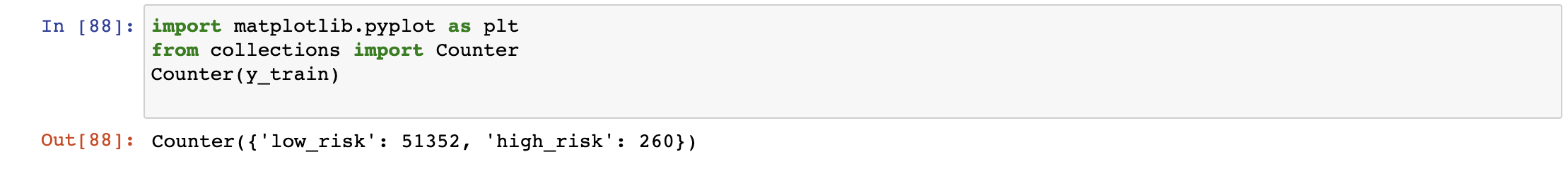
**Overview**

Credit risk classification a problem for credit card organisations because the data is usually uneven and good loans far outnumber the dangerous loans. As a result, additional strategies are needed to train and assess models with unbalanced classes.

There were in total 260 out of 51,612 that were high risk clients and rest of them were low credit risk.

****

The data was sampled using various techniques like oversampling and under sampling and combination of both: RandomOverSampler and SMOTE algorithms which were used as Oversampling methodology of minority class, the ClusterCentroids algorithm was used for under sampling the majority class, SMOTEENN algorithm was used as a combinatorial strategy of over- and under-sampling. This was done utilising the credit card credit dataset from LendingClub, a peer-to-peer lending services company.

To predict the risk two machine learning models were used to eliminate bias, BalancedRandomForestClassifier and EasyEnsembleClassifier.

The purpose of the analysis was to find the best sampling methodology as well as best classification algorithm that gave the best precision, sensitivity and F1 score for finding possible future credit card risk using historical data.

The results covered in this report are as follow:

* The oversampling methodology: RandomOverSample and SMOTE algorithm
* The under sampling methodologies: ClusterCentroids
* The combination of over and under sampling: SMOTEEN
* Machine learning models: BalancedRandomForestClassifier
* Machine learning models: EasyEsembleClassifier

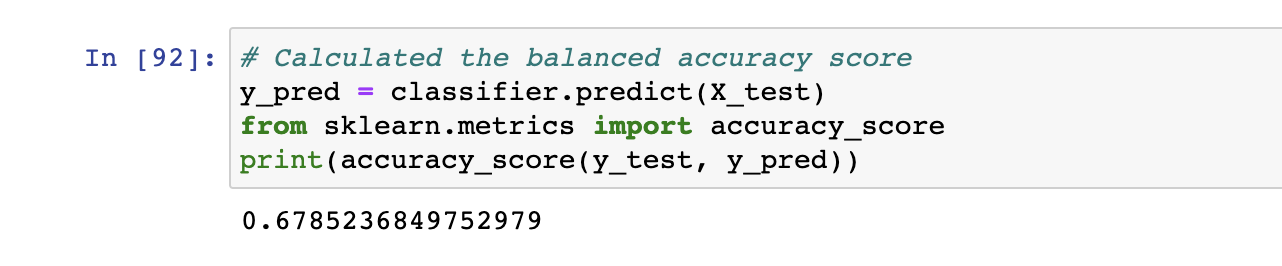
**Oversampling methodologies**

Oversampling techniques like RandomOverSampler and SMOTE algorithm was applied to minority group to make it same size as the majority group

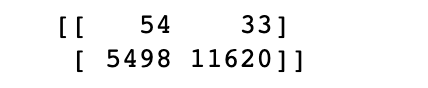
**RandomOverSampler**

Accuracy Score:

The accuracy score was 0.67852, 67.8% times the predicted value was equal to actual value using test data.

