**Credit Risk Classification (Supervised Machine Learning Model)**

**Overview of the Analysis**

* The purpose of the analysis is to build a model that can identify the creditworthiness of borrowers.
* A dataset of historical lending activity from a peer-to-peer lending services company.
* We used various techniques to train and evaluate a model based on loan risk.
* We used a Dependent variable (y value) in this analysis which was the "loan status" indicating if a loan is healthy or at risk.
* Our Independent Variables (x values) were loan size, interest rate, borrower income, debt to income ratio, number of accounts and derogatory marks.
* In this analysis, we first split our data to training and test sets. Then, defined our dependent and independent variables. Next, we created logistic regression model and fit our original data to this model. We printed the classification model.
* Logistic Regression model was created by using the original data set. In the end, the results -which was gathered evaluated and analysis was written accordingly.

**Results**

Logistic Regression Model with Original Data

*Balanced Accuracy Score: 0.9440909335978631*

**Table

Description automatically generated**

Logistic Regression Model with Randomly Oversampled Data

*Balanced Accuracy Score: 0.9945995452624425*

**Table

Description automatically generated**

**Summary**

Results show that the collected data can be effectively used to train and test the Machine Learning Classification Model. For better predictions, solving the imbalance sampling issue is necessary.

Randomly oversampling the data helps us to get higher balanced accuracy and recall scores. With higher recall value, the model can predict risky loans accurately.

With incorrect predictions we have two issues:

* False positives (where users are falsely flagged as risky)
* False negatives (where users are not flagged as risky but possess high risk)

both cases can create great loss. It is important to predict both 1s and 0s. Therefore, model should have good accuracy in terms of both cases.