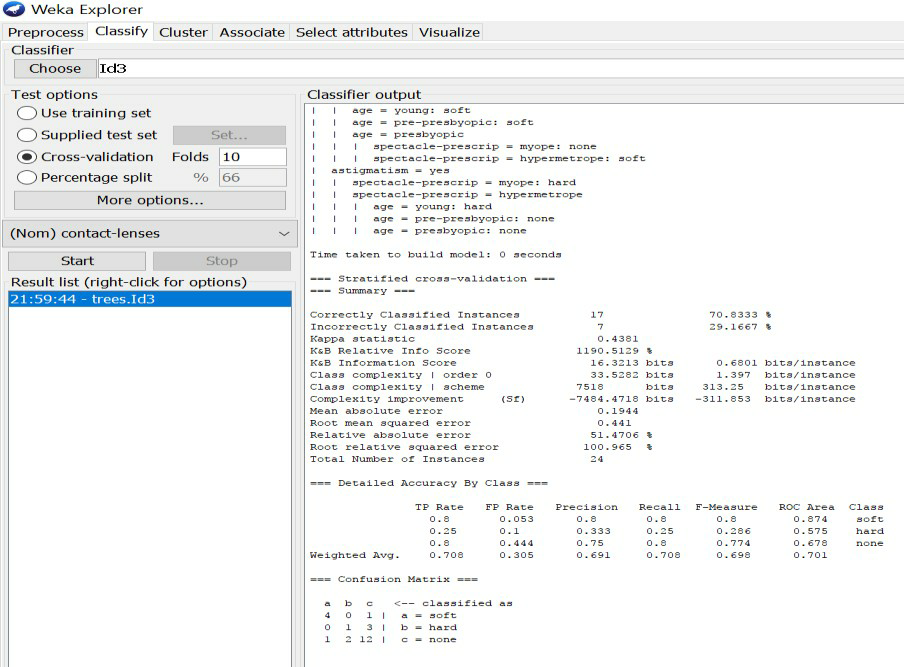
Experiment 3

**3. Demonstrate performing classification on data sets**

**A. Load each dataset into Weka and run Id3, J48 classification algorithm. Study the classifier output. Compute entropy values, Kappa statistic.**

**Procedure for Id3:**

1. Load the dataset (Contact-lenses.arff) into weka tool
2. Go to classify option & in left-hand navigation bar we can see different classification algorithms under tree section.
3. In which we selected Id3 algorithm, in more options select the output entropy evaluation measures& click on start option.
4. Then we will get classifier output, entropy values& Kappa Statistic as represented below.



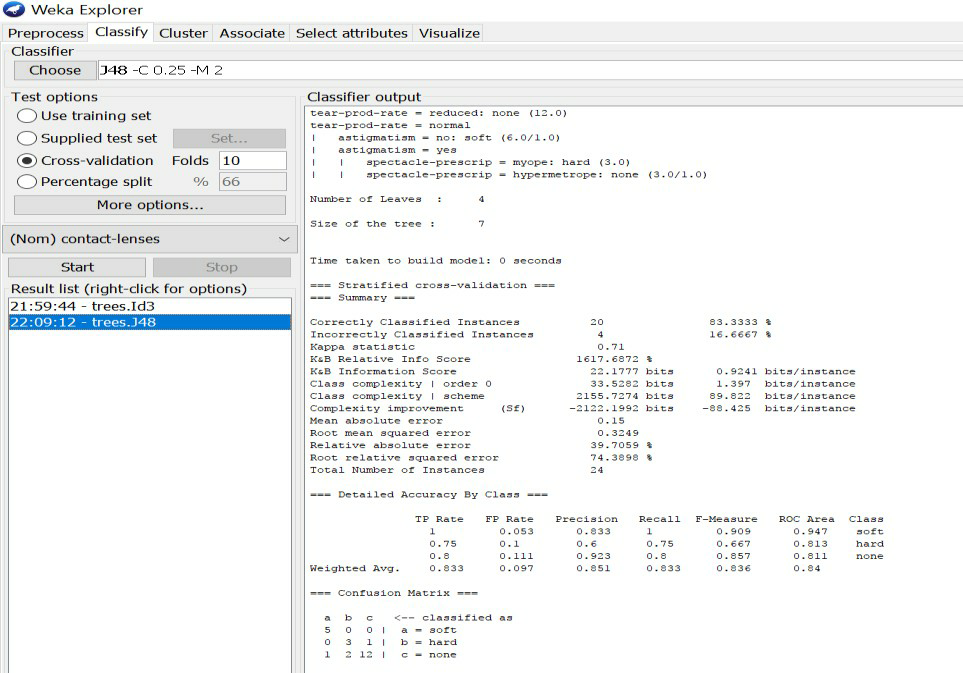
1. In the above screenshot, we can run classifiers with different test options (Cross-validation, Use Training Set, Percentage Split, Supplied Test set).

