

Classification and Testing)

Explanations:

The feature selection and classification system used was the same as question 3 however there was not multiple sheets the data was randomised by row, and split into 3 sections, training, Test1, Test2 it was then saved to individual csv files to save for Weka use. The classification model was creating using the training set and then tested on all 3. The details of the code and functions used are in question 3. The following code shuffles the rows within the Iris dataset.

```
split_data <- function(datafile)
{
  set.seed(11)
  datafile <- datafile[sample(1:nrow(datafile)),]
  datafile <- datafile[sample(1:nrow(datafile)),]
  write.csv(datafile[1:50,],"Results/TrainingSet.csv")
  write.csv(datafile[51:100,],"Results/TestSet1")
  write.csv(datafile[101:150,],"Results/TestSet2.csv")
  |
}
```

Features Selected:

```
Boruta performed 9 iterations in 0.1596181 secs.
5 attributes confirmed important: CLASS, PetalLengthCm, PetalWidthCm, SepalLengthCm, SepalWidthCm;
No attributes deemed unimportant.
      x
1      CLASS
2  SepalWidthCm
3 PetalLengthCm
4  PetalWidthCm
5  SepalLengthCm
```

Boruta determined that all the features within the dataset are important for determining the CLASS a flower belongs to, so the original dataset will be used for classification.

However Wekas classification makes use of only PetalWidth, and Pedal Length, so the features selected can be narrowed down to PetalWidth and Pedal Length.

Classification:

Weka outputs the classification system below, with the pruned tree being the actual system that will be used to test on the different datasets.

J48 pruned tree

```
PetalWidthCm <= 0.4: Iris-setosa (15.0)
PetalWidthCm > 0.4
|   PetalLengthCm <= 4.7: Iris-versicolor (15.0)
|   PetalLengthCm > 4.7: Iris-virginica (20.0/3.0)
```

Number of Leaves : 3

Size of the tree : 5

Time taken to build model: 0 seconds

Testing:

Training Set:

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances	47	94	%
Incorrectly Classified Instances	3	6	%
Kappa statistic	0.9099		
Mean absolute error	0.068		
Root mean squared error	0.1844		
Relative absolute error	15.3405 %		
Root relative squared error	39.1699 %		
Total Number of Instances	50		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.833	0.000	1.000	0.833	0.909	0.873	0.956	0.919	Iris-versicolor
	1.000	0.091	0.850	1.000	0.919	0.879	0.955	0.850	Iris-virginica
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Iris-setosa
Weighted Avg.	0.940	0.031	0.949	0.940	0.940	0.913	0.969	0.920	

=== Confusion Matrix ===

```
a b c <-- classified as
15 3 0 | a = Iris-versicolor
0 17 0 | b = Iris-virginica
0 0 15 | c = Iris-setosa
```

Test Set 1:

=== Evaluation on test set ===

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

=== Summary ===

Correctly Classified Instances	47	94	%
Incorrectly Classified Instances	3	6	%
Kappa statistic	0.9092		
Mean absolute error	0.072		
Root mean squared error	0.2035		
Relative absolute error	16.1239 %		
Root relative squared error	42.9074 %		
Total Number of Instances	50		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.923	0.054	0.857	0.923	0.889	0.849	0.953	0.822	Iris-versicolor
	0.944	0.031	0.944	0.944	0.944	0.913	0.957	0.912	Iris-virginica
	0.947	0.000	1.000	0.947	0.973	0.958	0.974	0.967	Iris-setosa
Weighted Avg.	0.940	0.025	0.943	0.940	0.941	0.914	0.962	0.910	

=== Confusion Matrix ===

```
a b c <-- classified as
12 1 0 | a = Iris-versicolor
1 17 0 | b = Iris-virginica
1 0 18 | c = Iris-setosa
```

Test Set 2:

=== Evaluation on test set ===

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

=== Summary ===

Correctly Classified Instances	47	94	%
Incorrectly Classified Instances	3	6	%
Kappa statistic	0.9098		
Mean absolute error	0.066		
Root mean squared error	0.1926		
Relative absolute error	14.8809	%	
Root relative squared error	40.8934	%	
Total Number of Instances	50		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.895	0.032	0.944	0.895	0.919	0.872	0.957	0.902	Iris-versicolor
	1.000	0.057	0.882	1.000	0.938	0.912	0.971	0.882	Iris-virginica
	0.938	0.000	1.000	0.938	0.968	0.954	0.969	0.958	Iris-setosa
Weighted Avg.	0.940	0.029	0.944	0.940	0.940	0.910	0.965	0.914	

=== Confusion Matrix ===

```
a b c <-- classified as
17 2 0 | a = Iris-versicolor
 0 15 0 | b = Iris-virginica
 1 0 15 | c = Iris-setosa
```

Results:

	Training	Test AD	Test MCI
Sensitivity	0.94	0.94	0.94
Specificity	0.97	0.98	0.97
Accuracy	94%	94%	94%
F1-Score	0.94	0.94	0.94
MCC	0.913	0.914	0.910
Youden's J	0.91	0.92	0.91

Evaluation:

The classification system created by the training set was proven to be highly effective when tested on the test sets.