

Compare the accuracies of the fused classifier with AdaBoost Ensemble with Decision Tree as the base learner.

For this comparison, the base learners are chosen in such a manner that they provide the highest possible accuracy. For instance, the AdaBoost Classifier uses 1000 base learners. Additionally, I took three neighbours for our k-NN classifier, 1000 base learners for Logistic Regression and Random forest, respectively.

The accuracy of the AdaBoost Ensemble Classifier with Decision Tree as the base learner was as follows:

```
The Confusion Matrix is:  
[[1895  282]  
 [ 156 1256]]  
The per class classification accuracy of ham and spam are: [0.87046394 0.88951841]  
The accuracy is 0.8779604346614656
```

The accuracy of the fused classifiers was as follows:

```
The Confusion Matrix is:  
[[2076  101]  
 [ 143 1269]]  
The per class classification accuracy of ham and spam are: [0.95360588 0.89872521]  
The accuracy is 0.9320144887155196
```

Observations:

The fusion model came up with a 93.2% accuracy rating, higher than that of AdaBoost by 5.41%. Furthermore, the fusion model correctly predicted 2076 ham emails and 1279 spam emails, resulting in higher per-class numbers than AdaBoost. Overall, the fused classifier performed better.

Impact of training sample size on the accuracies of the fused classifier and the AdaBoost Ensemble with Decision Tree as the base learner.

1) 50-50% training-test split:

Fused Classifier:

```
Average accuracy for 30 iterations is 0.9382732145444009
The average confusion matrix for 30 iterations is [[1346.86666667  48.26666667]
 [ 93.76666667  812.1      ]]
Average per class classification accuracy for 30 iterations is [0.96542294 0.89654917]
```

AdaBoost Classifier:

```
Average accuracy for 30 iterations is 0.9239316239316239
The average confusion matrix for 30 iterations is [[1307.2      83.63333333]
 [ 91.4      818.76666667]]
Average per class classification accuracy for 30 iterations is [0.93984052 0.89966315]
```

2) 60-40% training-test split:

Fused Classifier:

```
Average accuracy for 30 iterations is 0.9412456998008328
The average confusion matrix for 30 iterations is [[1073.36666667  39.33333333]
 [ 68.83333333  659.46666667]]
Average per class classification accuracy for 30 iterations is [0.96464616 0.90554327]
```

AdaBoost Classifier:

```
Average accuracy for 30 iterations is 0.9363751584283903
The average confusion matrix for 30 iterations is [[1059.9      56.13333333]
 [ 61.      663.96666667]]
Average per class classification accuracy for 30 iterations is [0.94976099 0.9158879 ]
```

3) 70-30% training-test split:

Fused Classifier:

```
Average accuracy for 30 iterations is 0.9418054549843108
The average confusion matrix for 30 iterations is [[811.23333333  27.9      ]
 [ 52.46666667  489.4      ]]
Average per class classification accuracy for 30 iterations is [0.96676101 0.9032789 ]
```

AdaBoost Classifier:

```
Average accuracy for 30 iterations is 0.9410089307265266
The average confusion matrix for 30 iterations is [[796.36666667  34.63333333]
 [ 46.83333333  503.16666667]]
Average per class classification accuracy for 30 iterations is [0.9583216  0.91480683]
```

4) 80-20% training-test split:

Fused Classifier:

```
Average accuracy for 30 iterations is 0.9435396308360479
The average confusion matrix for 30 iterations is [[541.43333333  18.66666667]
 [ 33.33333333  327.56666667]]
Average per class classification accuracy for 30 iterations is [0.96671798 0.90767091]
```

AdaBoost Classifier:

```
Average accuracy for 30 iterations is 0.9496923633731451
The average confusion matrix for 30 iterations is [[538.23333333  18.33333333]
 [ 28.          336.43333333]]
Average per class classification accuracy for 30 iterations is [0.9670283  0.92303655]
```

Observations:

The accuracy of the fused classifier increased from 93% to 94% while using 50-50% split and 60-40% split, respectively. After that, the fused classifier had approximately very similar accuracy for the remaining splits and varied by decimal points. A similar trend can be observed in the per-class classification results as well.

However, the AdaBoost Ensemble accuracy increased from 92.3% to 94.1% for the 50-50% and 80-20%, respectively. The classifier displayed a similar trend in the per-class classification accuracy of ham and spam.

When the training-testing ratio was set to 50-50% and 70-30%, the fused classifier displayed a higher accuracy than the AdaBoost ensemble. In contrast, the AdaBoost ensemble had a higher accuracy than the fused classifier for the training-test ratio of 60-40% and 80-20%.