Taimoor khan

Id: `13579

Data Mining

Sir: Zain shukat

Classification	Regression
Classification is the discovery of model or functions where the mapping of objects is done into predefined classes.	Regression is a devised model in which the mapping of objects is done into values.
It involves prediction of discrete values	It involves prediction of continuous values
Nature of the predicted data is unordered	Nature of the predicted data is ordered
Algorithms used are Decision tree, logistic regression, etc.	Algorithms used are Regression tree (Random forest), Linear regression, etc.
Method of calculation is measuring accuracy	Method of calculation is Measurement of root mean square error

Classification Scenario Example:

Suppose from your past data (train data) you come to know that your best friend likes the above movies. Now one new movie (test data) released. Hopefully, you want to know your best friend like it or not. If you strongly confirmed about the chances of your friend like the move. You can take your friend to a movie this weekend.

If you clearly observe the problem it is just whether your friend like or not. Finding a solution to this type of problem is called as classification. This is because we are classifying the things to their belongings (yes or no, like or dislike). Keep in mind here we are forecasting target class (classification) and the other thing this classification belongs to supervised learning. This is because you are learning this from your train data.

In this case, the problem is a binary classification in which we have to predict whether output belongs to class 1 or class 2 (class 1 : yes, class 2: no). As we have discussed earlier we can use classification for predicting more classes too. Like (colour prediction: red,green,blue,yellow,orange).

Regression Scenario Example:

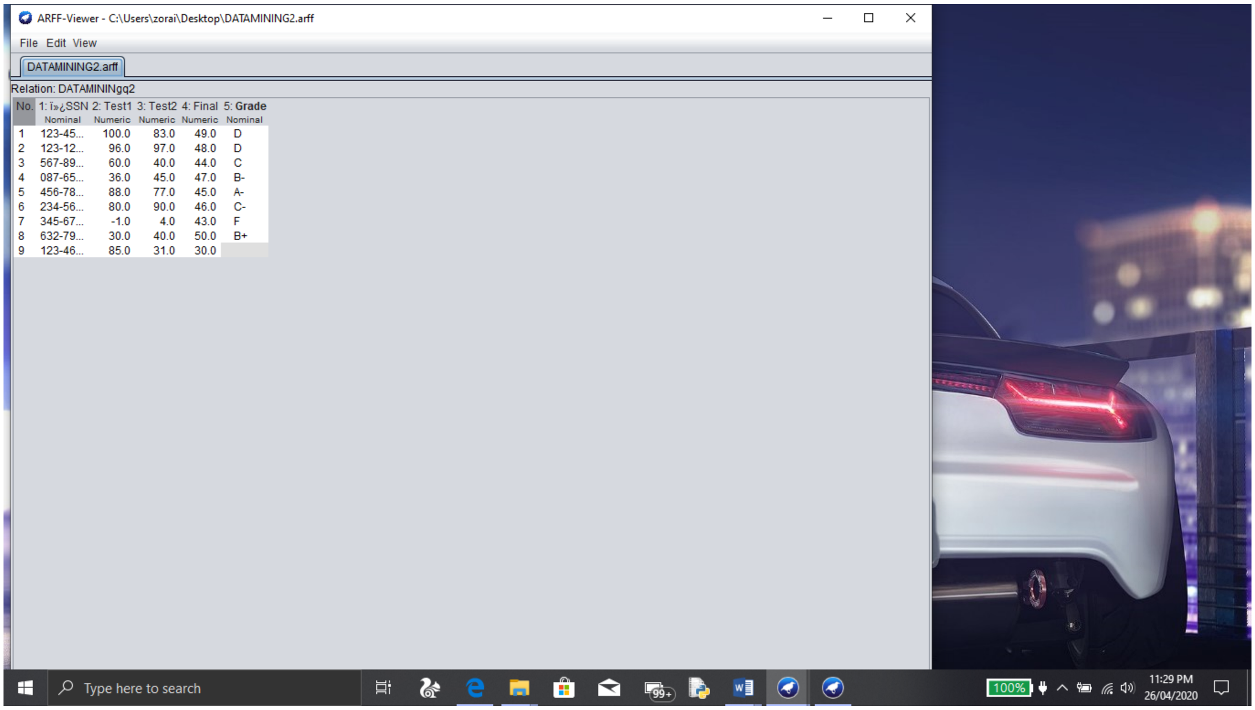
Suppose from your past data (train data) you come to know that your best friend likes the above movies. You also know how many times each particular movie seen by your friend. Now one new movie (test data) released. Now you are going to find how many times this newly released movie will your friend watch. It could be 5 times, 6 times,10 times etc

Name M Zoraiz Ali
Id 14413
Data mining
Sir Zain Shaukat

If you clearly observe the problem is about finding the count, sometimes we can say this as predicting the value. Keep in mind, here we are forecasting a value (prediction) and the other thing this prediction also belongs to supervised learning. This is because you are learning this from you train data.

Q2

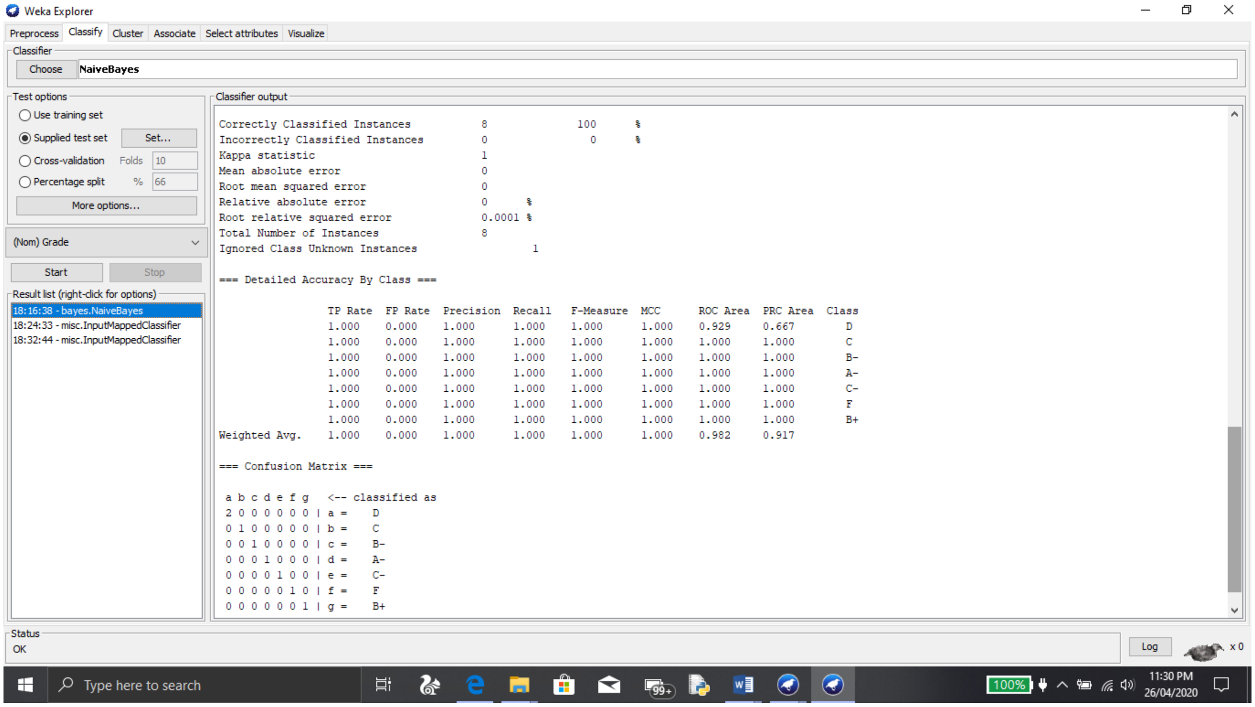
Ans:



Relation: DATAMININgq2

No.	1: ï»¿SSN Nominal	2: Test1 Numeric	3: Test2 Numeric	4: Final Numeric	5: Grade Nominal
1	123-45...	100.0	83.0	49.0	D
2	123-12...	96.0	97.0	48.0	D
3	567-89...	60.0	40.0	44.0	C
4	087-65...	36.0	45.0	47.0	B-
5	456-78...	88.0	77.0	45.0	A-
6	234-56...	80.0	90.0	46.0	C-
7	345-67...	-1.0	4.0	43.0	F
8	632-79...	30.0	40.0	50.0	B+
9	123-46...	85.0	31.0	30.0	

