

Predictive Modeling for Customer Churn Using Behavioral Engagement Signals

Abstract

Customer churn significantly impacts revenue in subscription-based and service-driven businesses. This project presents an end-to-end machine learning pipeline designed to predict customer churn using behavioral and engagement-based signals. The solution integrates preprocessing, feature engineering, model training, evaluation, and deployment through a Streamlit web application.

1. Introduction

Customer churn refers to customers discontinuing services or reducing engagement to the extent that they no longer provide value. Early detection of churn enables proactive retention strategies. This project aims to build an interpretable and deployable churn prediction model.

2. Problem Statement

Given historical customer data with churn labels, the objective is to predict the probability that a customer will churn. The model must handle missing data, mixed feature types, and class imbalance while maintaining interpretability.

3. Methodology

The dataset is preprocessed by separating numerical and categorical features. Numerical features undergo median imputation and standard scaling, while categorical features use most-frequent imputation and one-hot encoding. Logistic Regression with class balancing is selected due to its interpretability and probability output capability.

4. Evaluation Metrics

- Accuracy
- Precision
- Recall
- F1-Score

- ROC-AUC
- Confusion Matrix

5. Deployment

The trained model is deployed using Streamlit, enabling users to upload customer data, generate churn probabilities, adjust decision thresholds, and download predictions. This ensures accessibility and practical application for business teams.

6. Conclusion

This project demonstrates a complete churn prediction pipeline from model development to deployment. By leveraging engagement-based signals and probability-driven outputs, organizations can prioritize high-risk customers and implement proactive retention strategies.