****

Confusion Matrix:

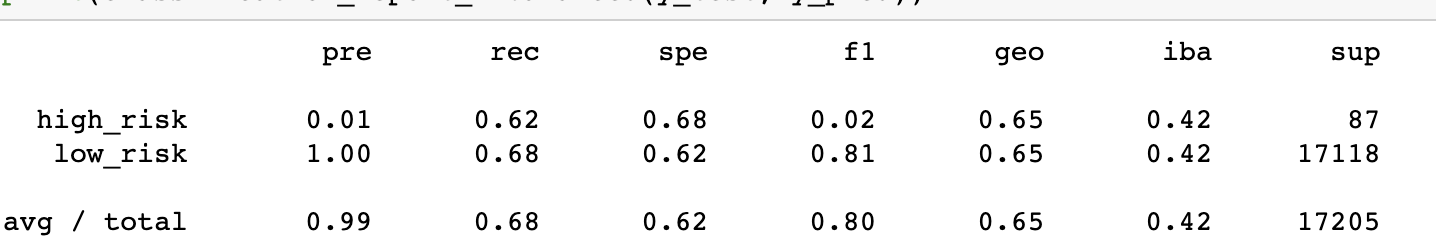


Predicted 54 times that it was high risk and it was high risk – TP

Predicted 5498 times that it was high risk and it was not high risk – FP

Predicted 33 times it was NOT high risk but it was high risk -FN

Predicted 11620 it was NOT high risk and it was NOT high risk - TN



**High Risk**

* Recall:0.62
* Precision:0.01
* F1-Score:0.02
* Summary on Statistics: Very low precision and even F1 score for high risk customers

**Low Risk**

* Precision:1
* Recall:0.68
* F1-Score:0.81
* Summary on Statistics: Acceptable for low risk but not acceptable for high risk customers

**SMOTE Algorithm**

**Accuracy Score:**

The accuracy score was slightly lower than RandomOversampler 67.8% vs 65.6%



**Confusion Matrix:**



Predicted 55 times that it was high risk and it was high risk – TP

Predicted 5879 times that it was high risk and it was not high risk – FP

Predicted 32 times it was NOT high risk but it was high risk -FN

Predicted 11239 it was NOT high risk and it was NOT high risk - TN



**High Risk**

* Precision: 0.01
* Recall:0.63
* F1-Score:0.02
* Summary on Statistics: Slightly better than RandomOverSampler for high risk but still pretty low and not good to predict high risk credit card customers

**Low Risk**

* Precision:1
* Recall:0.68
* F1-Score:0.81
* Summary on Statistics: Acceptable for low risk but not acceptable for high risk customers

**Undersampling methodologies**

Undersampling techniques ClusterCentroids was applied to majority group to make it same size as the minority group

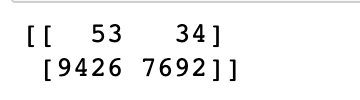
****

**Accuracy Score:**

****

RandomOversampler 67.8% vs SMOTE was 65.6% **vs ClusterCentroid 45.01%**

**Confusion Matrix:**

****

Predicted 53 times that it was high risk and it was high risk – TP

Predicted 9426 times that it was high risk and it was not high risk – FP

Predicted 34 times it was NOT high risk but it was high risk -FN

Predicted 7692 it was NOT high risk and it was NOT high risk – TN



**High Risk**

* Precision: 0.01
* Recall:0.61
* F1-Score:0.01
* Summary on Statistics: Similar to SMOTE but little less F1 score 0.01 vs 0.02

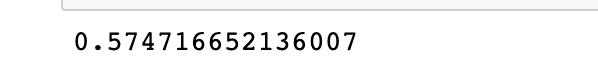
**Low Risk**

* Precision:1
* Recall:0.68
* F1-Score:0.62
* Summary on Statistics: Performs poorly with recall and not predicting low risks well as well as high risk, poor F1 score and recall reflects that along with accuracy score of 45%.

