Amr Mohamed Salah (6287)

Nour El-Din Hazem (6261)

Data Classification

1) Adaboost:

Adaboost with best k = 295

Adaboost Classifier Measures:

	precision	recall	f1-score	support
g	0.82	0.83	0.82	2006
h	0.83	0.82	0.82	2007
accuracy			0.82	4013
macro avg	0.82	0.82	0.82	4013
weighted avg	0.82	0.82	0.82	4013

Accuracy: 82.30750062297533 %

Confusion Matrix:

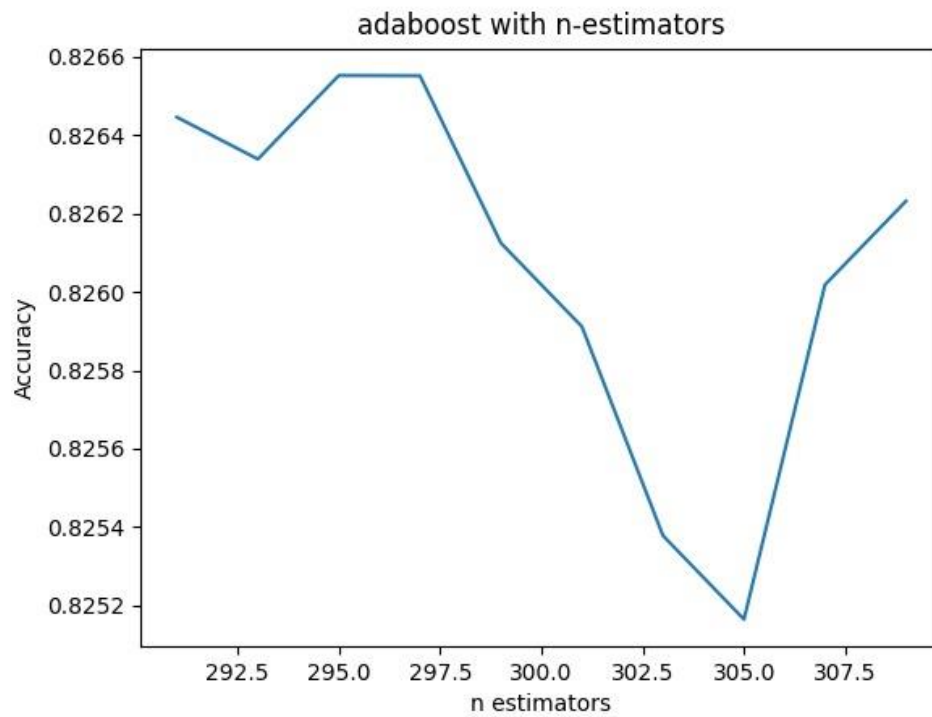| TP:1664	FP:342		[[1664 342]
FN:368	TN:1639		[368 1639]]

Hyperparameter Tuning:

From n = 0 to n = 350
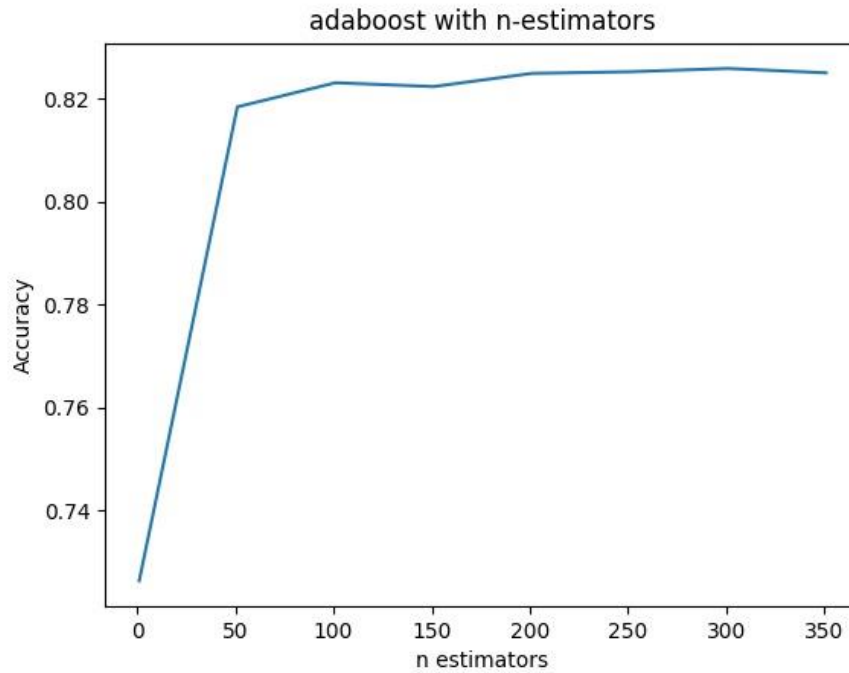
From n = 400 to n = 1000
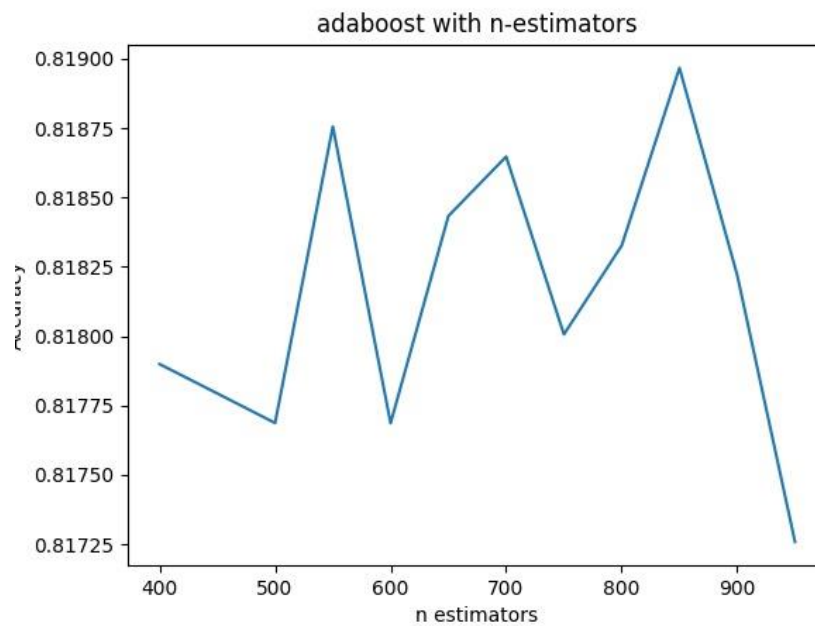
Comment: the Accuracy seems constant when increasing value of n_estimators, maximum value found in the range 290 to 310.

adaboost with n-estimators

2) KNN:

KNN with best k = 9

K Nearest Neighbors Classifier Measures:

	precision	recall	f1-score	support
g	0.73	0.84	0.78	2007
h	0.81	0.70	0.75	2006
accuracy			0.77	4013
macro avg	0.77	0.77	0.77	4013
weighted avg	0.77	0.77	0.77	4013

Accuracy: 76.65088462496885 %

Confusion Matrix:

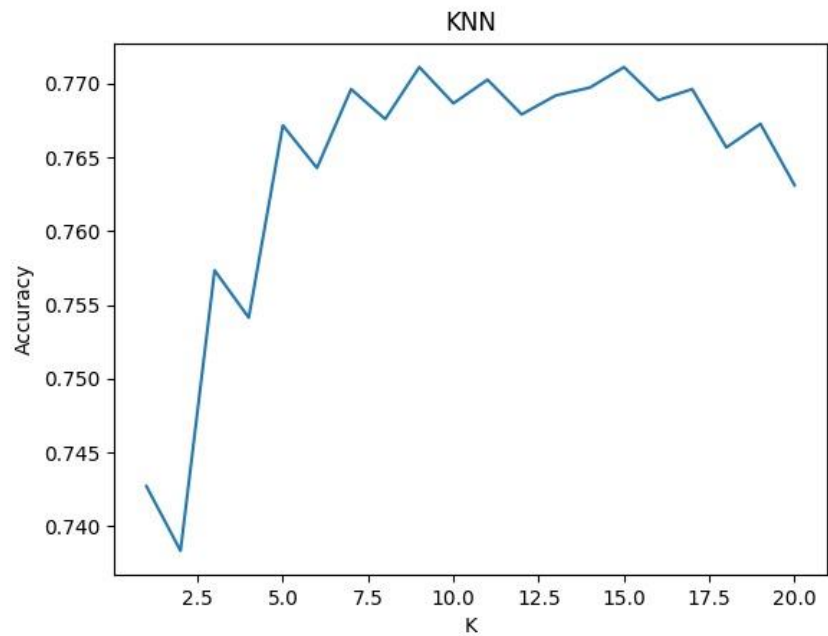|		
TP:1681	FP:326	[[1681 326]
FN:611	TN:1395	[611 1395]]

