Birla Institute of Technology & Science, Pilani
Hyderabad Campus

CS F415: DATA MINING
SECOND SEMESTER 2018-19
Assignment-3

INSTRUCTOR: DR. ARUNA MALAPATI

Submission date and time: 19th April, 2019 23:59 hrs

Maximum marks: 30

The goal of this assignment is to apply density based outlier detection algorithms to find fraudulent credit card transactions. You need to implement **both DBSCAN and Location Outlier Factor (LOF) techniques** on the chosen dataset.

Datasets:
- [Credit Card Fraud Detection](#)
- [German Credit Data Set](#)

You are free to you any other dataset provided that it is of comparable size as the datasets provided.

Programming Languages: C, C++, Java, Python

Team Size: 4

Report:
- Name and ID of team members.
- Dataset used.
- Pre-processing done on the data(if any).
- 2D Plot of the dataset for different parameters (epsilon for DBSCAN and k for LOF) highlighting outliers (NOTE: For plotting, you can use dimension reduction libraries to reduce to two dimensions and then plot the new dataset)

- Accuracy (last attribute in each row of the dataset is the actual classification as outlier or not)

Submission Files:
- Source code files
- Plots
- Report in PDF format
- README

Remarks:
- All submission documents should be zipped together and submitted to CMS through one of the group member's account before deadline. Name of the file should be DM_ASSN3_201x0xxx_201x0xxx_201x0xxx_201x0xxx.zip
- All source codes will be checked for plagiarism on Moss (for a Measure of Software Similarity). Any kind of plagiarism will lead to severe penalization.
- You are expected to demo your code and present your results as per the schedule that will be made available on CMS later.

Evaluation:
- Code & comments (15 marks)
- Output files (5 Marks)
- Report (5 marks)
- Viva (5 marks)

Please contact the following teaching assistants for any queries:

1. Harshith Thonupunoori(f20150071@hyderabad.bits-pilani.ac.in)