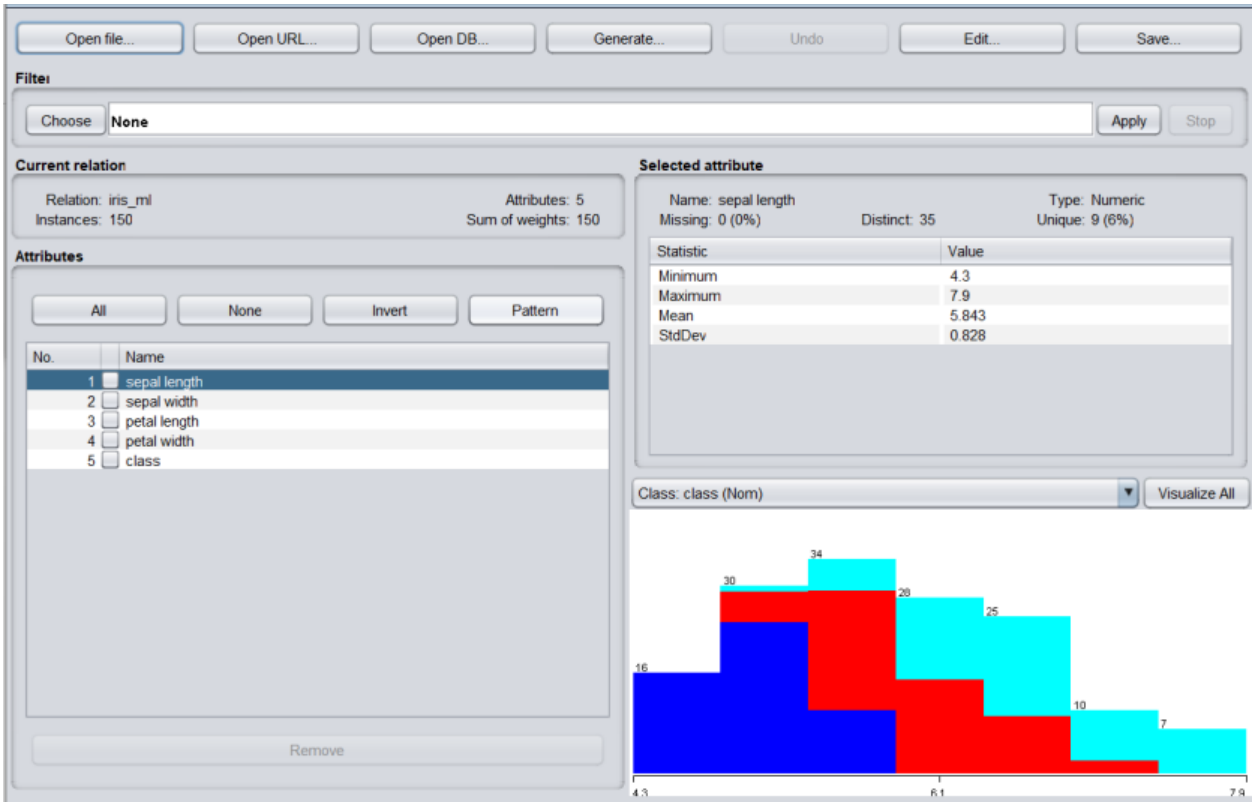
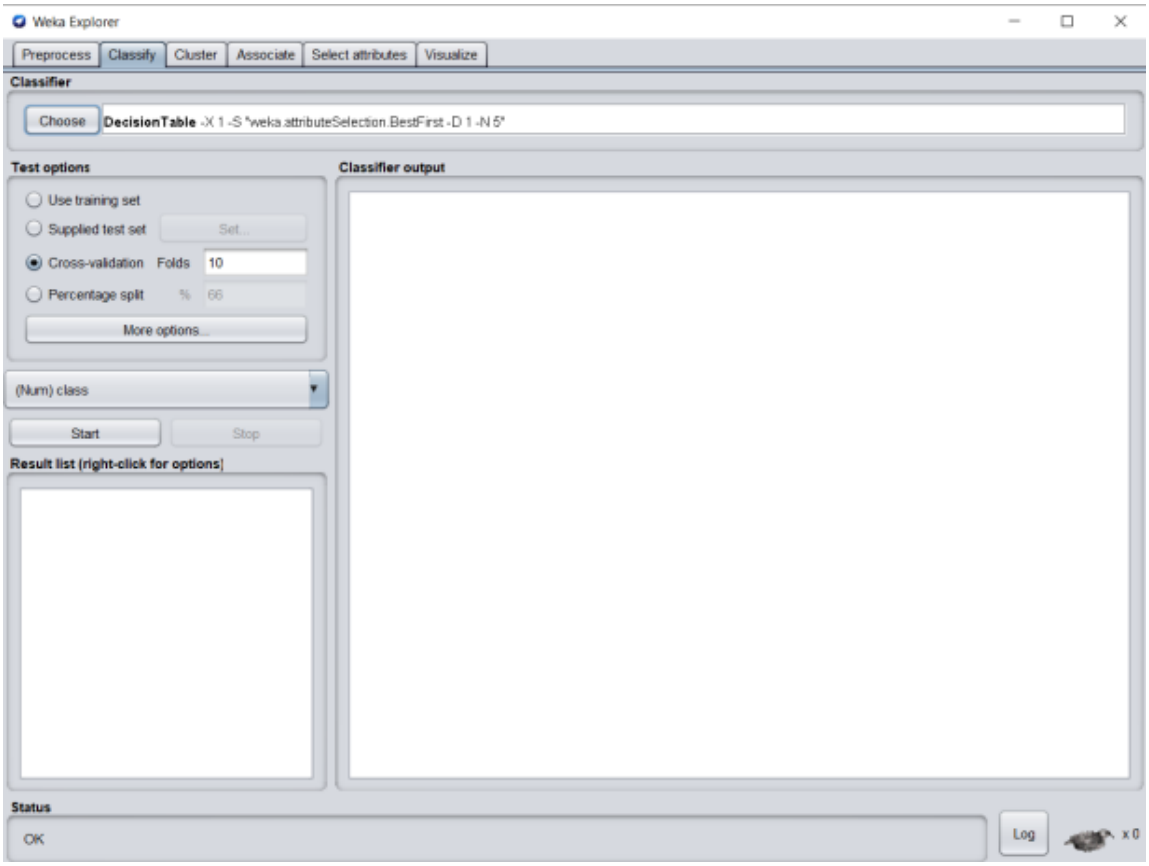
ILLUSTRATION OF 3 ALGORITHMS USING WEKA TOOL

