

Customer Churn Prediction Using Machine Learning: A Comparative Study

Abstract

Customer churn prediction is a critical task for subscription-based businesses, as retaining existing customers is often more cost effective than acquiring new ones. This study presents a comprehensive machine learning based framework for predicting customer churn using baseline classification models, benchmarking their performance, and analyzing feature importance. The work focuses on understanding customer behavior patterns that lead to churn and evaluating the effectiveness of various machine learning algorithms. Performance is assessed using standard classification metrics, and explainability is incorporated through feature importance analysis to provide actionable business insights.

Introduction

In highly competitive markets such as telecommunications, customer retention has become a major challenge. Customer churn refers to the phenomenon where customers discontinue a service within a given time period. Predicting churn enables organizations to proactively identify at risk customers and take preventive measures. With the availability of large scale customer data, machine learning techniques have gained prominence in churn prediction tasks. This study explores multiple baseline machine learning models, compares their performance, and investigates key factors influencing churn decisions.

Objectives

The primary objectives of this study are:

- To analyze customer churn behavior using historical customer data.
- To build baseline machine learning models for churn prediction.
- To compare model performance using standard evaluation metrics.
- To identify and interpret the most influential features contributing to customer churn.
- To provide insights that can support data-driven retention strategies.

Dataset Description

The dataset used in this study is the Telco Customer Churn dataset, which contains information about customers of a telecommunications company. The dataset includes demographic details, service usage information, contract types, billing methods, and churn labels. Key attributes include tenure, monthly charges, total charges, contract type, payment method, and service subscriptions. The target variable indicates whether a customer has churned.

Dataset Link - [Click Here](#)

Methodology

The methodology followed in this study consists of several stages. First, data preprocessing is performed, including handling missing values, encoding categorical variables, and feature scaling. Exploratory data analysis is conducted to understand data distributions and churn patterns.

Next, multiple baseline machine learning models such as Logistic Regression, Decision Trees, Random Forests, and Support Vector Machines are trained on the dataset. Model performance is evaluated using metrics including accuracy, precision, recall, F1-score, and ROC-AUC.

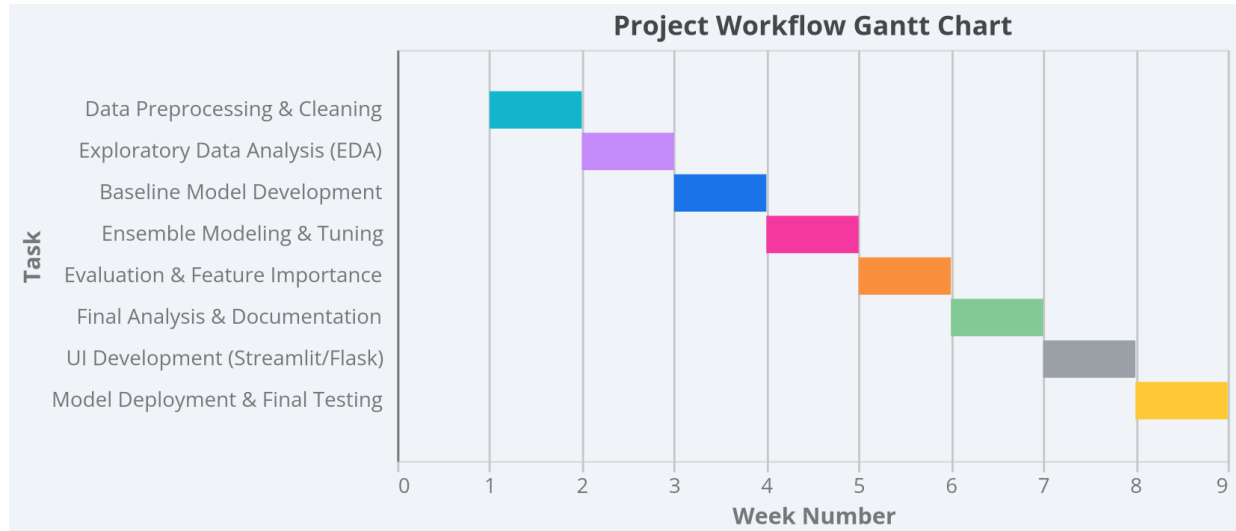
Finally, feature importance analysis is carried out using model specific techniques to identify key drivers of customer churn and enhance model interpretability.

Expected Results and Conclusion

This study is expected to identify a robust machine learning model capable of accurately predicting customer churn, with ensemble based approaches such as Random Forests anticipated to outperform simpler baseline models. Feature importance analysis is expected to highlight critical factors influencing churn behavior, including contract type, customer tenure, monthly charges, and payment methods.

Overall, the findings demonstrate the effectiveness of machine learning techniques for customer churn prediction. By systematically benchmarking multiple models and incorporating feature importance analysis, the proposed framework not only achieves reliable predictive performance but also provides interpretable and actionable insights. These results can support data driven decision making in customer relationship management and enable organizations to design targeted and effective customer retention strategies.

Workflow



References

- [1] Amin, Adnan, et al. "Customer churn prediction in telecommunication industry using data certainty." *Journal of Business Research* 94 (2019): 290-301.
- [2] Lu, Ning, et al. "A customer churn prediction model in telecom industry using boosting." *IEEE Transactions on Industrial Informatics* 10.2 (2012): 1659-1665.
- [3] Singh, Adesh Kumar, et al. "Customer Churn Prediction in Telecom Industry Using Data Mining Techniques."
- [4] Dahiya, Kiran, and Surbhi Bhatia. "Customer churn analysis in telecom industry." *2015 4th International conference on reliability, infocom technologies and optimization (ICRITO)(trends and future directions)*. IEEE, 2015.
- [5] Umayaparvathi, V., and K. Iyakutti. "A survey on customer churn prediction in telecom industry: Datasets, methods and metrics." *International Research Journal of Engineering and Technology (IRJET)* 3.04 (2016): 1065-1070.