greedy algorithm to find the best split.

©IJRASET: All Rights are Reserved

II. METHODOLOGY

In this section, the steps taken to predict customer churn are first described in general, then the actions taken in each step are reviewed. In the first step, the dataset is pre-processed. In this step, actions such as deleting rows with empty values, converting data format to processable format for algorithms, data normalization is performed and in the feature extraction step, some features of the table are extracted. Using balancing data algorithms, the data is balanced and the models begin to be trained and then perform the test phase. Finally, the performance of the models is compared based on the three balancing methods. Fig.1 shows the steps that have been taken.



1) About the Dataset The data used in this study was published on the Kaggle website[11], which is owned by a company in the telecom industry. In this dataset, there are 3333 rows and 21 columns, and the target column is called Churn. In this dataset, the customer specifications are presented, and the amount of customers' use of each of the company's services, which includes Voice Mail, SMS, and calls, is mentioned during the day and night.
Data preprocessing A dataset needs to be preprocessed before entering the models. Data preprocessing was performed in the following steps: By exploring the dataset, the null and missing values identified and Each row containing these values were deleted.

2) Feature extraction Feature selection is a process of searching for meaningful features for analysis. Using this method, it is possible to remove features that have duplicate information, and by recognizing the effective features, reduce the complexity of the data and increase the computational speed of the machine learning model. At this stage, some features have been removed.

3) Data balancing In classification models, it is assumed that the number of samples is evenly distributed in different classes. If this is not the case and the samples are imbalanced divided between the classes.

4) Output we'll predict percentage of customer churn.On basis on analysis find flaws to minimize the churn prediction.

III. ALGORITHMS

A. Support Vector Machine(SVM)

Support Vector Machine It is one of the most popularized Supervised Learningalgorithm, which is used for Classification as well as Regression problems. However,basically,itisusedforClassificationproblemsinMachineLearningscenario.Theintent of the SVM algorithm is to create the best decision boundary that can segregaten-dimensional space into classes so that it can easily put the new data point in thecorrectcategoryinthefuture.ThisbestdecisionboundaryiscalledahyperplaneofSVM.

B. Random Forest Algorithm

Random Forest classifier is a learning method that operates by constructing multipledecision trees and the final decision is made based on the majority of the trees andis chosen by the random forest.It is a tree-shaped diagram used to determine acourseofaction.Eachbranchofthetreerepresentsapossibledecision,instance,orreaction. Using of Random Forest Algorithm is one of the main advantages is thatitreducestheriskofoverfittingandtherequiredtrainingtime.Additionally,italsooffers a high level of accuracy. It runs efficiently in large databases and producesal most accurate predictions by approximating missing data.

IV. LITERATURE SURVEY

A. 10th International Conference on Computer [IEEE 2020]

Customer churn is a serious problem in the telecommunications industry and occurs more often. The cost of maintaining existing customers is much lower than attracting new customers, and the literature stated that five times the cost of maintaining existing customers have to be spent on attracting new customers. Random forest is one of the algorithms developed By above 90% accuracy.

B. 5th International Conference on Cloud Computing and Big Data Analytics [IEEE 2020]

1) Customer churn is a prominent issue facing companies. Preventing customer churn, trying to retain and retain customers has become an important issue for business operations and development. Most of the current customer churn predictions use a single prediction model, which makes it difficult to accurately predict customer churnneural network algorithm

2) The Accuracy 90%

C. International Conference on Computer Communication and Informatics [IEEE 2021]

1) In current days, the customers are getting more attracted towards the quality of service (QoS) provided by the organizations. However, the current era is evidencing higher competition in providing technologically advanced QoS to the customers. support vector machine algorithms

V. CONCLUSION

In the present competitive market of telecom domain, churn prediction is a significant issue of the CRM to retain valuable customers by identifying similar groups of customers and providing competitive offers/services to the respective groups. The obtained results show that our proposed churn model performed better by using machine learning techniques. In future, we will further investigate eager leaning and lazy learning approaches for better churn prediction. The study can be further extended to explore the changing behavior patterns of churn customers by applying ArtificialIntelligence techniques for predictions and trend analysis.

REFRENCES

[1] S. H. Dolatabadi and F.Keynia, ”Designing of customer and employee churn prediction model based on data mining method and neural predictor,” 2017 2nd International Conference on Computer and Communication Systems (IC- CCS), Krakow, 2017, pp.74-77.

