CS 6375 Assignment – 4:

Suchith Natraj Javali (SXJ200024)

Harichandana Neralla (HXN210036)

PART -1:

Implement Bagging and AdaBoost based on the decision tree code that you developed in Programming Assignment 1. (Bagging, 20 points) Construct four models for each combination of maximum depth d = 3,5 and bag size (k = 10,20). Report the confusion matrix for these four settings. b. (Boosting, 20 points) Construct four models for each combination of maximum depth d = 1,2 and bag size (k = 20,40). Report the confusion matrix for these four settings.

Output:

Bagging: Max Depth 3 and Bag Size 10

	Classifier Positive	Classifier Negative
Actual Positive	132.0	84.0
Actual Negative	12.0	204.0

Bagging: Max Depth 3 and Bag Size 20

	Classifier Positive	Classifier Negative
Actual Positive	168.0	48.0
Actual Negative	48.0	168.0

Bagging: Max Depth 5 and Bag Size 10

	Classifier Positive	Classifier Negative
Actual Positive	170.0	46.0
Actual Negative	6.0	210.0

Bagging: Max Depth 5 and Bag Size 20

	Classifier Positive	Classifier Negative
Actual Positive	196.0	20.0
Actual Negative	13.0	203.0

Boosting: Max Depth 1 and Number of Stumps 20

	Classifier Positive	Classifier Negative
Actual Positive	156.0	60.0
Actual Negative	96.0	120.0

Boosting: Max Depth 1 and Number of Stumps 40

	Classifier Positive	Classifier Negative
Actual Positive	156.0	60.0
Actual Negative	96.0	120.0

Boosting: Max Depth 2 and Number of Stumps 20

	Classifier Positive	Classifier Negative
Actual Positive	180.0	36.0
Actual Negative	48.0	168.0

Boosting: Max Depth 2 and Number of Stumps 40

	Classifier Positive	Classifier Negative
Actual Positive	180.0	36.0
Actual Negative	48.0	168.0

Sklearn Bagging: Max Depth 3 and Bag Size 10

	Classifier Positive	Classifier Negative
Actual Positive	192.0	24.0
Actual Negative	0.0	216.0

Sklearn Bagging: Max Depth 3 and Bag Size 20

	Classifier Positive	Classifier Negative
Actual Positive	216.0	0.0
Actual Negative	0.0	216.0

Sklearn Bagging: Max Depth 5 and Bag Size 10

	Classifier Positive	Classifier Negative
Actual Positive	189.0	27.0
Actual Negative	2.0	214.0

Sklearn Bagging: Max Depth 5 and Bag Size 20

	Classifier Positive	Classifier Negative
Actual Positive	204.0	12.0
Actual Negative	21.0	195.0

Sklearn Boosting: Max Depth 1 and Number of Stumps 20

	Classifier Positive	Classifier Negative
Actual Positive	141.0	75.0
Actual Negative	53.0	163.0

Sklearn Boosting: Max Depth 1 and Number of Stumps 40

	Classifier Positive	Classifier Negative
Actual Positive	140.0	76.0
Actual Negative	57.0	159.0

Sklearn Boosting: Max Depth 2 and Number of Stumps 20

	Classifier Positive	Classifier Negative
Actual Positive	216.0	0.0
Actual Negative	0.0	216.0

Sklearn Boosting: Max Depth 2 and Number of Stumps 40

	Classifier Positive	Classifier Negative
Actual Positive	216.0	0.0
Actual Negative	0.0	216.0

```
Test Error for Bagging = 5.54 %, for depth = 3 and stump = 10.
Test Error for Bagging = 11.21 %, for depth = 3 and stump = 20.
Test Error for Bagging = 5.51 %, for depth = 5 and stump = 10.
Test Error for Bagging = 10.75 %, for depth = 5 and stump = 20.
Test Error for Boosting = 0.69 %, for depth = 1 and stump = 20.
Test Error for Boosting = 0.69 %, for depth = 1 and stump = 40.
Test Error for Boosting = 1.15 %, for depth = 2 and stump = 20.
Test Error for Boosting = 1.15 %, for depth = 2 and stump = 40.
Test Error for Sklearn Bagging = 8.33 %, for depth = 3 and stump = 10.
Test Error for Sklearn Bagging = 7.41 %, for depth = 3 and stump = 20.
Test Error for Sklearn Bagging = 6.71 %, for depth = 5 and stump = 10.
Test Error for Sklearn Bagging = 5.32 %, for depth = 5 and stump = 20.
Test Error for Sklearn Boosting  = 29.63 %, for depth = 1 and stump = 20.
Test Error for Sklearn Boosting  = 30.79 %, for depth = 1 and stump = 40.
Test Error for Sklearn Boosting  = 0.00 %, for depth = 2 and stump = 20.
Test Error for Sklearn Boosting  = 0.00 %, for depth = 2 and stump = 40.
```

What can you say about the quality of your implementation's performance versus scikit's performance? Comparing the test errors for our implementation and Sklearn for bagging and boosting, we can see that Sklearn generally performs better than our implementation, with lower test errors. For bagging, our implementation tends to have higher test errors than Sklearn across all depths and stumps, with the biggest difference in test errors occurring at depth=5 and stump=20, where our implementation had a test error of 10.75% while Sklearn had a test error of 5.32%.For boosting, our implementation had lower test errors than Sklearn for all cases except for depth=2 and stump=20 and depth=2 and stump=40, where they had the same test error of 1.15%. Overall, it seems that our implementation may not be as robust as Sklearn and may need further tuning and optimization to improve its performance.