**Question 1**

**Description Of The Data:** The data given is of campaigns run by a retail bank to sell debit card. Please go through the following to understand every variable:

* **serial\_number:** Unique identifier for each entity.
* **age\_in\_years:** Age in number of years of the individuals targeted.
* **job\_description:** Categorical variable which describes the job held by each individual.
* **marital\_status:** Categorical variable which describes the marital status of the individual.
* **educational\_status:** Categorical variable which describes the educational status of the individual
* **has\_default:** Binary variable which describes if the individual has a previous loan default.
* **balance\_in\_account:** Numerical variable which describes the average yearly balance of the account. (units of currency)
* **housing\_status:** Binary variable which describes if the individual has previous housing loan existing.
* **previous\_loan:** Binary variable which describes if the individual has previous previous personal loan existing.
* **phone\_type:** Type of phone on which the individual was contacted.
* **date:** Day of the month on which the individual was contacted.
* **month\_of\_year:** Categorical variable which describes the month of the year on which the individual was contacted.
* **call\_duration:** Contact duration in number of seconds on the last campaign call.
* **campaign\_contacts:** Number of days from the previous campaign after which the individual was contacted. -1 signifies that no previous contact is on record.
* **days\_passed:** Total number of contacts to this individual before the current campaign started.
* **previous\_contact:** Number of previous contacts on the previous marketing campaign
* **poutcome\_of\_the\_campaign:** Categorical variable that describes the outcome of the previous campaign.
* **outcome:** Binary outcome of the current campaign.

**Objective Of The Problem:** Objective of the problem is to predict the categorical outcome of each entity in the test data set. “outcome” variable is to be predicted as per “serial\_number” variable of the test data set. Please note that the predictions are to be written to a csv file which contains same headers as the sample submission file and same number of rows as per the sample submission file.

**Evaluation Metric:** The evaluation metric for this binary classification problem is accuracy. Please go through the following link for a better understanding of the evaluation metric: [Scikit\_Accuracy](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html" \t "_blank)

**Procedure Of Creating Data Models:** Personal computing resource is to be used for data modelling. Please note all data modelling is to be done on the training data set and all predictions are to be made for the serial\_number variable of the test data set. Please write all predictions to a csv file just as the sample submission file and upload the same.

**Submission Limit:** Please note that total submission limit for a team is 15. Individual submission limit is shown to individual users but if the total team submission limit is reached no more submissions would be allowed even if individual submission limit has not been reached.

**Code and Concept Note Submission:** Please form a brief concept note on what are the steps that has been used to reach a certain accuracy. Please zip together your concept note and notebook file or code file of the solution together and upload the same on the code submission stage. Please note this is compulsory for each team. Concept notes and code or notebook submissions would be manually evaluated. The complete evaluation procedure is not just on the accuracy but also on how the problem was solved.

**Evaluation Algorithm**

Accuracy