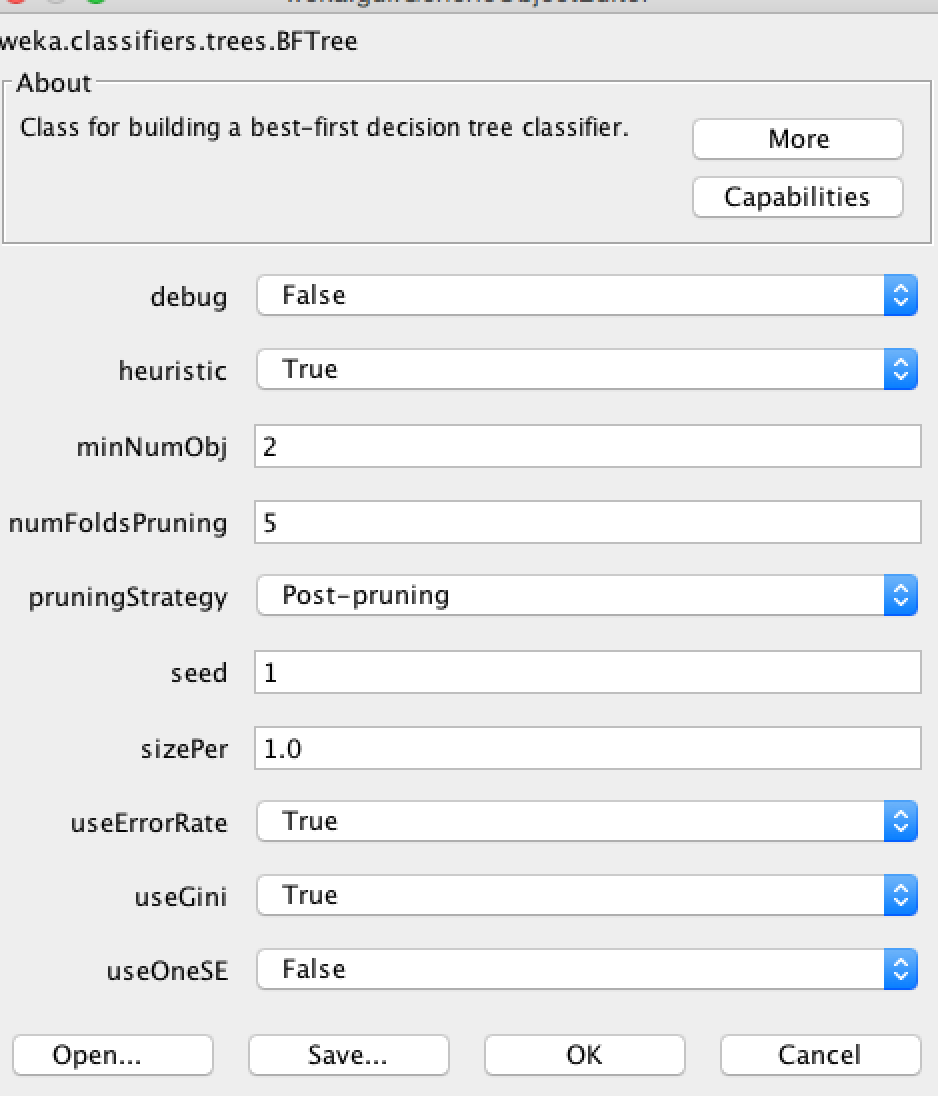
1. Preparation

In this assignment, we were using Weka (<http://www.cs.waikato.ac.nz/ml/weka/downloading.html>) to analyze our two data sets. The data sets are Iris and Congressional Voting Records. Our goal is to build models with different methods and different parameters in the methods. Then, we could obtain the best classification accuracy.