The result of applying the chosen classifier will be tested according to the optionsthat are set by clicking in the Test options box. There are four test modes:

* 1. **Use training set:** The classifier is evaluated on how well it predicts theclass of the instances it was trained on.
  2. **Supplied test set:** The classifier is evaluated on how well it predicts theclass of a set ofinstances loaded from a file. Clicking the Set... buttonbrings up a dialog allowing you to choose the file to test on.
  3. **Cross-validation:** The classifier is evaluated by cross-validation, usingthe number of folds that are entered in the Folds text field.
  4. **Percentage split**: The classifier is evaluated on how well it predicts acertain percentage of the data which is held out for testing. The amountof data held out depends on the value entered in the % field.

# Procedure for J48:

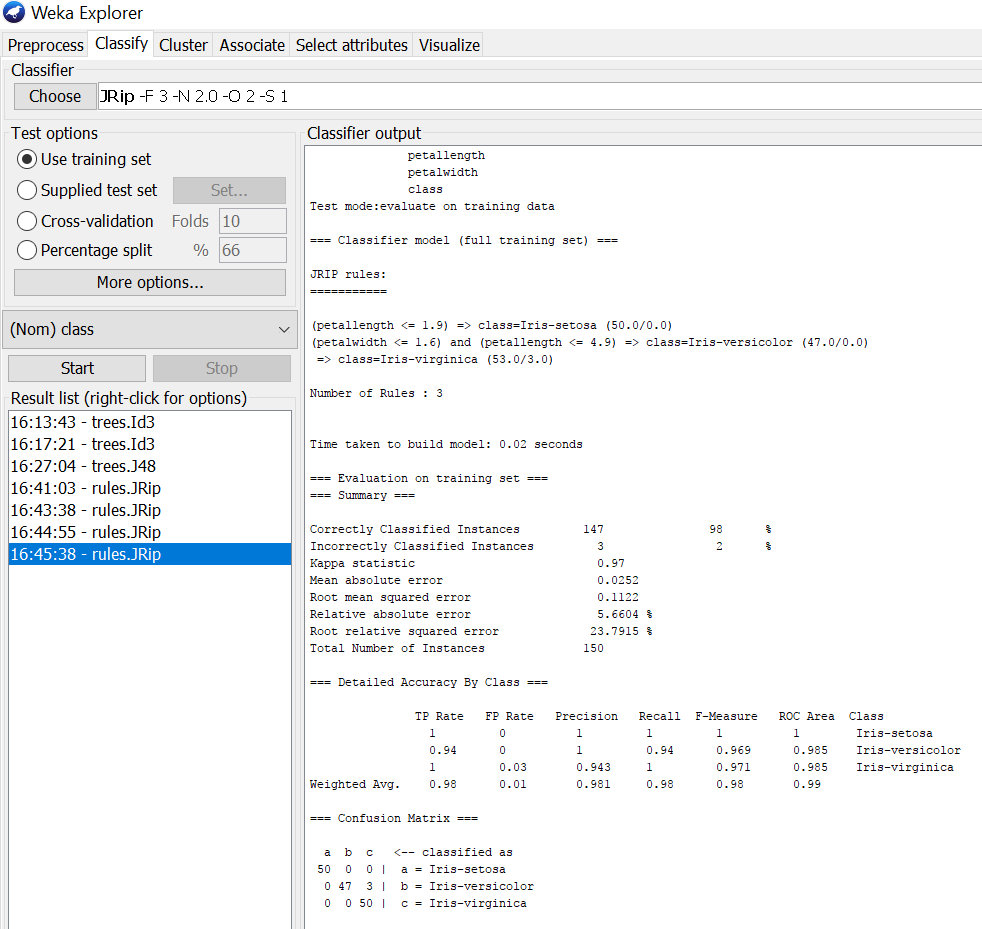
* + 1. Load the dataset (Contact-lenses.arff) into weka tool
    2. Go to classify option & in left-hand navigation bar we can see differentclassification algorithms under tree section.
    3. In which we selected J48 algorithm,in more options select the output entropy evaluation measures& click on start option.
    4. Then we will get classifier output, entropy values & Kappa Statistic as represented below.
    5. In the below screenshot, we can run classifiers with different test options (Cross-validation, Use Training Set, Percentage Split, Supplied Test set).



