



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 5, May 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Churn Prediction Using Machine Learning Technique

Jagruti S Patil, Prof C R Barde,

Department of Computer Engineering, GES's R.H.Sapat College of Engineering, Management Studies and Research,
Nashik, India

ABSTRACT: Telecom customer competition forecasting is an important task for telecom companies to retain customers. Churn is when a customer cancels their subscription or service from a communications company. Predicting customer churn helps telemarketing companies take steps to retain customers by identifying potential churn and providing effective retention strategies for them. This summary explains the context of the communication problem using machine learning. Contacts for problem prediction may include analysis of a customer's historical data, including demographic data, usage patterns, payment details, and service history, to predict whether a customer will leave in the future. Telecom customer churn prediction using machine learning involves processing customer history data, architectural design, selecting appropriate machine learning algorithms, effectively evaluating performance models using multiple metrics, and using best practices in a production environment. By using these methods, communications companies can reduce customer turnover and increase customer satisfaction.

KEY WORDS: Machine Learning, Churn Prediction, Churn, Telecom Churn Random Forest, Decision Tree, XGBoost, Churn, ANN.

I. INTRODUCTION

Customer retention in today's highly competitive telecommunications industry is a challenge for service providers. Customer churn, which is when customers cancel their subscriptions or services, is a major concern for telecommunications companies as it leads to loss of revenue and operations. To overcome this challenge, mobile companies are turning to machine learning technology to predict customer competition and take proactive steps to protect against it. Gaming is important for phone companies because it helps them retain customers and reduce customer churn rates. Telecommunications companies can prevent customer churn and increase customer satisfaction by identifying potential customer churn and offering them attractive retention strategies. Additionally, churn forecasting allows telephone companies to improve their marketing and sales by focusing on valuable customers and without affecting users. Telecom churn forecasting is an important task for telecom companies because it helps them retain existing customers and reduce customer churn. Customer loss. Churn is the process when a customer stops providing service to a communications company. Using machine learning algorithms to predict customers can help phone companies identify factors that drive customer churn and take proactive steps to retain customers. Telecom churn forecasting is an important task for telecom companies because it helps them retain existing customers and reduce customer churn. Mobile marketing companies can use machine learning algorithms to create accurate and effective churn prediction models. These models can be integrated into the production environment to provide instant predictions and help companies take proactive steps to retain customers.

A. Problem Statement

Telecom churn prediction is a crucial task for telecom companies as it helps them retain their existing customers and reduce customer attrition. Telecom Churn Prediction it helps telecom companies to retain their customers and minimize revenue loss due to customer churn. Churn refers to the percentage of customers who discontinue their subscription or service with a telecom company. Predicting customer churn is essential for telecom companies as it allows them to take proactive measures to retain customers, such as offering personalized promotions, discounts, or better services. Churn refers to the process of customers discontinuing their services with a telecom company. Predicting churn using machine learning algorithms can help telecom companies identify the factors that lead to customer churn and take proactive measures to retain them. Telecom companies can build accurate and effective churn prediction models using machine learning algorithms. The models can be deployed in a production environment to provide real-time predictions and help the company take proactive measures to retain customers. The predictive model should take into account various factors such as customer demographics, usage patterns, payment history, customer service interactions, and other relevant data points. It should be able to analyze historical data on customer churn and create a model that can predict future churn behavior.

B. Objectives of the study

1. The objective of this proposed system is to develop a machine learning model that can accurately predict customer churn in the telecom industry.
2. Churn refers to the phenomenon where a customer discontinues their subscription or service with a telecom company. Predicting churn is crucial for telecom companies as it enables them to take proactive measures to retain customers, reduce churn rates, and improve customer satisfaction.
3. To compare the algorithms that are effective in reducing churn rate in telecom companies
4. The scope of this proposed system includes collecting and preprocessing customer data, feature engineering, selecting and training appropriate machine learning algorithms, and evaluating the performance of the model. The system will also incorporate techniques such as data augmentation, ensemble learning, and hyper parameter tuning to improve the model's accuracy and robustness.
5. The proposed system will be developed using Python and the popular machine learning libraries. The system will be trained and tested on a large and diverse dataset of telecom customer data, which will be obtained from reputable sources such as Kaggle. The system will be designed to be scalable, efficient, and easy to deploy in a production environment

