Building the stock market prediction Engine

Saankhya, one of the Financial Analytics company wants to build a statistical algorithm for algorithmic trading and they hired you as the data scientist to help them.

Data Description

The data contains anonymized features relating to a single financial asset. The data contains 111 columns where the targets are y1 and y2.

First column is the timestamp which indicates the day at which the event occurred

"y1" (Numeric) is the percentage change (http://www.investopedia.com/ask/answers/03/100303.asp) in asset price of w.r.t. the previous day's asset price.

"y2" (Binary) is the volatility (http://www.investopedia.com/terms/v/volatility.asp) of the asset w.r.t. previous 2 weeks – 0 implies the asset is not volatile and 1 means it is.

You are expected to build at one classification model on "y2" and one regression model on "y1" using either of regression or time-series techniques.

Where do I see the data?

Click on the link given below.

http://grader.insofe.edu.in:3939/course/CSE7302-CUTe/CUTe-Code/data.zip

Each student is part of a 3-member group. To see which team you belong to, please refer to the Viva Schedule (PGP Batch 31-CUTE20170923 - Viva schedule.xls file). Once you download the data folder, suppose you belong to say Team12, take train12.csv from train folder and test12.csv from test folder and start!

Submission Instructions

1. Where do I submit?
   You can submit at grader.insofe.edu.in:3939. The course-ID is CSE7302-CUTe and Task is CUTe.

   http://grader.insofe.edu.in:3939/course/CSE7302-CUTe/CUTe-Code/data.zip

2. What should I submit?
   You need to submit the R code and a presentation in the PDF format.
3. Should I submit multiple times before submission?
   Yes, you are encouraged to do so to see the progress.
4. What is the submission format?
   The detailed instructions are given the tool itself. Please refer to them.

Deadline to submit: **23rd September, Saturday 12:00 Noon**

What happens after submission?

Prepare for Interview and refer to the schedule given.

What is the evaluation criteria?

We will evaluate on the following metrics.

S.No	Metric
1	Translation of business or technical case into Analytics or ML problem
2	Variety of models
3	Model performance and justification for the choice of the error metrics
4	Data Handling
5	Code correctness and Redability: Ability to code the model effectively
6	Code Readability: Functions, comments, etc.
7	Smooth flow of presentation and thought
8	Presentation Appearance: Font consistency, spell check, clean and informative slides
9	Project related Questions
10	Questions related to the CSE7302c module

What if you are absent to the Viva?

We will score your performance accordingly

What if you are not participating in the CUTe?

It is difficult to create such an exciting environment again. So, please ensue you have 100% participation. We will not reconduct the exam.

Questions related to Data and Algorithms

1. There are missing values in the data-frame, what do I do?
   It is up to you how to handle.
2. I'm getting a 404 error on the portal.
   Change localhost:8080 with grader.insofe.edu.in:3939
   E.g. at a 404 check your URL. If it looks like http://localhost:8080/course/CSE7135c/Probability
   change it to http://grader.insofe.edu.in:3939/course/CSE7135c/Probability
3. What's the viva format?
   Each team shall submit a .pdf of their presentation along with the R/python code.
4. What should be the contents of the presentation?
   The allotted presentation time is 5 minutes. Ensure the presentation reflects all your work.
5. Will the deadline be extended?
   Since this is automated tool, we don't have control to extend the deadlines. Ensure you submit them on time