Dynamic Churn Prediction System using Machine Learning Algorithms

Prof. S. N. Bhadane[1], Gauri Randhir[2], Mamta Borade[3], Sahil Bhatia[4], Gaurav More[5]

[1] *Professor, Information Technology, Pune Vidyarthi Griha's College of Engineering, Nashik, MH, India*
[2] *Student, Information Technology, Pune Vidyarthi Griha's College of Engineering, Nashik, MH, India*
[3] *Student, Information Technology, Pune Vidyarthi Griha's College of Engineering, Nashik, MH, India*
[4] *Student, Information Technology, Pune Vidyarthi Griha's College of Engineering, Nashik, MH, India*
[5] *Student, Information Technology, Pune Vidyarthi Griha's College of Engineering, Nashik, MH, India*

ABSTRACT

Abstract— Customer churn occurs when customers or subscribers stop doing business with a company or service. Customers in the telecom business have the option of selecting from several telecom operator and actively switching from one to the next. In this extremely competitive sector, the telecoms industry has an annual turnover rate of 15-30 percent. Individualized customer retention is difficult since most businesses have a huge number of customers and cannot afford to devote a significant amount of time to each of them. The greater revenue would be outweighed by the costs. However, if a company can predict which customers are likely to depart ahead of time, it can target customer retention efforts solely on these "high-risk" consumers. The goal is to broaden its coverage area and re-establish consumer loyalty. The client is at the heart of success in this market. Customer churn is an important indicator since retaining existing customers is substantially less expensive than acquiring new customers. To discover early warning indications of probable churn, one must first establish a comprehensive perspective of the consumers and their interactions across several channels. As a result, by managing churn, these companies may be able to not only maintain their market position but also grow and thrive. The greater the number of consumers in their network, the lower the cost of initiation and the greater the profit. As a result, lowering client attrition and creating an effective retention plan is the company's primary priority for success.

Keywords— Customer churn prediction, Churn in telecom, Machine learning, Feature selection, Classification.

---

1. INTRODUCTION

In the telecom sector, there are two payment models: postpaid (users pay a monthly/annual generated payment after using the services) and prepaid (customers pay/recharge a set of fixed charges in advance and then use the services). When customers desire to change to another service provider in the postpaid model, they normally notify the current operator to terminate the services, and you can immediately tell that this is an instance of churn.

Customers who want to migrate to another network, on the other hand, can simply cease using the services without any notification in the prepaid model, and it is difficult to tell whether someone has genuinely churned or is simply not using the services temporarily (e.g., someone may be on a trip abroad for a month or two and then intend to resume using the services again). As a result, churn prediction is frequently more crucial (and difficult) for prepaid clients, and the term "churn" should be defined carefully. Furthermore, prepaid is the most popular model in India and Southeast Asia.

2. LITERATURE SURVEY

Authors of [1] this study, introduced a brand-new approach that is recommended for non-contractual firms. Modeling consumer behavior using Beta-geometric Negative Binomial Distribution (BG/NBD) is one of the extensively utilized approaches for these kinds of firms in the literature. This strategy is known to perform poorly when it comes to categorizing specific consumers yet does well when estimating the population's overall turnover rate. Using machine learning, the BG/NBD model is intended to be improved. The suggested approach uses decision trees to extract features from the mathematical description of the BG/NBD model. Tested on two data sets with various features, it is found that the suggested strategy significantly improves performance.

For unbalanced data, such as bank clients, this research[2] develops a Borderline-SMOTE-random forest prediction model. It can better address the unbalanced data and the powerful anti-noise capacity of the random forest when combined with the oversampling approach. OOB error rate, AUC, Precision, Recall, and F-mean are utilized as the model's assessment metrics, and comparisons are made with KNN, decision trees, and Naive Bayes. According to the experimental findings, the Borderline-SMOTE-random forest prediction model is the most effective at resolving the issue of predicting bank customer turnover among the models, and its performance has increased by roughly 4%.

The authors[3] developed an Android application based on the Internet of Things to monitor environmental elements such as ammonia, methane, alcohol percentage, and TVOC. The ESP32, a well-known and popular development board, serves as the device's brain. Sensors such as the MQ135, MQ3, CCS811, and 162 LCD are connected to the Microcontroller board. ML is used to analyze the app's image and predict the condition of the food. The app includes a chatbot that provides food quality information. Based on the user's location, the app will recommend nearby organic stores.

The challenges associated with predicting client attrition in the auto insurance industry are illustrated in this survey work[4] along with several data mining techniques, including developments in deep learning and machine learning. Additionally, it places emphasis on the customer management cycle's churns in relation to the environment. A summary of the survey that was conducted in an organized manner shows how the churn prediction model was created, along with the various prediction techniques used and the business sector in which they were applied.

