Assesment Report

on

“Loan Predictor”

submitted as partial fulfillment for the award of

BACHELOR OF TECHNOLOGY DEGREE

SESSION 2024-25

By

“KUSHAGRA RASTOGI”

Under the supervision of

“Abhishek sir”

KIET Group of Institutions, Ghaziabad

Affiliated to

Dr. A.P.J. Abdul Kalam Technical University, Lucknow

(Formerly UPTU)

May, 2025

Introduction:

Loan default prediction is a critical aspect of financial risk assessment in the banking and lending industry. Accurately identifying whether a borrower is likely to default enables financial institutions to make informed decisions, minimize losses, and offer better services to creditworthy customers. This report presents a basic analysis using a sample dataset of loan applicants to simulate a loan default prediction scenario. For illustrative purposes, random predictions are used to demonstrate evaluation metrics such as the confusion matrix.

Methodology

The methodology followed in this report includes the following steps:

1. **Data Loading and Inspection:**
    - The dataset, presumed to contain information on loan applicants along with a target variable Default, is loaded using pandas.
    - Basic exploration is performed by displaying the dataset's columns and sample rows to understand its structure.
2. **Label Selection and Prediction Simulation:**
    - The actual loan default status is taken from the Default column, which acts as the ground truth for evaluation.
    - Since the focus is on demonstrating model evaluation metrics, synthetic predictions are generated randomly (binary values of 0 or 1) using NumPy for illustrative purposes.
3. **Evaluation Using Confusion Matrix:**
    - A confusion matrix is computed to compare actual labels and randomly generated predictions.

- The matrix quantifies true positives, true negatives, false positives, and false negatives, offering insight into classification performance.

4. **Visualization:**
    - A heatmap of the confusion matrix is plotted using seaborn, visually representing the prediction results to aid interpretation.

While this simulation uses random predictions, in a real-world application, the predicted values would result from a trained machine learning model. The confusion matrix serves as a fundamental evaluation tool to assess model accuracy and identify areas for improvement.

CODE:

```
# Install required libraries (if not already installed)

!pip install pandas scikit-learn seaborn matplotlib --quiet

# Import libraries

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

import numpy as np

from sklearn.metrics import confusion_matrix

# Load the dataset

data = pd.read_csv('/1. Predict Loan Default (1).csv')

# Preview the dataset to check columns

print("Columns in the dataset:")

print(data.columns)

# Display the first few rows

print("\nSample data:")

print(data.head())
```

```
# Use 'Default' as the actual labels

actual_labels = data['Default']

# Generate random predictions (0 or 1) for demonstration purposes

np.random.seed(42)  # for reproducibility

data['PredictedDefault'] = np.random.randint(0, 2, size=len(data))

# Use the generated predictions

predicted_labels = data['PredictedDefault']

# Create the confusion matrix

cm = confusion_matrix(actual_labels, predicted_labels)

# Create a heatmap for the confusion matrix

plt.figure(figsize=(6,4))

sns.heatmap(cm, annot=True, fmt='d', cmap='Blues')

plt.xlabel('Predicted')

plt.ylabel('Actual')

plt.title('Confusion Matrix Heatmap')

plt.show()
```

References/Credits :

1.Dataset Source: CHAT GPT .

2.Libraries Used: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn.

3.Images and Graphs generated using Python visualization libraries.