By: 64011748,
Ahmad Sohail
https://www.kaggle.com/datasets/radheshyamkollipara/bank-customer-churn/data

Project Scope and Objectives:
The aim of this project is to develop a predictive model to identify customers at high risk of churning from a bank. By predicting customer churn, the bank can take proactive measures to retain customers, thereby enhancing customer satisfaction and loyalty. This project seeks to analyze customer behavior, transaction patterns, and engagement levels to pinpoint the factors most indicative of churn risk.

Data Sources and Data Collection Methods:

The primary dataset for this project consists of customer churn records from a publicly available source on Kaggle.
It includes a variety of features such as CreditScore, Geography, Gender, Age, Tenure, Balance, and IsActiveMember.
The dataset comprises records of several thousand customers, making it a rich source for understanding the patterns associated with churn. Data quality will be assessed, with any necessary cleaning or preprocessing steps taken to ensure its readiness for analysis.

Tentative Analysis Approach:

- Exploratory Data Analysis (EDA): Conduct an initial review of the dataset to grasp the distributions of key features, relationships between variables, and any patterns or outliers present.
- Data Preprocessing: Clean the dataset by addressing missing values, encoding categorical variables, and scaling numerical features to prepare the data for modeling.
- Model Development: Build and train several machine learning models, such as Logistic Regression, Decision Trees, Support Vector Machines, and Random Forests, to predict customer churn. The performance of these models will be evaluated using metrics suitable for imbalance, such as the area under the receiver operating characteristic curve (AUC-ROC), precision, recall, and F1 score.
- Hyperparameter Tuning and Model Optimization: Apply techniques like grid search and random search to find the optimal model configurations that improve prediction accuracy and generalization to unseen data.

Project Expected Output/Outcome:

The project is expected to yield a highly accurate and reliable model for predicting the likelihood of a customer churning. This model will empower the bank to identify at-risk customers early and implement targeted retention strategies. Moreover, the analysis will uncover key drivers of churn, providing strategic insights that can guide the bank in enhancing customer satisfaction

and loyalty. A detailed report will be produced, summarizing the methodology, findings, and recommendations for both strategic and operational changes.