

CUSTOMER CHURN PREDICTION USING MACHINE LEARNING

ABSTRACT

Customer churn is a major challenge in the telecom industry due to intense market competition. This project focuses on predicting customer churn using Logistic Regression and deriving actionable business strategies. SHAP (SHapley Additive Explanations) was used for model interpretability, while SQL-based aggregation provided business insights across contract types and revenue categories. Customers were segmented into At Risk, Loyal, and Dormant groups to enable targeted retention strategies. The model achieved approximately 80% accuracy, demonstrating effective predictive capability.

INTRODUCTION

Customer retention is more cost-effective than customer acquisition. In the telecom sector, understanding churn drivers is essential for sustainable growth. Churn occurs when customers discontinue services due to pricing, service dissatisfaction, or better alternatives. This project aims to build a predictive model, analyze churn drivers using explainable AI, apply SQL aggregation for business insights, and segment customers for strategic decision-making.

TOOLS USED

- 1 Python (Pandas, NumPy)
- 2 Scikit-learn (Logistic Regression)
- 3 SHAP (Model Explainability)
- 4 SQLite (SQL Simulation)
- 5 Matplotlib & Seaborn (Visualization)
- 6 Jupyter Notebook

STEPS INVOLVED

1. Data Preprocessing: Loaded dataset, removed irrelevant columns, handled missing values, and performed one-hot encoding for categorical variables.
2. Model Development: Split dataset into training and testing sets, applied scaling, trained Logistic Regression model, and evaluated using accuracy and confusion matrix.
3. SHAP Analysis: Identified key churn drivers including Tenure Months, Total Charges, Contract Type, and Internet Service.

4. SQL Aggregation: Executed SQL queries to analyze churn by contract, revenue, and servicecategory.
5. Customer Segmentation: Classified customers into At Risk, Loyal, and Dormant segmentsbased on churn probability and tenure.

CONCLUSION

The churn prediction model achieved approximately 80% accuracy and successfully identified major churn drivers using SHAP explainability. SQL-based aggregation enhanced business-level insights, while customer segmentation enabled targeted retention strategies. By integrating machine learning, explainable AI, and SQL analytics, this solution provides both predictive accuracy and actionable business intelligence for telecom organizations.