Experiment-3.1

Student Name: Ashish Kumar
UID: 23MAI10008
Branch: CSE AIML
Section/Group: 23MAI-1
Semester: 01
Date of Performance: 23/10/2023
Subject Name: Artificial Intelligence Lab
Subject Code: 23CSH-621

Aim of the Experiment :

Aim of the Experiment is to train and test the prediction dataset (from UCI ML repository) using the 10 popular ML models and build your own ensemble using Majority voting and Stacking methods.

Objective of the Experiment :

Task to be done for this experiment is that we have to perform following tasks:

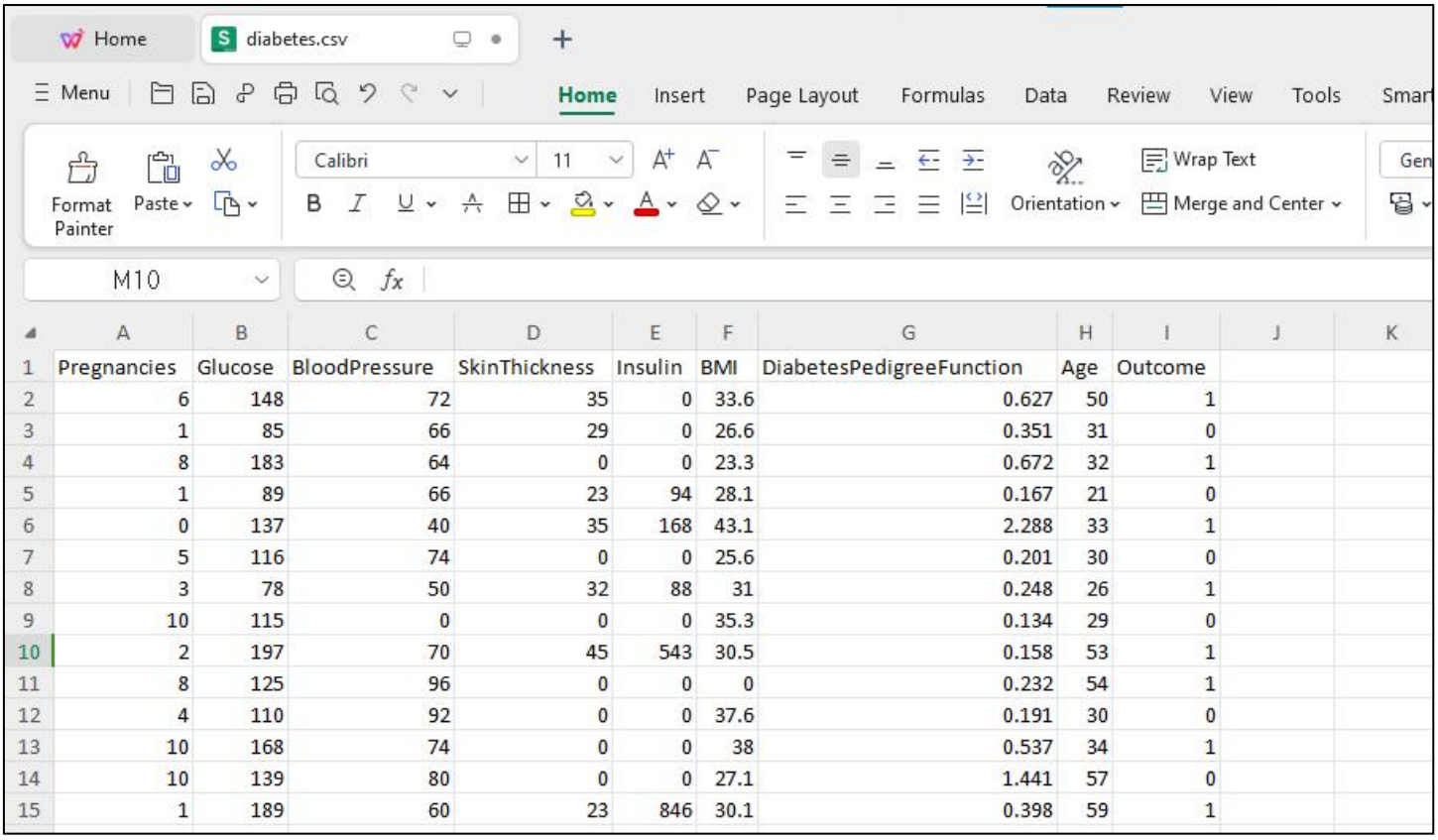
a) Train and test the prediction dataset (from UCI ML repository) using 10 popular ML models and build your own ensemble using Majority voting.

b) Build your own ensemble using Stacking.

Algorithm/ Steps for Experiment :

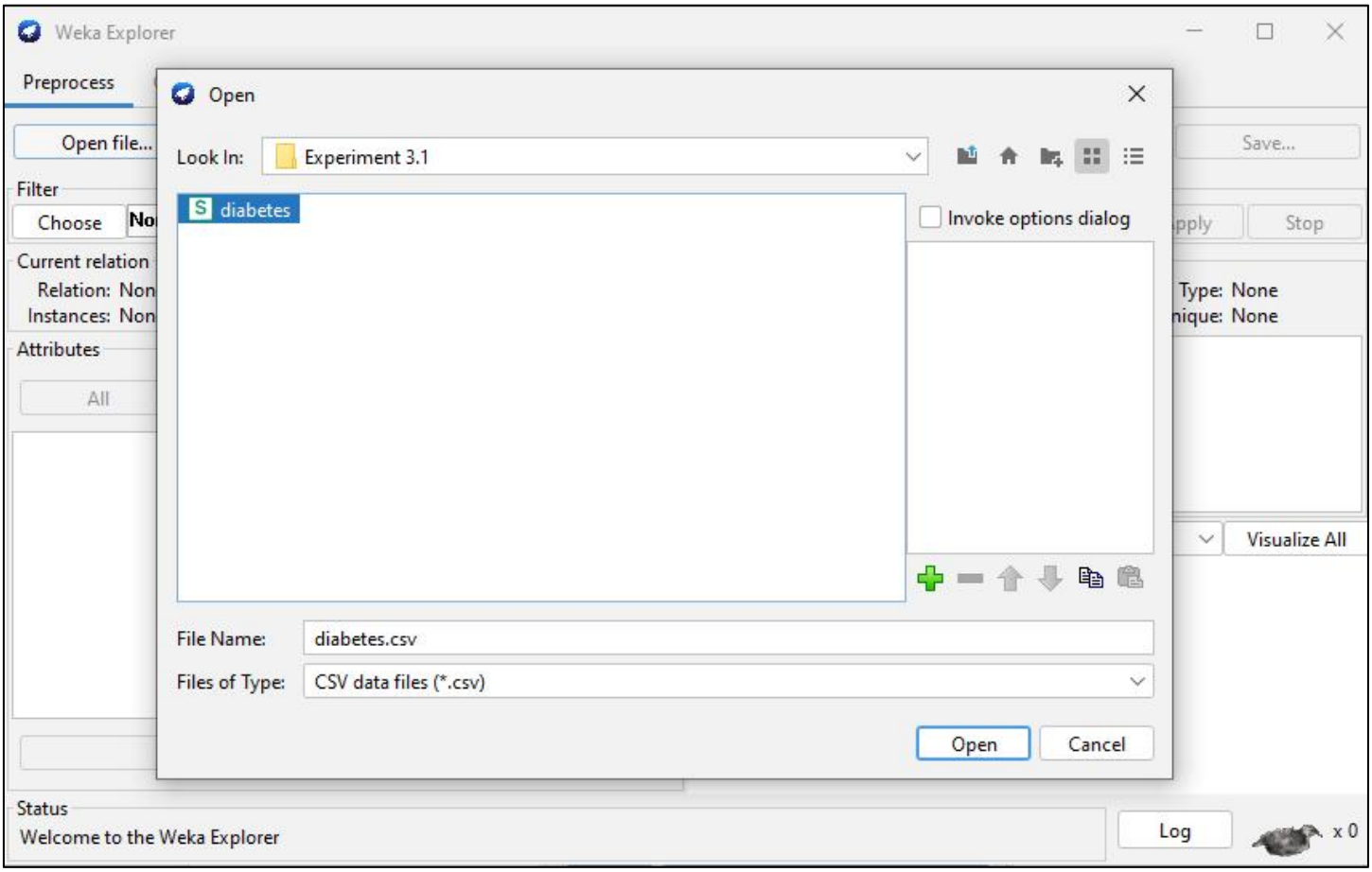
Step 1: Download the **Diabetes dataset** from UCI Machine Learning repository.