Name M Zoraiz Ali
Id 14413
Data mining
Sir Zain Shaukat



Classifier: NaiveBayes

Test options
○ Use training set
● Supplied test set
○ Cross-validation Folds 10
○ Percentage split % 66

(Nom) Grade

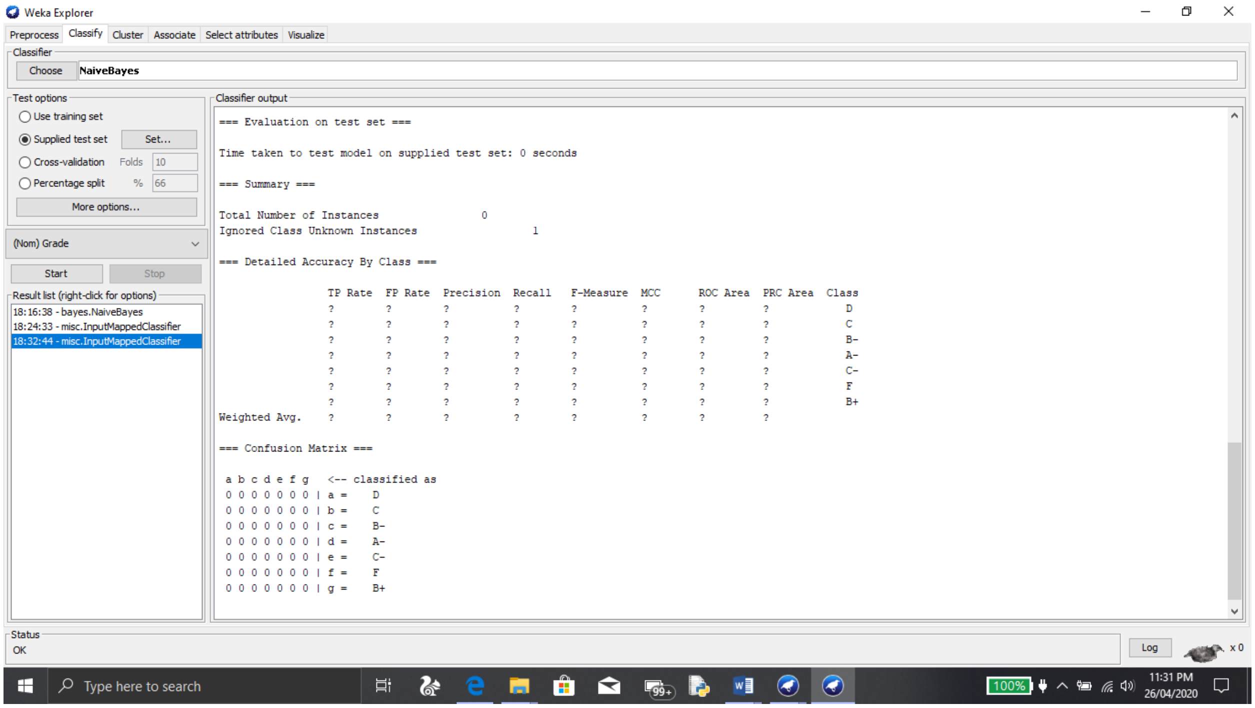
Result list (right-click for options)
18:16:38 - bayes.NaiveBayes
18:24:33 - misc.InputMappedClassifier
18:32:44 - misc.InputMappedClassifier

Classifier output

```
Correctly Classified Instances           8              100      %
Incorrectly Classified Instances         0                0      %
Kappa statistic                          1
Mean absolute error                      0
Root mean squared error                  0
Relative absolute error                  0      %
Root relative squared error              0.0001 %
Total Number of Instances                8
Ignored Class Unknown Instances                  1

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.000    1.000      1.000    1.000      1.000    0.929     0.667     D
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     C
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     B-
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     A-
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     C-
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     F
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     B+
Weighted Avg.    1.000    0.000    1.000      1.000    1.000      1.000    0.982     0.917

=== Confusion Matrix ===

 a b c d e f g   <-- classified as
 2 0 0 0 0 0 0 | a =    D
 0 1 0 0 0 0 0 | b =    C
 0 0 1 0 0 0 0 | c =    B-
 0 0 0 1 0 0 0 | d =    A-
 0 0 0 0 1 0 0 | e =    C-
 0 0 0 0 0 1 0 | f =    F
 0 0 0 0 0 0 1 | g =    B+
```

Status
OK

Name M Zoraiz Ali
Id 14413
Data mining
Sir Zain Shaukat



Classifier: NaiveBayes

Test options
○ Use training set
● Supplied test set
○ Cross-validation Folds 10
○ Percentage split % 66

(Nom) Grade

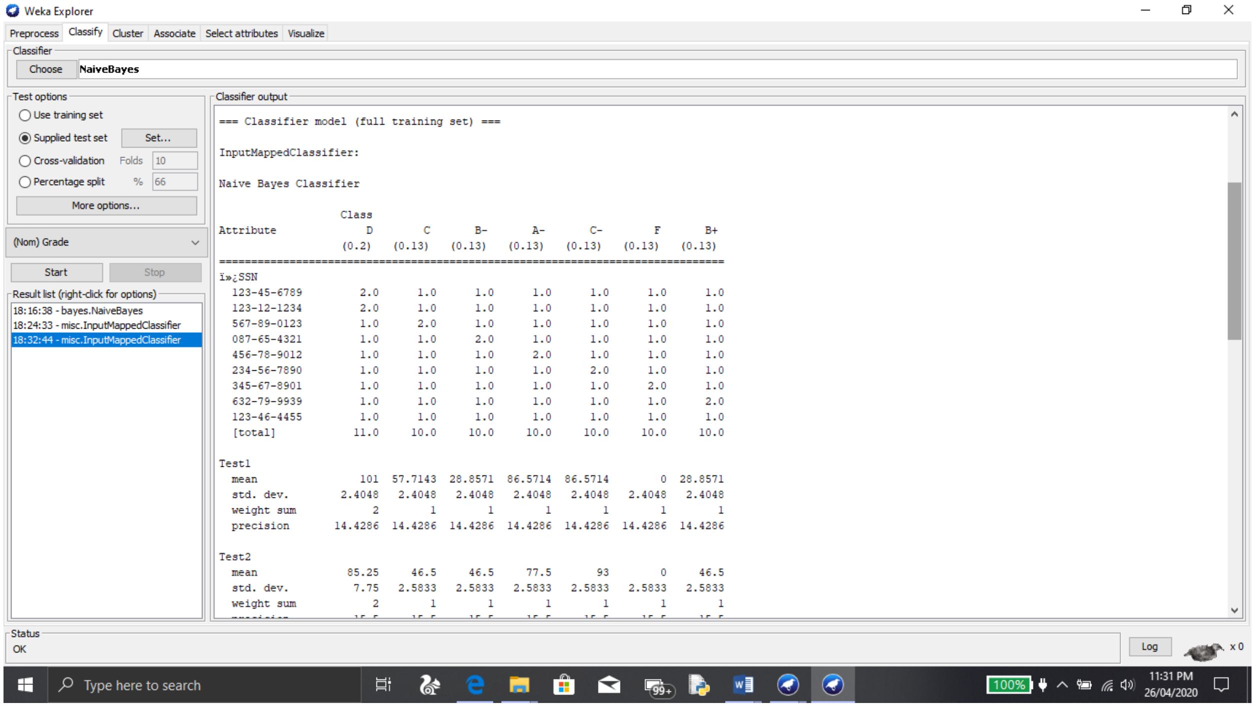
Result list (right-click for options)
18:16:38 - bayes.NaiveBayes
18:24:33 - misc.InputMappedClassifier
18:32:44 - misc.InputMappedClassifier

Classifier output

```
=== Evaluation on test set ===

Time taken to test model on supplied test set: 0 seconds

=== Summary ===

Total Number of Instances                0     
Ignored Class Unknown Instances                  1     

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 ?        ?        ?          ?        ?          ?        ?         ?          D
                 ?        ?        ?          ?        ?          ?        ?         ?          C
                 ?        ?        ?          ?        ?          ?        ?         ?          B-
                 ?        ?        ?          ?        ?          ?        ?         ?          A-
                 ?        ?        ?          ?        ?          ?        ?         ?          C-
                 ?        ?        ?          ?        ?          ?        ?         ?          F
                 ?        ?        ?          ?        ?          ?        ?         ?          B+
Weighted Avg.    ?        ?        ?          ?        ?          ?        ?         ?         

=== Confusion Matrix ===

 a b c d e f g   <-- classified as
 0 0 0 0 0 0 0 | a =    D
 0 0 0 0 0 0 0 | b =    C
 0 0 0 0 0 0 0 | c =    B-
 0 0 0 0 0 0 0 | d =    A-
 0 0 0 0 0 0 0 | e =    C-
 0 0 0 0 0 0 0 | f =    F
 0 0 0 0 0 0 0 | g =    B+
```

