



Customer Churn Rate Prediction Using Machine Learning

¹Billa Varsha, ²Maddela Pranitha, ³Sallam Preetham

¹UG Student, ²UG Student, ³UG Student

¹Computer Science and Engineering,

¹Sreenidhi Institute of Science and Technology, Hyderabad, India

Abstract: Customer loss or churn is an enormous issue for most firms currently. Regarding this, the computation of the churn rates will most definitely help the level of customers retained and also improve the level of profit generation. It explains the most often-used machine learning algorithms: logistic regression and random forests for modeling churn from historical data. It elaborates on specific features like demographics and transaction history, which help increase results on this model. That too, at a machine learning platform, if proactive retention exists, outputs are even more superior compared to normal practice. Further work might incorporate real-time analytics and deep learning to create even better predictions related to churn in the future.

Index Terms - Customer Churn Prediction, Machine Learning, Logistic Regression, KNN Algorithm, Random Forest Algorithm, ROC Curve.

I. INTRODUCTION

1.1 BACKGROUND

It deals with forecasting churn in the pattern of customers or their previous trends based on that of those likely to halt usage of service. This requires, based on machine learning techniques, analyzing and Churn is that customer who stops conducting any business with a company. Thus, minimization of churning is very important in order to ensure saving of all the revenues and subsequently sustaining the growth of business. Normally, the data carries demographic information regarding the customer, transactions, engagement, etc. The cleansing and feature engineering are pretty much important too. Machine Learning: Algorithms like Random Forest are highly in use for predicting churn as they can easily deal with both categorical and numerical data. They also give feature importance. They can work even when the datasets are imbalanced which is quite common in churn-related scenarios. Performance Metrics: Common metrics are precision, recall, F1 score, and AUC - ROC depending upon what business wants - false negatives. Applications: Recommendations provided by the model are used through which a targeted retention strategy, personalized offer, and a better experience for the customer are formulated.

1.2 Motivation

Proactive retention is one feature through machine learning in predicting customer churn by pointing out those most likely to leave before it actually happens. It gives better accuracy and scalability because it would be possible to handle large complex data. Information obtained identifying particular drivers for different customer segments provides for tailored retention strategies that target those customers that are bringing high risk. It will then optimize resources by only focusing efforts on customers labeled as either high-risk or high-value customers. It shall handle the intricate relationship among factors that may be churning. It's going to contain real-time predictions and interventions that are time-related too.

It will have analytics of the features of relative importance, areas for improvements in experience, and reduction drivers of churning. It shall definitely promise a competitive edge at re-attributing the existing customers, further enhanced branding loyalty.

1.3 Objectives

1. The objectives of the customer churn rate prediction are as follows:
Identify the high-risk customer bases who would most probably churn and develop retention plans focusing on those.
2. This serves to reduce the pain points that will lead to decreasing churn. Retain the high-value customers who are generating the highest revenue.
3. Interactively work with at-risk customers to avoid loss in the churn. Acquired Lower Acquisition Cost of New Customers Maintenance of existing customers Retaining CLV the highest customers are more valued effort and focus.

II. RELATED WORKS

2.1 Predictive Analytics and Customer Retention

Role of prediction of customer churn for retention strategies. Role of Predictive Analytics in Business Profitability Applications of Predictive Modeling in various Industries.

2.2 Machine Learning Algorithms for Churn Prediction

Comparison of logistic regression, random forests, and KNN algorithms for churn prediction strengths and weaknesses of the supervised learning techniques in modeling churn Hyperparameter Tuning to Enhance model performance.

2.3 Data Preprocessing for Churn Analysis

Working on categorical features such as geography and gender. Scaling of numeric features such as credit score and balance. Encoding the use of binary features, having a credit card, being an active member or otherwise.

2.4. Feature Engineering and Selection

Demographic along with transaction history plays major influences the churn. Feature Importance - use of random forest coefficients of logistic regression. Drawing up the feature derived by balance-to-salary, age groups, etc.

2.5. Evaluation Metrics of the Churn Prediction Model

ROC curves, precision, recall, and F1-score for model evaluation. Explanation of Area Under the Curve (AUC) for classification problem Cross validation for performance measurement.

2.6. Churn prediction Real World Applications

Bank and financial services: churn prediction from balance, tenure and credit score perspective. Telecom and retail: prediction based on transactional and behavioral patterns of churn.

2.7. Advanced Features in Churn Prediction

Real-time churn prediction with real-time analytics Deep learning models like neural networks for better accuracy Hybrid approach that combines traditional algorithms with ensemble techniques like XG Boost.

