Journal of Engineering Sciences

Vol 13, Issue 05, May /2022

ISSN NO:0377-9254

Telecom Churn Prediction Using ML

D . Mahendra Reddy[1], T . Kurdush Basha[2], M . Bharath[3], P . Jaya Teja Reddy[4]

Assistant Professor(A)[1], Student (U.G)[234], Department of Computer Science and Engineering,
Jawaharlal Nehru Technological University College of Engineering, Pulivendula,
Andhra Pradesh, India
mahendrareddy30@gmail.com [1], tkurdush2018@gmail.com [2], mopurubharath2000@gmail.com [3], jayatejareddypochimireddy@gmail.com [4]

Abstract:-- **Customers are the cornerstone of any business success, so companies are beginning to recognize the importance of achieving customer satisfaction. Customer attrition is a major issue and is recognized as one of the company's greatest concerns due to intensifying competition between businesses, the growing importance of marketing strategies, and the recent conscious behavior of customers. Organizations need to develop a variety of strategies to solve the churn problem that depends on the services they provide. The practice of customer attrition is essential in the highly competitive and rapidly evolving telecommunications industry. The process of migrating from one service provider to another Telecom service provider occurs because of the superior service or price, or the various benefits that competitors offer to customer at sign-up. Predicting customer attrition has become an integral part of the planning process and strategic decision-making in the telecommunications sector due to the high costs associated with acquiring new customers. Using machine learning techniques, we estimated the customer's churn probability . Logistic regression is widely used to estimate churn probability as a function of a set of variables or customer characteristics . This study uses the Kaggle website for datasets to predict and analyze churn. The results in the survey show that the predicted accuracy rate for consumer churn is 0.84 percent and the area under the curve is 0.82 percent.**

OBJECTIVES AND AIM

The primary and secondary objectives of the study are as follows:

Primary objectives

• To explore the customer churn prediction in telecom using machine learning

Secondary objectives

• To investigate the impact of customer churn in telecom industry as a whole

• To discuss the significance of customer churn models in telecom industry

• To compare the algorithms that are effective in reducing churn rate in telecom companies

INTRODUCTION

The telecommunications industry is becoming one of the most important sectors in the world, as a result of technological growth and the ever-evolving number of operators increasing the level of competition. Telecommunications companies are struggling to survive in this highly competitive market, and several steps have been put in place to generate huge revenues. To increase customer retention, it is important for businesses to reduce the potential for customer attrition, known as "customer migration from one service provider to another."

The basic principle of churn forecasting related to the telecommunications industry is to roughly calculate the number of subscribers who literally want to quit the company they were using and propose solutions to prevent significant churn. In recent years, in the fierce competition between companies, it has become necessary to estimate cancellation before cancellation. Due to the significant role played by the telecommunications industry, it has become increasingly important to build forecasting mechanisms along the line of churn forecasting. Retaining

customers in the telecommunications industry is already a nightmare as a result of fierce competition for services. Therefore, we proposed an advanced data mining method for finding customer attrition using machine learning algorithms, that is, SVM (Support Vector Machine), XGBOOST, and GRADIENTBOOST. The results emphasized that machine learning algorithms perform better in predicting customer attrition. The

Support Vector Machine (SVM) is a supervised learning algorithm that evaluates data and identifies patterns, especially for regression and classification analysis. Decision Tree (DT) is another machine learning application that does not have great ability to capture non-linear and complex relationships between functions. However, given the issue of customer attrition, the accuracy of decision trees based on data forms can be better.

IMPLEMENTATION

Data Collection

The dataset for this classification problem is provided by Kaggle and is sourced from IBM's sample dataset collection Use case pipeline construction begins with importing some basic libraries needed for the entire case. These include Pandas and Numpy for data processing and processing, and Matplotlib and Seaborn for visualization. In this exercise, the dataset (in .csv format) is downloaded to a local folder, loaded into a Jupyter notebook, and saved in PandasDataFrame.

