SMU Data Science Review

Volume 6
Number 2 *Summer 2022*

Article 10

2022

Stock Forecasts with LSTM and Web Sentiment

Michael Burgess
Southern Methodist University, mwburgess@smu.edu

Faizan Javed
Southern Methodist University, fjaved@mail.smu.edu

Nnenna Okpara
Southern Methodist University, nokpara@mail.smu.edu

Chance Robinson
Southern Methodist University, chance@mail.smu.edu

Stock Forecasts with LSTM and Web Sentiment

Chance Robinson[1], Nnenna Okpara[1], Michael Burgess[1], Faizan Javed[2]

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

Abstract. Traditional time-series techniques, such as auto-regressive and moving average models, can have difficulties when applied to stock data due to the randomness inherent to the markets. In this study, Long Short-Term Memory Recurrent Neural Networks, or LSTMs, have been applied to pricing data along with sentiment scores derived from web sources such as Twitter and other financial media outlets. The project team utilized this approach to complement the technical indicators observed at the end of each trading day for three stocks from the NASDAQ stock exchange over a 12-year span. A common benchmark to assess model performance on time series data is using the prior day's closing price of a given stock to predict the next day's closing value, which is a naive, but surprisingly accurate method when calculating the mean absolute error. The main objective of the paper is to use predictions from the various models assembled for the research, and then calculate whether the next day's closing price will rise or fall when compared against the last predicted value. All models showed on average a roughly 2% accuracy improvement over the largely balanced up and down movements for the tickers used in the study.

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.[1]

A Gallup poll conducted in August 2021 found that 56% of Americans own stock, which was consistent with prior measures taken in 2019 and 2020. This is slightly less that the levels that preceded the 2008 financial crisis.[2] 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 crypto-currencies, 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

[1] https://en.Wikipedia.org/wiki/List of countries by stock market capitalization
[2] https://news.gallup.com/poll/266807/percentage-Americans-owns-stock.aspx

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.

Stock prices are affected by several factors that determine if the stock market will go up or down. Sentiment analysis from social media, online news sources, and information provided by companies are some of the driving forces that can significantly affect these prices. According to Nguyen, Shirai and Velcin (2015) [1], understanding the moods from social media and incorporating it with company specific topics is a good novel approach for predicting prices. There is a correlation between investor sentiment and confidence, along with overall economic health, and the direction that stock prices will take.

Due to the complexity involved in the prediction of stock prices, researchers have developed methods like machine learning to solve the problem. Over the years, researchers have built prediction models to help with stock price prediction. These models help investors make business decisions to yield profits and avoid risks. Investors rely on historical data, market conditions and company performance to determine investment strategies to maximize profits. Stock market prediction is one topic that has attracted the attention of organizations, market traders, data scientists, investors. Everyone wants to know what the value of a given stock will be in the next 10 to 20 years. They also want to understand the stock market to know when to buy or sell shares to avoid incurring debts and high interest rates. Companies raise money for their corporation by trading stocks to the public hence the importance of stock market forecasting.

Investing in the stock market is a high-risk approach which requires one to evaluate the performance of the company before buying the stock. The main goal for investors is to maximize profits and yield a positive return. Accurate prediction of the behavior of a stock is essential to help investors make the best business decisions. Stock prices are constantly changing because of the opening and closing market price in the stock trades. The opening and closing market prices are not always the same due to the irregularities between supply and demand. Supply and demand control the rate at which stocks are bought and sold. Stock traders monitor the market because they buy and sell stocks the entire day. Their approach is to capitalize on market events where they can buy stocks at a lower rate or sell stocks at a high price. Stock market prediction is a challenging problem that can be solved with accurate prediction models.

Historically, analysts with a background in the financial domain could glean insights from corporate earnings statements, cash flows and balance sheets which could then be used in their trading strategies. But with the advent of the internet and more recently social media, the speed at which information is made available and processed is becoming an important factor in being able to act on that signal in a timely manner. Advances in Natural Language Processing (NLP) have provided a means to quickly assess the mood or sentiment of a text-based artifact, be it the more traditional corporate ones or social media platforms that have only been around since the turn of the century. The landscape for some of these newer NLP techniques is rapidly evolving and a method or framework that may have been state-of-the-art 5 to 10 years ago may no longer be the favored approach today.

This research paper will expand on the technique's other authors have previously applied in the field regarding the use of deep neural networks combined with natural language processing methods. Through the combined application of social media sentiment analysis and more advanced neural network approaches such as long short-term memory, this paper will improve the accuracy with which stock predictions are made. Further, previous papers and methodologies on this subject have focused predominantly on predicting future prices, but research was less prevalent for sign movement predictions. The team will attempt to use both techniques for our final up or down movement prediction.

