**Data Science: Campus Case Study Competition**

**Background:**

* The success of Debt Collection lies in timely intervention by the bank. If a customer doesn’t pay EMI, the aim is to recover the money within the same month to improve the cash flow of the bank and reduce the losses.
* It is imperative to select the right allocation channel for connecting with the customer (Customer visit / Calling / SMS), in order to minimize the collection cost of the company as well as for maintaining a good customer relationship.
* Thus, for efficient Collections, it is important to devise optimal allocation strategies based on the risk of the customer (probability of not paying EMI in next 1 month)

**Problem Statement:**

To build a collections model for predicting the probability of a customer not paying EMI in the next month

**Data Files:**

* Training dataset ‘Training\_Data.csv’ has been provided, consisting of independent attributes, along with the target attribute (Target\_Flag) for developing the model
* Test dataset ‘Test\_Data.csv’, consisting of only independent attributes has been shared. Predictions are to be made on this data
* Attribute dictionary consisting of the description of each attribute

**Expected Outcome:**

Predict the probability of a customernot paying EMI in the next one month. We will evaluate your submission basis the AUC of your model on the test data by comparing it with the actual event rate. We will also consider the methodology used for solving the business problem.

For the hackathon we will need you to submit the following:

1. A csv file with your name and institute as the file name.   
   It should have 3 columns: “Customer\_No”, “Prediction” and “Probability\_Prediction”.   
   The 1st column is the “Customer\_No” from test data, the 2nd column (“Prediction”) is the model prediction on the test data as a binary flag (0/1 flag) and the 3rd column “Probability\_Prediction” is the model probability score on the test data.  
     
   For example, if my name is Amit Kumar and I belong to IIT Madras, the file name will be amit\_kumar\_iitmadras.csv and will look this:

|  |  |  |
| --- | --- | --- |
| Customer\_No | Prediction | Probability\_Prediction |
| 90000 | 0 | .22 |
| 90001 | 1 | .98 |
| 90002 | 0 | .49 |
| ……. | ….. | …. |

1. The final code used for the building the model and testing it needs to be shared. We expect a code with proper formatting and commenting as it will be thoroughly checked.