DATASET: IRIS DATASET



CLASSIFICATION (DECISION TABLE)



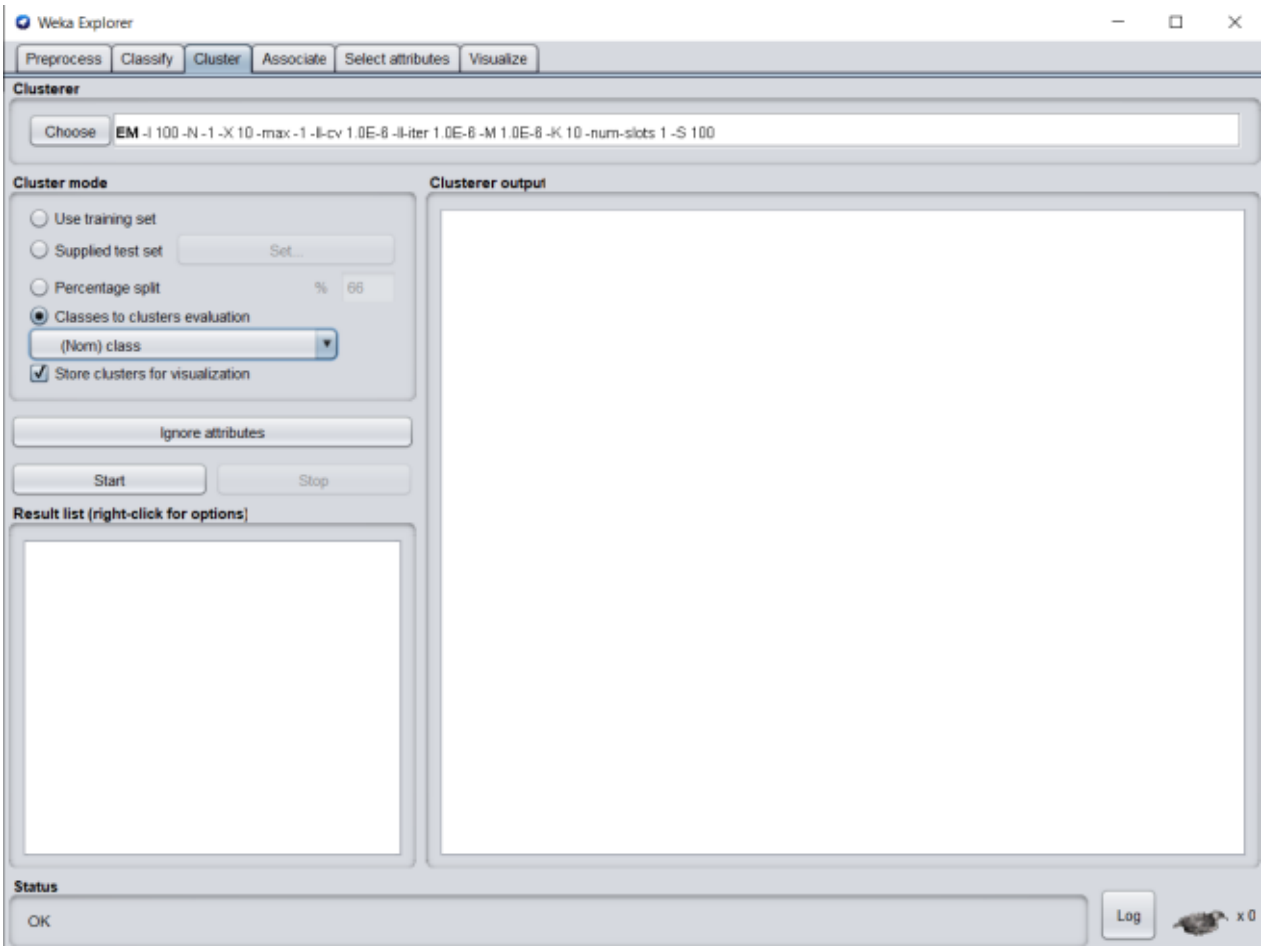
Classifier output

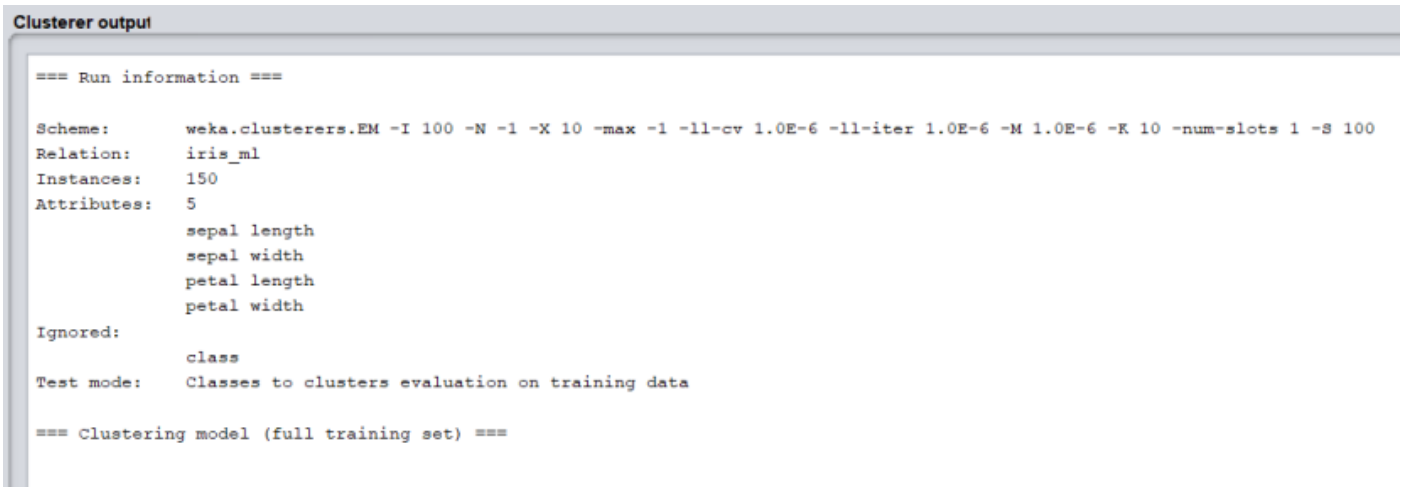
```
=== Run information ===

Scheme:       weka.classifiers.rules.DecisionTable -X 1 -S "weka.attributeSelection.BestFirst -D 1 -N 5"
Relation:     iris_ml
Instances:    150
Attributes:   5
              sepal length
              sepal width
              petal length
              petal width
              class
Test mode:    10-fold cross-validation

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

Decision Table:

Number of training instances: 150
Number of Rules : 3
Non matches covered by Majority class.
        Best first.
        Start set: no attributes
        Search direction: forward
        Stale search after 5 node expansions
        Total number of subsets evaluated: 12
        Merit of best subset found:   96
Evaluation (for feature selection): CV (leave one out)
Feature set: 4,5

Time taken to build model: 0.1 seconds
```