1. Data Process and Accuracy Comparison

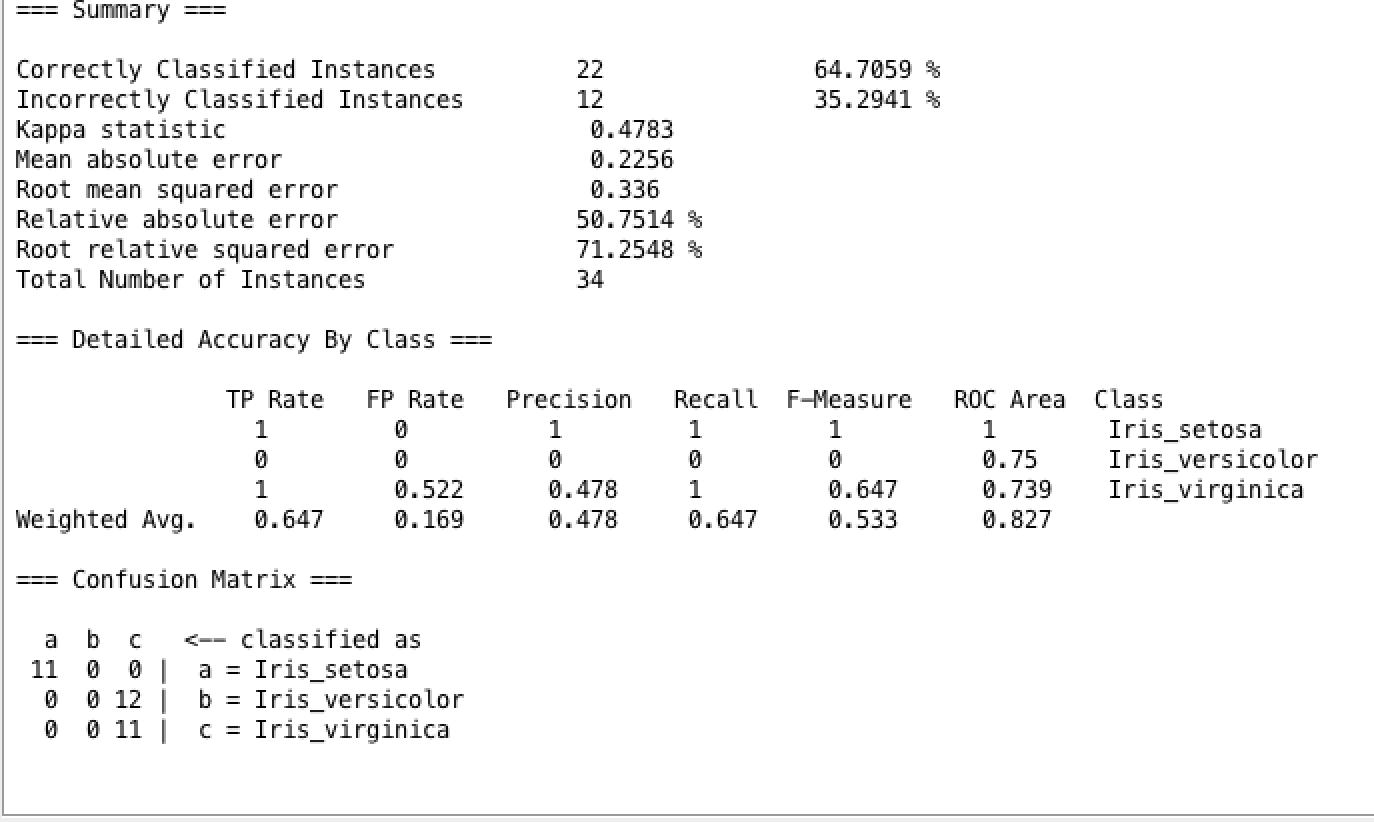
**For Iris**

First, we open a file and get the Iris data. Then we do the classification of Iris. Here we choose BFTree, decisionstump, ridor, decision table, simple cart, and zero.