# Extract if-then rules from the decision tree generated by the classifier, Observethe confusion matrix and derive Accuracy, F-measure, TPrate, FPrate, Precision and Recall values. Apply cross-validation strategy with various fold levels and compare the accuracy results.

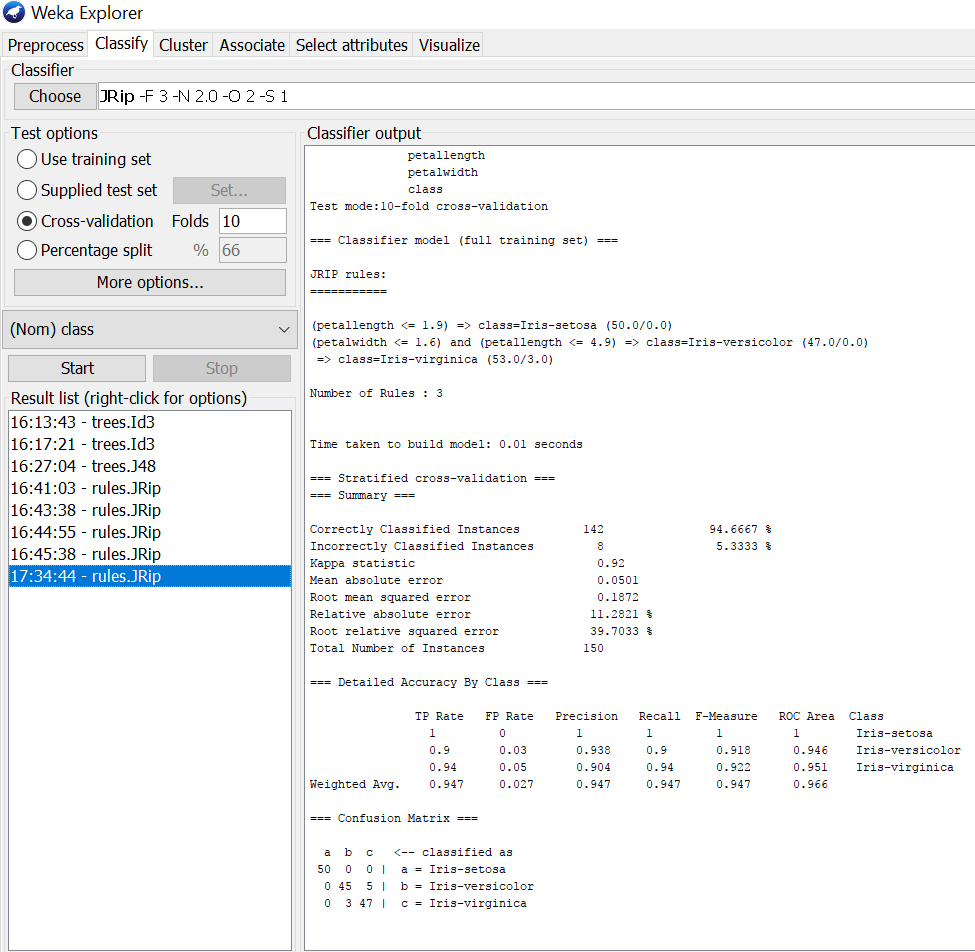
**Procedure:**

1. Load the dataset (Iris-2D. arff) into weka tool
2. Go to classify option & in left-hand navigation bar we can see differentclassification algorithms under rules section.
3. In which we selected JRip (If-then) algorithm & click on start option with ―use training set‖ test option enabled.
4. Then we will get detailed accuracy by class consists ofF-measure, TP rate, FP rate, Precision, Recall values& Confusion Matrix as represented below.



