CUSTOMER CHURN PREDICTION ANALYSIS

E Padma Priya ., S S Janashree .,P Sivaganga *

Artificial Intelligence and Data Science

Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering College, Avadi,

Chennai 600062- India.

Abstract

The Customers are the base of many successful businesses; thus, all the sectors are starting to understand how important it is to gain client satisfaction. The technical infrastructure has expanded quickly, changing how businesses operate. Due to growing business competition, the importance of marketing techniques, and customers' increasingly aware behaviour in recent years, leaving the organization is a crucial issue and it is one of the most crucial worries for most of the sectors. Different approaches must be developed by organizations to address the churn problems affecting the services they provide. To gain a deeper understanding of customer churn, this projectsummarizes churn prediction techniques. It also demonstrates that hybrid models, as opposed to single algorithms, provide the most accurate churn predictions, allowing telecom industries to better understand the needs of high- risk customers. Reduced client turnover is turning into a demand for service suppliers as a result of the potential impact that client neglect may wear business profitability. Prediction helps to search out users' World Health Organization area units doubtlessly to change from one organization to a different. The ever-rising churn rate in the medium could be a drawback. In lightweight of this, the present work makes use of a big-data platform and machine learning methodology. These medium firms are safeguarded with effective strategies for reducing the speed of churn because of machine learning algorithmic program techniques.

Keywords: Churn, Machine Learning, Random Forest,telecommunication,

I. INTRODUCTION

The telecommunications sector is growing to be one amongst the most important within the world, and as a result, the extent of competition has increased thanks to technological advancements and a rise within the variety of operators. To survive during this competitive trade, medium firms have developed methods that aim to get important sums of financial gain. It's crucial for businesses to cut back the probability of client churn, typically referred to as the migration of consumers betweenservice suppliers, to extend client retention rates. In commission industries with additional competitive services, client turnover is thought to be a heavy drawback. Varied analysis stressed that this circumstance is also foretold with growing accuracy by machine learning systems. The churn rate of consumers is best foretold by the machine learning algorithmic program. Supervised machine learning algorithms like random forest area units are often used in classification and regression problems. On numerous samples, it constructs call trees and uses their average for classification and majority vote for regression. The study found that the Random Forest algorithmic program out performs call trees and supply regression by a touch margin. The most problems facing the medium sector are unit user acquisition and retention. Each industry's marketplace is increasing quickly, that is resulting in a bigger subscriber base. Firms currently perceive however vital it's to stay in the present client base. Reduced client turnover is turning into a demand for service suppliers as a resultof the potential impact that client neglect may wear business profitability.

Telecom is presently dealing with an increasing of churn rate. As a necessary consequence, the latest research tends to make use of a machine learning algorithm on a big - data platform. Machine learning algorithm method allows the telecom companies to be secured with important means for lowering churn rates. Silent churn is a kind of churn which is difficult to predict because it's possible that such users will churn in the near future. The goal of decision-makers and marketers to decrease the churn proportion because it is a well-known fact that current users are more useful for businesses than reaching out to new. The primary purpose is exploring how the company use machine learning algorithm to analyze the customer churn from the organization. The secondary purpose is to look into how the leaving of a consumer affects the whole telecom sector, to tell about the importance of the churn prediction to the telecom sector and to study the algorithms used by telecom companies to lower the rate of customer churn. The limitatons of the churn prediction is the scope of the present analysis is restricted to the telecommunications sector and this analysis will not allow to use the alternative methods instead of machine learning techniques.

I. RELATED WORKS

Saran kumar A., Chandrakala D, gives a thorough analysis of the techniques employed in the process of predicting client attrition. For the first approach, they have constructed a dataset using real-world surveys and analyzed them using Deep Learning, Logistic Regression, and Naive Bayes algorithms. The second approach involves predicting customer attrition by looking at user-generated content (UGC), such as comments, postings, messages, and product or service evaluations. They employed sentiment analysis to determine the text's polarity (positive/negative) when assessing the user-generated content. The findings demonstrate that, while the algorithms were equally accurate, they differed in how they arranged the qualities according to their relative importance in the conclusion.