C. Features of Project

1. Real-time monitoring
2. Fraud detection
3. Historical Customer Data
4. Customer demographics
5. Customer service interactions
6. Network performance metrics
7. Competitor information
8. Social media sentiment analysis
9. Customer satisfaction surveys/NPS scores

II. RELATED WORK

1. Dr. O. Rama Devi, Sai Krishna Pothini has launched a model targeting individuals who use premium OTT platforms to stream video content on any device. This study used a survey to collect information from participants of each population. Data collection needs to go through many processes before it can be made suitable for machine learning models. Use this platform. This information is important for OTT companies to understand and optimize their marketing and retention processes. This process involves cleaning data, selecting important features, and training machine learning models. The model is then tested and validated through performance testing. The results can inform OTT companies to improve their customer acquisition and retention processes. [one]
2. The system proposed by QiuYing Chen and SangJoon Lee uses Orange3 software to create a churn prediction model for product delivery. Choosing the best gradient boosting algorithm for churn prediction on a food delivery platform. Model estimation using gradient boosting algorithm gives good and accurate results, easy to use. It also offers important features that make the use of gradient boosting methods more effective, unlike the results of general learning methods. Especially in ecommerce, it is more useful to use additional development methods to predict the customer without a contract. [2]
3. The system proposed by Weijie Yu and Weinan Weng aims to identify the factors affecting customer churn and create a good model to predict and analyze the data obtained from the results found. Customer churn prediction consists of several stages such as preliminary data, data analysis, measurement and evaluation, and machine learning algorithms. It also includes data preprocessing, data cleaning, transformation and classification. The selected machine learning methods are Logistic Regression, SVM, Random Forest, AdaBoost, GBDT, XGBoost, Light GBM and CatBoost. Classifiers are evaluated using performance metrics such as accuracy, precision, recall, AUC, and F1 score. According to the paper, the results show that Light GBM is better at identifying potential candidates than other classifications. [3] Chapte
4. Gavriel et al. To predict prepaid customer churn, an advanced data mining method is proposed using call data of 3333 customers with 21 characteristics and the associated churn parameter with two outcomes (yes/no). Some of the features include recordings of incoming and outgoing calls and voicemails for each customer. The authors used a keypoint analysis algorithm (PCA) to reduce the remaining data. Use three machine learning algorithms to estimate loss factors: neural networks, support vector machines, and Bayesian networks. The author uses AUC to measure the performance of the algorithm. The AUC values of the Bayesian network, neural network, and support

vector machine are 99.10%, 99.55%, and 99.70%, respectively. The data used in this study were small and had no missing values. [4]

