Credit Risk Classification Report

Overview of the Analysis

The purpose of this analysis is to evaluate a machine learning model for credit risk classification. The dataset includes financial information about loan applicants, and our goal is to predict whether a loan is high or low risk based on these features. The model aims to help financial institutions make informed lending decisions by identifying potentially risky loans.

Process:

1. **Data Preparation:**
    - The dataset was loaded and preprocessed.
    - The target variable (loan_status) was separated from the features.
2. **Model Selection:**
    - We applied a **Logistic Regression** model for classification.
3. **Model Evaluation:**
    - We assessed performance using accuracy, precision, and recall metrics.

Results

- **Machine Learning Model:** Logistic Regression
    - **Accuracy:** >99%
    - **Precision (for class 1):** 87%
    - **Recall (for class 1):** 95%
    - **Precision**, **Recall**, and **F-Score** are all 100% for class 0

Summary

The logistic regression model performed exceptionally well, achieving an accuracy of 99.24%. Given the nature of credit risk classification, recall is particularly important as it ensures we correctly identify as many high-risk loans as possible. With a recall score of 95%, the model effectively captures most of the high-risk cases, making it a viable choice for credit risk assessment.

Recommendation:

- Given its high accuracy and strong recall performance, we recommend using this model for credit risk classification.
- However, if the business prioritizes minimizing false positives (incorrectly classifying a loan as high risk), additional models like could be explored for better precision.