Hyperparameter Tuning:

from K = 0 to K = 200
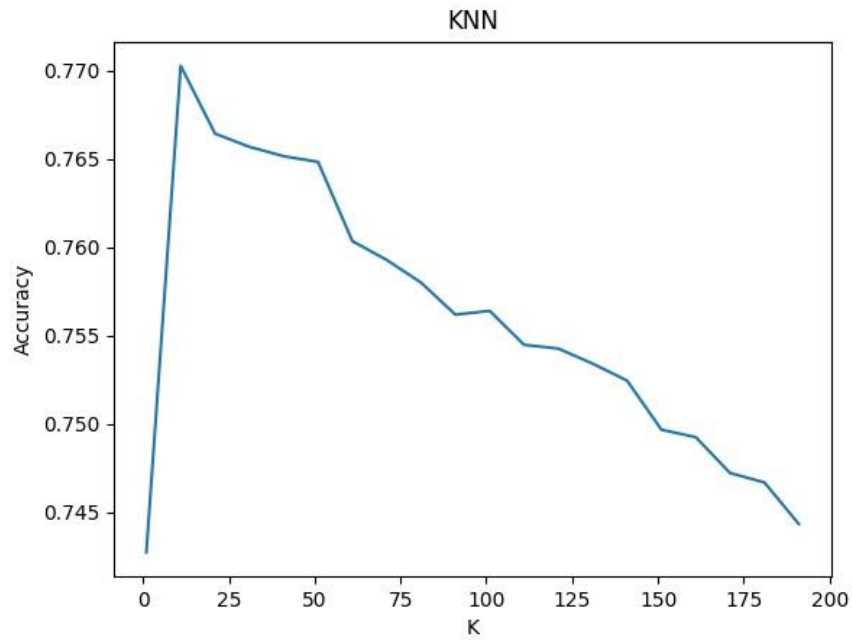
Highest Accuracy found in the range of 1 to 20

3) Random Forest

Random Forest with best k = 195

Random Forest Classifier Measures:

	precision	recall	f1-score	support
g	0.83	0.88	0.86	2006
h	0.88	0.83	0.85	2007
accuracy			0.85	4013
macro avg	0.86	0.85	0.85	4013
weighted avg	0.86	0.85	0.85	4013

Accuracy: 85.44729628706703 %
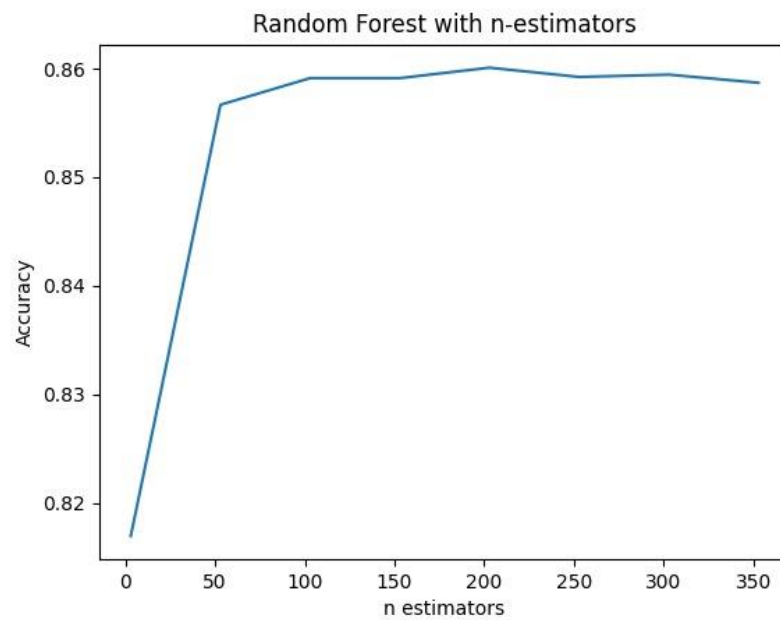
Confusion Matrix:
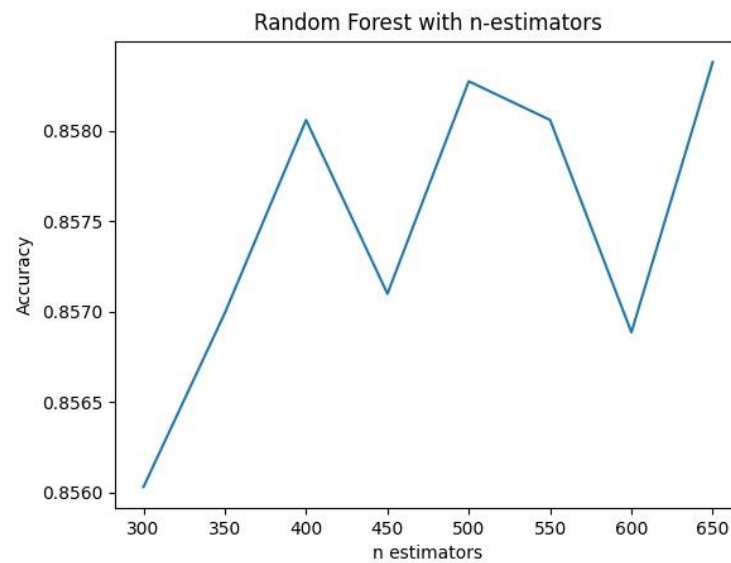
TP:1773   FP:233

