



Faculty of Engineering and Technology
Electrical and Computer Engineering Department

ENCS3340

Artificial Intelligence

Project #2

Machine Learning for Classification

Prepared by:

Diana Naseer ID: 1210363

Hala Jebreel ID: 1210606

Instructor's Name:

Dr. Yazan Abufaraha

Section: 4

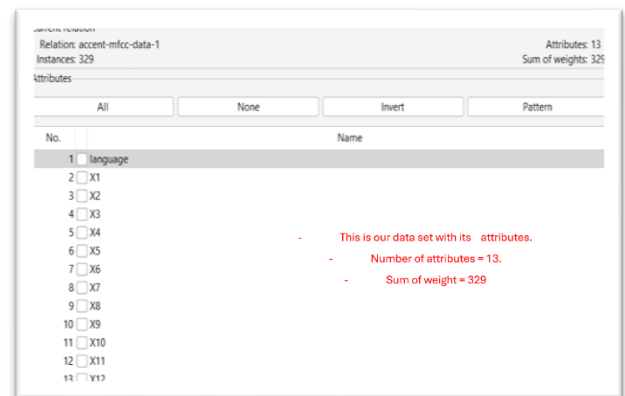
20th Jun 2024

Introduction:

In this project three machine learning algorithms were implemented using WEKA tool. The data set was Accent data set, and it was chosen based on the last digit of the least student id in

the team % 3 \rightarrow 1210363 \rightarrow 3%3 \rightarrow 0.

-so, we take the data set 0.



This data set contains:

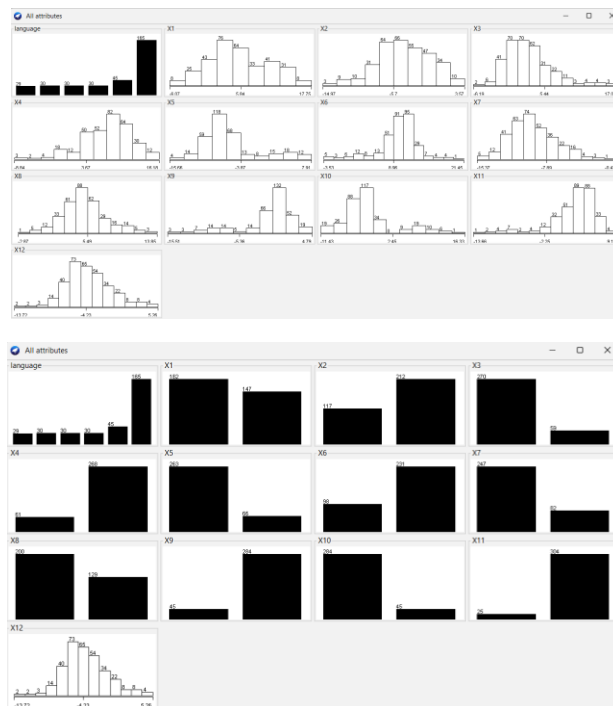
language: Object (categorical data representing the language).

X1 to X12: Float64 (numerical data representing different MFCC features).

The dataset contains a mixture of positive and negative values for each feature, indicating a wide range of MFCC (Mel-frequency cepstral coefficients) values for different languages. Each entry corresponds to a set of MFCC features for a particular language sample.