Step 2: Open the WEKA Tool and open the **'Explorer'** tab.

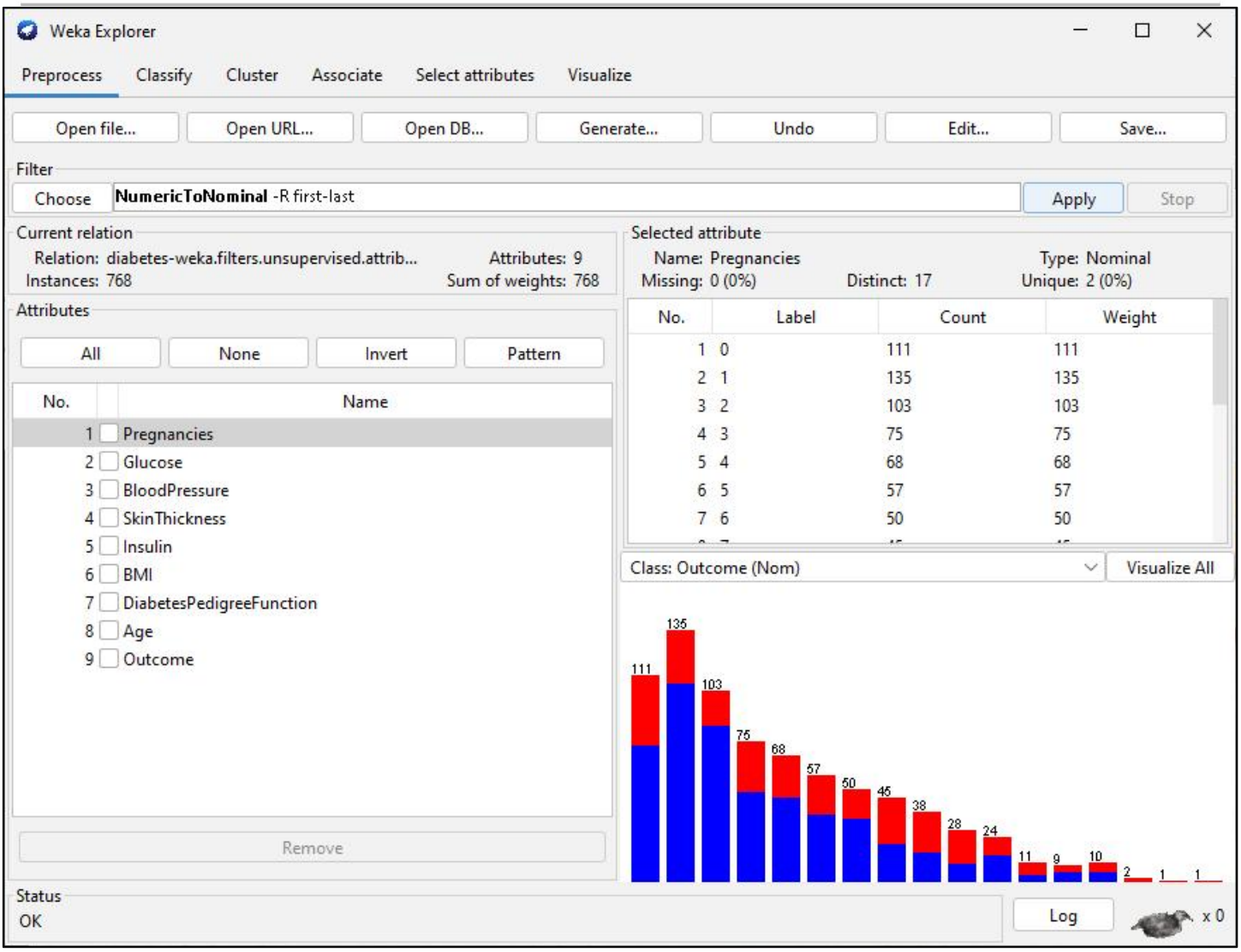


	A	B	C	D	E	F	G	H	I
1	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
2	6	148	72	35	0	33.6	0.627	50	1
3	1	85	66	29	0	26.6	0.351	31	0
4	8	183	64	0	0	23.3	0.672	32	1
5	1	89	66	23	94	28.1	0.167	21	0
6	0	137	40	35	168	43.1	2.288	33	1
7	5	116	74	0	0	25.6	0.201	30	0
8	3	78	50	32	88	31	0.248	26	1
9	10	115	0	0	0	35.3	0.134	29	0
10	2	197	70	45	543	30.5	0.158	53	1
11	8	125	96	0	0	0	0.232	54	1
12	4	110	92	0	0	37.6	0.191	30	0
13	10	168	74	0	0	38	0.537	34	1
14	10	139	80	0	0	27.1	1.441	57	0
15	1	189	60	23	846	30.1	0.398	59	1

Step 3: Click on the **'Open file'** Option >> Select Diabetes dataset >> Click on **Open**.



Step 4: Change the dataset from Numeric to Nominal. In **'Filter'** Section, click on **'Choose'** >> Unsupervised >> Attribute >> **Numeric to Nominal**. Click on **'Apply'** Option.