3. SYSTEM ARCHITECTURE AND METHODOLOGY

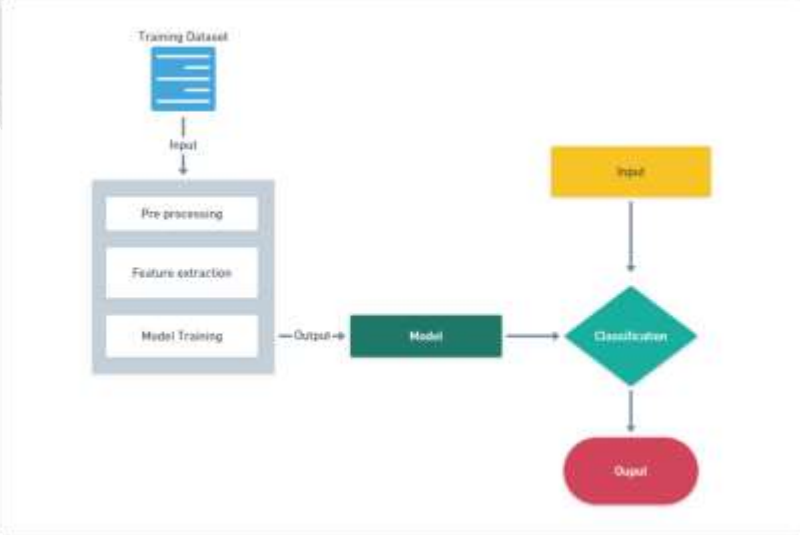

Fig -1: System Architecture

In the proposed system Python programming will be used to build the model for churn prediction. It is widely used among statisticians and machine learning developers for developing machine learning models and data analysis. Python is freely available and a powerful programming language. In the proposed work paper, we proposed different

machine-learning algorithms to analyze customer churn analysis. Which we can multiple different models are employed to accurately predict those churn customers in the data set. These models are Logistic Regression, Support Vector Machine, Random Forest, and Gradient Boosting Trees. Our Steps or Algorithm Steps are as follows–

1. **Dataset -** A telecom dataset is taken for predicting churn which to identify trends in customer churn at a telecom company and the data we have taken is in .csv format. The data given to us contains 7043 observations and 21 variables extracted from a dataset.
2. **Data Preparation** - Since the dataset acquired cannot be applied directly to the churn prediction models, so we can name each attribute.
3. **Data Preprocessing** - Data preprocessing is the most important phase in prediction models as the data consists of ambiguities, errors, redundancy, and transformation which need to be cleaned beforehand.
4. **Data Extraction** - The attributes are identified for the classifying process.
5. **Decision** - Based on data extraction and classification models we can take a decision whether the employee is a churner or not.

4. CONCLUSIONS

To retain existing customers, telecom companies must understand the causes of churn, which can be accomplished through knowledge acquired from telecom data. In this study, we are training four machine learning models: Logistic Regression, SVM, Random Forest, and Gradient Boosting. We may state that Gradient Boosting is the best of the four models, while Logistic Regression and Random Forest are average, and SVM is the worst. In the following phase, we will create the system's user interface and deploy the program to the cloud.

5. REFERENCES

[1] B. Bardük, "Modelling Time Statistics for Customer Churn Prediction," 2020 28th Signal Processing and Communications Applications Conference (SIU), 2020, pp. 1-4.

[2] L. Feng, "Research on Customer Churn Intelligent Prediction Model based on Borderline-SMOTE and Random Forest," 2022 IEEE 4th International Conference on Power, Intelligent Computing and Systems (ICPICS), 2022, pp. 803-807.

[3] X. Hu, Y. Yang, L. Chen and S. Zhu, "Research on a Customer Churn Combination Prediction Model Based on Decision Tree and Neural Network," 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), 2020.

[4] P. Gopal and N. B. MohdNawi, "A Survey on Customer Churn Prediction using Machine Learning and data mining Techniques in E-commerce," 2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), 2021.

[5] T. -Y. Tsai, C. -T. Lin and M. Prasad, "An Intelligent Customer Churn Prediction and Response Framework," 2019 IEEE 14th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), 2019, pp. 928-935.

[6] A. H. Ren and W. W. Zhao, "Electronic Commerce Based on Self-Organizing Data Mining Customer Churn Prediction Model", In 2013 International Conference on Advances in Social Science Humanities and Management (ASSHM-13), 2013, December.

[7] S. Akter and S. F. Wamba, "Big data analytics in E-commerce: a systematic review and agenda for future research", Electronic Markets, vol. 26, no. 2, pp. 173-194, 2016.

[8] L. Vanneschi, D. M. Horn, M. Castelli and A. Popovič, "An artificial intelligence system for predicting customer default in e-commerce", Expert Systems with Applications, vol. 104, pp. 1-21, 2018.

[9] A. Keramati, R. Jafari-Marandi, M. Aliannejadi, I. Ahmadian, M. Mozaffari and U. Abbasi, "Improved churn prediction in telecommunication industry using data mining techniques", Applied Soft Computing, vol. 24, pp. 994-1012, 2014.

[10] L. Zhao, Q. Gao, X. Dong, A. Dong and X. Dong, "K-local maximum margin feature extraction algorithm for churn prediction in telecom", Cluster Computing, vol. 20, no. 2, pp. 1401-1409, 2017.

[11] Bing Jia and Yongjian Yang, "The design of food quality supervision platform based on the Internet of Things," Proceedings 2011 International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE), 2011, pp. 263-266, doi: 10.1109/TMEE.2011.6199193.

[12] R. Rajamohamed and J. Manokaran, "Improved credit card churn prediction based on rough clustering and supervised learning techniques", Cluster Computing, vol. 21, no. 1, pp. 65-77, 2018.

[13] A. A. Ahmed and D. Maheswari, "An enhanced ensemble classifier for telecom churn prediction using cost based uplift modelling", International Journal of Information Technology, vol. 11, no. 2, pp. 381-391, 2019.

[14] A. Amin, S. Anwar, A. Adnan, M. Nawaz, K. Alawfi, A. Hussain, et al., "Customer churn prediction in the telecommunication sector using a rough set approach", Neurocomputing, vol. 237, pp. 242-254, 2017.