Stock Market Analysis

Stock Price Movements Based on Company Performance

Kristian Montoya
CSCI 4502
University of Colorado
Boulder, CO

Zaki Kidane
CSCI 4502
University of Colorado
Boulder, CO

INTRODUCTION

Understanding and predicting stock market movements is useful for individual stock market traders as well as investment firms and banks. So far, the task of successfully predicting stock prices has been done by people. We ask the question, "Can Machine Learning algorithms effectively outperform the average trader?" and seek to answer that question. By introducing the performance history of different stocks, we can visualize the data and look for correlations between stock attributes/features to help determine the current/future performance of any given stock. We intend to store the collected data in a dataframe for efficient retrieval and manipulation to be used by a machine learning model (possibly a linear regression).

RELATED WORK

There have been two main approaches with regards to stock market analysis and trading: fundamental analysis (also known as the 'Warren Buffet Style') and technical analysis (*Picasso et al., 2019)*. Fundamental analysis focuses on the company performance, financial conditions, operations and macroeconomic indicators to help decide whether to buy stocks in a company or not. This method of investing is generally not used for short trading spans as day-trading (where one buys and sells stocks within a day) or swing-trading (holding on to stocks for up to 20 days). It is used for long-term investment in a company stock for a year or more. On the other hand, technical analysis depends on historical stock price trends and makes the assumption that the trends generally repeat. This method is used for short-term, high frequency trading.

Some researchers have used technical analysis to predict price movements (*Chervello-Royo et al, 2015., Patel, 2015*). Some use fundamental analysis methods by mostly finding correlations between the company's yearly reports' features and the price movements, such as (*Chen et al, 2017)*. There are also other methods that have been attempted, such as the use of sentiment analysis from social media (*Nguyen,* 2015). The approach on this paper will be fundamental analysis or a hybrid of both fundamental or technical analyses.

PROPOSED WORK

The dataset we will be using throughout the project is the "New York Stock Exchange" that we found on Kaggle. The dataset is made up of four different csv files that relay an abundance of information. The first and main set of the data includes 79 different features that gather yearly data of companies over the span of four years. These features are basic information of the well-being of a given company and include several key components, including: cash ratio, gross profit, earnings per share, etc. Many of these features may prove to be a driving force behind the company's stock price at any given time. However, it is extremely likely that many of these features will prove to be completely useless and will only hurt our model's performance by adding unnecessary noise. To side-step this problem, we will find correlations between features and the overall stock price and reduce the dataset as we see fit. Doing so should allow our model to increase in performance and give more accurate results when testing our model.

The second and third dataset are composed of opening and closing prices of a given company over the course of a year that was recorded on a daily basis (over weekdays). Conveniently, the company's ticker symbol (TSLA corresponds to the company, Tesla) is consistent throughout all the datasets. This will allow us to build and collect all data in a single dataframe (or however we decide to collect/organize the data) to better understand the overall performance of a company. By combining the data of a given company from all the datasets, a better picture should be formed and connections/correlations should be easier to find/spot. Combining the datasets together in some coherent fashion will be key in boosting our model's overall performance; the company's annual performance and state of wellbeing, paired with daily opening/closing prices collected over several years, should give some indication on where the company stands performance wise.

The final dataset should introduce some interesting concepts that drive a company's stock price; the company's sector, sub-industry, and location are placed in this final file. This should allow us to section up companies based on their industry or sector to find trends over certain periods of times. For instance, we may find that companies within a certain industry have similar stock performance over a specified period; in other words, we should be able to identify trends over sectors/industries. This is something we could show off with visualizations and graphs; plotting relevant companies' features (whichever ones we decide to show off/compare).

We also need to engineer a feature so our implemented model has a goal feature to train/test off of. To do this, we were thinking of looking at whether or not a stock's price rises or shrinks over a period of time; feature value 1 means we should buy the stock; feature value 0 indicates that we should NOT buy the stock.

EVALUATION AND MILESTONES

As mentioned earlier, by combining the datasets in a clear manner should allow us some flexibility when cleaning and presenting the data through visualizations. Once we appropriately combine, reduce, and normalize our data we can start focusing on selecting an appropriate Machine Learning model that fits our data and would give the best results after training/testing. Such models could include linear regression, random forest, KNN (possibly based on sectors/industry), or a gradient boosting algorithm. We will most likely implement the random forest model to start with before testing other possible models. Of course, this will include hyper tuning different parameters within each model to extract better results, but this will NOT be the main focus of the project; our main goal is finding a concise way of combing, cleaning, and readying our data to find interesting trends. Once we train and gather the test results for each model, we can directly compare the accuracy between the models to find which one best suits our goals and grants the highest performance.

The performance we are currently aiming for is to predict whether prices will rise or fall (that is, whether a stock is a good buy or sell) in a given time frame accurately **85%** of the time. We came to that accuracy rate as a middle ground between generating enough profits (maximized at 100% accuracy) and realistic expectations for the predictability of the stock market, which can be based on other features we will not include such as the overall economic health of the Unites States economy.

REFERENCES

- Chen, Y., Chen, Y., & Lu, C. (2017). Enhancement of stock market forecasting using an improved fundamental analysis-based approach. Soft Computing, 21(13), 3735-3757. doi:10.1007/s00500-016-2028-y

- Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. Expert Systems with Applications, 42(24), 9603-9611. doi:10.1016/j.eswa.2015.07.052

- Picasso, A., Merello, S., Ma, Y., Oneto, L., & Cambria, E. (2019). Technical analysis and sentiment embeddings for market trend prediction. Expert Systems with Applications, 135, 60-70. doi:10.1016/j.eswa.2019.06.014

- Cervelló-Royo, R., Guijarro, F., & Michniuk, K. (2015). Stock market trading rule based on pattern recognition and technical analysis: Forecasting the DJIA index with intraday data. Expert systems with Applications, 42(14), 5963-5975.

- Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. Expert systems with applications, 42(1), 259-268.