Customer Churn Prediction – End-to-End ML Pipeline

Project Overview
This project builds a complete machine learning pipeline to predict customer churn for a telecom company. It automates data preprocessing, feature engineering, model selection, and evaluation using data-driven decisions.

Key Highlights
• End-to-end ML pipeline
• Automated technique selection
• Rich EDA visualizations
• Handles class imbalance
• ROC-AUC based model comparison
• Deployment-ready artifacts

Pipeline Steps
1. Data Loading & Cleaning
2. Missing Value Handling
3. Variable Transformation
4. Outlier Treatment
5. Categorical Encoding
6. Feature Selection
7. Data Balancing
8. Feature Scaling
9. Model Training & Tuning

Models Used
KNN, Naive Bayes, Logistic Regression, Decision Tree, Random Forest, SVM, XGBoost

Evaluation Metric
ROC-AUC Score with ROC Curve Visualization

Artifacts Generated
• churn_artifacts.pkl (model, scaler, features)
• scaler_path.pkl
• final_features.pkl

How to Run
python main.py

Tech Stack
Python, Pandas, NumPy, Scikit-Learn, XGBoost, Matplotlib, Seaborn

Author
Navi – Data Science & Machine Learning Intern