Intelligent Systems
Dr. Octavio Loyola-Gonzalez

Nicolas Gonzalez Albo Mendez
A01730763

H5 – Testing two resampling methods or classifiers for solving class imbalance problems

Introduction

Class imbalance is a serious and real problem in Machine Learning. For solving these kind of problems, two solutions are proposed, resampling methods and classifiers. In this homework, two different Resampling methods are used with a CostSensitiveClassifier.

Development & Results

Resample with CostSensitiveClassifier and LMT

The first step was using the Resample Method in Weka with the original dataset of Training.csv

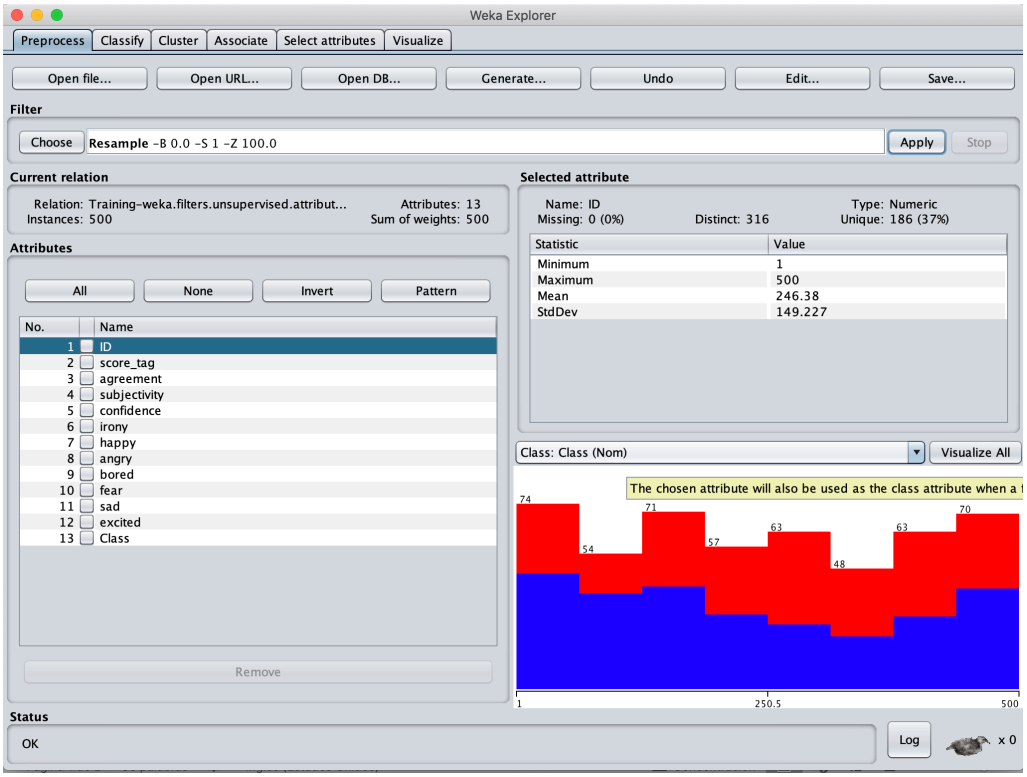

Figure 1. Resampling Training.csv in WEKA

Nicolas Gonzalez Albo Mendez
A01730763

Then, a CostSensitiveClassifier was applied with an LMT.

