Hire-a-thon Problem Statement

* **Source:** This dataset belongs to the BFSI Sector. The data has been fully masked in order to avoid any conflict of interest between the company and the Hire-a-thon.
* **Domain- specific knowledge:** We don’t want anyone to have advantage based on his/her domain specific knowledge, due to which we have even masked the columns names. The competition is designed to test your knowledge of Machine Learning in areas like Pre-processing, Feature Engineering, Feature Selection, Model Selection, Model Evaluation, Metrics etc.
* **Objective:** Accurately predict ‘Dependent\_Variable’
* **Dataset** 
  + X\_train
    - Unique\_ID: Represent the Unique Identifier
    - C1 to C8: Represents Categorical Columns, which are already encoded into numeric representation
    - N1 to N35: Represent Numerical Columns
  + y\_train
    - Unique\_ID: Represent the Unique Identifier
    - Dependent\_Variable: Represent the outcome or dependent variable
* **Metrics:** **ROC\_AUC Score** 
  + Documentation:
    - <http://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_curve.html#sklearn.metrics.roc_curve>
    - <http://scikit-learn.org/stable/auto_examples/model_selection/plot_roc.html#sphx-glr-auto-examples-model-selection-plot-roc-py>
  + Sample Code to calculate it in Python:  
    from sklearn.metrics import roc\_auc\_score, auc

fpr, tpr, thresholds = roc\_curve(y\_true, y\_score, pos\_label=1)

auc\_algo = auc(fpr, tpr)

* **Final Deliverable**
  + We will release the Testing Dataset on 24th November at 6:00pm with a sample submission file. The participants are expected to upload the following:
    - Predictions in the format of sample\_submisison\_file
    - Code
    - Solution Description (at least: 180 chars)
* **Note:**
  + ROC\_AUCuses probability estimates of the positive class for its calculation. Therefore, the final outcome is a probability estimate for the positive class (e.g. 0.82), not the class label (e.g. 0 or 1). Please refer to sample\_submission\_file.csv in case of any query.