Status
OK

Name M Zoraiz Ali
Id 14413
Data mining
Sir Zain Shaukat



Classifier
NaiveBayes

Test options
○ Use training set
● Supplied test set
○ Cross-validation Folds 10
○ Percentage split % 66

(Nom) Grade

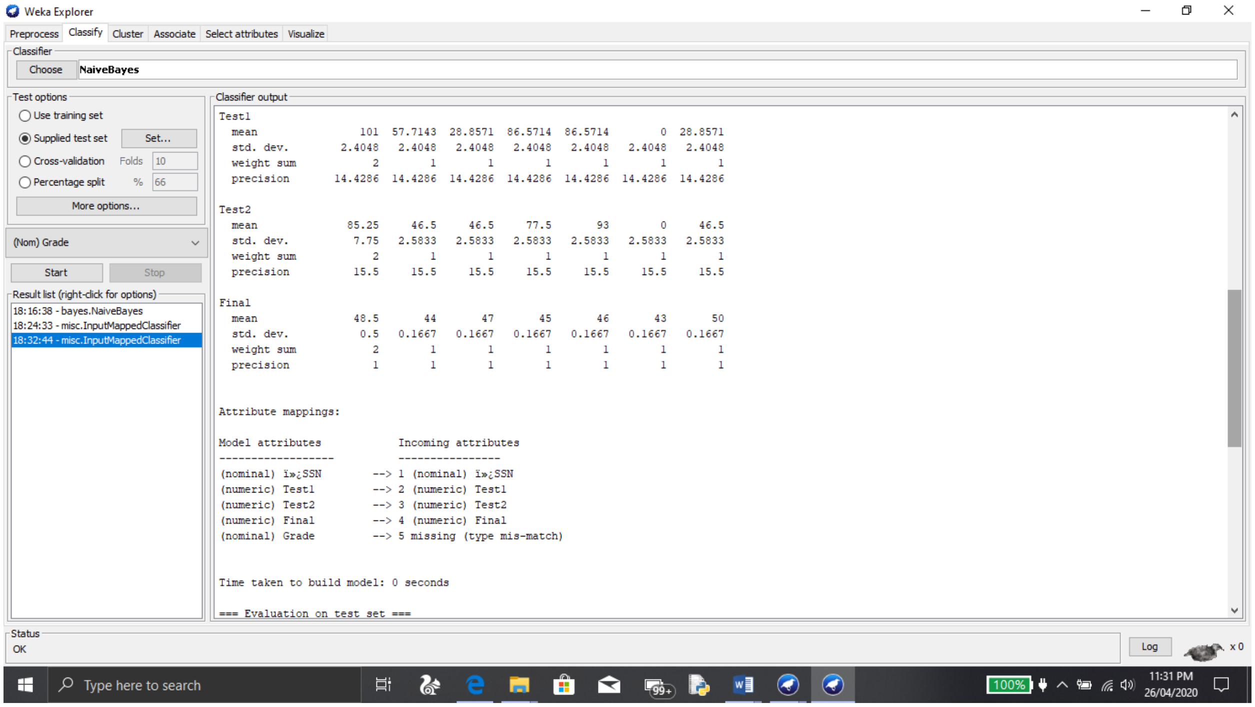
Result list (right-click for options)
18:16:38 - bayes.NaiveBayes
18:24:33 - misc.InputMappedClassifier
18:32:44 - misc.InputMappedClassifier

Classifier output

```
=== Classifier model (full training set) ===

InputMappedClassifier:

Naive Bayes Classifier

                   Class
Attribute              D        C       B-       A-       C-        F       B+
                   (0.2)   (0.13)   (0.13)   (0.13)   (0.13)   (0.13)   (0.13)
===============================================================================
ï»¿SSN
  123-45-6789        2.0      1.0      1.0      1.0      1.0      1.0      1.0
  123-12-1234        2.0      1.0      1.0      1.0      1.0      1.0      1.0
  567-89-0123        1.0      2.0      1.0      1.0      1.0      1.0      1.0
  087-65-4321        1.0      1.0      2.0      1.0      1.0      1.0      1.0
  456-78-9012        1.0      1.0      1.0      2.0      1.0      1.0      1.0
  234-56-7890        1.0      1.0      1.0      1.0      2.0      1.0      1.0
  345-67-8901        1.0      1.0      1.0      1.0      1.0      2.0      1.0
  632-79-9939        1.0      1.0      1.0      1.0      1.0      1.0      2.0
  123-46-4455        1.0      1.0      1.0      1.0      1.0      1.0      1.0
  [total]           11.0     10.0     10.0     10.0     10.0     10.0     10.0

Test1
  mean               101  57.7143  28.8571  86.5714  86.5714        0  28.8571
  std. dev.       2.4048   2.4048   2.4048   2.4048   2.4048   2.4048   2.4048
  weight sum           2        1        1        1        1        1        1
  precision      14.4286  14.4286  14.4286  14.4286  14.4286  14.4286  14.4286

Test2
  mean             85.25     46.5     46.5     77.5       93        0     46.5
  std. dev.         7.75   2.5833   2.5833   2.5833   2.5833   2.5833   2.5833
  weight sum           2        1        1        1        1        1        1
```

Status
OK

Name M Zoraiz Ali
Id 14413
Data mining
Sir Zain Shaukat



Classifier
Choose NaiveBayes

Test options
○ Use training set
● Supplied test set Set...
○ Cross-validation Folds 10
○ Percentage split % 66
More options...

(Nom) Grade

Result list (right-click for options)
18:16:38 - bayes.NaiveBayes
18:24:33 - misc.InputMappedClassifier
18:32:44 - misc.InputMappedClassifier

Classifier output

```
Test1
  mean              101  57.7143  28.8571  86.5714  86.5714        0  28.8571
  std. dev.      2.4048   2.4048   2.4048   2.4048   2.4048   2.4048   2.4048
  weight sum          2        1        1        1        1        1        1
  precision     14.4286  14.4286  14.4286  14.4286  14.4286  14.4286  14.4286

Test2
  mean            85.25     46.5     46.5     77.5       93        0     46.5
  std. dev.        7.75   2.5833   2.5833   2.5833   2.5833   2.5833   2.5833
  weight sum          2        1        1        1        1        1        1
  precision        15.5     15.5     15.5     15.5     15.5     15.5     15.5

Final
  mean             48.5       44       47       45       46       43       50
  std. dev.         0.5   0.1667   0.1667   0.1667   0.1667   0.1667   0.1667
  weight sum          2        1        1        1        1        1        1
  precision           1        1        1        1        1        1        1


Attribute mappings:

Model attributes           Incoming attributes
------------------         ----------------
(nominal) ï»¿SSN        --> 1 (nominal) ï»¿SSN
(numeric) Test1        --> 2 (numeric) Test1
(numeric) Test2        --> 3 (numeric) Test2
(numeric) Final        --> 4 (numeric) Final
(nominal) Grade        --> 5 missing (type mis-match)


Time taken to build model: 0 seconds

=== Evaluation on test set ===
```