The Accuracy, Confusion Matrix and the terms below can be used as performance metrics

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

Correctly Classified Instances         139               92.6667 %
Incorrectly Classified Instances        11                7.3333 %
Kappa statistic                          0.89
Mean absolute error                      0.092
Root mean squared error                  0.2087
Relative absolute error                 20.6978 %
Root relative squared error             44.2707 %
Total Number of Instances              150

=== 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    1.000     1.000     Iris-setosa
                 0.880     0.050     0.898       0.880   0.889      0.834    0.946     0.861     Iris-versicolor
                 0.900     0.060     0.882       0.900   0.891      0.836    0.947     0.869     Iris-virginica
Weighted Avg.    0.927     0.037     0.927       0.927   0.927      0.890    0.964     0.910

=== Confusion Matrix ===

  a  b  c   <-- classified as
 50  0  0 |  a = Iris-setosa
  0 44  6 |  b = Iris-versicolor
  0  5 45 |  c = Iris-virginica
```

CLUSTERING (EM)





Clusterer output

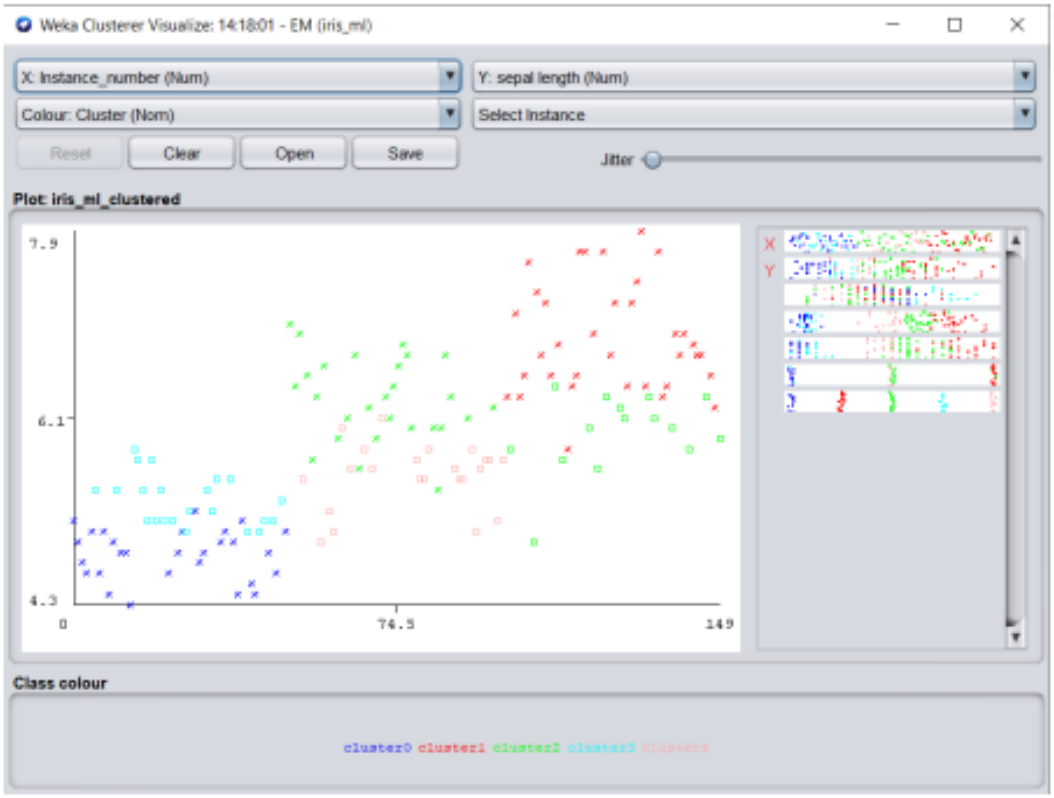
```
=== Run information ===