The remainder of this paper is organized as follows: Section 2 is a related literature review in which historical papers and research done on time series, time series analysis, stock predictions, and social media/sentiment analysis are explored and discussed. Section 3 details the methods used to conduct the research in this paper including the datasets used and techniques that were applied to model the data. Section 4 provides an analysis of the results of the models created in Section 3, while Sections 5 and 6 focus on a discussion of these results and potential ethical issues raised by this research. Finally, Section 7 summarizes the results of this paper's research and what conclusions can be drawn from that analysis.

2 Literature Review

Several authors have explored possible models for predicting both time series data as a whole and more specifically stock prices. Although there have been attempts to make these predictions with less advanced models such as Naive Bayes and Support Vector Machines, current research suggests that more advanced Neural Network models such as Long Short-Term Memory perform significantly better for these prediction tasks. Further, there have been some attempts to add sentiment or emotional data to these models and adding this content has been shown to improve model performance.

For example, Mehta, Pandya, and Kotecha (2021) [2] found in their research that incorporating news data, public sentiments, and historical stock prices into several different machine learning models suggested a relationship between news and stock market prices. The authors showed that positive news corresponded with higher stock prices while negative news was correlated with lower prices.

Similarly, Ji, Wang, and Yan (2021) [3] sought to prove that accurate prediction of stock prices required integrating a deep learning approach with social media and stock financial text features. The authors looked at news from 15 companies, as well as investor opinions, and extracted text features from this data using Doc2Vec. When combined with an LSTM model, the authors concluded that the integration of social media resulted in a model with better mean absolute error (MAE), root mean square error (RMSE), and r-squared performance.

Jing, Wu, and Wang (2021) [4] took a slightly different approach and used a convolutional neural network (CNN) to create sentiments for data extracted from stock oriented online forums. The authors also chose to use an LSTM model with this data as input for predicting the prices of thirty stocks on the Shanghai Stock Exchange. They found that this model outperformed models not using sentiment analysis when evaluated on Mean Absolute Percentage Error (MAPE) performance.

One of the unique challenges that must be addressed when performing sentiment analysis without automated feature extraction is the need for an accurate and relevant lexicon. The financial domain has many terms that are unique and specific only to it. Yang, Siy, and Huang (2020) [5] created a finance-domain focused version of the BERT model called FinBERT that emphasized these specific terms. The FinBERT model increased the number of tokens within the BERT model from three billion to five billion using training data such as a financial phrase bank, an analyst tone data set, and a FiQA data set. The authors claimed that some models showed up to a 29% improvement when compared to non-domain specific BERT models when used for specific finance related NLP tasks.

Building on similar research conducted by Oliveria, et al. (2014) [6], Yekrangi and Abdolvand (2020) [7] also worked to create a more applicable, financialspecific lexicon for making accurate financial sentiment classification. However, unlike the research that was done in the papers by both Oliveria, et al. (2014) [6], and Yang, et al. (2020) [5], Yekrangi and Adbolvand's model was created in several manual phases and did not employ the use of automatic or traditional machine learning techniques. In this instance, the authors found the most common words used in articles published by Bloomberg and Reuters and then verified that the words were related to finance. These words were then evaluated for context and classified as positive or negative. Finally, the lexicon was expanded by adding corpus and dictionary-based words related to the previously selected words. Yekrangi and Abdolvand found that their lexicon improved performance over the SentiWordNet (SWN) lexicon by 285%. The authors also had their lexicon evaluated by financial experts who found that 84% of the positive words and 88% of the negative words in the lexicon were classified correctly.

FinBERT and other similar lexicons have been rapidly adopted and have shown marked improvement over previous models. Ko and Chang (2021)) [8] created an LSTM model that used data that had been processed with a BERT model for sentiment analysis. The authors found a 12.05% accuracy improvement over a model that used only financial data from their study and no NLP techniques.

Liapis, Karanikola, and Kotsiantis (2021) [9] also found they could improve their models through the use of FinBERT. In their paper, the authors were able to show that data retrieved using the Twitter Intelligence Tool (TWINT) performed better when pre-processed using TextBlob, Vader, and FinBERT to assign sentiment scores. Interestingly, the authors also discovered that traditional univariate time series models performed better for predicting prices over a single day, but that LSTM models combined with the FinBERT analysis data performed better for seven- and fourteen-day forecasts.

The previous research done on this subject shows a clear correlation between sentiment analysis and an improvement in the accuracy of models for predicting stock prices. While most related research has been focused on long-term forecasting or historical review, in this paper the authors have sought to address the challenge of short-term, next-day forecasting.

3 Methods

In this section, we discuss the data sources and collection techniques used in our analysis and model building.

