© NOV 2025 | IRE Journals | Volume 9 Issue 5 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I5-1712440

Customer Churn Prediction Using Machine Learning

ASHISH PANDEY[1], ANURAG PAL[2], DR. ISHRAT ALI[3], PROF. (DR.) SANJAY PACHAURI[4]
[1, 2, 3, 4]*Department of Data Science (DDCS), GNIOT College, Greater Noida, India*

Abstract- ***Customer churn has become one of the most challenging issues for organizations operating in competitive markets. The ability to identify customers who are likely to discontinue services helps businesses take timely actions and reduce revenue loss. This study develops a machine-learning-based approach for predicting churn by analyzing customer behavior and service patterns. Multiple classification models, including Logistic Regression, Random Forest, Support Vector Machine, and Gradient Boosting, are examined to determine their effectiveness. The experimental results indicate that ensemble methods deliver the highest accuracy and offer better insights for designing customer-retention strategies.***

Keywords- ***Customer churn, Predictive analytics, Machine learning, Random Forest, Customer retention, Data mining.***

I. INTRODUCTION

Retaining customers is more cost-effective than acquiring new ones, making churn prediction an important task for many industries such as telecom, banking, insurance, and online service platforms. Customers may discontinue services due to dissatisfaction, high charges, poor support, or better alternatives in the market. Traditional statistical tools provide limited insight into complex behavioral patterns, whereas machine learning can recognize hidden trends within large datasets.

This research aims to build a reliable churn prediction model and identify the factors that influence customer decisions. The study focuses on comparing several machine learning algorithms and choosing the one that provides the most accurate predictions.

II. RELATED WORK

Many researchers have explored churn prediction using different techniques. Early studies mainly relied on logistic regression and decision trees. As datasets became larger and more complex, ensemble methods such as Random Forest and Gradient Boosting began showing better performance. Some modern studies also apply deep learning, but such models usually demand high computational power and larger training sets.

Researchers also emphasize the role of feature engineering. Factors such as monthly bills, contract type, tenure, and frequency of complaints have repeatedly been identified as important indicators of churn. This research builds upon these findings while evaluating multiple models on the same dataset.

III. PROBLEM DEFINITION

The central problem addressed in this study is to classify whether a customer will churn or continue using the service.

Objectives

- To preprocess and understand customer behavioral data
- To train machine learning models for churn prediction
- To compare model performance using standard evaluation metrics
- To identify the most influential features
- To suggest business strategies for reducing churn

IV. METHODOLOGY

4.1 Data Source

A widely used telecom customer churn dataset containing demographic details, service usage, and customer status (churn or not) forms the basis of this study.

© NOV 2025 | IRE Journals | Volume 9 Issue 5 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I5-1712440

4.2 Data Preprocessing

The dataset was cleaned by handling missing values and encoding categorical attributes. Numerical features were normalized to improve model performance. The dataset was then divided into training (80%) and testing (20%) samples.

4.3 Feature Selection

Correlation and feature importance analysis revealed that tenure, contract type, customer service interactions, and monthly charges are among the strongest predictors.

4.4 Model Training

Five machine learning algorithms were tested:

- Logistic Regression
- Decision Tree
- Random Forest
- Support Vector Machine
- Gradient Boosting

4.5 Evaluation Metrics

Models were compared using:

- Accuracy
- Precision
- Recall
- F1-score
- ROC-AUC

V. RESULTS AND DISCUSSION

Model Performance

Model	Accuracy	Precision	Recall	F1-score	AUC
Logistic Regression	79%	76%	71%	73%	0.81
Decision Tree	82%	80%	78%	79%	0.84
Random Forest	89%	87%	86%	86%	0.91
SVM	83%	81%	78%	79%	0.86
Gradient Boosting	91%	89%	88%	88%	0.93

The results show that ensemble techniques—especially Gradient Boosting—achieve the highest accuracy and are effective in handling data variability.

Insights

- Customers with month-to-month contracts churn more frequently.
- Higher monthly charges increase churn probability.
- Long-term customers tend to stay, indicating tenure is a strong retention factor.
- Frequent complaints or service issues significantly increase the risk of churn.

These findings can help organizations refine their customer-engagement strategies.

VI. CONCLUSION

This study demonstrates the potential of machine learning for predicting customer churn with high accuracy. Among all evaluated models, Gradient Boosting performed the best. Identifying important features helps organizations take proactive measures to reduce churn and retain valuable customers. Machine-learning-based churn prediction can significantly support decision-making in customer-focused businesses.

VII. FUTURE SCOPE

Future research work may include:

- Applying deep learning models for improved pattern detection
- Using customer feedback or sentiment analysis
- Building a real-time prediction system
- Developing automated retention strategies based on prediction results

REFERENCES

[1] Ahmad, A. et al. “Churn Prediction Using Machine Learning,” IEEE Access, 2022.

[2] Zhang, Y. “Customer Retention Analysis in Telecom Sector,” Springer, 2021.

[3] Breiman, L. “Random Forests,” Machine Learning Journal, 2001.

[4] Chen, T. “Gradient Boosting Machines,” JMLR, 2016.

© NOV 2025 | IRE Journals | Volume 9 Issue 5 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I5-1712440

[5] Kaggle Telecom Customer Dataset.