



BIRZEIT UNIVERSITY

FACULTY OF ENGINEERING AND TECHNOLOGY

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Artificial Intelligence ENCS4380

Project 2 : Machine Learning

Student's Names and ID's :

Raghad Afaghani-1192423

Sara Bader-1190889

Instructors:

Dr.Adnan Yahya

Dr.Yazan Abo Farha

Date:12/6/2022

Specifications

- Tool Used

We used WEKA 3.8.6 to preprocess, classify and simulate the testing results.

- Test set used

According to the least digit ID in our group, we used test set number zero which contains Data set featuring single English words read by speakers from six different countries for accent detection and recognition.

- Algorithms selection

We need to test three models: Decision Trees, Naïve Bayes and one more model of our choice which is Random Forest.

- Evaluation

The evaluation is based on the following metrics: Accuracy, Recall, Precision, and F-Score.

Accuracy = # of correct classified samples / # of total used samples

Precision (P) = TP / (TP + FP)

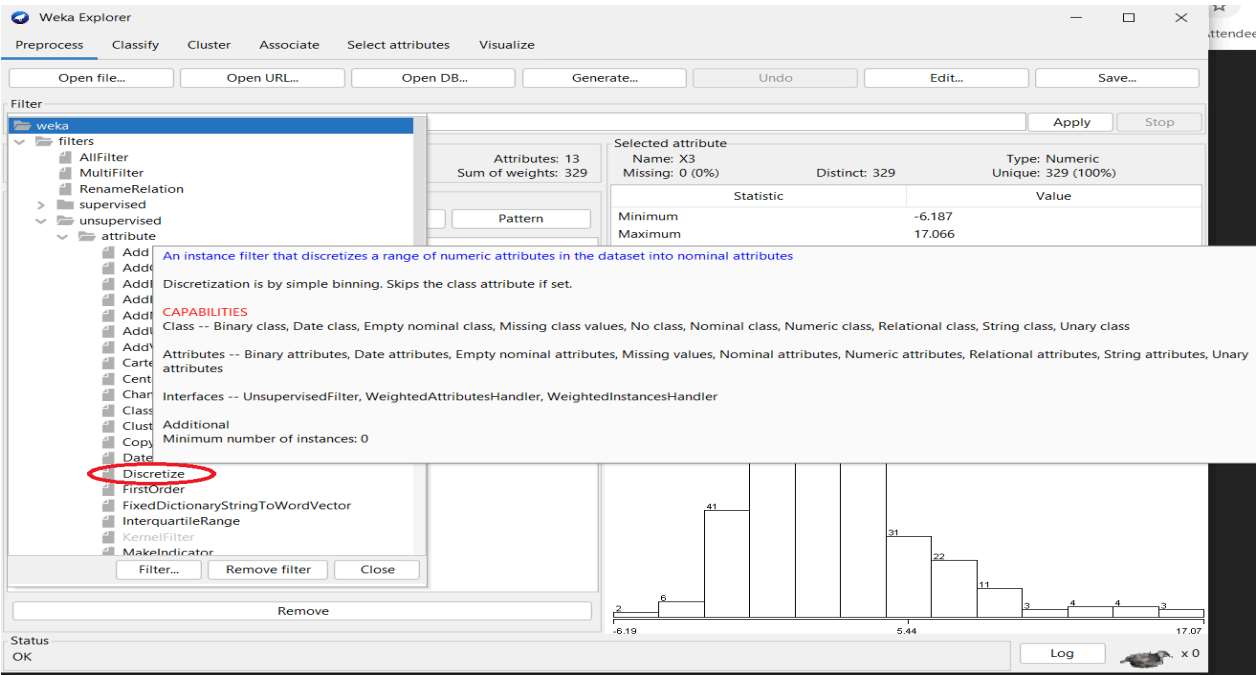
Recall (R) = TP / (TP + FN)

F-Score = 2PR / (P + R)

