Telecom Customer Churn Prediction and Analysis Project Report

Prepared by Anurag Chawla

Email anuragchawla7777@gmail.com

GitHub Repository https://github.com/anuragchawla27/customer-churn-prediction

Project Overview

This project presents an end to end analytics and machine learning solution for predicting telecom customer churn. It combines data cleaning, SQL analysis, dashboarding, and predictive modeling to support business decision making.

Objectives

Identify customers likely to churn, understand the key factors influencing churn, and provide actionable insights to reduce customer loss and improve retention.

Tools and Technologies

Python, Pandas, NumPy, Scikit learn, MySQL, Power BI, Matplotlib, Seaborn, Joblib, Git and GitHub.

Data Preparation

Performed missing value handling, type conversions, feature engineering such as tenure grouping and high charge flags, and one hot encoding for categorical variables.

SQL Analysis

Executed multiple queries to compute churn rates, analyze customer behavior by contract type, tenure, and payment method, and derive business insights.

Dashboard Development

Built an interactive Power BI dashboard displaying KPIs, churn trends, contract comparisons, and customer segmentation to assist stakeholders.

Model Development

Trained Logistic Regression and Random Forest models. Selected Logistic Regression with class balancing due to improved recall for churn customers.

Performance

Model achieved approximately 75 to 82 percent accuracy with strong recall for the churn class. The trained model was saved using Joblib for deployment readiness.

Project Outcome

The project demonstrates end to end capabilities in data analysis and data science including database querying, visualization, machine learning, and production ready model management.

Conclusion

This work reflects practical industry skills suitable for roles in Data Analytics and Data Science and provides a scalable foundation for customer retention strategies.