1- Decision Tree

1. Discretization:



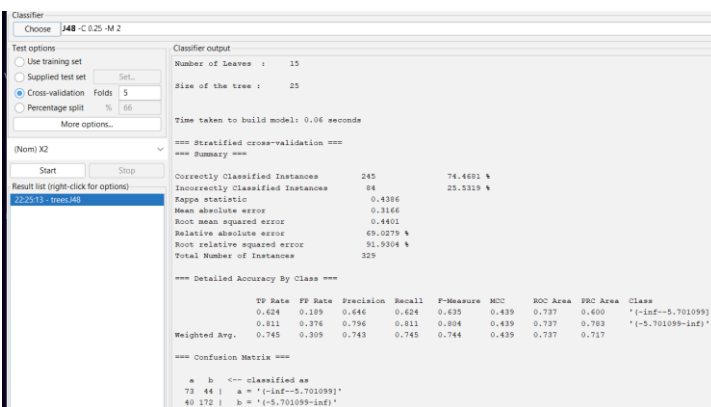
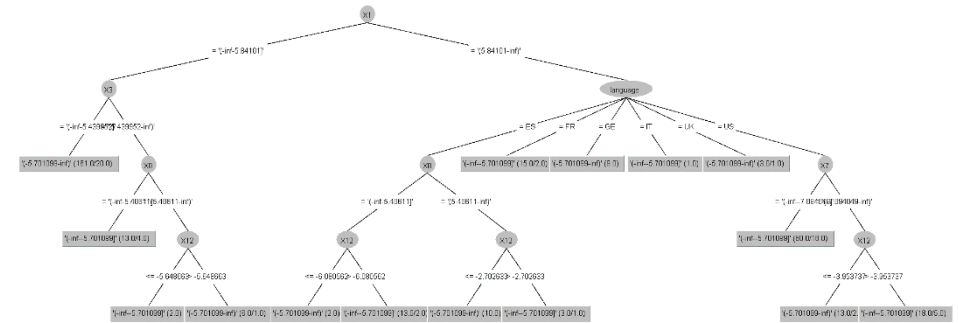
Here is the plot of the data before do anything on it.

Now, the plots after the discretization with bins=2 for all attributes.

Using the 5-fold cross validation to train the model and applying Decision Tree algorithm as a Classifier.

- In This case where bins are 2 for:

```
Number of Leaves :      15
Size of the tree :      25
```



- From the figure, it is shown that the correctly classified instances are 74.468%.

TP:73

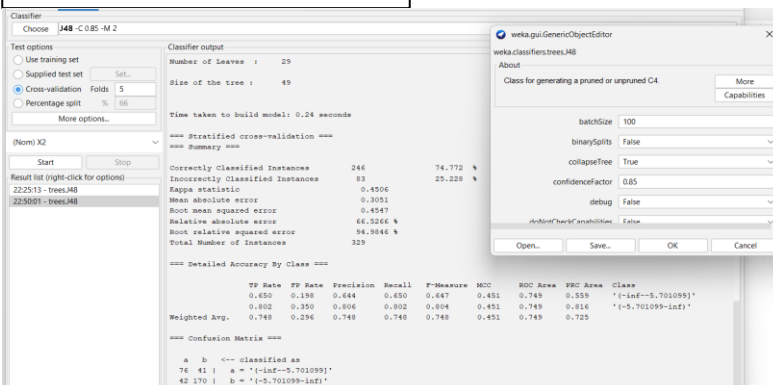
FN:44

FP:40

TN:172

As the precision is 0.646, recall 0.624, F-Measure = 0.635. These results could be acceptable.

Change the hyper-parameter:



Here we Changes hyper-parameter "Confidence Factor" from 0.25 to 0.85:

And as we see from the results:

- 1- The correctly classified change from 74.4681 to 74.772 and the incorrectly changed from 25.5319 to 25.228 , so we noticed that the performance is got better.

2- The new calculation is:

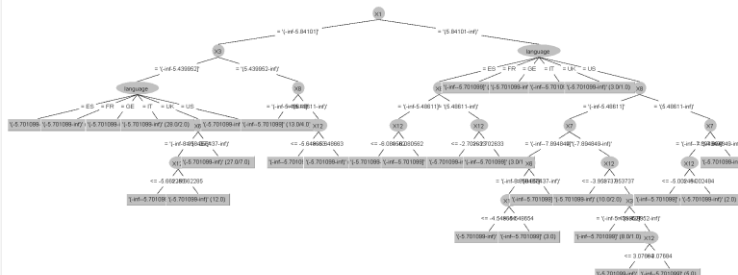
TP:76

FN:41

FP:42

TN:170

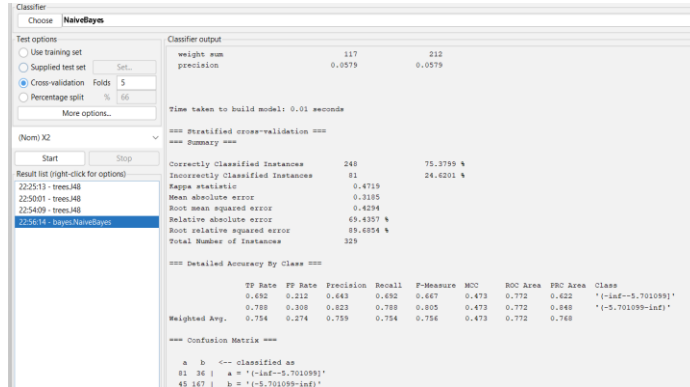
As the precision is 0.644, recall 0.650, F-Measure = 0.647. These are the new results and we noticed that the difference is not big, but it's still got better a little.



2-Naïve Bayes:

Using the 5-fold cross validation to train the model and applying Decision Tree algorithm as a Classifier.

1-Classification:



- From the figure, it is shown that the correctly classified instances are 75.3799%.

TP:81

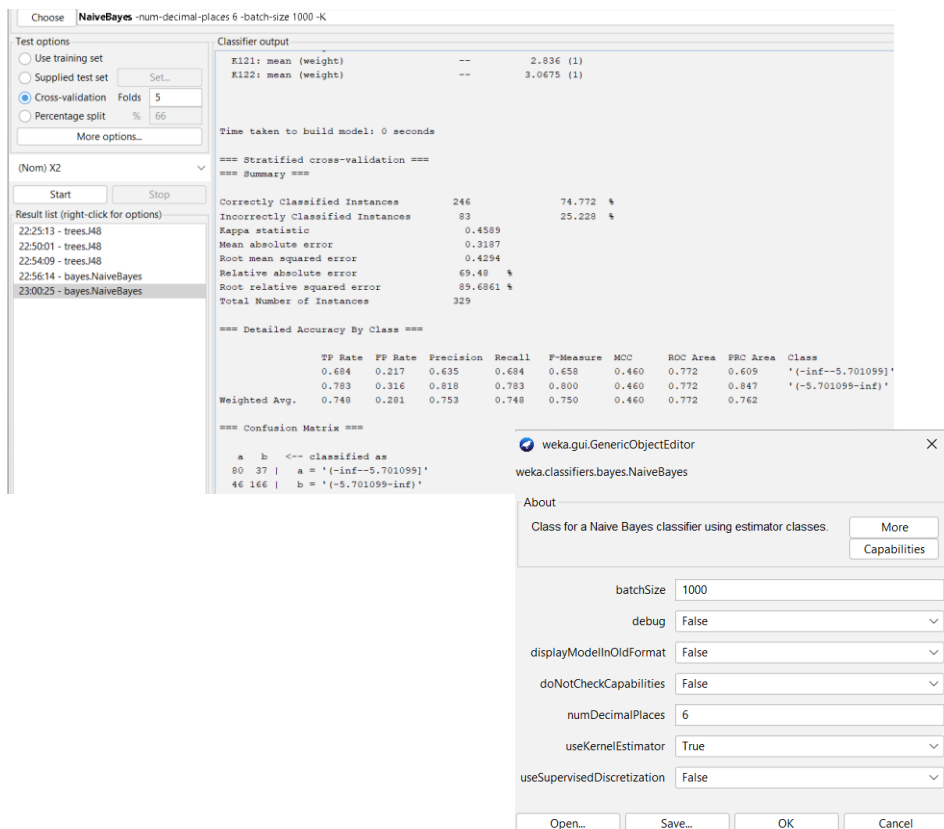
FN:36

FP:45

TN:167

As the precision is 0.643, recall 0.692, F-Measure = 0.667.

