**Module 20 Report**

**Overview of the Analysis**

* The purpose of this analysis was to use two machine learning algorithms to train and evaluate the model based on loan risk. Used a dataset of historical lending activity from a peer-to-peer lending services company to build a model that can identify the credit worthiness of borrowers.
* In this dataset there was a range of client financial details such as loan size, interest rate, borrower income, and many others. The primary objective was to predict clients’ loan status, categorized as healthy loans or high risk loans.
* The model initially separates the y variable and the X variable dropping the y variable column. It then employs the “value\_counts” function to assess label variable (y) balance.
* By using train\_test\_split, the dataset is split into training and testing. Then a logistic regression model is applied to the training data and predictions are made to the testing data.
* Balanced\_accuracy\_score, confusion\_matrix, and classification\_report was used to evaluate the model’s performance.
* To enhance prediction precision, the dataset undergoes oversampling using RandomOverSampler. The logistic regression model is reapplied, and the evaluation methods are employed once more to reassess the model's performance.

**Results**

* Machine Learning Model 1:
* Accuracy: 0.9442676901753825 • Precision: 0 (healthy loan) 1.00; 1 (high-risk loan) 0.87 • Recall: 0 (healthy loan) 1.00; 1 (high-risk loan) 0.89
* Machine Learning Model 2:
* Accuracy: 0.9959744975744975 • Precision: 0 (healthy loan) 0.99; 1 (high-risk loan) 1.00 • Recall: 0 (healthy loan) 0.87; 1 (high-risk loan) 1.00

**Summary**

The logistic regression model achieves a balanced accuracy of 0.944, excelling in predicting "healthy loans" with high precision. However, due to limited data on "high-risk loans," the model's precision is slightly lower for this category.

After oversampling the minority class, the model's precision improves. The balanced accuracy score rises to 0.996, and recall for "high-risk loans" increases to 1.00.

In summary, the model is precise in predicting "healthy loans," and for improved prediction rates in "high-risk loans," oversampling followed by logistic regression is recommended.