Ann. For. Res. 66(1): 212-233, 2023
ISSN: 18448135, 20652445

ANNALS OF FOREST RESEARCH
www.e-afr.org

UNDERSTANDING THE EFFECT OF INVESTORS' SENTIMENTS ON THE S&P 500 PRICE LEVELS: A DEEP LEARNING MODEL APPROACH

[1]Danielle Khalife, [2]Jad Yammine, and [3]Tatiana El Bazi

[1,2,3] Business School, Holy Spirit University of Kaslik, P.O. Box 446, Jounieh, Lebanon

E-mail: daniellekhalife@usek.edu.lb

Abstract

This paper aims to examine to what extent can investors' sentiments extracted from social media content, specifically Twitter, improve the predictability of the S&P 500 price levels. Two Recurrent Neural Network models were built; the first one is solely based on historical records and technical indicators. The second one includes the same variables as the first model, along with the outputs of the sentiment analysis, performed using TextBlob library. While assessing the performance of both models created, the second model hatched better outcomes, highlighting the critical role these digital platforms play in shaping the behavior of a specific asset.

Keywords: deep learning, equity market, sentiment analysis, social media content, time-series forecasting.

1. Introduction

Due to its numinous traits, the equity market has always been a key area that enticed and lured the curiosity of many researchers across different fields. One of the most valuable areas of financial analysis, yet among the most debatable topics of all times, is stock market trends analysis. However, stock market prediction remains one of the most challenging undertakings due to the market's dynamic, complex character and non-linearity. Equities, like any financial instruments, are swayed by many factors, split into technical and fundamental, including the macroeconomic environment, such as competitive pressures, key socio-economic circumstances, related companies' status, investors' preferences, and many more. Lately, with the emergence of social media, the fast-paced environment, and fleet spread of information, many researchers have been making use of companies' related data fetched from the web, such as financial news, reports, investors' preferences, and opinions being intrigued by the behavioral finance theory. A massive volume of information is accumulated on a daily basis, and the idea of combining these data into a predictive model to improve the prediction performance is excessively gaining traction. Even if we agree or not, we are all aware that we live in a media-driven environment, where content, information, and news are spread in no time. Their influence on our societies is on a steady and constant rise. Indeed, it imposed its existence and participated in shaping and affecting individuals' opinions, decisions, and actions. Financial markets may have one of the top shares of social media

content, especially Twitter since most investors and traders tend to express their financial opinions on this platform. It is said that such type of data can have an influence on the financial markets by affecting individuals investing choices. Kesavan et al. (2020) attempted to investigate the effect of inclusion of the content extracted from Twitter on the foreseeability Indian equity market. Their findings significantly highlighted that such content when included, it positively influences the accuracy of the models created. On the other hand, several academics attempted to base their analysis and try to forecast the price behavior of a specific asset using a single type of data, based explicitly on technical indicators. Yet, it has been demonstrated that the outcomes of their analysis can be improved by combining other source of data, such as the sentiment feature extracted from news and investors' behaviors. Li et al. (2020), proved that models that combine several types of data perform better than models based on a single type of data. Wu et al. (2021) proposed a new S_I_LSTM method to forecast future stock values. It considers multiple variables such as historical prices, technical indicators, forum posts related to stocks, and financial news, out of which they obtained investors' sentiments. The CNN method was employed to conduct the sentiment analysis and forecast the China Shanghai A-share market. They highlighted that investors tend to use multi-source data in terms of decision-making. The predicted closing price based on their method is near the actual closing price compared to the single data source. The impact of social media on the financial markets exceedingly branched out and evolved to become a go-to online trading consultant, where individuals gather virtually, seek investment information, and exchange trading ideas and recommendations.

Historically, statistical and mathematical methods were applied to time-series problems. According to these conventional statistical methods, the relationship between the dependent and independent variable is linear, which is not appropriate and accurate all the time. With the development of Artificial Intelligence, more efforts have indeed been directed to this field. This field has proven its potency and efficiency by mimicking and surpassing human intelligence. Indeed, all methods veil weaknesses and flaws; however, when employed in different debatable research areas, they hatched favorable, outcomes compared to the conventional methods. Non-linear models such as Machine or Deep Learning models are said to provide more forecasting power compared to traditional linear-parametric models; they generate strengthened outcomes and have proved that predicting the behavior of the financial markets is conceivable. Yet, it has been demonstrated that deep learning models outperforms the machine learning ones. Several distinguished studies approaching this same area of research, shall be addressed here. Matsubara et al. (2018), Sanboon et al. (2019), Naik and Mohan (2019), and Shen and Shafiq (2020), were among the authors that were capable of demonstrating that deep learning models have a better predictive ability than the conventional machine learning techniques. Additionally, Yu and Yan (2020), in this same field of analysis, conducted an impressive study where they have shown that deep neural network technique, blending the deep learning and neural networks' assortments, outperforms the conventional machine learning methods. Jiang (2021), presented in a thorough and detailed way the progress in the implementation of deep learning models in the equity market.

We are increasingly acquiring additional information regarding this specific topic due to the outcomes attained by several recent pieces of literature. The findings of these papers point out that content extracted from the different social platforms reflecting investors' moods and opinions, quantified through sentiment analysis, are being recognized among the inputs and features that are said to have a significant influence on the behavior of equities, and an increased emphasis is placed on such variables. In the lights of these findings, this paper provides two deep learning predictive models aiming to predict the S&P 500 daily closing price. The first model fits in historical records along with selected technical indicators as inputs. The second model includes the same features as the first model; however, it consists of including the sentiment feature to examine to what extent can such types of variables improve the predictability of the predictive model created and if their implementation improves the overall performance of the related model. The proposed models use the last 50 values of the inputs to try to predict the 51st value, the S&P 500 daily closing price.

The remainder of this paper involves several chapters, where each chapter tackles a specific area, and when combined they contribute to generate a well-structured empirical scientific research. The 'Literature Review' chapter covers formerly performed studies with similar backgrounds, followed by the methodology chapter that provides a thorough explanation of the chosen variables and the adopted process helping in answering our research questions. Thereafter, the outcomes attained were well portrayed, interpreted and discussed. Lastly, this analysis was concluded with a final chapter where the contributions of this study, its major findings, limitations and further research were displayed.

2. Literature Review

Traditional financial theories like the Random Walk (Malkiel, 1973) and Efficient Market hypothesis (Fama, 1981) proclaim that stocks' behaviors are unpredictable. However, modern computational technologies, such as artificial intelligence, contradicted these principles by demonstrating that this process is feasible. Machine learning and deep learning models can process a massive amount of complex and non-linear data, incentivizing researchers and analysts to employ them when tackling such topics at the expense of conventional analysis methods. Equities' behaviors are influenced on different levels, including the macroeconomic environment, the company's fundamentals, and technical indicators. Various forecasting models were built, relying on a single or multiple sort of features offering investors valuable assistance and guidance to attain effective financial resolutions.

