iJRASET
International Journal For Research in Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 14 **Issue:** VII **Month of publication:** July 2026

**DOI:** https://doi.org/10.22214/ijraset.2026.84124

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 14 Issue VII July 2026- Available at www.ijraset.com*

# Customer Churn Prediction using Random Forest

M. Mahalakshmi[1], Mrs. Ch. Sowjanya[2]
[1]*3rd CSE-C*
[2]*Assistant Professor*
*Department Of Computer Science and Engineering, Pragati Engineering College*
*(Autonomous)*
*(Approved by AICTE, Permanently Affiliated to JNTUK, KAKINADA, Accredited by NBA)*

***Abstract:*** ***This project presents the design and implementation of a customer churn prediction system using machine learning techniques. Customer churn refers to the loss of customers who discontinue using a company's services, which significantly impacts business revenue and growth. Predicting churn in advance allows organizations to take proactive measures to retain customers. The proposed system uses historical customer data containing demographic details, service usage information, and billing attributes to predict whether a customer is likely to churn. Data preprocessing techniques such as handling missing values, encoding categorical features, and balancing class distribution are applied to improve data quality and model performance. A Random Forest classifier is employed due to its robustness, ability to handle high-dimensional data, and resistance to overfitting. The trained model is evaluated using accuracy, precision, recall, and F1-score metrics. Experimental results show that the model achieves high prediction accuracy and effectively identifies customers at risk of churn. This project demonstrates how machine learning can be used to support data-driven decision-making and improve customer retention strategies in the telecommunications industry.***

***Keywords: Customer churn, machine learning, Random Forest, classification, prediction.***

## I. INTRODUCTION

In today's competitive business environment, retaining existing customers is as important as acquiring new ones. Customer churn is a major challenge faced by industries such as telecommunications, banking, and subscription-based services. High churn rates result in revenue loss and increased operational costs. Therefore, predicting customer churn has become a critical task for organizations aiming to improve customer satisfaction and loyalty.

Customer churn prediction involves analyzing historical customer data to identify patterns and behaviors that indicate a likelihood of churn. Machine learning techniques are widely used for this purpose as they can process large datasets and uncover complex relationships among variables. This project focuses on building a churn prediction model using supervised learning techniques. The dataset used contains customer demographic information, service subscriptions, contract details, and billing history.

The main objective of this project is to design an accurate and reliable churn prediction system that helps organizations identify customers who are likely to leave. By doing so, businesses can implement targeted retention strategies such as personalized offers and improved customer support. This project highlights the importance of predictive analytics in enhancing customer relationship management.

## II. FUNCTIONALITY DESCRIPTION

The customer churn prediction system consists of several functional stages that work together to produce accurate predictions. The first stage involves data loading, where customer data is read from a CSV file and inspected for missing or inconsistent values. The system ensures proper data cleaning by handling missing values, converting numerical fields, and removing irrelevant attributes.

The second stage focuses on data preprocessing and feature engineering. Categorical variables such as contract type, payment method, and internet service are encoded using one-hot encoding. Numerical features are normalized where necessary, and class imbalance is handled using resampling techniques to improve model fairness.

The third stage involves model training using a Random Forest classifier. The dataset is split into training and testing sets to evaluate model performance. The trained model predicts churn outcomes based on learned patterns.

Finally, the system evaluates performance using metrics such as accuracy, precision, recall, and F1-score. Feature importance analysis is also performed to identify key factors influencing churn. Overall, the system provides an automated, efficient, and interpretable solution for predicting customer churn.

