

Capstone Spring 2023:JPX Tokyo Stock Exchange Prediction

Assignment: Project Introduction

February 2023

1 Background

Success in any financial market requires one to identify solid investments. When a stock or derivative is undervalued, it makes sense to buy. If it's overvalued, perhaps it's time to sell. While these finance decisions were historically made manually by professionals, technology has ushered in new opportunities for retail investors. Data scientists, specifically, may be interested to explore quantitative trading, where decisions are executed programmatically based on predictions from trained models.

2 What is the problem?

There are plenty of existing quantitative trading efforts used to analyze financial markets and formulate investment strategies. To create and execute such a strategy requires both historical and real-time data, which is difficult to obtain especially for retail investors. This project will provide financial data for the Japanese market, allowing retail investors to analyze the market to the fullest extent.

3 Why is it interesting and important?

There are plenty of reasons for predicting stock exchanges. This is an important field because it has the potential to drive financial gains, mitigate risks, stabilize the economy and advance technological deviance gains.

One of the primary reasons people are interested in predicting stock prices is for financial gain. Accurately predicting the movement of stocks can help investors make better decisions and increase their profits.

It also helps with risk management. Accurately predicting the stock market can also help investors manage risk by allowing them to adjust their portfolios based on market conditions. This can help prevent losses and stabilize investments.

It is needed for economic stability. The stock market is often seen as an indicator of the overall health of the economy. Accurately predicting market movements can help policymakers and economists make better decisions and take actions to prevent economic instability.

Predicting the stock market requires the use of advanced technologies and techniques, such as machine learning and artificial intelligence. This has led to significant advancements in these fields, which have potential applications beyond stock market prediction.

4 Why is it hard?

The stock market is a complex system with many variables, including economic indicators, political events, and company-specific factors. It can be difficult to accurately capture and model all of these factors, making it challenging to predict future market movements.

The stock market is also inherently unpredictable, with a large degree of randomness and volatility. Even small changes in market sentiment or investor behavior can lead to significant price swings, making it challenging to predict future price movements.

Historical data can be a valuable tool for predicting future stock prices, but the data can be limited, and it may not accurately reflect future market conditions. Additionally, unexpected events can occur that can't be predicted based on past data.

Stock market movements are also influenced by human behavior, including emotions such as fear, greed, and panic. These factors can be challenging to predict and can cause prices to deviate from what might be expected based on fundamental analysis or historical trends.

Overall, stock market prediction is a challenging task due to the complexity of the market, randomness, limited data, and human behavior. While there have been advances in technology and techniques for predicting the stock market, accurate predictions remain difficult to achieve.

5 What is your plan?

With the given datasets, my goal is to predict the direction of the stock exchange. I will then choose a machine learning algorithm either linear regression, decision trees, random forests, or neural networks. I also have to split data into training and validation sets. Use the training set to train the model and the validation set to evaluate its performance. And then train and validate the model, fine-tune the model and then test the model.

Stock exchange prediction is a difficult task, and no model can predict the stock market with 100 percent accuracy. However, with careful data preprocessing, algorithm selection, and model training, I can create a model that can make reasonably accurate predictions.

Here are the datasets provided by Kaggle: `stock_prices.csv` The core file of interest. Includes the daily closing price for each stock and the target column.

`options.csv` Data on the status of a variety of options based on the broader market. Many options include implicit predictions of the future price of the stock market and so may be of interest even though the options are not scored directly.

`secondary_stock_prices.csv` The core dataset contains on the 2,000 most commonly traded equities but many fewer liquid securities are also traded on the Tokyo market. This file contains data for those securities, which aren't scored but may be of interest for assessing the market as a whole.

`trades.csv` Aggregated summary of trading volumes from the previous business week.

`financials.csv` Results from quarterly earnings reports.

`stock_list.csv` - Mapping between the `SecuritiesCode` and company names, plus general information about which industry the company is in.

<https://www.kaggle.com/competitions/jpx-tokyo-stock-exchange-prediction/data>

6 What are the performance metrics?

I don't have a decided performance metric, but I will choose from some common performance metrics that are used in stock market prediction.

Mean Squared Error (MSE): This metric measures the average squared difference between the predicted values and the actual values. It is a common metric for regression problems, where the goal is to predict a continuous value, such as stock prices.

Root Mean Squared Error (RMSE): This metric is like MSE, but the square root is taken to make the units of the error the same as the original values. **Mean Absolute Error (MAE):** This metric measures the average absolute difference between the predicted values and the actual values. It is another common metric for regression problems.

Directional Accuracy (DA): This metric measures the percentage of predictions that are correct in terms of the direction of the market movement (up or down).

7 GitHub Repository

<https://github.com/ssarowar/Stock-Exchange-Prediction>

8 References

[?] JPX Tokyo Stock Exchange Prediction. Kaggle. (n.d.). Retrieved February 16, 2023, from <https://www.kaggle.com/competitions/jpx-tokyo-stock-exchange-prediction/data>

[?] Biswal, A. (2023, February 16). Stock price prediction using machine learning: An easy guide: Simplilearn. Simplilearn.com. Retrieved February 16, 2023, from <https://www.simplilearn.com/tutorials/machine-learning-tutorial/stock-price-prediction-using-machine-learning>

[?] Yates, T. (2022, July 13). 4 ways to predict market performance. Investopedia. Retrieved February 16, 2023, from https://www.investopedia.com/articles/07/mean_reversion_martingale.asp

[?] Agrawal, R. (2022, July 21). Evaluation metrics for your regression model. Analytics Vidhya. Retrieved February 16, 2023, from <https://www.analyticsvidhya.com/blog/2021/05/know-the-best-evaluation-metrics-for-your-regression-model/>