**Undersampling and Oversampling – SMOOTEN**

**Accuracy Score:**

The accuracy score is 57.4% which is lower than undersampling and oversampling methodologies.



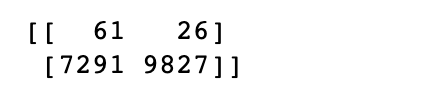
**Confusion Matrix:**

Predicted 61 times that it was high risk and it was high risk – TP

Predicted 7291 times that it was high risk and it was not high risk – FP

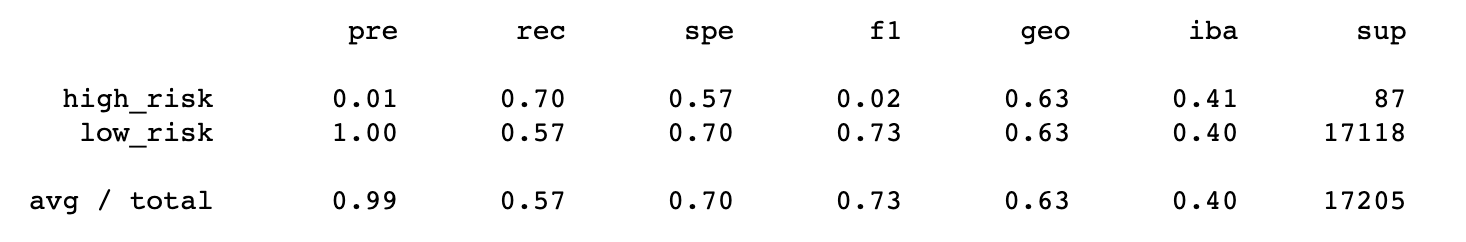
Predicted 26 times it was NOT high risk but it was high risk -FN

Predicted 9827 it was NOT high risk and it was NOT high risk – TN



**High Risk**

* Precision: 0.01
* Recall:0.70
* F1-Score:0.02
* Summary on Statistics: Better recall than over sampling and undersampling techniques for high risk and slightly better f1 score

****

**Low Risk**

* Precision: 1.00
* Recall:0.0.57
* F1-Score:0.73
* Summary on Statistics: Poorer recall and F1 score for SMOTENN algorithm in comparison with oversampling methodologies but slightly better than undersampling methodology.

**Ensemble Machine Learning Algorithms**

**Balanced Random Forest Classifier**

A balanced random forest randomly under-samples each bootstrap sample to balance it.

It involves first selecting random samples of a training dataset with replacement, meaning that a given sample may contain zero, one, or more than one copy of examples in the training dataset. This is called a bootstrap sample

**Accuracy Score:**

The accuracy score was 87.3% which was much higher than sampling methodologies.

