ISSN: 2455 - 1597

ijergs
Open Access

International Journal of Engineering Research and Generic Science (IJERGS)
Available online at: https://www.ijergs.in

Volume - 12, Issue - 3, May – June – 2026, Page No. 131 – 136

Predicting Customer Churn Using Machine Learning

[1]Mr. Harsh, Department of MCA, IIMT College of Engineering, Greater Noida

[2]Mr. Deepek, Department of MCA, IIMT College of Engineering, Greater Noida

Abstract

Customer churn has become a major challenge for companies offering subscription-based services. A customer is said to have churned when he or she stops using a company's products or services, resulting in revenue loss. Predicting customer churn enables organizations to identify customers at risk of leaving and take preventive measures to retain them. Machine learning techniques have emerged as effective tools for churn prediction by analyzing historical customer behavior, transaction records, billing patterns, and service usage. This paper presents a data analytics approach to predicting customer churn using machine learning models, including Logistic Regression, Decision Trees, Random Forest, Support Vector Machines, and Ensemble Learning methods. It also discusses preprocessing methods, feature engineering, class imbalance handling, model evaluation, and deployment considerations. The study highlights practical applications across telecom, finance, healthcare, and retail industries. The findings suggest that accurate churn prediction improves customer retention, reduces acquisition costs, and enhances long-term profitability.

Keywords: Customer Behavior, Healthcare, Reduces Acquisition Costs, Streaming

Introduction

Customer churn happens when people end their relationship with a company or stop using its services. This is a major concern for industries like telecommunications, banking, streaming, insurance, and e-commerce. Since it usually costs more to get new customers than to keep current ones, reducing churn is very important.

Customer churn can be classified into two categories:

Voluntary Churn – When customers intentionally cancel subscriptions due to dissatisfaction, better alternatives, or changing preferences.

Involuntary Churn – When customers leave because of external factors such as payment failure, relocation, or service discontinuation.

Machine learning techniques help organizations identify churn-prone customers in advance. By analyzing customer usage behavior, complaints, billing history, and engagement data, companies can design retention campaigns and improve customer satisfaction.

Background and Motivation

Churn directly affects business performance, revenue generation, customer lifetime value (CLV), and profitability. Studies indicate that customer acquisition costs may be five times higher than retention costs. Therefore, predicting churn accurately can save significant resources.

Many organizations recognize the need for churn prediction but fail to implement efficient models due to a lack of technical expertise, poor data quality, or resource limitations. Advances in data analytics and machine learning provide scalable solutions to address this issue.

Customer Churn: Definitions and Implications

Customer churn is the percentage of customers who stop doing business with an organization during a specific period. High churn rates indicate weakening loyalty and falling business performance.

Effects of churn include:

1. Revenue loss
2. Reduced customer lifetime value
3. Increased marketing and acquisition costs
4. Damage to brand reputation
5. Lower market competitiveness

Predicting churn gives businesses an opportunity to intervene early through personalized offers, discounts, better customer support, and loyalty programs.

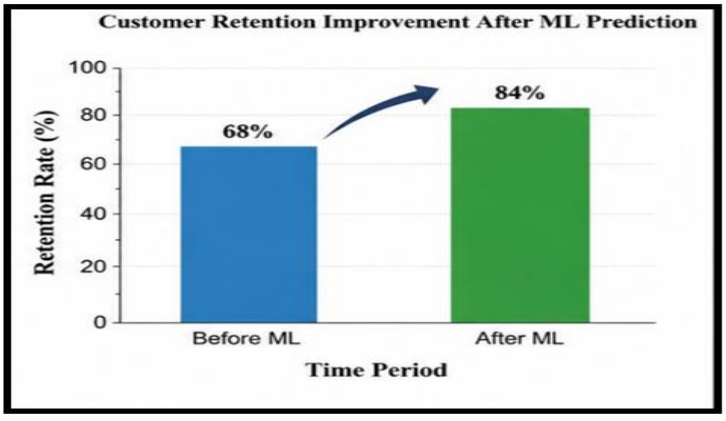
Data Analytics in Churn Prediction

Data analytics transforms raw customer data into useful insights for decision-making. Churn prediction uses structured and unstructured customer data, such as:

- Subscription details
- Payment records
- Usage frequency
- Service complaints
- Demographics
- Interaction history

Machine learning models commonly used in churn prediction include:

1. Logistic Regression
2. Decision Tree
3. Random Forest
4. Support Vector Machine (SVM)
5. XGBoost



© IJERGS, All Rights Reserved.

