BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
K. K. BIRLA GOA CAMPUS

First Semester 2018-19
Data Mining (CS F415)
Assignment-1
(Fraud Detection System)

One of the top banks of the country is worried about their credit card transactions and want's an expert opinion. Could you help them out?

Build a model (using clustering) so that, given an unknown tuple you can find whether it is fraudulent or not?

Answer the following questions:

1. Do you need data pre-processing? Justify your answer
2. Do you need Normalization? Justify your answer
3. Compare the clustering result(s) with the actual result(s) given in the class column. Show it using:
    a. Root Mean Squared Error
    b. Correlation between class and predicted column.
4. Which clustering algorithm could get you closest to the actual result and why?

Attribute List:

- Column 'Time' denotes the Number of seconds elapsed between this transaction and the first transaction in the dataset
- Columns V1 to V28 have been derived after applying PCA
- Column 'Amount' denotes the amount associated with the transaction
- Column 'Class' denotes 1 for fraudulent transactions, 0 otherwise

Assignment Submission Format:

A zip file consisting of the followings only:

- Portable source code (jupyter notebook):
    - Must contain all required packages/libraries.
    - Path for any required file(s) should not be local to your machine
    - Instructor should be able to run your code after direct download.
- Source Code (jupyter notebook pdf version)
    - Should contain all the intermediate steps + results to reach the conclusion
- README.txt
    - Step by step instructions to run your code.
    - ~~Download package 1, download xyz.jar, install MySQL~~
- Report in PDF format (max 2 pages. 11pt. Times New Roman.)
    - Insights, inferences, results and conclusions drawn from the assignment.
    - No source code or figures in this PDF
    - Proper references to the source code and figures.
- Figures (depends on the type of the assignment)
    - Self-explanatory caption to the figures. ~~1.jpg, q1.jpg, abc.jpg~~

Assignment Submission Policy:

- Submission accepted through **Photon only.**
- No assignment will be accepted by **email or after the deadline**.

Plagiarism: Plagiarism will be checked for every submission with Turnitin.

- The rule is very simple
- If **(Plagiarism % from Turnitin Report) > 30**
    - Will be awarded **"Component Maximum Marks * -1"**