Lab Report: Report On congressional voting record.

Course Code: CSE322

Course Title: Data Mining and Machine Learning Lab

Submitted to:

Nasima Islam Bithi

Lecturer,

Department of CSE, DIU.

Submitted by:

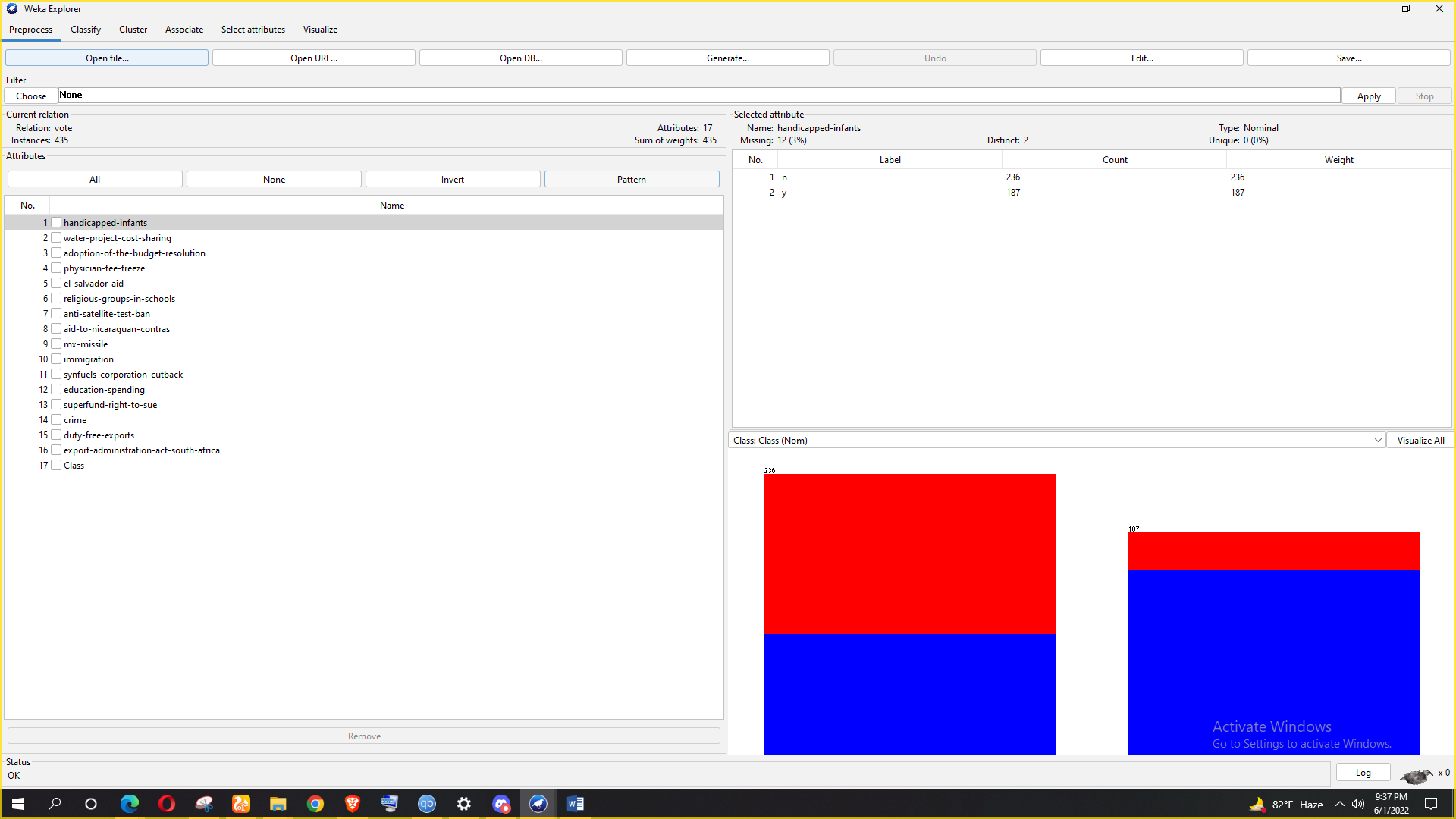
Abdur Rahman Alvi