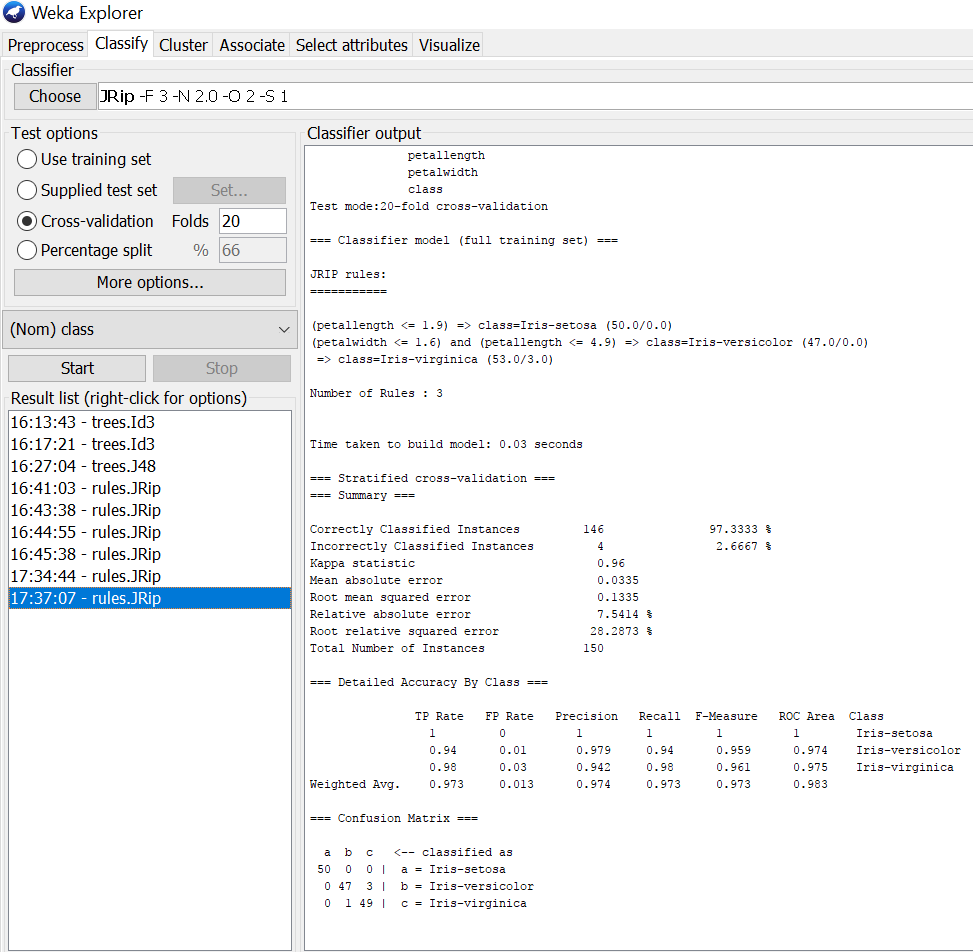
# Using Cross-Validation Strategy with 10 folds:

Here, we enabled cross-validation test option with 10 folds & clicked start button as represented below.



# Using Cross-Validation Strategy with 20 folds:

Here, we enabled cross-validation test option with 20 folds & clicked start button as represented below.