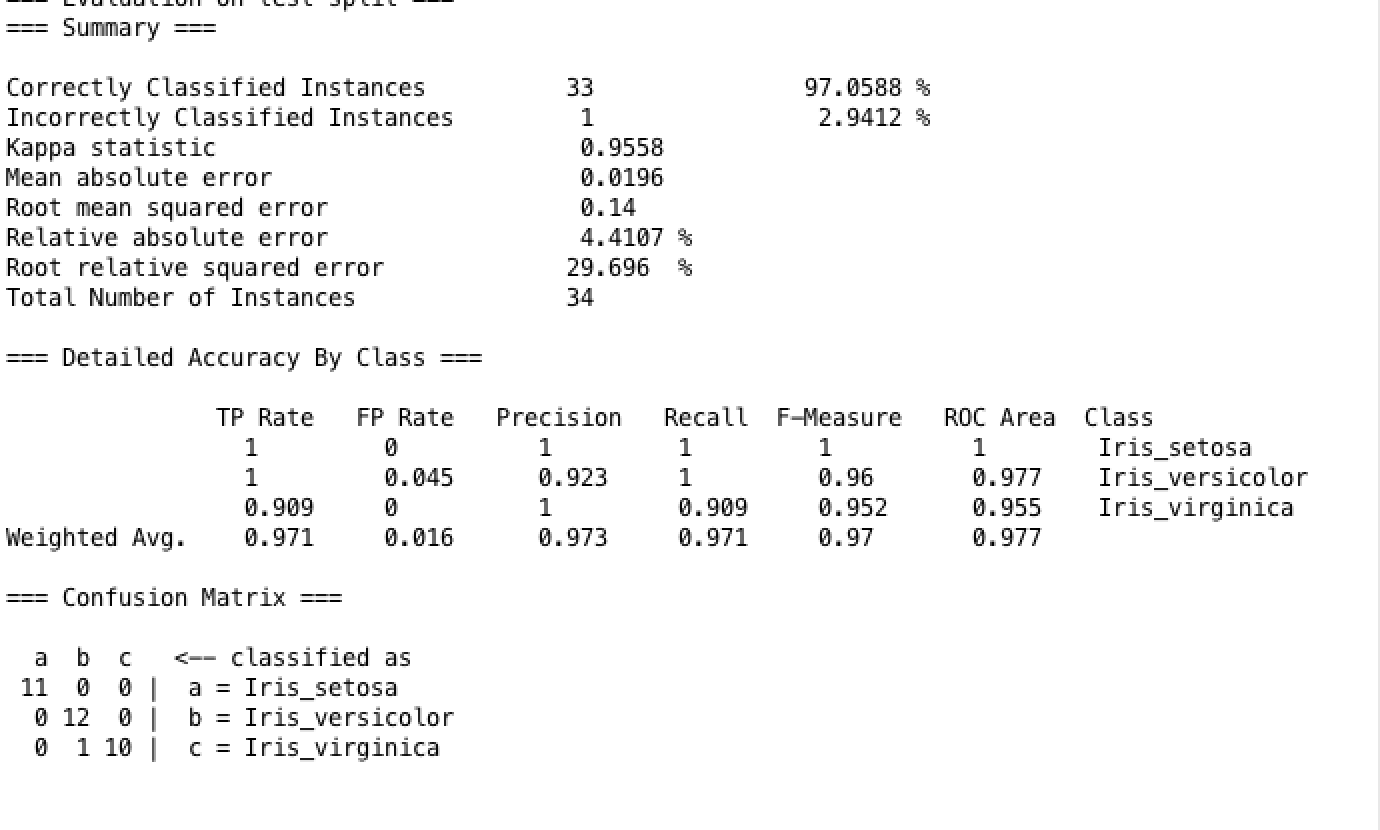


We tried different parameters in BFTree and we finally found our best accuracy which is 97.0588%. Also, we other methods that can give us fluctuating accuracy.

For decisionstump method, we got a pretty low accuracy of prediction which is just 66.6667%. (We tried several cases but 66.6667% is the best accuracy we have got since the second best accuracy is 64.7059%)



However, then we have new high accuracy with ridor method.



