**Data Warehousing & Data Mining**

**Vehicle Silhouettes**

**Description:**

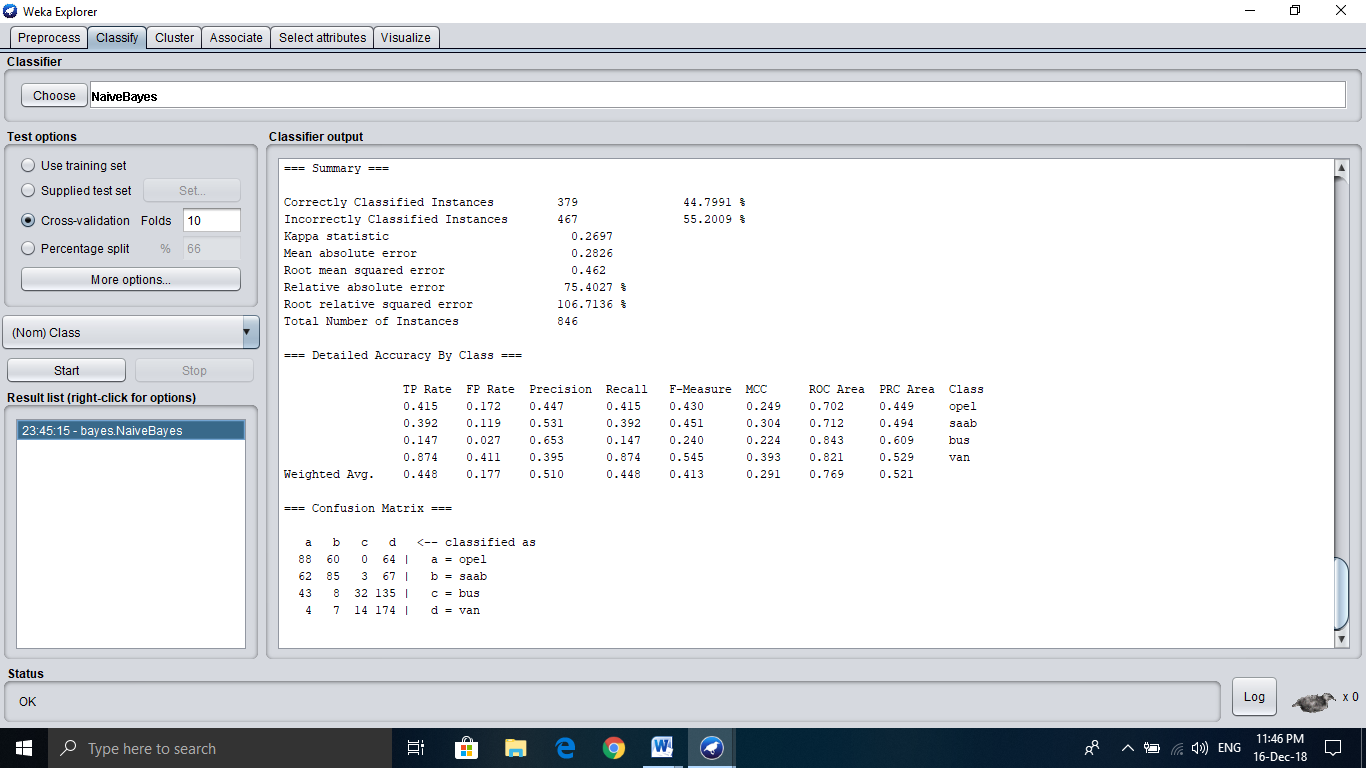
This data file contains details of four types of vehicle and its silhouettes. With this data we can try things like predicting the vehicle may be viewed from one of many different angles.

The dataset consists of 946 instances, 18 attributes and 4 classes. 100 instances are being kept by Strathclyde for validation. So Statlog partners will receive 846 examples.

