**Credit Risk Classification**

**Overview of the Analysis**

The analysis aimed to develop machine learning models for predicting loan default risk based on financial data. The purpose was to utilize various financial attributes to classify loans as either healthy or high-risk. The dataset included features such as loan size, interest rate, borrower income, debt-to-income ratio, number of accounts, derogatory marks, total debt, and loan status. The goal was to predict the loan status, which had two classes: 0 for healthy loans and 1 for high-risk loans. It's worth noting that the dataset exhibited a class imbalance, with healthy loans being significantly more prevalent than high-risk loans.

The stages of the machine learning process involved standard procedures: data preprocessing, feature selection, model selection, and evaluation. Several machine learning algorithms were explored, including Logistic Regression, as a baseline model, due to its simplicity and interpretability.

**Results**

**Machine Learning Model: (Logistic Regression):**

**A screenshot of a computer screen

Description automatically generated**

* Accuracy: 99%
* Precision (healthy loan - label 0): 100%
* Precision (high-risk loan - label 1): 85%
* Recall (healthy loan - label 0): 99%
* Recall (high-risk loan - label 1): 91%

**Summary**

The logistic regression model demonstrated exceptional performance in predicting both healthy and high-risk loans. For healthy loans, the model achieved near-perfect accuracy, precision, and recall scores, indicating highly accurate predictions with minimal false positives and false negatives. On the other hand, for high-risk loans, the model maintained a strong precision score of 85% and a recall score of 91%, suggesting effective identification of actual high-risk loans while keeping false positives relatively low.

Considering the performance metrics and the problem context, the logistic regression model appears to be the most suitable choice for predicting loan default risk. Its balanced performance across both classes makes it a reliable model for identifying both healthy and high-risk loans. However, it's essential to acknowledge that performance may vary depending on specific problem requirements. In some cases, such as when minimizing false positives is critical, further exploration of different algorithms or fine-tuning of existing models may be necessary.