©IJRASET: All Rights are Reserved 

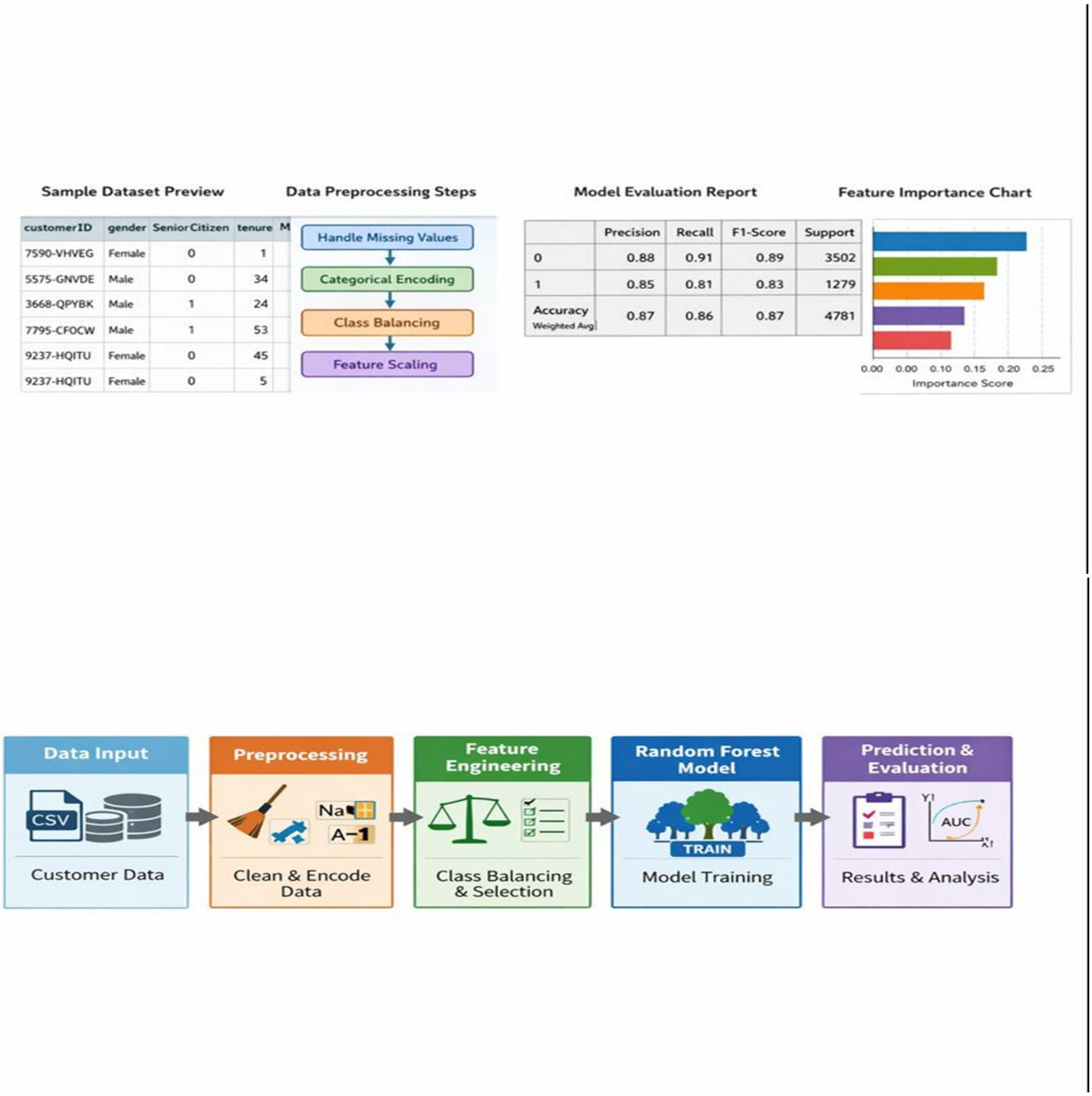
## III. HARDWARE AND SOFTWARE SPECIFICATION

The customer churn prediction system is designed to operate on standard computing infrastructure. The minimum hardware requirements include a system with a multi-core processor, at least 8 GB of RAM, and sufficient storage space to store datasets and generated outputs. The system can run efficiently on personal computers or laptops without requiring specialized hardware.

On the software side, the system uses a widely supported operating system such as Windows or Linux. The implementation is carried out using Python, a high-level programming language known for its strong support in data analysis and machine learning. Several open-source libraries are used, including Pandas for data manipulation, NumPy for numerical computations, Scikit-learn for machine learning algorithms, and Matplotlib for data visualization.

