Customer Churn Prediction in Telecom Industry

# Introduction

In the highly competitive telecom industry, customer retention has become a critical focus. Churn, or customer attrition, directly impacts revenue and long-term growth. This project aims to predict customer churn using machine learning and derive insights to inform strategic business decisions.

# Abstract

This project leverages machine learning techniques, particularly the Random Forest classifier, to analyze customer churn in the telecom industry. The dataset comprises 7,043 records of telecom customers, including demographics, account information, service usage, and churn labels. Additional synthetic features such as complaints, recharge frequency, and call duration were generated to enrich the dataset. Model performance was evaluated, and feature importance was interpreted using SHAP and ELI5. Customers were segmented based on churn probability into Loyal, At Risk, and Dormant categories. SQL was used to perform detailed segment-wise analysis for actionable insights.

# Tools Used

• Python (Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn)  
• SHAP, ELI5 for model interpretability  
• Imbalanced-learn (SMOTE) for handling class imbalance  
• DuckDB and SQL for analytical queries  
• Google Colab for development environment  
• Microsoft Excel and PowerPoint for reporting and presentation

# Steps Involved in Building the Project

1. Data Preprocessing: Cleaned the dataset and handled missing values in ‘TotalCharges’. Categorical variables were encoded.  
2. Feature Engineering: Added synthetic features such as complaints, recharge frequency, and call duration.  
3. Model Training: Split the data using stratified sampling and trained a Random Forest model.  
4. Evaluation: Evaluated model performance using accuracy and classification metrics.  
5. Interpretation: Identified top churn predictors using ELI5 and SHAP visualizations.  
6. Segmentation: Grouped customers by churn probability into Loyal, At Risk, and Dormant.  
7. SQL Analytics: Used DuckDB to analyze key metrics across segments for strategic insights.

# Conclusion

The churn prediction model successfully highlighted tenure, monthly charges, and complaints as critical churn indicators. Customers with high complaints and low tenure are highly likely to churn, especially in the Dormant segment. Strategic recommendations include targeted retention for high-risk segments, faster resolution of complaints, and pricing optimization. This solution provides a scalable framework for reducing churn and enhancing customer lifetime value in the telecom domain.