ID: 193-15-3002

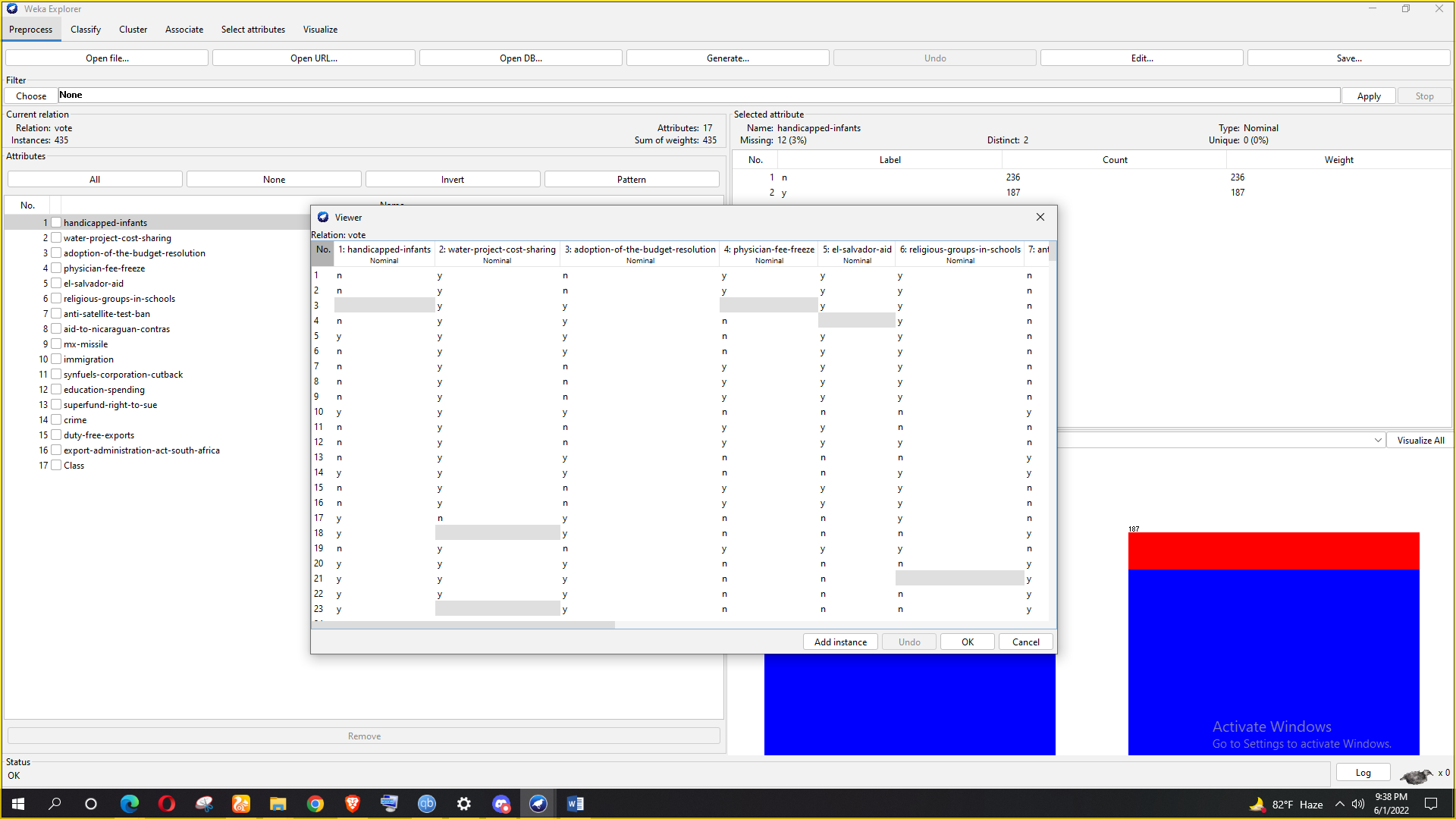
Section: PC-R

After downloading the dataset of the congressional voting record of 1984, I load the dataset(arff formate) on weka explorer

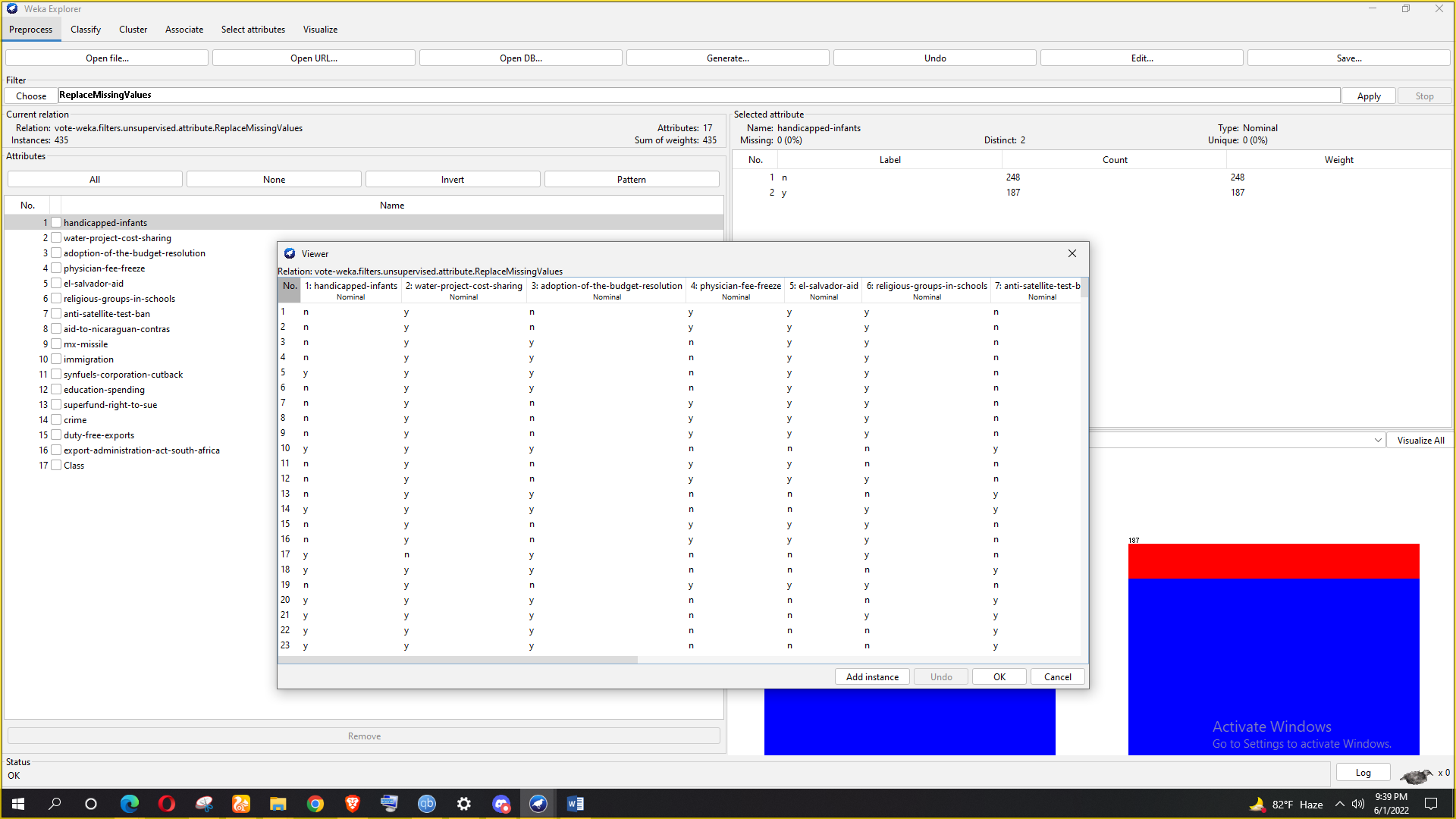
Load dataset:



**When it has missing value-**

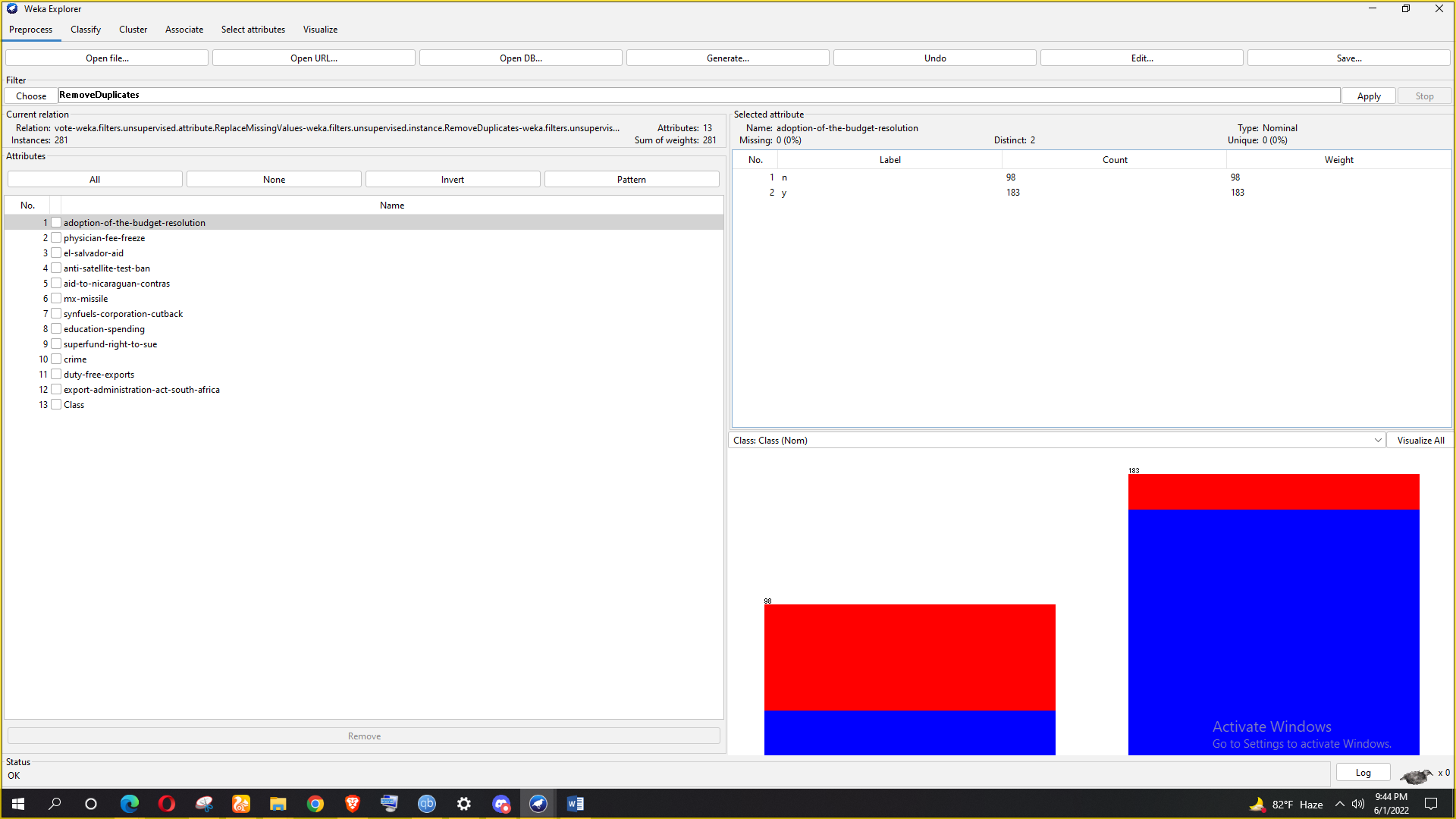


**After solving missing value-**

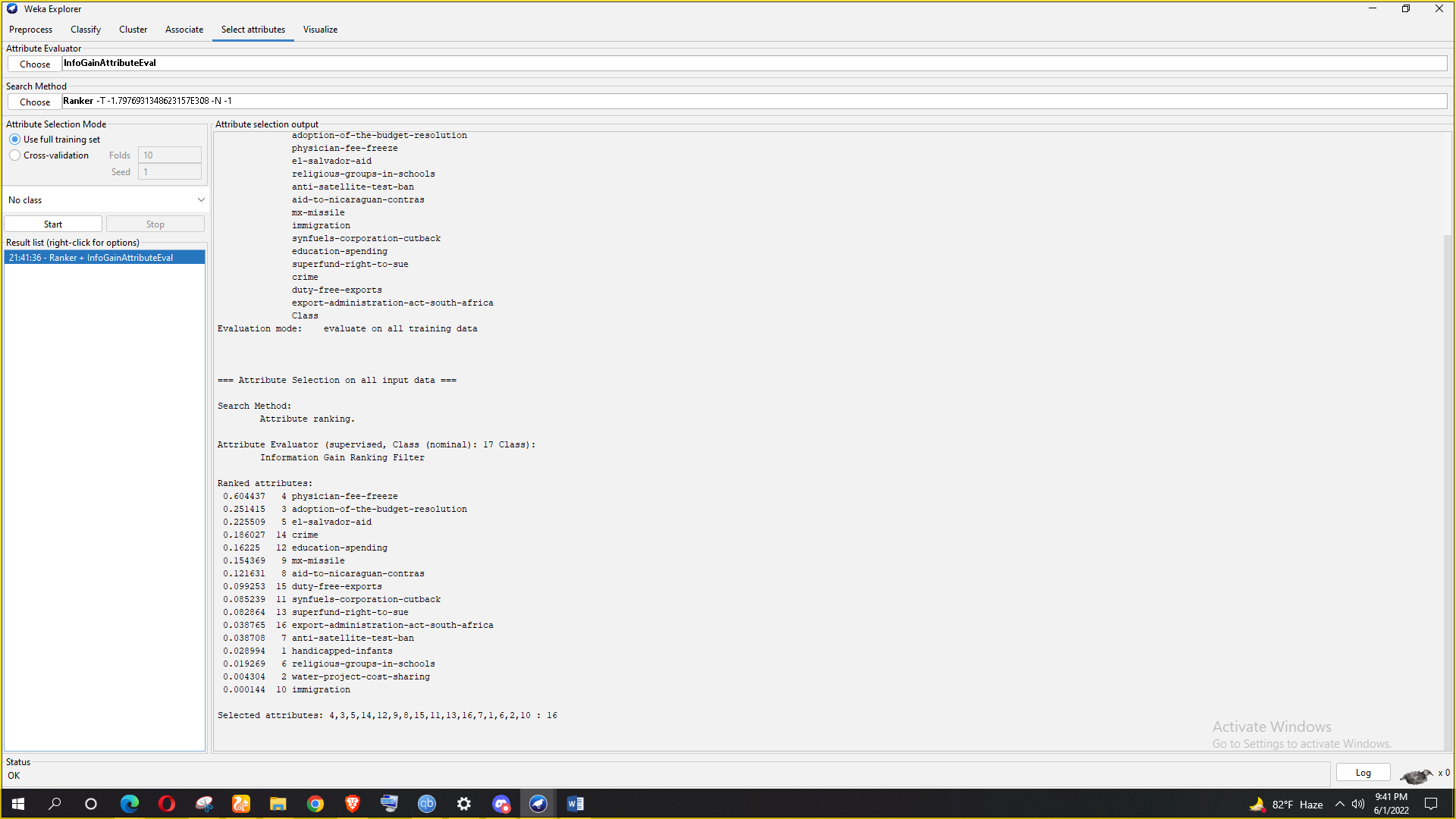


**Now we are going to remove duplicates by using RemoveDuplicates from filters>unsupervised>instances>RemoveDuplicates**

**After applying remove duplicates-**

