CS - Stock Market Prediction

Rubric

DS 4002 – Fall 2023 - Glenn Hogan

Submission format: Upload link to github repo and PDF to canvas

Individual Assignment

General Description: This assignment will have you practice predictive modeling by using stocks as the data.

Prerequisites - Introduction to Coding course in R or Python, or familiarity with either

Why am I doing this?

The goal of this assignment is to put you outside your comfort zone, work on experimenting with unfamiliar models, and make some kind of deliverable regardless of outcome. As a data scientist, often the parts of the real world you are asked to model aren't as simple as toy problems that are often given in classrooms. They often involve researching the current cutting edge or generally accepted models and trying to convert them for your own uses. Also, in future work or research you may be required to work with unfamiliar technologies or domains that will require you to go outside your comfort zone and experiment with tools you haven't worked with before. The results will also be presented in a document detailing your work and the steps you took. Being able to present your findings, successful or unsuccessful in a professional and understandable manner is crucial as a data scientist. This is your chance to show off the skills you have honed so far and to see where you can make further improvements for the future.

- Course Learning Objective: Build a model based on given data set
- Course Learning Objective: prepare findings and document them

What am I going to do?

First read the hook document within this folder, then read through the document explaining algorithmic trading within this folder. You may or may not want to read the article about the ARIMA model, depending on if that is the model you decide to do. That article is purely there to support you and provide some idea of what your code may look like, get as creative as you would like with your model. Outside research is encouraged! After that, familiarize yourself with the rubric. Then, you are going to use the provided data of 6 different stocks from NASDAQ, you may use more if you like, and build any machine learning or data science model you would like (feel free to get as creative as you wish), to predict the stock prices for each form 11/2022 to 11/2023. Do analysis on these predictions and create a document detailing your process and teh steps you took to find your results.

Deliverables include:

- Multipage PDF detailing your findings and the process you used to get them, including; hypothesis, data cleaning, analysis, conclusions, and next steps.
    - Cite all sources used within this PDF
- Online repository including figures and all used code

All of this will be submitted electronically via Canvas

Tips for success:

- Be bold. This is your chance to try something creative and fun in a low stakes environment
- Don't dwell on failures. This is a hard problem, one that remains unsolved, be ready to have models not work or not do as well as you like.
- Don't overthink it. A clear presentation of fundamentals is more valuable than an unclear presentation of cutting edge techniques.
- Talk to the professor and the TA. This is a creative assignment, and you are allowed to show ideas to people for comment.
    - Use them for ideas or trying to get yourself unstuck if you run into problems throughout the assignment
- Don't get lost in the sauce. If you end up overwhelmed by how difficult the code and assignment seems, work through it one step at a time. Models don't get built overnight.
- Do research. If you don't know where to start or have an idea but don't know how to execute it, use online resources, there is tons of pseudo code and useful documentation by very smart people that can help you out.

How will I know I have Succeeded? You will meet expectations on this CS

Spec Category	Spec Details
Formatting	● One Github Repository (submitted via link on canvas) ● Top level page should contain ○ A SRC folder ○ A DATA folder ○ A FIGURES folder ● Multipage PDF Document ○ Introduction/Hypothesis ○ Data Cleaning/Preprocessing ○ Analysis Plan ○ Tricky Analysis Decision ○ Results and Analysis ○ Conclusions and Next Steps ○ References
GitHub	- Goal: This repository should contain all source code, along with all data files and figures generated throughout the code. - SRC folder: all source code used in the project - DATA folder: include the data given, along with any additional supplemental data curated and used throughout the project - FIGURES folder: All figures generated by project - Include good titles and axis so the figure scan be read, or include a README.md detailing what the figures mean
PDF Document	- Goal: Multipage PDF document explaining the different steps you took to complete the case study along with analysis about the results, and the conclusions drawn from it. Should be professional enough you could show your boss at a trading firm why they should consider your idea. - Introduction/Hypothesis: Give context to the problem, describe some of what you learned from the explanation article or any other understanding you gained from outside research. Include your hypothesis and the goals for your model. - Data Cleaning/Preprocessing: Describe the steps you took to modify or clean the data, and why it was necessary to take those steps.

	- Analysis Plan: This should be the start of the meat of the document, should include details about the kind of model being used and how you plan on evaluating the results you get from using this model. Include a specific quantifiable goal that you can use as a finish line for your analysis. (Essentially this is your goalpost, it helps you to know when you have achieved your goal). Should let the reader know exactly how you went about executing your case study - Tricky Analysis Decisions: May be included in your analysis plan. Identify and describe a decision you made in the analysis that was non-trivial. Think about a judgment call you had to make in your project and describe why it matters and why it was tricky. - Results and Analysis: Address the hypothesis and talk about the performance of the model. Detail how you found these results and work through what they mean. No need to repeat the analysis plan, but go in depth about what these results mean to you. - Conclusions and Next Steps: Give your closing thoughts and overall main takeaways, take the opportunity to talk about what you improve in the future or what you would change and tweak if you had to do the case study all over again - References: Use IEEE Documentation style

Acknowledgements: Special thanks to Jess Taggart from UVA CTE for coaching on making this rubric. This structure is pulled directly from Streifer & Palmer (2020).