

Randomized response decision tree: I choose the randomized decision tree method mentioned in [1]. The below are the steps performed in the program.

1. Firstly, trained the decision tree on the first $X=\{100,75,50,30,6,4,1,0.5\}$ percentage of the original data, for several values of X .
2. Ran the above decision tree on the original test data and evaluated the results.
3. Randomized the original data using theta values Modified the code to work on randomized data
4. Then, trained the modified decision tree on the first $X=\{100,75,50,30,6,4,1,0.5\}$ percentage of the randomized original data, for several values of X .
5. Ran the above decision tree on the randomized original test data and evaluated the results.
6. The steps 3 to 5 are performed for different values of θ such as 0.8, and 0.7.
7. The below are the output results for same.

	Unmodified tree			Modified tree	
		$\theta = 0.8$	$\theta = 0.7$	$\theta = 0.8$	$\theta = 0.7$
	Unmodified data	Randomized data	Randomized data	Randomized data	Randomized data
%Training Data	%Accuracy	%Accuracy	%Accuracy	%Accuracy	%Accuracy
100	61	58	52	47	51
75	61	58	52	47	51
50	61	58	52	47	51
30	61	58	52	47	51
10	61	58	52	47	51
6	47	44	40	47	51
4	39	38	52	47	51
1	77	20	60	47	51
0.5	25	24	60	47	51

Impact of the value of θ and the Size of the training data: In the case of unmodified data using unmodified tree, the accuracy started decreasing after 10 percentage of training data. For different values of θ , i have got the same accuracy for all the percentage of randomized training data. I can say that this decision tree accuracy depends somehow on the data with which I trained. Also, we are checking if θ is less than random value, then only i am modifying all the columns data. I observed that most of the generated values using `Math.random()` is giving values less than 0.6. Also, we don't get the same accuracy if we run the program for same values of θ as it uses `Math.random()` function. We may expect to see a more consistent results if the attribute splitting method is used accurately. Though results are not highly accurate, I can say that attributes play a key role in classification as we use them in entropy and, gain methods to construct the decision tree. We can also improve the accuracy results if we use more size of the training data.

Reference [1]: Du, Wenliang, and Zhijun Zhan. "Using randomized response techniques for privacy-preserving data mining." *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2003.