# Credit Score Prediction Report

## 1. Introduction

The financial industry heavily relies on credit scores to assess the creditworthiness of individuals. Accurate credit scoring helps financial institutions mitigate risk, optimize loan approval processes, and enhance financial decision-making. This project aims to develop a predictive model for credit scoring based on customer financial data, leveraging machine learning techniques to classify customers into different credit risk categories.

## 2. Dataset Overview

The dataset consists of financial and demographic attributes of customers. It includes records of loan details, credit behavior, outstanding debts, and other key indicators that influence credit scores. The primary goal is to analyze these variables and develop a model that accurately predicts credit scores. Below is a summary of the key attributes considered in this study:

\*\*Customer Identification & Demographics:\*\*

- Customer ID, Name, Age, Occupation, SSN

\*\*Financial Information:\*\*

- Annual Income, Monthly In-Hand Salary, Number of Bank Accounts, Number of Credit Cards

\*\*Credit & Loan Behavior:\*\*

- Number of Loans, Loan Types, Interest Rates, Outstanding Debt, Credit Mix, Credit Utilization Ratio

\*\*Payment Behavior & Credit History:\*\*

- Number of Delayed Payments, Payment Behavior, Minimum Amount Payment Status, Monthly Balance

\*\*Credit Score Indicator:\*\*

- Credit Score (Target Variable)

## 3. Data Preprocessing

Data preprocessing is crucial to ensure model accuracy and reliability. The following steps were taken:

- \*\*Handling Missing Values:\*\* Imputed missing values using statistical techniques such as mean, median, and mode.

- \*\*Encoding Categorical Variables:\*\* Converted categorical features like 'Occupation' and 'Type\_of\_Loan' into numerical representations.

- \*\*Outlier Detection & Removal:\*\* Identified and handled outliers in income, loan amounts, and delayed payments.

- \*\*Feature Scaling:\*\* Applied normalization techniques to standardize numerical features.

## 4. Exploratory Data Analysis (EDA)

Exploratory analysis was conducted to identify patterns and relationships between variables. Some key findings:

- Customers with a high \*\*number of delayed payments\*\* had lower credit scores.

- A \*\*higher credit utilization ratio\*\* was correlated with increased financial risk.

- \*\*Payment behavior and outstanding debt\*\* significantly influenced credit scores.

- Customers with a diverse \*\*credit mix\*\* (loans and credit cards) had better credit scores.

- Aged credit history played a \*\*positive role\*\* in credit score improvement.

## 5. Model Development & Results

Various machine learning models were trained to classify customers based on their credit scores. The models evaluated included:

- \*\*Logistic Regression:\*\* Baseline model for binary classification.

- \*\*Decision Tree:\*\* Provided interpretability but had high variance.

- \*\*Random Forest:\*\* Improved accuracy and reduced overfitting.

- \*\*Gradient Boosting (XGBoost):\*\* Delivered the highest performance with optimized hyperparameters.

Among all models, \*\*Gradient Boosting (XGBoost) achieved the best performance\*\* with high accuracy, precision, and recall. Feature importance analysis revealed that payment behavior, outstanding debt, and number of delayed payments were the strongest predictors of credit score.

## 6. Conclusion & Recommendations

The analysis confirms that financial habits, loan repayment history, and credit utilization play a significant role in determining credit scores. The final machine learning model provides a reliable predictive framework for classifying customers based on risk factors.

\*\*Recommendations for Financial Institutions:\*\*

- Implement real-time credit monitoring to assess risk dynamically.

- Encourage customers to maintain a low credit utilization ratio for better credit scores.

- Provide personalized loan and credit limit recommendations based on customer behavior.

- Improve customer education on the impact of delayed payments on creditworthiness.

Future improvements could involve integrating alternative data sources, such as transactional behavior and digital footprints, to enhance prediction accuracy.