3.1 Data

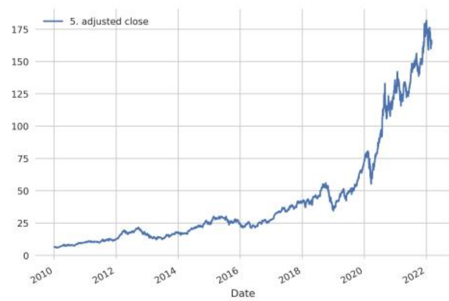
The financial time series data has been collected from the Alpha Vantage API, which provides a variety of services for both historical and current day price history. An academic-use API key was granted to the project team by the Alpha Vantage company, which allows for access to premium endpoints and higher daily usage rates. The premium-only daily adjusted service for example, covers over 20 years of historical records for a given equity. It includes attributes for the Open, High, Low, Close, Adjusted Close, Volume, Dividend Amount and Split Coefficients. The intraday extended history service, another premium endpoint, allows for pulling up to two years of historical data which can be segmented by minute to provide a more granular level of pricing detail. This paper will be focused primarily on the daily historical information that is available through Alpha Vantage's service. However, the higher frequency trading data might provide an interesting case study for future research. [3]

As for the text-based Sentiment Analysis component, the data has been obtained from two sources. The first source for financial tweets utilizes a Python library named TWINT, which is short for Twitter Intelligence Tool. This was chosen over another widely used Python library, Tweepy, as it is not rate limited and allows for searching of Cashtags, which are more specific to financial news.

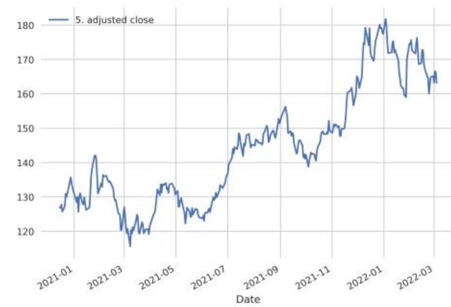
[3] https://www.alphavantage.co/documentation/

Cashtags are also available to search through Tweepy, but an enterprise search API is required. Another added benefit of the TWINT library is that it does not require a Twitter developer account.

Benzinga is a news aggregator that provides access to both real-time and historical news headlines. Benzinga offers a Stock News API with multiple channels that cover areas such as fintech (or financial technology, payments and capital markets), regulations, rumors and news relating to headlines about the US.



(a) Full Time Series



(b) Last 3 Months

Fig.1: Daily Adjusted Time Series for Apple (AAPL) stock

Government among others. It is unique in that it creates the content in-house as opposed to relying on second-hand sources of news feeds that often rely on web-scraping techniques. The project team chose to discard some of the financial pricing data obtained through Alpha Vantage due to the fact that Benzinga was launched in 2010. Therefore, all analysis was performed on data collected from January 1st, 2010 through March 4th, 2022. [4]

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 pre-trained on huge corpora of text that can be further refined on a given domain.

Labeled data for textual features can be costly and hard to attain. The project team has opted instead to use methods that allow for direct computation of sentiment on a body of text. As the team does not have access to labeled training data to feed into the neural network, an attempt will be made to leverage transfer learning techniques on entirely unlabeled data and compare performance to

[4] https://www.benzinga.com/apis/

models not making use of any investor sentiment. To the project team's knowledge, this approach has not been used based on the other research papers cited here, but it is likely more applicable in real world situations where recent labeled data may be rather difficult to attain.

FinBERT is an example of a transfer learning model, which is an extension of BERT specific to the financial domain. With the combination of the technical indicators from Alpha Vantage and the sentiment scores produced by FinBERT, the research team hopes to improve upon what one would be able to achieve with an LSTM recurrent neural network over using just one approach versus the