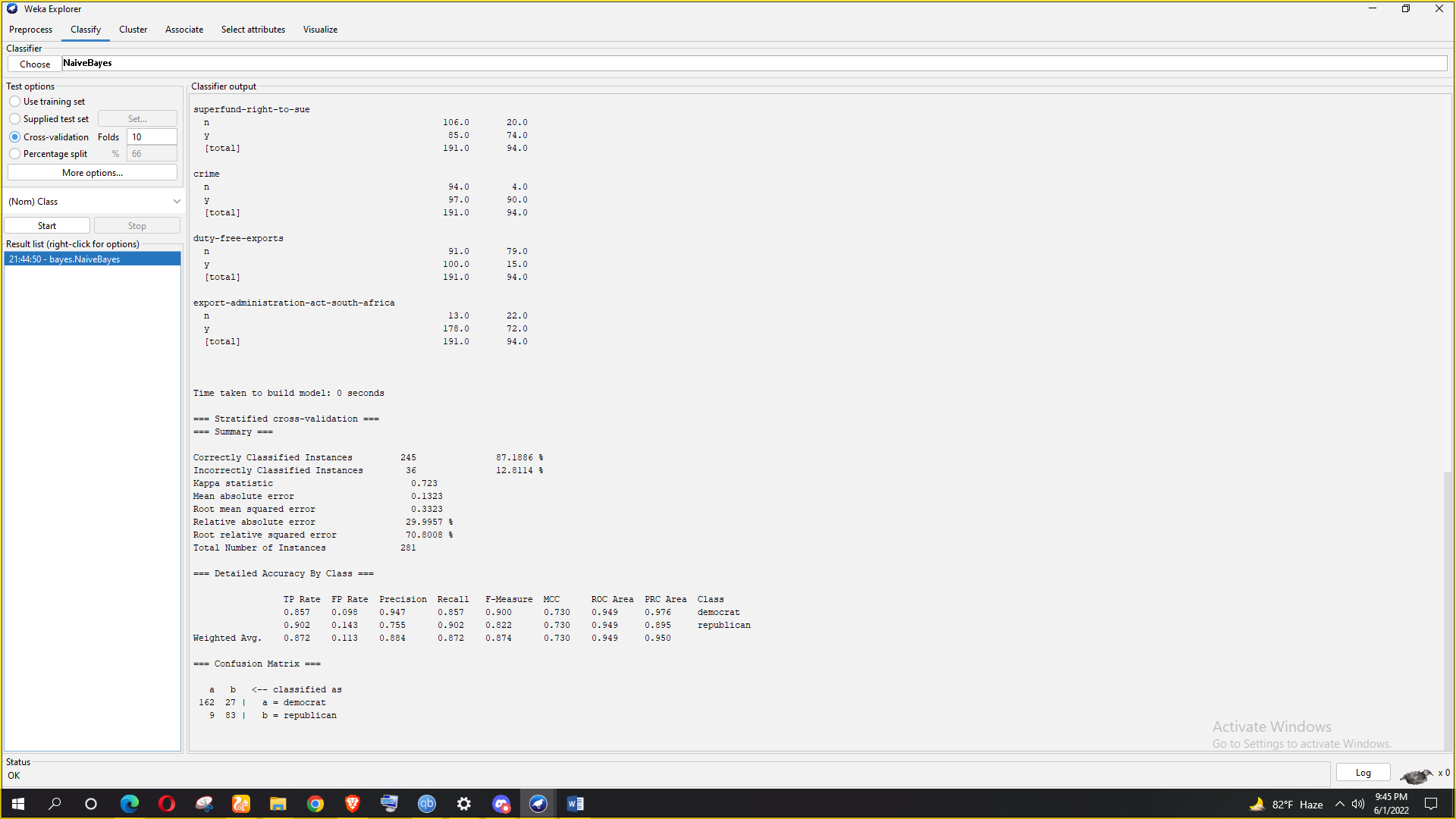


**After preprocessing the dataset. We will now select the 12 best attributes from the congressional voting record dataset and apply InfoGainAttributeEvaluator and Ranker as search methods.**



**After that I applied ByesNet, IBK and two tree algorithm like