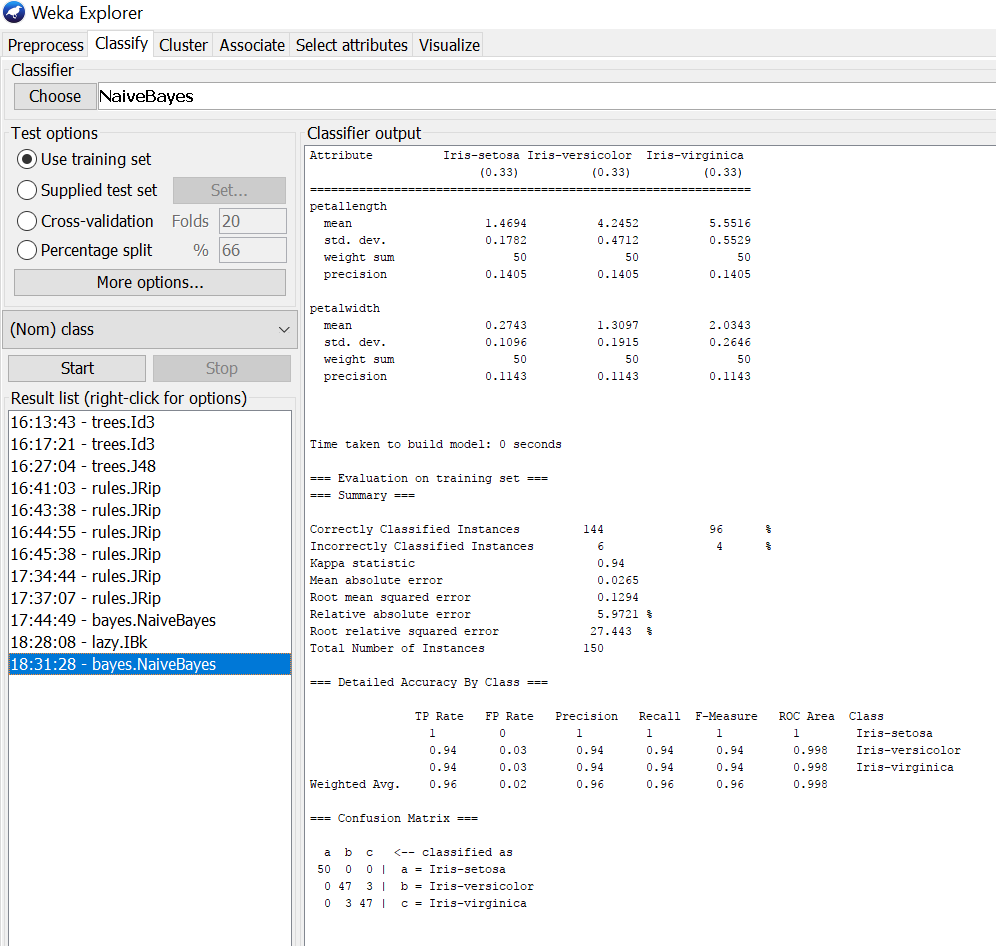


If we see the above results of cross validation with 10 folds & 20 folds. As per our observation the error rate is lesser with 20 folds got 97.3% correctness when compared to 10 folds got 94.6% correctness.

# Load each dataset into Weka and perform Naive-bayes classification and k-Nearest Neighbour classification. Interpret the results obtained.

**Procedure for Naïve-Bayes:**

1. Load the dataset (Iris-2D. arff) into weka tool
2. Go to classify option & in left-hand navigation bar we can see differentclassification algorithms under bayes section.
3. In which we selected Naïve-Bayes algorithm & click on start option with ―use training set‖ test option enabled.
4. Then we will get detailed accuracy by class consists of F-measure, TP rate, FP rate, Precision, Recall values& Confusion Matrix as represented below.



# Procedure for K-Nearest Neighbour (IBK):

1. Load the dataset (Iris-2D. arff) into weka tool
2. Go to classify option & in left-hand navigation bar we can see differentclassification algorithms under lazy section.
3. In which we selected K-Nearest Neighbour (IBK) algorithm & click on start option with ―use training set‖ test option enabled.
4. Then we will get detailed accuracy by class consists of F-measure, TP rate, FP rate, Precision, Recall values& Confusion Matrix as represented below.

