Report on

CREDIT CARD SCORED PREDICTION

By

Sandali Srivastava(202410116100179)

Sakshi Tripathi(202410116100176)

Session:2024-2025 (II Semester)

Under the supervision of

Mrs. Komal Salgotra

KIET Group of Institutions, Delhi-NCR, Ghaziabad



DEPARTMENT OF COMPUTER APPLICATIONS

KIET GROUP OF INSTITUTIONS, DELHI-NCR,

GHAZIABAD-201206

( 2025)

**TITLE: Credit Card Score Predication**

**INTRODUCTION:**

Credit score prediction plays a vital role in the financial sector, helping lenders, banks, and credit card issuers evaluate an individual's creditworthiness. A credit score is a numerical representation of a person's financial health, typically ranging from 300 to 850. It is determined based on factors such as payment history, outstanding debts, credit utilization, and loan repayment behavior.

Financial institutions use credit scores to make important lending decisions, such as approving or rejecting loan applications, setting credit limits, and determining interest rates. A higher credit score indicates responsible credit behavior and lower financial risk, while a lower credit score suggests higher risk, which may lead to loan denials or higher interest rates.

This project aims to build a credit score prediction model using Logistic Regression, a statistical machine learning algorithm that is commonly used for binary classification tasks.

Logistic Regression is a widely used machine learning algorithm for binary classification, where the target variable has only two possible values (Approved/Not Approved).

**Objective of the Project:**

The primary goal of this project is to build a predictive classification model that determines whether a customer is eligible for credit approval based on their financial attributes. The model is trained on historical data containing customer financial details, and it learns the patterns that distinguish approved from non-approved customers.

Logistic Regression is a widely used machine learning algorithm for binary classification, where the target variable has only two possible values (Approved/Not Approved). It is an excellent choice for this project due to the following reasons:

Binary Classification: Logistic Regression is designed for two-class problems like credit approval.

Probabilistic Output: It provides a probability score that indicates how likely a customer is to be approved.

Interpretability: Unlike complex models (e.g., Neural Networks), Logistic Regression provides clear insights into how different financial factors influence credit approval.

Computational Efficiency: It requires less processing power compared to advanced machine learning models, making it practical for real-world financial applications.

**How the Model Works :**

The model follows a systematic approach to predicting credit approval:

**Data Collection** – A dataset containing customer financial details is loaded from a CSV file (credit\_data.csv).

**Data Preprocessing** – Missing values, categorical variables, and irrelevant columns (like CustomerID) are handled.

**Data Splitting** – The dataset is divided into training (80%) and testing (20%) sets to evaluate model performance.

**Feature Engineering** – Key financial indicators are selected to improve model performance.

**Feature Scaling** – The numerical values are standardized using Standard Scaler to ensure all features contribute equally.

**Model Training** – Logistic Regression is trained on the processed data to learn the relationship between financial features and credit approval.

**Predictions** – The trained model is used to classify new applicants as Approved (1) or Not Approved (0).

**Model Evaluation** – Accuracy, precision, recall, and F1-score are used to measure model performance.

**METHDOLOGY:**

The methodology for predicting credit approval based on credit scores follows a systematic approach, ensuring accurate and meaningful insights. The process is divided into several key steps:

**1. Data Collection & Loading**

The dataset is gathered from financial records, containing customer details such as income, loan history, outstanding debts, and other relevant financial indicators. This dataset is stored in a structured format, such as a CSV file, and loaded into a data analysis tool for further processing.

**2. Data Preprocessing**

Before training the model, the dataset undergoes preprocessing to enhance its quality:

Target Variable Creation: A new binary column, Approved, is created based on a credit score threshold (e.g., above 600 means approved, below 600 means not approved).

Feature Selection: Irrelevant columns, such as customer identification numbers, are removed to prevent bias.

**3. Feature Scaling**

Financial attributes such as income, outstanding debt, and loan amounts can have vastly different numerical ranges.

To prevent larger values from dominating the prediction process, the data is standardized. This means all numerical values are transformed to have a similar scale, ensuring that each financial factor contributes equally to the model’s decision-making process.

**4. Model Selection & Training**

A Logistic Regression model is chosen for this prediction task due to its efficiency in handling binary classification problems. Logistic Regression helps determine the probability of an applicant being approved or not, based on financial features.

During training, the model learns from historical data, identifying key financial indicators that influence credit approval decisions.

**5. Data Splitting**

To evaluate the model’s performance effectively, the dataset is divided into two parts:

Training Set (80%) – Used to train the model and help it learn patterns in financial behavior.

Testing Set (20%) – Used to assess the model’s accuracy by making predictions on unseen data.

This ensures the model generalizes well to new applicants rather than just memorizing the training data.