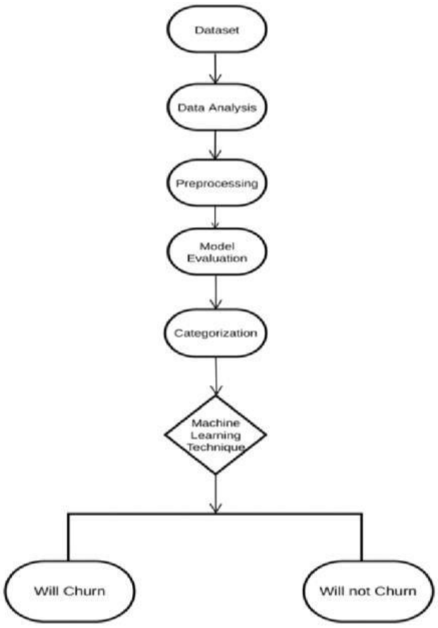
Nadeem Ahmad Naz, Umar Shoaib and M. Shahzad Sarfraz performed customer attrition predictions for the telecom industry. The article lists numerous characteristics that the many paper reviewers utilized to create a customer churn prediction model in practise. These characteristics include segmentation, account information, billing information, call dialup kinds, line information, payment information, complainant information, service provider information, and services information. About ten modelling strategies, including LR, NNM, DT, FL, CMC, SVM, and DME, are discussed in this study for the aim of churning. The review demonstrates that predicting customer churn depends on the goals of the decision makers. For example, DT and SVM with a low ratio are utilised if interested in both the real and false churn rates.

Navid Forhad, Md. Shahriar Hussain, Rashedur M Rahman they have presented the churn analysis in a telecom industry. The analysis aims to anticipate churn is based on categorization using rules. Though the dataset's incompleteness makes it difficult to develop a predictive design The results show that we must employ a comprehensive and substantial dataset if we intend to do any kind of precision. The paper provides a sample churn analysis method that might be used more widely in creating a fresh viewpoint for creating better churn analysis methods. Additionally, from a marketing standpoint, what can be done to keep them. It's also interesting to consider how long the data should be collected for

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 02 | February - 2023 Impact Factor: 7.185 ISSN: 2582-3930

SYSTEM DESIGN

The paper provides a sample churn analysis method that might be used more widely in creating a fresh viewpoint for creating better churn analysis methods. Whether or not churning clients are worthwhile to keep is an interesting question from the viewpoint of a business. Additionally, from a marketing standpoint, what can be done to keep them.
It's also interesting to consider how long the data should be collected for.



The dataset that was used in this study has a large number of features, from which we selected the ones that were most important to improving performance assessment and helping us make decisions. Classification performance improves only if the dataset only contains highly predictable features. As a result, improving classification performance requires fewer unrelated features and a greater emphasis on important properties. For the purpose of predicting consumer churn in the telecommunications industry, numerous methodologies have been put forth. Hereby, by using the Random Forestwe can predict the customer churn rate. For example the consumers who are most likely to stop using the service or endtheir subscription.

III MATERIAL AND METHODS

PRE-REQUISITES FOR BUILDING A CHURN PREDICTION MODEL

We use Kaggle's Telco Customer Churn dataset for this analysis. Use Jupyter Notebook for visualization.
Make sure the following libraries are installed - Pandas, Matplotlib, pickle, joblib and shap.

REVIEWING THE DATASET

We load our data frame into Python using the pandas library. The file was saved as "customer_churn.csv". data frame has 21 columns related to subscriber behavior of telecom users. Each customer is recognized by a individual user identifier. There are 19 unique features which is used to predict the target characteristic - customer leaving. In our dataset, the churn rate will be calculated on the basis of total number of users who have left organizations in the previous month.

```
# import telecom dataset into a pandas data frame
df = pd.read_csv("C:/Data/Telco-Customer-Churn.csv")#
visualize column names
df_telco.columns
```

DEMOGRAPHIC DETAILS

Gender of the client: If the client is a man or a woman (Female, Male)
Senior Citizen: If the customer is a senior citizen or not (0.1)
Partners: If the user has a co-conspirator or not (yes, no)
Dependents: If the client is supported by others or not (yes, no).