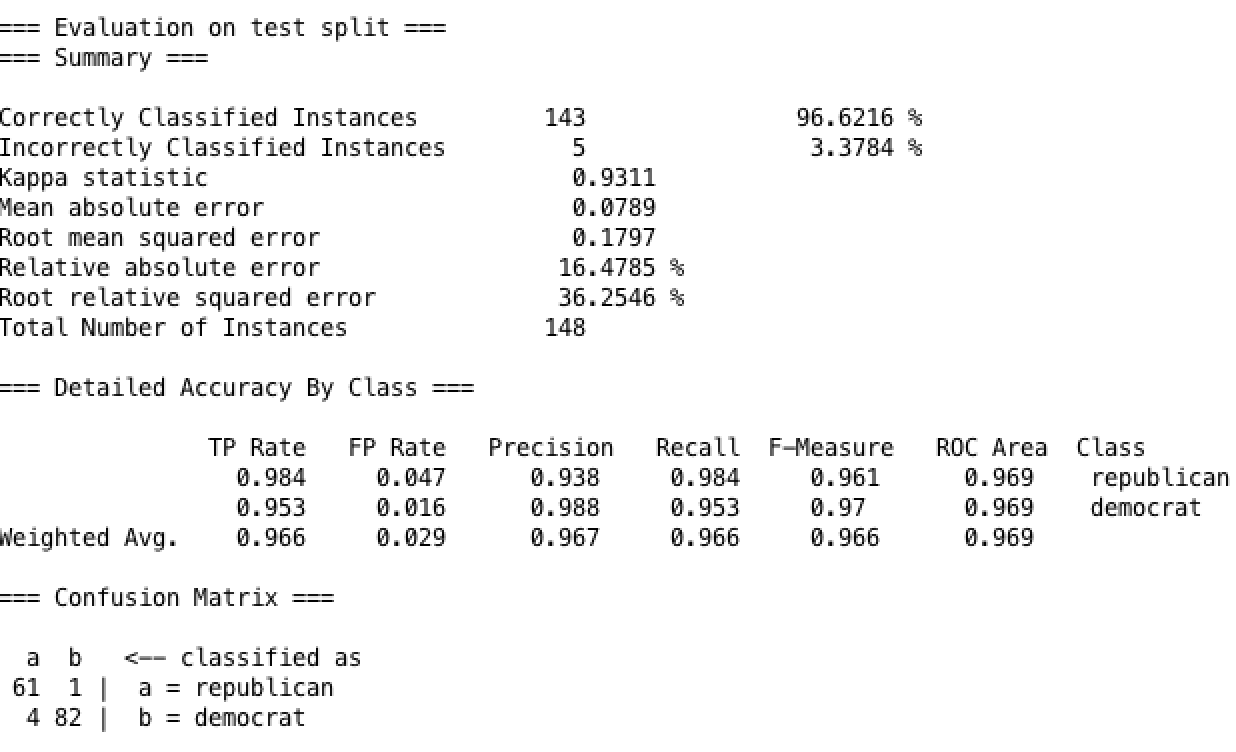
After we tried several methods we can create a convincible table for methods and accuracy.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | BFTree | Decisionstump | Ridor | Decision table | Simple cart | Zero |
| Accuracy | 97.0588% | 66.6667% | 97.0588% | 96.0784% | 96.0784% | 33.3333% |