Numerous studies focused on predicting stocks price levels using technical indicators. Jiao and Jakubowicz (2017) employed more than 200 technical features obtained using historical prices of the constituents of the S&P 500 along with eight global indices. The analysis was done based on four machine learning models. Their findings revealed that it is a challenging task to forecast individual stock prices based on their records, and the S&P500 index by itself is much easier to predict than its constituents. Pahwa and Agarwal (2019), computed several technical indicators that served as inputs for training the classifier created. Their results were purely based on numbers,

their model provided a significant accuracy. Ravikumar and Saraf (2020) utilized raw data extracted from Yahoo Finance to derive essential features such as stock momentum, index volatility, and sector momentum. These parameters served as inputs to feed the machine learning models created, divided into classification and regression, and their findings pointed out that the Logistic Regression model generated the highest accuracy.

Deep learning is a branch of the machine learning sphere; however, the performance of the deep learning algorithms is better than the machine learning ones when the amount of data increases. Mehtab et al. (2020) built eight machine learning models along with four LSTM deep learning models and processed historical index prices to obtain key technical features. The authors were able to demonstrate that out of all the models built; the deep learning LSTM models performed way better than the machine learning models. Moohambikai and Bhakirathi (2020) built several Recurrent Neural Network models based on various time steps and different numbers of Long Short-Term Memory (LSTM) layers. The LSTM layers were used as a tool to improve the accuracy and performance of the model. According to their findings, all models performed pretty well; however, the model with 60-time steps and four LSTM layers was the best performer. Additionally, Chen and Ge (2019), Feng et al. (2019), Sachdeva et al. (2019), Sethia and Raut (2019), Zhang et al. (2019), were amongst the authors that attempted to forecast the equity market using a combination of historical records and technical indicators, and were able to prove that deep learning models generate competitive outcomes. Owing to their ability to grasp the chronological correlations inherent in time-series datasets dynamically, deep learning models have broadly endorsed their efficiency in such forecasting tasks and are indeed an efficient approach.

However, employing a single class of data does not result in improved outcomes (Mohan et al., 2019). After being intrigued by behavioral finance theory (Tversky and Kahneman, 1989), researchers, started shedding light on the importance of employing a feature that reflects investors' behaviors as there is a common consensus that investors are not consistently rational and markets are inefficient contradicting the conventional financial theories that their assumptions are sincerely arduous to accept. Content extracted from social media played a significant role in this area, as the advent of these platforms has resulted in a massive volume of information being collected on a daily basis. The use of such information is becoming more prevalent and has demonstrated that it increases the models' performance. Various studies included the sentiment feature as one of the factors affecting the stock price level; the majority of the findings affirmed that it impacts the assets' prices. In their study, Bollen et al. (2011) implemented an emotion tracking tool, to examine textual data retrieved from social media, specifically tweets, to check if investors' overall sentiments have an influence and affect the value of Dow Jones Industrial Index, and employed the outcomes to forecast percentage change of this index. According to these aforementioned authors, when predicting future stock prices, the accuracy of the models can significantly improve by implementing public mood parameters. Thereafter, several academics attempted to investigate the possibility of forecasting equity prices using behavioral, psychological, and cognitive data relevant to the market players. Liu (2018) proposed a Bidirectional, Attention-based LSTM deep neural model to forecast directional movements of the S&P 500 and other stocks based on financial

news, companies' related data, and their historical prices. The outcomes of the author's research proved that this technique is competitive. Attanasio et al. (2019) worked on building models that detect whether a trend reversal would occur the next day and used the outcomes as trading signals for trade management. As per their research paper, the strategies that combine sentiment analysis along with stock price-related parameters averagely perform better than all the other combinations. Deng (2020), rather than considering all the companies of the S&P 500, only considered historical prices, financial news headlines, and social media twits of the top five biggest companies of the prementioned index, including, Apple, Microsoft, Amazon, Google that is owned by alphabet, and Facebook. The author used Vader and Word2vec to extract the sentiments from the financial news and social media content, and these outputs were combined with the technical features to forecast short-term stock price movements using machine learning methods, that revealed good accuracy and F1 score. Lee (2020) went for including Google Trends records related to the pandemic searches, along with the Daily News Sentiment Index to check their influence on eleven selected sectors indices and the related industries' return. According to their findings, information extracted from social platforms is valuable especially when uncertainty reigns on the stock market. Song et al. (2020) conducted a distinctive study. Their suggested method consists of measuring the reports released by analysts and splitting the findings into two indicators: report attention and rating sentiment. The investors' sentiment was computed based on specific external market factors, and they used the LSTM method for stock price prediction. Their results revealed that the LSTM model performed better than the traditional SVM model. Furthermore, and most importantly, the inclusion of the investors' sentiment parameter improved the model accuracy. Ko and Chang (2021) used the BERT model to conduct sentiment analysis on financial news and posts retrieved from forums related to specific stocks. The authors used the sentiment analysis results as an input for an LSTM neural network model created to forecast the stock's next opening price. As per the authors' findings, the sentiment analysis is an important parameter to include when forecasting stock prices as it reduces the model's RMSE when included. Nemes and Kiss (2021) also used BERT as a primary technique and compared it with three other Natural Language Processing techniques (NLP), to extract the sentiments out of economic news headlines. Their findings revealed that news headlines have an influence on stocks' fundamental values and pointed out a significant difference between the multiple NLP techniques used. Sidogi et al. (2021), using LSTM architecture, were able to study the impact of headlines of financial articles on the foreseeability of the equity market. Intraday equity records were employed along with particular latencies among both posted news titles and the equities price levels. To execute the sentiment analysis the authors employed FinBERT. Two models were created and their findings implied that when employing the sentiment feature, the forecasting power of the model enhanced considerably. Wang et al. (2021) attempted to predict the final stock market trend by including the sentiment feature using multiple machine learning techniques. Their findings emphasized the weaknesses of the different machine learning methods employed and proved that social media and news articles improve the performance of the forecasting model. There are several additional impressive studies conducted by several authors in this specific area of research, including, Vanstone et al. (2019), Jin et al.

(2020), Aasi et al (2021), Gupta et al. (2022), Swathi et al. (2022), highlighting the prominent importance of the inclusion of the sentiment feature.

3. Methodology

Since no process can be completed effectively without consistent inputs, this chapter displays data related information, and provides a thorough and detailed explanation of the adopted methodology, including the steps undertaken to efficiently answer the raised research questions.