CUSTOMER ACCOUNT INFORMATION

- Tenure: The length of time a consumer has been a client of the business (Multiple different numeric values).
- Contract: Identifies the nature of the customer's present contract (Monthly basis, One year, Two year).
- Paperless Billing: If the customer uses online payment method or not. (Yes, No).
- Payment Method: The chosen payment method by the customer (credit card, automatic bank transfer,electronic check, paper check).
- Monthly Charges: The sum billed to the client each month (Multiple different numeric values).
- Total Charges: The entire sum that has been charged to the client (Multiple different numeric values)

SERVICES DETAILS

- Phone service – If the user has tele/smart phone service or not (yes or no).
- Multiple lines – If the user has a group of analog tele lines with one number or not. (No smartphone carrier, yes, no)
- Internet Services – If the user is accede to internet carrier with organizations.
- Online security – If the user has network safety or not (no carrier, yes, no)
- Online backup – If the user has network backup or not. (no network carrier, yes, no)
- Tech support – If the user has technical assistance or not (no network carrier, yes, no)
- Streaming TV – If the user has telecasting TV or not (no network carrier, yes, no)
- Streaming Movies – If the user has broadcasting films or not (no network carrier, yes, no)

EXPLORATORY DATA ANALYSIS FOR CUSTOMER CHURN PREDICTION

- The purpose of exploratory data study is to get a better understanding of the columns and rows in your dataset and how they relate to customer churn. Most of the customers in the dataset are young people without families. Users' gender and marital status are evenly distributed. In the world, customer tends to leave the wireless carrier; start using another service if they consider the monthly subscription charge to be excessive.