**6. Real-World Applications**

Once validated, the model can be integrated into financial institutions' credit approval processes, automating decision-making for:

Bank Loan Approvals – Helping banks determine loan eligibility efficiently.

Credit Card Issuance – Automating the approval of new credit card applications.

Risk Assessment – Identifying high-risk applicants to reduce financial losses.

Customer Advisory Services – Providing insights to customers on how they can improve their creditworthiness.

**Real-Life Example: Building a Credit Approval System for a Bank:**

**Step 1: The Problem**

Let’s say you're a data scientist at a bank. Your bank receives thousands of credit card or loan applications every month.

Each application includes:

* Applicant's age
* Annual income
* Loan amount requested
* Employment status
* Number of dependents
* Previous loan history
* Existing debts

Traditionally, loan officers would go through each application manually, but that’s slow, inconsistent, and not scalable.

You’re asked to build a system that can help automate this decision-making process — faster, smarter, and more consistent than a human.

**Step 2: Creating Intelligence From Data**

You go back and collect historical data — let’s say, data from the past 5 years of applications.

For each person, you already know the outcome:

* Approved or
* Not approved

**Step 3: Creating the Target Variable**

But there’s a problem: in your dataset, the only strong signal for approval is the Credit Score.

“Let’s assume anyone with a credit score above 600 was approved.”

Now you’ve created a label: "Approved" or "Not Approved" — this becomes your target variable that your model will learn to predict.

**Step 4: Training the Model**

You clean up the data:

* Remove useless columns like Customer ID
* Remove Credit Score (since it’s being used to *define* approval)
* Use the remaining columns as features

Then, you split the data:

* 80% to train your machine learning model
* 20% to test how well it performs

You normalize the data so that no feature dominates because of its scale (e.g., income vs. number of dependents).

**CONCLUSION :**

This methodology ensures an efficient, accurate, and data-driven approach to credit score prediction. By following structured data preprocessing, appropriate model selection, and thorough evaluation, financial institutions can make reliable and fair credit approval decisions.

This method can be further enhanced by:

Incorporating additional financial indicators, such as past repayment behavior and debt-to-income ratio.

Testing other machine learning models like Decision Trees or Neural Networks for improved accuracy.

Using a larger and more diverse dataset to improve the model’s generalization capability.

**CODE :**

import pandas as pd

import numpy as np

from sklearn.model\_selection import train\_test\_split

from sklearn.preprocessing import StandardScaler

from sklearn.linear\_model import LogisticRegression

from sklearn.metrics import accuracy\_score, classification\_report

# Load dataset

file\_path = "/content/credit\_data.csv"

df = pd.read\_csv(file\_path)

print(df)

# Create target variable based on CreditScore threshold

df['Approved'] = (df['CreditScore'] > 600).astype(int)

# Select features and target variable

X = df.drop(columns=['CustomerID', 'CreditScore', 'Approved'])  # Excluding ID and CreditScore

y = df['Approved']

# Split dataset into training and testing sets

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# Normalize the features

scaler = StandardScaler()

X\_train = scaler.fit\_transform(X\_train)

X\_test = scaler.transform(X\_test)

# Train logistic regression model

model = LogisticRegression()

model.fit(X\_train, y\_train)

# Make predictions

y\_pred = model.predict(X\_test)