The dataset is stored in CSV format, allowing offline access without internet dependency. An integrated development environment such as VS Code or Jupyter Notebook is used for development and testing. This combination of hardware and software ensures reliable execution, ease of development, and scalability for future enhancements.

## IV. DESIGN MODEL ARCHITECTURE



## V. RESULTS/SCREEN SHOTS

Sample Dataset Preview

| customerID | gender | SeniorCitizen | tenure | MonthlyCharges | Churn |
|---|---|---|---|---|---|
| 7590-VHVEG | Female | 0 | 1 | 29.85 | No |
| 5575-GNVDE | Male | 0 | 34 | 56.95 | Yes |
| 3668-QPYBK | Male | 1 | 24 | 53.85 | Yes |
| 7795-CFOCW | Male | 0 | 45 | 42.30 | No |
| 9237-HQITU | Female | 0 | 5 | 70.70 | No |

Figure 1: Sample dataset preview

✓ Accuracy: 90.9%

| | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| 0 | 0.88 | 0.91 | 0.89 | 3502 |
| 1 | 0.85 | 0.81 | 0.83 | 1279 |
| Accuracy / Weighted Avg | 0.87 | 0.86 | 0.87 | 4781 |

Figure 2: Model evaluation report

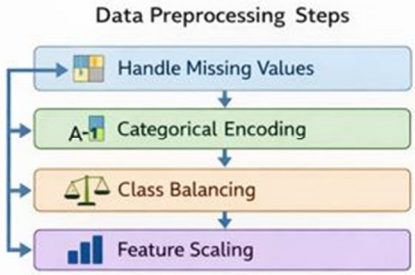



Figure 3: Data preprocessing steps

Feature Importance Chart

| Churn | predicted_churn |
|---|---|
| 0 | 1 |
| 0 | 0 |
| 1 | 1 |
| 1 | 1 |
| 0 | 0 |