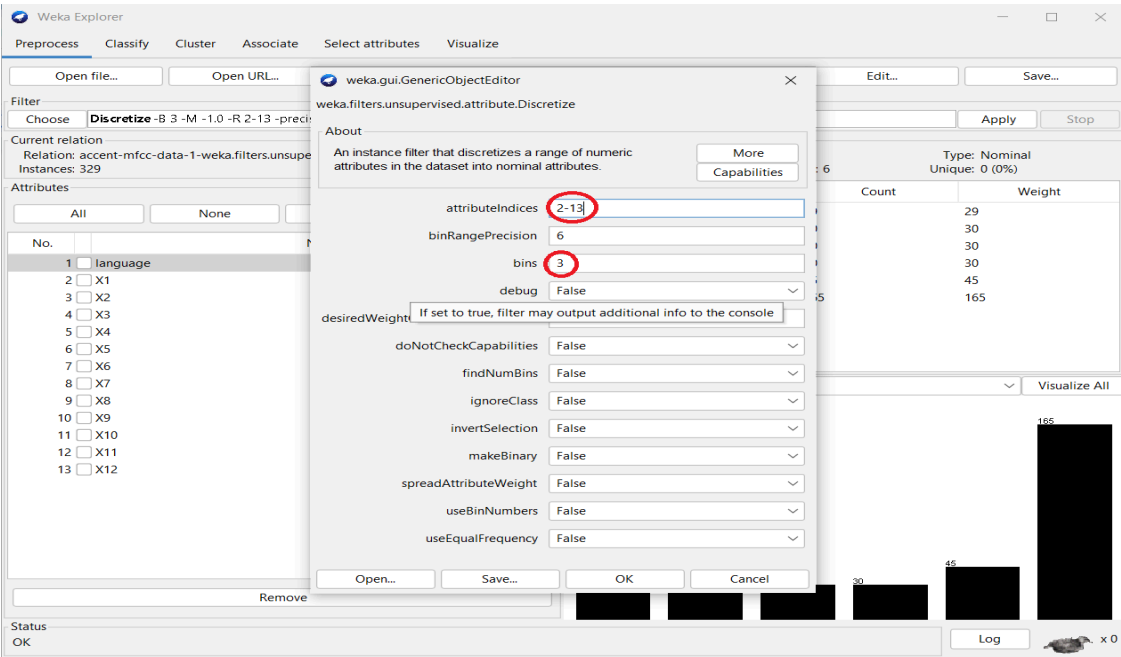
Results

- Preprocessing

Discretization