**Background**

Since 2008 recession incidence when lots of loans were being defaulted, it is crucial to predict a risk healthy loan before it happens using Machine Learning.

**Result**

Logistic Regression prediction provided 0.99 accuracy with 56 false negative and 102 false positive events, providing the following classification report:

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

Since the original data was imbalanced, RandomOverSample module was employed, followed by Logistic Regression which provided 4 false negative and 116 false positive events. It also gave 0.99 accuracy, below was the classification report.

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**

Logistic Regression is suitable for classifying and predicting credit risks with 0.99 accuracy. Further analysis using RandomOverSample module from imbalanced -learn is appropriate since the original data was imbalanced with 75036 numbers of healthy loans and 2500 high-risk loans.

Upon using RandomOverSample module in the analysis, the false negative events dropped from 56 to 4, which means 4 loans identified as healthy turned out to be high-risk compared to 16 misclassified as healthy but turned out to be otherwise previously. 102 healthy loans upon RandomOverSample were misclassified compared to 116. This is a more conservative approach and more appropriate since data was originally imbalanced; being conservative is especially good during poor economy.