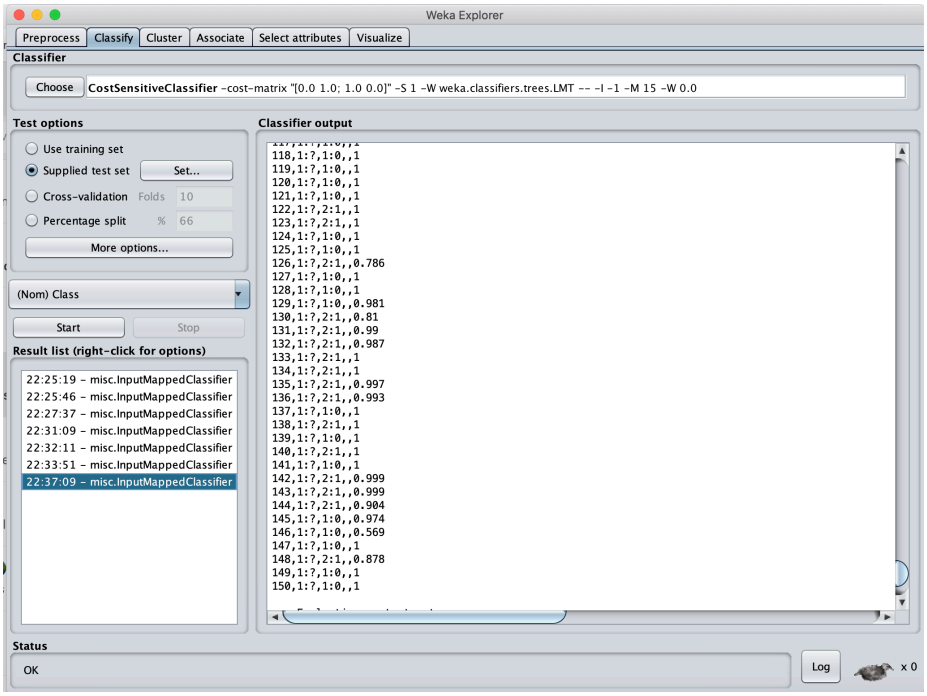

Figure 2. Results of the LMT in WEKA

After uploading the WEKA's Output to Kaggle, it can be appreciated that sadly it did not scored better than by highest score.

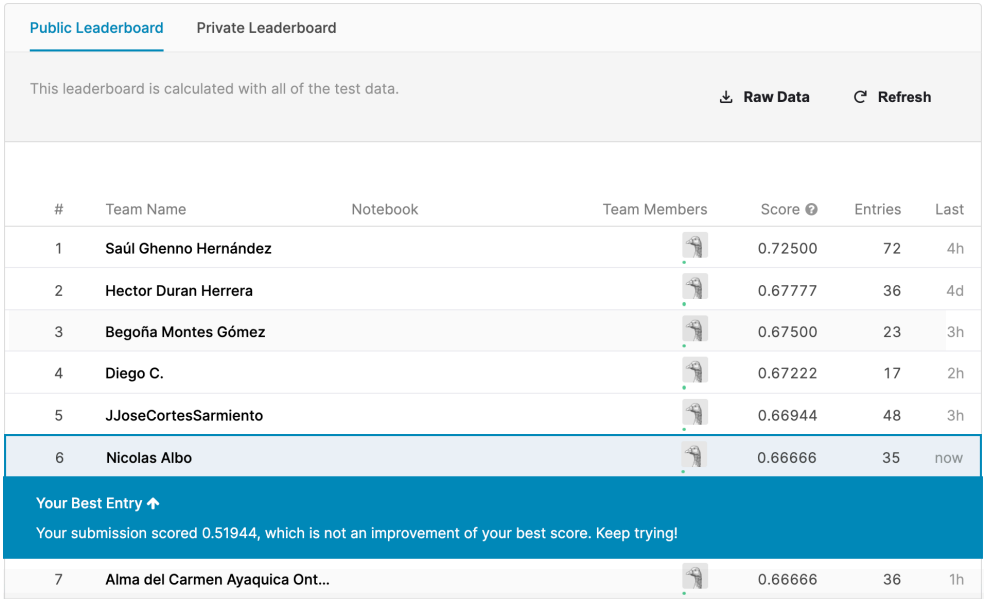

Figure 3. Score in Kaggle

SpreadSubsample with CostSensitiveClassifier and LMT

In this case, the first step was using the SpreadSubsample Method in Weka with the original dataset of Training.csv