RandomForest, J48:**

**ByesNet:**

****

**Confusion matrix :**

|  |  |
| --- | --- |
| 162 | 27 |
| 9 | 83 |

Hence TP = 162

TN = 27

FP = 9

FN = 83

Accuracy = (162 + 83) / (162 + 27 + 9 + 83)

= 0.8718861209964412

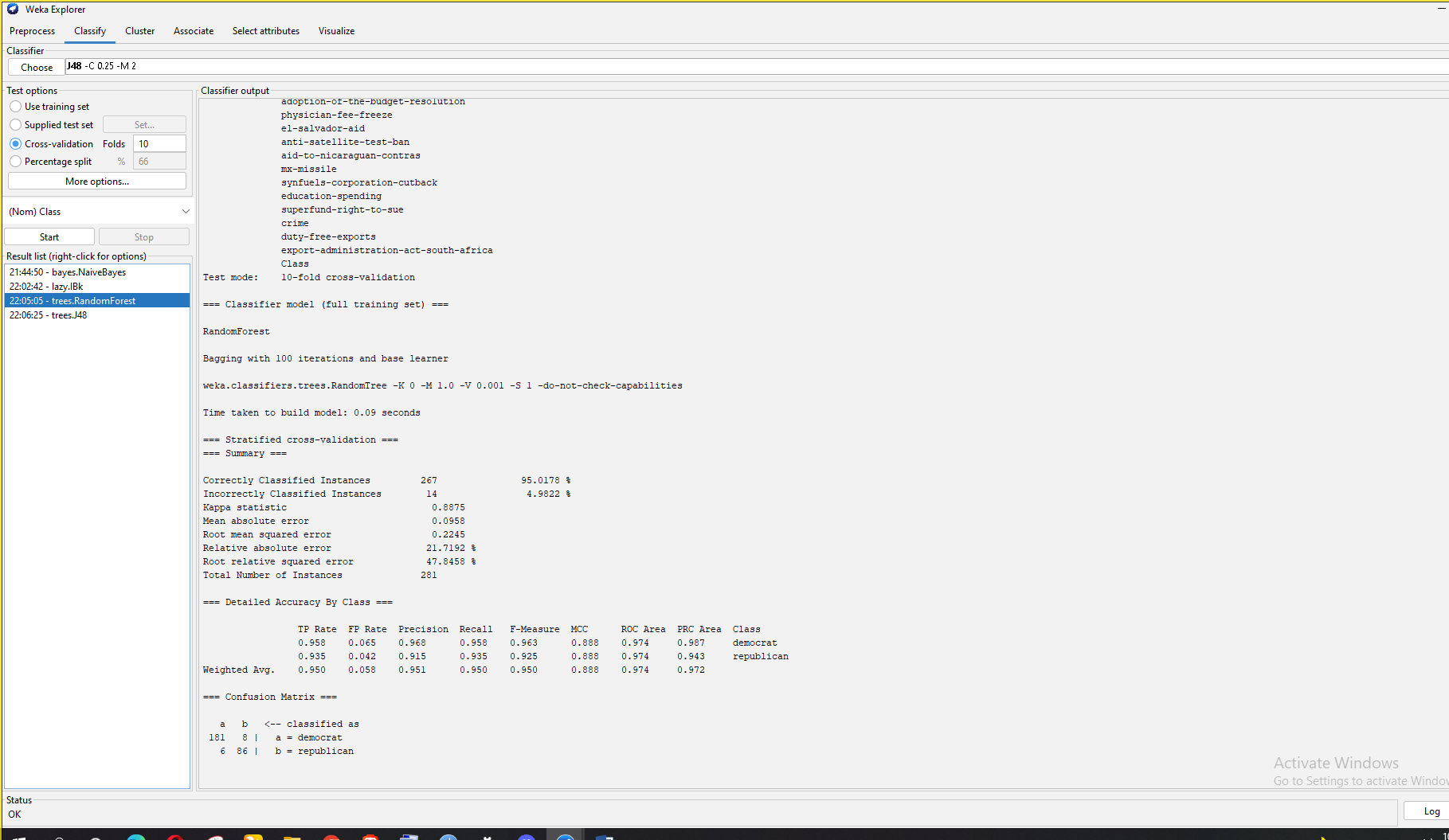
Precision = 162 / (162 + 9) = 0.9473684210526315

