**Topic Submission Form**

This form should be submitted by the mentioned deadline.

**Name: Rachit Dev**

**Student Number**:

**Course: Master of Science in Data Science**

**Fill your topic/s below**

**Project Title/Area 1:** **An Approach to optimize credit card fraud detection.**

**Dataset:** https://www.kaggle.com/mlg-ulb/creditcardfraud

**Description:** Fraud is one of the major ethical issues in the credit card industry. The main aims are, firstly, to identify the different types of credit card fraud, and, secondly, to review alternative techniques that have been used in fraud detection.

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

It contains only numerical input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, we cannot get the original features and more background information about the data. Features V1, V2, ... V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature 'Amount' is the transaction Amount, this feature can be used for example-dependent cost-sensitive learning. Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise.

I have already done some part of this problem earlier and would like to take it to the next level. My GitHub project can be found here:

<https://github.com/rachitdev/MLStudy/blob/master/CreditCardFraudDetection/CreditCardFraudDetectionMadhuRachitFinal.ipynb>

**Project Title/Area 2**: **Stock Market Prediction using Time Series, Deep Learning and Sentiment Analysis.**

**Dataset**: yfinance API can directly make the dataset. Any news API such as Bloomberg or Bing can be used to generate the news which can be used for sentiment analysis in the project. My GitHub project link below uses yfinance API and generates a dataset.

<https://github.com/rachitdev/TimeSeriesAnalysis/blob/master/C365/Time%20Series%20Analysis-6%20Forecasting.ipynb>

**Description:** Stockprediction via market data analysis is an attractive research topic. Both stock prices and news articles have been employed in the prediction processes. However, how to combine technical indicators from stock prices and news sentiments from textual news articles, and make the prediction model be able to learn sequential information within time series in an intelligent way, is still an unsolved problem. In this research, we will try to build up a stock prediction system and propose an approach that

1. Represents numerical price data by technical indicators via technical analysis, and represents textual news articles by sentiment vectors via sentiment analysis.

2. Setup a layered deep learning model to learn the sequential information within market snapshot series which is constructed by the technical indicators and news sentiments.

3. Setup a fully connected neural network to make stock predictions.

In the above-mentioned link of the dataset there is a forecasting approach using Time Series Analysis which I have implemented in predicting the stock closing prices. I can take this approach to the next level.

**Project Title/Area 3: Predicting Political Instability in any country.**

**Dataset:** There are 2 datasets being used in this model:

1. <https://fragilestatesindex.org/excel/>
2. https://datacatalog.worldbank.org/dataset/neighborhood-fragility-index-2017

**Description:** Label - Fragile States Index (FSI). We obtained our high-risk classifier data from the Fragile States Index (FSI) which is released annually from The Fund for World Peace[2]; each country is given a score based on Social, Economic, Political and Military indicators from a variety of quantitative and qualitative sources where a low  
score indicates high stability (i.e. Finland is 18) and a high score indicates low stability (i.e. Somalia is 113). In order to conduct tests with binary classification, we categorized an unstable state as that having an FSI score above 100. We chose this particular ranking system because it is a popular standard in political science and the product of two reputable organizations, Foreign Policy Magazine and The Fund for Peace.

**Fill in this section if a member of staff has agreed to be your supervisor:**

Member of Staff:                                                                                   \_\_\_\_

If you have found a supervisor then you and the member of staff who agreed to supervise your project should sign below.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_                                                                        \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Student Signature                                                                         Supervisor Signature

\_\_\_\_\_\_\_\_\_\_\_\_\_                                                                            \_\_\_\_\_\_\_\_\_\_\_\_

Date                                                                                               Date