of continuous attributes:



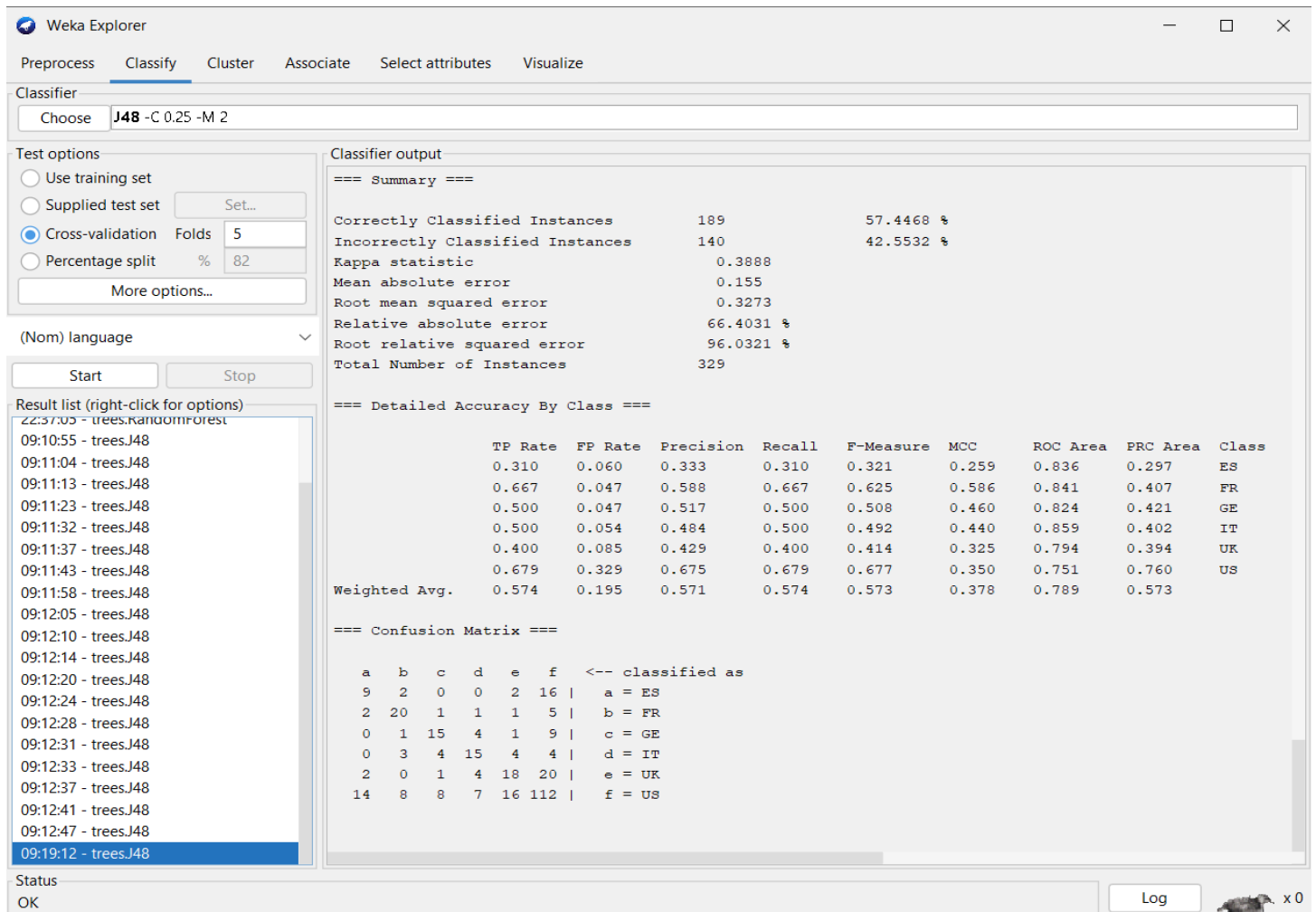
Figure[1]