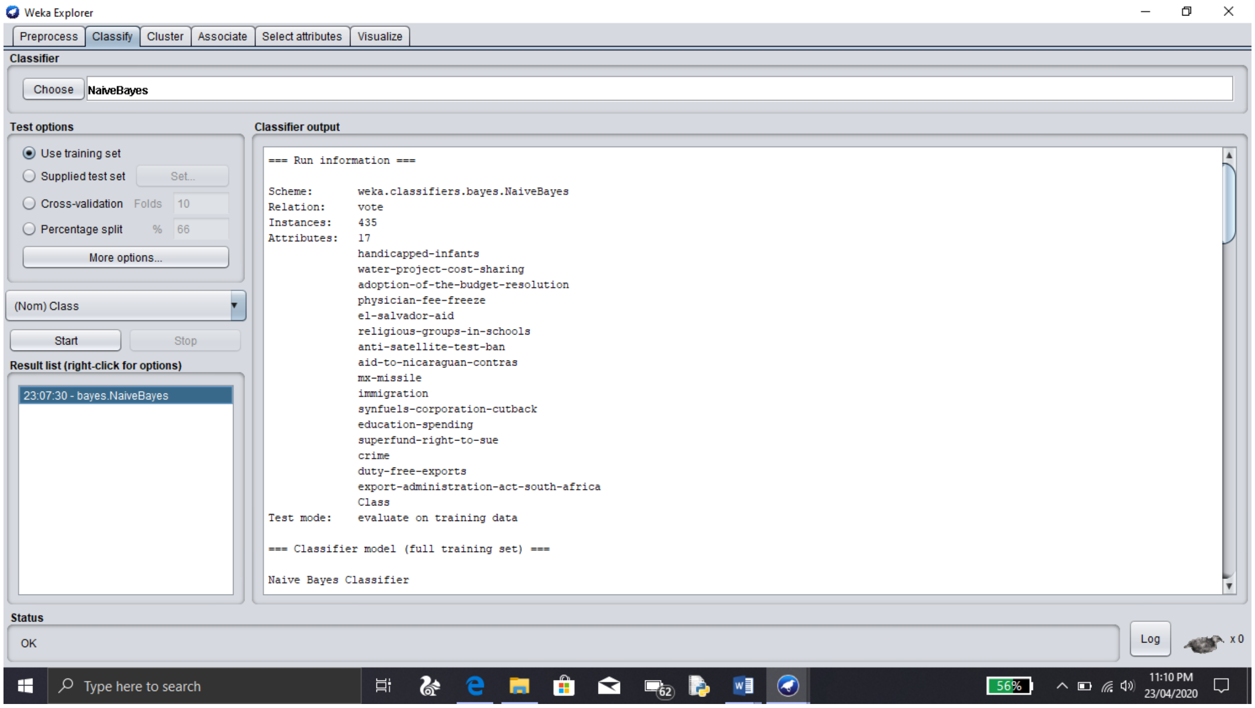
Status
OK

Type here to search
100%
11:31 PM
26/04/2020

Q3

Ans:

Name M Zoraiz Ali
Id 14413
Data mining
Sir Zain Shaukat



Classifier: NaiveBayes

Test options:
- ● Use training set
- ○ Supplied test set Set...
- ○ Cross-validation Folds 10
- ○ Percentage split % 66

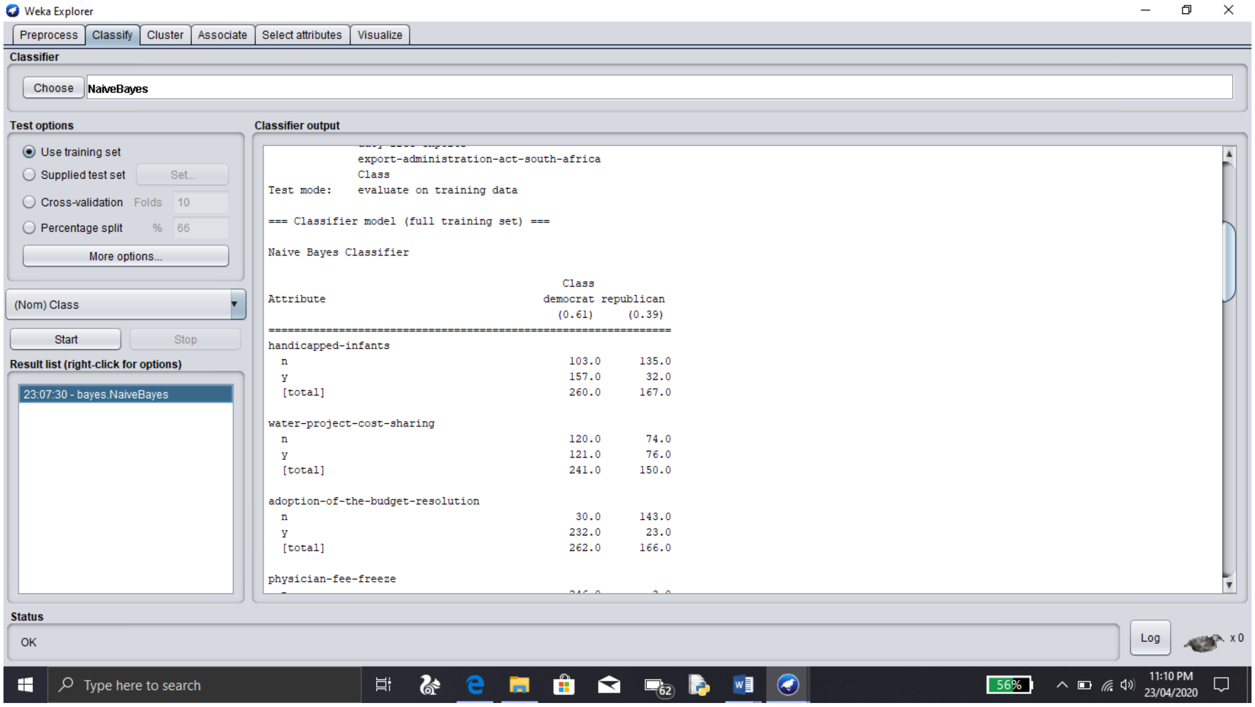
(Nom) Class

Result list: 23:07:30 - bayes.NaiveBayes

Status: OK

```
=== Run information ===

Scheme:       weka.classifiers.bayes.NaiveBayes
Relation:     vote
Instances:    435
Attributes:   17
              handicapped-infants
              water-project-cost-sharing
              adoption-of-the-budget-resolution
              physician-fee-freeze
              el-salvador-aid
              religious-groups-in-schools
              anti-satellite-test-ban
              aid-to-nicaraguan-contras
              mx-missile
              immigration
              synfuels-corporation-cutback
              education-spending
              superfund-right-to-sue
              crime
              duty-free-exports
              export-administration-act-south-africa
              Class
Test mode:    evaluate on training data

=== Classifier model (full training set) ===

Naive Bayes Classifier
```

Name M Zoraiz Ali
Id 14413
Data mining
Sir Zain Shaukat



Classifier: NaiveBayes

Test options
● Use training set
○ Supplied test set
○ Cross-validation Folds 10
○ Percentage split % 66

(Nom) Class

Result list (right-click for options)
23:07:30 - bayes.NaiveBayes

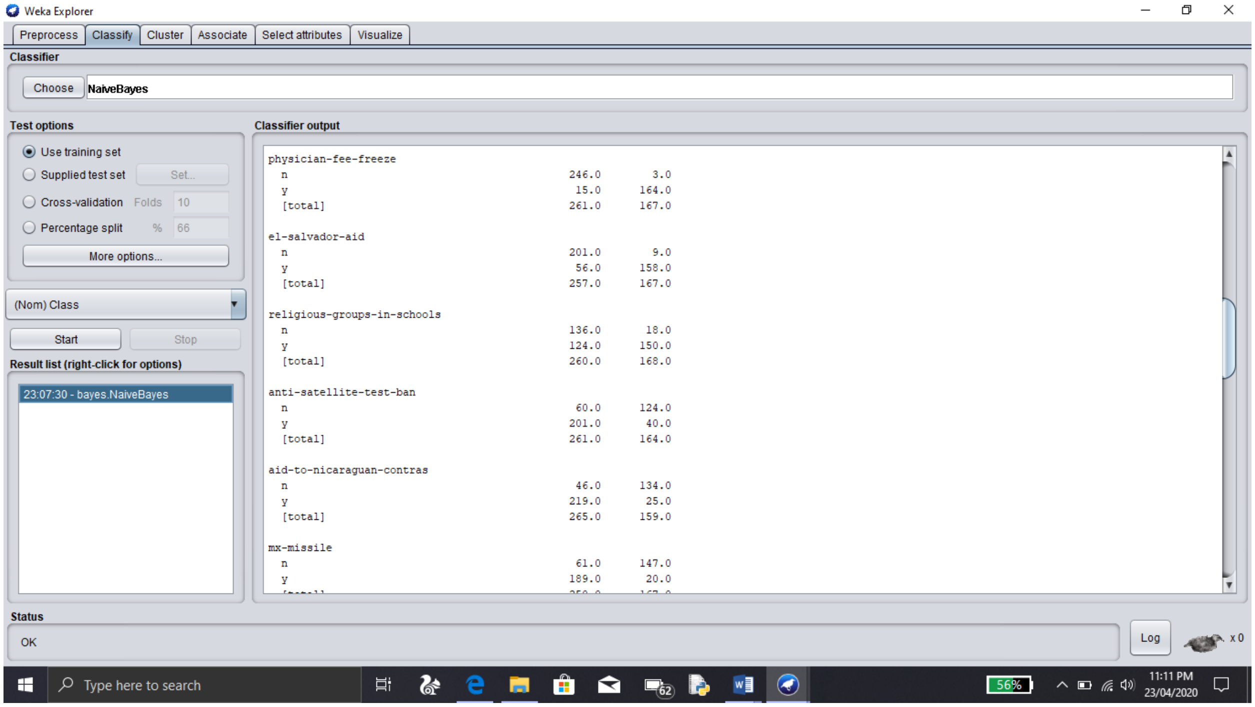
Classifier output

```
             export-administration-act-south-africa
             Class
Test mode:    evaluate on training data

=== Classifier model (full training set) ===

Naive Bayes Classifier

                                          Class
Attribute                              democrat republican
                                         (0.61)      (0.39)
================================================================
handicapped-infants
  n                                        103.0      135.0
  y                                        157.0       32.0
  [total]                                  260.0      167.0

water-project-cost-sharing
  n                                        120.0       74.0
  y                                        121.0       76.0
  [total]                                  241.0      150.0

adoption-of-the-budget-resolution
  n                                         30.0      143.0
  y                                        232.0       23.0
  [total]                                  262.0      166.0

physician-fee-freeze
```

