

Credit Scoring Model Using Machine Learning

1. Introduction

Credit scoring is a crucial process for financial institutions to evaluate the creditworthiness of loan applicants. This project implements a **Logistic Regression** model to predict credit scores using customer data, helping financial institutions make informed lending decisions.

2. Problem Statement

Determining a customer's creditworthiness is a complex process that involves analyzing multiple financial and personal factors. Traditional methods are often time-consuming and inefficient. This project aims to develop a logistic regression model that accurately predicts a customer's credit score based on historical data.

3. Objectives

- To preprocess and analyze a dataset of credit score points.
- To build and evaluate a **Logistic Regression** model for credit scoring.
- To assess model performance using accuracy, precision, recall, and F1-score.

4. Proposed System

The system utilizes a **Logistic Regression** model to predict a customer's credit score based on multiple financial attributes. The dataset is preprocessed, trained, and tested to determine how well the model can classify applicants.

5. Software & Hardware Requirements

Software Requirements:

- Python
- Jupyter Notebook / Google Colab
- Scikit-Learn
- Pandas, NumPy, Matplotlib
- Machine Learning Algorithm: **Logistic Regression**

Hardware Requirements:

- Computer with at least 4GB RAM
- High-performance processor (i5 or above recommended)

6. Methodology

- 1. **Data Collection:** The dataset is obtained from **data_train.csv**, containing customer financial attributes.
- 2. **Data Preprocessing:** Handling missing values, feature scaling, and standardization.
- 3. **Feature Selection:** Identifying the most relevant features for credit scoring.
- 4. **Model Training:** Training a **Logistic Regression** model.
- 5. **Testing & Evaluation:** Measuring accuracy, F1-score, precision, recall, and confusion matrix.

7. Results & Conclusion

The trained **Logistic Regression** model successfully predicts credit scores with high accuracy. The classification report provides insights into model performance, and the results validate its effectiveness in credit risk assessment.

	Accuracy (Train)	Accuracy (Test)	F1-Score	Precision	Recall
0	0.904092	0.881745	0.862658	0.854406	0.881745