Intelligent Systems
Dr. Octavio Loyola-Gonzalez

Nicolas Gonzalez Albo Mendez
A01730763

H7 – Testing the Bayes Network method

Introduction

A Bayesian Network is a probabilistic model where variables and their conditional dependencies are represented through a DAG (directed acyclic graph). They are useful for predicting the factors that caused a given problem. In this homework, two different Bayesian Networks are going to be implemented in order to find if news is real of fake.

Development & Results

First of all, a new feature (News) was added to the ones used in the previous homework, the feature was extracted using the Intent API of Parallel Dots, both for the Training and Testing datasets.

Bayes Network with SimpleEstimator and K2 Search Algorithm

For this case, BayesNet is selected as the classifier along with SimpleEstimator and K2 as search algorithm in WEKA.

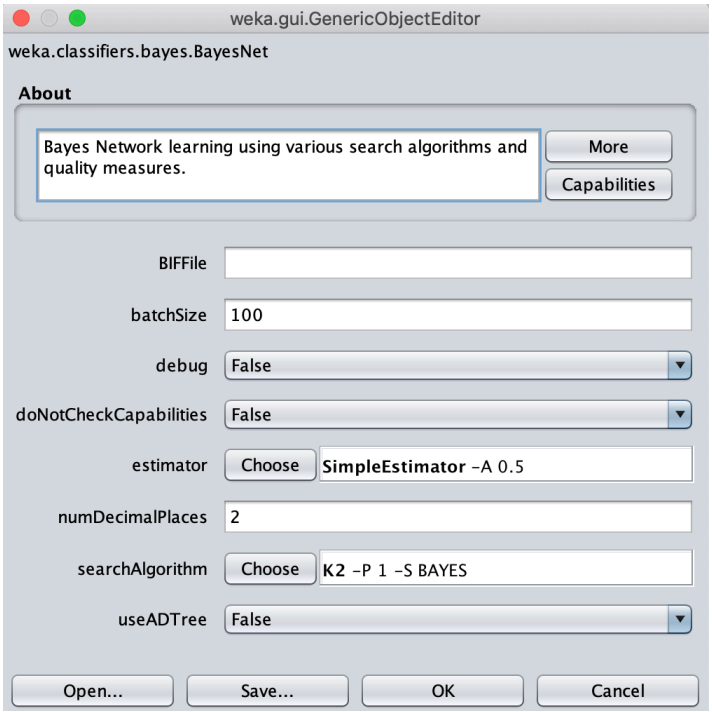

Figure 1. Configuration for a Bayes Network classifier

Nicolas Gonzalez Albo Mendez
A01730763

Then, the following data is obtained.

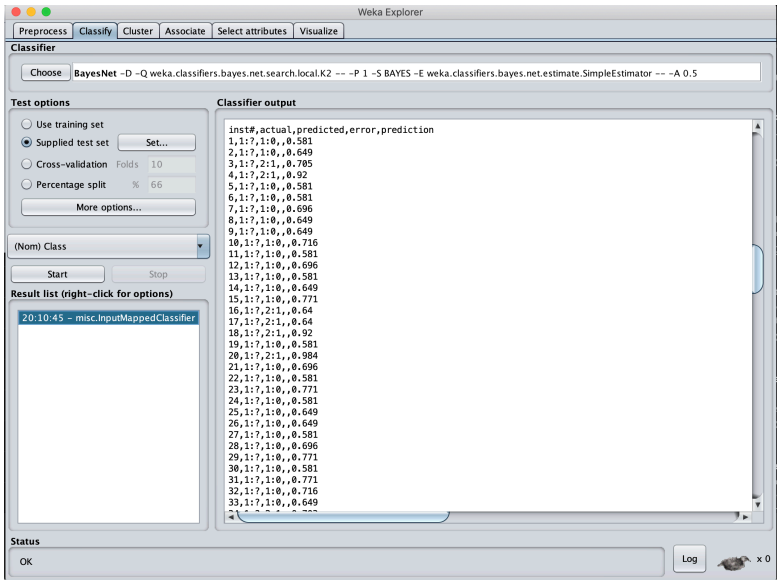

Figure 2. Results obtained from our Bayes Network

Using a Bash script named Filter.sh, the data is given the format required by Kaggle. After that, the results are uploaded to Kaggle.