2.8. Ethical and Business Considerations

Ethical use of customer data for churn prediction Balancing model interpretability with complexity for business take-up Cost-benefit analysis of implementing churn-prediction systems.

2.9. Visualization and Insights

Graphical representation of feature distribution to understand the churn patterns Dashboard design for real-time churn monitoring The Use of Data Storytelling: Presenting Churn Insights in an Effective Manner.

III. SYSTEM ARCHITECTURE

It is very crucial to make the data useful because unwanted or null values can cause unsatisfactory results or may lead to producing less accurate results. In the data set, there are a lot of incorrect values and missing values. We analyzed the whole dataset and listed out only the useful features. The listing of features can result in better accuracy and contains only valuable features as to come up with the specific information like the owner, place of registration, address. pictorial representation of system architecture is shown in which includes various phases, namely, Data pre-processing and feature selection, Splitting of Pre-processed Data into train and test set, training and testing of models respectively.

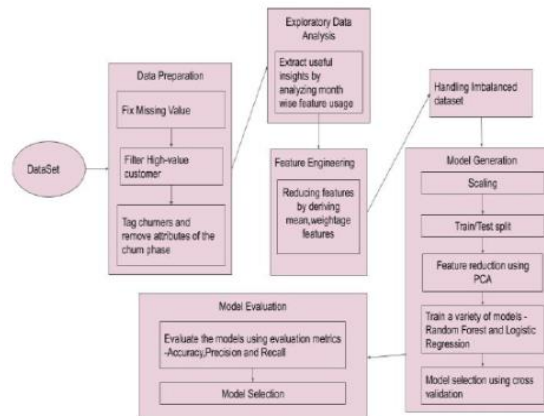


Figure 1: System Architecture

IV. PROPOSED SYSTEM

Millions of pieces of information are created daily by such vast numbers of clients. Decision-makers and business experts believe that keeping them is costly in comparison to acquiring new clients for the organization overall. In order to understand why there are consistently customer attrition and how habits develop from the churned data, CRM teams need consult with business analysts.

The suggested approach identifies the reasons behind customer attrition and detects churn customers using classification and clustering algorithms. The characteristics are categorized using correlation-based attributes scoring filters and information gain. Using classification methods, the algorithm first classifies the churn customer data; RF fared better with 84.023% of correctly categorized occurrences.

To give group-level retention offers, the same churn customers are categorized using cosine similarity after segmentation. It assists CRM teams create targeted retention plans and boosts marketing effectiveness by providing a specific client group with comparable prizes based on identical conduct patterns. Finding the underlying reasons for client turnover is made easier with the help of significant churn indicators. CRM teams may use the data to efficiently develop marketing campaigns and tailor promotions. Criteria like accuracy, reliability, Gradient Boosting and The ROC area are applied to this churn prediction model. In order to incorporate a strategy for maintaining clients and business growth, the model may use an RF algorithm to classify churn and then use the k-means clustering algorithm to further categorize customers into groups. Rules derived from an attribute-selected classification algorithm are then used to further identify the factors that contribute to churn.

V. METHODOLOGY

5.1 Imagery Acquisition & Pre-processing:

For gathering information on customer churn prediction, it has to extract previous customer data from the database of a company. Thousands of records, comprising features like CreditScore, Age, Balance, and Estimated Salary, constitute the dataset. Such pre-processing procedures include handling any

missing values, scaling numerical features, encoding categorical information about gender and geography, and splitting the dataset into the training and testing sets for validation purposes.

5.2 Feature Extraction:

It determines the relevant characteristics of the dataset which contribute to customer attrition. The important features are Credit Score, Age, Tenure, Balance, and other transaction-based variables like NumOfProducts and IsActiveMember. By one-hot encoding, the categorical factors such as gender and geography are encoded. Exited is the target variable, representing customer attrition. There are multiple feature selection techniques used for determining the most influential predictors.

5.3 Model Building:

Then, developed dataset will be used to build some machine learning models such as Random Forest, KNN, and Logistic Regression. All the models will then have their hyperparameters tuned to enhance the capability to predict. Main objectives of the project are to design explainable models and run complex algorithms in comparison.

5.4 Training & Forecasting:

With models, developments on the training set identify patterns and relationships in the data. Forecasts are done on the test set to gauge the performance of the model. Some of the metrics used to judge how good the predictions performed are accuracy score, precision, recall.

5.5 Result Integration:

The best models serve as a guideline for the decision-making process. The model insights may be used proactively in retaining customers and in increasing profitability for the business. Deep learning models further enhance the real-time integration of data, thus making it even more accurate.