3.1. Tools and Techniques of Data Collection

This paper aims to build multiple deep learning models trying to predict future S&P 500 levels based on currently available data and investors' sentiments, and investigate if the latter can cause stock prices to deviate systematically from their fundamental value. This analysis was performed over a period of ten years, starting from the 28th of December, 2011, until the 30th of December, 2021. The aim was to capture a large timespan and extend the period as much as possible to be able to feed the recurrent neural network (RNN) models a substantial amount of data, to achieve better outcomes. The main reason why it did not cover 2022 is due to the lack of data, as the goal is to base the analysis on a full year and not a fragment of a year.

To competently conduct this study, three features of fundamental prominence need to be carefully addressed. Firstly, the equity market share of the research was represented by the S&P 500 index. Daily financial figures were employed. Specifically, 2538 daily observations of the S&P 500 were directly fetched using Python Software from Yahoo Finance, with the help of a built-in library called 'yfinance'. The second type of feature is the sentiment one. As mentioned previously, it is represented by social media content, specifically, content extracted from Twitter. The choice of Twitter was not random, yet it was founded on extensive readings. Twitter's portion of financial content is the highest among other social media platforms, placing it among the elites, where investors post and seek investment ideas and advices. Tweets were directly extracted from Twitter via their Twitter API v2 product, dedicated for academic researchers. Specific keywords were used to efficiently target the index, as neglecting this matter can undoubtedly lead to a spurious analysis. The keywords employed were 'SP500', the most frequently used hashtag of the addressed index, and 'equities', a general term widely used when posting content related to the stock market. A total of 5,067,618 tweets were retrieved and quantified. However, they were downsized to 3,657 entries, as the homogeneity of the data is crucial. The frequency of the dataset was altered to daily by averaging all the values of a given day to end up with a single value representing the average daily sentiment. Concerning the technical indicators, the Moving Average, Relative Strength Index, and Bollinger Bands were chosen. These indicators were directly computed in Python software and not accessed from any other source using the S&P 500 historical records were used, particularly the closing prices. The process terminated with 2,519 observations for the Moving Average, 2,524 entries for the Relative Strength Index, and 2,519 entries for each Bollinger band. When merging all features into a single data frame, the number of

entries was scaled down significantly. The final dataset, which served as input for both models created, comprised 2,519 rows and nine features, representing daily figures over a ten years period.

Table 1. Inputs of both models.

	Technical Model	**Sentiment Model**
Dependent Variable	Daily Closing Price of the S&P500	Daily Closing Price of the S&P500
Independent Variables	**Historical records of the S&P500:** • Opening • Closing • High • Low **Technical Indicators:** • 20D Moving Average • Relative Strength Index • Bollinger Bands	**Historical records of the S&P500:** • Opening • Closing • High • Low **Technical Indicators:** • 20D Moving Average • Relative Strength Index • Bollinger Bands **Sentiment Feature:** • Quantified Twitter Content

3.2. Empirical Framework

In order to efficiently meet the objectives of this research paper and answer our research questions, a specific approach comprising several crucial steps was undertaken. The entire process was executed using Python Programming Language, with the help of Google Colaboratory.

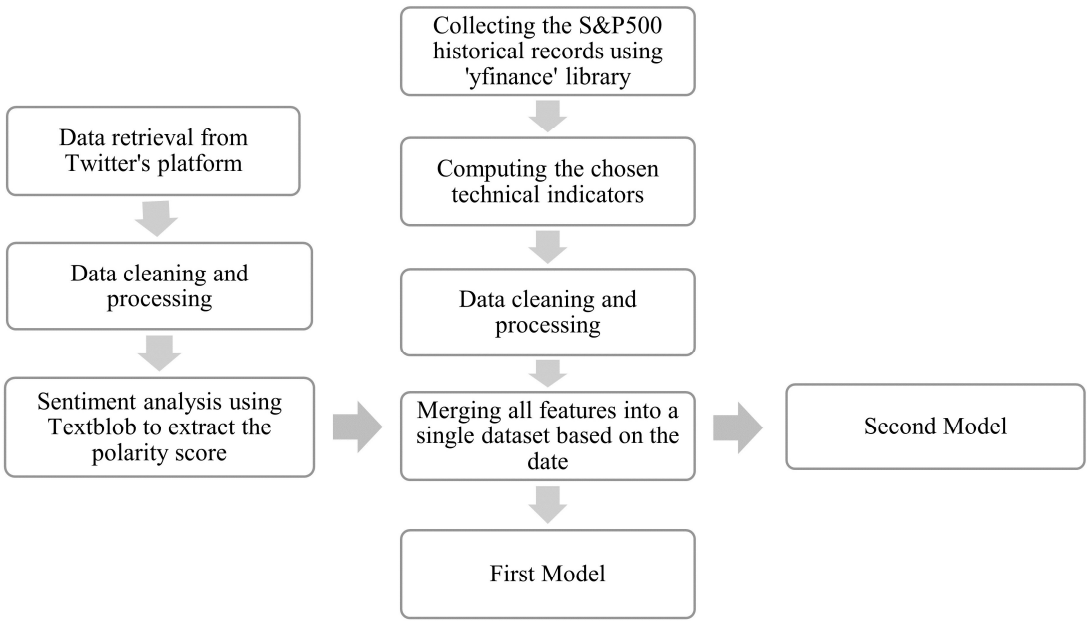


Figure 1. Data analysis process.

The figure above highlights the different phases that shall be implemented to successfully reach the goals of this study. These steps will be well elaborated in the following section.

3.2.1. Twitter data retrieval

The first step through the procedure is regarded as the most challenging one. It consists of extracting historical tweets directly through Twitter API v2. product. The maximum number of tweets that can be extracted on a monthly basis is limited to 10,000,000. When acquiring access to this product, a bearer token is granted. It is the key to be able to retrieve the data. This token is critical, confidential, and shall not be disclosed, to prevent anyone from obtaining the data. Thus, the ultimate way for preserving it is to store it in an environment variable. After preserving it that way, two functions were defined, the first one is a function to call bearer token back so that we can kick off the data retrieval process, and another one that outputs headers that are needed to reach the Twitter API. After successfully reaching the API, the query in favor of the endpoint that will be utilized shall be created, as well as the arguments or parameters that the retrieval process will be based on. The function includes two main components. The first component is the search URL because this specific API has several endpoints, thus, determining the one we need to collect the data from is crucial. As for us, the endpoint that is employed is the 'full-archive search' granting us full archival access to public tweets since March 2006, the date that marks the appearance of the very first tweet. The second component is related to the parameters, which are the parameters provided by the chosen endpoint. These queries are entirely customizable relative to the outputs we are willing to obtain. The desired parameters are split into three main categories. The first category is related to the components that require to be managed and curbed. The second

category is mainly discretionary and relative, our main concern is to retrieve the text content and the date and time each tweet was posted, thus, the query parameters that were specified pivoted around two main areas, the time component, and the text-only. And finally, the following token query parameter is employed to set off the subsequent request to retrieve additional tweets as the maximum result per request is 500 tweets only. Since all the required steps are set, a new function is built that fits in the API website, headers, and queries, to link them to the endpoint. The function's prominent role is to send a request, and once all is in order, the response will be reverted as a JavaScript Object Notation or JSON form. Indeed, CSV files, compared to JSON files, are easier to handle when it comes to data analysis and processing. Hence, using the JSON response, a new function is built that consists of appending all of the obtained content into a CSV format. Eventually, that way, everything is set to initiate the retrieval process, however, they need to be fit jointly in a way to successfully complete this task. The main aim of this study is to collect data throughout the period of ten years, thus, we shall automate the process so that it loops over the prementioned timespan. However, ensuring a fair distribution among the specified duration is a must to prevent inadequate dispersion of data. To achieve that, we must specify a cap for the number of tweets we want to retrieve per interval of time, and if this number is hit, directly it proceeds to the subsequent request. Thus, we opted for a six-hour interval with a maximum result of 500 tweets.