Step 5: Click on the **'Classify'** Tab >> Choose and select different models from **'Classifier'** Section.

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier

Choose **ZeroR**

Test options

- Use training set
- Supplied test set — Set...
- Cross-validation — Folds 10
- Percentage split — % 66

More options...

(Nom) Outcome

Start | Stop

Result list (right-click for options)

18:33:35 - rules.ZeroR

Classifier output

```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         500               65.1042 %
Incorrectly Classified Instances       268               34.8958 %
Kappa statistic                          0
Mean absolute error                      0.4545
Root mean squared error                  0.4766
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              768

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               1.000    1.000    0.651      1.000    0.789      ?        0.497     0.650     0
               0.000    0.000    ?          0.000    ?          ?        0.497     0.348     1
Weighted Avg.  0.651    0.651    ?          0.651    ?          ?        0.497     0.544

=== Confusion Matrix ===

   a   b   <-- classified as
 500   0 |   a = 0
 268   0 |   b = 1
```

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier

Choose **J48** -C 0.25 -M 2

Test options

- Use training set
- Supplied test set — Set...
- Cross-validation — Folds 10
- Percentage split — % 66

More options...

(Nom) Outcome

Start | Stop

Result list (right-click for options)

18:33:35 - rules.ZeroR
18:34:54 - trees.J48

Classifier output

```
Time taken to build model: 0.05 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         500               65.1042 %
Incorrectly Classified Instances       268               34.8958 %
Kappa statistic                          0
Mean absolute error                      0.4544
Root mean squared error                  0.4766
Relative absolute error                 99.971  %
Root relative squared error            100      %
Total Number of Instances              768

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               1.000    1.000    0.651      1.000    0.789      ?        0.497     0.650     0
               0.000    0.000    ?          0.000    ?          ?        0.497     0.348     1
Weighted Avg.  0.651    0.651    ?          0.651    ?          ?        0.497     0.544

=== Confusion Matrix ===

   a   b   <-- classified as
 500   0 |   a = 0
 268   0 |   b = 1
```

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier

Choose **RandomForest** -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Test options

- Use training set
- Supplied test set — Set...
- Cross-validation — Folds 10
- Percentage split — % 66

More options...

(Nom) Outcome

Start | Stop

Result list (right-click for options)

- 18:33:35 - rules.ZeroR
- 18:34:54 - trees.J48
- 18:35:14 - trees.RandomForest

Classifier output

```
Time taken to build model: 1.26 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         498               64.8438 %
Incorrectly Classified Instances       270               35.1563 %
Kappa statistic                          0.0872
Mean absolute error                      0.4216
Root mean squared error                  0.4649
Relative absolute error                 92.7581 %
Root relative squared error             97.5441 %
Total Number of Instances              768

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.906     0.832     0.670      0.906     0.770      0.108    0.640     0.773     0
               0.168     0.094     0.489      0.168     0.250      0.108    0.640     0.458     1
Weighted Avg.  0.648     0.575     0.607      0.648     0.589      0.108    0.640     0.663

=== Confusion Matrix ===

   a   b   <-- classified as
 453  47 |   a = 0
 223  45 |   b = 1
```

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier

Choose **NaiveBayes**

Test options

- Use training set
- Supplied test set — Set...
- Cross-validation — Folds 10
- Percentage split — % 66

More options...

(Nom) Outcome

Start | Stop

Result list (right-click for options)

- 18:33:35 - rules.ZeroR
- 18:34:54 - trees.J48
- 18:35:14 - trees.RandomForest
- 18:35:46 - bayes.NaiveBayes

Classifier output

```
Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         521               67.8385 %
Incorrectly Classified Instances       247               32.1615 %
Kappa statistic                          0.2916
Mean absolute error                      0.3456
Root mean squared error                  0.4663
Relative absolute error                 76.0311 %
Root relative squared error             97.8216 %
Total Number of Instances              768

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.754     0.463     0.752      0.754     0.753      0.292    0.737     0.848     0
               0.537     0.246     0.539      0.537     0.538      0.292    0.737     0.560     1
Weighted Avg.  0.678     0.387     0.678      0.678     0.678      0.292    0.737     0.747

=== Confusion Matrix ===

   a   b   <-- classified as
 377 123 |   a = 0
 124 144 |   b = 1
```

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier

Choose **IBk** -K 1 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Test options

- Use training set
- Supplied test set — Set...
- Cross-validation — Folds 10
- Percentage split — % 66

More options...

(Nom) Outcome

Start | Stop

Result list (right-click for options)

- 18:33:35 - rules.ZeroR
- 18:34:54 - trees.J48
- 18:35:14 - trees.RandomForest
- 18:35:46 - bayes.NaiveBayes
- 18:36:09 - lazy.IBk

Classifier output

```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         502               65.3646 %
Incorrectly Classified Instances       266               34.6354 %
Kappa statistic                          0.1477
Mean absolute error                      0.3941
Root mean squared error                  0.5302
Relative absolute error                 86.7083 %
Root relative squared error            111.2312 %
Total Number of Instances              768

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.856     0.724     0.688      0.856     0.763      0.160    0.604     0.710     0
               0.276     0.144     0.507      0.276     0.357      0.160    0.604     0.433     1
Weighted Avg.  0.654     0.522     0.625      0.654     0.621      0.160    0.604     0.613

=== Confusion Matrix ===

   a   b   <-- classified as
 428  72 |   a = 0
 194  74 |   b = 1
```

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier

Choose **DecisionTable** -X 1 -S "weka.attributeSelection.BestFirst -D 1 -N 5"

Test options

- Use training set
- Supplied test set — Set...
- Cross-validation — Folds 10
- Percentage split — % 66

More options...

(Nom) Outcome

Start | Stop

Result list (right-click for options)

- 18:33:35 - rules.ZeroR
- 18:34:54 - trees.J48
- 18:35:14 - trees.RandomForest
- 18:35:46 - bayes.NaiveBayes
- 18:36:09 - lazy.IBk
- 18:36:32 - rules.DecisionTable

Classifier output

```
Time taken to build model: 0.32 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         499               64.974  %
Incorrectly Classified Instances       269               35.026  %
Kappa statistic                          0.1537
Mean absolute error                      0.4253
Root mean squared error                  0.4708
Relative absolute error                 93.5742 %
Root relative squared error             98.7649 %
Total Number of Instances              768

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.834     0.694     0.692      0.834     0.756      0.162    0.626     0.747     0
               0.306     0.166     0.497      0.306     0.379      0.162    0.626     0.450     1
Weighted Avg.  0.650     0.510     0.624      0.650     0.624      0.162    0.626     0.643

=== Confusion Matrix ===

   a   b   <-- classified as
 417  83 |   a = 0
 186  82 |   b = 1
```

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier

Choose **RandomTree** -K 0 -M 1.0 -V 0.001 -S 1

Test options

- Use training set
- Supplied test set — Set...
- Cross-validation — Folds 10
- Percentage split — % 66

More options...