2- Change hyper-parameter:



Here we Changes hyper-parameter "Batch Size" from 100 to 1000, " numDecimalPlaces "from 2 to 6, and "use kernel estimator" form false to true :

And as we see from the results:

- the correctly classified change from 75.37999 to 74.772 and the incorrectly changed from 24.6201 to 25.228.

2-

TP:80

FN:37

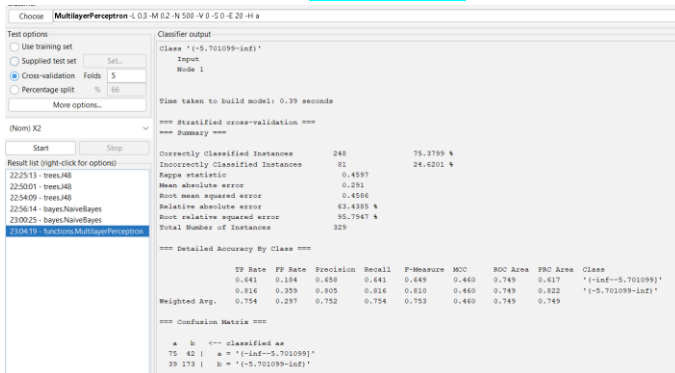
FP:46

TN:166

- As the precision is 0.635, recall 0.684, F-Measure = 0.658 . These is the new results and we noticed that the difference not big.

3- MLP

1- Classification:



- From the figure, it is shown that the correctly classified instances are 75.3799 %.

TP:75

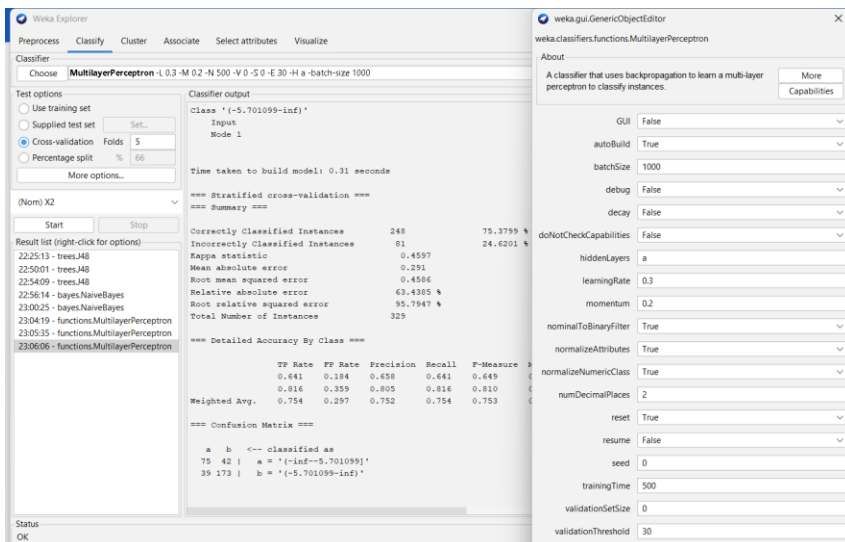
FN:42

FP:39

TN:173

As the precision is 0.658, recall 641, F-Measure = 0.649. These results could be acceptable.

2- Change hyper-parameter:



Here we Changes hyper-parameter "Batch Size" from 100 to 1000 :

And as we see from the results:

1- the correctly classified still the same = 75.3799

2-

TP:75

FN:42

FP:39

TN:173

- As the precision still the same.
- So, the results not changed.

Conclusion:

From the table, we got that the better performance for MLP & Naïve with 75.3799%.

Model	Decision Tree	Naïve Bayes	MLP
Performance before changing the hyper parameter	74.468	75.3799	75.3799
Performance after changing the hyper parameter	74.772	74.772	75.3799