

## Article

# Unleashing the Power of Tweets and News in Stock-Price Prediction Using Machine-Learning Techniques

Hossein Zolfagharinia <sup>1,\*</sup> , Mehdi Najafi <sup>1</sup>, Shamir Rizvi <sup>1</sup> and Aida Haghighi <sup>2</sup>

<sup>1</sup> Global Management Studies Department, Ted Rogers School of Management, Toronto Metropolitan University, Toronto, ON M5B 2K3, Canada; najafi.mehdi@torontomu.ca (M.N.); shamir.rizvi@torontomu.ca (S.R.)

<sup>2</sup> School of Occupational and Public Health, Faculty of Community Services, Toronto Metropolitan University, Toronto, ON M5B 2K3, Canada; aida.haghighi@torontomu.ca

\* Correspondence: h.zolfagharinia@torontomu.ca; Tel.: +1-416-979-5000 (ext. 557532)

**Abstract:** Price prediction tools play a significant role in small investors' behavior. As such, this study aims to propose a method to more effectively predict stock prices in North America. Chiefly, the study addresses crucial questions related to the relevance of news and tweets in stock-price prediction and highlights the potential value of considering such parameters in algorithmic trading strategies—particularly during times of market panic. To this end, we develop innovative multi-layer perceptron (MLP) and long short-term memory (LSTM) neural networks to investigate the influence of Twitter count (TC), and news count (NC) variables on stock-price prediction under both normal and market-panic conditions. To capture the impact of these variables, we integrate technical variables with TC and NC and evaluate the prediction accuracy across different model types. We use Bloomberg Twitter count and news publication count variables in North American stock-price prediction and integrate them into MLP and LSTM neural networks to evaluate their impact during the market pandemic. The results showcase improved prediction accuracy, promising significant benefits for traders and investors. This strategic integration reflects a nuanced understanding of the market sentiment derived from public opinion on platforms like Twitter.

**Keywords:** stock-price prediction; neural network; LSTM; multi-layer perceptron; news count



**Citation:** Zolfagharinia, H.; Najafi, M.; Rizvi, S.; Haghighi, A. Unleashing the Power of Tweets and News in Stock-Price Prediction Using Machine-Learning Techniques. *Algorithms* **2024**, *17*, 234. <https://doi.org/10.3390/a17060234>

Academic Editor: Jesús Ángel Román Gallego

Received: 8 March 2024

Revised: 17 May 2024

Accepted: 24 May 2024

Published: 28 May 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The dynamic landscape of the global stock market plays a significant role in shaping economies, influencing individual financial decisions, and driving continuous innovation in investment strategies. The indisputable significance of the stock market is underscored by the growth in total global capitalizations, which exceeded USD 109 trillion in 2023—a remarkable threefold increase from the 2009 figure of USD 25 trillion [1,2]. As per World Bank data from 2017, stock trading's substantial impact on the American economy has been evident since 2013, with the total value of stocks traded on US markets consistently surpassing 200% of the nation's annual GDP. The New York Stock Exchange (NYSE)—a symbolic hub of financial activity—boasts an average market capitalization of approximately USD 29 trillion, highlighting its central role in global financial markets [3]. The daily average of about USD 169 billion in stocks traded on the NYSE further emphasizes the fluidity and dynamism of the market [4].

Beyond its macroeconomic influence, the movement of stocks on a micro level is crucial in determining the financial market's overall well-being. Notably, studies such as that of Chan and Woo [5] reveal that stock-market price booms can drive long-term economic growth. The involvement of novice investors in the stock market adds another layer of complexity to the financial ecosystem. In particular, there are now over 54% of US adults having some form of investment in the stock market [6]. This increase in individual investors can be attributed to the post-2008 financial crisis [7]. In this evolving landscape,

the need for reliable stock-price prediction models increases, especially as financial products and services become more accessible to smaller investors.

Moreover, price prediction tools play a significant role in small investors' behavior. As such, numerous research studies (e.g., [8–23]) have been conducted to develop appropriate models and techniques that would widely be employed in algorithmic trading. Algorithmic trading, in essence, leverages computational power and mathematical formulas to illuminate buy or sell decisions for financial securities on an exchange and incorporates complex formulas and models with human oversight [24–27]. These techniques are integral to institutional firms' trading algorithms, as they aid in minimizing transaction costs and market risks [28]. The rise of artificial intelligence (AI) in both the stock market and financial firms has significantly contributed to the growth of the algorithmic trading market. Companies like Sentient have developed AI-powered algorithmic traders, thereby showcasing the potential for these advanced algorithms to function as standalone entities [29,30].