This approach contributed to gathering a total of 5,067,618 tweets based on two specific keywords, that are, 'SP500', the most frequent hashtag used, and 'equities', a popular notion frequently included when tackling the stock market, and the language was set to English to prevent any data that can be misleading to the analysis later on.

3.2.2. Twitter sentiment analysis

The ultimate approach to making use of textual data is to quantify them in order to be fed to the models and generate appropriate results. Therefore, performing the sentiment analysis on the data retrieved from Twitter is essential.

The content extracted, like any textual data, is nothing more than a string of terms, specifically, a string of letters. However, using techniques like Natural Language Processing (NLP) for sentiment analysis requires considering each term composing each sentence by itself rather than depicting each character alone. Thus, the cleaning process shall assertively be performed to improve the quality of our data for a better analysis. In this paper, it was done by creating a 'cleaning function' aiming to eliminate any unnecessary content that may deceive the analysis, starting by making all characters lowercases and removing non-alphabetical characters, retweets, and any hyperlinks. Once the cleaning process is done and properly applied, we can proceed to the sentiment analysis process.

Since the data extracted from Twitter are raw, unlabelled, the TextBlob library is utilized to extract the 'Sentiments' out of the tweets to employ them later on into our model. It is one of many open-source libraries dedicated to data processing that performs sentiment analysis, using a step of predetermined criteria and a dictionary. It assigns to each textual content a value helping

in computing two main features, the subjectivity, and polarity of each given statement. The subjectivity that ranges between zero and 1, from subjective to objective, refers to personal opinion, mood, or judgment, in addition to the polarity, a float that ranges between -1 and 1 highlighting the positive and negative statements. The value is determined based on the textual data's lexical direction along with the amplitude of every term forming the entire phrase. Thus, this process necessitates the use of a predetermined vocabulary that categorizes each term. Textual data are often analyzed as a bag of words. When allocating values to each term forming the sentence, the overall score is derived using an aggregating averaging technique to end up with a single value.

Nevertheless, the homogeneity of the dataset is crucial. Therefore, after obtaining the outcomes of the sentiment analysis, the frequency of the data frame shall be changed to daily, as the main aim of this analysis is to predict the daily closing price. By taking this step, the entries were massively downsized from 5,067,618 entries to 3,657 only by computing the average value of all the given sentiments per day.

3.2.3. Technical analysis

Figure 2. S&P 500 daily candlesticks chart.

Historical records are directly loaded from Yahoo Finance with the help of 'yfinance' library. This library grants users access to any asset's past data. For the S&P 500, only the Opening, High, Low, and Closing prices were retrieved. With the help of the 'Plotly' library, we were able to illustrate these historical records in a single candlesticks chart which is represented in the figure above. These records not only served as inputs for the models built, but they were also used to compute the chosen technical indicators that are the following, the 20 days Simple Moving Average, Relative Strength Index, and Bollinger Bands, including the lower and upper band. Undoubtedly, and as mentioned previously, this process generated null values. There are several methods to treat missing values, however, the most suitable option, in this case, was to drop them instead of filling them in an unrealistic way. All the historical figures covered a period of ten years same as the Twitter retrieval process, however, an additional month was added to prevent losing any data and preserve the predetermined timespan since this process generated missing values.

3.2.4. Deep learning models

After properly preparing the datasets, both data frames shall be merged; historical data, technical indicators, along with the outcomes of the sentiment analysis, to proceed for the models' creation. However, before that, data must be well prepared and processed. Starting with selecting the features that shall be included in the related model, along with proceeding to the scaling process, this step is very important as scaled data are deemed to have a favorable influence on the overall performance of the model. Then, after effectively completing this initial step, these scaled values shall be transformed into a three-dimensional shape, as it is the shape employed to train multivariate regression models.

Data were split into 80% for training and 20% for testing. The training period started on the 28th of December, 2011, and ended on the 1st of February, 2020. On the other hand, the testing period rendered from the 2nd of February, 2020, until the 30th of December, 2021. Two Recurrent Neural Network (RNN) models were built. The first one is based on historical data and technical indicators solely. The second one will combine these data with the outputs of the sentiment analysis to train the neural network model helping in predicting future price trends. Besides their extensive ability to process large amount of data, the reason why we opted for deep learning and recurrent neural network (RNN), is that this is a time-series task, and such type of models has throughout the years proven its efficacity in treating sequential problems.

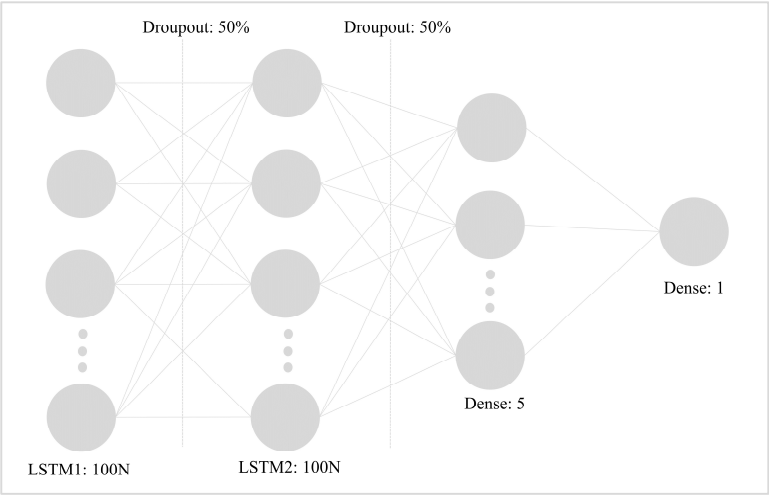


Figure 3. Both models' architecture.