Figure[2]

- Testing the model using 5-fold cross validation and reporting the confusion matrix, accuracy, precision, recall, and F1-score.

1. Decision Tree



Figure[3]

$$\text{accuracy} = 189/329 = 57.4468\%$$

$$\text{Precision} = 0.571$$

$$\text{Recall} = 0.574$$

$$\text{F-measure} = 0.573$$

=== Confusion Matrix ===

a b c d e f <-- classified as

9 2 0 0 2 16 | a = ES

2 20 1 1 1 5 | b = FR

0 1 15 4 1 9 | c = GE

0 3 4 15 4 4 | d = IT

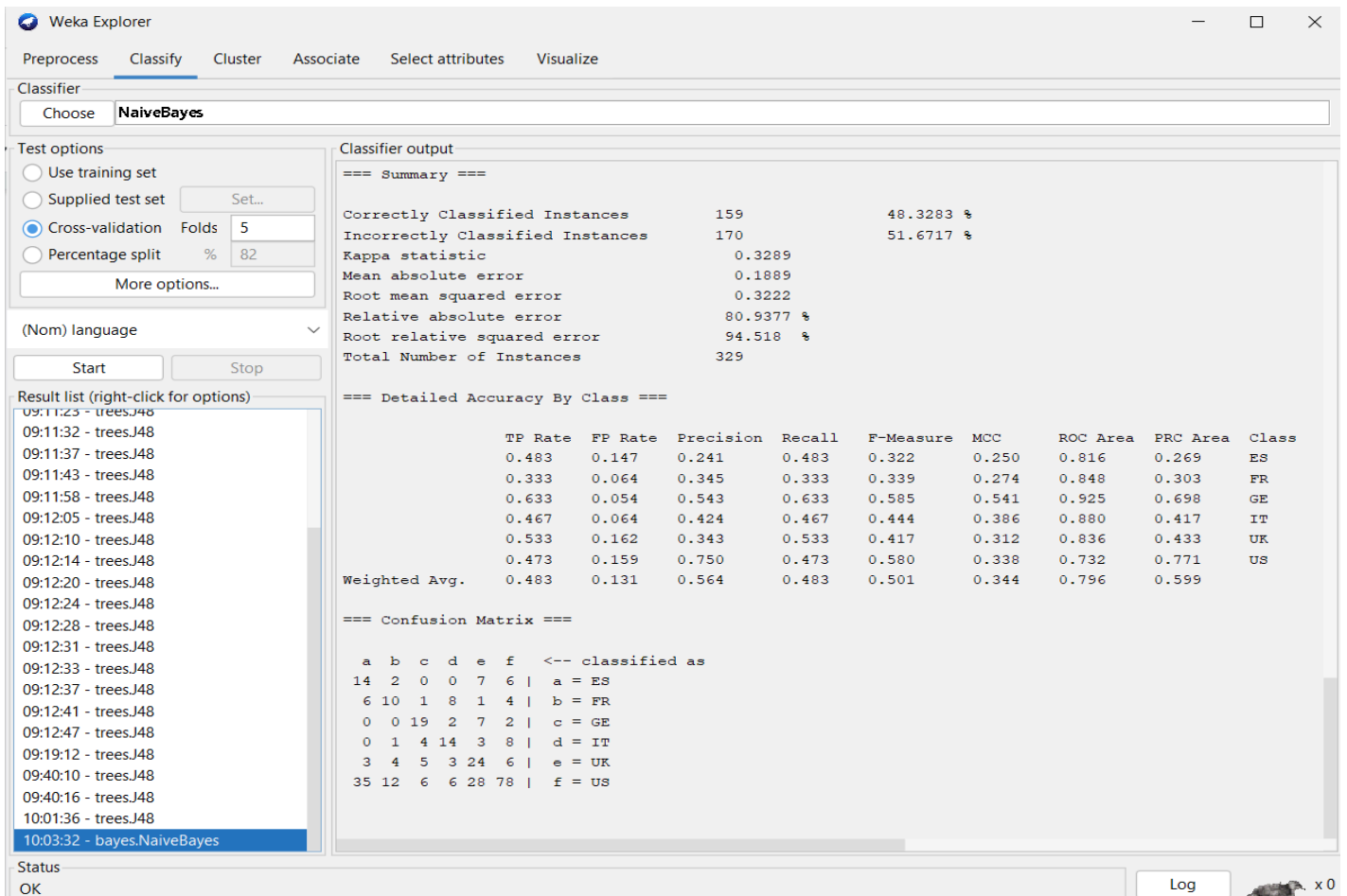
2 0 1 4 18 20 | e = UK

14 8 8 7 16 112 | f = US

- **9 : it's ES and classified as ES**
- **2 : it's ES and classified as FR**
- **2 : it's ES and classified as UK**
- **16 : it's ES and classified as US**
- 2 : it's FR and classified as ES
- 20 : it's FR and classified as FR
- 1 : it's FR and classified as GE
- 1 : it's FR and classified as IT
- 1 : it's FR and classified as UK
- 5 : it's FR and classified as US

And so on,,,,

2. Naïve Bayes



Figure[4]

