**HW-5 REPORT**

1. The stopping criteria were:
   1. When there are less than 15 examples, then out of those examples, we do vote and return the majority class
   2. If all data belongs to same class, we return that class
   3. If there are no examples, we just calculated majority of parent examples.
   4. In case of random forest, one extra condition is that we stop at depth 50.

Heuristic used was just entropy.

1. No new attributes added
2. Test set performance: Decision Tree: 80.6%, Random Forest: 81.065% (with slight variation due to randomness in the algorithm)

Best Kaggle Score: 81.4%

1. First Test Data was used. The decisions made were:

'feature 1(‘pain’) <= 0'

'feature 2(‘private’) <= 0'

'feature 3 (‘bank’)<= 0'

'feature 4 (‘money’)<= 0',

'feature 5 (‘drug’)<= 0'

'feature 6 (‘spam’)<= 0'

'feature 7 (‘prescription’)<= 0'

'feature 8 (‘creative’)<= 0'

'feature 9 (‘height’)<= 0'

'feature 10(‘featured’)<= 0'

'feature 11(‘differ’) <= 0'

'feature 12 (‘width’)<= 0'

'feature 13 (‘other’)<= 0'

'feature 14(‘energy’)<= 0'

'feature 15(‘business’) <= 0'

'feature 16(‘message’)> 0'

'feature 17(‘volume’)<= 0'

'feature 18(‘revision’)<= 0'

'feature 19(‘path’) <= 0'

'feature 20(‘meter’) <= 0’

'feature 21(‘memo’) <= 0'

'feature 22(‘planning’) <= 0'

'feature 23(‘pleased’)<= 0'

'feature 24(‘record’)<= 0'

'feature 25(‘out’) <= 0'

'feature 26(‘;’) <= 0'

'feature 27(‘$’) <= 0'

'feature 28(‘#’) <= 0'

'feature 29(‘!’) > 0'

'feature 30(‘(’) <= 0'

'feature 31(‘[’) <= 0'

1. feature 19(‘path’) >0 (1 tree)

feature 22(‘planning’)>0 (3 trees)

feature 4(‘money’)>0 (2 trees)

feature 27(‘$’)>0 (4 trees)

feature 9(‘height’)>0 (1 tree)

feature 7(‘prescription’)>0 (1 tree)

feature 30(‘(’)>0 (1 tree)

feature 15(‘business’)>0 (1 tree)

feature 2(‘private’)>0(1 tree)

feature 23(‘pleased’)>0(5 trees)

feature 5(‘drug’)>0 (1 tree)

feature 28(‘#’)>0 (1 tree)

feature 10(‘featured’)>0 (1 tree)

feature 8(‘creative’)>0 (2 tree)

feature 31(‘[’)>0 (5 trees)

feature 18(‘revision’)>0(2 trees)

feature 16(‘message’)>0(2 trees)

feature 21(‘memo’)>0(4 trees)

feature 3(‘bank’)>0(1 tree)

feature 26(‘;’)>0(1 tree)

feature 24(‘record’)>0(2 trees)

feature 6(‘spam,’)>0(3 trees)

feature 29(‘!’)>0(1 tree)

feature 11(‘differ’)>0(1 tree)

feature 32(‘&’)>0(1 tree)

feature 14(‘energy’)>0(2 tree)

(f) Optional question not attempted

(g) No external source used