  i) Missing values and data types
- The pandas.DataFrame.info perform is beneficial within the early stages of EDA after we need to be told the maximum amount as we will concerning the {info - the information}. This methodology produces aquick outline of the info frame that has the names of the columns, the info sorts of the columns, and also memory consumption of info frame. The info set contains 7043 observations and twenty-one columns.
- # summary of the data framedf_telco.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 21 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   customerID        7043 non-null   object
 1   gender            7043 non-null   object
 2   SeniorCitizen     7043 non-null   int64
 3   Partner           7043 non-null   object
 4   Dependents        7043 non-null   object
 5   tenure            7043 non-null   int64
 6   PhoneService      7043 non-null   object
 7   MultipleLines     7043 non-null   object
 8   InternetService   7043 non-null   object
 9   OnlineSecurity    7043 non-null   object
 10  OnlineBackup      7043 non-null   object
 11  DeviceProtection  7043 non-null   object
 12  TechSupport       7043 non-null   object
 13  StreamingTV       7043 non-null   object
 14  StreamingMovies   7043 non-null   object
 15  Contract          7043 non-null   object
 16  PaperlessBilling  7043 non-null   object
 17  PaymentMethod     7043 non-null   object
 18  MonthlyCharges    7043 non-null   float64
 19  TotalCharges      7043 non-null   object
 20  Churn             7043 non-null   object
dtypes: float64(1), int64(2), object(18)
memory usage: 1.1+ MB
```

PREPROCESSING DATA FOR CUSTOMER CHURN PREDICTION MODEL

The information selected to answer the problem needs to be transformed into a machine-learning format. The model's performance, and thus the quality of the insights it produces, is dependent on the quality of the data. Thus, the primary goal is to ensure that every knowledge point is delineated with equivalent logicand is consistent across the data set.

CUSTOMER CHURN PREDICTION MODEL

The primary goal of this project is to create a churn prediction model. A specialist typically trains, tunes, evaluates, and tests a large number of models to define and detects the desired accuracy from the trained data model. Random forest models shows more accurate than other algorithms.

- Customer demographics, such as age, education, geography, and income, as well as other essentialdetails about the client.
- Features of user behavior, describes the lifecycle stage and number of subscriptions used by users ofthe service or product.
- Support elements that describe communications with customer support, such as satisfaction ratings
- Context features that represent other contextual information about the customer.

DEPLOYMENT AND MONITORING

This step puts the resulting models into practice and sets continuous data mining. The developed model need to be integrated for more accuracy and performance. Testing and monitoring model performance of the features will help to improve the model's accuracy. In the case of mobile services, monitoring and testing couldimply logging customer interactions and reviews.

IV RESULTS AND DISCUSSIONS

PROPOSED SYSTEM

To predict customer churn, we trained a model using various algorithms like logistic regression, XGBoost, and random forest to find an accurate value of customer lifetime. The analysis revealed that the random forest algorithm exceeds decision trees and logistic regression in terms of accuracy. The model was implemented bytraining and testing the dataset, resulting in the greatest number of correct values. Figure 1 depicts and describes the proposed model for churn prediction. The first step is data preprocessing, which involves filtering data and converting it into a similar format before selecting the feature. From the steps given above, we predictand classify the churn using the Random Forest algorithm. The data set was trained and tested in the model toobserve and analyse customer behaviour.

FUTURE ENHANCEMENT

The current level of this project is at a stage where we can only predict the churn result or a telecom service. We also intend to give a more specific analysis result with the help of advanced concepts of machine learning and AI interactions. We plan to enhance this project to the point where it can be a common interface for any domain or field in action. This project concept has a great scope in the growing industrial era, and it will be quite useful for any kind of service offered in the mark

V CONCLUSION

Customers can switch from one provider to another easily because the telecommunications sector is one of the most competitive fields and has many organizations to provide service for the company. The task of analyzing churn is to determine which customers are going to churn the organization. Consumer churn is a crucial problem for most industries, as customers will not hesitate to leave the company unless they get what they want. In exchange for the money they paid, the customer will expect good service. Because of the service given by the sector, customers are only leaving the company if they are dissatisfied. The cost of customer acquisition is well known to be higher than the cost of customer retention, making it difficult for a prototyping business to retain customers. There is no standard solution to the churning problem that global service providers face in the telecommunications industry. Big data analysis using machine learning techniques is applied to customer churn, providing customers with warning signals before damage occurs and allowing operators to take preventive measures. By developing models and analyzing historical data, you can use these techniques to determine customer churn. Run experiments from the perspective of the end user, collect networkviews, normalize data, preprocess datasets, use function selection, remove imbalanced class and missing rows and columns, and use derived variables by modifying existing data to help the telecom industry retain customers very efficiently. We can infer that machine learning is an effective method for identifying client churn.