$$\text{accuracy} = 159/329 = 48.3283\%$$

$$\text{Precision} = 0.564$$

$$\text{Recall} = 0.483$$

$$\text{F-measure} = 0.501$$

=== Confusion Matrix ===

a b c d e f <-- classified as

14 2 0 0 7 6 | a = ES

6 10 1 8 1 4 | b = FR

0 0 19 2 7 2 | c = GE

0 1 4 14 3 8 | d = IT

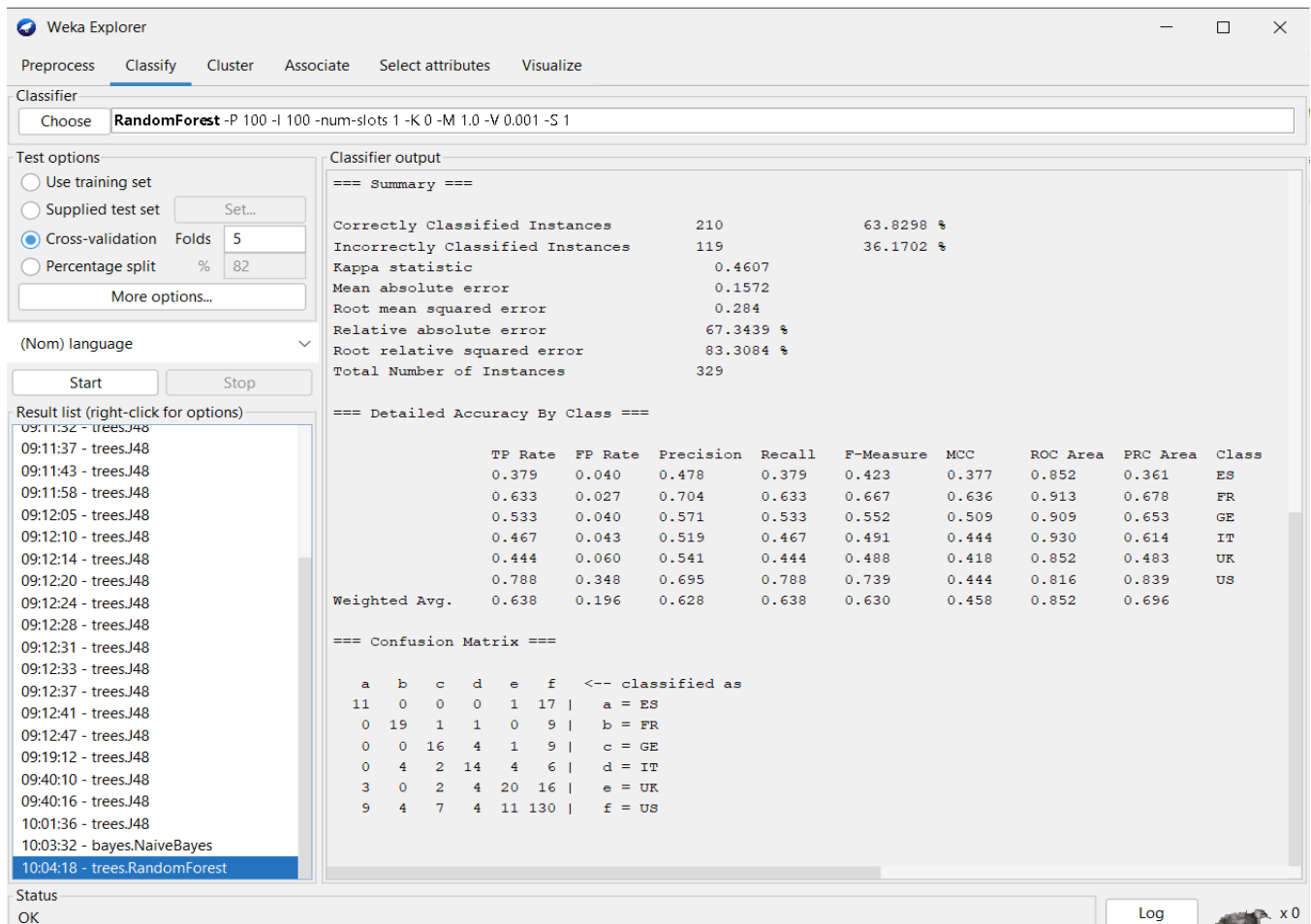
3 4 5 3 24 6 | e = UK

35 12 6 6 28 78 | f = US

- 14 : it's ES and classified as ES
- 2 : it's ES and classified as FR
- 7: it's ES and classified as UK
- 6 : it's ES and classified as US
- 6 : it's FR and classified as ES
- 10 : it's FR and classified as FR
- 1 : it's FR and classified as GE
- 8 : it's FR and classified as IT
- 1 : it's FR and classified as UK
- 4 : it's FR and classified as US

And so on,,,,

3. Random Forest



Figure[5]