Both models have an identical architecture. Same steps are repeated, no modifications to the models' architecture are made. However, features are solely altered to reach the goal of this study and examine to what extent can the inclusion of the sentiment feature affect the S&P 500 price prediction. The models consist of two LSTM layers, 100 neurons each; in between each one of them, dropout layers of 50% were added to prevent overfitting by randomly dropping ineffective units from the neural network, along with a final dense layer of a single neuron,

generating the forecasted values. When fitting, epochs were specified at 40 and a batch size of 128. And the window for making a sole forecast is set to 50, thus, in other words, the past 50 values were used to try to predict the 51st closing price.

As for evaluating the models' performance, several evaluation metrics were chosen, namely, the Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Median Absolute Percentage Error (MDAPE). Their formulas are respectively listed below:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}|\frac{y_i - \hat{y}_i}{y_i}|$$

$$MDAPE = \frac{1}{n}\sum_{i=1}^{n}|\frac{y_i - \hat{x}_i}{y_i}|$$

where yi represents the predicted value, ŷi represents the actual value, and x̂i represents the median of the sample. Starting with the Mean Absolute Error, which is a measure of accuracy, a metric that assesses, in an absolute way, the aggregate degree of errors in a series of projections, sans considering their direction, the mean absolute percentage error is the percentage value of the mean absolute error. And finally, the Median Absolute Percentage Error, (MdAPE) an error unit, that reveals the level of error as a ratio, it is computed by sorting the percentage error values, without considering their directions, in an absolute way, and attributing the median as the midpoint.

4. Results and Discussion

A detailed explanation of the outcomes of each process performed is provided in this chapter aiming to give an explicit explanation of the attained results. Additionally, a specific part dedicated to exploring and interpreting these findings.

4.1. Results from quantitative analysis

4.1.1. Twitter data retrieval

	created_at	tweet
0	2021-12-31 23:59	RT @BigBreakingNow: $AMC Block Deals\nhttps://...
1	2021-12-31 23:57	@SilverChartist Is it too late to get in? Or s...
2	2021-12-31 23:53	@CyclesFan Does risk off in crypto through 202...
3	2021-12-31 23:53	RT @BArmstrong_TDA: #SP500 That's a wrap on 20...
4	2021-12-31 23:52	And our year in the markets is complete. SP500...

Figure 4. Sample of Twitter's dataset.

Figure 4 represents a sample of the final dataset acquired after completing the data retrieval process from Twitter through their Twitter API v2. product. As seen in the figure above, tweets are raw and uncategorized. Unlabelled and unprocessed data are one of the main challenges that data scientists face. In our instance, our major goal is to evaluate and extract the overall mood out of these tweets. Thus, in this regard, we opted for TextBlob public library to process these textual data and perform the sentiment analysis by extracting investors' moods out of the tweets and label them to serve later on as inputs for the models that shall be created.

4.1.2. Sentiment analysis

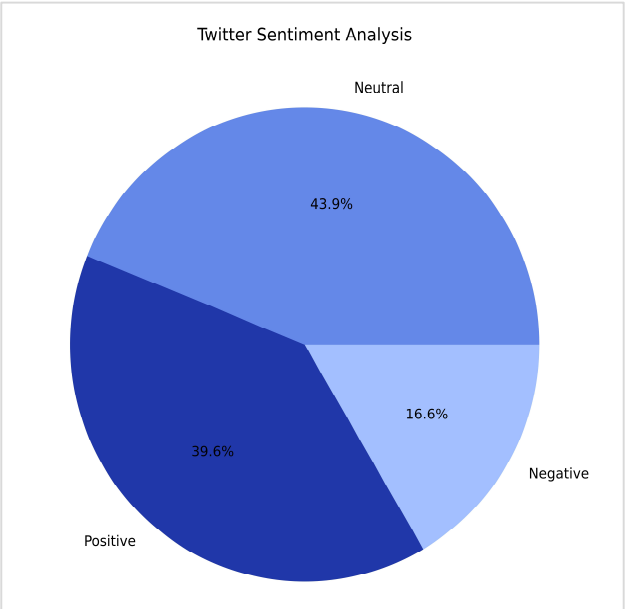


Figure 5. Twitter sentiment analysis.

Results of the sentiment analysis revealed that 39.6% of the tweets are positive, 16.6% are

negative, and 43.9% of tweets are neutral. The neutral category may seem alarming and eventually, if left untreated, may be misleading, as feeding the models unbiased data is futile, will hinder the training process since the models will not efficiently learn. Nevertheless, this weakness will be surmounted, as this specific class will disappear when resampling the data, and we will end up with only two categories, positive and negative.

As mentioned previously, the homogeneity of the inputs is crucial to be able to efficiently conduct this analysis. However, Twitter content was retrieved randomly based on a six-hours interval over a period of ten years, and the sentiment analysis was done accordingly. Thus, the frequency of the data frame shall be converted to daily.

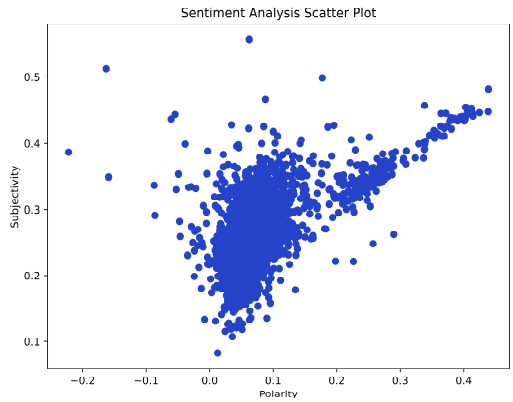
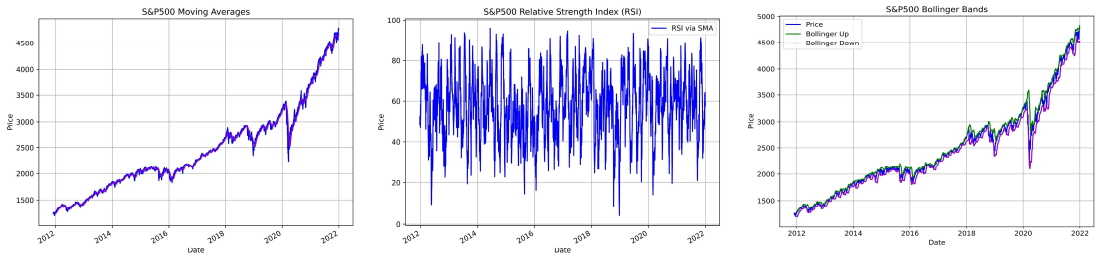


Figure 6. Sentiment analysis scatter plot.

The resampling process not only led to downsizing the data from 5,067,618 entries to 3,657, but it also contributed to getting rid of the 'Neutral' category. The outcome was illustrated using the scatter plot represented in the figure above, which portrayed the Polarity of the data, and the Subjectivity of the given judgments rendering from objective to subjective. Data visualization, generally speaking, helps in emphasizing analytical insights. Hence, this plot is of significant importance, as it helps us deepen our understanding of the given inputs. It reveals that the majority of the tweets are skewed to and pivot around the 'Positive' area. Moreover, it highlights the existence of some outliers, emphasizing the presence of highly positive or negative tweets or exceedingly objective or subjective statements.