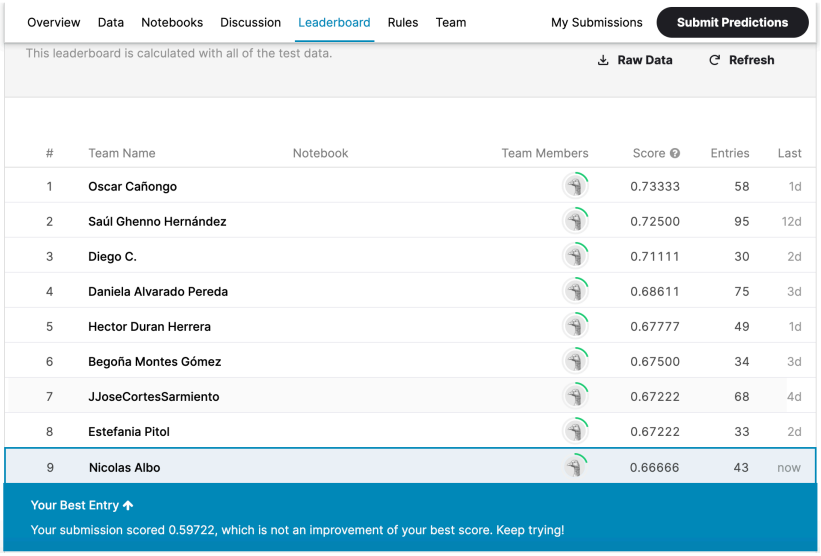

Figure 3. Performance of the results in Kaggle

Sadly, the results obtained by this classifier didn't perform well, in fact it is below my best score.

Nicolas Gonzalez Albo Mendez
A01730763

NaiveBayes Classifier

The second variant of the Bayes Network classifier used in this homework is NaiveBayes.

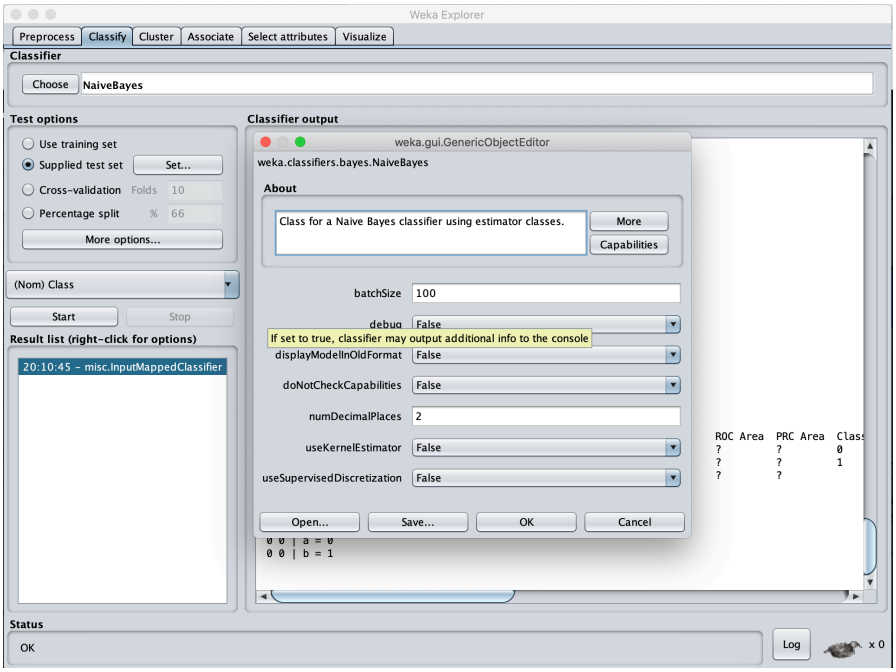

Figure 4. Configuration for NaiveBayes

The following data is obtained by the classifier.

