Project Summary – Credit Default Prediction Using Machine Learning

1. Project Objective:

- The primary goal of this project is to **predict whether a customer will default (Y)** using machine learning techniques.
- This helps businesses assess credit risk and make informed financial decisions.

2. ML Pipeline Overview:

1. Exploratory Data Analysis (EDA)

- Analyzed distributions of key features like **age, credit limit, bill amounts, and repayment status**.
- Identified trends showing how **payment behavior (PAY_0 to PAY_6)** affects default probability.
- Detected missing values and outliers for later cleaning.

2. Data Cleaning & Preprocessing

- Handled missing/incorrect values.
- Scaled numerical columns like credit limit and bill amounts.
- Encoded categorical features where needed.
- Split the dataset into **train and test sets**.

3. Model Training

Implemented and trained multiple ML models:

- **Logistic Regression**
- **Gradient Boosting Classifier**
- **XGBoost Classifier**

Each model was trained to classify whether a customer will default based on previous repayment patterns and financial behavior.

4. Model Evaluation & Comparison

Used metrics such as:

- **Accuracy**
- **Precision & Recall**
- **Confusion Matrix**
- **AUC-ROC**

XGBoost / LightGBM delivered the **best performance**, with strong ability to capture nonlinear patterns and interactions in financial features.

3. Key Insights & Business Takeaways:

Customer Risk Patterns

- Features such as **credit limit**, **age**, and **repayment history** (PAY_0 to PAY_6) are strong predictors of default.
- Customers with:
  - ❖ irregular repayment pattern
  - ❖ higher previous balances
  - ❖ repeated late payments

  are more likely to default.

Early Risk Identification

- The model helps detect risky customers early so banks can:
  - ❖ send reminders and payment alerts
  - ❖ adjust credit limits
  - ❖ offer flexible payment plans
  - ❖ manage collections efficiently

4. Business Implications:

For Banks & Financial Institutions

- **Reduced financial losses** by detecting defaulters early.
- Increased **collection rate** with targeted interventions.
- Ability to **segment customers** based on risk and personalize offers.
- Improved **customer satisfaction** through smarter credit management.

Operational Advantages

- Automated risk scoring.
- Faster and more accurate loan approval processes.
- Transparent criteria for identifying high-risk customers.

5. Recommended Model for Deployment:

XGBoost / LightGBM

These models are ideal due to:

- High accuracy
- Fast training and inference
- Natural handling of imbalanced data
- Strong performance on financial dataset

They provide a reliable and scalable solution for real-world credit risk scoring systems.

6. Conclusion:

This project demonstrates the complete end-to-end development of a credit default prediction system — from EDA to deployment-ready ML models.
The insights and model outputs offer actionable strategies that directly improve financial decision-making, operational efficiency, and customer experience.