**Analysis of machine learning models**

**Purpose of analysis**

The analysis is aimed to build machine learning models to predict loan statuses based upon financial information. The goal is to accurately classify loans as healthy 0 or high risk 1.

**Financial information and the prediction target**

The data is about the loans and the objective was to predict whether it would be healthy or risky.

**Variable overview**

The original dataset loan status has class 0 healthy loans:75036 instances and in 1 there are 2500 instances. Data exploration and processing are done using value counts and random over sampler. And for model building and evaluation train\_test\_split is used to split data into training and training sets. Trained logistic regression as categorized as X\_train, y\_trian. Made prediction in testing data and calculated accuracy.

**Metrics and Model results**

Balanced accuracy score, precision, recall, confusion matrices, and classification reports.

**Machine learning model 1**

Balanced accuracy score 0.95

Precision (0):1

recall scores (0): 0.99

Precision (1):0.85

recall scores (1): 0.91

The weighted average showed extremely high accuracy i.e., 99% whereas precision is 85%. balanced accuracy score is 95.2. precision for loans is nearly perfect.

**Machine learning model 2**

Balanced accuracy score 0.99

Precision (0):1

recall scores (0): 0.99

Precision (1):0.84

recall scores (1): 0.99

**Summary (based on the above data)**

Both machine learning models 1 and 2 generate strong performance in predicting the loan status with high balance accuracy score, precision, and recall values. However, model 2 with random sampling achieves slightly better precision and recall scores for 1 i.e., high-risk loan.

Overall, model 2 has a higher recall for high-risk loans suggesting it may be more suitable for identifying loans at risk of default. For instance, if minimizing false negative (miss high loan) is important model 2 is preferred. Same as if achieving high precision is important model 1 is preferred.

Considering the balance accuracy score model 2 is preferred. for the oversampled, the model is accurate as per the result as it shows 99% accuracy, and there are only 4 false positives for oversampled data. It has 116 false negatives out of 19384 loans that are generated in the classification report of the model.

Model 2 is highly reliable I recommend it for the prediction of loans and for other purposes. Model 2 can be considered better as it has more accuracy and recall score than model 1 which helps in finding potential defaults.