5. Huang et al. The problem of customer churn on the big data platform was investigated. The researcher's goal is to show that big data improves the process of predicting customers based on the volume, variety and speed of data. At China's largest telecommunications company, processing information from the support office and business support requires a large amount of information to resolve the problem. Use the random forest algorithm and measure using AUC. [5]
6. This article describes our work on churn analysis and forecasting for these types of services. We work on data mining techniques to accurately and effectively predict whether users will switch (lose) to other providers offering the same or similar services. The information we use is evidence and facts compiled by Orange Telecom for the KDD 2009 competition. Many groups score high on this information, which requires the use of important information. Our goal is to find other ways to match or improve the high scores recorded by using more efficient and effective resources. In this study, we focused on a set of meta-classifiers that were individually trained and selected based on their performance.
7. Idris proposed a method based on AdaBoost genetic programming to simulate the problem of loss in communication. The model was tested on two data sets. One is provided by Orange Telecom and the other is provided by cell2cell. The accuracy of the Cell2cell dataset is 89%, while the accuracy of the other dataset is 63%. [7]
8. He et al. A prediction model based on neural network algorithm is proposed to solve the customer churn problem of a large telecommunications company in China with approximately 5.23 million customers. The model is very accurate, reaching 91.1% accuracy.
9. AZhang Y proposed a combination method to build a binary classifier. This method is a combination of the knearest neighbor algorithm and the logistic regression method. The knearest neighbor algorithm works by dividing m onedimensional data into mdimensional data sets. This hybrid KNNLR classifier improves the classification accuracy of logistic regression in some cases where the predictor and target variables result in a nonlinear relationship. Experimental results on four test data show that the quality of the method compares with wellknown classification algorithms such as C4.5 and RBF. It also shows the results of its application in customer churn prediction based on real customer data.
10. ShinYuan Hung proposed a different methodology to develop predictive models for interactive customer engagement. Following recommendations from previous research by Wei and Chiu (2002) , we included customer service and customer dissatisfaction inputs in the model. We examined the effects of insufficient information on design. Our empirical analysis shows that data mining techniques can help mobile service providers meet customer needs effectively.
11. Combine dynamic gradient boosting with physical data to estimate losses. Selected data were used in this study from the WSDM Cup 2018 competition of KKBOX, a music streaming service. This study used supervised machine learning with decision trees to develop a reliable testing model to predict customer churn in the existing XGBoost library. Improved accuracy by combining the LightGBM library with the main XGBoost model. The standard model outperforms other models submitted to the competition due to its high accuracy. Future research will include further optimization of the XGBoost and LightGBM models using Stack Net search and unprecedented performance models.
12. Dolatabadi and Keynia presented various neural studies and data mining techniques. The data was collected over a year and a half and provides information on all employees and customers. By comparing methods such as decision trees, naive Bayes, support vector machines, and neural networks, this article concludes that support vector machines can be used to produce reliable and accurate consumer products. Future work will include analyzing customer and employee performance to improve predictive models and increase the reliability of participating companies
13. Hoppner et al. proposed the development phase model of decision tree, introduced a new loss classification called ProfTree, and developed an evolutionary algorithm to optimize EMPC [10]. The data used is a real churn dataset from different communication providers, containing 889 customers and 10 different descriptions. ProfTree, EvTree, CART, ctree etc. It is generally the most awardedmodel that measures the complexity and effectiveness of other tree methods such as. By creating and collecting various valuable trees, the value of the property will increase.
14. Fei et al proposed a new method, a Naive Bayes classifier, to predict users using the K method. K shows the better accuracy and precision of the combination Naive Bayes classification method compared to the Naive Bayes classification method combined with EWD. However, they have weaknesses in predicting true negative and positive outcomes. This information is collected from clean data received from telecommunication companies. The selected data included five thousand callers and twentyone features. The project aims to build models by learning

- temporal bounds that affect the future learning rate of a classifier. The technology can be enhanced with different algorithms such as support vector machine, decision tree and Bayesian networks.
15. Kumar, A.S. and Chandrakala, D. proposed an improvement method such as AdaBoost classification using vector machines to overcome many classification problems, using detection tools as predictions, combining AdaBoost SVM, NBTree and support removal vector machine classification. The data set used is the bank. To effectively predict customer churn rate, the plan is to get a more comprehensive distribution.
 16. Xia and Jin, Customer churn prediction demand based on support vector machine (SVM). Data analysis from the University of California Machine Learning and Home Telecommunications repository. Compared with BPANN, C4.5 decision tree, logistic regression, and naive Bayesian classification systems, SVM has better predictive ability, greater reproducibility, and better similarity.

III. METHODOLOGY

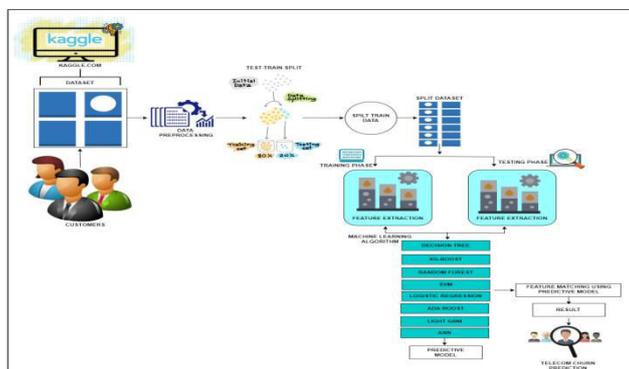


Fig.1.Representation Of The Methodology.

The simple formula for predicting future loss is historical data. We look at profiles of customers who have already stopped using the app (responders) and their characteristics/behaviors before churn occurs (predictors). The data set includes customers' details, their total costs and the services they received from the company. There is churn data for many customers from Kaggle from more than 21 features. This method falls into the category of educational supervision.