****

**Confusion Matrix:**

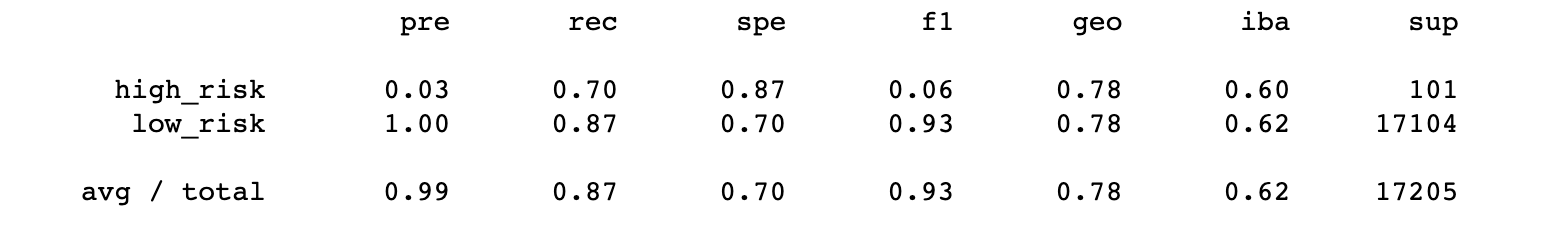
****

Predicted 71 times that it was high risk and it was high risk – TP

Predicted 2153 times that it was high risk but it was not high risk – FP

Predicted 30 times it was NOT high risk but it was high risk -FN

Predicted 14951 it was NOT high risk and it was NOT high risk - TN

****

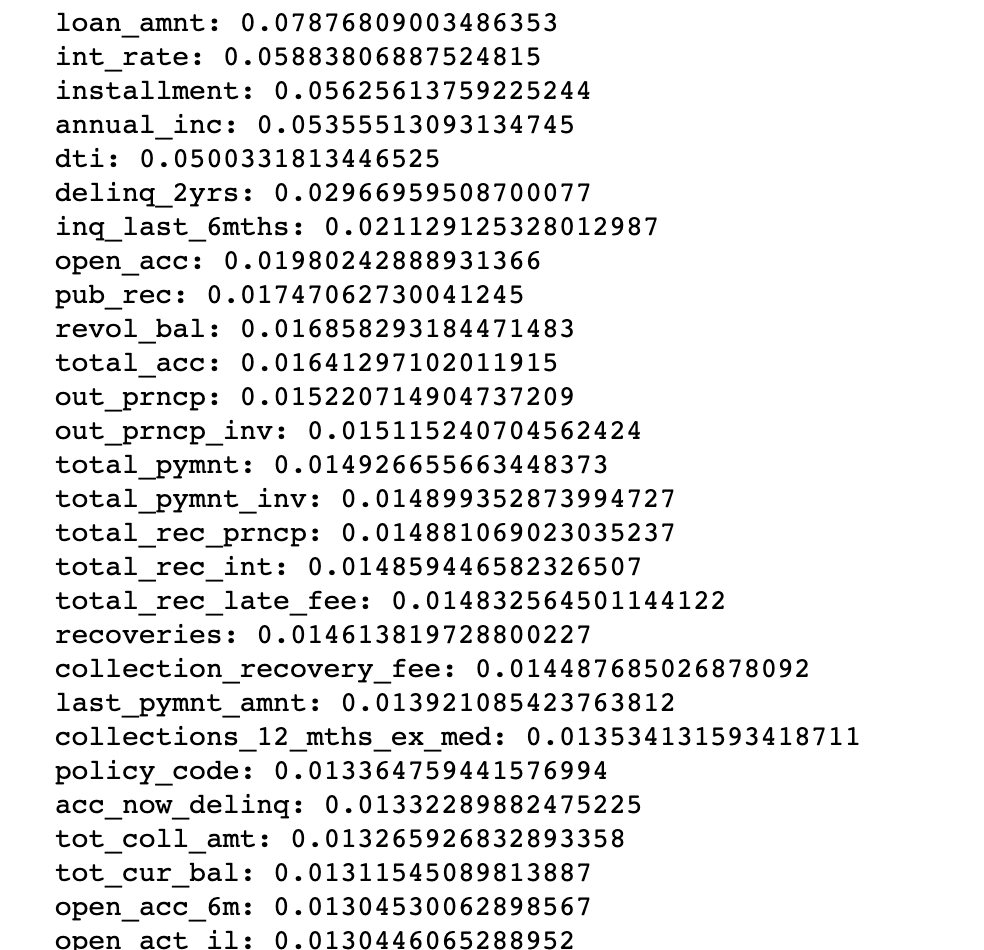
**High Risk**

* Precision: 0.03
* Recall:0.70
* F1-Score:0.06
* Summary on Statistics: Much better recall from sampling methodologies and better f1 score but still too many high risks predicted when they were not, 2153 hence low precision

**Low Risk**

* Precision:1
* Recall:0.87
* F1-Score:0.93
* Summary on Statistics: Much better predictions and improvements from sampling methodologies.

