* Overview of the loan prediction risk analysis:

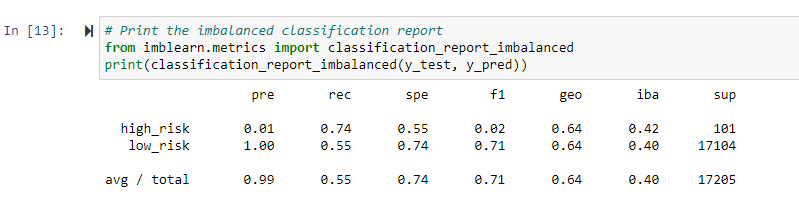
There are different techniques to train and evaluate models with unbalanced classes. This project focused on using imbalance-learn and scikit- learn libraries to build and evaluate models using resampling. The procedure adopted for the project are:

* Oversampling the data using the Randomoversampler and the SMOTE algorithms
* Undersampling the data using the ClusterCentroids algorithm.
* Use a combinatorial approach of over- and Undersampling using the SMOTEENN algorithm.
* Results:

**Naive RandomOverSampler model**

Text

Description automatically generated

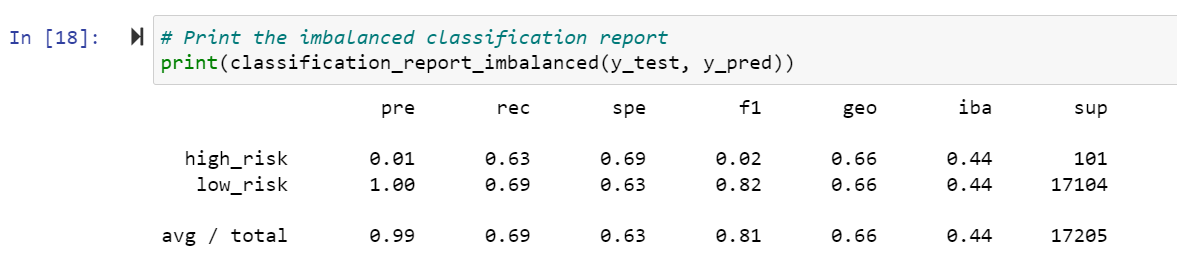


The balanced accuracy score is 65%.  
The high-risk precision is about 1% with 74% sensitivity, which only makes an F1 of 2%.  
Due to the high number of the low-risk population, its precision is almost 100%, with a sensitivity of 55%.

**SMOTE Oversampling**

Text

Description automatically generated

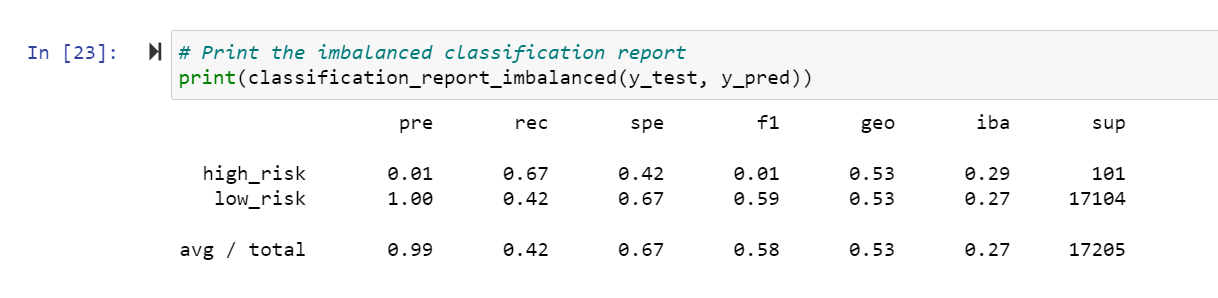


The balanced accuracy score is 64%.  
The high-risk precision is about 1% only with 63% sensitivity which makes an F1 of 2% only.  
Due to the high number of the low-risk population, its precision is almost 100%, with a sensitivity of 69%.

**Undersampling**:

Graphical user interface, text

Description automatically generated

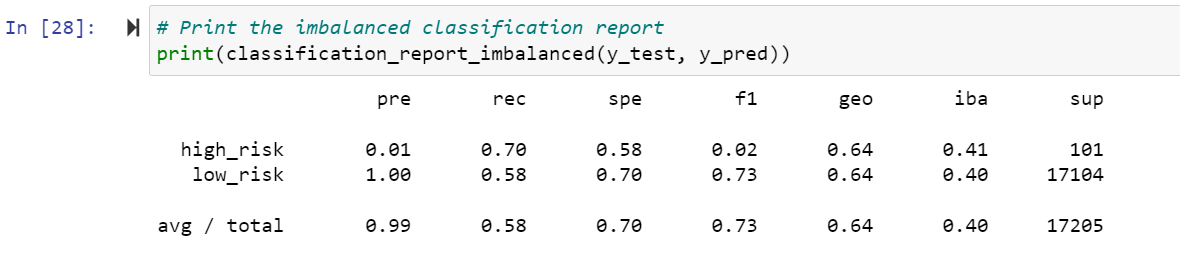


Here the balanced accuracy score is down to about 54%.  
The high\_risk precision is still 1% only with 67% sensitivity which makes an F1 of 1%.  
Due to the high number of false positives, the low-risk sensitivity is only 42%.

**Combination Sampling**

Text

Description automatically generated



The balanced accuracy score is about 64%.  
The high-risk precision is still 1% only with 70% sensitivity which makes an F1 of only 2%.  
Due to the high number of false positives, the low-risk sensitivity is 58%.

* Summary:

All the models used to perform the credit risk analysis show weak precision in determining if credit risk is high.  
With a low precision, many low-risk credits are still falsely detected as high risk, which would penalize the bank's credit strategy and infer its revenue by missing those business opportunities.  
For those reasons, I would not recommend the bank use any of these models to predict credit risk.