Question 6

In this question there is a file named random_forest.py which can be runned using :

Python random_forest.py gain/gini : Here gain/gini specifies the entropy function to be used. (python version = 2)

Now as the code does oob and test accuracy both for 100 trees it takes approx 5 minutes to run the code.

6a)

For this part the function used was gini as sklearn has a default function as gini and the classifier was trained for 70-30 split. Here for this part 100 trees were made and the resultant test results are:

accuracy is : 94.20

Time taken : 5min23sec

Now results for sklearn:

Time taken : Approx 1 sec

Accuracy : 95.72

6b)

Here for this part the sensitivity for m was checked for sqrt(n), 2*sqrt(n) and 0.5*sqrt(n) where the number of trees where set to 50.(n is number of attributes available)

1. **sqrt(m) = 0.9384**
2. **2*sqrt(m) = 0.9420**
3. **0.5*sqrt(m) = 0.9290**

6c)

Here the oob was founded for different sensitivity of m as above with 50 trees