General Instructions:

1. Read the problem statement carefully before answering. In case of any ambiguity please contact the invigilator

2. Write the solution in a python notebook named DNADD_B3_<<your Infosys id>>; for example, DNADD_B3_HSBC2019_Alex and upload the notebook at http://192.168.1.62/uploadfile/ upon completion of the exam.

3. Before you submit the notebook, please ensure that you have saved the last modified code. Please use File ➔ Save and Checkpoint option to do this.

4. Access to your previous notebooks, instructor notes, and internet is permitted. This is an open book exam.

5. Discussion among peers and any form of malpractice is prohibited.

---- All the best ----

Problem Statement:

This case is about a bank (Stark Banks) which has a growing customer base. Majority of these customers are liability customers (depositors) with varying size of deposits. The number of customers who are also borrowers (asset customers) is quite small, and the bank is interested in expanding this base rapidly to bring in more loan business and in the process, earn more through the interest on loans. In particular, the management wants to explore ways of converting its liability customers to personal loan customers (while retaining them as depositors). A campaign that the bank ran last year for liability customers showed a healthy conversion rate of over 9% success. This has encouraged the retail marketing department to devise campaigns with better target marketing to increase the success ratio with minimal budget.

The department wants to build a model that will help them identify the potential customers who have higher probability of purchasing the loan. This will increase the success ratio while at the same time reduce the cost of the campaign.

The file Bank_Personal_Loan_Modelling.csv contains data on 5000 customers. The data include customer demographic information (age, income, etc.), the customer's relationship with the bank (mortgage, securities account, etc.), and the customer response to the last personal loan campaign (Personal Loan). Among these 5000 customers, only 480 (= 9.6%) accepted the personal loan that was offered to them in the earlier campaign.

Marketing Campaign for Personal loans

Perform the following actions in your dataset:

1. Read the column description and ensure you understand each attribute well

2. Study the data distribution in each attribute, share your findings. (5 points)

3. Study the target column distribution and state your insights.

4. Split the data into training and test set in the ratio of 70:30 respectively

5. Use different classification models (Logistic, K-NN) to predict the likelihood of a liability customer buying personal loans (10 points)

6. Print the confusion matrix for all the above models (10 points)

7. Give your reasoning on which is the best model in this case and state your conclusions (5 points)

_____