FN:351   TN:1656
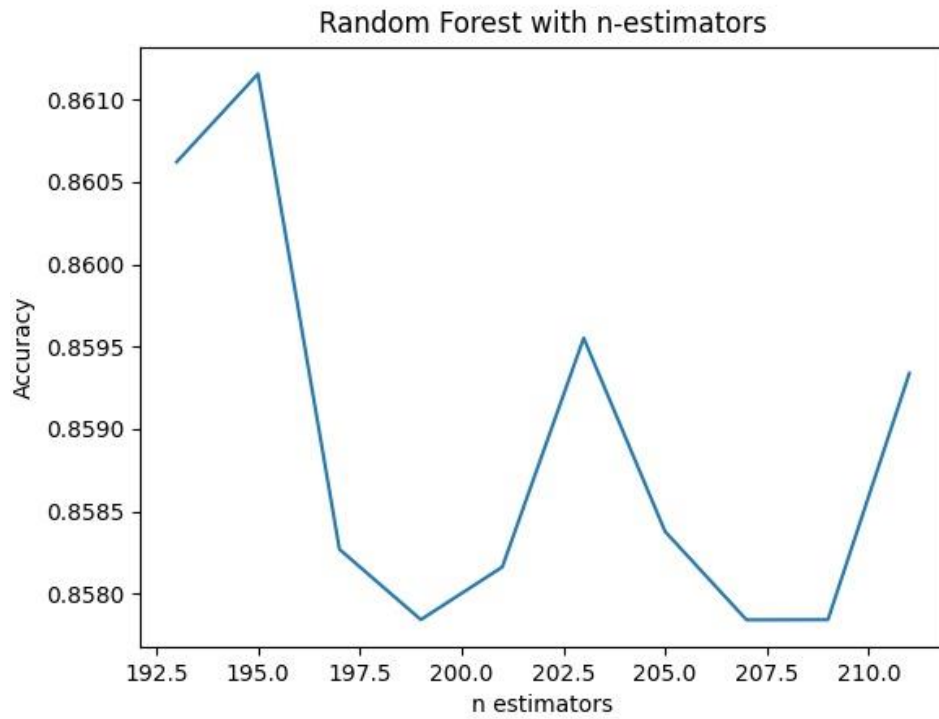
[[1773 233]

 [351 1656]]

Hyperparameter Tuning:

from n = 0 to n = 400

from n = 300 to n = 700

Comment: the Accuracy seems constant when increasing value of n_estimators, maximum value found in the range 190 to 210.

Random Forest with n-estimators

4) Decision Tree:

Decision Tree Classifier Measures:

	precision	recall	f1-score	support
g	0.79	0.80	0.79	2006
h	0.79	0.78	0.79	2007
accuracy			0.79	4013
macro avg	0.79	0.79	0.79	4013
weighted avg	0.79	0.79	0.79	4013

Accuracy: 79.04310989284824 %

Confusion Matrix:

TP:1598   FP:408

FN:433   TN:1574

[[1598 408]
 [433 1574]]

5) Naïve Bayes:

Naive Bayes Classifier Measures:

	precision	recall	f1-score	support
g	0.61	0.91	0.73	2007
h	0.82	0.42	0.56	2006
accuracy			0.67	4013
macro avg	0.72	0.67	0.64	4013
weighted avg	0.72	0.67	0.64	4013

Accuracy: 66.50884624968852 %

Confusion Matrix:

TP:1821 FP:186

FN:1158 TN:848

[[1821 186]
 [1158 848]]

Comments:

- Naïve Bayes resulted in the worst accuracy due to the assumption that all the features are independent which was not the case.
- Ensemble classifiers are expected to produce higher results than the single classifiers as they are a combination of several weak classifiers.