

Final Project

Time-series data and application to stock market

Le Nam Dong

May 28, 2023

I. Overview

The stock market is a highly complex and dynamic system that is influenced by many factors, from macro factors, i.e. seasonal trend, pandemic, economic indicators to micro factors like company's performance. Given its unpredictability and non-linearity, predicting stock prices has been a challenging task. One approach to this problem is using time-series analysis, which lays a foundation for forecasting. By gathering patterns in the time-series data, I hope to gain insights into the price movement of the stock market, along with identifying the trading points or portfolio management of the chosen market. The primary objective of the project is to apply time-series analysis techniques to predict stock market prices. Specifically, the report aims to answer the following main questions: Nasdaq stock price prediction on a time window, Vietnam stock price prediction on a time window, Vietnam trading point identification, and Vietnam risk management with deeper sub-questions. Also, I would include the decision-making process for the solution of each question where applicable.

The report would contain the following sections: Data Preprocessing & Exploratory Data Analysis, which takes a brief look over the datasets, Modeling Techniques to answer above questions, Results & Challenges that I discover during the project, and a final Conclusion for key takeaways that I have learned throughout the project.

II. Data Preprocessing

Regarding the dataset, both Nasdaq and Vietnam datasets have the same set of attributes, including Open, High, Low, Close, Volume, TradingDate, along with an AdjustedClose field for Nasdaq data.

To ensure the dataset quality and system capacity for the analysis, I have performed several preprocessing steps as below:

- Filter popular tech companies (Nasdaq)
- Filter Records from the last 2 years (Nasdaq & Vietnam)
- Handling Missing and Infinite Values (Nasdaq & Vietnam)
- Filter out new companies that operate less than a year (Vietnam)
- Exploratory Data Analysis (Nasdaq & Vietnam)
- Time Window for Data Feature (Nasdaq & Vietnam)
- Split dataset (Nasdaq & Vietnam)
- Normalization (Nasdaq & Vietnam)

III. Modeling Techniques

- Model Architecture: The model architecture used for both Nasdaq and Vietnam datasets consists of LSTM (Long Short-Term Memory) layers, dropout layers for regularization, and a dense output layer.
- Model Training & Evaluation: The model is compiled with the Mean Squared Error (MSE) loss function and trained using the training data. Early stopping is implemented to monitor validation loss and stop training if the model does not improve over epochs. After that, MSE is calculated on the test set as the evaluation criterion. Keep in mind that the test data is also normalized for the model to work.

IV. Results & Interpretations

In this section, I would like to address the limitations and challenges encountered during the project, including the issues with the model and the impact on the results.

Additionally, I would like to express my accountability to the situation and the willingness to learn from this experience.

- **Model Performance and Unexpected Errors**

Despite my initial progress in achieving low Mean Squared Error (MSE) values during the development of the model, I have encountered unforeseen errors at the last minute when I re-run the whole program. The MSE and loss values for all epochs across all my models increase very high abnormally. Unfortunately, I was unable to identify and troubleshoot the errors within the given time.

- **Investigation into Errors**

Upon reflection, I suspect that the errors may be related to the implementation of MinMax normalization. I encountered error messages that seemed to be associated with this step. While I did not have sufficient time to review each line of codes for any mistakes I could have made, I plan to thoroughly review this after the submission deadline to understand the root causes of the errors.

- **Impact on the Project**

The errors had an immense impact on the project, making all of my models unusable with totally inaccurate predictions. As a result, the outcomes and conclusions drawn from the project are incomplete and do not meet the expected standards of the project.

- **Lessons Learned**

This experience has taught me a lesson in the importance of time management and avoiding rushed work on coding projects. Working under time pressure limited my ability to address the difficulties encountered and seek timely support. Moving forward, I will plan ahead for a coding project and allocate enough time for testing, troubleshooting, and seeking assistance when faced with similar challenges.

- **Accountability and Responsibility**

I take full responsibility for the mistakes and errors that occurred during the project, even though they were unintentional. I would hope to get your valuable feedback on my code and methodology, if you have enough time to review it. Your insights will help me identify the underlying issues and try to improve it again in the summer and definitely help me in the upcoming projects.