**Methods:**

* Naive Bayes
* IBK (KNN)
* J48 (Entropy)
* BFTree (Gini Index)
* Random Forrest

**Naïve Bayes:**



**=== Summary ===**

Correctly Classified Instances 379 44.7991 %

Incorrectly Classified Instances 467 55.2009 %

Kappa statistic 0.2697

Mean absolute error 0.2826

Root mean squared error 0.462

Relative absolute error 75.4027 %

Root relative squared error 106.7136 %

Total Number of Instances 846

**=== Confusion Matrix ===**

a b c d <-- classified as

88 60 0 64 | a = opel

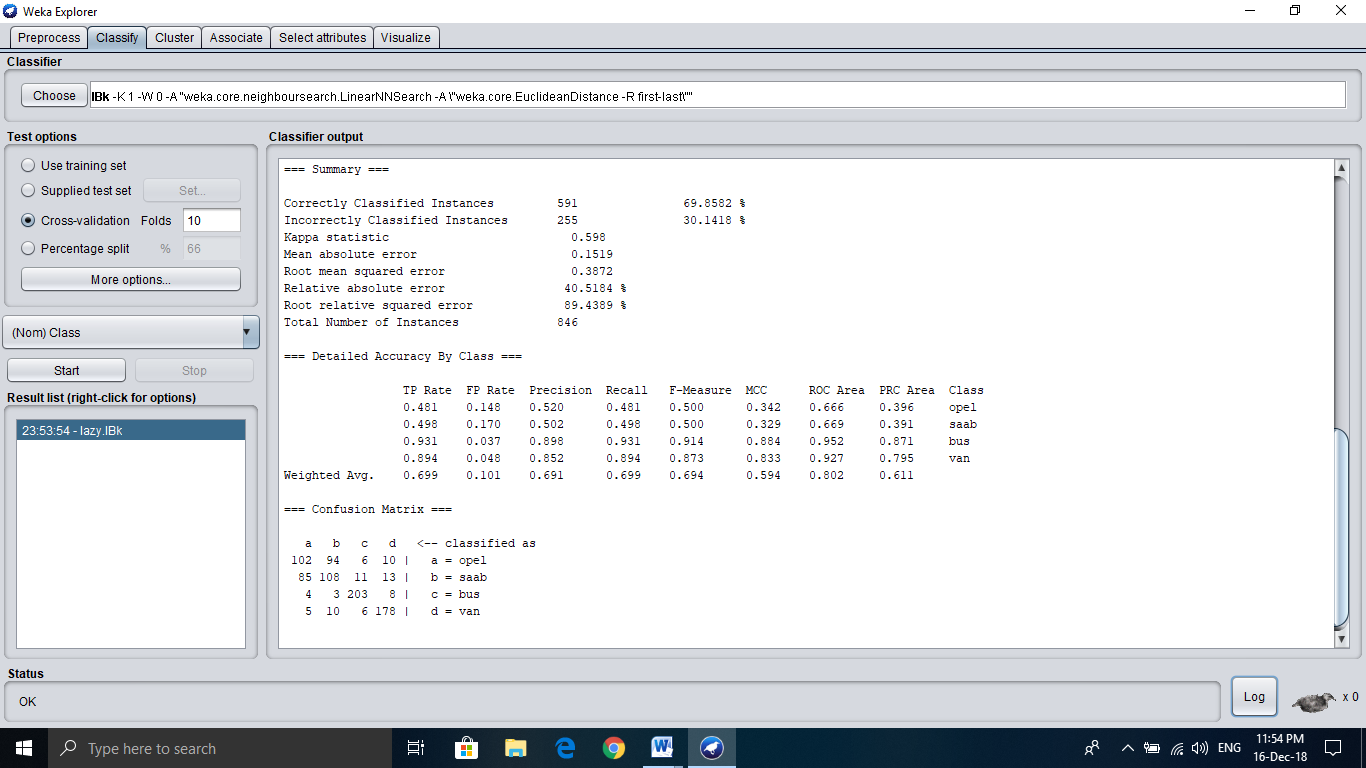
62 85 3 67 | b = saab

43 8 32 135 | c = bus

4 7 14 174 | d = van

Here the weighted average of true positive and false positive are 0.448 and 0.177.