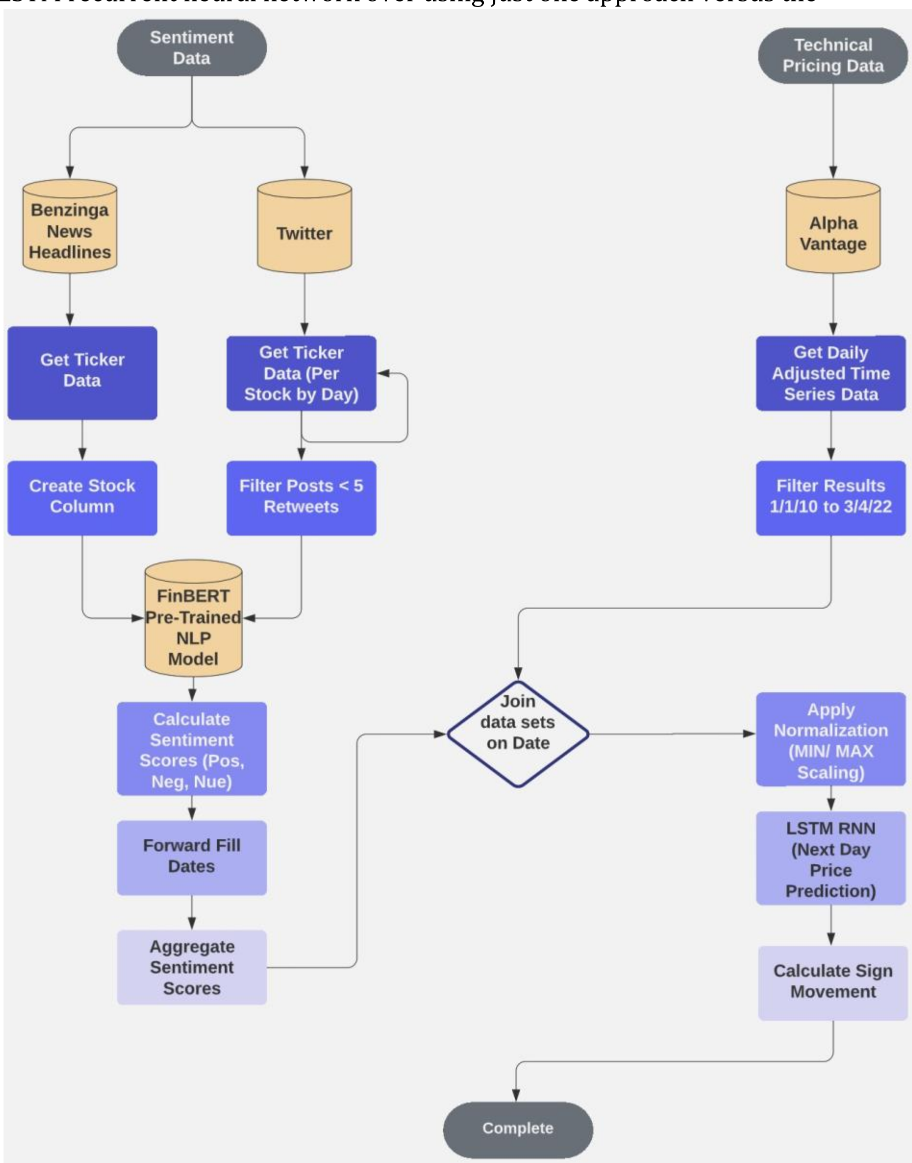


Fig.2: Model Architecture

other. Aristotle is quoting as having said that "the whole is greater than the sum of its parts", and this study hopes to prove that applies here as well. [5]NASDAQ stock data is typically available on weekdays, from 9:30 a.m. to 4 p.m. Eastern Time. Due to the way in which the data was collected from the daily adjusted Alpha Vantage service, there are over three thousand samples for the nearly 12 year span of historical records. There are no duplicate entries for a given equity and the day with which the stock pricing details were recorded. The financial news articles and tweet data however often have many unique records for the same point in time or day, and can obviously be posted outside of the normal Wall Street trading hours. The articles were scored individually using the FinBERT library previously mentioned, and aggregated to show an average score for each of the recorded days. This was done separately for both the financial news data and tweets to compare and contrast their contributions to the model.

Any missing sentiment values were imputed with prior scores so as to carry forward content that may have been posted over the weekend as an example. Similarly, tweets posted over the weekend were merged with dates corresponding to the following Monday. As the TWINT API also provides the number of likes and retweets for each of the tweets, daily totals were calculated for those columns as well to serve as additional features for the model. However, as this feature was not found to be useful in our initial tests it was excluded from subsequent model comparisons.

4 Results

In this research, the goal is 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. As noted, a common benchmark for evaluating the performance of a stock related time series model is to use the current day's observed target value, the adjusted closing price in the case of this paper, to predict what the next day's adjusted closing price would be. This will serve as the baseline model for which all others are measured against. Surprisingly, the naive baseline model outperforms the advanced models that are included in this research paper. This could also indicate that labeled training data could help with our model's ability to fully leverage the sentiment available in our news and social media data, and that the

[5] https://huggingface.co/ProsusAI/finbert

aggregated sentiment scores from the FinBERT library are not as effective as training on labeled data.

One can also surmise that even neural network based models make use of the prior day's value as evidenced in the chart below. You can see that some of the predicted values are close representations of the prior day's closing price. Figure 4 illustrates that the performance of the baseline AAPL model outperforms most of the complex LSTM-based recurrent neural network models which use a variety of uni-variate and multivariate input features. The additional input features appeared to have negatively impacted the model's performance from a Mean Absolute Error standpoint. Perhaps this would not be the case other data

