

Stock Price Forecasting with Time Series and Sentiment Analysis

Chance Robinson¹, Nnenna Okpara², Michael Burgess³

¹Master of Science in Data Science
Southern Methodist University
Dallas, Texas USA
{chance, nokpara, mwburgess}@smu.edu

Abstract. This paper seeks to explore the use of traditional time-series based methods for prediction of future stock prices and combine them with sentiment signals that might be available in social media and other web content. For example, tweets for a particular company or ticker and news headlines from financial sources such as the Wall Street Journal, Reuters, etc.... Our scope will be limited to 10 of the largest stocks in the NASDAQ Composite index. One way in which we hope to contribute to the research that has already been done in this domain is to leverage resources that allow you to analyze current day price and news feeds and utilize an online inference mechanism to make near real-time predictions. We hope to prove that the ability to act on new information as it happens can contribute to a better performance and more accurate predictive model.

1 Introduction

Stock Markets have been a cornerstone of personal investments for many Americans for over a century now. The market capitalization of domestic companies in the U.S. is over \$40 trillion dollars as of 2020 according to the World Bank.

A Gallup poll conducted in August 2021 found that that 56% of Americans own stock, which was consistent with prior measures taken in 2019 and 2020. This is slightly less than the levels that preceded the 2008 financial crisis. However, with inflation hitting levels not seen since the 1980s, many are looking at the market to hedge against inflation. And with the volatility of cryptocurrencies, traditional stock indexes may be primed to expand in the coming months.

It is estimated that 55% of all trades in U.S equity markets are algorithmic or high-frequency trades, which are executed by computer programs. Traditional day-traders may be at a disadvantage if not looking beyond the standard financial performance indicators and seeking to act on available information more quickly. Finding additional sources of “alpha”, or signal that can be attained from alternative sources of information, can provide investors an edge over other participants in the market. Being able to adapt to emerging market news or other external factors such as a natural disaster or geopolitical event could provide timely clues and allow investors to gain an edge, providing them the ability to make informed decisions to take up buy, sell or hold positions.

This research paper will expand on the technique's others have already applied in the field regarding the use of deep neural networks combined with natural language processing methods. We hope to also provide some insights into making use of cloud services for "just-in-time" predictions which may prove to be advantages depending on how quickly markets react to the news.

2 Literature Review

2.1 Stock price forecasting model based on modified con-volution neural network and financial time series analysis

According to Cao and Wang (2019) [1] study, a stock market prediction model was conducted using convolution neural network (deep learning algorithm) and financial time series data. The convolution neural network (CNN) parameters included the number of convolution layers, the size of down sampling layers, convolution kernels sizes/numbers and samples length. The combination of convolution neural network (CNN) and financial time series analysis were effective to forecast stock market index.

2.2 Machine Learning for Algorithmic Trading

Jansen, S., author. (2020). [2] Offers a unique perspective on both Machine Learning and trading. Starts by covering some of the core concepts behind market behavior and expands into how to utilize cutting edge techniques from the field of ML to gain additional signal from non-traditional sources of financial related data.

2.3 LSTM-based sentiment analysis for stock price forecast

Ko, C., Chang, H. (2021). [3] found that their methods produced a 12.05 accuracy improvement, which entailed the use of BERT for sentiment analysis, and an LSTM neural network for analyzing and processing time series data. The improvement was compared to what they defined as LSTMP, which only included the financial data from the study, and no NLP techniques. [Opening price, Closing price, Highest price, Lowest price and Volume] They used a 20-day sliding window to predict the 21st observation across their data. Their sentiment analysis came from both news stock-oriented forums and news articles scraped from the web.

2.4 A Multi-Method Survey on the Use of Sentiment Analysis in Multivariate Financial Time Series Forecasting

Liapis, Karanikola, and Kotsiantis (2021) [4] sought to explore the use of sentiment analysis in multivariate time series forecasting for stock prices. Using the Twitter Intelligence Tool (TWINT) they scraped Twitter for tweets related to

specific stocks and grouped them by day. The data were pre-processed using Re and sentiment scores were created by using TextBlob, Vader, and FinBERT. The authors found that for predicting prices over a single day traditional univariate analysis performed best, but that for seven and fourteen day forecasts a Long Short-Term Memory network (LSTM) with FinBERT sentiment analysis outperformed the traditional models.

2.5 Harvesting social media sentiment analysis to enhance stock market prediction using deep learning

Mehta, Pandya and Kotecha (2021) [5] conducted research on stock market prediction incorporating social media sentiment analysis using deep learning approach. The deep learning methods used for the research were Support Vector Machine, Naïve Bayes, linear regression, MNB classifier and Long Short-Term Memory. They incorporated news data, public sentiments, and historical stock prices with stock trend prices. They concluded that there was a relationship between the news and the stock market because positive news correlates high market values while negative news correlates with low market values.

2.6 Sentiment analysis on social media for stock movement prediction. Expert Systems with Applications

Nguyen, Shirai and Velcin (2015) [6] presented a novel approach to consolidate the sentiments in social media for the prediction of stock price movement. They used Support Vector Machine (SVM) to systematically evaluate the effectiveness of sentiment analysis from social media (message boards). Some of the models incorporated in this study were historical prices, human sentiment, sentiment classification, LDA-based method, JST-based method, and Aspect-based sentiment. They concluded that the Aspect-based sentiment was a better method in predicting over 18 stock prices with an average accuracy of 54.4.