Status
OK

11:10 PM
23/04/2020

Name M Zoraiz Ali
Id 14413
Data mining
Sir Zain Shaukat



Classifier: NaiveBayes

Test options
● Use training set
○ Supplied test set
○ Cross-validation Folds 10
○ Percentage split % 66

(Nom) Class

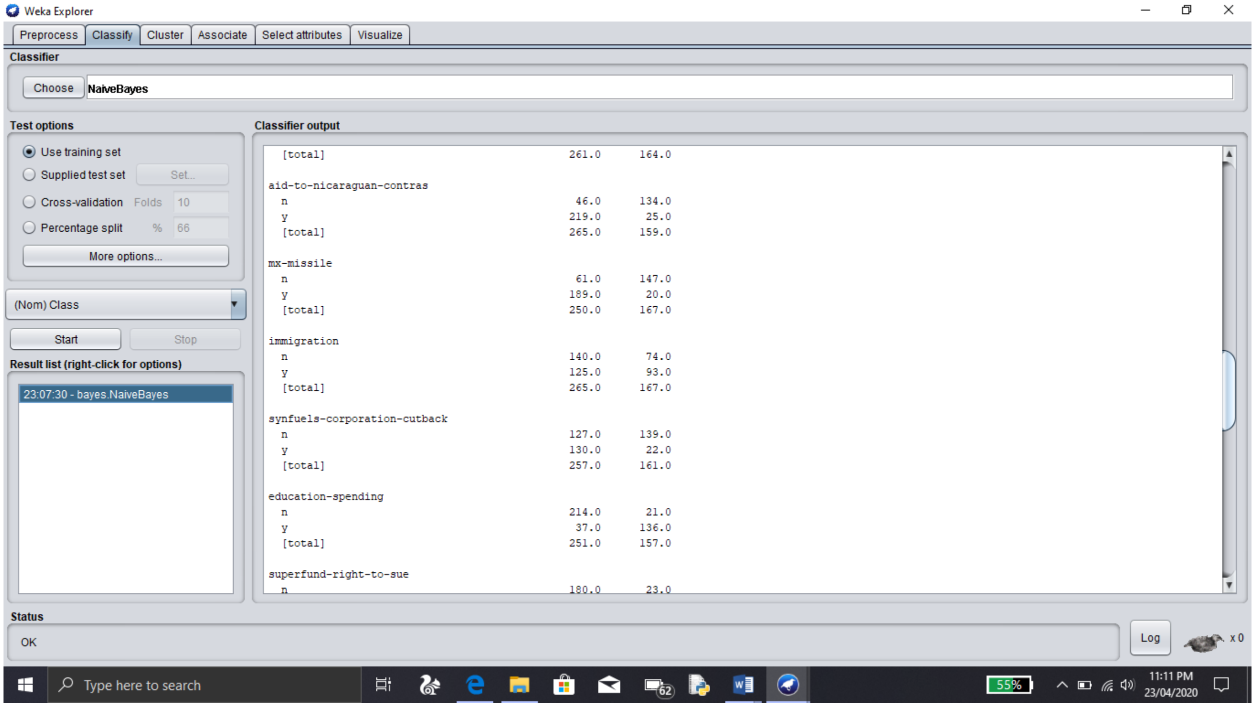
Result list (right-click for options)
23:07:30 - bayes.NaiveBayes

Classifier output

```
physician-fee-freeze
  n                                       246.0       3.0
  y                                        15.0     164.0
  [total]                                 261.0     167.0

el-salvador-aid
  n                                       201.0       9.0
  y                                        56.0     158.0
  [total]                                 257.0     167.0

religious-groups-in-schools
  n                                       136.0      18.0
  y                                       124.0     150.0
  [total]                                 260.0     168.0

anti-satellite-test-ban
  n                                        60.0     124.0
  y                                       201.0      40.0
  [total]                                 261.0     164.0

aid-to-nicaraguan-contras
  n                                        46.0     134.0
  y                                       219.0      25.0
  [total]                                 265.0     159.0

mx-missile
  n                                        61.0     147.0
  y                                       189.0      20.0
```

Status
OK

Name M Zoraiz Ali
Id 14413
Data mining
Sir Zain Shaukat



Classifier: NaiveBayes

Test options:
● Use training set
○ Supplied test set
○ Cross-validation Folds 10
○ Percentage split % 66

(Nom) Class

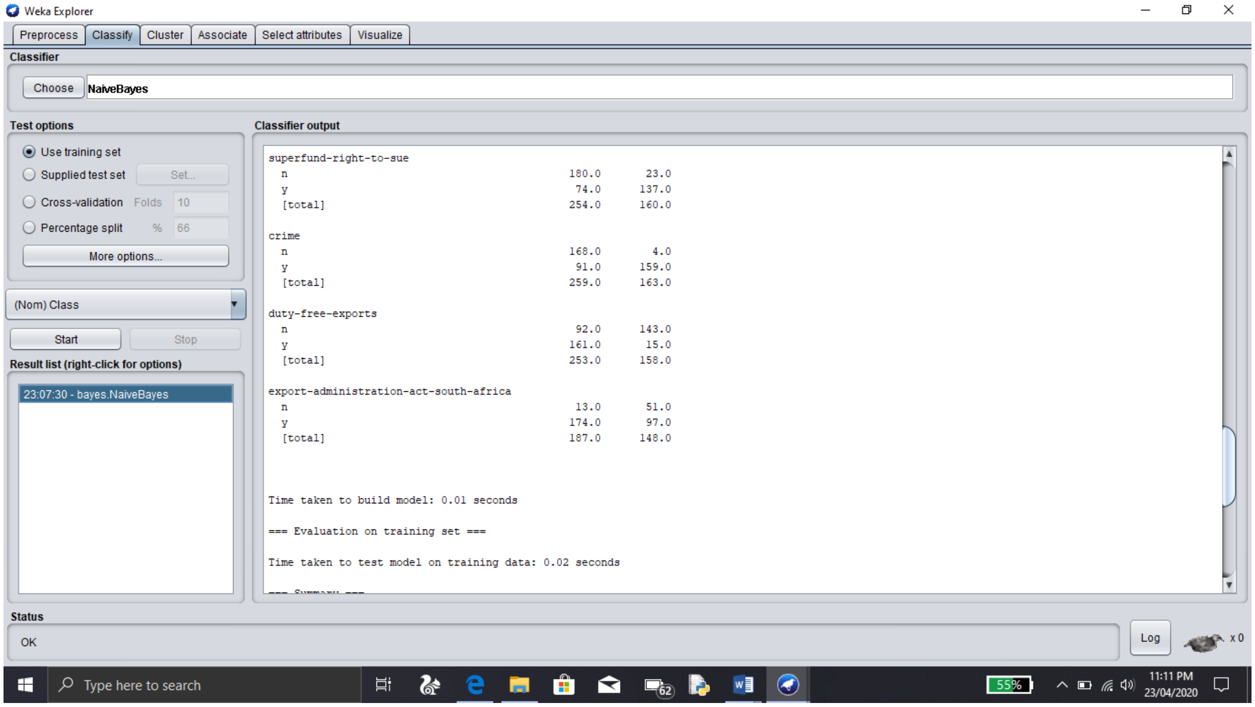
Result list (right-click for options): 23:07:30 - bayes.NaiveBayes

