# AI Test 1

Solve the following machine learning problems with Python Machine learning frameworks. You can do up to any level (A, B, and C) and submit the assignment. Those who are doing up to C will get better scores. Make any assumptions required to fill in potential gaps in completion of the problem statement requirements. However, clearly explain your assumptions and comment your code well.

Key assessment criteria include data manipulations, data manipulation logic,

algorithm selection/algorithm parameter choice, Network Architecture (for Neural Networks), accuracy metrics of the model, and code design.

Comment code extensively with information related to your choices: E.g. Choice of Algorithm, hyperparameters, reasons for regularization(if any), design of processing techniques, etc.

**Level A**

**Dataset:**

Transaction data in a financial network. Headers are self-explanatory.

Training dataset: Aml\_train.csv

Eval dataset: aml\_eval.csv

**Problem Statement:**

Create predictions for evaluation dataset and attach them to the final values. Show prediction scores. Explain choice of metric to evaluate classification.

**Hint:** Catching suspicious transactions is far more important that normal transactions.

* A Training program should create a machine learning model as a file for output.
* Training program should also write its accuracy results to file. (Choice of accuracy metric is important)
* Another Program should then read the created model file, and use it to predict evaluation data.

**Level B**

**Description:**

Attached is a dataset of consumer complaints to the U.S Consumer Financial Protection Bureau (**complaints.csv**). This contains a list of complaints made to the Financial Protection Bureau. The dataset and its detailed explanation can also be viewed on this link.

In many cases, consumers can give detail on their complaints, which are in the column "Consumer Complaint Narrative". As you can see close to 80% of this data is empty. You are required to filter off only the data that contains these values for the assignment.

**Problem Statement:**

Using this dataset create a text classification model that can predict the "Product" column using the "Consumer complaint narrative" data as input.The model should take a line of text as input and predict the "Product" the complaint is about.

**Level C**

**Description:**

Production level Machine learning requires models to be able to provide their predictions with APIs. In this problem, expand the original text classification into a Multi-Classification model that can predict 4 types of categories (mentioned below). You can filter the data to only classify for items that have at least a 100 examples.

**Problem Statement:**

Create a multi-intent classification model and serve it on a Python server. Specifications below:

* Using the data in "Consumer complaint narrative" column create a text classification model that can predict the "Product", "Sub-Product" ,"Issue", and "Sub-Issue".
* Serve this model on a Python server with a POST API that accepts a text body as input payload. On receiving this text, it creates a prediction and saves the prediction and text into a csv file.
* E.g: payload for API: *"Experian has allowed 6 inquiries on my credit report that I am disputing because they were not authorized. I did not receive calls from these companies even though I have a fraud alert on my credit profile..... "*
* Final CSV created by API:
* *"Experian has allowed 6 inquiries on my credit report that I am disputing because they were not authorized. I did not receive calls from these companies even though I have a fraud alert on my credit profile....."*, Credit reporting, credit repair services, or other personal consumer reports",Credit reporting,Incorrect information on your report,Information belongs to someone else