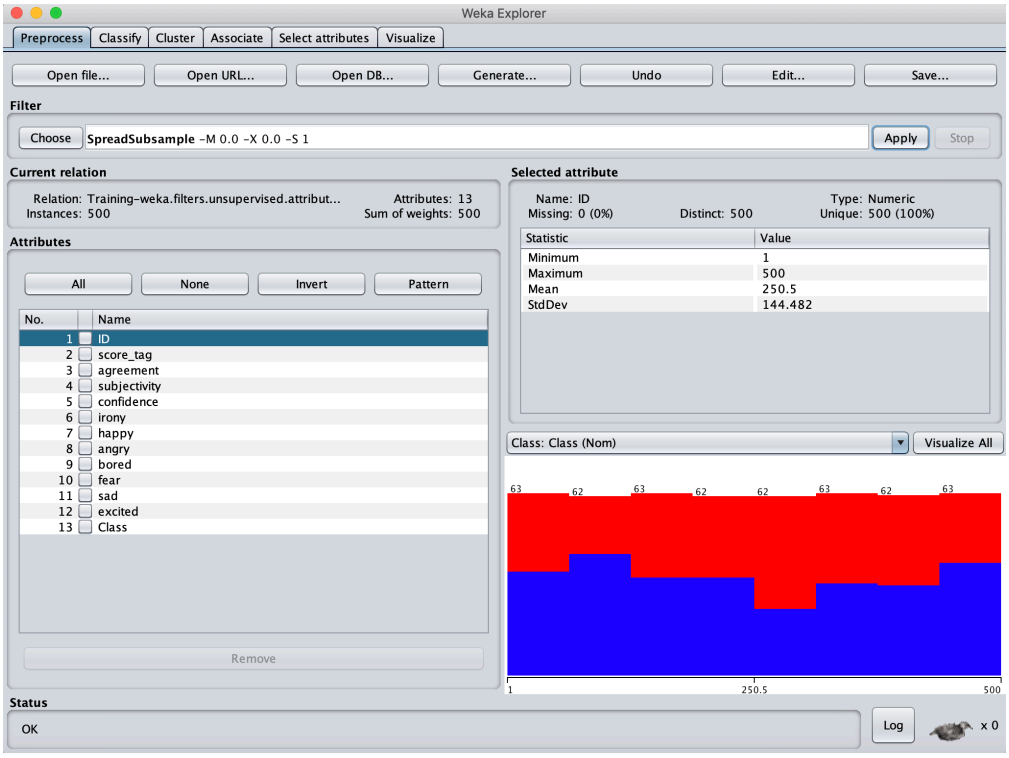

Figure 4. SpreadSubsample method applied in Training.csv in WEKA

Nicolas Gonzalez Albo Mendez
A01730763

Then, a CostSensitiveClassifier was applied with an LMT.

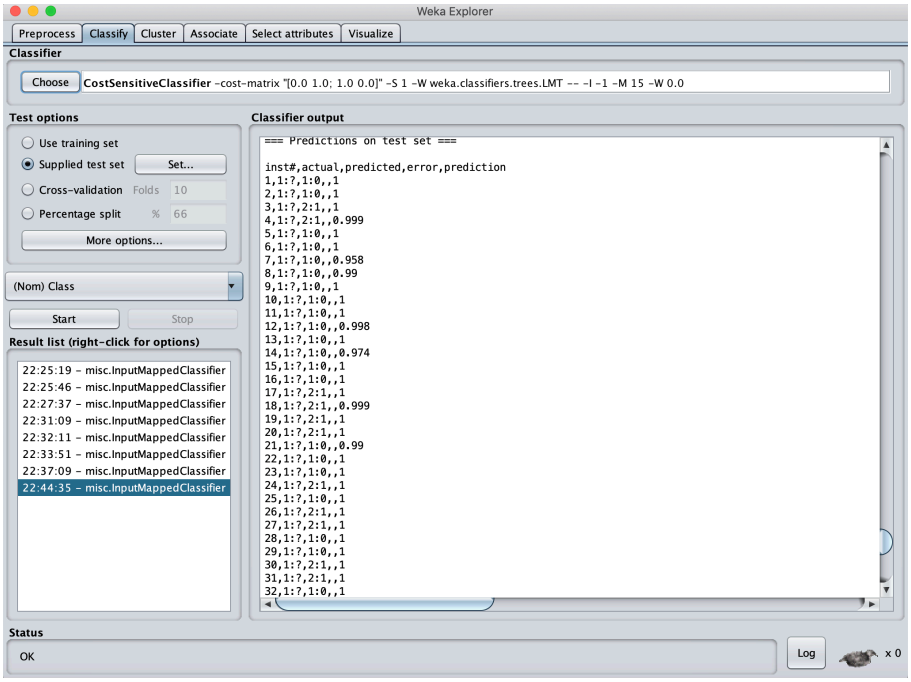

Figure 5. Results of the LMT in WEKA

After uploading the WEKA's Output to Kaggle, it can be appreciated that sadly it did not scored better than by highest score.

