**Deliverable:**

Submit max 10 slides and your codebase. Your slides will be acting as your report, so include as much necessary information as you want, such as:

General goal of your project: obtaining insights from data, parameter tuning, algorithm improvement, etc.

2. Specific problem you chose to study: justify your choice. Why is it important in practice? You may cite research papers, white papers, or business reports to support your claims.

3. Dataset description: describe your datasets and include links if publicly available. Include information such as: the number of rows, a few sample rows, column names and datatypes.

4. Describe the machine learning that you used: what are the features, what is the class label, etc. Justify why your datasets are appropriate for training a model to solve your problem. 5. Results and comparative analysis.

5. Bibliography. Provide the title, authors, journal/conference name and page numbers of all the works you cited.

**Code Base To-do:**

* Data cleaning TO BE DONE FIRST [John Greenough](mailto:johngreenough79@gmail.com)
  + Missing values and outliers
  + Dummy variables and other transforms needed
* Data Visualization - [williamgirdwood98@gmail.com](mailto:williamgirdwood98@gmail.com)
  + Bar charts, scatter plots and histograms to visualise variables
  + Plots that differentiate between fraud and non fraud transactions
* Train Model - [Senan Gaffori](mailto:SenanGaf44@gmail.com)
  + Regression models and forest model, maybe neural net
  + Model improvement with cross validation techniques
* Model Evaluation - [DS S](mailto:dev4school23@gmail.com)
  + Confusion matrix to evaluation, precision, accuracy and recall
  + ROC and AUC measurements

**Slide Presentation To-do:**

* General Goal/ Specific Problem - [Senan Gaffori](mailto:SenanGaf44@gmail.com)
  + Obtain insights into identifying fraud
  + Identifying financial fraud
  + Explain why
  + Give motivating examples (i.e numbers, people losing, etc.)
* Data breakdown - [DS S](mailto:dev4school23@gmail.com)
  + Results
  + Predicting factors
  + Etc
* Model Selection
  + Rationale behind picking certain model
    - Probably gonna be Logistic regression
  + Techniques used and why
    - Test sets n stuff
* Model effectiveness
  + Show statistics of performance
* Bibliography - [Senan Gaffori](mailto:SenanGaf44@gmail.com)
  + List all sources
  + Talk about interesting papers/ things you found