**Data Science Case Study**

**Time Duration:**

**Problem Statement:**

You are provided with two datasets in the attached zip archive in the CSV file format.

**The first file**, public\_train.csv, contains the training data with which you can train and validate your model.  It consists of multiple rows with the following column headers, where the class\_col column contains the correct class label for that row:

 row\_id, variable\_0, variable\_1, variable\_2, variable\_3, variable\_4, variable\_5, variable\_6, variable\_7, variable\_8,  variable\_9, class\_col

**The second file**, public\_test.csv, contains unclassified data with the following headers:

 row\_id, variable\_0, variable\_1, variable\_2, variable\_3, variable\_4, variable\_5, variable\_6, variable\_7, variable\_8,  variable\_9



**Exercise**

1. Your task is to build and train a classifier. **You will be evaluated on how well your classifier classifies this second set of data.**
2. You must submit the results of running your classifier on the public\_test.csv  data, in a CSV file as an ordered list of probabilities (to 3 decimals) that represent your model's prediction of whether each row in the test data belongs to Class 1.  Your csv submission should have two columns **row\_id** and **yhat.** In addition to the header row, there should be one row in your submission for each row in the public\_test.csv data.  Note that the row\_id is a zero-based index.  For example, if there were only 3 rows in the test data, your output might look like this:

**row\_id,yhat**

0, 0.6

1, 0.873

2, 0.22

1. In addition, you are asked to upload a Zip file containing:
   1. Brief (1 page) write-up of your approach and the tools you used as a Text, PDF, PPT or Word Doc of max-size 1MB, and
   2. Code you wrote as part of building, training, and evaluating your classifier.

**Instructions:**

1. Please complete the above Data Science exercise and submit your results and write up to the following website:
2. You should use the exact email address to which this email was sent when making your submission