

Predicting Customer Churn Using ML Model

DR V KANIMOZHI¹, ANEESH AHAMED M²

¹Associate Professor, Department of Management Sciences, Hindusthan College of Engineering and Technology.

²Student, Department of management sciences Hindusthan college of engineering and technology

Abstract- This study focuses on customer churn prediction using machine learning techniques to help organizations identify customers who are likely to discontinue services. In today's competitive business environment, retaining customers is more cost-effective than acquiring new ones, making churn management a strategic priority. The research utilizes customer demographic, behavioral, transactional, and interaction data to analyze patterns influencing churn. Various machine learning models, including Logistic Regression, Decision Trees, Random Forest, and Gradient Boosting, are applied and compared to determine the most effective approach. The study follows a structured methodology involving data preprocessing, feature engineering, model training, and evaluation using metrics such as accuracy, precision, recall, and ROC-AUC. The findings aim to provide actionable insights for businesses to develop targeted retention strategies, improve customer satisfaction, and reduce financial losses. Overall, the research contributes to data-driven decision-making and enhances customer relationship management through predictive analytics.

Index Terms- Customer Churn, Machine Learning, Customer Retention, Customer Behavior, Data Analysis, Logistic Regression, Churn prediction model

I. INTRODUCTION

Customer churn has become a critical challenge for organizations operating in highly competitive and dynamic markets. Churn refers to the loss of customers who discontinue using a company's products or services over a given period. With the rapid growth of digital platforms and increasing customer expectations, businesses face constant pressure to retain their existing customers. Studies indicate that acquiring new customers is significantly more expensive than retaining existing ones, making customer retention a key strategic priority.

In the modern data-driven environment, organizations generate vast amounts of customer data

through transactions, service interactions, and digital activities. This data provides valuable insights into customer behavior and preferences. However, traditional analytical methods often fail to capture complex patterns and relationships within such large datasets. As a result, machine learning has emerged as an effective solution for predicting customer churn by identifying hidden trends and risk factors.

This study focuses on applying machine learning techniques to analyze customer data and predict churn behavior. By utilizing models such as Logistic Regression, Decision Trees, Random Forest, and Gradient Boosting, the research aims to classify customers based on their likelihood of leaving. The findings of this study will help organizations develop targeted retention strategies, improve customer satisfaction, and enhance overall business performance through data-driven decision-making.

II. STATEMENT OF THE PROBLEM

- Organizations struggle to accurately identify customers who are likely to churn despite having large volumes of data.
- Traditional analytical methods fail to capture complex behavioral patterns influencing customer attrition. This leads to ineffective retention strategies, financial losses, and reduced customer lifetime value.
- There is a need for an efficient machine learning-based approach to predict churn and support proactive decision-making.

III. NEED FOR THE STUDY

- Increasing competition makes customer retention more important than acquisition
- High customer churn leads to financial losses for organizations

- Businesses have large amounts of customer data but lack proper analysis
- Traditional methods fail to capture complex customer behavior
- Need for accurate prediction of customers likely to churn
- Machine learning provides better insights and prediction accuracy
- Helps in developing targeted and effective retention strategies
- Supports data-driven decision-making and improves business performance

IV. OBJECTIVES OF THE STUDY

Primary Objective:

1. To develop an accurate and reliable machine learning model for predicting customer churn using customer behavioral and demographic data.
Secondary Objectives:
2. To identify the Problem why the customers or reducing randomly by consuming the products.
3. To compare the performance by machine learning algorithms (such as Logistic Regression, Random Forest, Decision Trees, and Gradient Boosting) to determine the most effective model for churn prediction.
4. To analyze the problem to find out a solution to overcome.

SCOPE OF THE STUDY

This study focuses on analyzing customer churn using a descriptive data analysis approach, aiming to understand the underlying patterns, behaviours, and factors contributing to customer attrition. The scope includes examining historical customer data, such as demographic details, usage patterns, complaints, Kand payment behaviour, to identify visible trends and churn characteristics. Descriptive analytics is used to summarize the data through charts, frequency distributions, and correlation insights without applying predictive or causal modelling. The study is limited to identifying what is happening and how churn is distributed across customer segments, rather than predicting future churn or establishing cause-effect relationships. This scope supports organizations in understanding churn profiles, highlighting key problem areas, and forming a baseline for future predictive or prescriptive analyses.

V. RESEARCH METHODOLOGY

Research Design: The study adopts a Quantitative research design focused on analyzing numerical customer data using statistical and machine learning techniques. It includes Exploratory Data Analysis (EDA) to identify patterns, trends, and relationships influencing customer churn. The design is predictive in nature, aiming to classify customers into churners and non-churners using various machine learning models.