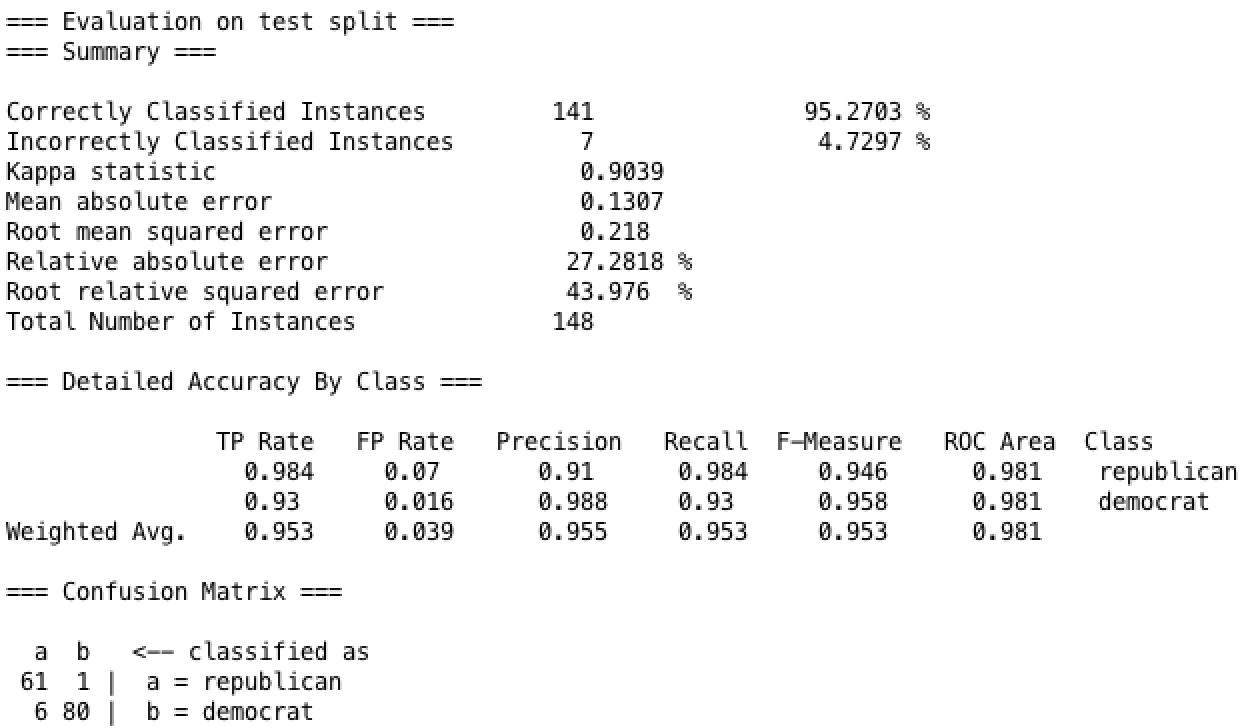
**For Congressional Voting Records**

For this data set, we tried to do the classification with decision table, simple cart, conjunctive rule, BFTree, and ridor. Also, we split data in different ways, for example, we used cross-validation folds equals to 10 in conjunctive rule, we used percentage split equals to 66% in simple cart.