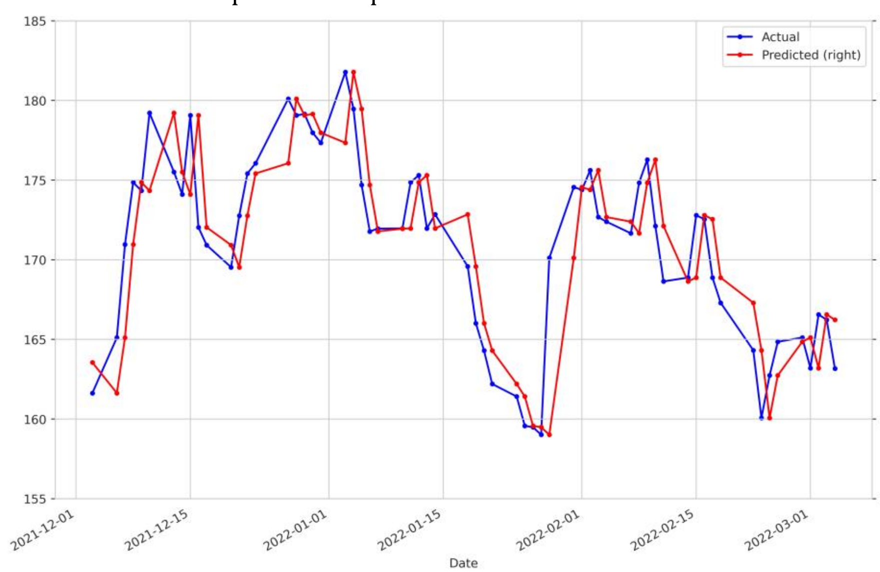


Fig.3: Actual vs. Predicted Adjusted Closing Price for the last three months of the naive baseline model with Apple (AAPL) stock.

sets, but as it stands, the time series encoded data for the current day adjusted closing price appears to not benefit from the additional information provided by either the pricing or sentiment features for next day predictions for the majority of our stocks and models.

The model names have been condensed for display purposes and have the following mappings. (UV = Uni-variate Time Series, MV = Multivariate Time Series with technical indicators only, NWS = Multivariate Time Series with Benzinga news article sentiment, TW = Multivariate Time Series with Twitter sentiment, ARIMA = Auto Regressive Integrated Moving Average)

4.1

Model	MAE (Scaled)	MAE (Re-Scaled)
Baseline	0.0140	1.7012
LSTM (UV)	0.0095	1.6797
LSTM (MV)	0.0103	1.8127
LSTM (MV+NWS)	0.0105	1.8421
LSTM (MV+TW)	0.0105	1.8465

4.2

ARIMA	0.0092	1.5698

Apple Stock

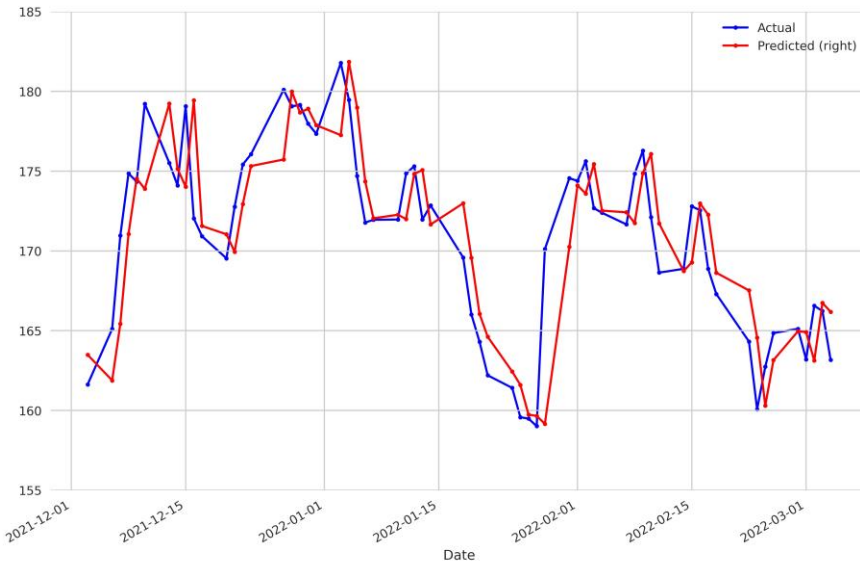


Fig.4: Actual vs. Predicted Adjusted Closing Price for the last three months of the uni-variate model with Apple (AAPL) stock.

Table 1: Comparison of models by Mean Absolute Error for both the scaled and re-scaled test sets for Apple (AAPL) stock.