Recall = 162 / (162 + 27) = 0.8571428571428571

F1-Score = (2 \* Precision \* Recall) / (Precision + Recall)

=(2 \* 0.9473684210526315 \* 0.8571428571428571) /(0.9473684210526315 + 0.8571428571428571) = 0.963

**RandomForest:**

****

**Confusion matrix :**

|  |  |
| --- | --- |
| 181 | 8 |
| 6 | 86 |

Hence TP = 181

TN = 8

FP = 9

FN = 86

Accuracy = (181 + 86) / (181 + 8 + 6 + 86) = 0.9501779359430605

Precision = 181 / (181 + 6) = 0.9679144385026738

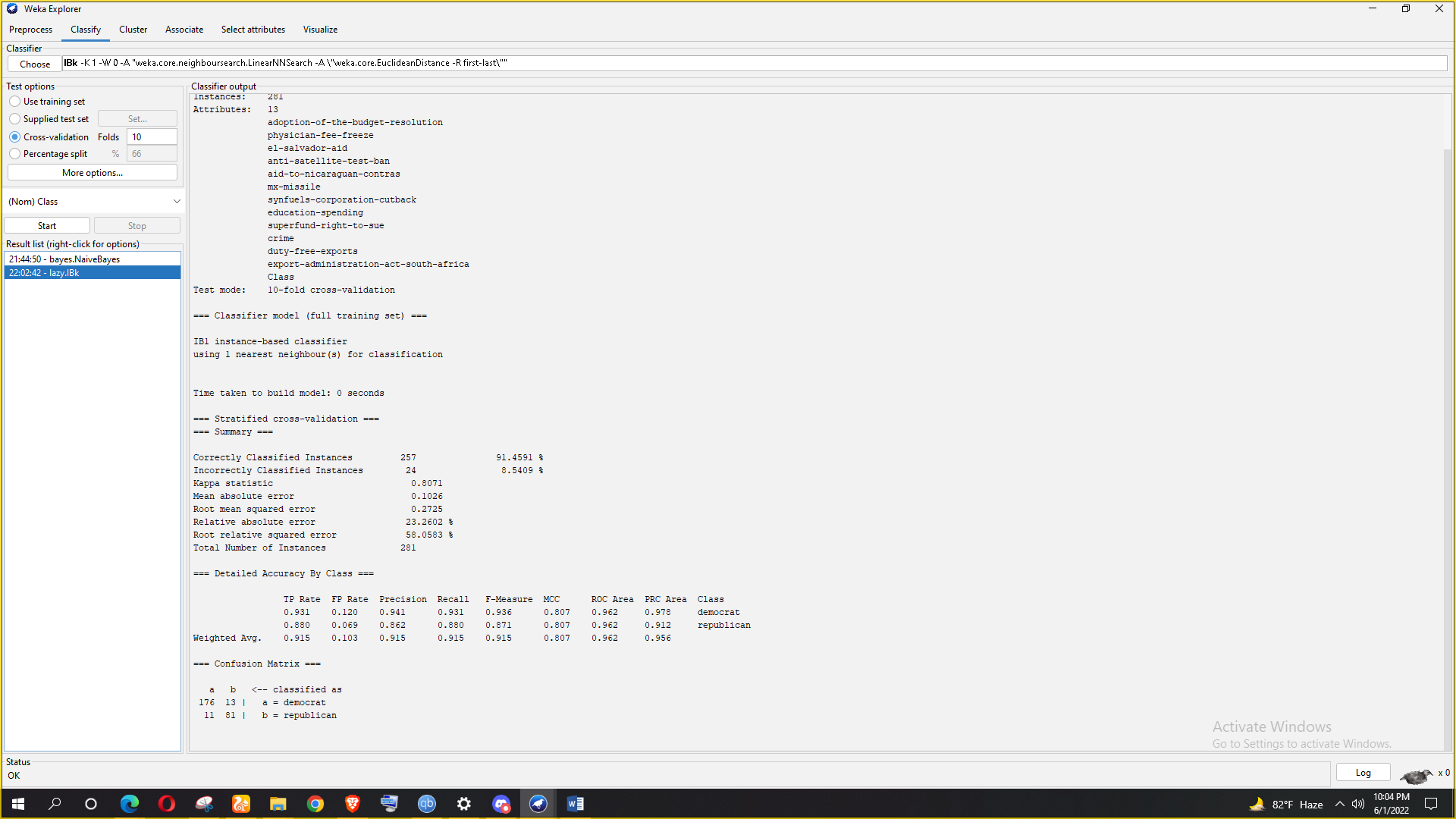
Recall = 181 / (181 + 8) = 0.9576719576719577