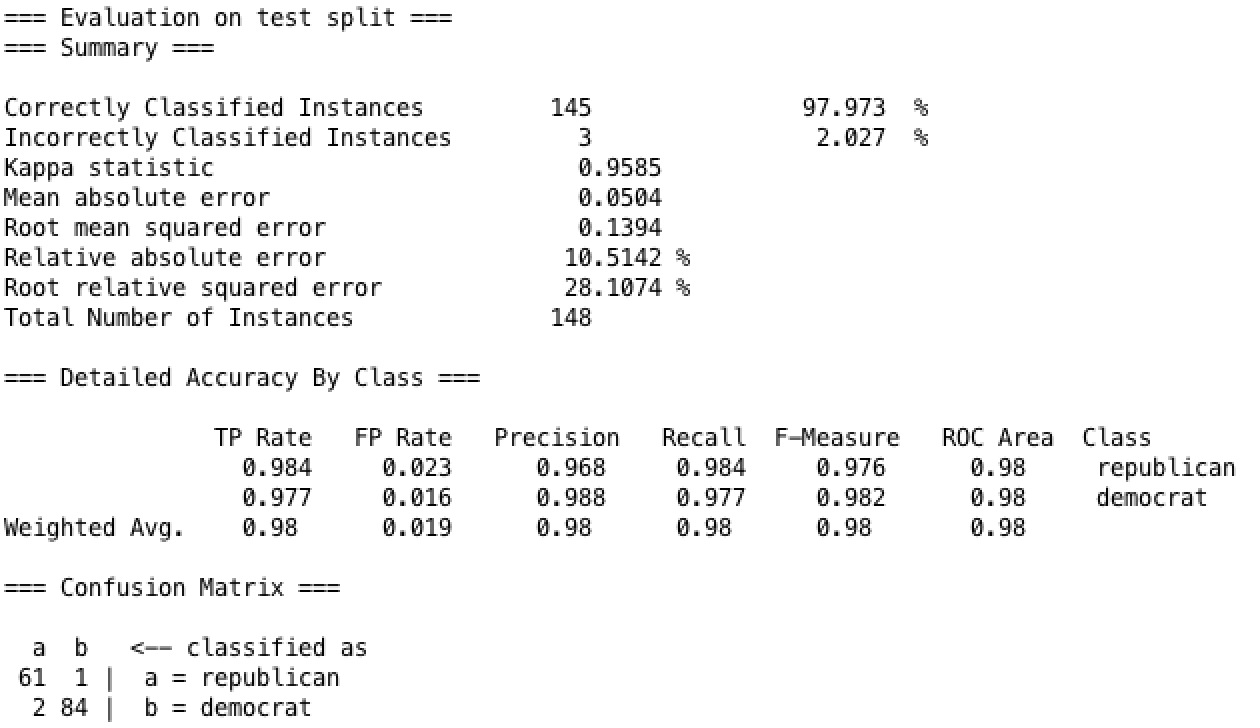
For conjunctive rule



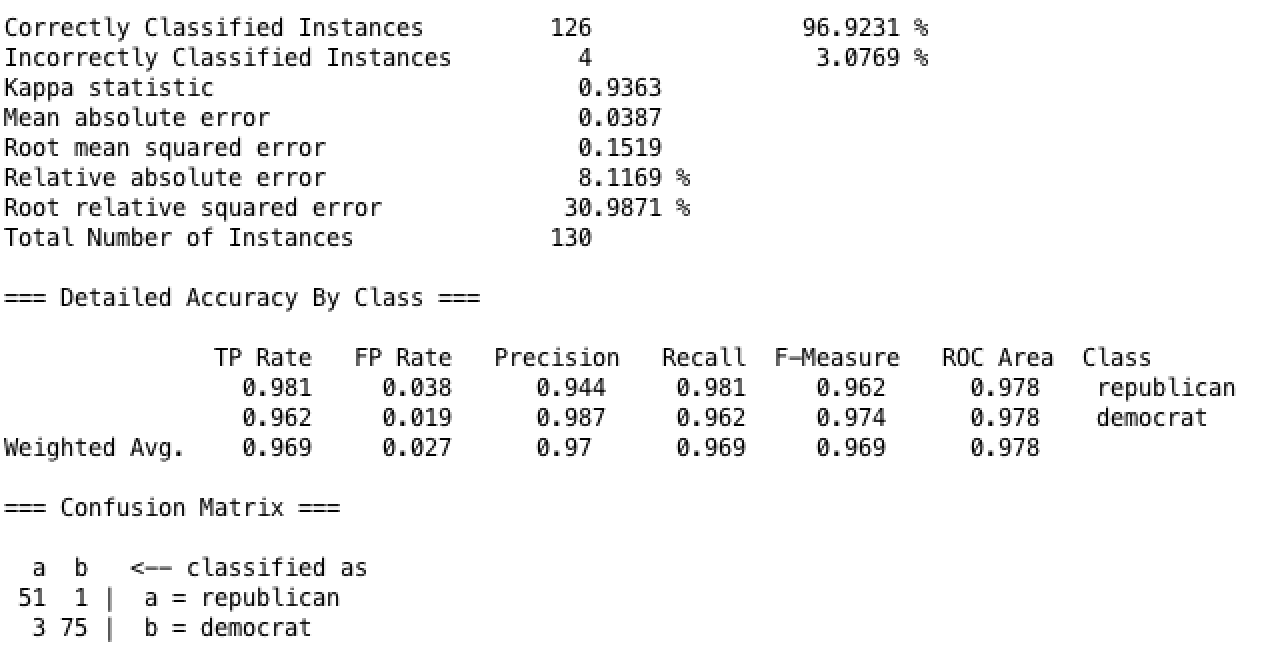
Decision table



Simple cart



BFTree



Also, we have a convincible table for accuracy of prediction.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Conjunctive rule | Decision Table | Simple Cart | Ridor | BFTree |
| Accuracy | 96.6216% | 95.2703% | 97.973% | 95.3846% | 96.9231% |

1. Conclusion

After we tried different methods with different parameters in different data sets, we find some value results. First, change in parameters would cause accuracy change differently. For example, doing decisionstump method, we have 66.6666% and 64.7059% accuracy. Second, we could have different results with different methods but there is not the best method for all the data sets, which means we have to try different learning algorithms to improve our accuracy. Third, splitting data could have impact on results. We have tried 10 folds cross validation and 435 folds cross validation in splitting Congressional Voting Records and we got two slightly different results since the main purpose of using cross validation is to avoid overfitting instead of to improve accuracy.