REFERENCES

[1] Alkhayrat M, Aljnidi M, Aljoumaa K (2020) A comparative dimensionality reduction study in telecom customer segmentation using deep learning and PCA. J Big Data

[2] Saran Kumar A, Chandrakala D “A Survey on Customer Churn Prediction using Machine Learning Techniques” November 2016, International Journal of Computer Applications, 154(10): 13–16, DOI: 10.5120/ijca2016912237 volume Article number: 28 (2019).

[3] Abdelrahim Kasem Ahmad, Assef Jafar & Kadan Aljoumaa “Customer churn prediction in telecom using machine learning in big data platform” Journal of Big Data, Article number: 28 (2019).

[4] Kriti, “Customer Churn: A Study of Factors Affecting Customer Churn using Machine Learning”, Iowa State University Ames, Iowa 2019.

[5] A. K. Ahmad, A. Jafar and K. Aljoumaa, "Customer churn prediction in telecom using machine learning in big data platform", Journal of Big Data, vol. 6, no. 1, 2019.

[6] M. Malyar, M. M. Robotyshyn and M. Sharkadi, "Churn Prediction Estimation Based on Machine Learning Methods", 2020 IEEE 2nd International Conference on System Analysis & Intelligent Computing (SAIC), pp. 1-4, 2020, October.

[7] M. Maw, S. C. Haw and C. K. Ho, "Customer Churn Prediction in Telecommunication: An Analysis on Issues Techniques and Future Trends", 2019 IEEE Conference on Sustainable Utilization and Development in Engineering and Technologies (CSUDET), 2019.

[8] A. Ben, "Enhanced Churn Prediction in the Telecommunication Industry", International Journal of Innovative Research in Computer Science & Technology (IJIRCST), vol. 8, 2020.

[9] L. C. Cheng, C. C. Wu and C. Y. Chen, "Behavior analysis of customer churn for a customer relationship system: an empirical case study", Journal of Global Information Management (JGIM), vol. 27.1, pp. 111-127, 2019.

[10] N. Alboukaey, A. Joukhadar and N. Ghneim, "Dynamic behavior based churn prediction in mobile telecom", Expert Systems with Applications, vol. 162, pp. 113779, 2020.

[11] R. Manivannan, R. Saminathan and S. Saravanan, "An improved analytical approach for customer churn prediction using Grey Wolf Optimization approach based on stochastic customer profiling over a retail shopping analysis: CUPGO", Evolutionary Intelligence, vol. 14.2, pp. 479-488, 2021.

[12] E. Jamalian and R. Foukerdi, "A hybrid data mining method for customer churn prediction", Engineering Technology & Applied Science Research, vol. 8.3, pp. 2991-2997, 2018.

[13] S. F. Sabbeh, "Machine-learning techniques for customer retention: A comparative study", International Journal of Advanced Computer Science and Applications, vol. 9, no. 2, 2018

[14] Saheed, Y.K. and Hambali, M.A., 2021, October. Customer Churn Prediction in Telecom Sector with Machine Learning and Information Gain Filter Feature Selection Algorithms. In 2021 International Conference on Data Analytics for Business and Industry (ICDABI) (pp. 208–213). IEEE.

[15] Dias, J., Godinho, P. and Torres, P., 2020, July. Machine learning for customer churn prediction in retail banking. In International Conference on Computational Science and Its Applications (pp. 576–589). Springer, Cham.

[16] Domingos, E., Ojeme, B. and Daramola, O., 2021. Experimental Analysis of Hyperparameters for Deep Learning-Based Churn Prediction in the Banking Sector. Computation 2021, 9, 34.

[17] Vo, N.N., Liu, S., Li, X. and Xu, G., 2021. Leveraging unstructured call log data for customer churn prediction. Knowledge-Based Systems, 212, p.106586.

[18] Shirazi, F. and Mohammadi, M., 2019. A big data analytics model for customer churn prediction in

the retiree segment. International Journal of Information Management, 48, pp.238–253.

[19] Kolomiiets, A., Mezentseva, O. and Kolesnikova, K., 2021. Customer churn prediction in the software by subscription models it business using machine learning methods. In CEUR Workshop Proc(Vol. 3039, pp. 119–128).

[20] Amuda, K.A. and Adeyemo, A.B., 2019. Customers churn prediction in financial institution using artificial neural network. arXiv preprint arXiv: 1912.11346.

[21] Jagadeesan, A.P. Bank customer retention prediction and customer ranking based on deep neural networks. *Int. J. Sci. Dev. Res.***2020**, *5*, 444–449

[22] Amuda, K.A.; Adeyemo, A.B. Customers churn prediction in financial institution using artificial neural network. *arXiv***2019**, arXiv:1912.11346

[23] Iranmanesh, S.H.; Hamid, M.; Bastan, M.; Hamed Shakouri, G.; Nasiri, M.M. Customer churn prediction using artificial neural network: An analytical CRM application. In Proceedings of the International Conference on Industrial Engineering and Operations Management, Bangkok, Thailand, 5–7 March 2019; pp. 23–26.

[24] Olaniyi, A.S.; Olaolu, A.M.; Jimada-Ojuolape, B.; Kayode, S.Y. Customer churn prediction in banking industry using K-means and support vector machine algorithms. *Int. J. Multidiscip. Sci. Adv. Technol.***2020**, *1*, 48–54.

[25] Kaya, E.; Dong, X.; Suhara, Y.; Balcisoy, S.; Bozkaya, B. Behavioral attributes and financial churn prediction. *EPJ Data Sci.***2018**, *7*, 41.