(Nom) Outcome

Start | Stop

Result list (right-click for options)

- 18:33:35 - rules.ZeroR
- 18:34:54 - trees.J48
- 18:35:14 - trees.RandomForest
- 18:35:46 - bayes.NaiveBayes
- 18:36:09 - lazy.IBk
- 18:36:32 - rules.DecisionTable
- 18:36:56 - trees.RandomTree

Classifier output

```
Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         471               61.3281 %
Incorrectly Classified Instances       297               38.6719 %
Kappa statistic                          0.0493
Mean absolute error                      0.4351
Root mean squared error                  0.5526
Relative absolute error                 95.7361 %
Root relative squared error            115.9306 %
Total Number of Instances              768

=== Detailed Accuracy By Class ===

                 TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.824     0.780     0.663      0.824     0.735      0.053    0.542     0.683     0
                 0.220     0.176     0.401      0.220     0.284      0.053    0.542     0.373     1
Weighted Avg.    0.613     0.569     0.572      0.613     0.578      0.053    0.542     0.575

=== Confusion Matrix ===

   a   b   <-- classified as
 412  88 |   a = 0
 209  59 |   b = 1
```

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier

Choose **KStar** -B 20 -M a

Test options

- Use training set
- Supplied test set — Set...
- Cross-validation — Folds 10
- Percentage split — % 66

More options...

(Nom) Outcome

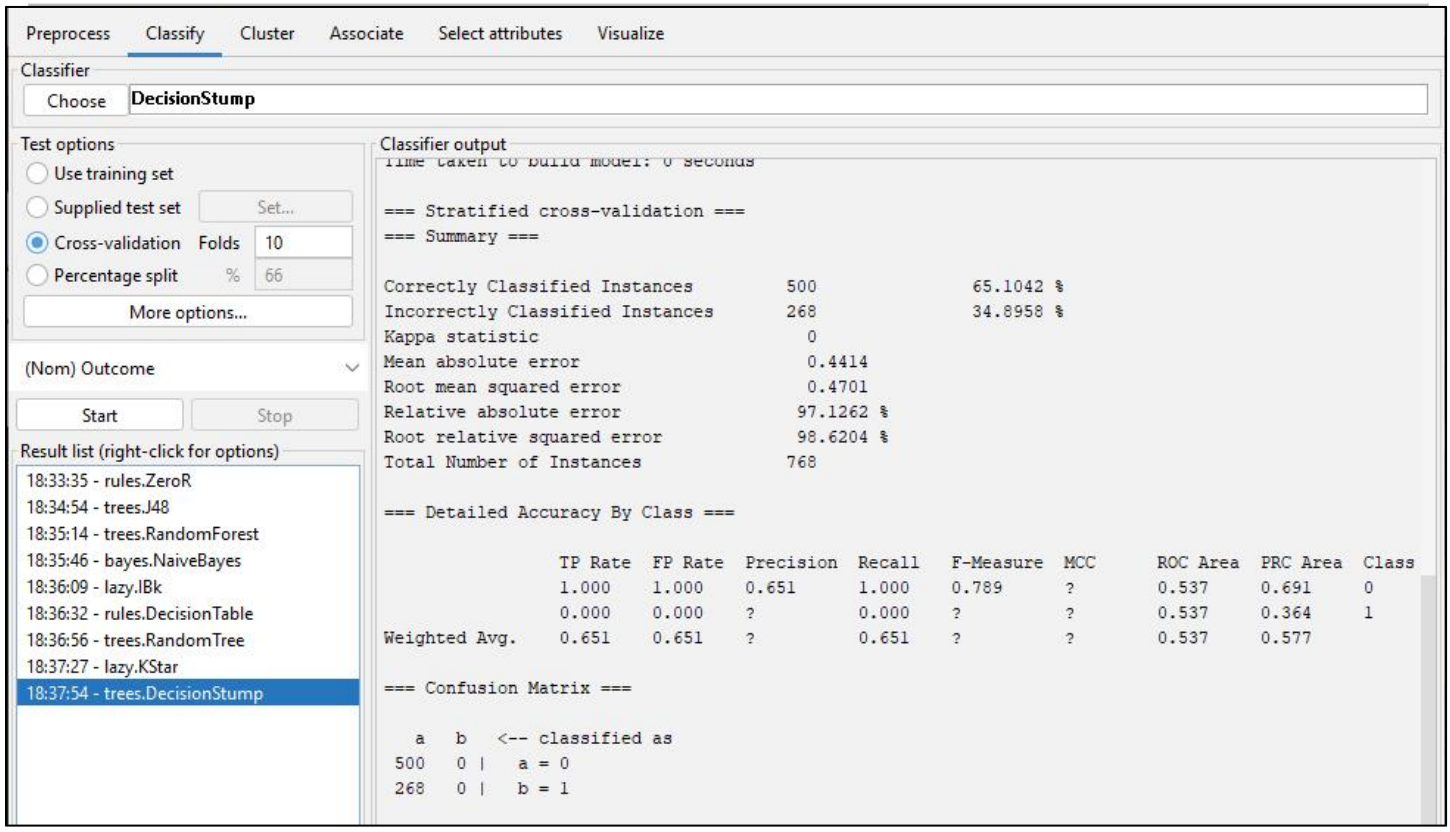
Start | Stop

Result list (right-click for options)

- 18:33:35 - rules.ZeroR
- 18:34:54 - trees.J48
- 18:35:14 - trees.RandomForest
- 18:35:46 - bayes.NaiveBayes
- 18:36:09 - lazy.IBk
- 18:36:32 - rules.DecisionTable
- 18:36:56 - trees.RandomTree
- 18:37:27 - lazy.KStar

Classifier output

```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         506               65.8854 %
Incorrectly Classified Instances       262               34.1146 %
Kappa statistic                          0.149
Mean absolute error                      0.4072
Root mean squared error                  0.4699
Relative absolute error                 89.5881 %
Root relative squared error             98.594  %
Total Number of Instances              768

=== Detailed Accuracy By Class ===

                 TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.874     0.743     0.687      0.874     0.769      0.166    0.653     0.775     0
                 0.257     0.126     0.523      0.257     0.345      0.166    0.653     0.463     1
Weighted Avg.    0.659     0.527     0.630      0.659     0.621      0.166    0.653     0.666

=== Confusion Matrix ===

   a   b   <-- classified as
 437  63 |   a = 0
 199  69 |   b = 1
```



Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier
Choose **DecisionStump**

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

More options...

