Assignment 5
Due Date: Sunday, 31 Jan 2020

Important: Only PDF outputs and IPYNB files will be considered for evaluation.

.py files will NOT be graded: you will be awarded a zero score if you end up submitting a .py file!

Classification task: Use the HR Analytics Dataset to predict attrition.

Q1. Perform 4 visualizations either using matplotlib or seaborn. Write what you can infer from the visualization. Note: If you do not write what you can infer from the visualization then you will not score just for showing different graphs/ charts. (20 points)

Q2. Build three models using different classification algorithms after splitting your dataset into 60% train and 40% test. (30 points)

Q3. Using your trained Models predict your values of Attrition using the test data. Compare the following between the three models:

(A) Confusion Matrix: Select Accuracy, Precision, Recall and F1 score and write down what you can infer from the results- which model performs the best? (25 points)
(B) Generate ROC curve, AUC value and Gini value for your model and explain the results. (25 points)