Work Groups:

You may work in teams with a maximum of 5 people per team.
All submitted documentation should indicate the people on the team

Data Files:

data.csv – Historical data to be used to create the classification model.
numeric columns: 0,2,4,10,11,12
categorical columns: 1,3,5,6,7,8,9,13
target column: 14
futures.csv - Feature data for which you are to predict target.
In both files, question mark (?) indicates missing value

Work to be done:|

1. Use the **data.csv** file to create a classification model.

2. Use your model on the data contained in the **futures.csv** file to predict SMALL/LARGE for the each line contained therein.

What to submit:

1. A "hard copy" professional technical report, which describes the process and results of creating the classification model. This report is by far most important part of the project. You may put the report in my department mailbox if I am not available in my office.

   Again let me emphasize:
   - This is to be a hard copy. It is what I will read.
   - It should be organized in a professional manner.
     A good reference is:
     https://www.aresearchguide.com/writing-a-technical-report.html

2. A .zip file that contains the following files:
   - A file containing the prediction for each line of the **futures.csv** file.
     You file should contain the lines in the **futures.csv** file with
     an additional column appended so that your submission is similar to the
     data.csv file. Note: be sure I can read this file with the same code used to read the data.csv file.
     Failure will result in grade reduction.

   - All python files (and associated brief descriptions of same) used in the creation of the classification model and for the prediction of the "futures". Note: I should be able to run these files and recreate your results if I desire.
     These files could be part of the Appendix of your professional report.

   - A electronic copy of your technical report. This is for archival backup.
     I will not be reading this.