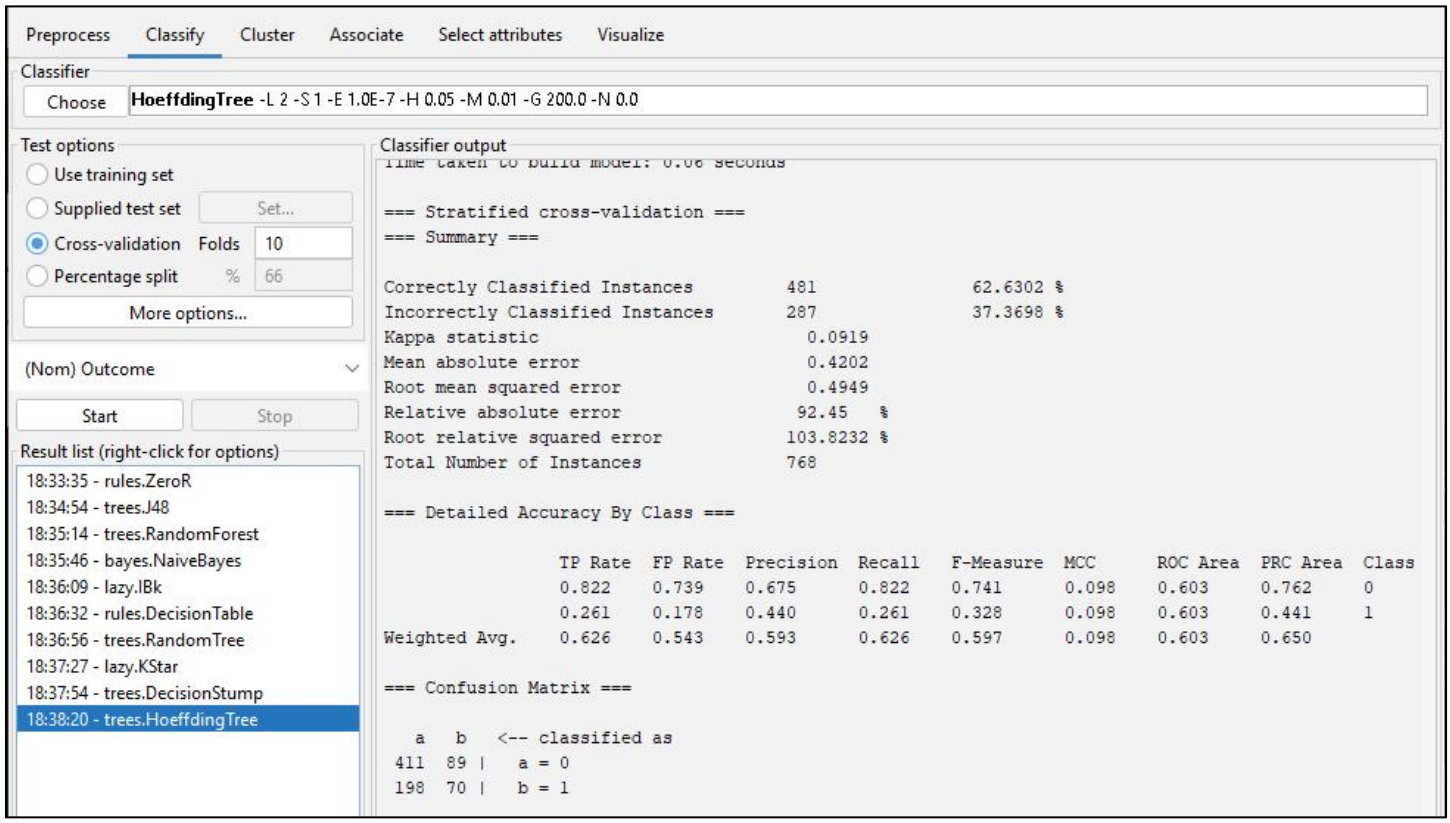
(Nom) Outcome

Start | Stop

Result list (right-click for options)
- 18:33:35 - rules.ZeroR
- 18:34:54 - trees.J48
- 18:35:14 - trees.RandomForest
- 18:35:46 - bayes.NaiveBayes
- 18:36:09 - lazy.IBk
- 18:36:32 - rules.DecisionTable
- 18:36:56 - trees.RandomTree
- 18:37:27 - lazy.KStar
- 18:37:54 - trees.DecisionStump

Classifier output

```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         500               65.1042 %
Incorrectly Classified Instances       268               34.8958 %
Kappa statistic                          0
Mean absolute error                      0.4414
Root mean squared error                  0.4701
Relative absolute error                 97.1262 %
Root relative squared error             98.6204 %
Total Number of Instances              768

=== Detailed Accuracy By Class ===

                 TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.651      1.000    0.789      ?        0.537     0.691     0
                 0.000    0.000    ?          0.000    ?          ?        0.537     0.364     1
Weighted Avg.    0.651    0.651    ?          0.651    ?          ?        0.537     0.577

=== Confusion Matrix ===

   a   b   <-- classified as
 500   0 |   a = 0
 268   0 |   b = 1
```



Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier
Choose **HoeffdingTree** -L 2 -S 1 -E 1.0E-7 -H 0.05 -M 0.01 -G 200.0 -N 0.0

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

More options...