While numerous studies have developed models and techniques utilizing AI methods in stock-price prediction, they often focus on technical indices and general data in a normal market context. More specifically, these studies frequently overlook the impact of news on the decisions of small traders—particularly in market-panic scenarios. The profound impact of the media—most notably the news and Twitter—on investors' decisions in relation to stock buying and selling is evident. This has greatly shaped the modern financial landscape. News that is disseminated through traditional media or online platforms can swiftly influence investor sentiment by providing critical information about companies, industries, and the broader economic landscape. For example, during the COVID-19 pandemic, the demand for oil rapidly declined due to business closures and travel restrictions; this had, inevitably, caused the futures price for West Texas Intermediate (WTI) to plummet from USD 18 a barrel to around –USD 37 a barrel [31]. Real-time updates on corporate earnings, geopolitical events, and market trends can trigger immediate reactions and prompt investors to make rapid decisions in reassessing their positions. The social media platform Twitter, in particular, has become a dynamic space for financial discussions and for disseminating market-related information. Tweets from influential market analysts, financial experts, and even company executives can rapidly circulate and thereby influence investor perceptions and drive fluctuations in stock prices. The accessibility and speed of information on both news outlets and Twitter have made it imperative for investors to stay vigilant, as their decisions are now increasingly shaped by the instantaneous flow of news and opinions.

Evidently, both the news and tweets can potentially impact small traders' trading decisions. As such, this study aims to investigate whether or not these parameters can affect price prediction performance. More specifically, we aim to answer the following research questions:

- Can considering the news and tweets improve the stock-price prediction accuracy?
- Does the impact of the news and tweets on the price prediction differ under normal and market-panic conditions?

Given these research questions, the primary aim of the research is to investigate whether considering news and tweets can enhance stock-price prediction accuracy. This impact is crucial for investors, financial institutions, and algorithmic trading firms, as more accurate predictions can lead to better investment decisions, reduced risks, and improved profitability. Furthermore, by examining the impact of news and tweets on stock-price prediction, the research seeks to deepen understanding of the dynamic interactions between market information, investor sentiment, and stock-price movements. This understanding can help develop more accurate prediction models that capture the complex interaction of factors influencing financial markets. Following the understanding of these impacts, this research aims to investigate the stability of these impacts under panic conditions. In other words, it aims to consider whether and how these impacts may alter during periods of market distress. To this end, we chose the COVID-19 pandemic as the case for our analysis, given its status as one of the most significant instances of market panic in recent

history [32–34]. Similarly, considering the widespread utilization of LSTM and MLP models in stock-price prediction (e.g., [12,19,21,35–45]), we opted to evaluate these two methods specifically in the context of incorporating tweets and news count as predictive variables, particularly during market distress.

The remainder of this paper is organized as follows. In Section 2, we review the relevant literature to identify extant knowledge gaps and highlight the contributions of our current study. Next, in Section 3, we define the problem under investigation and outline the proposed model for price prediction. Then, in Section 4, we develop the solution algorithms. Moreover, in Section 5, we analyze the results to address the research questions. Lastly, in Section 6, we offer concluding remarks and insights and suggest future avenues for research.

