I) Random Forest Classifier

The following results are obtained by setting the number of features to be considered as **auto** and varying the number of base learners.

1) For n = 10

```
The Confusion Matrix is:
 [[2074  103]
 [ 162 1250]]
The per class classification accuracy of ham and spam are: [0.95268718 0.88526912]
The accuracy is 0.9261632766787405
```

2) n = 50

```
The Confusion Matrix is:
 [[2055  122]
 [ 127 1285]]
The per class classification accuracy of ham and spam are: [0.94395958 0.91005666]
The accuracy is 0.930621342992477
```

3) n = 100

```
The Confusion Matrix is:
 [[2077  100]
 [ 130 1282]]
The per class classification accuracy of ham and spam are: [0.95406523 0.90793201]
The accuracy is 0.935915296740039
```

4) n = 500

```
The Confusion Matrix is:
 [[2071  106]
 [ 125 1287]]
The per class classification accuracy of ham and spam are: [0.95130914 0.91147309]
The accuracy is 0.9356366675954305
```

5) n = 1000

```
The Confusion Matrix is:
 [[2071  106]
 [ 120 1292]]
The per class classification accuracy of ham and spam are: [0.95130914 0.91501416]
The accuracy is 0.9370298133184731
```

The following results are obtained by setting the number of features to be considered as **sqrt** and varying the number of base learners.

1) n = 10

```
The Confusion Matrix is:
 [[2074  103]
 [ 183 1229]]
The per class classification accuracy of ham and spam are: [0.95268718 0.8703966 ]
The accuracy is 0.9203120646419616
```

2) n = 50

```
The Confusion Matrix is:
 [[2064  113]
 [ 137 1275]]
The per class classification accuracy of ham and spam are: [0.94809371 0.9029745 ]
The accuracy is 0.9303427138478685
```

3) n = 100

```
The Confusion Matrix is:
 [[2069  108]
 [ 130 1282]]
The per class classification accuracy of ham and spam are: [0.95039045 0.90793201]
The accuracy is 0.9336862635831708
```

4) n = 500

```
The Confusion Matrix is:
 [[2069  108]
 [ 127 1285]]
The per class classification accuracy of ham and spam are: [0.95039045 0.91005666]
The accuracy is 0.9345221510169964
```

5) n = 1000

```
The Confusion Matrix is:
 [[2065  112]
 [ 127 1285]]
The per class classification accuracy of ham and spam are: [0.94855305 0.91005666]
The accuracy is 0.9334076344385622
```

The following results are obtained by setting the number of features to be considered as **log2** and varying the number of base learners.

1) n = 10

```
The Confusion Matrix is:
 [[2065  112]
 [ 127 1285]]
The per class classification accuracy of ham and spam are: [0.94855305 0.91005666]
The accuracy is 0.9334076344385622
```

2) n = 50

```
The Confusion Matrix is:
 [[2075  102]
 [ 135 1277]]
The per class classification accuracy of ham and spam are: [0.95314653 0.90439093]
The accuracy is 0.9339648927277793
```

3) n = 100

```
The Confusion Matrix is:
 [[2066  111]
 [ 125 1287]]
The per class classification accuracy of ham and spam are: [0.9490124  0.91147309]
The accuracy is 0.9342435218723879
```

4) n = 500

```
The Confusion Matrix is:
 [[2072  105]
 [ 127 1285]]
The per class classification accuracy of ham and spam are: [0.95176849 0.91005666]
The accuracy is 0.935358038450822
```

5) n = 1000

```
The Confusion Matrix is:
 [[2072  105]
 [ 128 1284]]
The per class classification accuracy of ham and spam are: [0.95176849 0.90934844]
The accuracy is 0.9350794093062135
```

Observations:

The model's accuracy increased with the increase in base learners, starting from as low as 92% approx and taking the maximum value of 93% approx. The accuracy improved by changing the number of features considered during each split, where 'sqrt' gave the least accuracy of 93.34% and 'auto' showed the highest accuracy of 93.70%.

Setting n=1000 and max features = auto as the parameters for the Random Forest Classifier gives us the highest accuracy of 93.70% and per-class classification accuracy of 95.13% for ham and 91.50% for spam.

II) Boosting Ensemble with logistic regression base learner

The boosting ensemble used for this case is the AdaBoost ensemble using logistic regression as a base learner. The following outputs result from varying the number of base learners.