Exploratory Data Analysis:

After collecting the data, several steps are taken to examine the data. The goal of this step is to understand the data structure, perform initial preprocessing, clean up the data, identify patterns and inconsistencies in the data (skewness, outliers, missing values, etc.) and make hypotheses. Is to verify.

Feature Engineering

Feature extraction begins with an initial set of measurement data, builds derived values intended to be informative and non-redundant, facilitat subsequent learning and generalization steps, and in some cases better human interpretation. be connected. Feature extraction is related to dimensionality reduction. Feature extraction reduces the number of resources required to describe large amounts of data.

Train And Split Data

Train test partitioning is a technique for assessing the performance of machine learning algorithms. It can be used for classification or regression problems and can be used for supervised learning algorithms. The procedure is to get the dataset and divide it into two subsets.

Model Selection and Evaluation

Model selection and evaluation are very important operations in a machine learning workflow. This is the workflow section for analyzing the model. Examine the more insightful statistics of its performance to determine the actions that need to be taken to improve this model. Evaluating your model gives you better insight into what the model predicts well and what it doesn't. This helps transform the model from the model that predicts the dataset.

Training and Split Data

Training the model means learning or determining the appropriate values for all weights and biases from the labeled example. Machine learning uses data to answer questions. Therefore, prediction or reasoning is a step that can answer some questions. This is the point of all this work where the value of machine learning is recognized.

MACHINE LEARNING MODELS :

Random Forest Algorithm

Random forest is a popular machine learning algorithm that belongs to the supervised learning method. "Random forests are classifiers that take a set of decision trees for different subsets of a particular dataset and then take the average to improve the predictive accuracy of that dataset."

Logistic regression

If the dependent variable is binary, logistic regression is a good regression analysis model. Logistic regression is a predictive study used to explain the relationship between an independent set of variables and dependent binary variables. Logistic regression is based on a math-oriented method for investigating the effects of variables on other variables. Predictions are made by creating a set of equations that associate input values with output fields.

Examine customer records, include regression equations, and perform a scoring process for each customer in the dataset. If the consumer's

p value is greater than the predefined value, the consumer is at risk of churn.

Support Vector Machine(SVM)

The goal of the SVM algorithm is to create optimal lines or decision boundaries that can divide n-dimensional space into classes so that new data points can be easily placed in the correct category in the future. This best decision boundary is called the hyperplane.

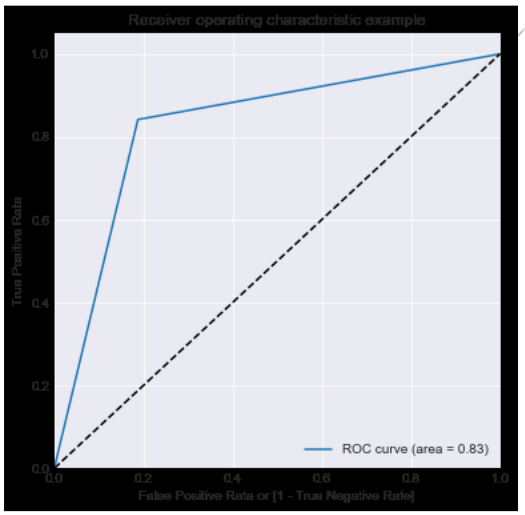
Decision Tree

Decision trees are the most powerful and popular tool for classification and prediction. The decision tree is a flow chart-like tree structure, where each internal node specifies a test for the attribute, each branch represents the result of the test, and each leaf node (terminal node) contains a class label.