accuracy = $210/329 = 63.8298\%$

Precision= 0.628

Recall = 0.638

F-measure = 0.630

=== Confusion Matrix ===

a b c d e f <-- classified as

11 0 0 0 1 17 | a = ES

0 19 1 1 0 9 | b = FR

0 0 16 4 1 9 | c = GE

0 4 2 14 4 6 | d = IT

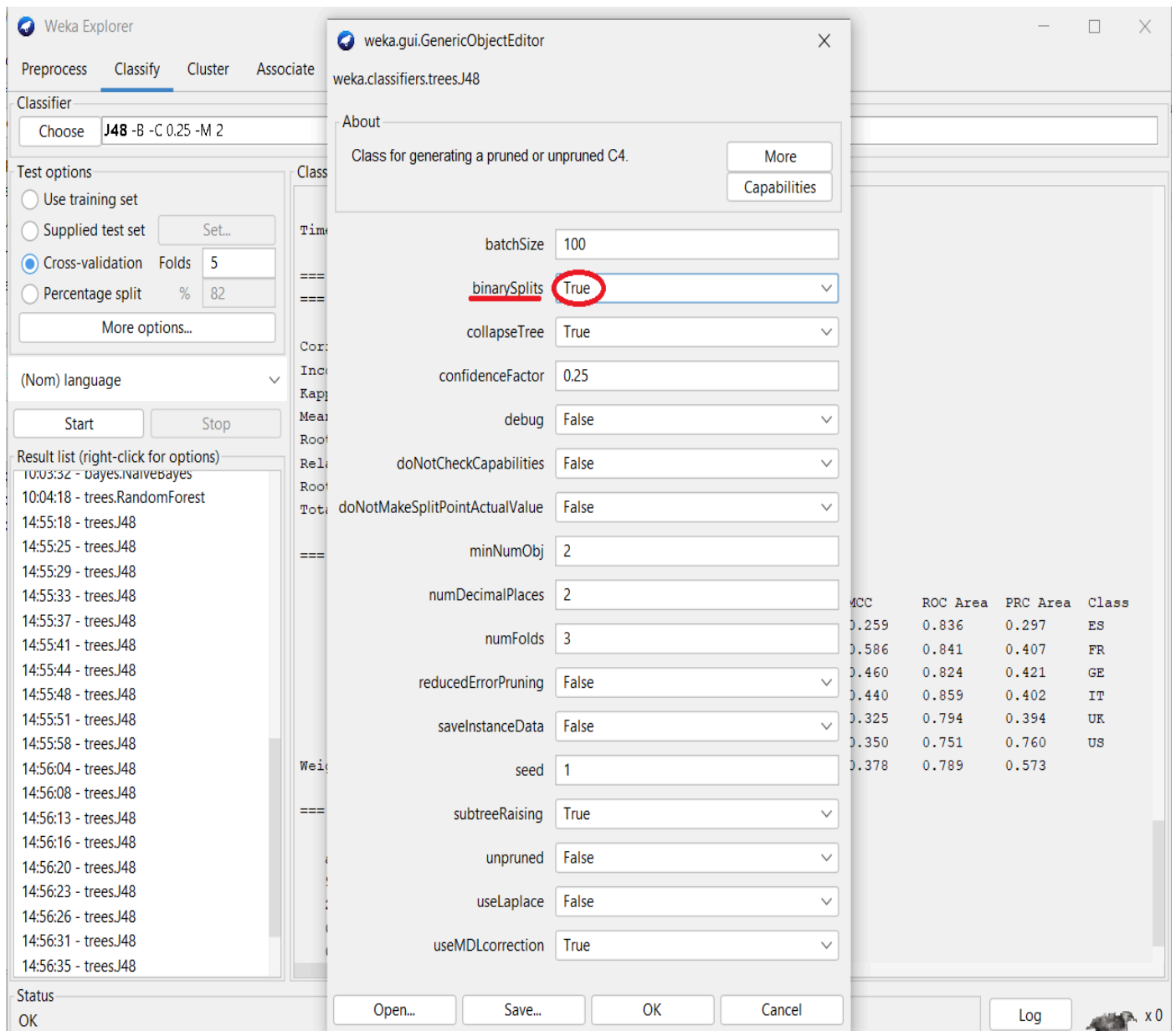
3 0 2 4 20 16 | e = UK

9 4 7 4 11 130 | f = US