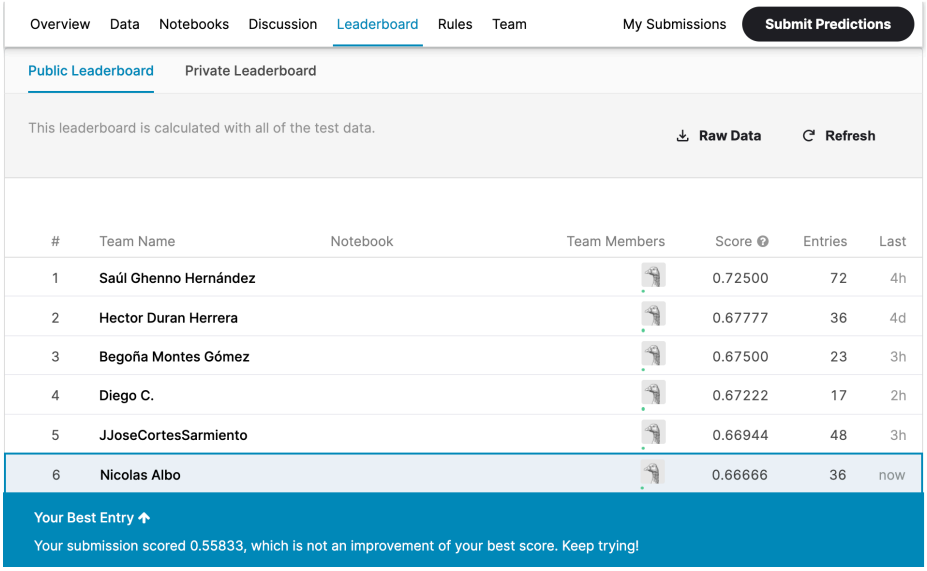

Figure 6. Score in Kaggle

Resample with CostSensitiveClassifier and Hoeffding Tree

In this last example, the first step was using the Resample Method in Weka with the original dataset of Training.csv

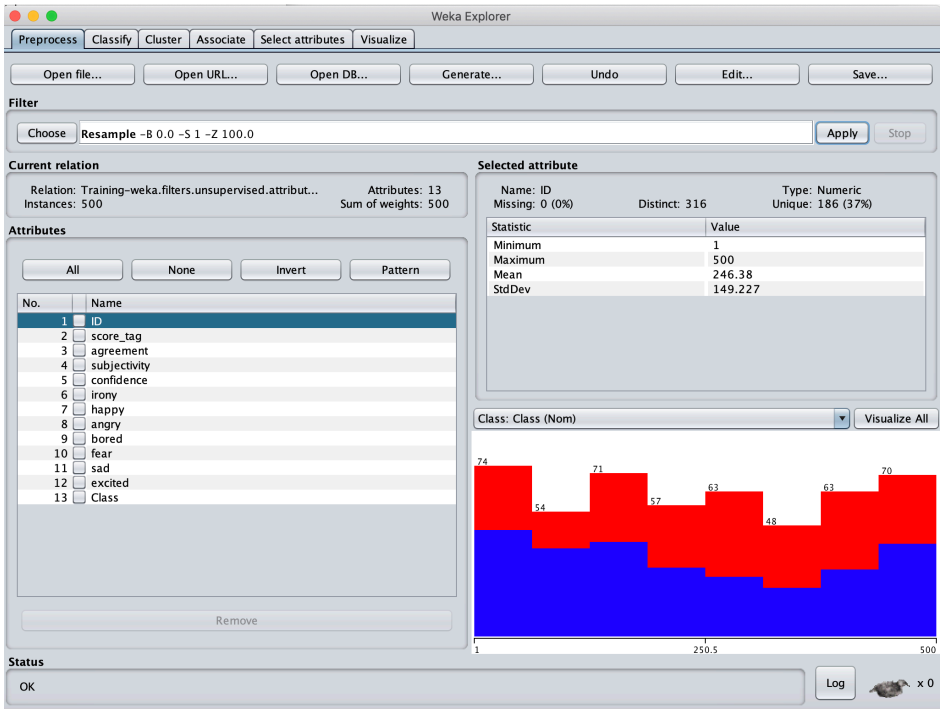

Figure 7. Resampling Training.csv in WEKA

Then, a CostSensitiveClassifier was applied with a Hoeffding Tree.

