Instructions

This project is worth 30 marks and you are expected to answer all questions.

Marks are allocated as follows:
- Data handling (7 marks)
- Model development, evaluation and performance (9 marks)
- Discussions (14 marks)

Your Task

Along with this instructions, you have been provided with an **imperfect** dataset on students' academic success. Your objective is to train an ML model capable of making predictions about students academic success. Towards this, your tasks are as follows:

1. Split your data into training and testing sets.
2. Develop a model that is able to make predictions on students academic performance using the training set you have been provided.
3. Evaluate your model
4. Discuss the following:
   i. Your treatment of the provided data, covering how you split the data into training and testing datasets.
   Ii. Your observations about the data and their implications on the quality of the model being built and what you did to resolve it.
   Iii. Your choice of model and

Note: The provided dataset consists of two tabular data in CSV format. **data-description.csv** explains the data and the expected range of data while **data-export.csv** contains the labelled data.

Deliverables

1. Training and Testing Datasets: Once you split the data for training and testing. Export these datasets as respective CSV files, making it clear which file contains the training data and which contains the testing data.
2. A Jupyter Notebook: This notebook should contain your solution to the above problem and your answers to the questions above. Please ensure each section is labelled clearly.