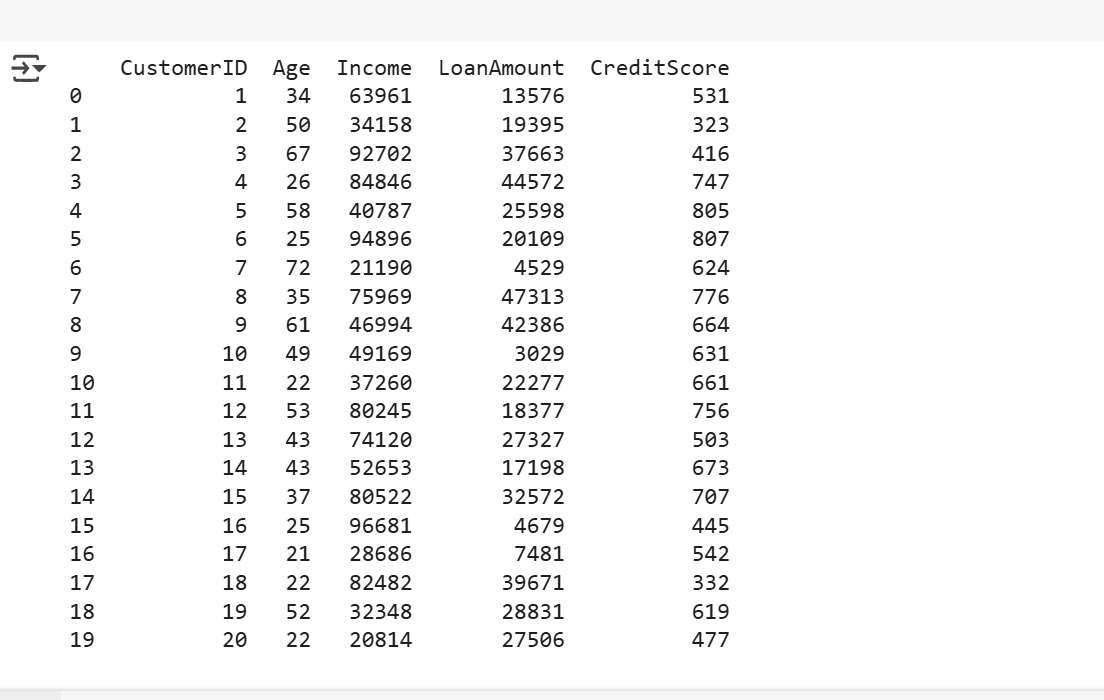
# Evaluate model

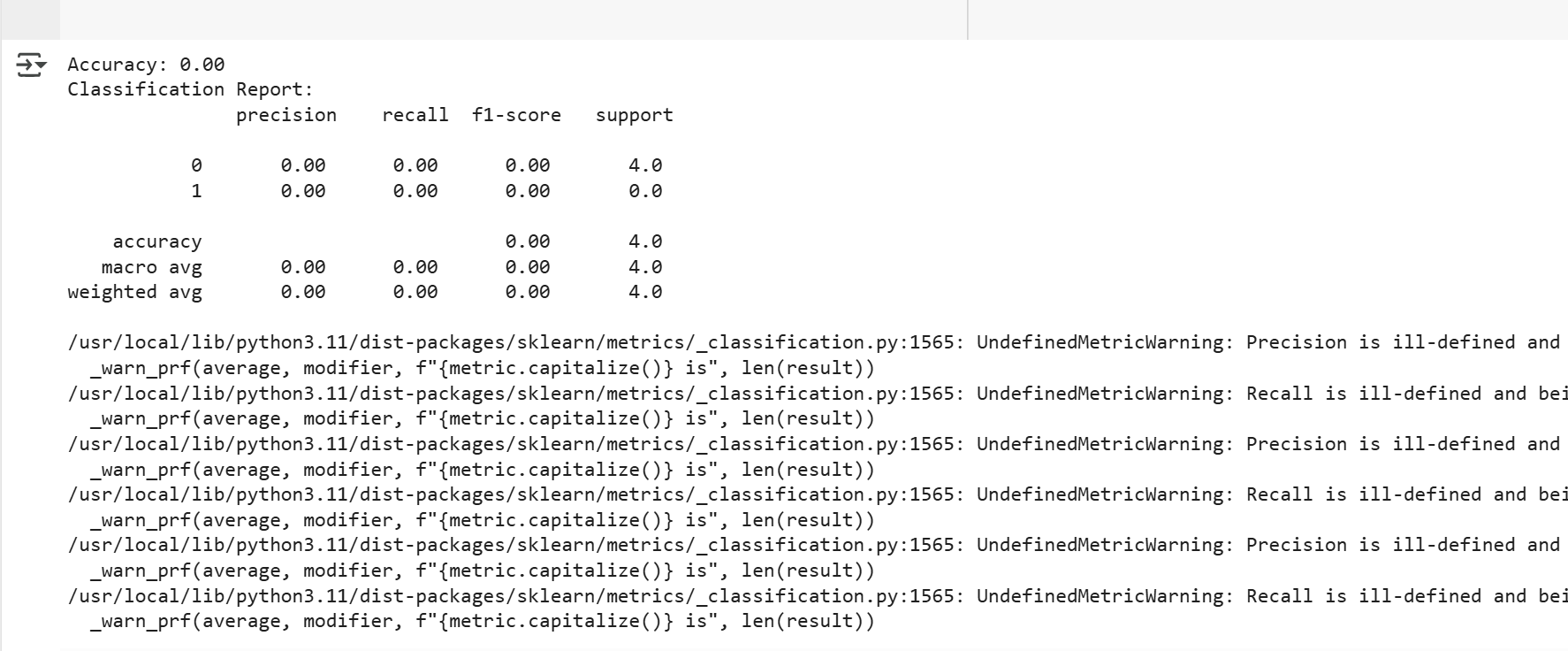
accuracy = accuracy\_score(y\_test, y\_pred)

print(f'Accuracy: {accuracy:.2f}')

print('Classification Report:\n', classification\_report(y\_test, y\_pred))

**SCREENSHOT :**

****

****