MiningLab WEI Chen
Nov. 2nd
The average accuracy of the Naive-Bayes classifier on my collection of documents.

16 text documents have been collected dealing with the fruit 'apple', and other 16 dealing with Apple company. All documents are collected from online news or online business website.

By k-fold cross validation, in my case k=4, the indices of documents in test sets of 4 evaluations are listed as follows:

test set 1=[2, 1, 23, 12, 3, 13, 4, 31]

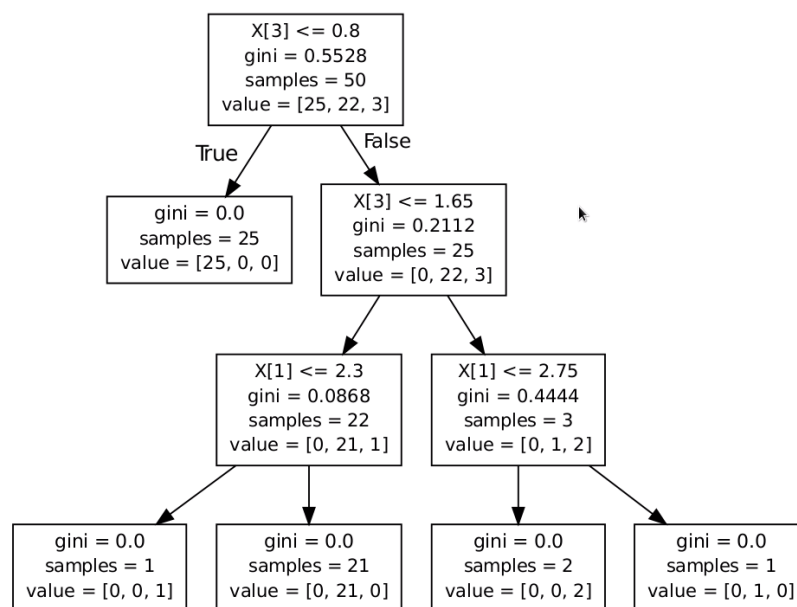test set 2= [6, 10, 19, 20, 24, 28, 5, 18]

test set 3= [25, 29, 21, 17, 7, 8, 9, 11]

test set 4= [14, 26, 15, 22, 0, 16, 27, 30]

And the average accuracy is computed as 63.0%

Decision Tree
Visualization of the decision tree with overfitting:

The confidence interval with overfitting is obtained:
(0.44182378391560051, 0.63817621608439956)