F1-Score = (2 \* Precision \* Recall) / (Precision + Recall)

= (2 \* 0.9679144385026738 \* 0.9576719576719577) / (0.9679144385026738 + 0.9576719576719577)

= 0.98

**IBk :**

****

**Confusion matrix :**

|  |  |
| --- | --- |
| 176 | 13 |
| 11 | 81 |

Hence TP = 176

TN = 13

FP = 11

FN = 81

Accuracy = (176 + 81) / (176 + 13 + 11 + 81) = 0.9145907473309609

Precision = 176 / (176 + 11) = 0.9411764705882353

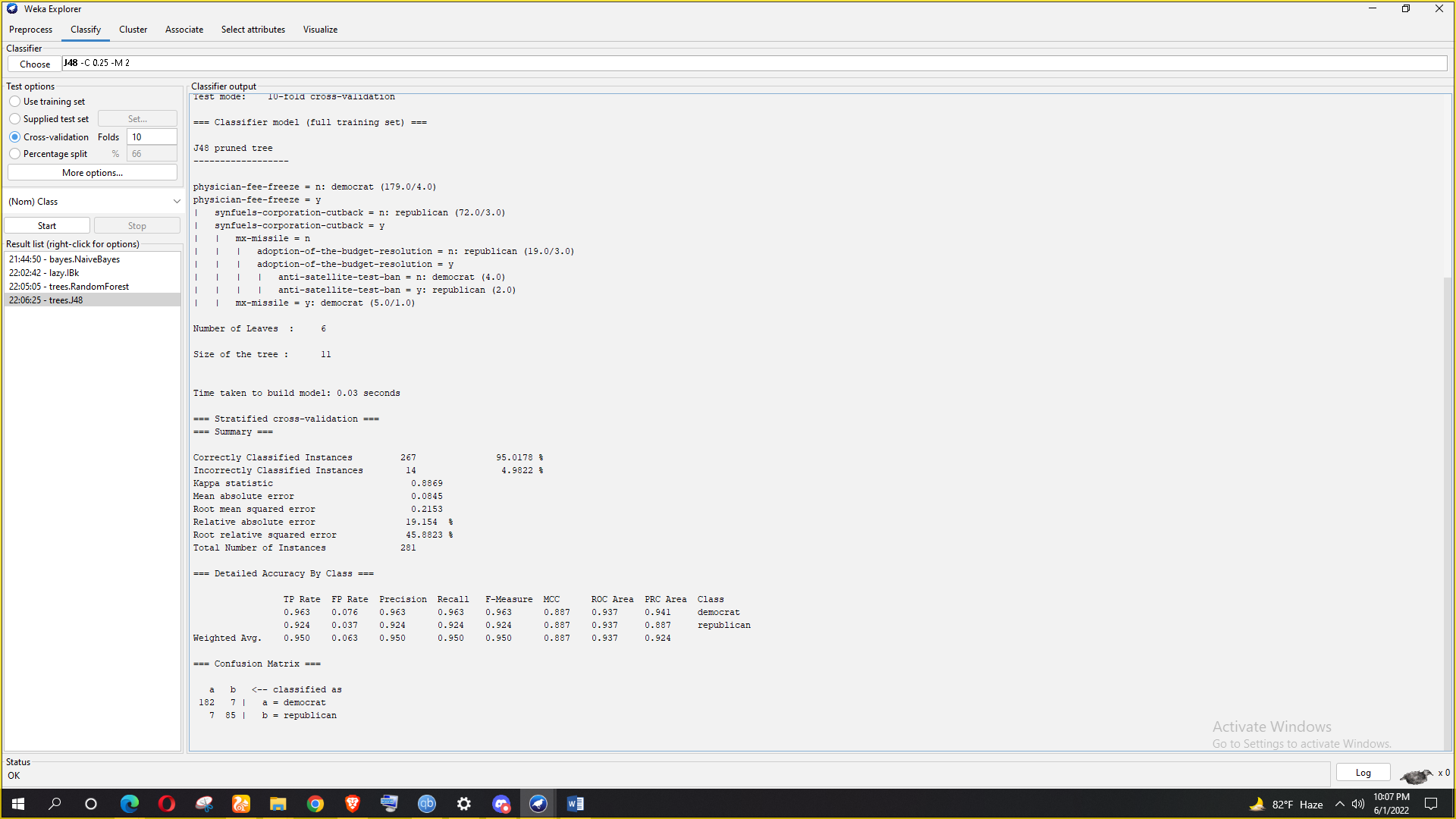
Recall = 176 / (176 + 13) = 0.9312169312169312

F1-Score = (2 \* Precision \* Recall) / (Precision + Recall)