(Nom) Outcome

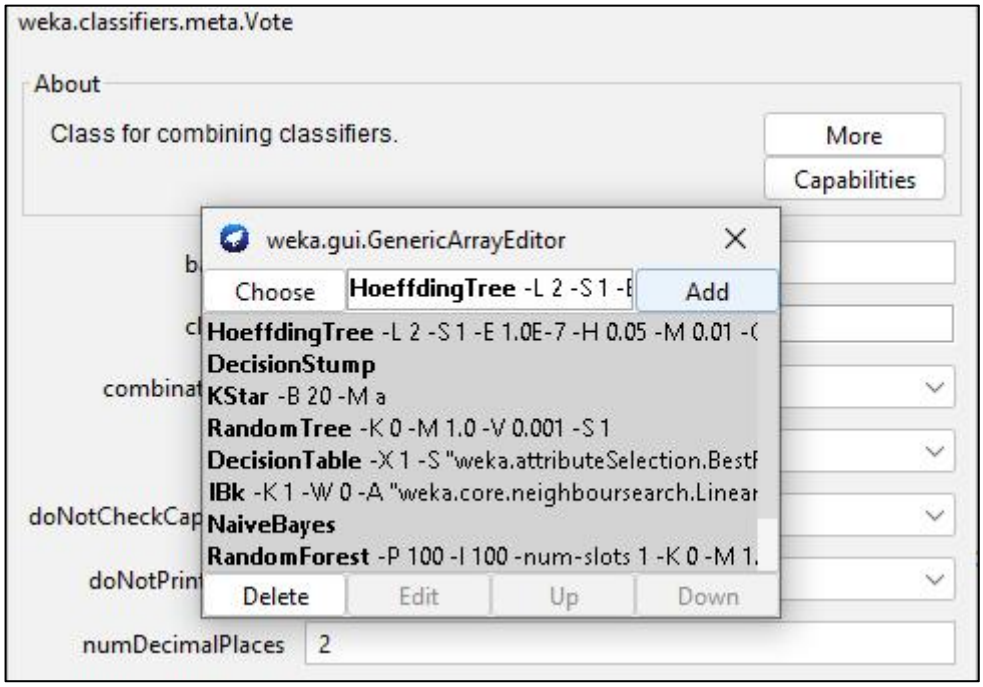
Start | Stop

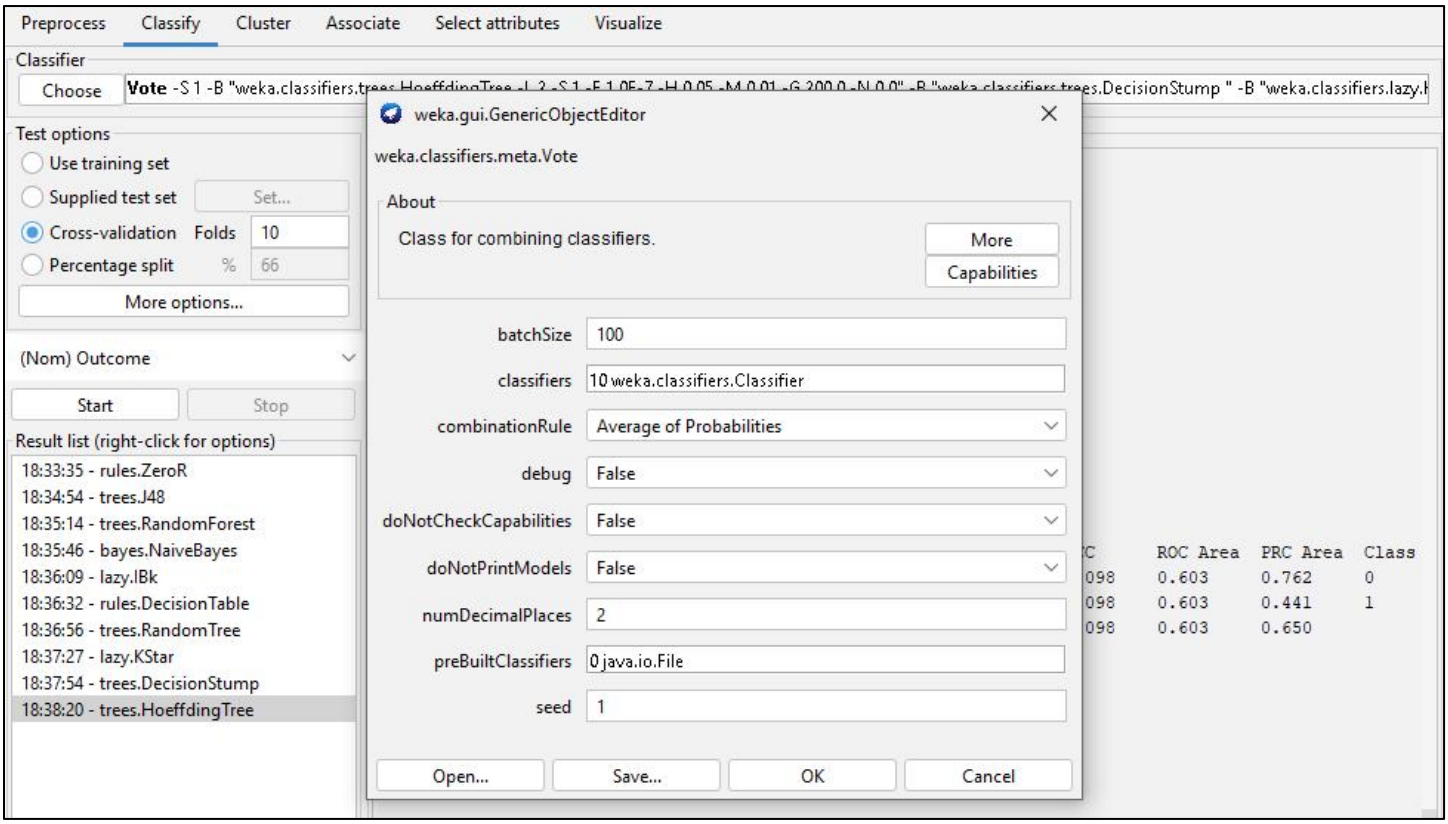
Result list (right-click for options)
- 18:33:35 - rules.ZeroR
- 18:34:54 - trees.J48
- 18:35:14 - trees.RandomForest
- 18:35:46 - bayes.NaiveBayes
- 18:36:09 - lazy.IBk
- 18:36:32 - rules.DecisionTable
- 18:36:56 - trees.RandomTree
- 18:37:27 - lazy.KStar
- 18:37:54 - trees.DecisionStump
- 18:38:20 - trees.HoeffdingTree