2.7 Stock Price Forecasting via Sentiment Analysis on Twitter

Kordonis et.al (2016) [7] developed an approach to evaluate how tweets correlate with stock price trends using machine learning. They analyzed the tweets that were collected using Twitter's search API and evaluated the effectiveness of Naïve Bayes and SVM model for public sentiments. They also collected historical stock data using Yahoo finance API to predict the future of stock prices. From the research, it was concluded that changes in the public sentiments from tweets affect the stock market. There was a correlation between public sentiments from tweets and stock prices.

2.8 Deep Learning with Long-Short Term Memory Networks for Financial Market Predictions

Fischer and Krauss (2018) [8] applied a Long Short-Term Memory (LSTM) network to the problem of predicting stock prices for the SP 500 for the years 1992

– 2015, reporting that they were able to use this model to generate daily returns of 0.46 percent and a Sharpe ratio of 5.8. They found that the LSTM model outperformed more traditional models including Random Forest, Deep Neural Net, and Logistic Regression. While the model outperformed the general market from 1992 to 2010, additional volatility in the market resulted in near 0 returns for 2010 to 2015. The authors explored additional causes of volatility and probability and were able to create a rules-based strategy that improved yields to 0.23 percent while reducing their exposure to risk when compared to the 3 other competing models.

2.9 LSTM Neural Network with Emotional Analysis for Prediction of Stock Price

In this paper, Zhuge, Xu, and Zhang (2017) [9] sought to create a new time series learning model for stock prediction. They divided the model into two parts, with the first consisting of a Naïve Bayes emotional content model derived from forum message data from stock related posts on the Eastmoney forum dated between June 2, 2008, and June 5, 2015. The authors then fed this emotional data into a LSTM time series model. They further divided the models into an external space called Network Public Opinion and internal space called Actual Behavior with 15 variables as input. For the 6 stock tickers the authors looked at, they found that the model that included emotional content data improved MSE performance by an average of 25.96 percent over a model that only used historical exchange data.

2.10 Theme

Traditional time-series methods may not be best suited to stock market data due to various other factors that go into the valuation of a given stock. Neural networks may be better equipped to infer complex relationships between different features of a data set, and when combined with other text-based data such as sentiment, this can be improved upon even further.

3 Methods

3.1 Data

Our financial time series data will be collected mainly from the Alpha Vantage API, which provides a variety of services for both historical and current day price history. The daily adjusted service, for example, covers over 20 years of historical records for a given equity. As for the text-based Sentiment Analysis component, we will collect data from the Twitter API as well as other news aggregators that provide free access to historical news headlines.

3.2 Techniques

Recent research in this area has applied deep neural networks to time series-based data using Long Short-Term Memory (LSTMs) recurrent neural networks.

Sentiment Analysis is a sub-field of Natural Language Processing, and some of the more cutting-edge research in this area makes use of transfer learning, or more specifically transformer-based models such as BERT, which have been pretrained on huge corpora of text that can be further refined on a given domain.

FinBERT for example, is an extension of BERT specific to the financial domain. With the combination of these two methods, we hope to improve upon what one would be able to achieve by just using one or the other. Aristotle is quoting as having said that “the whole is greater than the sum of its parts”, and we hope to prove that applies here as well.

4 Results

In this research we hope to demonstrate that applying sentiment analysis to different time series models results in a marked improvement in performance over the base models. Historical research shows that both LSTM and Sentiment Analysis can greatly improve predictions and long-range forecasts for stock pricing, but we have not seen this analysis applied in real time as we are seeking to do.

5 Conclusion

In conclusion, we hope to confirm our hypothesis that the sequential time series data provided by daily or even hourly stock market prices can be complimented with external sources of signal such as content from social media or financial news sites.

6 Acknowledgments

A special thanks to our research project advisors, Dr. Biven Sadler and Dr. Faizan Javed, for their guidance in the Time Series and Natural Language Processing domains.

References

1. Cao, J., Wang, J. (2019). Stock price forecasting model based on modified convolution neural network and financial time series analysis. *International Journal of Communication Systems*, 32(12), e3987-n/a. 10.1002/dac.3987
2. Jansen, S., author. (2020). *Machine Learning for Algorithmic Trading - Second Edition*. Packt Publishing.
3. Ko, C., Chang, H. (2021). LSTM-based sentiment analysis for stock price forecast. *PeerJ.Computer Science; PeerJ Comput Sci*, 7, e408. 10.7717/peerj-cs.408

4. Liapis, C. M., Karanikola, A., Kotsiantis, S. (2021). A Multi-Method Survey on the Use of Sentiment Analysis in Multivariate Financial Time Series Forecasting. *Entropy* (Basel, Switzerland); *Entropy* (Basel), 23(12), 1603. 10.3390/e23121603
5. Mehta, P., Pandya, S., Kotecha, K. (2021). Harvesting social media sentiment analysis to enhance stock market prediction using deep learning. *PeerJ Computer Science*; *PeerJ Comput Sci*, 7, e476. 10.7717/peerj-cs.476
6. 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. 10.1016/j.eswa.2015.07.052
7. Stock Price Forecasting via Sentiment Analysis on Twitter (2016). . ACM. 10.1145/3003733
8. Fischer, T., Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654-669. 10.1016/j.ejor.2017.11.054
9. Zhuge, Q., Xu, L., Zhang, G. (2017). LSTM Neural Network with Emotional Analysis for Prediction of Stock Price. *Engineering Letters* 25(2), 167-175. http://www.engineeringletters.com/issues_v25/issue2/EL25209.pdf