Overview of the Analysis

Purpose of this analysis was to evaluate a model based on loan risk. The goal was to identify the "creditworthiness" of borrowers. This module took its data from a csv data frame named 'lending\_data'. Within this csv, there is the columns with its corresponding data of loan\_size, interest\_rate, borrower\_income, debt\_to\_income, num\_of\_accounts, derogatory\_marks, total\_debt, and loan\_status. Accuracy rate was predicted twice, first creating a logistic regression model with the original data, the second with the resampled training data.

When creating the logistic regression model with the original data, the following steps were used:

* create the y label from the "loan\_status" column, and the X data frame from the remaining columns.
* Split the data into training and testing dataset using "train\_test\_split" (code: X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, random\_state=1).
* Start creating the model, using LogisticRegression.
* Creating and saving the prediction, by testing X\_test. The accuracy scored is gotten by using the balanced\_accuracy method.
* Afterwards, the classification report for the model is printed to be able to further assess the model.

For when creating the logistic regression model but with the resampled data, the following steps were used:

* Using the RandomOverSampler to resample the original data.
* A model is created again using the LogisticRegression module but using the resampled data created one step before.
* A prediction is created by using the testing data.
* The balanced accuracy score is calculated.
* Finally, the classification report model is printed.

Results

\* Machine Learning Model 1:

\* Balance Accuracy: 0.9520479254722232

precision recall f1-score support

0 1.00 0.99 1.00 18765

1 0.85 0.91 0.88 619

accuracy 0.99 19384

macro avg 0.92 0.95 0.94 19384

weighted avg 0.99 0.99 0.99 19384

\* Machine Learning Model 2:

\* Balance Accuracy: balanced\_accuracy: 0.9936781215845847

precision recall f1-score support

0 1.00 0.99 1.00 18765

1 0.84 0.99 0.91 619

accuracy 0.99 19384

macro avg 0.92 0.99 0.95 19384

weighted avg 0.99 0.99 0.99 19384

Summary

In this challenge, Model 2 using the resampled data would be better a choice due to it have a higher balanced score, making it more reliable on not making mistakes and accurately predicting healthy and non-healthy loans.