4.3 Amazon Stock

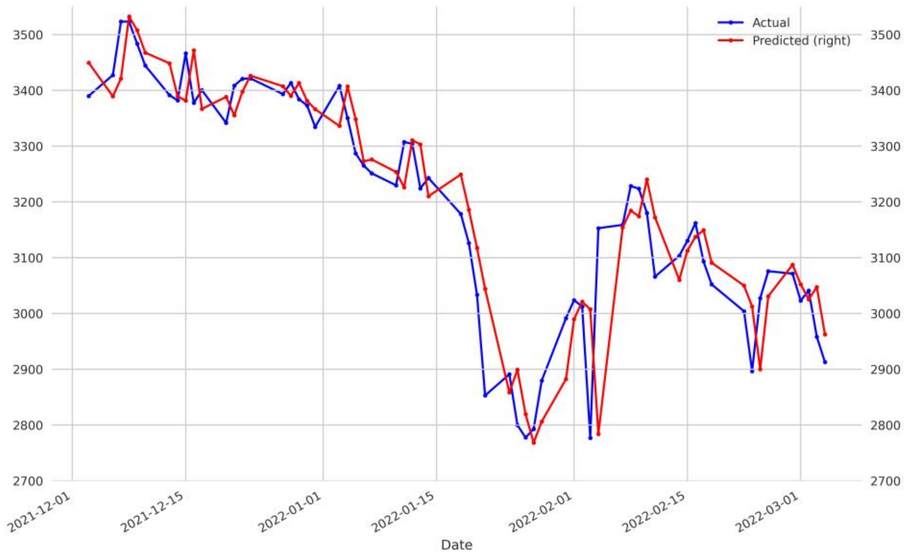


Fig.5: Actual vs. Predicted Adjusted Closing Price for the last three months of the uni-variate model with Amazon (AMZN) stock.

Table 2: Comparison of models by Mean Absolute Error for both the scaled and re-scaled test sets for Amazon (AMZN) stock.

Model	MAE (Scaled)	MAE (Re-Scaled)
Baseline	0.0122	41.9158
LSTM (UV)	0.0117	42.3601
LSTM (MV)	0.0125	45.1714
LSTM (MV+NWS)	0.0125	45.2038

LSTM (MV+TW)	0.0121	43.8257
ARIMA	0.0112	40.7389

4.4 Microsoft Stock

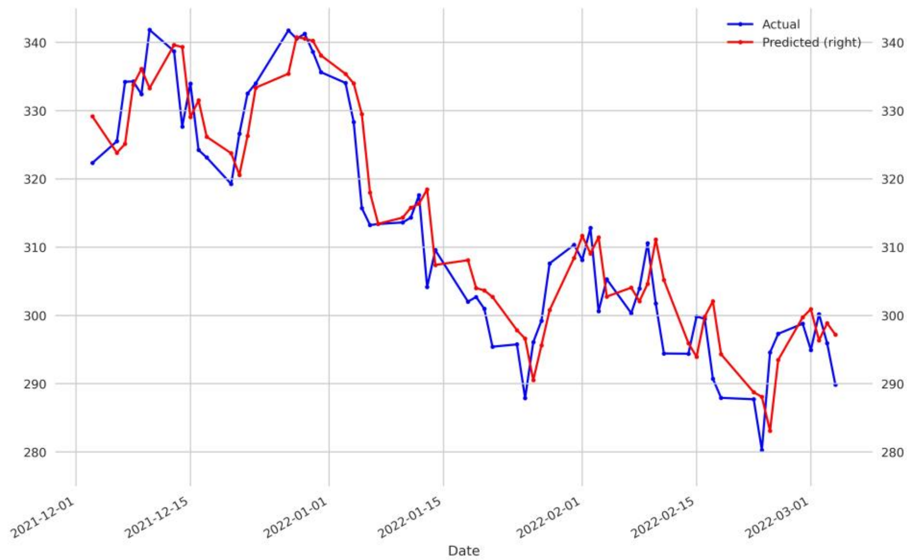


Fig.6: Actual vs. Predicted Adjusted Closing Price for the last three months of the uni-variate model with Microsoft (MSFT) stock.

Table 3: Comparison of models by Mean Absolute Error for both the scaled and re-scaled test sets for Microsoft (MSFT) stock.

Model	MAE (Scaled)	MAE (Re-Scaled)
Baseline	0.0143	3.0943
LSTM (UV)	0.0100	3.2478

4.5

LSTM (MV)	0.0107	3.4869
LSTM (MV+NWS)	0.0104	3.3835
LSTM (MV+TW)	0.0104	3.3853
ARIMA	0.0094	3.0166

If, however the objective is not just a reduced mean absolute error from predicting a continuous output such as the adjusted closing price, but instead whether the price will go up or down on the next trading day then the additional features derived from social media and other news sources do appear to have more predictive value as shown in the tables below.

Although the sources and their usefulness are not always consistent between stocks, the majority finding was that the sentiment scores show an improvement over not having them in the model for sign-based predictions. In some cases, these percentages beat out the uneven distribution of the up or down movements for the given stock. The naive baseline model for the mean absolute error objective has little value if the desired output is a directional prediction, as it is only capable of projecting out the last day's observed value, which is neither up nor down but the same obviously. One could argue that a sign movement is a much more practical desired result and one that investors may prefer.

In lieu of this, the baseline model has been replaced with the actual distribution of the up and down movement trends for the hold-out test data set.

Table 4: Comparison of models by their ability to predict the movement increase or decrease for Apple (AAPL) stock.

Model	↑ Accuracy %	↓ Accuracy %	Overall Accuracy %
Baseline	0.5267	0.4733	0.5000
LSTM (UV)	0.9063	0.1826	0.5638
LSTM (MV)	0.7188	0.3130	0.5267
LSTM (MV+NWS)	0.5547	0.4870	0.5226
LSTM (MV+TW)	0.6406	0.4261	0.5391
ARIMA	0.6328	0.6609	0.6461

Table 5: Comparison of models by their ability to predict the movement increase or decrease for Amazon (AMZN) stock.