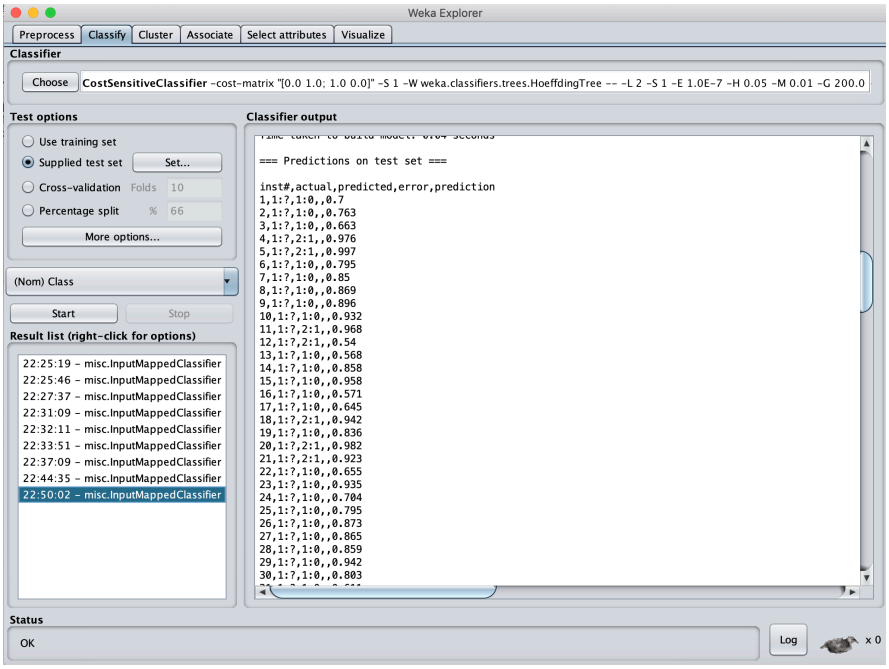

Figure 8. Results of the Hoeffding Tree in WEKA

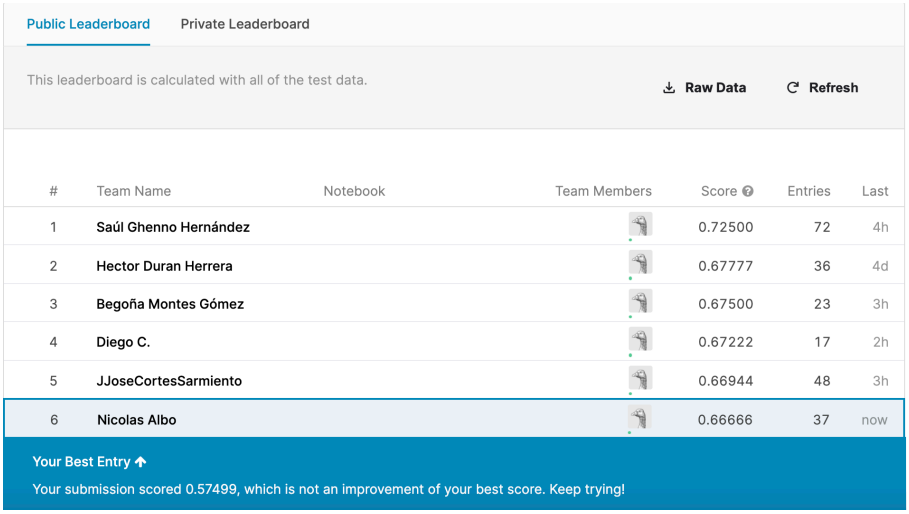

Figure 9. Score in Kaggle

Nicolas Gonzalez Albo Mendez
A01730763

Conclusion

Resampling methods or classifiers are a proper solution for class imbalance problems. But, in this case, these methods did not help me to improve my high score in Kaggle. Maybe I did not use an appropriate resampling method that could have helped me to rank higher in Kaggle.