Asset Pricing Research Proposal

Tia Kungwani, Andrew Kaminer, Siddharth Vijay

We chose to replicate the article "Stock Closing Price Prediction using Machine LearningTechniques". In this article, the focus is on predicting the closing price for five companies: Nike, Goldman Sachs, Johnson and Johnson, Pfizer and JP Morgan Chase and Co. This data for this article was collected from Yahoo Finance and included each stock's High, Low, Open, Close, Adjacent close, Volume, and closing price. In addition, the following variables were considered for the model:

1. Stock High minus Low price (H-L)
2. Stock Close minus Open price (O-C)
3. Stock price's seven days' moving average (7 DAYS MA)
4. Stock price's fourteen days' moving average (14 DAYS MA)
5. Stock price's twenty one days' moving average (21 DAYS MA)
6. Stock price's standard deviation for the past seven days (7 DAYS STD DEV)

For this research, the authors chose to use an Artificial Neural Network Model (ANN) and a Random Forest Model (RFM). The dataset was trained for each company from 04/06/2009- 04/03/2017 and was tested for the timeframe 04/04/2017 – 04/05/2019. After training and testing each model, the ANN model seemed to perform better as it gives Root Mean Squared Error (RMSE): 0.42, Mean Absolute Percentage Error (MAPE): 0.77 and Mean Bias Error (MBE): 0.013.

We chose to replicate this study in particular because we were initially looking to replicate a model that a trader could use to predict stock price changes concurrently, and could invoke this algorithm when making trading decisions. Furthermore, we were looking to use a regression model to analyze stock price changes compared with its historical data (this is done through the Random Forest Model). As a result, all of these criteria are met in this article and we find it extremely interesting and applicable to the current stock market.

For this article, we plan on changing the five companies listed above to companies that have a higher volatility in today's market. Furthermore, we will obtain data from these companies using various sources, such as XBRL, Yahoo Finance, etc.

In order to evaluate the effectiveness of the models, we'll conduct a comparative analysis on the two techniques applied to the five different sector companies. Predicted closing prices are subject to Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Mean Bias Error (MBE), and the model minimizing these errors will prove to be more accurate. Therefore, we anticipate that the ANN model will prove to be a better estimation technique as it is likely to give lower root mean square error (RMSE) and mean absolute percentage error (MAPE) values, indicating that it is more accurate and efficient in predicting stock closing price.

Source: Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock closing price prediction using Machine Learning Techniques. Procedia Computer Science, 167, 599–606. https://doi.org/10.1016/j.procs.2020.03.326

Individual Interests

Andrew Kaminer:
My particular interests lie in finding novel ways of outperforming the market: the area with the most potential for generating alpha is machine learning. Thus, I am involved with constructing one of the machine learning models. I am also in charge of making sure that our results match the results of the original paper and presenting those results in a way that is both useful and aesthetically appealing. I hope to go into quantitative financial research, so the ability to construct models, produce results, and present those results in a meaningful way is an invaluable skill for me to work on. By plotting useful metrics, I hope to also gain a better understanding of statistics, which I can leverage going forward into bigger projects.

Siddharth Vijayasankar:
My interests are in performing statistical analysis on the dataset that will be used to train and test the ML model. I am interested in doing this using python and using the python libraries, including pandas and numpy. Furthermore, I am also interested in learning about and contributing to both ML models (Artificial Neural Network and Random Forest Model). I plan on helping build both models and connecting the ML models with the datasets to make sure the output is appropriate, applicable, and meets the anticipated expectation of what is to be displayed.

Tia Kungwani:
My particular interests lie in working with specific machine learning techniques and algorithms to analyze and help predict the direction of the stock market prices. Therefore, my individual role in this project is going to revolve around conceptually understanding how specific machine learning models work and the best ways to develop them. I'll also specifically be involved in developing the neural network model and helping tune our models to fit our particular study. Through this work, I hope to gain a better understanding of machine learning techniques and models as well as of the statistical analysis methods that are used to train the model.

6 Week Plan

Week 10:
Tia: Get a conceptual idea of how to make the models
Sid: Collect data
Andrew: Look at study results and figure out how to compare our results from the study's

Week 11:
Tia: Start work on neural network
Sid: Clean the data + Descriptive statistics
Andrew: Start work on decision tree

Week 12:
Tia: Continue working on neural network
Sid: Jump on to whichever model needs the most work
Andrew: Continue working on decision tree

Week 13:
Everyone: First attempts at replicating paper and plotting

Week 14:
Tia: Tune models
Sid: Tune models
Andrew: Make the plotting good and making sure we replicate the main results of the paper

Week 15:
Tia: Slideshow
Sid: Report
Andrew: Report