Model	↑ Accuracy %	↓ Accuracy %	Overall Accuracy %
Baseline	0.5021	0.4979	0.5000
LSTM (UV)	0.3802	0.7049	0.5432
LSTM (MV)	0.5619	0.4754	0.5185
LSTM (MV+NWS)	0.5041	0.5246	0.5144
LSTM (MV+TW)	0.4711	0.5902	0.5309
ARIMA	0.5372	0.6311	0.5844

Table 6: Comparison of models by their ability to predict the movement increase or decrease for Microsoft (MSFT) stock.

Model	↑ Accuracy %	↓ Accuracy %	Overall Accuracy %
Baseline	0.5185	0.4815	0.5000

LSTM (UV)	0.8730	0.1880	0.5432
LSTM (MV)	0.5555	0.4274	0.4938
LSTM (MV+NWS)	0.5555	0.5043	0.5309
LSTM (MV+TW)	0.5635	0.3846	0.4774
ARIMA	0.5476	0.5641	0.5555

5 Discussions

There were several areas in which additional exploration would have been beneficial to the project. The first of which is the ability to fine-tune the pre-trained FinBERT model with labeled data. The team could have attempted to perform this internally, but this would have been a large undertaking and not something that would have been practical to do with the large volume of texts that were collected, nor a realistic objective if attempting to use this strategy on a day to day a basis.

Second, only the daily technical indicators collected from the Alpha Vantage API were used in the analysis even though the team had access to up to two years of intra-day frequency trading information at various intervals. This was mainly due to the fact that the news headline and tweet data was much sparser when compared to the additional observations provided by the intra-day frequencies. Nonetheless, it would have been worth pairing the closing prices with sentiment scores collected within a reasonable time-frame leading up to the specified interval recording. This might have allowed for a more concrete assessment of whether or not the sentiment scores contributed more to the predictive power of the model when compared to those without any corresponding scores.

Additionally, investigation of cloud services such as those provided by Amazon Web Services (AWS), Microsoft Azure or Google Cloud Platform (GCP), for “just-in-time” predictions may have proven to be advantageous depending on how quickly markets reacted to events and updates that were trending on social media and news sites at the time. Or even just to have the scalability and infrastructure needed to support a high volume data pipeline.

And lastly, additional data preparation and cleansing could have helped with the sentiment data, especially those sourced from twitter. It is likely that a large percentage of tweets were false-positives/ negatives reducing the accuracy of the models. It is estimated that nearly 25% of all tweets from AAPL for example were posted by bots and didn’t necessarily reflect market sentiment by actual investors. Adding a component to classify tweets as being legitimately posted or sourced from bots or other non-reliable sources could serve as an suitable method to filter out the deceptive posts.

6 Ethical Considerations

It should be noted that this and similar research could be seen as giving an unfair competitive advantage to brokerage firms and companies when compared to individual retail investors. Certainly the average retail investor lacks access to the knowledge and resources necessary to build the models or to apply the techniques suggested in this paper. It may be possible to mitigate these concerns by developing this research into an open source library and available to anyone who wishes to use it, as has been done with similar models and libraries such as FinBERT, as opposed to limiting access to this paper or creating a licensed product for purchase.

There are also concerns raised by the US Securities and Exchange Commission's (SEC) code of ethics. It is possible, although unlikely, that the sentiments and models created in this paper could fall under the SEC's definition of material information, that is any information an investor would consider important in making their investment decision. [6] If the research in this paper, or the results of the use of this research, were found to constitute material information, it would be the legal and ethical duty of the corporation or individual using these models to disclose their results publicly.

There are ethical concerns that need to be considered when working with public data like Twitter. There is a need to protect the privacy of the author of the tweets collected. Although the tweets are made public, it is still ethical to understand the risks and protect the data. For the analysis, the team chose to protect the privacy of the tweets by removing any personal information. The team ensured personal information was not stored on any open-source software.

Furthermore, the prediction model developed by the team also permits the exploitation of people. The exploitation could happen in the sense of persuading people to invest more or circulating wrong information. The goal of the team's prediction model is to help people make better decisions when investing in stocks.

7 Conclusion

In conclusion, using only historical pricing information to predict future stock movements is not a realistic goal as there are so many additional factors that go into the price fluctuations for any given stock. However, complimenting the technical information with additional sources of "alpha", such as sentiment from tweets, can be another tool in the belt of a day trader or interested machine learning practitioner with an eye on the markets.

The research team found that there was potential for the techniques outlined in the paper to predict sign movement with more than 50% accuracy, albeit on a relatively small sample size with the three NASDAQ tickers used in our analysis.

