

Overview of the Analysis

The purpose of this analysis is to evaluate the performance of machine learning models in predicting credit risk. The financial data analyzed includes variables such as loan size, interest rate, borrower income, debt-to-income ratio, number of accounts, derogatory marks, and total debt. The primary goal is to predict the `loan_status`, which indicates whether a loan is healthy (0) or high-risk (1).

The machine learning process involved several stages:

1. **Data Preprocessing:** This included loading the dataset, separating the features (x) from the target variable (y), and splitting the data into training and testing sets.
2. **Model Training:** A logistic regression model was trained using the training data.
3. **Model Evaluation:** The performance of the logistic regression model was evaluated using the testing data, focusing on accuracy, precision, recall, and F1-score metrics.

Results

- **Logistic Regression Model:**
 - Accuracy: 99%
 - Precision:
 - Healthy Loan (0): 1.00
 - High-Risk Loan (1): 0.85
 - Recall:
 - Healthy Loan (0): 0.99
 - High-Risk Loan (1): 0.91
 - F1-Score:
 - Healthy Loan (0): 1.00
 - High-Risk Loan (1): 0.88

Summary

The logistic regression model demonstrates high performance in predicting credit risk, particularly for healthy loans (0). It achieves near-perfect precision and recall for healthy loans, indicating that the model can accurately identify loans that are not at risk. While the model performs well in predicting high-risk loans (1), with good precision and recall scores, there is still some room for improvement to reduce the number of false negatives.

Recommendation: Based on the results, the logistic regression model is recommended for predicting credit risk due to its high overall accuracy and excellent performance in identifying healthy loans. The model's precision and recall for high-risk loans are also strong, making it a reliable tool for credit risk analysis. This model can assist the company in minimizing financial risk by accurately identifying loans that are likely to default. However, continuous monitoring and potential model adjustments should be considered to further enhance its performance in identifying high-risk loans.