Classifier output

```
  [total]                                            261.0      164.0

aid-to-nicaraguan-contras
  n                                                   46.0      134.0
  y                                                  219.0       25.0
  [total]                                            265.0      159.0

mx-missile
  n                                                   61.0      147.0
  y                                                  189.0       20.0
  [total]                                            250.0      167.0

immigration
  n                                                  140.0       74.0
  y                                                  125.0       93.0
  [total]                                            265.0      167.0

synfuels-corporation-cutback
  n                                                  127.0      139.0
  y                                                  130.0       22.0
  [total]                                            257.0      161.0

education-spending
  n                                                  214.0       21.0
  y                                                   37.0      136.0
  [total]                                            251.0      157.0

superfund-right-to-sue
  n                                                  180.0       23.0
```

Status
OK

Name M Zoraiz Ali
Id 14413
Data mining
Sir Zain Shaukat



Classifier: NaiveBayes

Test options:
● Use training set
○ Supplied test set
○ Cross-validation Folds 10
○ Percentage split % 66

(Nom) Class

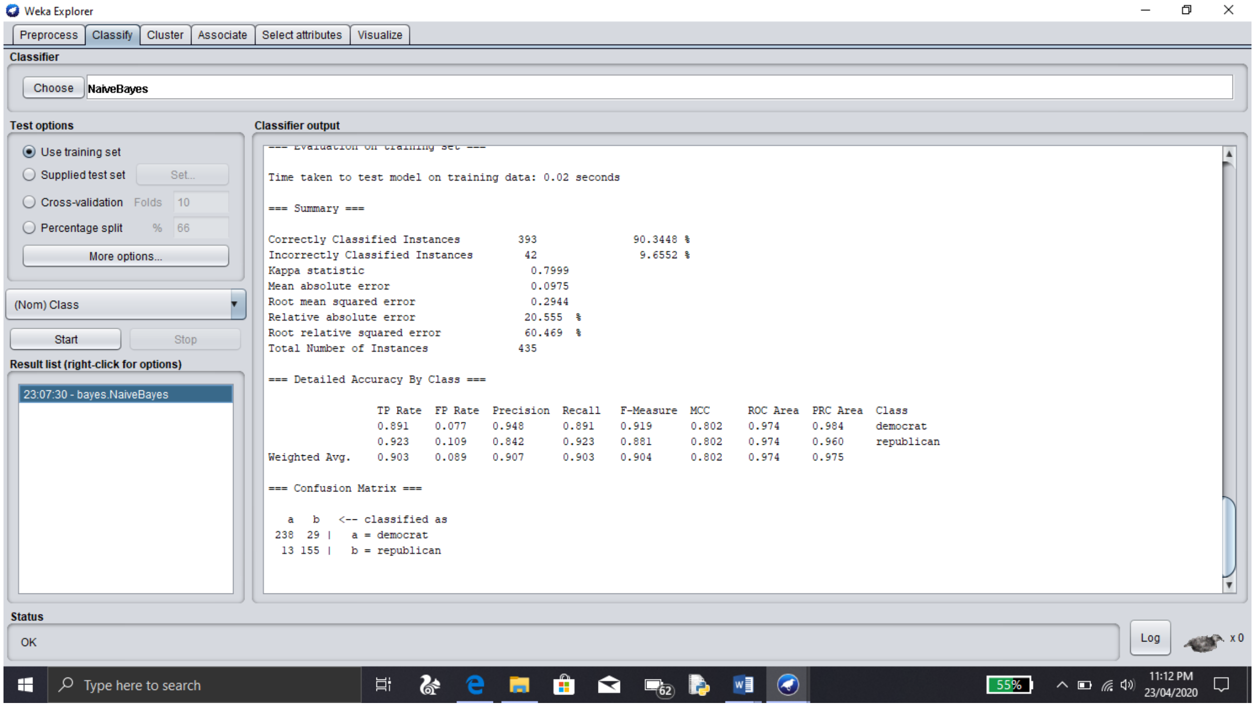
Result list (right-click for options)
23:07:30 - bayes.NaiveBayes

Classifier output

```
superfund-right-to-sue
  n                                          180.0      23.0
  y                                           74.0     137.0
  [total]                                    254.0     160.0

crime
  n                                          168.0       4.0
  y                                           91.0     159.0
  [total]                                    259.0     163.0

duty-free-exports
  n                                           92.0     143.0
  y                                          161.0      15.0
  [total]                                    253.0     158.0

export-administration-act-south-africa
  n                                           13.0      51.0
  y                                          174.0      97.0
  [total]                                    187.0     148.0



Time taken to build model: 0.01 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.02 seconds

=== Summary ===
```

Status
OK

Name M Zoraiz Ali
Id 14413
Data mining
Sir Zain Shaukat



Classifier: NaiveBayes

Test options:
● Use training set
○ Supplied test set
○ Cross-validation Folds 10
○ Percentage split % 66

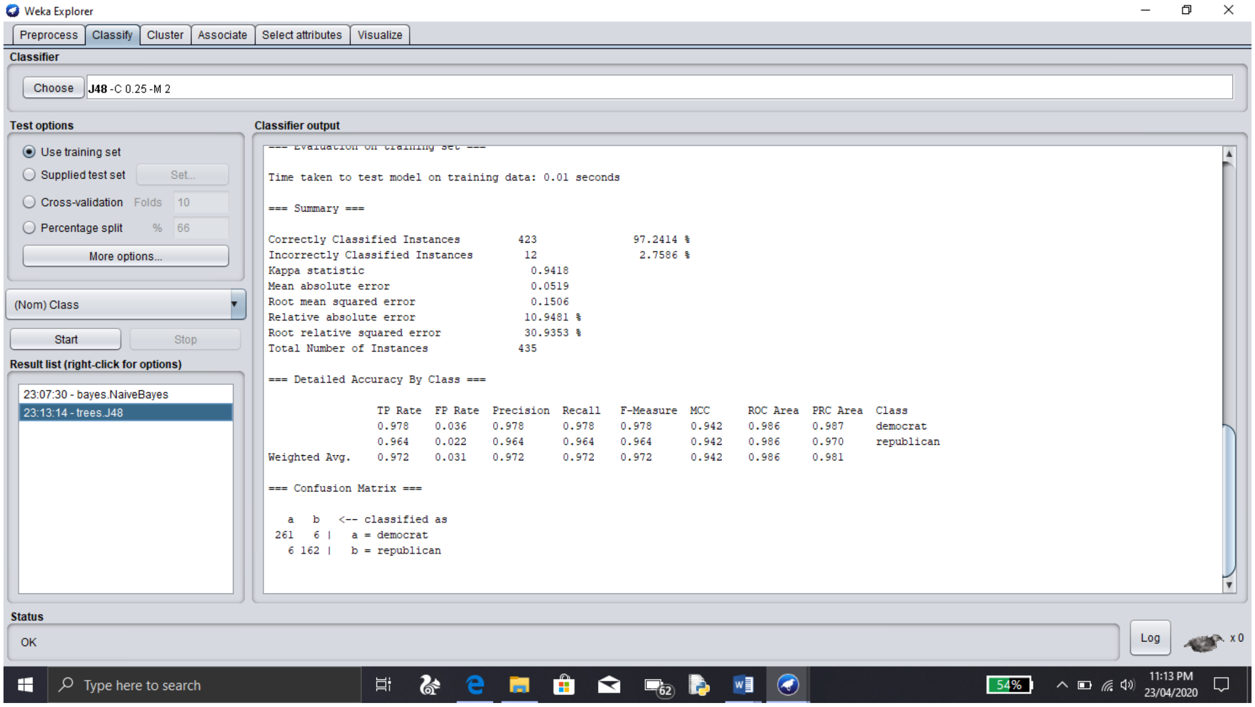
(Nom) Class

Result list: 23:07:30 - bayes.NaiveBayes

```
=== Evaluation on training set ===

Time taken to test model on training data: 0.02 seconds

=== Summary ===

Correctly Classified Instances         393               90.3448 %
Incorrectly Classified Instances        42                9.6552 %
Kappa statistic                          0.7999
Mean absolute error                      0.0975
Root mean squared error                  0.2944
Relative absolute error                 20.555  %
Root relative squared error             60.469  %
Total Number of Instances              435

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.891    0.077    0.948      0.891    0.919      0.802    0.974     0.984     democrat
                 0.923    0.109    0.842      0.923    0.881      0.802    0.974     0.960     republican
Weighted Avg.    0.903    0.089    0.907      0.903    0.904      0.802    0.974     0.975

=== Confusion Matrix ===

   a   b   <-- classified as
 238  29 |   a = democrat
  13 155 |   b = republican
```