**IBK (KNN):**



**=== Summary ===**

Correctly Classified Instances 591 69.8582 %

Incorrectly Classified Instances 255 30.1418 %

Kappa statistic 0.598

Mean absolute error 0.1519

Root mean squared error 0.3872

Relative absolute error 40.5184 %

Root relative squared error 89.4389 %

Total Number of Instances 846

**=== Confusion Matrix ===**

a b c d <-- classified as

102 94 6 10 | a = opel

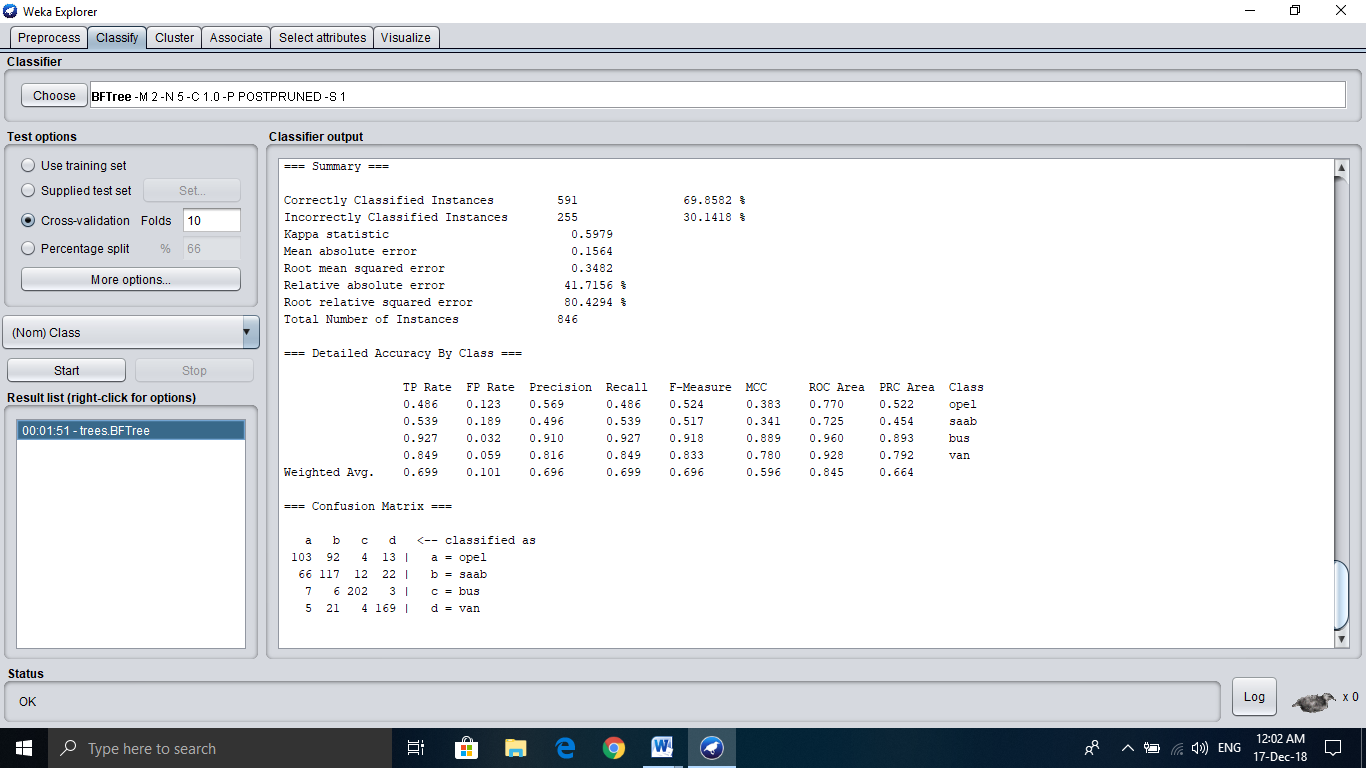
85 108 11 13 | b = saab

4 3 203 8 | c = bus

5 10 6 178 | d = van

Here the weighted average of true positive and false positive are 0.699 and 0.101.