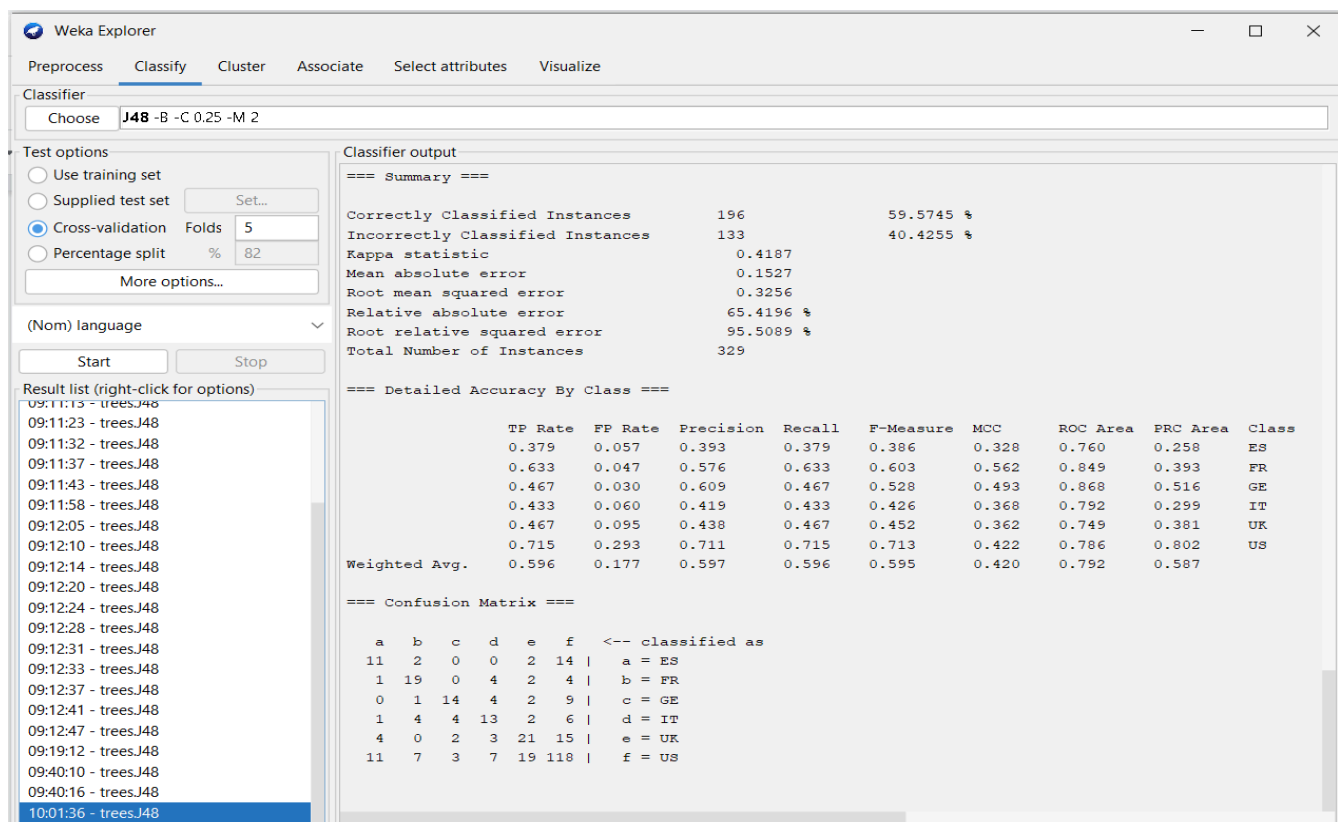
- 11 : it's ES and classified as ES
- 1 : it's ES and classified as UK
- 17 : it's ES and classified as US
- 19 : it's FR and classified as FR
- 1 : it's FR and classified as GE
- 1 : it's FR and classified as IT
- 9 : it's FR and classified as US

And so on,,,,

- Changing one hyper parameter



Figure[6]



Figure[7]

