**Overview of the loan prediction risk analysis**

In this project, as data analysts, we are required by a company to predict credit risk on a set of data. We work on this analysis using Machine Learning methods. This involved resampling the data using imbalanced-learn in python. In addition, the combination approach was used to over and undersample the data. This was done using SMOTEENN. In the end, we used both the machine learning models to compare to reduce bias. These two models were easyensembleclassifier and BalancedRandomForestClassifier

**Results**

* **Naive Random Oversampling results:** The balanced accuracy test is 64%,

Graphical user interface, application

Description automatically generated with medium confidence

* **Graphical user interface, application

  Description automatically generated with medium confidenceSmote Oversampling results:** Their accuracy score is 65%
* Graphical user interface, text, application

  Description automatically generated**Undersampling results:** The balanced accuracy test is 54%
* **Combination (over and under) sampling results:** The balanced accuracy score is 64%.

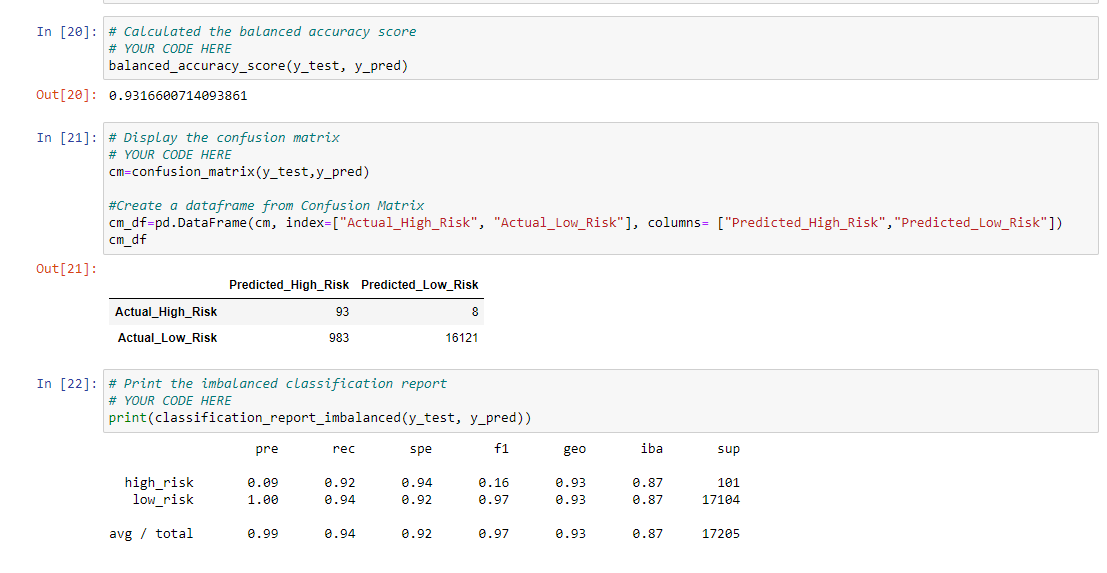
Text

Description automatically generated with medium confidence

* **Balanced Random Forest Classifier results**: The balanced accuracy score is 78%

Graphical user interface, text, application, email

Description automatically generated

* **Easy Ensemble AdaBoost Classifier results:** The balanced accuracy score is 93% and the precision is 99%

**Summary**

Based on all the results, Easy Ensemble AdaBoost Classifier provided the highest balanced accuracy results of 93%. This would in identifying loans categorized as high risk.