Customer Churn Prediction

■ Project Overview

This project predicts customer churn for a telecommunications company using machine learning models. Customer churn occurs when a customer stops using a company's services. The goal is to identify customers who are likely to churn so that the company can take proactive measures to retain them.

■ Data Source

The dataset used is the Telco Customer Churn dataset, which is publicly available on Kaggle: https://www.kaggle.com/blastchar/telco-customer-churn. It contains customer demographic information, account details, and whether the customer churned or not.

■ Workflow

- Data Preprocessing: Handle missing values and encode categorical variables.
- Exploratory Data Analysis (EDA): Visualize churn distribution and check correlations.
- Model Training: Logistic Regression, Random Forest, XGBoost (optimized).
- Model Optimization: Hyperparameter tuning using GridSearchCV/RandomizedSearchCV.
- Evaluation Metrics: Accuracy, Precision, Recall, F1-Score, Confusion Matrix, ROC-AUC Curve.

■ Results

Logistic Regression: Accuracy ≈ 79.91%
Random Forest: Accuracy ≈ 78.42%
XGBoost (optimized): Accuracy ≈ 80.62%

The XGBoost model achieved the best performance with the highest ROC-AUC score.

■ Conclusion

Churn prediction can help telecom companies focus on customers who are at risk. The most important features influencing churn include tenure, monthly charges, and contract type. Proactive strategies can reduce customer loss and improve revenue.

■ Author

Mohamed Kasm