IV. CONCLUSION

In conclusion, churn prediction is a critical challenge for telecom companies as it affects revenue, market share, and customer satisfaction. The challenges involved in churn prediction include dealing with large volumes of historical customer data, handling noisy, incomplete, and highly dimensional data, and addressing imbalanced datasets. Various machine learning algorithms such as Random Forest, Decision Tree, XGBoost, etc. can be used to solve this problem depending on the nature of the data and the specific requirements of the telecom company. By implementing machine learning algorithms for churn prediction, telecom companies can reduce churn rates, improve customer satisfaction, and gain a competitive advantage in the market.

REFERENCES

1. M Dr. O. Rama Devi, Sai Krishna Pothini, "Customer Churn Prediction using Machine Learning: Subscription Renewal on OTT Platforms", IEEE Xplore Part Number: CFP23BC3-ART; ISBN: 978-1-6654-5630-2, 978-1-6654-5630-2/23/\$31.00 2023 IEEE
2. QiuYing Chen, Sang-Joon Lee, "A Machine Learning Approach to Predict Customer Churn of a Delivery Platform", 2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)—978-1-6654-5645-6/23/\$31.00 2023 IEEE — DOI: 10.1109/ICAIIIC57133.2023.10067108
3. Weijie Yu, WeinanWeng, "Customer Churn Prediction Based on Machine Learning" 2022 4th International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM), 978-1-6654-6399-7/22/\$31.00 c2022 IEEE
4. Brandusoiu I, Todorean G, Ha B. Methods for churn prediction in the prepaid mobile telecommunications industry. In: International conference on communications. 2016. p. 97–100.

5. Huang F, Zhu M, Yuan K, Deng EO. Telco churn prediction with big data. In: ACM SIGMOD international conference on management of data. 2015. p .607–18.
6. Yabas, U, Chankya, H.C. (2013). Churn prediction in subscriber management for mobile and wireless communications services. IEEE Publications.
7. Idris A, Khan A, Lee YS. Genetic programming and adaboosting based churn prediction for telecom. In: IEEE international conference on systems, man, and cybernetics (SMC). 2012. p. 1328–32.
8. He Y, He Z, Zhang D. A study on prediction of customer churn in fixed communication network based on data mining. In: Sixth international conference on fuzzy systems and knowledge discovery, vol. 1. 2009. p. 92–4.
9. Zhang, Y.; Qi, J.; Shu, H.; Cao, J. A hybrid KNN-LR classifier and its application in customer churn prediction. In Proceedings of the 2007 IEEE International Conference on Systems, Man and Cybernetics, Montr´eal, QC, Canada, 7–10 October 2007; pp. 3265–3269.
10. Shin-Yuan Hung a, David C. Yen b, Hsiu-Yu Wang, “Applying data mining to telecom churn management”, Expert Systems with Applications 31 (2006) 515–524,
11. B. Gregory, “Predicting Customer Churn: Extreme Gradient Boosting with Temporal Data,” ArXiv180203396 Cs Stat, Feb. 2018, Accessed: Apr. 24, 2021. [Online]. Available: <http://arxiv.org/abs/1802.03396>
12. S. H. Dolatabadi and F. Keynia, “Designing of customer and employee churn prediction model based on data mining method and neural predictor,” in 2017 2nd International Conference on Computer and Communication Systems (ICCCS), Jul. 2017, pp. 74– 77, doi: 10.1109/CCOMS.2017.8075270.
13. S. Hoppner, E. Stripling, B. Baesens, S. vandenBroucke, and T. Verdonck, “Profit driven decision trees for churn prediction,” Eur. J. Oper. Res., vol. 284, no. 3, pp. 920–933, Aug. 2020, doi: 10.1016/j.ejor.2018.11.072.
14. T. Y. Fei, L. H. Shuan, L. J. Yan, G. Xiaoning, and S. W. King, “Prediction on Customer Churn in the Telecommunications Sector Using Discretization and Naive Bayes Classifier,” p. 13.
15. Kumar, A. S. and Chandrakala, D., “An Optimal Churn Prediction Model using Support Vector Machine with Adaboost,” Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol., vol. 2, no. 1, 2017.
16. G. Xia and W. Jin, “Model of Customer Churn Prediction on Support Vector Machine,” Syst. Eng. - Theory Pract., vol. 28, no. 1, pp. 71–77, Jan. 2008, doi: 10.1016/S1874- 8651(09)60003-X.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details