VI. EXPERIMENTAL RESULTS

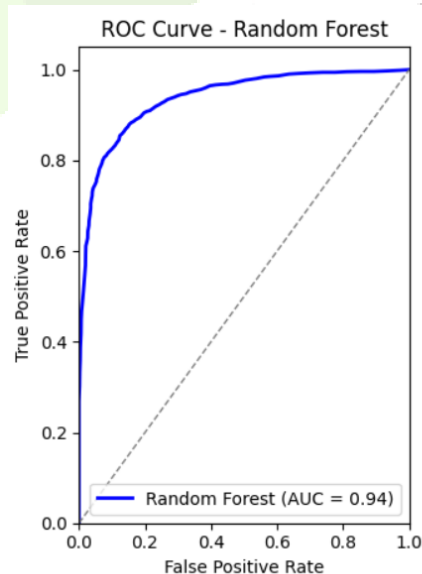


Figure 2: ROC Curve

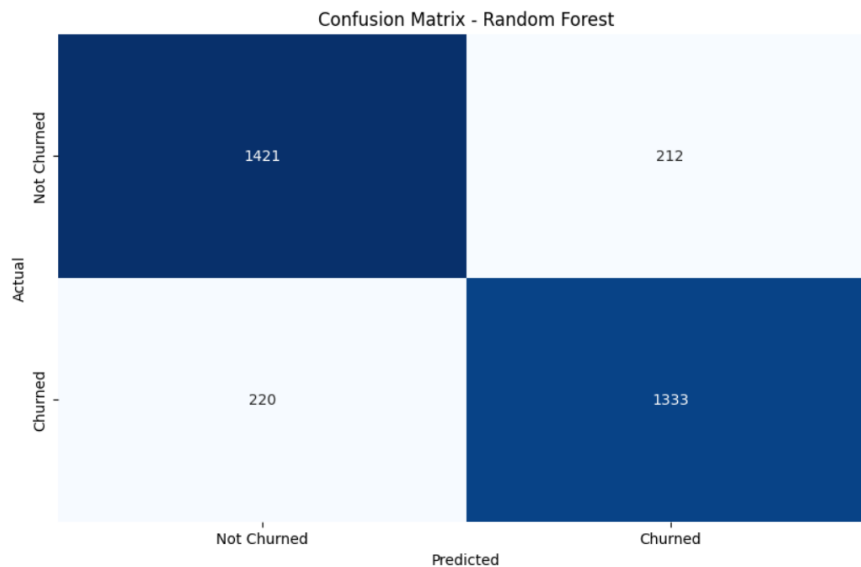


Figure 3: Confusion Matrix

VII. CONCLUSION

To successfully manage the churn prediction challenge, different data mining methods are put forward by different researchers. The main data mining methods are based on neural networks, statistical based techniques, decision trees, covering algorithms, Regression Analysis, K means etc. This paper provides a detailed study on the methods used for the process of customer churn prediction. Each of the above churn prediction models has its own advantages and disadvantages. Hence a good prediction model is required in order to avoid the customer churn problem. This can be achieved by considering a method which can process large inputs with higher dimension and complex attributes for future work for Churn prediction. Good prediction models have to be constantly developed and combination of the proposed methods has to be used. Churn prediction for a company tends to be a very tedious task and as of many upcoming company's and startups there is a tough competition in the market to retain the customers by providing services that are beneficial to both sides. It is very difficult to predict genuine customers of the company. In future, with the upcoming concepts and frameworks in the field of reinforcement learning and deep learning sector, machine learning is proving to be one of the most efficient way to address problems like churn prediction with better accuracy and precision.

VIII. REFERENCE

- [1] T. Vafeiadis, K.I. Diamantaras, G. Sarigiannidis, K. Chatzisavvas" Customer churn prediction telecommunications", Simulation Modelling: Practice and Theory 55 (2015) 1-9
- [2] Abbasimehr H, Setak M, Tarokh M (2011) A neuro-fuzzy classifier for customer churn prediction. International Journal of Computer Applications 19(8):35–41
- [3] Abhishek and Ratnesh, "Predicting Customer Churn Prediction in Telecom Sector Using Various Machine Learning Techniques", In the proceedings of 2017 International Conference on Advanced Computation and Telecommunication, Bhopal, India, 2017.
- [4] Wei CP, Chiu IT (2002) Turning telecommunications call details to churn prediction: a data mining approach. Expert systems with applications 23(2):103–112.
- [5] Sharma H, Kumar S (2016) A survey on decision tree algorithms of classification in data mining. International Journal of Science and Research (IJSR) 5(4):2094–2097