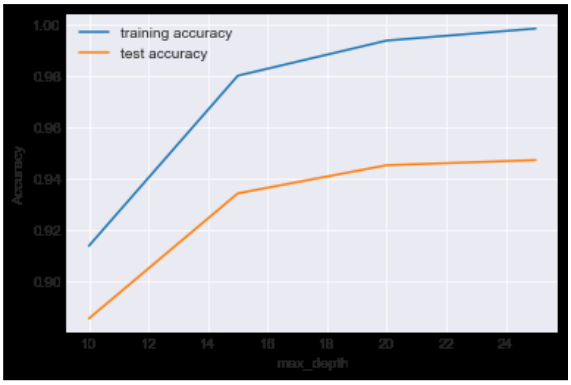
XG Boost

XGBoost stands for Extreme Gradient Boosting proposed by researchers at the University of Washington. This is a library written in C ++ that optimizes gradient boosting training.

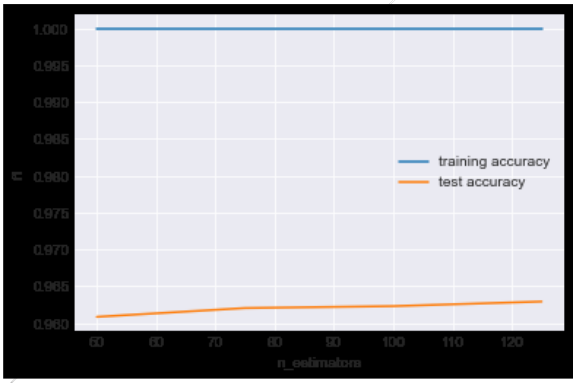
RESULT



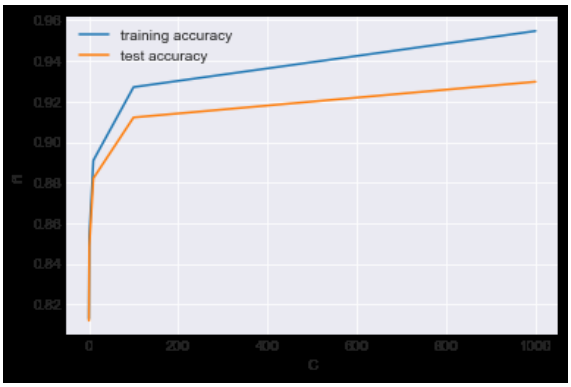
The logistic regression roc curve is lying in the top left corner which is a sign of a good fit.



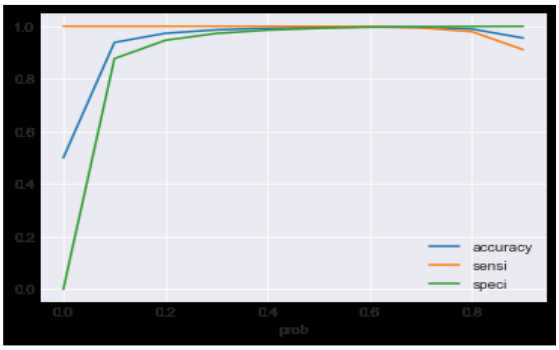
Test f1-score almost becomes constant after max_depth=20



Selecting n_estimators = 80



SVM with hyperparameter tuning



Accuracy, speed and Precision using XGBoost Clasifier

Comparison of used Machine Learning Algorithms in project based on precision, specificity, Auc score and Sensitivity/Recall is shown in the following table:

Model/Metrics	Train	Test
Logistic Regression (cut-off = 0.45)		
Roc_auc_score	82.11%	81.21%
Sensitivity/Recall	86.48%	84.40%
Specificity	77.75%	78.02%
precision	79.54%	25.04%
DecisionTree (cut-off = 0.4)		
Roc_auc_score	82.41%	76.57%
Sensitivity/Recall	89.79%	78.13%
Specificity	75.03%	75%
precision	78.24%	21.38%
Random Forest (cut-off = 0.45)		
Roc_auc_score	85.60%	96.53%
Sensitivity/Recall	88.70%	77.57%
Specificity	82.50%	81.73%
precision	83.52%	26.97%
GBC		
Roc_auc_score	96.11%	80.84%
Sensitivity/Recall	100.00%	79.87%
Specificity	92.21%	81.81%
precision	92.78%	28.52%
XGB (cut-off = 0.2)		
Roc_auc_score	97.24%	80.76%
Sensitivity/Recall	99.99%	76.13%
Specificity	94.49%	85.38%
precision	94.78%	32.13%
SVM (linear C = 1000)		
Roc_auc_score	81.33%	82.62%
Sensitivity/Recall	79.91%	78.40%
Specificity	82.75%	86.85%
precision	82.25%	35.14%