Scheme:       weka.clusterers.EM -I 100 -N -1 -X 10 -max -1 -ll-cv 1.0E-6 -ll-iter 1.0E-6 -M 1.0E-6 -K 10 -num-slots 1 -S 100
Relation:     iris_ml
Instances:    150
Attributes:   5
              sepal length
              sepal width
              petal length
              petal width
Ignored:
              class
Test mode:    Classes to clusters evaluation on training data

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

```
Number of clusters selected by cross validation: 5
Number of iterations performed: 16


               Cluster
Attribute            0       1       2       3       4
                (0.18)  (0.23)  (0.28)  (0.15)  (0.15)
=========================================================
sepal length
  mean           4.7748  6.8585  6.1613  5.2823  5.5432
  std. dev.      0.2405  0.5228  0.4138  0.2407  0.3159

sepal width
  mean           3.1789  3.0862  2.8547  3.7037  2.5786
  std. dev.      0.2599  0.2891  0.2687  0.2857  0.2512

petal length
  mean           1.4194  5.7859  4.7484  1.5173   3.863
  std. dev.      0.1692  0.4745  0.3193  0.1592  0.3516

petal width
  mean           0.1948  2.1327  1.5757  0.3028  1.1696
  std. dev.      0.0557  0.2359  0.2196  0.1212  0.1351



Time taken to build model (full training data) : 1.02 seconds
```

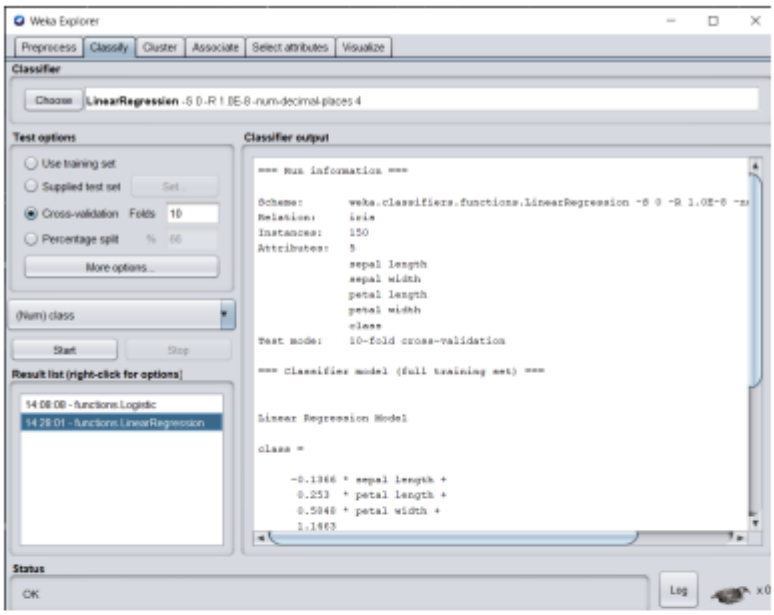
```
=== Model and evaluation on training set ===

Clustered Instances

0       28 ( 19%)
1       35 ( 23%)
2       42 ( 28%)
3       22 ( 15%)
4       23 ( 15%)


Log likelihood: -1.60803
```

Log likelihood can be used as a performance metric

```
Class attribute: class
Classes to Clusters:

  0  1  2  3  4  <-- assigned to cluster
 28  0  0 22  0 | Iris-setosa
  0  0 27  0 23 | Iris-versicolor
  0 35 15  0  0 | Iris-virginica

Cluster 0 <-- Iris-setosa
Cluster 1 <-- Iris-virginica
Cluster 2 <-- Iris-versicolor
Cluster 3 <-- No class
Cluster 4 <-- No class

Incorrectly clustered instances :       60.0     40       %
```



REG RESSION IN LINEAR




```
=== Run information ===

Scheme:       weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4
Relation:     iris
Instances:    150
Attributes:   5
              sepal length
              sepal width
              petal length
              petal width
              class
Test mode:    10-fold cross-validation

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

```
Linear Regression Model

class =

     -0.1366 * sepal length +
      0.253  * petal length +
      0.5848 * petal width +
      1.1463

Time taken to build model: 0.19 seconds
```

Root mean squared error and the other terms below can be used as performance metrics

```
=== Cross-validation ===
=== Summary ===

Correlation coefficient                  0.9611
Mean absolute error                      0.1733
Root mean squared error                  0.2254
Relative absolute error                 25.6225 %
Root relative squared error             27.4199 %
Total Number of Instances              150
```