NUS Fintech Society (ML) Project 1: Stock Price Prediction

Start Date: 10 October 2200Hrs
End Date: 24 October 2359Hrs (NO extension due to the nature of project)
Learning Objectives:
- Time series analysis and application of ML models
- Some CI/CD considerations
- Portfolio building
- Economic indicators (inclusion is welcome but not compulsory)

Task: Predict the closing stock price of NASDAQ: NVDA for trading days from 25 October - 7 November (inclusive). There will be no restrictions on the data sources used (e.g. you may use relevant macro-economic indicators). Predictions will be compared with actual (after 7 November, but you have to submit before 24 October) using RMSE. The lower the RMSE, the more accurate the prediction.

You may find the historical price of NVDA here:
https://finance.yahoo.com/quote/NVDA/history?p=NVDA

Basic expectations:
- Good coding practices (documentation etc)
- Explanation on the data sources used

Implementation 1: Machine learning models
- Ipynb or pdf should include:
    - why that particular model was chosen
- Example of ML Models: SVM, Deep Learning, LSTM etc.

Implementation 2: ARIMA models
- Ipynb or pdf should include:
    - why that particular ARIMA model(p,d,q) was chosen over others

Your submissions will mainly be graded on RMSS. If you included economic indicators, for example, the impact that interest rate increases have had on the stock market (https://www.investopedia.com/investing/how-interest-rates-affect-stock-market/), you can be given bonus marks.

Submission details:
1. Upload predictions csv on Kaggle [JOIN COMPETITION USING LINK]
    a. https://www.kaggle.com/t/9543b6e890f749eab78f551d389f6e23
    b. Sample submission format can be found on Kaggle
2. Submit Github link on this Google Form:
    a. https://forms.gle/VSvxmyg1GfE3jZwk8
    b. Github repo should include:
        i. Ipynb file name for implementation 1: <FULLNAME>_Impl_1.ipynb

       ii.   Ipynb file name for implementation 2: <FULLNAME>_Impl_2.ipynb
      iii.   [Optional] pdf on what method or data was used / any interesting considerations / model comparisons.

Follow the naming conventions strictly.

QnA:

1. What should we record in our documentation - are there any key things to explain or any best practices?
   a. In terms of coding, google documentation style works (when defining functions etc); in terms of this project, we are expecting justification on why a certain model is chosen and other considerations when introducing datasets
2. is there any sample csv for input data?
   a. No, we would like you to source your own datasets that you think can give the best predictions, similar to what you would do when building your own trading model. We would however be providing a sample submission csv on Kaggle for standardization.
3. by deploying model in real time, do you mean like so we prepare the model beforehand, then on 21 oct for example, we retrieve a more recent set of historical data to start training the model In preparation of the 25 Oct to 7 Nov period's price?
   a. Yes, exactly!
4. Would you recommend automating the ARIMA model selection? Are there situations where human selection would be better?
   a. I would recommend automating to find an order that you can compare with your manually selected one. I believe in some cases human selection might be better.