Status: OK

Name M Zoraiz Ali
Id 14413
Data mining
Sir Zain Shaukat



Classifier

J48 -C 0.25 -M 2

Test options

● Use training set
○ Supplied test set
○ Cross-validation Folds 10
○ Percentage split % 66

(Nom) Class

Result list (right-click for options)

23:07:30 - bayes.NaiveBayes
23:13:14 - trees.J48

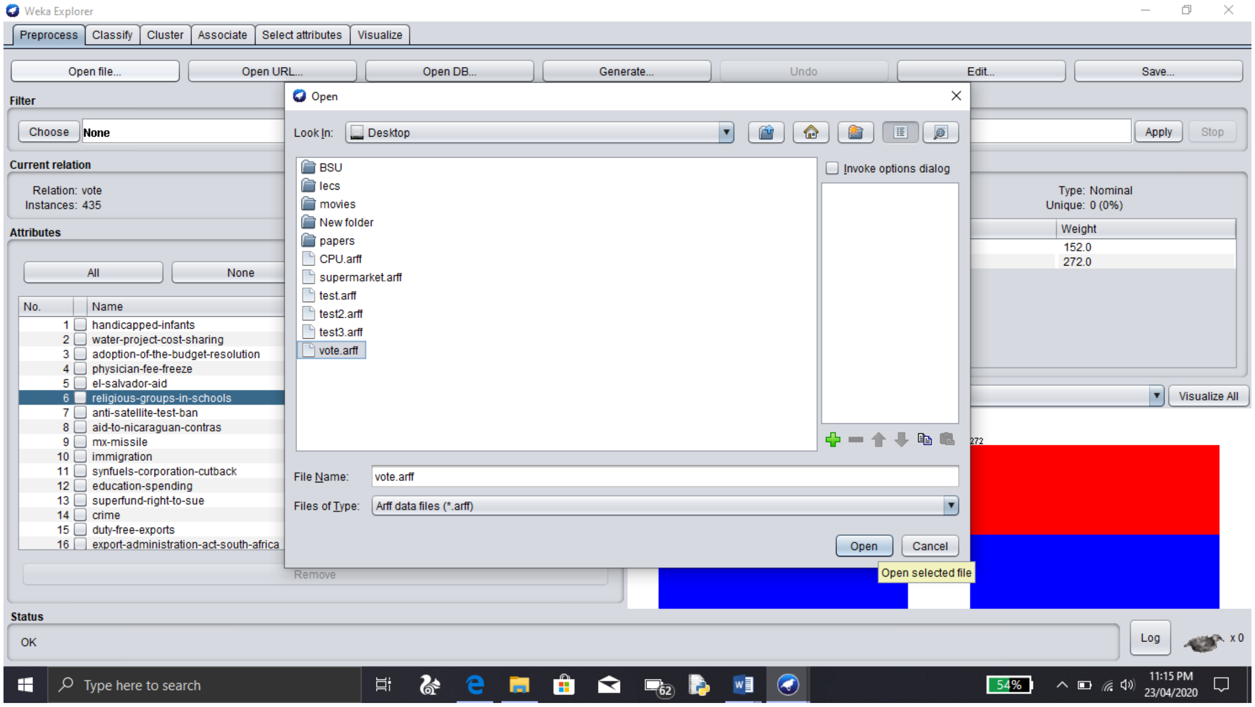
Classifier output

```
=== Evaluation on training set ===

Time taken to test model on training data: 0.01 seconds

=== Summary ===

Correctly Classified Instances         423               97.2414 %
Incorrectly Classified Instances        12                2.7586 %
Kappa statistic                          0.9418
Mean absolute error                      0.0519
Root mean squared error                  0.1506
Relative absolute error                 10.9481 %
Root relative squared error             30.9353 %
Total Number of Instances              435

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.978    0.036    0.978      0.978    0.978      0.942    0.986     0.987     democrat
                 0.964    0.022    0.964      0.964    0.964      0.942    0.986     0.970     republican
Weighted Avg.    0.972    0.031    0.972      0.972    0.972      0.942    0.986     0.981

=== Confusion Matrix ===

   a   b   <-- classified as
 261   6 |   a = democrat
   6 162 |   b = republican
```

Status

OK

Name M Zoraiz Ali
Id 14413
Data mining
Sir Zain Shaukat



Filter: None

Relation: vote
Instances: 435

No.	Name
1	handicapped-infants
2	water-project-cost-sharing
3	adoption-of-the-budget-resolution
4	physician-fee-freeze
5	el-salvador-aid
6	religious-groups-in-schools
7	anti-satellite-test-ban
8	aid-to-nicaraguan-contras
9	mx-missile
10	immigration
11	synfuels-corporation-cutback
12	education-spending
13	superfund-right-to-sue
14	crime
15	duty-free-exports
16	export-administration-act-south-africa

Type: Nominal
Unique: 0 (0%)

Weight
152.0
272.0

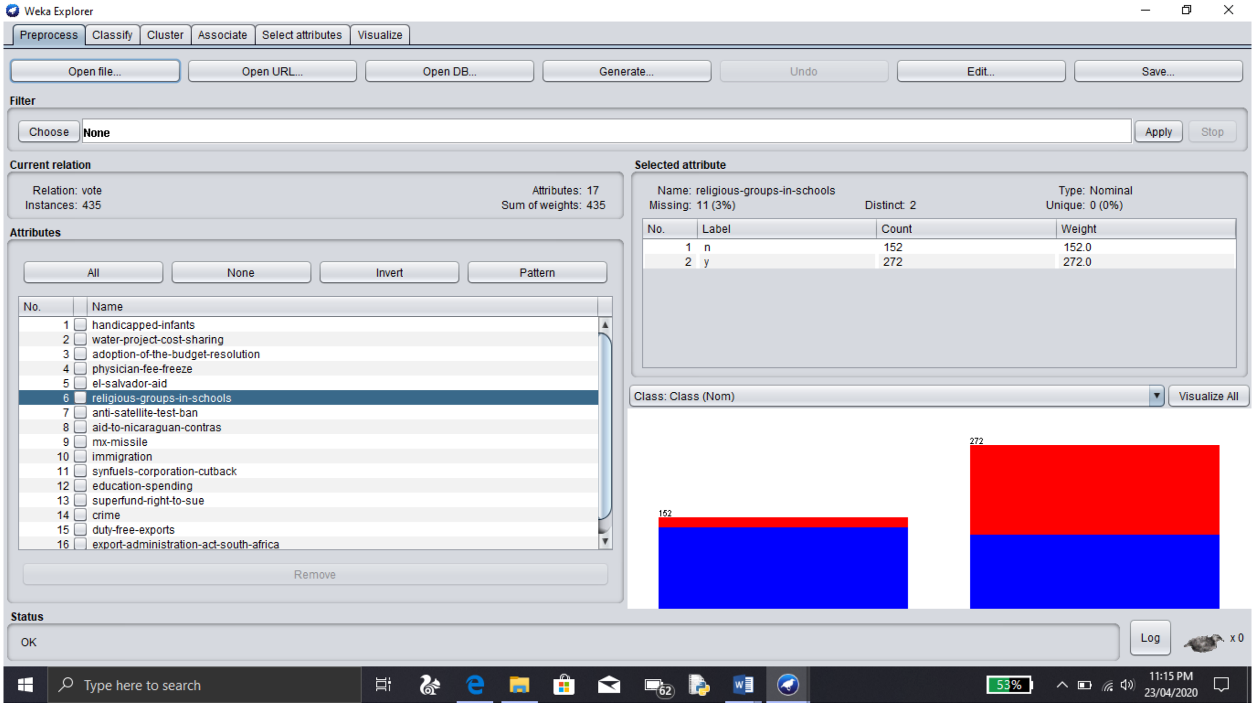
Look In: Desktop

- BSU
- lecs
- movies
- New folder
- papers
- CPU.arff
- supermarket.arff
- test.arff
- test2.arff
- test3.arff
- vote.arff

