Machine learning Minor

ISE 2. 20 M

Due Date: 2/12/22@10a

**Credit Card Fraud Detection**

**Dealing with imbalanced dataset (Given on Kaggle data repository)**

The datasets contains transactions made by credit cards in September 2013 by european cardholders. This dataset presents transactions that occurred in two days, where you may observe 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.

**Objective:**

* Understand the imbalanced dataset and will perform various approaches like **undersampling/ oversampling, choosing right metrics of ROC- AUC** to deal with imbalanced dataset.
* After trying different approaches and training different models (LR, KNN, SVM, Random Forest, XGboost, Naivebays) you will compare their results and decide the one which fits best for your application.

Submit python code and relevant observation and conclusion in pdf format on your github account by 2nd Dec 22/10am