SUMMARY

The results of the survey focus on predicting the termination of customers who discontinue services in the telecommunications sector. By determining the nature of the customer, the enterprise can improve customer support and improve the overall performance of the telecommunications sector. The expected results of this study show an improvement in the classification accuracy of the proposed system. From this we can conclude that the accuracy of the logistic regression classifier is much better than SVM, Random Forest, etc.

CONCLUSION

It is a known fact that the customer acquisition cost is larger than customer retention cost that makes the retention a difficult prototype of business. There is no standard approach which resolves the churning problems of worldwide service providers of telecom industry accurately. Machine learning technique is used for customer churn which sets warning bells for customers before any damage could occur, providing telecom firms the chance to take precautionary steps. These techniques are used to find the churn in customers by constructing models and studying from historical information. Conducting trials with perspective of end users, collecting their views on network, normalization of data, data set preprocessing, using feature selection, removing missing values and class imbalance and changing existing variables with derived variables develops the churn prediction accuracy which supports the telecom sector to retain their customers much efficiently. In the highly competitive telecommunications sector, mobile communications standardization and public policy make it a highly competitive market for customers to easily switch from one provider to another. Predicting churn, the task of identifying customers who may discontinue use of the service, is a lucrative and essential topic in the telecommunications sector. Customer churn is often a serious problem for the telecommunications sector, as customers do not hesitate to leave unless they anticipate what they are looking for. Above all, customers need cost performance, competitive costs, and higher quality service. Customer attrition is directly related to customer satisfaction. It is a known fact that the customer acquisition cost is larger than customer retention cost that makes the retention a difficult prototype of business. There is no standard approach which resolves the churning problems of worldwide service providers of telecom industry accurately. Machine learning technique is used for customer churn which sets warning bells for customers before any damage could occur, providing telecom firms the chance to take precautionary steps. These techniques are used to find the churn in customers by constructing models and studying from historical information. Conducting trials with perspective of end users, collecting their views on network, normalization of data, data set preprocessing,

using feature selection, removing missing values and class imbalance and changing existing variables with derived variables develops the churn prediction accuracy which supports the telecom sector to retain their customers much efficiently. We can conclude that machine learning is an effective way to detect customer attrition.

In the telecommunications sector, it is important to identify and manage potential customers who are constantly characterized by volatile markets and fierce competition. Proper customer management with potential churn can increase the profits of the telecommunications sector while reducing the likelihood of churn. Telcos recognizes the importance of predicting customer attrition as an opportunity to generate enormous profits. Building a churn forecasting model simplifies the customer retention process, making the telecommunications sector more successful and growing in a highly competitive market.

The churn prediction model relies heavily on the process of data mining and data mining techniques due to the improved performance generated by machine learning algorithms. Customer churn is used to create models that accurately include customer hazards and customer survival capabilities. Machine learning algorithms have been used for accurate predictions in the telecommunications field. Telecom churn forecasting has emerged as a diverse area of churn forecasting applications compared to other telecom-based sectors due to its size, bias, and dataset diversity. From this, we can conclude that the customer attrition model is one of the most important solutions for staying competitive in the telecommunications sector.