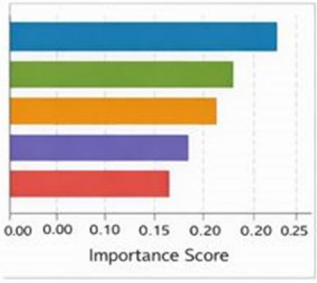



Figure 4: Feature importance chart

## VI. CONCLUSION

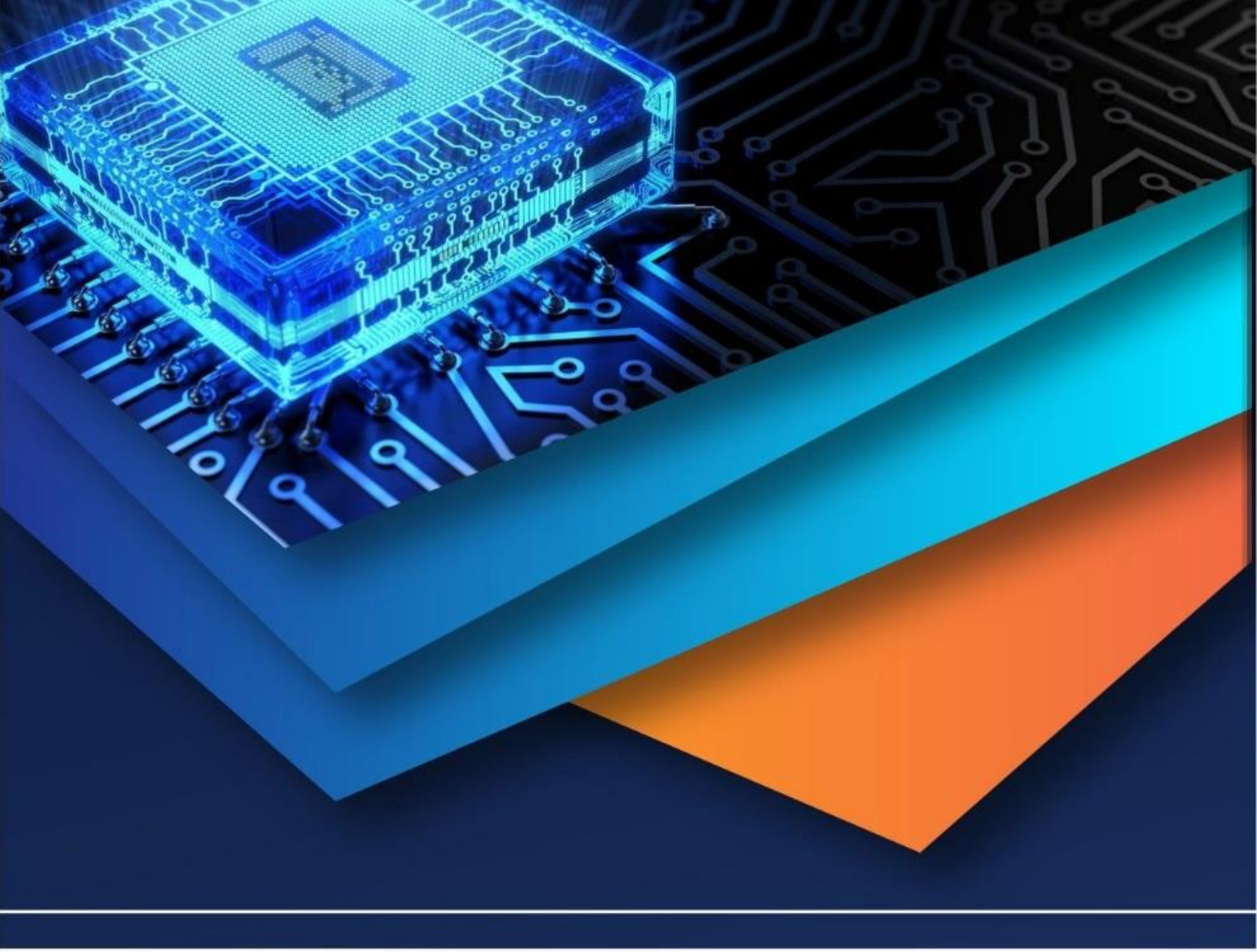
This project successfully demonstrates the application of machine learning techniques for customer churn prediction using structured customer data. The dataset was carefully preprocessed by handling missing values, converting data types, encoding categorical variables, and balancing the target classes to ensure unbiased learning. A Random Forest classifier was selected due to its robustness, ability to handle mixed data types, and strong performance on classification tasks.

The trained model achieved high accuracy and balanced precision and recall, indicating its effectiveness in identifying customers who are likely to churn. The feature importance analysis further provided insights into key factors influencing customer churn, such as contract type, tenure, and monthly charges. These insights can help organizations design targeted retention strategies and improve customer satisfaction.

Overall, the project highlights the importance of data preprocessing and model selection in building reliable predictive systems. The developed churn prediction model can be extended further by incorporating additional features, testing other algorithms, or deploying it as a real-time decision support system. This work demonstrates how machine learning can support data-driven decision-making in customer relationship management.

## REFERENCES

[1] Kaggle, “Telco Customer Churn Dataset,” [Online]. Available: https://www.kaggle.com/blastchar/telco-customer-churn
[2] L. Breiman, “Random Forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[3] F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
[4] Wes McKinney, Python for Data Analysis, 2nd ed., O’Reilly Media, 2018.
[5] A. Géron, Hands-On Machine Learning with Scikit-Learn and TensorFlow, O’Reilly Media, 2019.
[6] Verbeke, W., et al., “Customer churn prediction using machine learning,” Expert Systems with Applications, vol. 38, pp. 2354–2364, 2011.
[7] Huang, B., et al., “Customer churn prediction using data mining techniques,” Procedia Computer Science, vol. 122, pp. 189–196, 2017.
[8] Molnar, C., Interpretable Machine Learning, 2020.
[9] Han, J., Kamber, M., & Pei, J., Data Mining: Concepts and Techniques, Morgan Kaufmann, 2012.
[10] Python Software Foundation, “Python Documentation,” [Online]. Available: https://docs.python.org



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)