4.1.3. Technical analysis



(a) (b) (c)

Figure 7. Technical indicators (a) 20d Moving Average, (b) Relative Strength Index, (c) Bollinger Bands.

Concerning the technical analysis results, using the S&P 500 historical prices, we were able to manually compute the selected technical indicators based on their formulas, in particular the 20 days Simple Moving Average, the Relative Strength Index, and Bollinger bands. The preceding charts highlight the outcomes of this process.

4.1.4. Deep learning models

Date	Open	High	Low	Close	20d	RSI	bollinger_up	bollinger_down	Subjectivity	Polarity
2011-12-28 00:00:00+00:00	1265.380005	1265.849976	1248.640015	1249.640015	1241.805988	46.990317	1277.831737	1205.780238	0.275533	0.079378
2011-12-29 00:00:00+00:00	1249.750000	1263.540039	1249.750000	1263.020020	1242.608990	58.162977	1279.815060	1205.402921	0.280493	0.042938
2011-12-30 00:00:00+00:00	1262.819946	1264.119995	1257.459961	1257.599976	1243.259991	50.752241	1281.062119	1205.457864	0.268859	0.049750
2012-01-03 00:00:00+00:00	1258.859985	1284.619995	1258.859985	1277.060059	1244.898993	62.611068	1285.617362	1204.180624	0.234749	0.066993
2012-01-04 00:00:00+00:00	1277.030029	1278.729980	1268.099976	1277.300049	1245.909998	67.140838	1288.845546	1202.974449	0.249881	0.066580

Figure 8. Merged dataset.

The figure above is a representation of the final dataset that will be fed to the models. After properly processing the datasets, the outcomes of the sentiment analysis, historical records, along the results of the technical analysis were merged, forming a single data frame. All the features are quantitative, combined based on a single criterion, the date, which starts on the 28th of December, 2011. This dataset is of paramount importance, as its components shall serve as inputs for the models created, which justifies the substantial consideration devoted to the data processing part in order to obtain homogenous and compelling data to feed the models.

Models were built based on the same framework and hyperparameters, with no amendments in their architecture, except for the features. As stated before, the features of the first model were the index's technical indicators along with its historical records. On the other hand, the second model included the 'Polarity' feature, which is the additional feature that will aid in answering our research questions.

Table 2. Evaluation metrics.

	Technical Model	Combined Model
MAE	126.57	72.44
MAPE	3.37%	2.16%
MDAPE	2.95%	1.30%

In a holistic view, the evaluation metrics of both models, used as a tool to help in assessing the performance of the models created, are considerable while revealing substantial outcomes. However, it is essential to compare the models' metrics to get a clearer view of how the inclusion of the sentiment feature affects each model's performance and therefore if it significantly impacts

the equity market represented by the S&P 500 index. These outcomes achieved are important and considerable. Nonetheless, the second model generated better outcomes on all levels. The MAE of the second model is nearly 50 percent lower than the MAE of the technical model. And this also applies to the MAPE and the MDAPE, both metrics dropped significantly when including the sentiment feature, showcasing the notable and crucial role that this specific feature plays.

Nevertheless, building models and reaching a significant performance on their own are not enough, yet, meaningless. It should be paired with multiple tasks to prove their efficiency and portray their performance. As numbers remain dull, we shall find an efficient alternative way to prove their convenience.

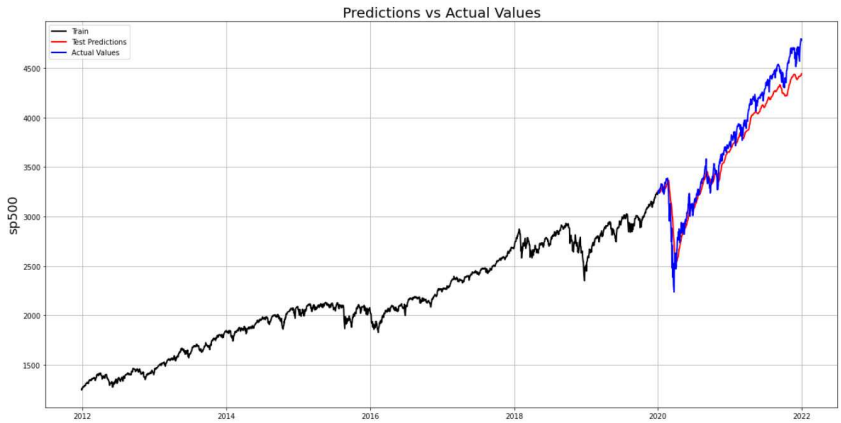


Figure 9. Predictions vs. actual values, first model.

Figure 9 represents a line chart comparing the projected values generated by the first model that is based on technical features solely, against the actual values or real closing prices of the S&P 500. The entries visualized were split training and testing throughout the entire time span of the analysis. As demonstrated by the chart above, the projected values are near the real numbers. However, we can observe discrepancies, especially throughout courses of growing uncertainty, instability, and volatility, where they are the strongest. On the other hand, during consolidation or steady phases, when the market remains and is described as hesitant, the spread between the predicted prices and the actual ones is low.

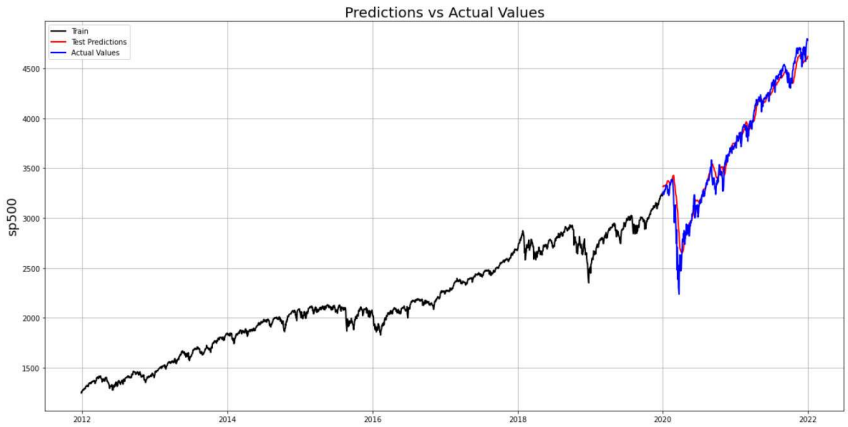


Figure 10. Predictions vs. actual values, second model.