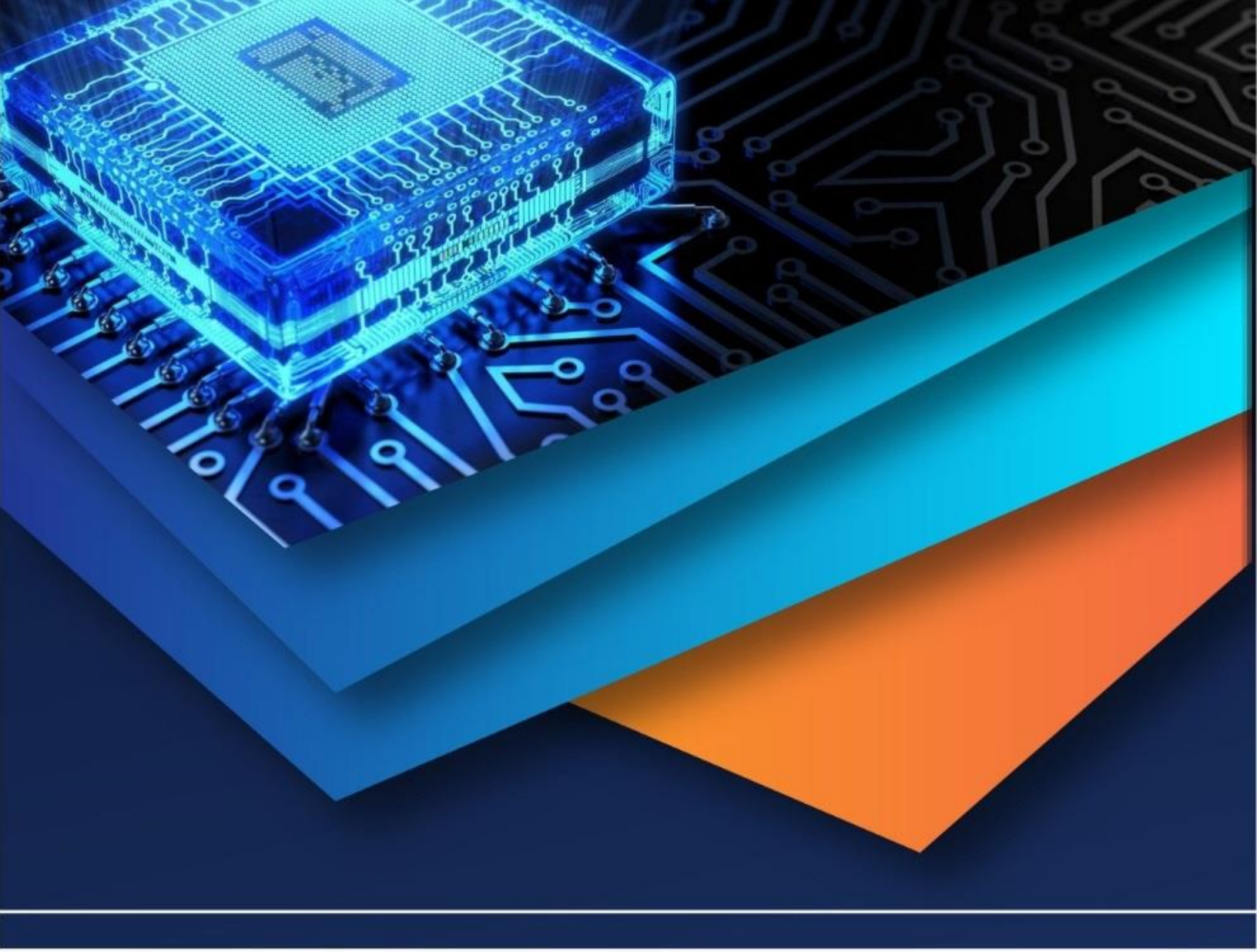
[2] A. D. Caigny, K. Coussement, and K. W. D. Bock, ” A new hybrid classifica- tion algorithm for customer churn prediction based on logistic regression and decision trees,” European Journal of Operational Re- search,vol.269(2),pp.760–772,September

[3] K. Coussement, S. Lessmann, and G Verstraeten, ”A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry,” Decision Support Systems,vol.95,pp.27- 36,March2017.

[4] K. Mishra and R. Rani, ”Churn prediction in telecommunication using ma- chine learning,” International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai,pp.2252-2257, 2017.

[5] X. Wang, K. Nguyen, and B. P. Nguyen,” Churn Prediction using Ensemble Learning. In ¡i¿Proceedings of the 4th International Conference on Machine Learning and Soft Computing,” Association for Computing Machinery, New York, NY, USA, pp.56–60, January2020.

[6] K. G. M. Karvana, S. Yazid, A. Syalim and P. Mursanto, ”Customer Churn Analysis and Prediction Using Data Mining Models in Banking Industry,” In- ternational Workshop on Big Data and Information Security (IWBIS), Bali, Indonesia, pp.33-38,2019.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)