Classifier output

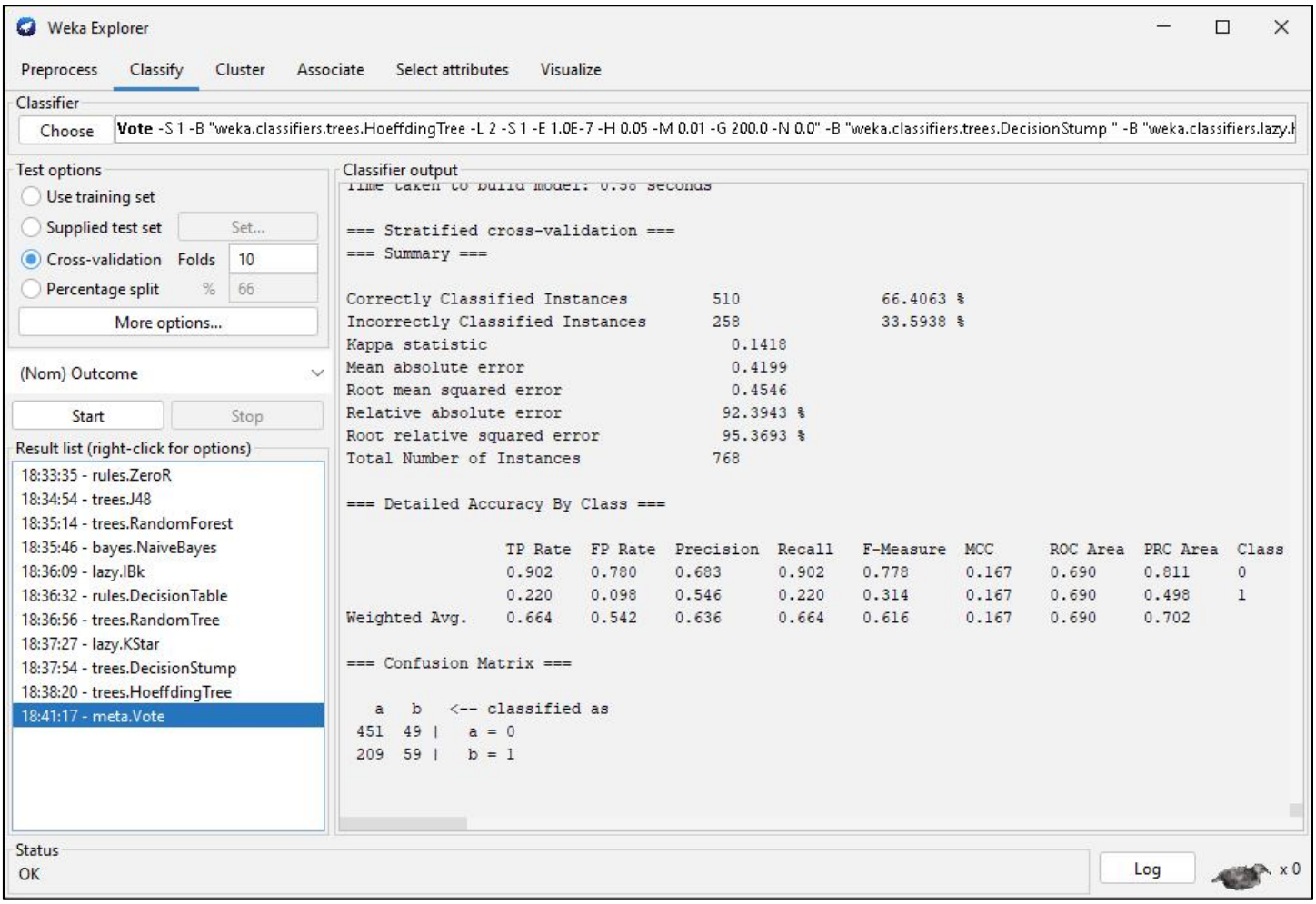
```
Time taken to build model: 0.06 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         481               62.6302 %
Incorrectly Classified Instances       287               37.3698 %
Kappa statistic                          0.0919
Mean absolute error                      0.4202
Root mean squared error                  0.4949
Relative absolute error                 92.45   %
Root relative squared error            103.8232 %
Total Number of Instances              768

=== Detailed Accuracy By Class ===

                 TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.822    0.739    0.675      0.822    0.741      0.098    0.603     0.762     0
                 0.261    0.178    0.440      0.261    0.328      0.098    0.603     0.441     1
Weighted Avg.    0.626    0.543    0.593      0.626    0.597      0.098    0.603     0.650

=== Confusion Matrix ===

   a   b   <-- classified as
 411  89 |   a = 0
 198  70 |   b = 1
```

Step 6: Select the **"Vote"** model for majority voting. Click on the model then click on Classifiers to select 10 different classifier models.