When visualizing the outcomes based on the second model, which included the sentiment feature along with the technical ones, it contributed to generating better outcomes, this was proved earlier through all the evaluation metrics used to assess the performances of both models. According to the chart above, we can observe that the predictions are closer to the real values when compared to the forecasts of the first model, especially at normal times. Nevertheless, this does not mean that there are no deviations during such periods and periods of increased uncertainty; however, the chart makes it conspicuous that the performance of the second model outdoes the performance of the model that relies purely on technical indicators. Since the second model is more efficient and provides better results, we attempted to project the next day's closing price of the S&P 500 based on it using the last 50 values. The predicted price deviated from the original price by -2.79%. This difference may seem notable to some and acceptable to others, yet it is crucial to consider and bear in mind that predicting the stock market is feasible, however, it has been and will always remain a challenging and complex task. Modern computational techniques and the inclusion of several features have played an extensive role in this area, nevertheless, the main aim is to refine and improve to reach better outcomes.

4.2. Discussions

All the above work was executed in order to answer this paper's research questions pivoting around to what extent can sentiments extracted from social media platforms improve the predictability of the S&P 500 price levels by checking if it improves the model's performance.

Primarily, it is prominent that the first model was built to serve as a tool helping in answering the prespecified research questions and reaching the primary goal of this paper. In terms of evaluation metrics, this model generated favorable outcomes. This denotes that the assets' historical features may play a significant role when attempting to forecast their fundamental value. This aligns with the outcomes of several pieces of literature sharing this exact topic, including the findings of Singh et al. (2021), that based their analysis on technical indicators and proved that it is an efficient tool for stock price forecasting. Nevertheless, the choice of the technical indicators shall not be random and shall be based on thorough research, as they may have an adverse impact on the model's overall performance, thus, it is advised to correctly pick such variables. Oriani and Coehlo (2016) highlighted how lagging technical features can improve the prediction power of the forecasting model. Concerning the second model created, the implementation of the sentiment variable played a significant role in reaching this research's objectives. It contributed to proving that daily sentiment scores obtained by quantifying the S&P 500's related tweets have a substantial ability to affect the latter price movements. Reaching this assessment was due to the comparison that was made between the chosen evaluation metrics of both models. The results highlighted an improvement in the model's performance when the sentiment feature was involved in the forecasting process. This aligns with the findings of the majority of research papers related to this topic, especially the results of one of the most important papers of all time written by Bollen et al.

(2011), highlighting the enhancement of the models' overall performance when the public mood parameters derived from Twitter content are included. However, when conducting the sentiment analysis on textual data, the selected library that will perform the Natural Language Processing technique (NLP) may fail to catch the adverse sentiments, thus, the choice of such tools is critical and has to be founded on extensive readings, as it affects the outcomes of the analysis, and not choosing the proper library may be misleading and deteriorate the performance of the created model.

5. Conclusion

This paper focuses on examining to what extent sentiments extracted from social platforms improve the predictability of the S&P 500 price levels, and checking if the implementation of such data improves the model's aggregate performance. To be able to lead this study efficiently, two models were created, one that takes the technical features as inputs and the second one that combines the inputs of the first model with the sentiment feature. Thus, the contributions that this paper offers are various. Firstly, by presenting two deep learning hybrid forecasting models aiming to anticipate the price movement of the S&P 500 that blends social media content with technical indicators. Secondly, this paper contributes to proving that the social media sentiment feature is one of the key players that affects and influences the stocks' fundamental values or price levels. And thirdly, one of the main potencies of this paper is the notable performances of the two proposed models, which were not achieved arbitrarily yet based on extensive research and several experiments.

While examining the performance of both models created, the second model that included the sentiment feature achieved better results than the model purely based on technical indicators. The improvement was portrayed through the evaluation metrics; improved results were seen in all the metrics used, highlighting the considerable significance of behavioral finance. Additionally, the outcomes of this study are coherent with past findings. This research's findings are of significant importance to this field since it proposes predictive models based on advanced computational methods when compared to conventional statistical approaches.

Concerning the limitations of this study, to collect data from Twitter, a specific code was written. However, the execution process, when running successfully, and if it did not encounter any errors, is heavily time-consuming. Moreover, the process of data cleaning by itself required to be performed with extra caution since we are handling large amount of data.

As for the further research, it is interesting to perform the analysis on different market indices or stocks to check how they will influence the models' performance. Additionally, in order to determine the efficacity of the proposed model, this analysis can be extended by including a program that consists of generating trading signals based on the forecasted values of the sentiment model. Flags shall be originated founded on predetermined criteria, and according to the movement of the asset's predicted price, trades are executed. Following such strategies, decent returns can be

generated. Indeed, market simulation must be carried out before proceeding to reveal how well the strategy, if implemented, can perform and to highlight its drawdowns.

References

Aasi, B., Imtiaz, S. A., Qadeer, H. A., Singarajah, M., & Kashef, R. (2021, April). Stock Price Prediction Using a Multivariate Multistep LSTM: A Sentiment and Public Engagement Analysis Model. In *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)* (pp. 1-8). IEEE. doi:10.1109/IEMTRONICS52119.2021.9422526

Attanasio, G., Cagliero, L., Garza, P., & Baralis,E. (2019, November). Combining news sentiment and technical analysis to predict stock trend reversal. In *2019 International Conference on Data Mining Workshops (ICDMW)* (pp. 514-521). IEEE. doi:10.1109/ICDMW.2019.00079

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of computational science*, *2*(1), 1-8. doi:10.1016/j.jocs.2010.12.007

Chen, S., & Ge, L. (2019). Exploring the attention mechanism in LSTM-based Hong Kong stock price movement prediction. Quantitative Finance, 19(9), 1507-1515. doi:10.1080/14697688.2019.1622287

Deng, J. (2020). *Stock Prediction by Analyzing Financial News Sentiment and Investor Mood of Social Media*. University of California, Santa Cruz.

Fama, E. F. (1981). Stock Returns, Real Activity, Inflation, and Money. *The American Economic Review*, *71*(4), 545–565. http://www.jstor.org/stable/1806180

Feng, F., He, X., Wang, X., Luo, C., Liu, Y., & Chua, T. S. (2019). Temporal relational ranking for stock prediction. *ACM Transactions on Information Systems (TOIS)*, *37*(2), 1-30. doi:10.1145/3309547

Gupta, I., Madan, T. K., Singh, S., & Singh, A. K. (2022). HiSA-SMFM: Historical and Sentiment Analysis based Stock Market Forecasting Model. *arXiv preprint arXiv:2203.08143*. doi:10.48550/arXiv.2203.08143

Jiang, W. (2021). Applications of deep learning in stock market prediction: recent progress. *Expert Systems with Applications*, *184*, 115537. doi:10.1016/j.eswa.2021.115537

Jiao, Y., & Jakubowicz, J. (2017, December). Predicting stock movement direction with machine learning: An extensive study on S&P 500 stocks. In *2017 IEEE International Conference on Big Data (Big Data)* (pp. 4705-4713). IEEE. doi:10.1109/BigData.2017.8258518