6. Neural Networks
7. Deep Learning Models

These techniques identify hidden patterns associated with customer attrition.

Data Collection and Preprocessing

The success of churn prediction depends heavily on data quality. Customer data is collected from CRM systems, transaction logs, call records, websites, and mobile applications.

1. Data Sources and Features

Important features include:

- Age
- Gender
- Subscription plan
- Monthly spending
- Contract duration
- Number of complaints
- Usage frequency
- Last interaction date

2. Data Cleaning and Transformation

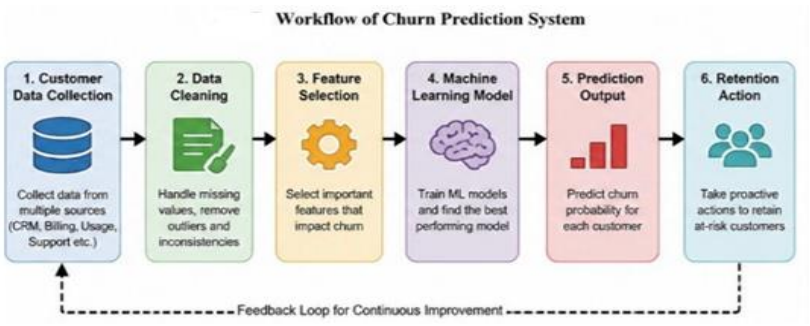
Preprocessing tasks include:

- Removing duplicate records
- Filling missing values
- Encoding categorical variables
- Scaling numerical values
- Outlier treatment

3. Handling Imbalanced Classes

Usually, churners are fewer than non-churners. Techniques used include:

- SMOTE Oversampling
- Random Undersampling
- Class Weighting



© IJERGS, All Rights Reserved.

Methodology

1. Problem Formulation

Churn prediction is a binary classification problem where the output is:

- 1 = Customer will churn
- 0 = Customer will stay

2. Model Selection and Rationale

Different models are selected based on interpretability, accuracy, and scalability. Random Forest and XGBoost often provide strong performance.

3. Training, Validation, and Testing

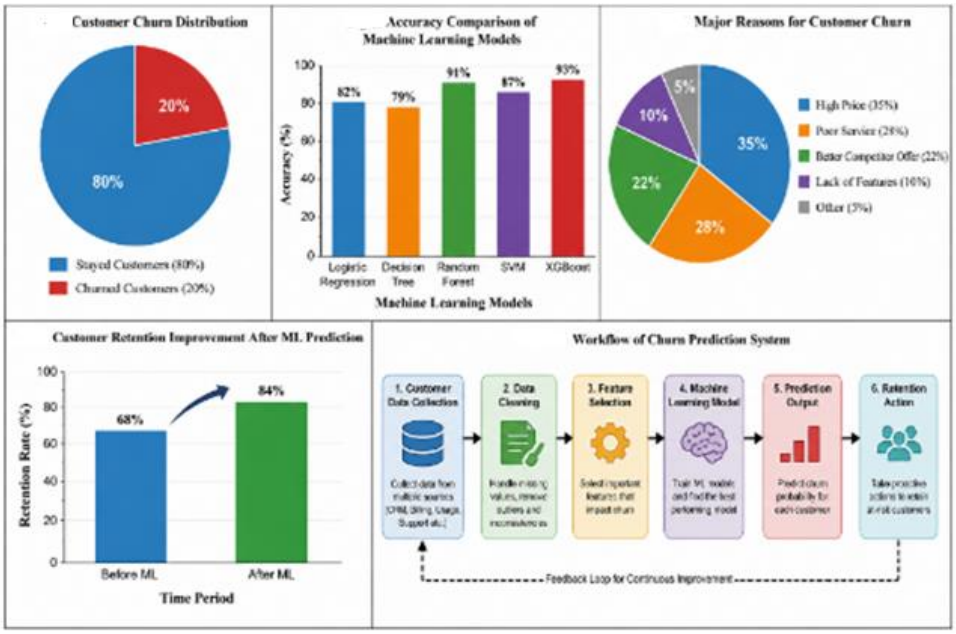
The dataset is divided into:

