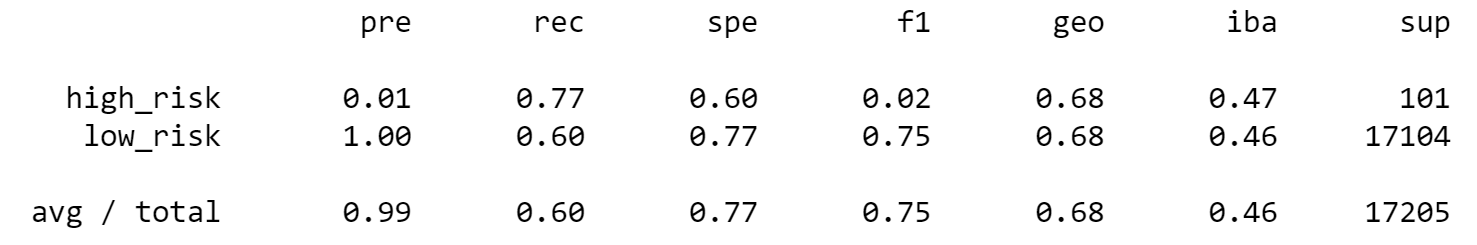
**Credit Risk Analysis:**

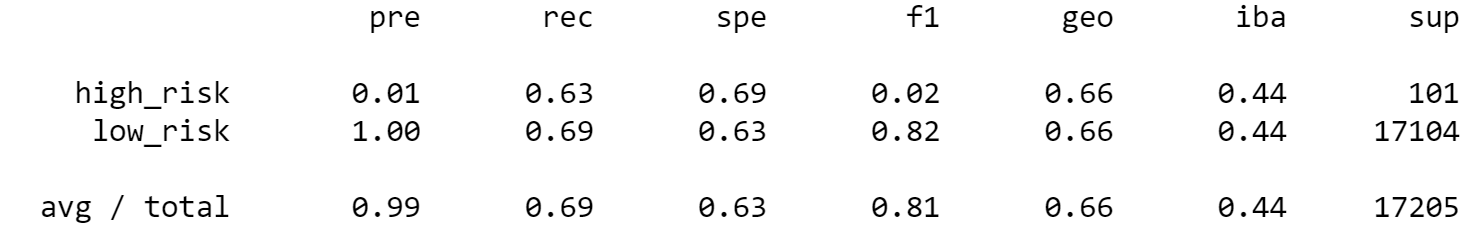
In order to predict credit risks from LendingClub data sets, Machine Learning was utilized to fit, train, and test data to provide predictive results for data set accuracy and precision. RandomOverSampler, SMOTE, ClusterCentroid, SMOTEENN, Balanced Random Forest, and Easy Ensemble Classifiers were utilized to predict credit risks. Based on the analysis provided from each tested algorithm, the Easy Ensemble ADA Boost classifier was identified as the best result with a 93.16% balanced accuracy.

RandomOverSampling using logistic regression was utilized to identify the accuracy, precision, and sensitivity of the data passing through fitted, trained, and tested data sets for prediction results. Balanced accuracy was 68.70%, precision high risk (0.01), precision low risk (1.00), recall high risk (0.77), and recall low risk (0.60) see Figure 1. This model has a high precision rate for low risk credit scores, but the scenarios show that this model is overfitted with a 100% precision score for low risk models. The recall (sensitivity), has a drop in correlation for low risk predictions, which also reflects in the F1 scores.



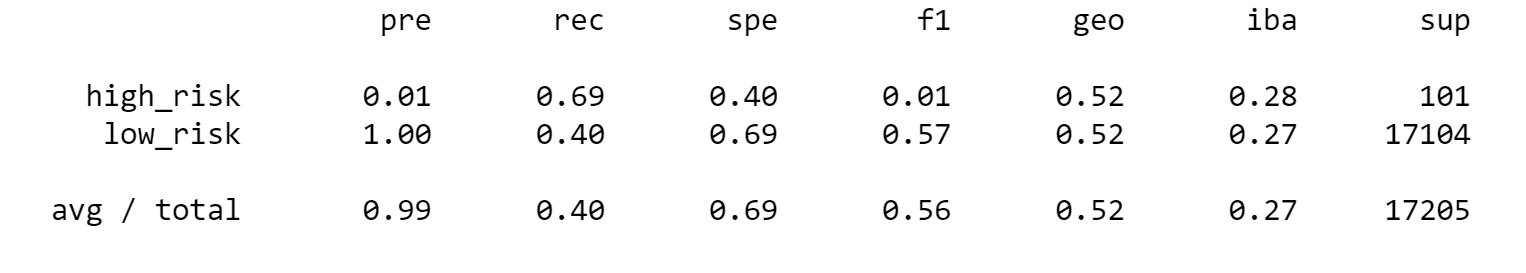
**Figure 1**

SMOTE Oversampling using logistic regression was utilized to identify the accuracy, precision, and sensitivity of the data passing through fitted, trained, and tested data sets for prediction results. Balanced accuracy was 66.25%, precision high risk (0.01), precision low risk (1.00), recall high risk (0.63), and recall low risk (0.69) see Figure 2. This model has a high precision rate for low risk credit scores, but the scenarios show that this model is overfitted with a 100% precision score for low risk models. The recall (sensitivity), has a drop in correlation for low risk predictions, which also reflects in the F1 scores.



