Problem 1

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     iris
Instances:    150
Attributes:    5
               sepal_length
               sepal_width
               petal_length
               petal_width
               species
Test mode:    10-fold cross-validation

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

J48 pruned tree
------------------

petal_width <= 0.6: setosa (50.0)
petal_width > 0.6
|   petal_width <= 1.7
|   |   petal_length <= 4.9: versicolor (48.0/1.0)
|   |   petal_length > 4.9
|   |   |   petal_width <= 1.5: virginica (3.0)
|   |   |   petal_width > 1.5: versicolor (3.0/1.0)
|   petal_width > 1.7: virginica (46.0/1.0)

Number of Leaves  :     5

Size of the tree :   9


Time taken to build model: 0.02 seconds

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

Correctly Classified Instances        144                96        %
Incorrectly Classified Instances        6                 4        %
Kappa statistic                        0.94
Mean absolute error                    0.035
Root mean squared error                0.1586
Relative absolute error                7.8705 %
Root relative squared error           33.6353 %
Total Number of Instances              150

=== Detailed Accuracy By Class ===

                  TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                  0.980    0.000    1.000      0.980   0.990      0.985    0.990     0.987     setosa
                  0.940    0.030    0.940      0.940   0.940      0.910    0.952     0.880     versicolor
                  0.960    0.030    0.941      0.960   0.950      0.925    0.961     0.905     virginica
Weighted Avg.     0.960    0.020    0.960      0.960   0.960      0.940    0.968     0.924

=== Confusion Matrix ===

 a  b  c   <-- classified as
49  1  0 |  a = setosa
 0 47  3 |  b = versicolor
 0  2 48 |  c = virginica
```
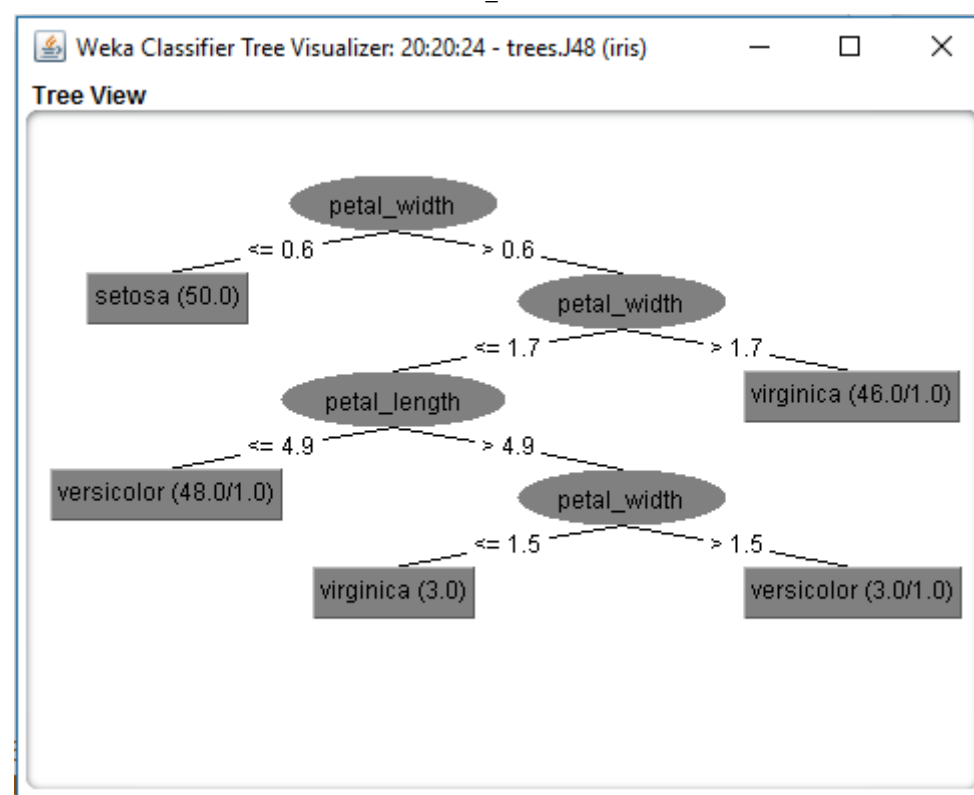
1. How many misclassified instances exist when 10-fold cross validation is used?

   6

2. What is the percentage of Accuracy?

   96%

3. How many instances are classified under setosa, virginica, versicolor?

   49, 51, 50

4. For petal width <= 0.6, what is the class label under which the instances are classified?

   setosa

5. For petal width <= 1.7, how many instances are classified under versicolor?

   51

6. For petal width > 1.7, how many instances are classified under virginica?

   46

Problem 2

1. Choose the breast-cancer dataset from Weka viz. breast-cancer.arff and perform decision tree learning using Random Forest classifier
2. In Classifier Output on the right, it is mentioned that Correctly Classified Instances a.k.a Accuracy is 69.58%. The fraction x/y = 69.58%. What are the values of x and y in this case that led to 69.58%

   x = 199, y = 286

3. Set the number of iterations options numIterations to 50, 100, 150, 200, 250. List Accuracy for each of these options. What is the trend that is observed in accuracy values.

   68.88%; 69.58%; 70.63%; 70.28%; 68.58%

4. Perform classification on breast-cancer dataset using J48. Which among Random Forest and J48 has better accuracy

   J48 has accuracy of 75.52%, which is better.

```
=== Run information ===

Scheme:          weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation:        breast-cancer
Instances:       286
Attributes:      10
                      age
                      menopause
                      tumor-size
                      inv-nodes
                      node-caps
                      deg-malig
                      breast
                      breast-quad
                      irradiat
                      Class
Test mode:       10-fold cross-validation

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

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.29 seconds

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

Correctly Classified Instances          199                    69.5804 %
Incorrectly Classified Instances         87                    30.4196 %
Kappa statistic                         0.1736
Mean absolute error                     0.3727
Root mean squared error                 0.4613
Relative absolute error                89.0857 %
Root relative squared error           100.9171 %
Total Number of Instances               286

=== Detailed Accuracy By Class ===

                     TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                     0.871    0.718    0.742      0.871   0.801      0.184    0.634     0.798     no-recurrence-
events
                     0.282    0.129    0.480      0.282   0.356      0.184    0.634     0.409     recurrence-events
Weighted Avg.        0.696    0.543    0.664      0.696   0.669      0.184    0.634     0.682

=== Confusion Matrix ===

   a    b    <-- classified as
 175   26 |   a = no-recurrence-events
  61   24 |   b = recurrence-events
```

Problem 3

1. What is the Weka command to construct a tree by not pruning it

J48 -U -M2

2. What is the Weka command command to prune a tree by allocating a minimum of 10 objects to a leaf node

J48 -C 0.25 -M10

3. What is percentage accuracy of decision tree if pruning is allowed and a minimum of 10 objects are to be allocated to a leaf node

74.82%