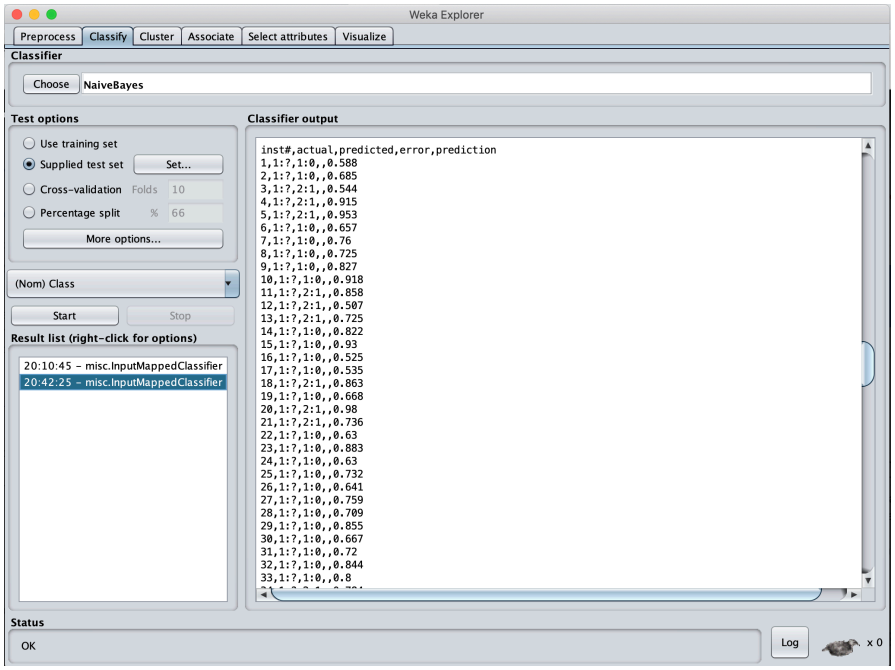

Figure 5. Results obtained by the NaiveBayes classifier.

Nicolas Gonzalez Albo Mendez
A01730763

Like in the previous variant, the script Filter.sh was used for preparing the data for Kaggle. Then the results where uploaded to Kaggle.

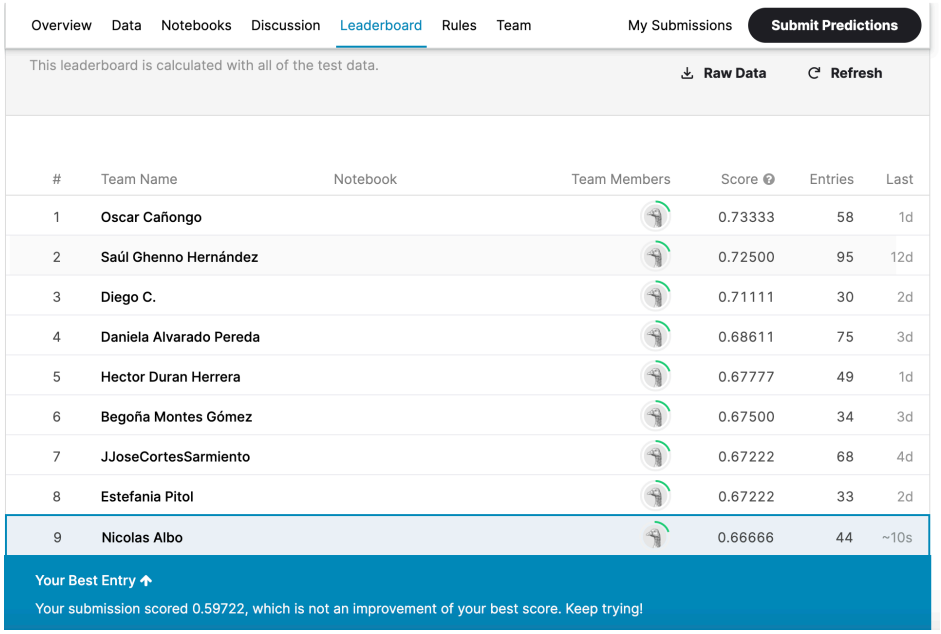

Figure 6. Performance of the results in Kaggle

Unfortunately, like the previous attempt, it didn't rank higher than my best score in Kaggle.

Conclusion

Bayesian Networks are powerful classifiers, but they have some disadvantages like the time that they require for training is excesive, an average of two hours is required for training these classifiers. The good news is that in WEKA there is the option to choose Bayesian Networks classifiers that have been already trained like BayesNet and NaiveBayes that are useful in situations where we don't have enough time for training.