Jin, Z., Yang, Y., & Liu, Y. (2020). Stock closing price prediction based on sentiment analysis and LSTM. *Neural Computing and Applications*, *32*(13), 9713-9729. doi:10.1007/s00521-019-04504-2

Kesavan, M., Karthiraman, J., Ebenezer, R. T., & Adhithyan, S. (2020, May). Stock market prediction with historical time series data and sentimental analysis of social media data. In *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 477-482). IEEE. doi:10.1109/ICICCS48265.2020.9121121

Ko, C. R., & Chang, H. T. (2021). LSTM-based sentiment analysis for stock price forecast. *PeerJ Computer Science*, *7*, e408. doi:10.7717/peerj-cs.408

Lee, H. S. (2020). Exploring the initial impact of COVID-19 sentiment on US stock market using big data. *Sustainability*, *12*(16), 6648. doi:10.3390/su12166648

Li, X., Wu, P., & Wang, W. (2020). Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong. *Information Processing & Management*, *57*(5), 102212. doi:10.1016/j.ipm.2020.102212

Liu, H. (2018). Leveraging financial news for stock trend prediction with attention-based recurrent neural network. *arXiv preprint arXiv:1811.06173*. doi:10.48550/arXiv.1811.06173

Malkiel, B. G. (1973). A random walk down Wall Street. W. W.

Matsubara, T., Akita, R., & Uehara, K. (2018). Stock price prediction by deep neural generative model of news articles. *IEICE TRANSACTIONS on Information and Systems*, *101*(4), 901-908. doi:10.1587/transinf.2016IIP0016

Mehtab, S., Sen, J., & Dutta, A. (2020, October). Stock price prediction using machine learning and LSTM-based deep learning models. In *Symposium on Machine Learning and Metaheuristics Algorithms, and Applications* (pp. 88-106). Springer, Singapore. doi:10.1007/978-981-16-0419-5_8

Mohan, S., Mullapudi, S., Sammeta, S., Vijayvergia, P., & Anastasiu, D. C. (2019, April). Stock price prediction using news sentiment analysis. In *2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)* (pp. 205-208). IEEE. doi:10.1109/BigDataService.2019.00035

Moohambikai, M.M., & Bhakirathi, B. (2020). PREDICTION OF STOCK PRICE WITH ENHANCED DEEP LEARNING MODEL.

Naik, N., & Mohan, B. R. (2019, May). Stock price movements classification using machine and deep learning techniques-the case study of indian stock market. In *International Conference on Engineering Applications of Neural Networks* (pp. 445-452). Springer, Cham. doi:10.1007/978-3-030-20257-6_38

Nemes, L., & Kiss, A. (2021). Prediction of stock values changes using sentiment analysis of stock news headlines. *Journal of Information and Telecommunication*, 1-20. doi:10.1080/24751839.2021.1874252

Oriani, F. B., & Coelho, G. P. (2016, December). Evaluating the impact of technical indicators on stock forecasting. In *2016 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1-8). IEEE. doi:10.1109/SSCI.2016.7850017

Pahwa, K., & Agarwal, N. (2019, February). Stock market analysis using supervised machine learning. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)* (pp. 197-200). IEEE. doi:10.1109/COMITCon.2019.8862225

Ravikumar, S., & Saraf, P. (2020, June). Prediction of Stock Prices using Machine Learning (Regression, Classification) Algorithms. In *2020 International Conference for Emerging Technology (INCET)* (pp. 1-5). IEEE. doi:10.1109/INCET49848.2020.9154061

Sachdeva, A., Jethwani, G., Manjunath, C., Balamurugan, M., & Krishna, A. V. (2019, March). An effective time series analysis for equity market prediction using deep learning model. In 2019 International Conference on Data Science and Communication (IconDSC) (pp. 1-5). IEEE. doi:10.1109/IconDSC.2019.8817035

Sanboon, T., Keatruangkamala, K., & Jaiyen, S. (2019, February). A deep learning model for predicting buy and sell recommendations in stock exchange of thailand using long short-term memory. In *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)* (pp. 757-760). IEEE. doi:10.1109/CCOMS.2019.8821776

Sethia, A., & Raut, P. (2019). Application of LSTM, GRU and ICA for stock price prediction. In *Information and communication technology for intelligent systems* (pp. 479-487). Springer, Singapore. doi:10.1007/978-981-13-1747-7_46

Shen, J., & Shafiq, M. O. (2020). Short-term stock market price trend prediction using a comprehensive deep learning system. *Journal of big Data*, *7*(1), 1-33. doi:10.1186/s40537-020-00333-6

Sidogi, T., Mbuvha, R., & Marwala, T. (2021, October). Stock Price Prediction Using Sentiment Analysis. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 46-51). IEEE. doi:10.1109/SMC52423.2021.9659283

Singh, A., Gupta, P., & Thakur, N. (2021). An Empirical Research and Comprehensive Analysis of Stock Market Prediction using Machine Learning and Deep Learning techniques. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1022, No. 1, p. 012098). IOP Publishing. doi:10.1088/1757-899x/1022/1/012098

Song, H., Peng, D., & Huang, X. (2020). Incorporating research reports and market sentiment for stock excess return prediction: a case of mainland china. *Scientific Programming*, *2020*. doi:10.1155/2020/8894757

Swathi, T., Kasiviswanath, N., & Rao, A. A. (2022). An optimal deep learning-based LSTM for stock price prediction using twitter sentiment analysis. *Applied Intelligence*, 1-14. doi:10.1007/s10489-022-03175-2

Tversky, A., & Kahneman, D. (1989). Rational choice and the framing of decisions. In *Multiple criteria decision making and risk analysis using microcomputers* (pp. 81-126). Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-74919-3_4

Vanstone, B. J., Gepp, A., & Harris, G. (2019). Do news and sentiment play a role in stock price prediction? *Applied Intelligence*, *49*(11), 3815-3820. doi:10.1007/s10489-019-01458-9

Wang, Z., Zhenda, H. U., Fang, L. I., & Seng-Beng, H. O. (2021). Learning-Based Stock Market Trending Analysis by Incorporating Social Media Sentiment Analysis. doi:10.21203/rs.3.rs-181424/v1

Wu, S., Liu, Y., Zou, Z., & Weng, T. H. (2021). S_I_LSTM: stock price prediction based on multiple data sources and sentiment analysis. *Connection Science*, 1-19. doi:10.1080/09540091.2021.1940101

Yu, P., & Yan, X. (2020). Stock price prediction based on deep neural networks. *Neural Computing and Applications*, *32*(6), 1609-1628. doi:10.1007/s00521-019-04212-x

Zhang, Z., Zohren, S., & Roberts, S. (2019). Deeplob: Deep convolutional neural networks for limit order books. IEEE Transactions on Signal Processing, 67(11), 3001-3012. doi:10.1109/TSP.2019.2907260