Although this number may seem underwhelming, it is known that many stocks exhibit a random walk behavior and anything above and beyond a 50% threshold could prove to be of value. Perhaps one could focus on only positions that offer a higher confidence or those whose gains would exceed the transnational costs for smaller trades. LSTM Recurrent Neural Networks combined with NLP techniques like Sentiment

[6] https://www.sec.gov/Archives/edgar/data/789280/000119312513307005/d571345dex99p.htm

Analysis can be a useful asset for executing algorithmic trades based on data, and not one's personal emotion.

8 Acknowledgments

A special thanks to our research project advisor, Dr. Faizan Javed, for his guidance in the Time Series and Natural Language Processing domains.

References

1. Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment Analysis on Social Mediafor Stock Movement Prediction. Expert Systems with Applications, 42(24), 96039611. 10.1016/j.eswa.2015.07.052
2. 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
3. Ji, X., Wang, J., & Yan, Z. (2021). A Stock Price Prediction Method Based onDeep Learning Technology. International Journal of Crowd Science, 5(1), 55-72. 10.1108/IJCS-05-2020-0012
4. Jing, N., Wu, Z., & Wang, H. (2021). A Hybrid Model Integrating Deep Learningwith Investor Sentiment Analysis for Stock Price Prediction. Expert Systems with Applications, 178, 115019. 10.1016/j.eswa.2021.115019
5. Yang, Y., UY, M. C. S., & Huang, A. (2020). FinBERT: A Pretrained LanguageModel for Financial Communications.
6. Oliveria, N., Cortez, P., & Areal, N. (2014). Automatic Creation of Stock Market Lexicons for Sentiment Analysis Using StockTwits Data. Paper presented at the 18th International Database Engineering & Applications Symposium, 115-123. 10.1145/2628194.2628235
7. Yekrangi, M., & Abdolvand, N. (2020). Financial Markets Sentiment Analysis: Developing a Specialized Lexicon. Journal of Intelligent Information Systems, 57, 127146. 10.1007/s10844-020-00630-9
8. 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
9. Liapis, C. M., Karanikola, A., & Kotsiantis, S. (2021). A Multi-Method Survey onthe Use of Sentiment Analysis in Multivariate Financial Time Series Forecasting. Entropy (Basel, Switzerland); Entropy (Basel), 23(12), 1603. 10.3390/e23121603
10. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation; Neural Comput, 9(8), 1735-1780. 10.1162/neco.1997.9.8.1735
11. 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
12. Jansen, S., author. (2020). Machine Learning for Algorithmic Trading - SecondEdition. Packt Publishing.
13. 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
14. Kordonis, J., Symeonidis, S., Arampatzis, A. (2016). Stock Price Forecasting via Sentiment Analysis on Twitter (2016). PCI '16: Proceedings of the 20th Pan-Hellenic Conference on Informatics, Article 36, 1-6. 10.1145/3003733.3003787
15. 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
16. 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/issue 2/EL 25 2 09.pdf

17. Ranco, G., Aleksovski, D., Caldarelli, G., Grˇcar, M., & Mozetiˇc, I. (2015). TheEffects of Twitter Sentiment on Stock Price Returns. PloS One; PLoS One, 10(9), e0138441. 10.1371/journal.pone.0138441
18. Jing, N., Wu, Z., & Wang, H. (2021). A Hybrid Model Integrating Deep Learningwith Investor Sentiment Analysis for Stock Price Prediction. Expert Systems with Applications, 178, 115019. 10.1016/j.eswa.2021.115019
19. Ho, T., & Huang, Y. (2021). Stock Price Movement Prediction Using Sentiment Analysis and CandleStick Chart Representation. Sensors (Basel, Switzerland), 21(23), 7957. 10.3390/s21237957
20. Derakhshan, A., & Beigy, H. (2019). Sentiment Analysis on Stock Social Mediafor Stock Price Movement Prediction. Engineering Applications of Artificial Intelligence, 85, 569-578. 10.1016/j.engappai.2019.07.002.
21. Hentschel, M., & Alonso, O. (2014). Follow the Money: A Study of Cashtags onTwitter. First Monday, 19(8) 10.5210fm.v19i8.5385
22. 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. 10.1016/j.eswa.2019.06.014
23. Schmidt, A. (2019). Sustainable News – A Sentiment Analysis of the Effect of ESGInformation on Stock Prices. Elsevier BV. 10.2139/ssrn.3809657
24. Peng, Y., & Jiang, H. (2016). Leverage Financial News to Predict Stock PriceMovements Using Word Embeddings and Deep Neural Networks. Association for Computational Linguistics. 10.18653/v1/n16-1041
25. Joosery, B., & Deepa, G. (2019). Comparative analysis of time-series forecasting algorithms for stock price prediction. Paper presented at the 1-6. 10.1145/3373477.3373699 http://dl.acm.org/citation.cfm?id&61;3373699
26. Nelson, D. M. Q., Pereira, A. C. M., & de Oliveira, R.,A. (2017). Stock market'sprice movement prediction with LSTM neural networks. Paper presented at the 1419-1426. 10.1109/IJCNN.2017.7966019