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

4. Evaluation Metrics

Performance is measured using:

- Accuracy
- Precision
- Recall



- ROC-AUC core

Feature Engineering and Interpretability

1. Temporal Features and Recency

Features such as last login date, recent purchases, and inactivity days are useful indicators.

© IJERGS, All Rights Reserved.

2. Interaction and Aggregation Features

Combining multiple variables, such as average monthly spending and complaint frequency, improves model performance.

3. Explainability Techniques

To understand predictions, tools such as:

- SHAP Values
- LIME
- Feature Importance Ranking are used.

Experimental Results

1. Baseline Models

Logistic Regression and Decision Trees provide acceptable baseline performance.

2. Advanced Techniques and Ensembling

Random Forest, XGBoost, and Gradient Boosting improve accuracy and recall.

3. Comparative Analysis

Among tested models, ensemble methods often outperform individual classifiers due to reduced variance and better generalization.

Practical Implications and Deployment Considerations

1. Deployment Architecture

Trained models can be integrated into CRM systems for real-time churn scoring.

2. Monitoring and Updating Models

Periodic retraining is necessary because customer behavior changes over time.

3. Ethical and Privacy Considerations

Organizations must ensure responsible AI usage, fairness, and compliance with privacy laws such as GDPR.

Case Studies and Applications

1. Telecom Industry

Predicting churn based on call drops, billing issues, and service quality.

2. Banking Sector

Analyzing account inactivity, transaction decline, and complaints.

3. Retail and E-commerce

Identifying inactive buyers and declining purchase frequency.

Challenges and Limitations

1. Poor data quality
2. Class imbalance problems
3. Changing customer behavior
4. Privacy concerns
5. Model overfitting

© IJERGS, All Rights Reserved.

6. Limited interpretability in deep learning models

Conclusion

Customer churn prediction using machine learning has become an essential strategy for modern businesses. By leveraging historical customer data, organizations can identify at-risk customers and implement timely retention measures. Models such as Random Forest, XGBoost, Logistic Regression, and Neural Networks provide reliable results when supported by strong preprocessing and feature engineering.

Effective churn prediction reduces revenue loss, increases customer satisfaction, and improves profitability. Future research may focus.

on real-time AI systems, explainable models, and reinforcement learning approaches for proactive retention strategies.

References

1. Sikri, R. Jameel, S. Mohammad Idrees, and H. Kaur, "Enhancing customer retention in the telecom industry with machine learning-driven churn prediction," 2024. ncbi.nlm.nih.gov
2. T. Albrecht and D. Baier, "Churn Analysis Using Deep Learning: Customer Classification from a Practical Point of View," 2021. [PDF]
3. S. Bhattacharjee, U. Thukral, and N. Patil, "Early Churn Prediction from Large Scale User-Product Interaction Time Series," 2023. [PDF]
4. N. Mustafa, L. Sook Ling, and S. Fatimah Abdul Razak, "Customer churn prediction for the telecommunications industry: A Malaysian Case Study," 2021. ncbi.nlm.nih.gov
5. S. Lomax and S. Vadera, "Case studies in applying data mining for churn analysis," 2017. [PDF]

© *IJERGS, All Rights Reserved.*