Credit Risk Analysis Report

Purpose of the Analysis:

This analysis evaluates how well a **logistic regression model** predicts the risk of loans being **healthy (0)** or **high-risk (1)**. The dataset includes information such as loan size, interest rate, borrower income, debt-to-income ratio, number of accounts, derogatory marks, total debt, and loan status. The model's performance is measured using **accuracy**, **precision**, and **recall**, which show how well it can separate healthy loans from high-risk loans. This helps the company decide if the model is suitable for automating loan approvals and risk assessments.

Model Performance: Accuracy, Precision, and Recall

- **Accuracy: 99%**
    - The model correctly predicts 99% of the loan labels, indicating strong overall performance.

- **Precision:**
    - "0"- Healthy Loans : 1.00
        - The model perfectly identifies healthy loans, with no false positives.
    - "1"- High-Risk Loans: 0.87
        - The model correctly classifies 87% of high-risk loans, with 13% being false positives .

- **Recall:**
    - "0"-Healthy Loans:1.00
        - The model perfectly identifies all healthy loans, with no false negatives.
    - "1"- High-Risk Loans: 0.95
        - The model correctly identifies 95% high-risk loans, but misses 5% false negatives.

Summary and Recommendation:

Results:

- The logistic regress model performs excellently in predicting healthy loans, with perfect precision and recall. It is very effective in identifying healthy loans, which constitutes most of the dataset.
- The model is also strong in prediction of high-risk loans, with a recall of 95% means it flags most high-risk loans accurately. But there is still a slight drop in precision (87%) for high-risk loans, indicating a few healthy loans are misclassified as high-risk.

Recommendation:

This model is **recommended for used by the company** for the the following reasons:

- High accuracy(99%) makes it a reliable tool for automating loan risk assessment.
- Strong performance on the healthy loans (“0”) ensures that the majority of loans are correctly classified, reducing operational risks.
- High recall for high-risk loans (95%) ensures that the majority of risk loans are flagged, minimizing the risk of loan defaults.

Justification: The model works well overall, especially in identifying high-risk loans (1), which is important for preventing financial losses. While precision for high-risk loans could be improved, the high recall reduces the risk of missing important loans.

Improvement Opportunity:

The company could consider addressing the class imbalance issue by applying techniques such as class weight adjustments or resampling to increase the model’s precision for the high-risk loans. However, even without these changes, the model is still valuable for the company.