**##** **Overview of the Analysis**

In this section, describe the analysis you completed for the machine learning models used in this Challenge. This might include:

\* Explain the purpose of the analysis.

\* Explain what financial information the data was on, and what you needed to predict.

\* Provide basic information about the variables you were trying to predict (e.g., `value\_counts`).

\* Describe the stages of the machine learning process you went through as part of this analysis.

\* Briefly touch on any methods you used (e.g., `LogisticRegression`, or any resampling method).

**Provide an overview that explains the purpose of this analysis. (5 points).**

Purpose: To access the credit-score of borrowers that can impact the decision-making process for banks to approve or reject the financial proposal.

Financial information data:

* Loan size
* Interest Rate
* Borrower Income
* Debt to Income Ratio
* Number of Accounts
* Derogatory Marks
* Total Debt
* Loan Status

Creating a model to appropriately predict the financial loan risk based on the dataset containing financial attributes of borrowers. The predictive variable is categorised into ‘healthy’ and ‘high-risk’ categories.

The machine learning process includes splitting of the data from the CSV file into training and testing sets. The training of machine learning models including Logistic Regression. The model is evaluated using metrics like accuracy, precision, recall, f1-score and support.

**Using a bulleted list, describe the accuracy, precision, and recall scores of the machine learning model. (5 points)**

In the confusion matrix, the result shows that out of the 18 765 loan status that are healthy, the model predicted 18 649 as healthy correctly and 116 healthy incorrectly.

Whereas out of 619 loan status that are non-healthy (high-risk), the model predicted 615 as non-healthy correctly and 4 as non-healthy incorrectly.

|  | **Predicted Healthy Loans (low-risk)** | **Predicted Non-Healthy Loans (high-risk)** |
| --- | --- | --- |
| **Actual Healthy Loans (low-risk)** | 18649 | 116 |
| **Actual Non-Healthy Loans (high-risk)** | 4 | 615 |

In Model 1, class 0 (healthy loan) The performance report:

precision recall f1-score support

0 1.00 0.99 1.00 18765

**Precision**: It measures the proportion of correctly predicted instances among all instances predicted as class 0. In this model, the precision score is 1.00, indicating that all instances predicted as healthy loans were healthy loans.

**Recall**: Also known as sensitivity, measures the proportion of correctly predicted instances of class 0 out of all actual instances of class 0. Here, the recall score is 0.99, indicating that 99% of actual healthy loans were correctly identified by the model.

**F1-score**: The F1-score is the harmonic mean of precision and recall and provides a balance between the two metrics. In this case, the F1-score is 1.00, reflecting the model's ability to achieve high precision and recall simultaneously.

In Model 2, class 1 (high-risk loan) The performance report:

precision recall f1-score support

1 0.85 0.91 0.88 619

**Precision**: Precision measures the proportion of correctly predicted instances among all instances predicted as class 1. In this case, the precision score is 0.85, indicating that 85% of instances predicted as high-risk loans were high-risk loans.

**Recall**: Also known as sensitivity, measures the proportion of correctly predicted instances of class 1 out of all actual instances of class 1. Here, the recall score is 0.91, indicating that 91% of actual high-risk loans were correctly identified by the model.

**F1-score**: The F1-score is the harmonic mean of precision and recall and provides a balance between the two metrics. In this case, the F1-score is 0.88, reflecting the model's ability to achieve a balance between precision and recall for class 1.

accuracy 0.99 19384

macro avg 0.92 0.95 0.94 19384

weighted avg 0.99 0.99 0.99 19384

The provided accuracy score of 0.99 indicates the overall measure of how well the model is performing across all classes. It signifies that the model correctly predicts the outcome 99% of the time on the total dataset of 19,384 instances.

**Summarise the results from the machine learning model. Include your justification for recommending the model for use by the company. If you don’t recommend the model, justify your reasoning. (10 points)**

**Justification for Recommendation:**

Based on all 4 of the evaluation metrics of classification report, the high-performance result from each model appears to be well-suited for predicting the creditworthiness of borrowers. The accuracy of both models shows strong performance across these evaluation metrics that indicates the effectiveness of the model to capture the underlying patterns in the data and making accurate predictions.

While the accuracy of Class 1 is slightly lower than Class 0 status, it is still within a reasonable accuracy of positive predictions. The recall scores indicate that the models are still able to capture a significant portion of actual high-risk loans. This is essential for risk-management purpose in financial institutes to identify and mitigate potential risk associated with high-risk loans.

In summary, the model is recommended for use by the company, especially if the primary goal is to identify healthy loans accurately while maintaining a good balance for high-risk loans. The high precision for Class 0 ensures that the company can be confident in approving loans predicted as healthy by the model. However, the company should be aware that, as with any model, continuous monitoring and evaluation are necessary to ensure its effectiveness over time in response to changing market conditions and regulatory requirements.