= (2 \* 0.9411764705882353 \* 0.9312169312169312) / (0.9411764705882353 + 0.9312169312169312)

= 0.93

**J48:**

****

**Confusion matrix :**

|  |  |
| --- | --- |
| 182 | 7 |
| 7 | 85 |

Hence TP = 182

TN = 7

FP = 7

FN = 85

Accuracy = (182 + 85) / (182 + 7 + 7 + 85) = 0.9501779359430605

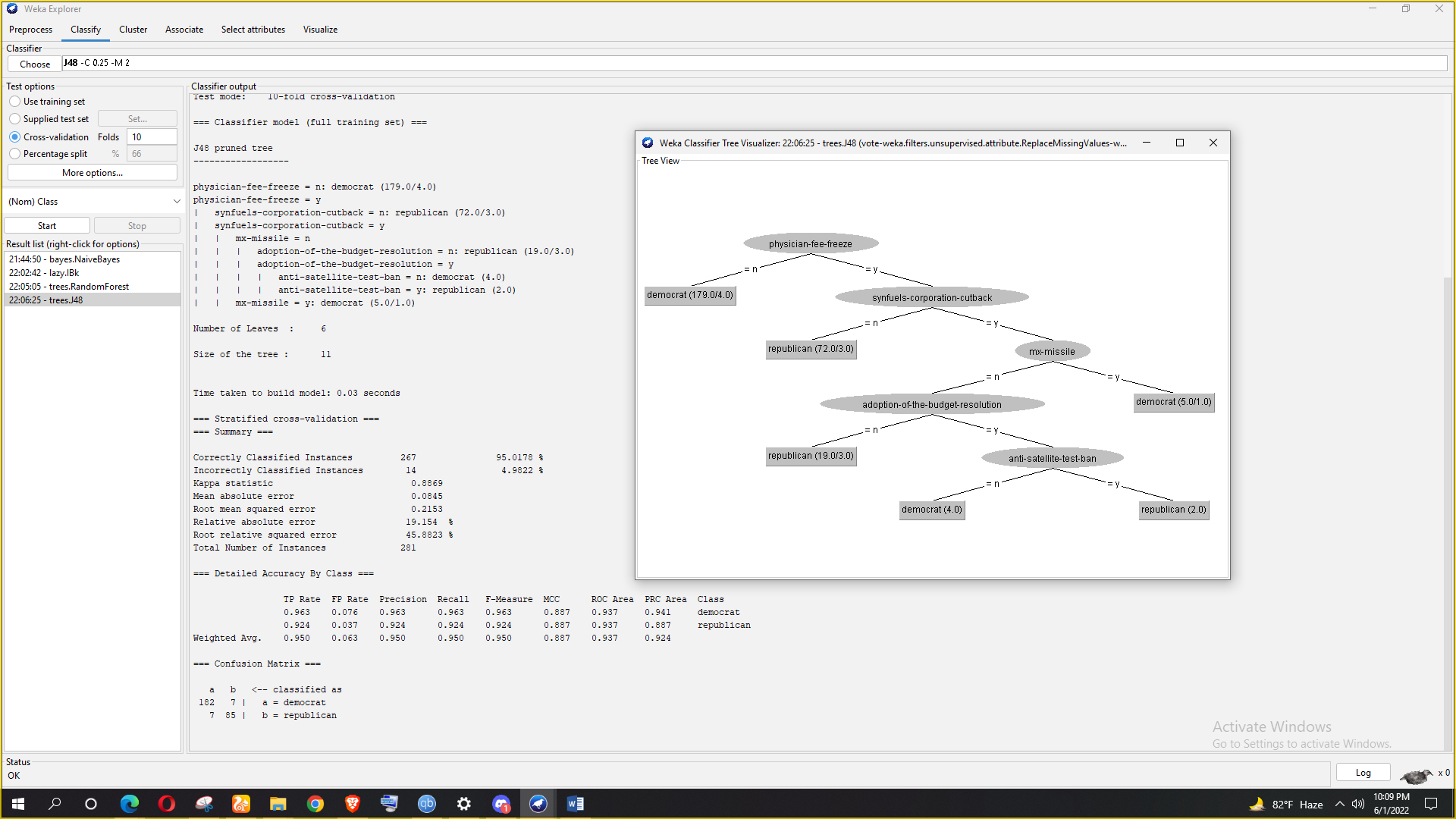
Precision = 182 / (182 + 7) = 0.9629629629629629

Recall = 182 / (182 + 7) = 0.9629629629629629

F1-Score = (2 \* Precision \* Recall) / (Precision + Recall)

= (2 \* 0.9629629629629629 \* 0.9629629629629629) / (0.9629629629629629 + 0.9629629629629629) = 0.963

**From the calculation of previous accuracy of algorithm, If I sorting as best, second best and third best then it will be RandomForest, J48 and IBk. Because the accuracy percentage are : 0.98, 0.963,0.98** .

****

**If I analysis from the decision tree generated by J48 algorithm, three characteristic of democrat voter are physician-fee-freez, mx-missile and anti-satellite-test-ban.**

The association rules that I have used are given below:

1.physician-fee-freeze=n 179 ==> Class=democrat 175.

2.physician-fee-freeze=n export-administration-act-south-africa=y 173 ==> Class=democrat 168.

3.adoption-of-the-budget-resolution=y 183 ==> export-administration-act-south-africa=y 178

4.physician-fee-freeze=n 179 ==> export-administration-act-south-africa=y Class=democrat 16