Summary Report

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 16 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 bad economy.