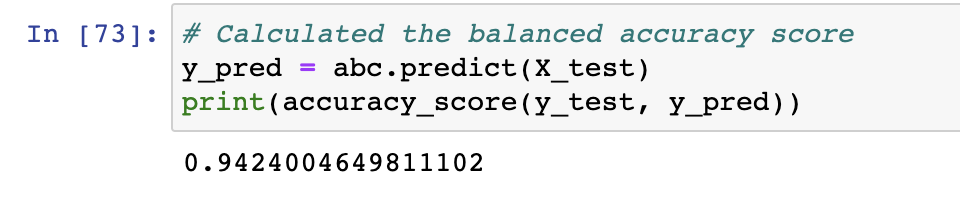
Loan amount, interest rates and instalments were the key variable which showed to be top 3 most important variables.



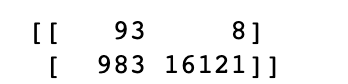
**Easy Ensemble AdaBoost Classifier**

**Accuracy Score:**

The accuracy score was 94.2% vs 87.3% for BalancedClassifier .



**Confusion Matrix:**

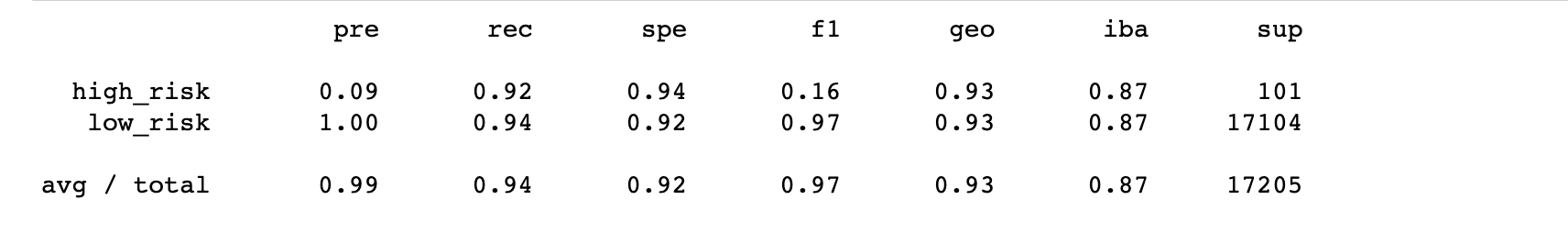
****

Predicted 93 times that it was high risk and it was high risk – TP

Predicted 983 times that it was high risk but it was not high risk – FP

Predicted 8 times it was NOT high risk but it was high risk -FN

Predicted 16121 it was NOT high risk and it was NOT high risk – TN



**High Risk**

* Precision: 0.09
* Recall:0.92
* F1-Score:0.16
* Summary on Statistics: Much better recall 0.92 and f1 score but still too many high risks predicted when they were not, 983 hence low precision but it still categorises much better and is the best performing one

**Low Risk**

* Precision: 1
* Recall:0.94
* F1-Score:0.97
* Summary on Statistics: Best F1-Score out of all 6 algorithms

**SUMMARY**

**Resampling Techniques**

Resampling is a widely used approach for coping with very imbalanced datasets. It entails eliminating samples from the majority class and/or adding more examples from the minority class (under-sampling) (over-sampling).

In our case we used Oversampling, undersampling and over and under sampling and found out that Oversampling performed better than SMOTTEN. Both SMOTE and RandomOverSampler performed in line and even though there F1 score was very low F1=0.02 and Precision was only 0.02 for both, but they still performed better than SMOTTEN.

**Ensemble ML algorithms**

Easy Ensemble AdaBoost Classifier performed much better than all the algorithms Logistic regressions (after re-sampling) and balanced classifier. The accuracy score was 94.2% for Ensemble AdaBoost vs 87.3% for BalancedClassifier . The F1 score for high risk was 0.16 but recall had a much better improvement from all the other algorithm used with 0.92.