CS Rubric – Stock Prediction

DS 4002 - Fall 2023 - Justin Brady
Due: End of Semester
Submission Format: Upload link to github repository

Individual Assignment

General Description: Submit to canvas a link to your github repository

Preparatory Assignments - Everything in the course

Why am I doing this? This assignment is a way for you to take what you have learned in data science and apply it to a real world scenario. It is also a way to practice synthesizing information provided to you by another source and using it to help you succeed. In the real world you'll often have to operate on what you are told by others, so this will be a useful skill down the road. You will also be able to practice communicating the results of your work. Not just building the model, but discussing how it can be used is a communication skill that can always be improved.

What am I going to do? First read the attached hook document to understand the context around the project. Take some time to read over the provided reference material and gain a deeper understanding of how the model you are working with works, especially if this is your first time working with time series data. Then move on to the provided data and use it in producing the following:

- Graphs displaying the stationarity, seasonality, and autocorrelation of each stock index
- A document containing discussion of each of the graph and how each will be used in building the model
- A graph showing the prediction of the future stock prices over the final year of the data
- A two page document discussing the accuracy of the predictions, and how you would use your model to make trading decisions in the future

All of this will be submitted electronically via a link to a github repository.

Tips for Success

- Take time to read the materials provided and gain a deep understanding of the context of the project.
- Don't be afraid to do some of your own research as well to expand upon what has been provided.
- Use your peers, professor, and teaching assistants as resources when you don't understand something. Never be afraid to ask for help!
- Be confident in your analysis. You don't have to be perfect, but make a decision and be ready to defend it in your discussion of how you would use your model

How will I know I have succeeded? You will meet expectations on CS - Stock Prediction when follow the criteria in the rubric below

Spec Category	Spec Details
Formatting	● Repository - A github repo containing the following materials ○ A DATA folder ○ A PREPROCESSING folder ○ A SRC folder ○ An ANALYSIS folder ○ A README.md file (which auto displays)
DATA	● Goal: This folder contains all of the data for this project ● If your data fits in github place all of it here ● If your data doesn't fit in GitHub use a single file explaining the process to obtain the dataset ● You should include the provided data along with any other data you wish to include to conduct your analysis
PREPROCESSING	● Goal: This folder contains your graphs and analysis before building your model. The contents should be as follows: ○ Stationarity graph (.png) ○ Seasonality graph (.png) ○ Autocorrelation graph (.png) ○ Analysis document (.pdf) ● Stationarity graph: ○ A graph displaying the stationarity of the data (one per stock ticker) ○ Should have a proper title, proper axes labels, and the file should be named with the format ticker_stationarity.png (i.e. AAPL_stationarity.png) ● Seasonality Graph: ○ A graph displaying the seasonality of the data (one per stock ticker) ○ Should have a proper title, proper axes labels, and the file should be named with the format ticker_seasonality.png (i.e. AAPL_seasonality.png) ● Autocorrelation graph: ○ A graph displaying the Autocorrelation of the data (one per stock ticker) ○ Should have a proper title, proper axes labels, and the file should be named with the format ticker_autocorrelation.png (i.e. AAPL_autocorrelation.png) ● Analysis Document ○ Pdf document assessing the stationarity, seasonality, and autocorrelation ○ Explain the tests and how they impact building an ARIMA model at a high level

	○ For each stock ticker describe what conclusions you draw about the data. Is the data stationary or not, does it have seasonality, and is it highly autocorrelated or not. ○ Also describe how you came upon the lag value for each of the stocks ○ Explain how the results of these graphs influence the code you wrote to build the model ○ Document should be named analysis.pdf
SRC	● Goal: This folder contains all the source code for your project ● This should include well commented .r or .py files you used to produce the graphs in PREPROCESSING, the model for analysis, and the final prediction graph in ANALYSIS ● Make sure to include in the comments how you arrived on the p, d, and q values ○ This could be through trial and error, using premade functions like auto_arima, or another method you find through research ● Code can be split by stock or all contained in all in one file ● The high-level documentation for this code lives in the README.md
ANALYSIS	● Goal: This folder contains the stock prediction graphs as well as your analysis document discussing results. The contents should be as follows: ○ Stock prediction graph (.png) ○ Analysis document (.pdf) ● Stock prediction graph ○ This graph should display the predictions of stock price over the final year of the provided data ○ It should also include the actual prices of the data over that period for purposes of comparison ○ Be sure to include proper axes labels and a proper title ○ The file should be named ticker_prediction.pdf (i.e. AAPL_prediction.pdf) ● Analysis document ○ PDF document assessing the results of the predictions, providing analysis on their accuracy, and discussing how the results can be used in future trading decisions ○ Use the graph of each stock ticker to discuss how the overall predictions fit into the actual results over the course of the year. Include the accuracy of the model as well as the error bounds that the model falls inside ○ Discuss if this model should be used to make trading decisions in the future. ○ If it should discuss the format those trading decisions should take. I.e. should a person use it to only conduct day trades, or should they use it to take long term investments on stocks ○ If the model should not be used, include information as to

	why not. ○ Discuss potential improvements that could be made to the model to improve its accuracy or validity in making decisions. ○ Be sure to provide evidence when answering the questions from both the graph model, prediction accuracy, and information from your research
README.md	● Goal: This file serves as an orientation to everyone who comes to your repository, it should enable them to get their bearings. ● Use markdown headers to divide content ● Preprocessing section ○ Short description of the files contained in this folder ● Data section ○ Data dictionary ● Src section ○ Description of code setup/installation ○ Description of code usage (including any edits that must be made when running code) ● Analysis section ○ Short description of the files contained in this folder ● References section ○ Include all references used in completing this project in IEEE documentation style