CS5011 - Machine Learning
Software Assignment - 2

PROBLEM STATEMENT

The assignment aims at performing classification on the Bank Marketing dataset given training and test split, over various classification methods.

TASKS

Two classification models (as specified in Table 1) have to be trained using the training data to predict the class label for the features in the test data. Compare the prediction accuracies between the two prediction models allotted.

INPUT DATA

Bank Marketing Data Set: Please visit https://bit.ly/2NaqAWG for the dataset. The dataset is 16 - dimensional with 45211 samples in total. (Use 70% of the data for training, 15% for validation and the rest for testing.)

OUTPUT

Compare the classification accuracies for the two methods allotted in Table 1. Justify why one classifier outperforms the other on the given dataset.

HINTS FOR EXCELLENCE

Additional observations and visualizations of the data and the attributes of the trained models will be given extra credit.

Coding language: **Python**

Deadline: 15th April, 2020

Serial No.	Roll No.	Classifier 1	Classifier 2
1	CS19M502	SVM	LDA
2	CS19M503	k-NN	Decision Tree
3	CS19M504	Adaboost	Random Forest
4	CS19M505	LDA	Naive Bayes
5	CS19M506	GMM	k-NN
6	CS19M507	Naive Bayes	Random Forest
7	CS19M508	LDA	Adaboost
8	CS19M509	Decision Tree	SVM
9	CS19M510	Random Forest	Naive Bayes
10	CS19M511	GMM	k-NN
11	CS19M512	LDA	Random Forest
12	CS19M513	k-NN	SVM
13	CS19M514	Decision Tree	LDA
14	CS19M515	Random Forest	k-NN
15	CS19M516	Naive Bayes	GMM
16	CS19M517	SVM	Naive Bayes
17	CS19M518	GMM	Adaboost
18	CS19M519	LDA	Decision Tree
19	CS19M520	GMM	SVM
20	CS19M521	LDA	Adaboost
21	CS19M522	SVM	Naive Bayes
22	CS19M523	Random Forest	Bayes
23	CS19M524	k-NN	GMM
24	CS19M525	GMM	LDA
25	CS19M526	Decision Tree	Random Forest
26	CS19M527	Naive Bayes	SVM
27	CS19M528	Decision Tree	GMM
28	CS18M504	Naive Bayes	LDA
29	CS18M505	Adaboost	GMM
30	CS19M501	Random Forest	SVM
31	CS18M518	LDA	k-NN