Correctly Classified Instances 196 59.5745 %

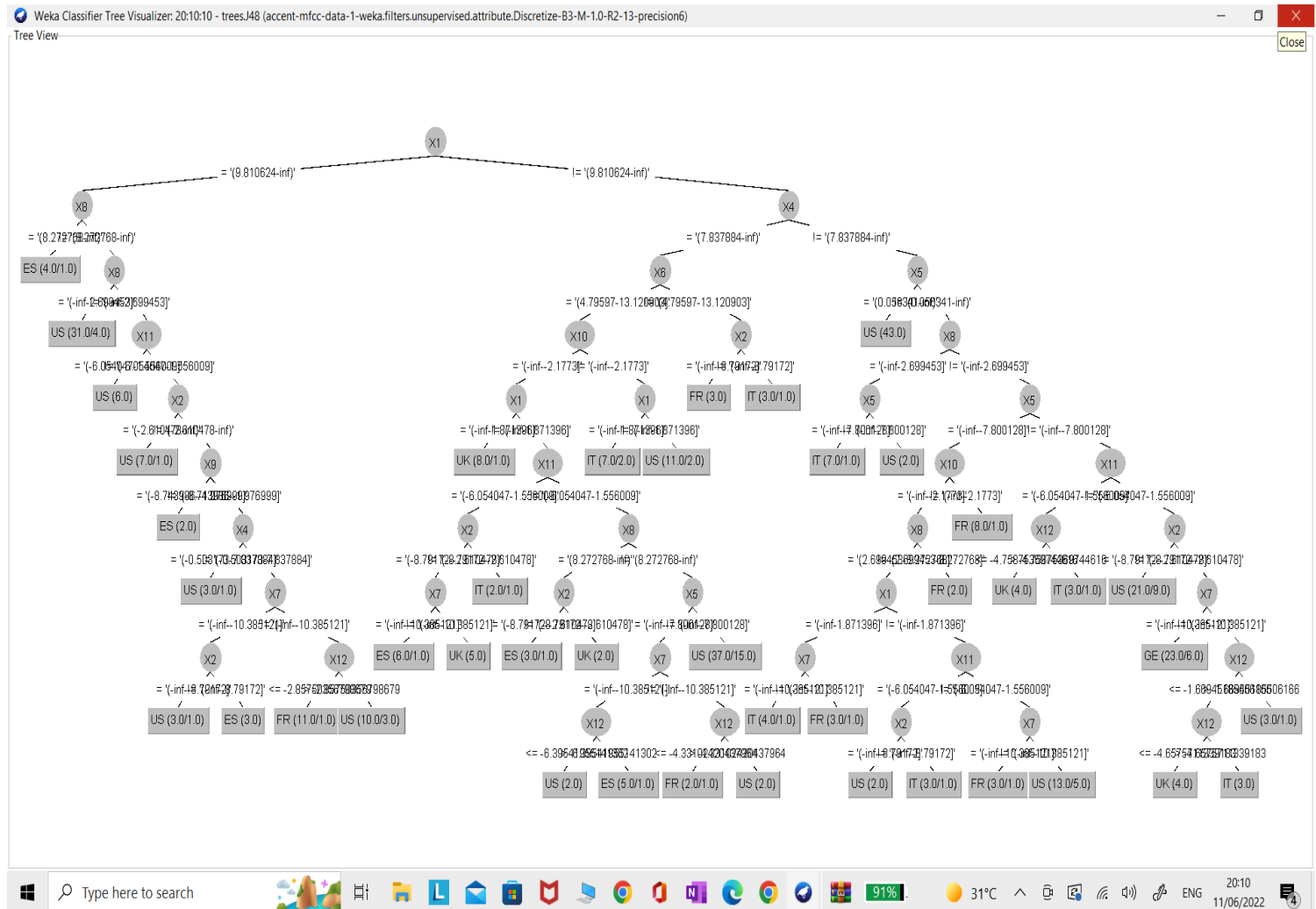
Incorrectly Classified Instances 133 40.4255 %

Note : The accuracy increased to 59.5745 % where it was 57.4468%.

	accuracy	Precision	Recall	F-measure
Decision Tree	57.4468%	0.571	0.574	0.573
Decision Tree after change one hyper parameter	59.5745 %	0.597	0.596	0.595

And the tree as shown in Figure[8].

Right Click on the classifier name in the Results list—>Visualize Tree



Conclusions

We noticed that Random Forest has the highest accuracy between all supervised learning methods, which was with accuracy = 63.8298%, then the Decision Tree was at the second position after Random Forest with accuracy = 57.4468% and finally was Naïve Bayes with accuracy = 48.3283%.

The F-Score for the was 0.573 for Decision Tree, 0.501 for Naïve Bayes, 0.630 for Random Forest . Again Random Forest is the best, followed by Decision Tree, Naïve Bayes.