Sample Size: The sample size is derived from the organization's CRM and operational datasets, consisting of a large number of customers (typically 1000+ records) with sufficient demographic, transactional, and behavioural data. The dataset is further split into:

- Training set (70%)
- Validation set (15%)
- Testing set (15%)

Sampling Technique:

- Stratified Random Sampling – to ensure representation across customer segments (plan, region, tenure)
- Systematic Sampling (optional) – for large datasets

Data Sources:

The research primarily uses secondary data, including:

- Customer demographic data
- Transaction and billing records
- Customer support and complaint records
- CRM databases
- Historical churn data

Analysis Tools:

The study uses advanced analytical tools and technologies:

- Python (main programming tool)
 - Pandas & NumPy – data processing
 - Matplotlib & Seaborn – data visualization
 - Scikit-learn – machine learning models
- Machine learning models applied:
- Logistic Regression

Evaluation metrics:

- Accuracy

VI. LIMITATIONS OF THE STUDY

This study has several limitations that may affect the generalizability and accuracy of the findings. First, the research relies primarily on secondary data, which may contain missing values, inconsistencies, or limited variables, reducing overall data quality. Second, the study focuses on historical customer data, which may not fully capture dynamic changes in customer behavior over time. Third, the results are dependent on the selected machine learning models, and different algorithms or parameter settings may produce varying outcomes. Additionally, class imbalance in churn data may influence model performance despite the use of balancing techniques. The study is also limited to specific industries or datasets, which restricts its applicability across different sectors. Furthermore, external factors such as economic conditions, competitor strategies, and changing customer preferences are not fully considered in the analysis. Lastly, the interpretability of complex machine learning models can be challenging, making it difficult for managers to directly understand and apply the insights in practical decision-making processes.

Findings of the Study

- Customer churn is mainly influenced by behavioral factors such as low usage frequency and reduced engagement.
- High number of complaints and poor customer service experience significantly increase churn probability.
- Transactional factors like delayed payments and plan downgrades are key indicators of churn.
- Demographic variables such as customer tenure and location have a moderate impact on churn behavior.
- Exploratory Data Analysis (EDA) reveals clear patterns differentiating churned and non-churned customers.
- Machine learning models, especially Random Forest and Gradient Boosting, provide higher accuracy than traditional methods.

- The study confirms that predictive analytics helps identify high-risk customers and supports effective retention strategies.

Suggestions

- Organizations should implement machine learning-based churn prediction systems for early identification of high-risk customers.
- Customer complaints must be addressed quickly through improved support services and faster issue resolution.
- Companies should offer personalized retention strategies such as discounts, loyalty rewards, and special offers.
- Regular monitoring of customer usage patterns and payment behaviour should be carried out to detect churn signals.
- Businesses should continuously update and improve their datasets to enhance prediction accuracy and decision-making.
- Future studies should include advanced machine learning techniques and real-time data analysis for better churn prediction.

REFERENCES

- [1] Abbas Keramati, H. G. (2016). Developing a prediction model for. Keramati et al. *Financial Innovation*, 13.
- [2] Damandeep Singh, V. A. (2021). Survey Paper on Churn Prediction on Telecom. SSRN, 9.
- [3] David L. Garcia, A. N. (2017). *Intelligent Data Analysis Approaches to*. SPRINGER NATURELINK, 47.
- [4] Eva Ascarza, S. A. (2018). In Pursuit of Enhanced Customer Retention Management: Review, Key Issues, and Future. SSRN, 46.
- [5] JAEHYUN AHN, J. H. (2020). A Survey on Churn Analysis in Various. *IEEE*, 24.
- [6] Krishnakumar, S. A. (2023). A review of churn prediction models using different machine learning and deep learning approaches in cloud environment. *Journal of Current Science and Technology*, 26.
- [7] Maryam Shahabikargar, A. B. (2025). A comprehensive survey on customer churn analysis studies. *JOURNAL OF*

INFORMATION AND TELECOMMUNICATIONS, 47.

- [8] PEDRO SOBREIRO, D. D. (2022). A SLR on Customer Dropout Prediction. IEEE, 19.
- [9] Sucki, O. (2019). Predicting the customer churn with machine learning methods - CASE: private. LUT University, 80.
- [10] Uday Bhale, H. S. (2024). Customer Churn Construct: Literature Review and Bibliometric Study. SSRN, 12.