**BFTree (Gini Index):**



**=== Summary ===**

Correctly Classified Instances 591 69.8582 %

Incorrectly Classified Instances 255 30.1418 %

Kappa statistic 0.5979

Mean absolute error 0.1564

Root mean squared error 0.3482

Relative absolute error 41.7156 %

Root relative squared error 80.4294 %

Total Number of Instances 846

**=== Confusion Matrix ===**

a b c d <-- classified as

103 92 4 13 | a = opel

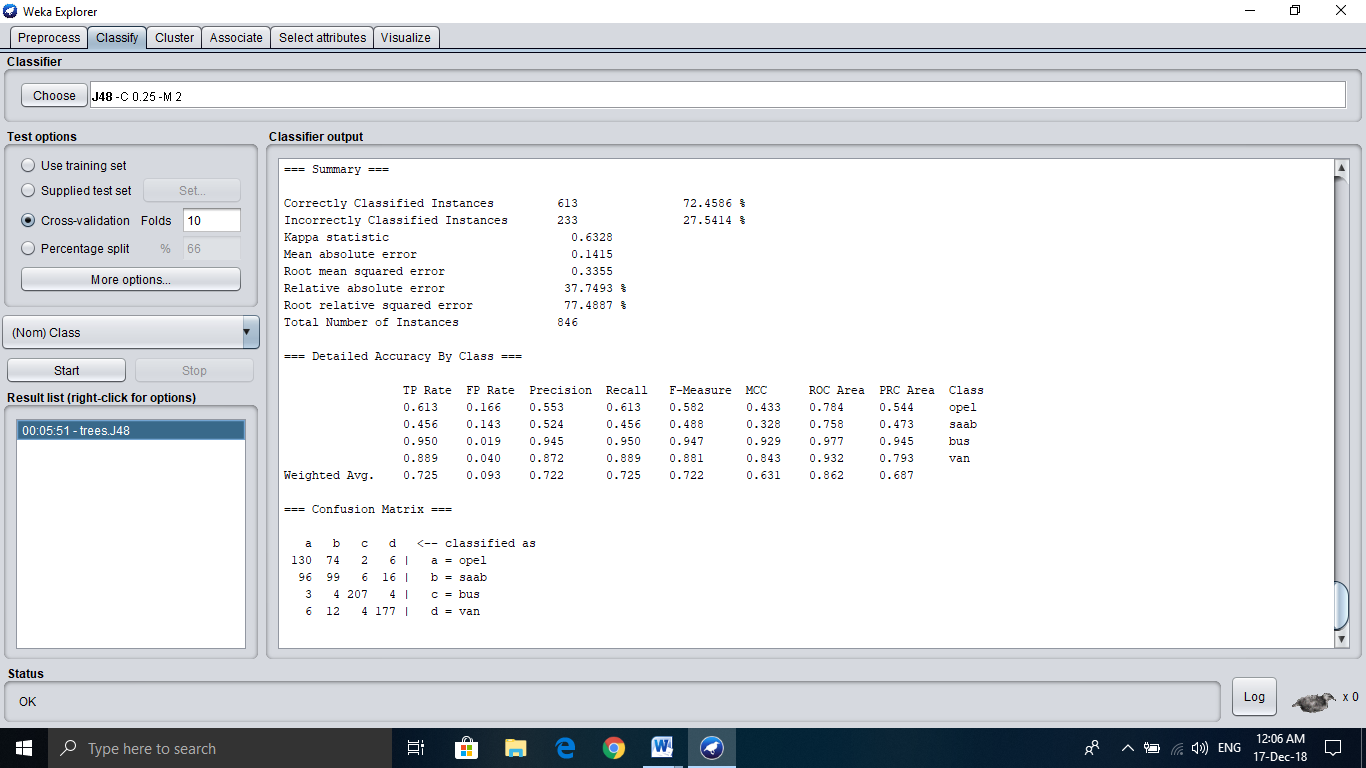
66 117 12 22 | b = saab

7 6 202 3 | c = bus

5 21 4 169 | d = van

Here the weighted average of true positive and false positive are 0.699 and 0.101