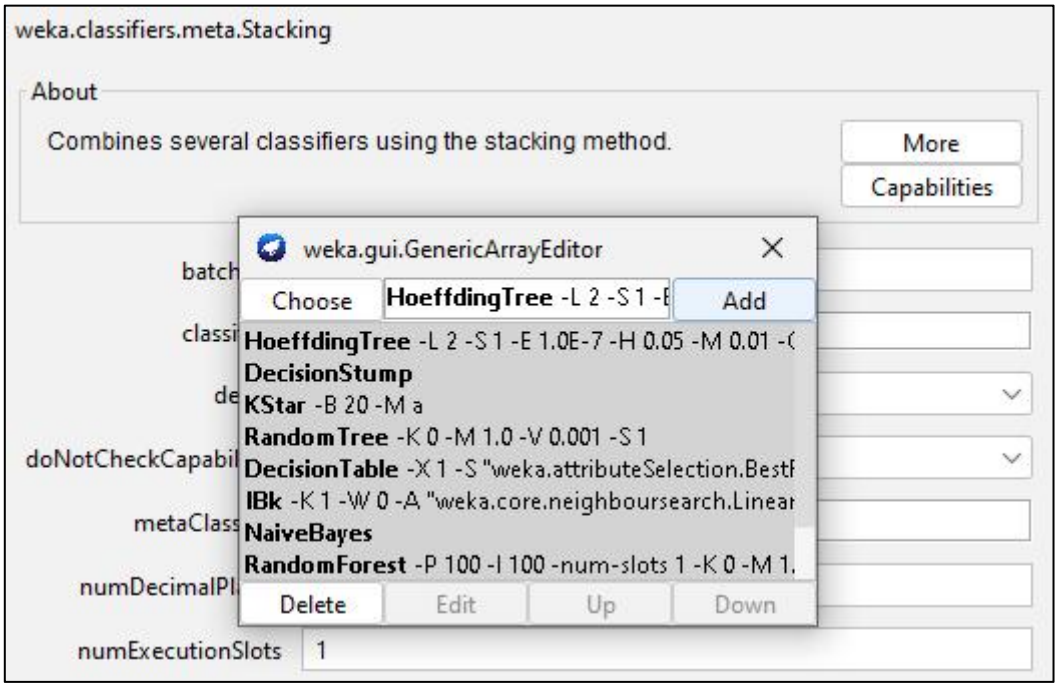


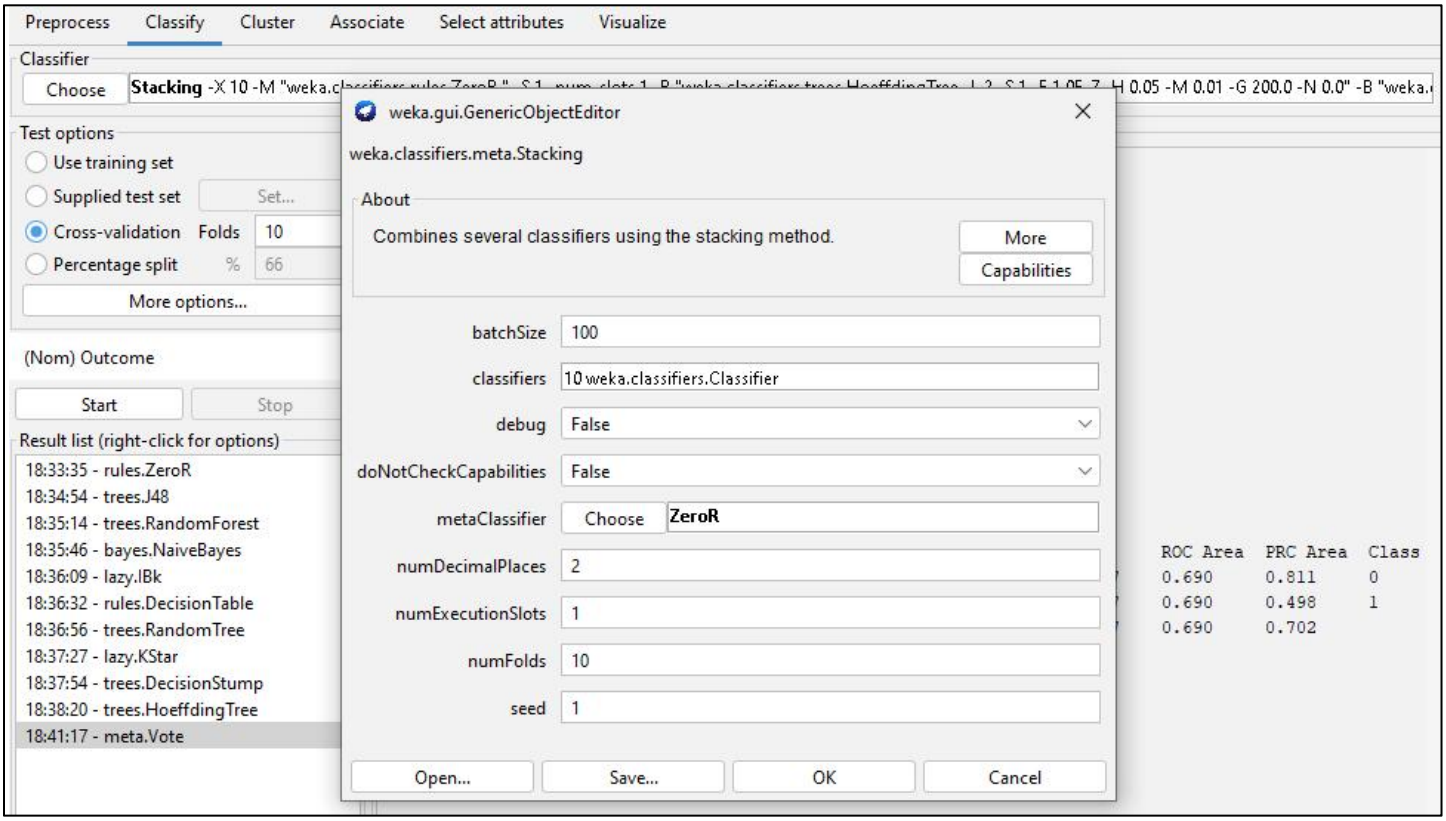


Step 7: Click Start to build the model. Classifier output shows the model evaluation parameters.

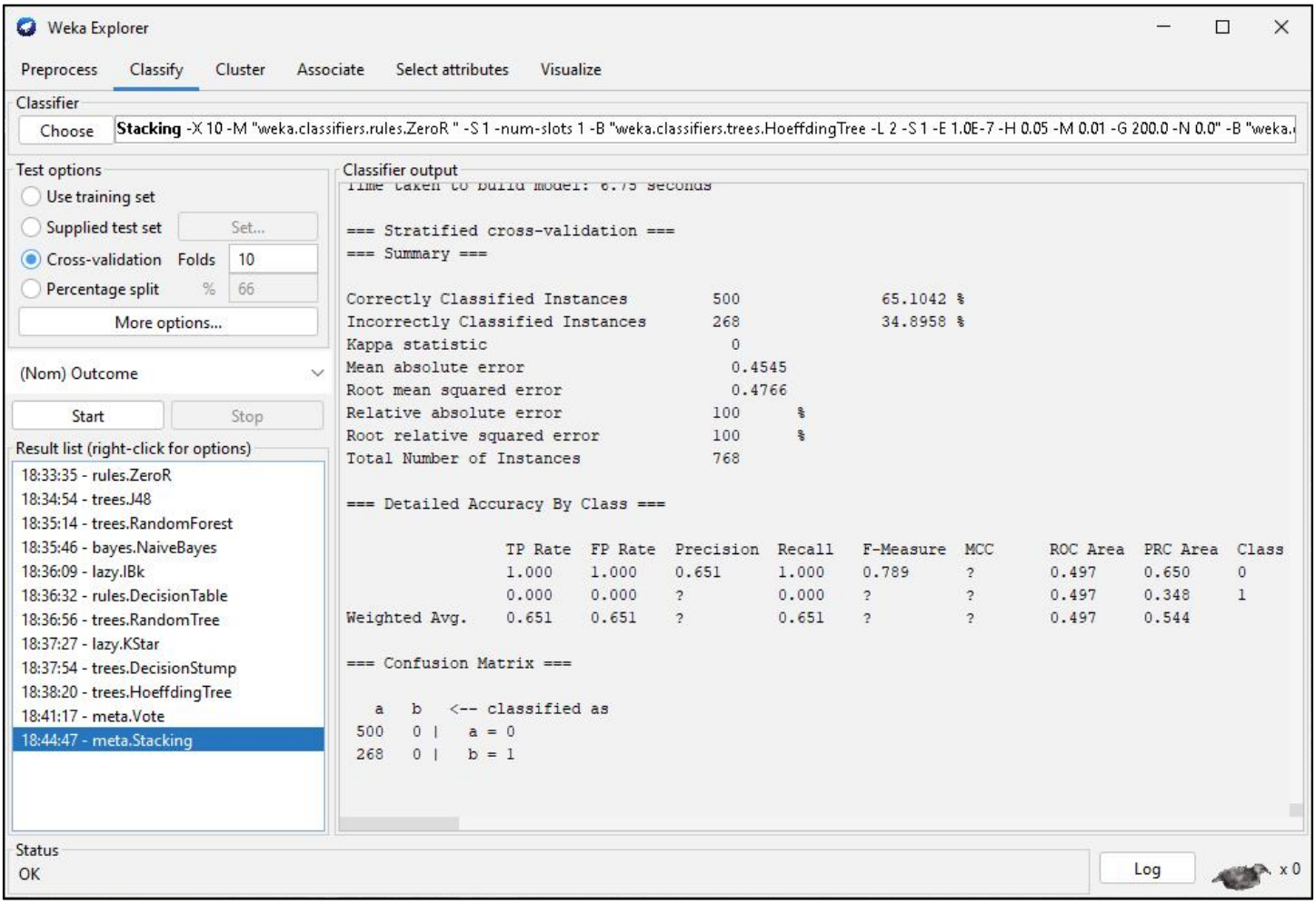


Step 8: Select the **"Stacking"** model for majority voting. Click on the model then click on Classifiers to select 10 different classifier models.





Step 9: Click Start to build the model. Classifier output shows the model evaluation parameters.



From the results we can summarize that ensemble models can have lower mean absolute error and root mean squared error as compared to other models such as Decision Table, Linear Regression, and Random Forest etc.

It means both “Majority Voting” and “Stacking” models have good ability to accurately predict data as compared to other models.

Learning outcomes (What I have learnt):

1. I learnt about the WEKA Tool and its applications.
2. I learnt about how to use Explorer Tab in WEKA Tool.
3. I learnt about how to change attributes from Numeric to Nominal.
4. I learnt about how to use the Vote and Stacking method in WEKA Tool.
5. I learnt about how to compare the accuracy of different models.