Reminder: This is an individual project.

Data Files:

- **Project_data.csv** – Historical data to be used to create the classification model.
  In this file:
  - `V0->V11` are the feature variables.
  - `target` is what you are to predict.
- **new_obs.csv** – Contains feature data (`V0->V11`) which you will use for new predictions.

Work to be done:

1. Use the project_data.csv file to create a classification model.

2. Use your model on the feature data contained in the new_obs.csv file to make a prediction for each line contained therein.

What to submit:

You will submit 3 files. **Failure to adhere to these requirements will result in a grade reduction.**

- **report.pdf** - A technical report, which describes the process and results of creating the classification model. It should be organized in a professional manner. A good reference is:
  https://www.aresearchguide.com/writing-a-technical-report.html

- **predictions.csv** - This file should contain the lines in the **new_obs.csv** file with an additional column (`target`) appended which is the prediction for that line. i.e., The file should have the identical format of **data.csv**.
  Note: be sure I can read this file with a pandas `read_csv` statement and extract your predictions from the column labeled 'target'**. Failure will result in a significant grade reduction**

- **programs.zip** – A zip file of all python files that you used to complete the project.
  These files should be briefly described in your technical report.
  Note: I should be able to run these files and recreate your results if I desire.

Grading:

The grade for the project will be composed of two parts:

- 50% - Overall quality of the work done in creating the model and the documenting of same in the technical report. You should emphasize anything you feel is "special" and you want to be sure I notice.
  Inevitably, you will be compared to your peers so do not leave anything out of your report.
  At the same time, an overly inflated report (bloviating) will not be looked upon kindly.

- 50% - Your rank in a scoring competition.
  You will be predicting new values for `target`. A value of 1 is a positive and a 0 is a negative.
  Your predictions will be compared to true values which are known by me.
  Based upon your predictions a score will be computed according to the following formula:

  Penalty Score = 50 (number of False Positives) + 10 (number of False Negatives)

  Scores will be computed for all students and ranked in ascending order. Obviously, you want a low score.