Project Description

This project aims to **predict whether a customer will default on a credit card payment** in the following month, helping financial institutions manage risk and make informed lending decisions.

Workflow

1. **Data Loading & Cleaning:**

   o Loaded dataset containing customer demographics, payment history, and credit utilization.

   o Removed duplicates, handled missing values, and converted categorical variables into numerical format.

2. **Exploratory Data Analysis (EDA):**

   o Studied relationships between age, gender, education, marital status, payment delay history, and default probability.

   o Found that **past payment delays** are a strong predictor of future defaults.

   o Higher **bill amounts and credit utilization rates** tend to increase default risk.

3. **Feature Engineering:**

   o Derived features from payment status (PAY_* variables) and bill amounts (BILL_AMT_*) for better prediction.

   o Scaled numerical data to normalize differences in value ranges.

4. **Model Training & Evaluation:**

   o Models tested: **Logistic Regression**, **Decision Tree**, **Random Forest**, **XGBoost Classifier**.

   o Performance evaluated using Accuracy, Precision, Recall, and F1-score.

   o **XGBoost Classifier** achieved the highest accuracy and recall, making it the preferred model.

Conclusion

- **Key Insights**:

   o **Payment history** is the most important factor in predicting defaults.

- Customers with repeated late payments have a significantly higher chance of defaulting.

- High bill amounts relative to the credit limit indicate increased financial stress and risk.

- **Business Value**:

    - The model allows credit card issuers to **proactively identify high-risk customers** and take preventive measures such as adjusting credit limits, sending early reminders, or offering financial counseling.

    - Reduces potential losses by enabling **data-driven risk management**.

- **Outcome**:

    - The **XGBoost Classifier** offers a reliable solution for predicting credit defaults.

    - Implementation of this model can improve portfolio quality, reduce bad debt, and enhance profitability for financial institutions.