1) n = 10

```
The Confusion Matrix is:
 [[2001  176]
 [ 224 1188]]
The per class classification accuracy of ham and spam are: [0.9191548  0.84135977]
The accuracy is 0.8885483421565896
```

2) n = 50

```
The Confusion Matrix is:
 [[1995  182]
 [ 151 1261]]
The per class classification accuracy of ham and spam are: [0.91639871 0.89305949]
The accuracy is 0.9072164948453608
```

3) n = 100

```
The Confusion Matrix is:
 [[2005  172]
 [ 140 1272]]
The per class classification accuracy of ham and spam are: [0.92099219 0.90084986]
The accuracy is 0.9130677068821399
```

4) n = 500

```
The Confusion Matrix is:
 [[1994  183]
 [ 128 1284]]
The per class classification accuracy of ham and spam are: [0.91593937 0.90934844]
The accuracy is 0.9133463360267484
```

5) n = 1000

```
The Confusion Matrix is:
 [[1989  188]
 [ 124 1288]]
The per class classification accuracy of ham and spam are: [0.91364263 0.9121813 ]
The accuracy is 0.9130677068821399
```

Observations:

While the model's accuracy increased when the base learners increased from 10 to 50 to 100, it remained almost the same for 100, 500 and 1000.

Setting n=1000 as the parameters for the AdaBoost ensemble with logistic regression as a base learner gives us the highest accuracy of 91.30% and per-class classification accuracy of 91.36% for ham and 91.21% for spam.

III)AdaBoost Ensemble with Decision Tree base learner

The following outputs result from varying the number of base learners.

1) n = 10

```
The Confusion Matrix is:
 [[1906  271]
 [ 168 1244]]
The per class classification accuracy of ham and spam are: [0.87551677 0.88101983]
The accuracy is 0.877681805516857
```

2) n = 50

```
The Confusion Matrix is:
 [[1904  273]
 [ 160 1252]]
The per class classification accuracy of ham and spam are: [0.87459807 0.88668555]
The accuracy is 0.8793535803845082
```

3) n = 100

```
The Confusion Matrix is:
 [[1896  281]
 [ 155 1257]]
The per class classification accuracy of ham and spam are: [0.87092329 0.89022663]
The accuracy is 0.8785176929506826
```

4) n = 500

```
The Confusion Matrix is:
 [[1908  269]
 [ 142 1270]]
The per class classification accuracy of ham and spam are: [0.87643546 0.89943343]
The accuracy is 0.8854834215658958
```

5) n = 1000

```
The Confusion Matrix is:
 [[1927  250]
 [ 154 1258]]
The per class classification accuracy of ham and spam are: [0.88516307 0.89093484]
The accuracy is 0.8874338255781554
```

Observations:

The model's accuracy increased with the increase in base learners, gaining the highest value of 88.7%, with 1000 base learners compared to lower base learners. At n=1000, the per-class classification accuracy for ham is 88.51% and 89.09% for spam.

While considering the accuracy of the decision tree alone, as shown below, the overall accuracy is 88.82%, with entropy as the criterion set.

```
The Confusion Matrix is:
 [[1922  255]
 [ 146 1266]]
The per class classification accuracy of ham and spam are: [0.88286633 0.89660057]
The accuracy is 0.8882697130119811
```

Taking the parameters that give the highest accuracy from the observations mentioned above, it can be observed that:

1) Random forest classifier made the highest accurate predictions for the class spam with 1292 correct predictions, followed by 1288 correct predictions made by the AdaBoost classifier with Logistic Regression base learner, and 1266 correct predictions made by the decision tree classifier. The AdaBoost Ensemble Classifier with Decision trees as the base learner made the least number of correct predictions of 1258 out of 1412 data spam instances.

2) The per-class accuracy using the AdaBoost Ensemble Classifier with Decision trees and Decision Tree Classifiers with entropy as the criterion have very close values. The per-class accuracy for ham is 88.5% using AdaBoost ensemble closely followed by 88.2% using Decision Tree Classifiers, similarly the per-class accuracy for spam is 89.09% using AdaBoost ensemble and 89.66% using Decision Tree Classifiers.

While the per-class classification accuracy for ham and spam using the AdaBoost Classifier with Logistic Regression as a base learner was 91.36% and 91.2%, respectively, the Random Forest Classifier obtained the highest per-class accuracy with 95.1% accuracy for ham and 91.5% per-class accuracy for spam.

3) The AdaBoost Ensemble Classifier and Decision Tree classifier gave approximately similar accuracy of 88.7% and 88.8%, respectively. The AdaBoost Ensemble Classifier closely follows this with an accuracy of 91.30%, whereas the Random Forest Classifier has the highest classification accuracy of 93.70%.

4) The AdaBoost Ensemble Classifier with Decision trees as base learners produce results such as per-classification accuracy and accuracy almost similar to the Decision Tree Classifier results as opposed to the AdaBoost Ensemble Classifier with Logistic Regression as a base learner.

5) The AdaBoost Ensemble classifier produces better results with Logistic Regression as a base learner instead of Decision Trees as the base learner.

6) Lastly, the Random Forest Classifier produced the best results considering the given parameters among the four classifiers used in this project to classify the given data as 'ham' or 'spam'.