File Name: vote.arff
Files of Type: Arff data files (*.arff)

OK

Name M Zoraiz Ali
Id 14413
Data mining
Sir Zain Shaukat



Filter

Choose None

Current relation

Relation: vote
Instances: 435
Attributes: 17
Sum of weights: 435

Attributes

No.		Name
1	□	handicapped-infants
2	□	water-project-cost-sharing
3	□	adoption-of-the-budget-resolution
4	□	physician-fee-freeze
5	□	el-salvador-aid
6	□	religious-groups-in-schools
7	□	anti-satellite-test-ban
8	□	aid-to-nicaraguan-contras
9	□	mx-missile
10	□	immigration
11	□	synfuels-corporation-cutback
12	□	education-spending
13	□	superfund-right-to-sue
14	□	crime
15	□	duty-free-exports
16	□	export-administration-act-south-africa

Selected attribute

Name: religious-groups-in-schools
Missing: 11 (3%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)

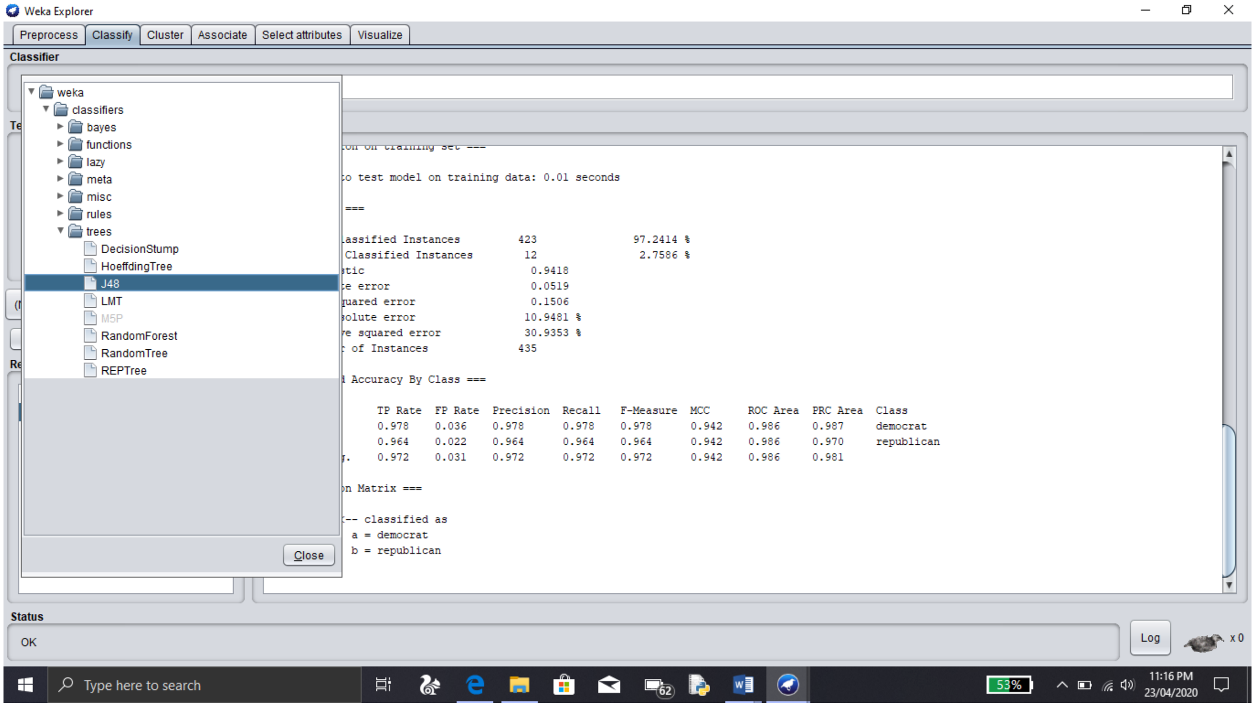
No.	Label	Count	Weight
1	n	152	152.0
2	y	272	272.0

Class: Class (Nom)

Status

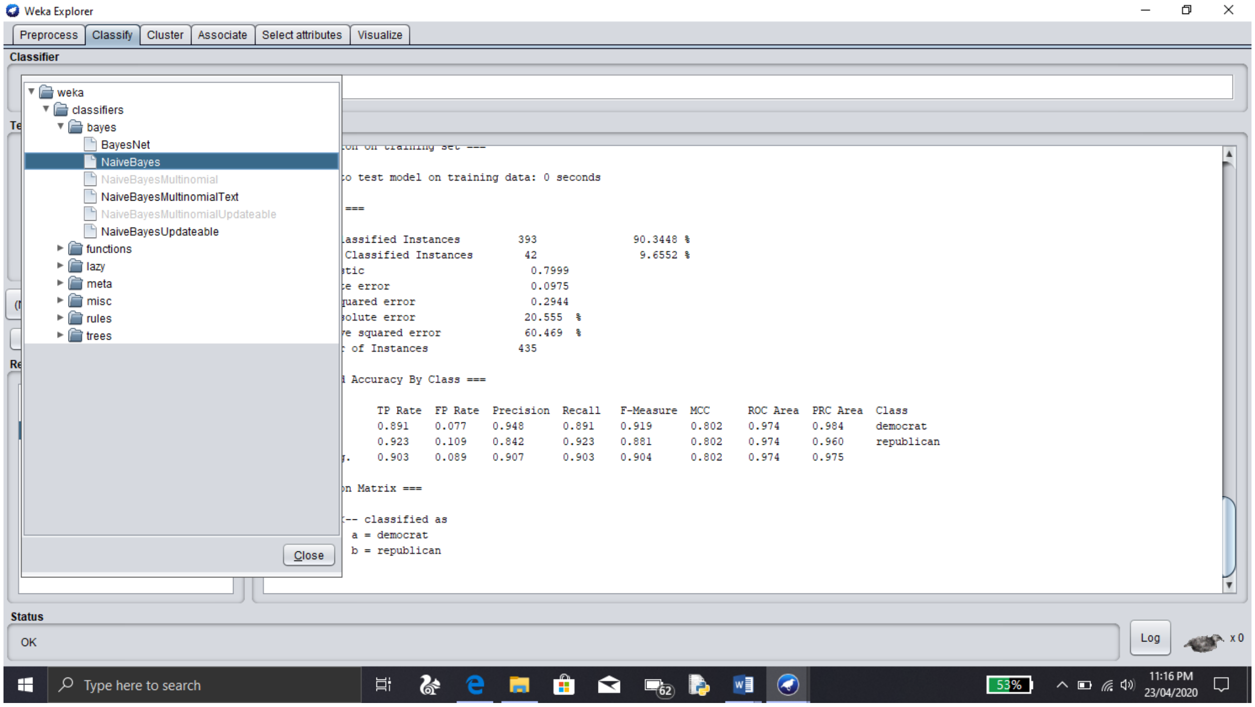
OK

Name M Zoraiz Ali
Id 14413
Data mining
Sir Zain Shaukat



```
ion on training set ===

to test model on training data: 0.01 seconds

 ===

lassified Instances        423               97.2414 %
 Classified Instances       12                2.7586 %
stic                         0.9418
te error                     0.0519
quared error                 0.1506
solute error                10.9481 %
ve squared error            30.9353 %
t of Instances             435

d Accuracy By Class ===

     TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
     0.978    0.036    0.978      0.978    0.978      0.942    0.986     0.987     democrat
     0.964    0.022    0.964      0.964    0.964      0.942    0.986     0.970     republican
g.   0.972    0.031    0.972      0.972    0.972      0.942    0.986     0.981

on Matrix ===

<-- classified as
 a = democrat
 b = republican
```

Name M Zoraiz Ali
Id 14413
Data mining
Sir Zain Shaukat



```
ion on training set ===

to test model on training data: 0 seconds

 ===

lassified Instances         393               90.3448 %
 Classified Instances        42                9.6552 %
stic                          0.7999
te error                      0.0975
quared error                  0.2944
bolute error                 20.555  %
ve squared error             60.469  %
t of Instances              435

d Accuracy By Class ===

     TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
     0.891    0.077    0.948      0.891    0.919      0.802    0.974     0.984     democrat
     0.923    0.109    0.842      0.923    0.881      0.802    0.974     0.960     republican
g.   0.903    0.089    0.907      0.903    0.904      0.802    0.974     0.975

on Matrix ===

<-- classified as
 a = democrat
 b = republican
```