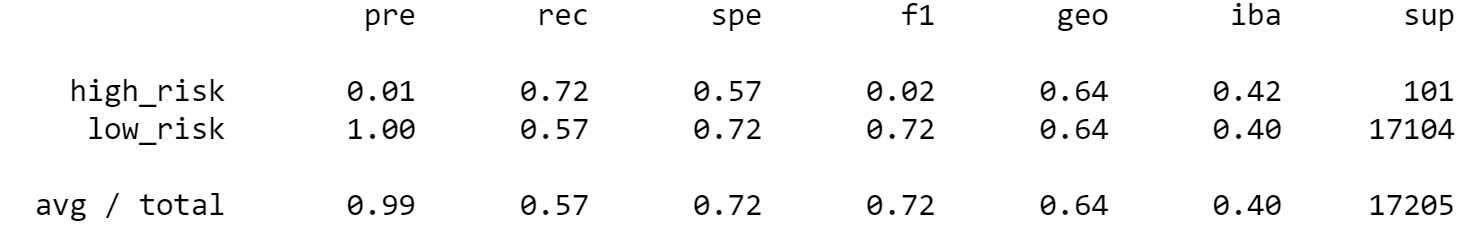
**Figure 2**

Undersampling using logistic regression was utilized to identify the accuracy, precision, and sensitivity of the data passing through fitted, trained, and tested data sets for prediction results. Balanced accuracy was 66.25%, precision high risk (0.01), precision low risk (1.00), recall high risk (0.69), and recall low risk (0.40) see Figure 3. This model has a high precision rate for low risk credit scores, but the scenarios show that this model is overfitted with a 100% precision score for low risk models. The recall (sensitivity), has a drop in correlation for low risk predictions, which also reflects in the F1 scores. This model is better than random sampling, but not as effective as SMOTE oversampling.



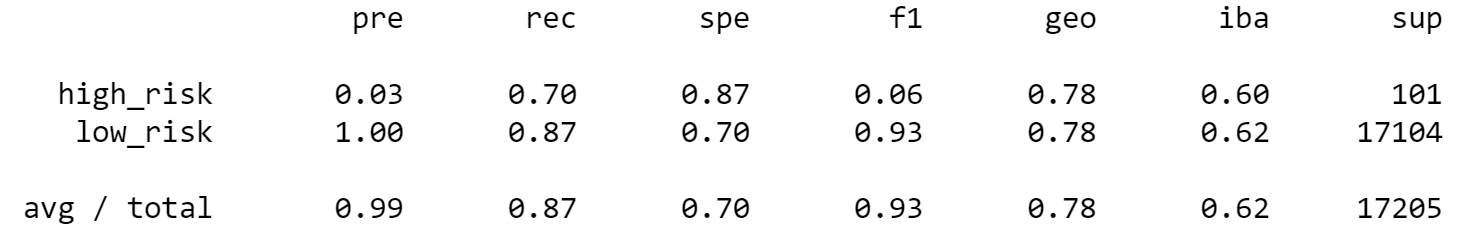
**Figure 3**

Combination Sampling (SMOTEENN) had a balanced accuracy of 54.43%, precision high risk (0.01), precision low risk (1.00), recall high risk (0.72), and recall low risk (0.57) see Figure 4. This model has a high precision rate for low risk credit scores, but the scenarios show that this model is overfitted with a 100% precision score for low risk models. The recall (sensitivity), has a drop in correlation for low risk predictions, which also reflects in the F1 scores.



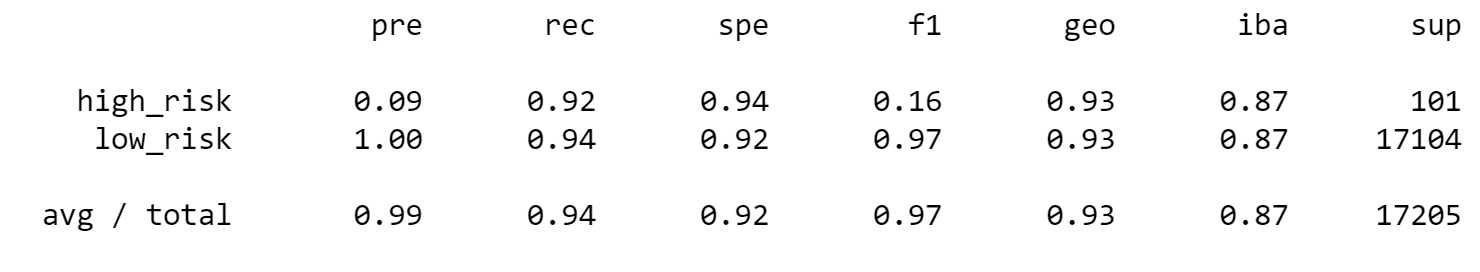
**Figure 4**

Balanced Random Forest had a balanced accuracy of 78.85%, precision high risk (0.03), precision low risk (1.00), recall high risk (0.70), and recall low risk (0.87) see Figure 5. This model has a high precision rate for low risk credit scores, but the scenarios show that this model is overfitted with a 100% precision score for low risk models. The recall (sensitivity), has a drop in correlation for low risk predictions, which also reflects in the F1 scores.



**Figure 5**

Easy Ensemble ADABoost had a balanced accuracy of 93.17%, precision high risk (0.09), precision low risk (1.00), recall high risk (0.92), and recall low risk (0.94) see Figure 6. This model has a high precision rate for low risk credit scores, but the scenarios show that this model is overfitted with a 100% precision score for low risk models. This model is the ideal method to use for credit analysis as it has the highest values for accuracy, precision, and recall compared to the others.



**Figure 6**

In conclusion, Easy Ensemble ADA Boost methodology is the correct model to utilize for credit risk machine learning. Although all models fell near the same realm for prediction values for high and low risk scores, this model was the only one that had above a 90% balanced accuracy and recall.