**J48 (Entropy):**



**=== Summary ===**

Correctly Classified Instances 613 72.4586 %

Incorrectly Classified Instances 233 27.5414 %

Kappa statistic 0.6328

Mean absolute error 0.1415

Root mean squared error 0.3355

Relative absolute error 37.7493 %

Root relative squared error 77.4887 %

Total Number of Instances 846

**=== Confusion Matrix ===**

a b c d <-- classified as

130 74 2 6 | a = opel

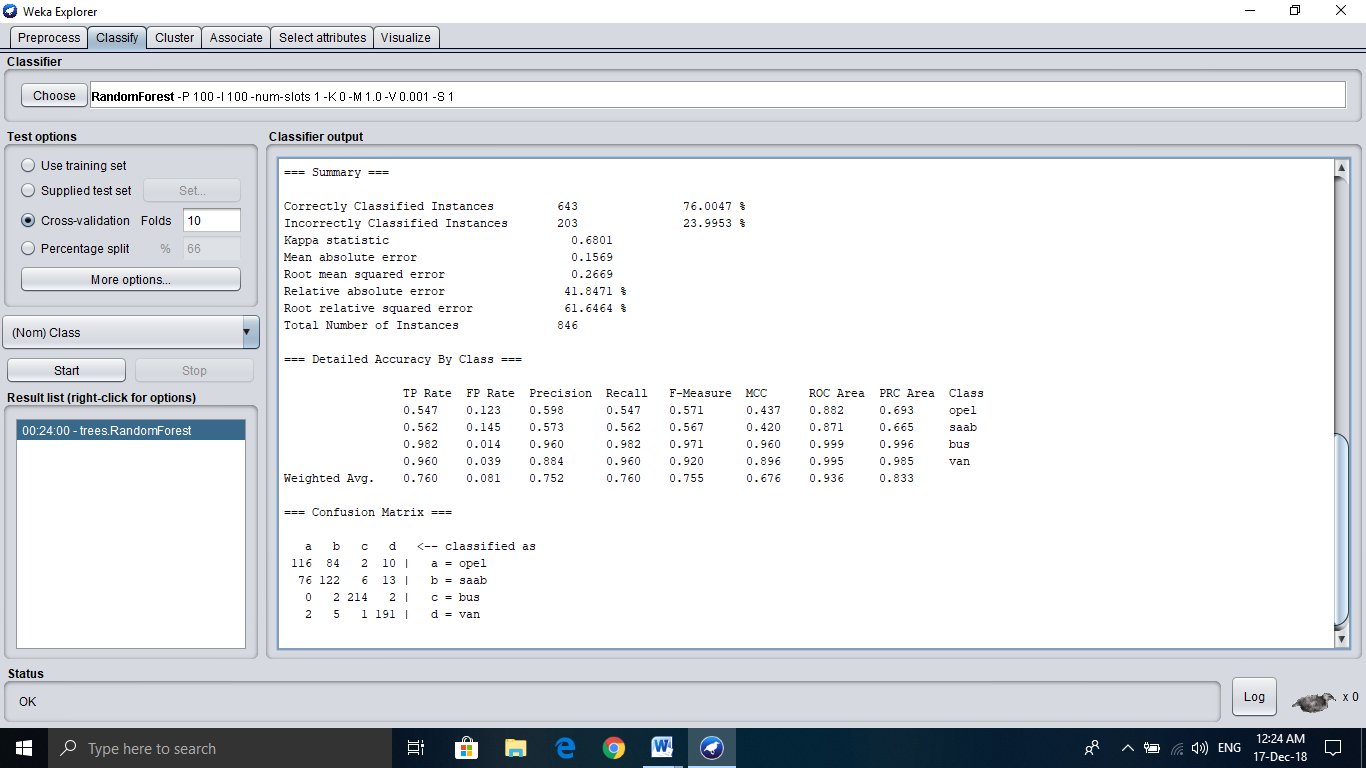
96 99 6 16 | b = saab

3 4 207 4 | c = bus

6 12 4 177 | d = van

Here the weighted average of true positive and false positive are 0.725 and 0.093

**Random Forest:**



**=== Summary ===**

Correctly Classified Instances 643 76.0047 %

Incorrectly Classified Instances 203 23.9953 %

Kappa statistic 0.6801

Mean absolute error 0.1569

Root mean squared error 0.2669

Relative absolute error 41.8471 %

Root relative squared error 61.6464 %

Total Number of Instances 846

**=== Confusion Matrix ===**

a b c d <-- classified as

116 84 2 10 | a = opel

76 122 6 13 | b = saab

0 2 214 2 | c = bus

2 5 1 191 | d = van

Here the weighted average of true positive and false positive are 0.760 and 0.081

**ROC Graph:**

**Comment:**

I have performed five different random algorithm or classifier for this dataset (Vehicle Silhouettes). To choose a best classifier, we take Best value. Here, RandomForest is very close to Best. RandomForest gives highest TPR 0.76 and the lowest FPR 0.081.

So, for the prediction to viewing Vehicle Silhouettes in different angle I choose RandomForest as the Best classifier.