In today's competitive business and digital world, customer attrition is an integral part of service providers' efforts to build profitable, long-term relationships with specific customers. Factors that influence customer switching behavior are call quality, satisfaction, rates, brand image, mobile phones, seniority, and revenue. Some customers are very cost sensitive and switch to another telecommunications provider when they get better costs and choose a service provider of their choice for friends and family. Other factors that influence customers when choosing a service provider are cost and communication, followed by service complaints. Quality is also one of the factors that allow customers to switch from one service provider to another. Call charges also play a key role in deciding to switch to another service provider, followed by network coverage, customer support, and value-added services. Family influence is another major factor in churn rates in the telecommunications sector. Factors that influence a customer's chances of entering a rivalry include inadequate or slow response to billing errors and complaints. Several other factors, such as packaging costs, poor characteristics, and older technologies, can also affect customer churn and hurt competition.

From this, we can conclude that telecommunications providers need to provide their customers with many incentives for the above factors in order to flexibly and effectively reduce churn rates.

The cost of acquiring a new customer can be higher than the cost of retaining a customer. One of the best ways to retain customers is to reduce their customer churn rates. Termination refers to the transfer of a customer from one service provider to another, or the termination of a particular service over a given time period for multiple predictable reasons. If you use machine learning techniques that allow your company to inspect data records and find customers who may cancel. There are several algorithms available to reduce the churn rate of telecommunications companies. Telecom service providers use sophisticated analysis algorithms to screen large amounts of customer data. This algorithm is smart enough to detect hidden characteristics and figure out which customers are most likely to churn. Data mining plays a key role in telecommunications companies, and efforts to reduce overall churn develop good marketing strategies, detect fraud and consumers, and better manage their networks. .. The appropriate algorithm is selected based on the nature of the problem and the feasible data. From this we can conclude that machine learning algorithms are considered one of the best solutions for the telecommunications sector to reduce churn rates.

REERENCES

1. Adwan O, Faris H, Jaradat K, Harfoushi O, Ghatasheh N (2014) Predicting customer churn in telecom industry using multilayer preceptron neural networks: Modeling and analysis. Life Science Journal11(3):75–81

2. Umayaparvathi V, Iyakutti K (2016) A survey on customer churn prediction in telecom industry:Datasets, methods and metrics. International Research Journal of Engineering and Technology (IRJET)4(4):1065–1070

3. Hadden J, Tiwari A, Roy R, Ruta D (2006) Churn prediction: Does technology matter. International Journal of Intelligent Technology 1(2):104–110

4. Coussement K, De Bock KW (2013) Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning. Journal of Business Research 66(9):1629–1636

5. Kirui, C., Hong, L., Cheruiyot, W., Kirui, H.: Predicting customer churn in mobile telephony industryusing probabilistic classifiers in data mining. International Journal of Computer Science Issues (IJCSI)10(2 Part 1), 165 (2013).

6. Adebiyi, S.O., Oyatoye, E.O., Amole, B.B. (2016). Relevant Drivers for Customers` Churn and Retention Decision in the Nigerian Mobile Telecommunication Industry. Journal of Competitiveness, 8(3), pp. 52-67.

7. Ahmad, A.M., Jafar, A, Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. Journal of Big Data, 6(28), pp. 1-24

8. Ahna, J.H., Hana, S.P., Lee, Y.S. (2006). Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry", Telecommunications Policy, 30(10– 11), pp. 552–568.

9. Almana, A.M., Aksoy, M.S., Alzahrani, R. (2014). A Survey on Data Mining Techniques in Customer Churn Analysis for Telecom Industry. Int. Journal of Engineering Research and Applications, 4(5), pp. 165-171.

10. Alwin, P.K.D.N.M., Kumara, B.T.G.S., Hapuarachchi, H.A.C.S. (2018). Customer Churn Analysis and Prediction in Telecommunication for Decision Making. 2018 International Conference on Business Innovation (ICOBI), pp. 40- 45.

11. Amin, A, Anwar, S, Adnan, A, Nawaz, M, Alawfi, K, Hussain, A, Huang, K. (2016). Customer churn prediction in the telecommunication sector using a rough set approach. Neurocomputing, volume 237, pp. 242-254.

12. Asthana, P. (2018). A comparison of machine learning techniques for customer churn prediction. International Journal of Pure and Applied Mathematics, 119(10), pp. 1149-1169.