Customer Churn Analysis for Telecom Industry

Introduction

This project aims to predict telecom customer churn and provide business insights through machine learning, explainability, and segmentation.

Abstract

We use historical telecom data to train a Random Forest model for churn prediction. SHAP is applied for model explainability. Customers are segmented into actionable groups such as 'At Risk' and 'Loyal' based on churn probability and behavior.

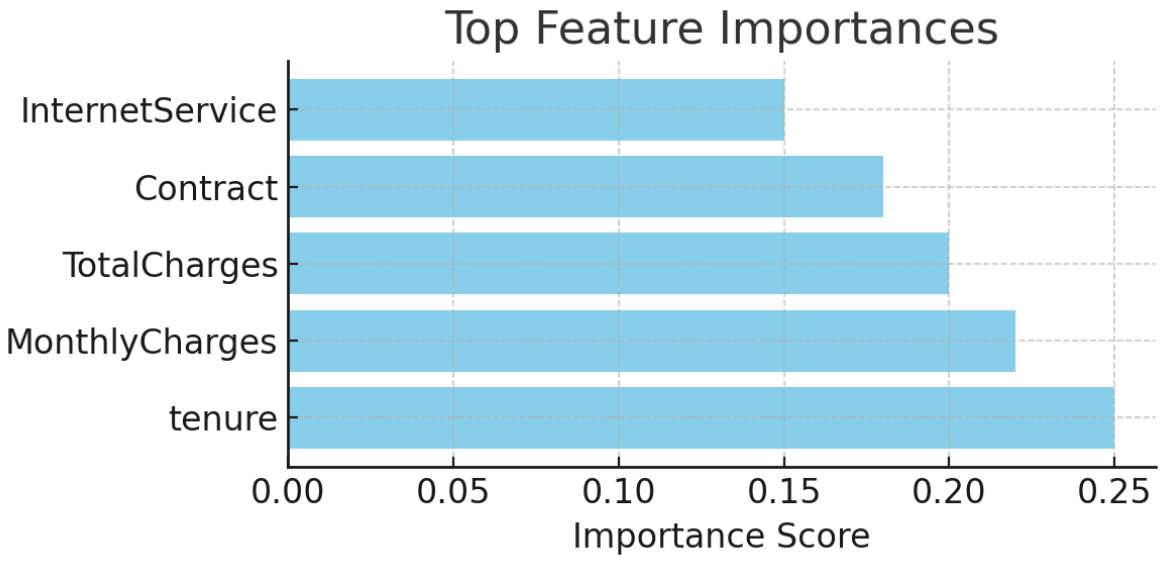
Tools Used

- Python
- Pandas, NumPy
- Matplotlib, SHAP
- Scikit-learn
- Jupyter Notebook / Google Colab

Steps Involved

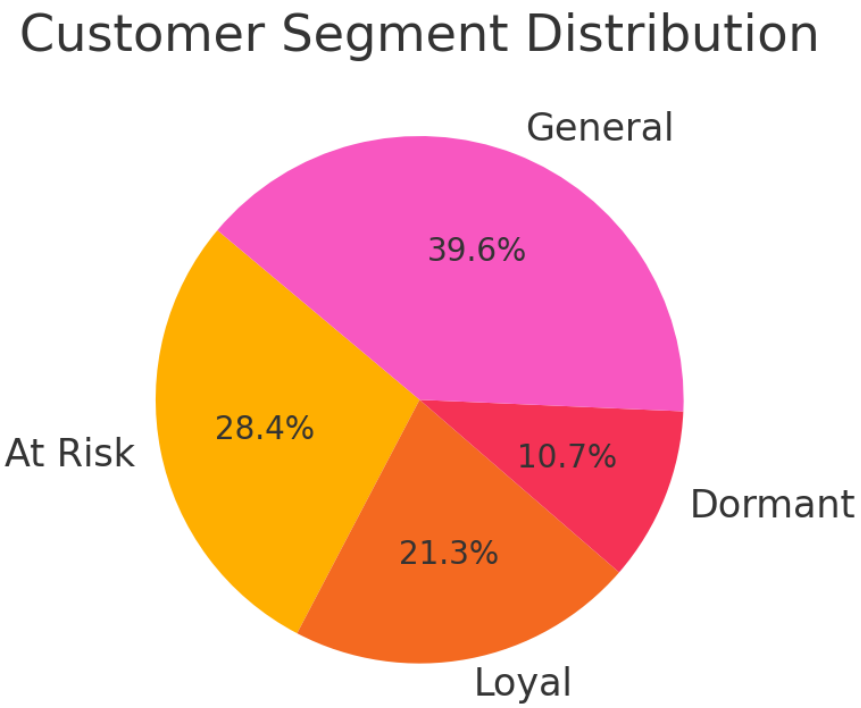
- Data Cleaning: Handle missing values and convert 'TotalCharges'.
- Encoding: Label encode categorical columns, drop irrelevant features.
- Model Training: Train a Random Forest classifier.
- Evaluation: Assess model with classification metrics.
- Explainability: Use SHAP for prediction explanation.
- Segmentation: Classify customers by behavior and churn risk.

Key Feature Importance



Customer Churn Analysis for Telecom Industry

Customer Segment Distribution



Conclusion

The churn prediction model and segmentation strategy empower the telecom provider to make informed retention decisions. Explainable AI tools like SHAP enhance trust in predictions and provide direction for targeted interventions.