## 2. Literature Review

A stock market is a place for publicly listed companies to trade stocks and other financial instruments; likewise, the price of shares is termed the stock price [46]. Initial studies (e.g., [47]) held a view of the stock market as stochastic and, hence, non-predictable. However, later studies (e.g., [22,35,48,49]) argued that the stock market may be predictable to an extent when it is examined from a behavioral economics and socioeconomic theory of finance point of view. Therefore, many research studies have developed different models to predict stock prices and parameters in the stock market. These various approaches and techniques can be categorized into three main categories in terms of strategy: (i) technical analysis, (ii) fundamental analysis, and (iii) sentiment-based analysis [20,22,30]. Technical analysis is the most popular approach. It defines a couple of indicators, such as open price, close price, low price, high price, and trading volume, and applies mathematical and statistical techniques to structured data to predict the stock price based on trends in the past and present stock prices [22,48,50–52]. On the contrary, the fundamental analysis is concerned with the company that underlies the stock itself instead of the actual stock [53–55]. The data that are used by the fundamental analyst are usually unstructured and, thus, pose some challenges. However, this type of data has occasionally been shown to be a good predictor of stock-price movement [56]. Moreover, this approach is utilized by financial analysts daily, as it incorporates various factors, such as economic forecasts, the efficiency of management, business opportunities, and financial statements [57]. Fundamental analysis can, in essence, be defined as a method of finding a stock's inherent value based on financial analysis. Lastly, sentiment-based analysis is based on linguistic feature extraction (e.g., [58–60]). This approach has not been as popular, given the difficulty of developing reliable and efficient sentiment analysis tools. This is mainly due to the design complexity and relevant source selection (which is of utmost importance).

Given the popularity of the technical analysis approach, many studies have employed different statistical and mathematical techniques to predict stock prices. These techniques can be generally categorized into three groups: (i) statistical model (STAT), (ii) evolutionary algorithm (EA), and (iii) machine learning (ML), including neural networks (NN). To utilize these techniques, decision-makers should decide on the type of target variable the model aims to predict. These variables can either be the stock price, stock direction, index price, or index direction. The stock price refers to an individual numerical value that the model believes the stock will be priced at soon. The stock direction refers to the direction (i.e., up or down) in which the stock price will move. Furthermore, the index could be another target variable. Unlike the case of stocks, where the data pertains to that of an individual stock, an index measures a section of the stock market by combining price data from multiple stocks. A few well-known indices are the S&P 500 and the Dow Jones Industrial Average.

Exploring the existing literature reveals that diverse techniques have been used for stock-market prediction. Table 1 summarizes studies that have used various techniques for stock-market prediction. To compile this table, we utilized two prominent research databases, namely the Web of Science (WOS) and Scopus, which are known for their comprehensive coverage across various disciplines. Additionally, we supplemented our search by exploring the reference lists of selected studies not initially found in these databases. Our survey focused primarily on papers published after 2000, particularly

those published within the last decade. Subsequently, we meticulously reviewed the available literature to identify studies specifically addressing stock-price prediction through technical or sentimental analysis techniques. Each identified article underwent a thorough investigation, enabling us to categorize them based on the type of analysis and techniques employed for stock-price prediction. As shown in Table 1, individual stock outputs are significantly more popular than index-based computations. Likewise, ML and NN are the most popular techniques that have been used in recent years. Hence, given the popularity of these techniques, we will focus on reviewing research studies using ML and NN techniques in order to position our work in the extant literature.

**Table 1.** A summary of the related literature [61–130].

#	Reference	Appr.	Technique	#	Reference	Appr.	Technique	#	Reference	Appr.	Technique
1	Armano et al. [8]	TECH	NN and EA	37	Leigh et al. [62]	TECH	NN and EA	73	Nie and Jin [98]	TECH	ML (SVM)
2	Wang et al. [9]	TECH	NN and STAT	38	Kim [63]	TECH	ML (SVM)	74	Chen and Pan [99]	TECH	ML (SVM) and EA
3	Kao et al. [10]	TECH	ML (SVM)	39	Wang [64]	TECH	STAT (FUZZY)	75	Lahmiri [100]	TECH	ML (SVM) and EA
4	Liu and Lv [11]	TECH	NN and EA	40	Chen and Leung [65]	TECH	NN	76	Qiu et al. [101]	T/F	NN
5	Guo et al. [12]	TECH	STAT	41	Pai and Lin [66]	TECH	STAT (ARIMA)	77	Qiu and Song [102]	TECH	NN
6	Mahmud and Meesad [13]	TECH	Fuzzy	42	Enke and Thawornwong [67]	FUND	NN	78	An and Chan [103]	TECH	EA
7	Wang et al. [14]	TECH	EA	43	Kim et al. [68]	TECH	Genetic Alg	79	Shynkevich et al. [104]	TECH	ML (KNN)
8	Zhang et al. [15]	TECH	ML (RF)	44	Schumaker and Chen [69]	SENT	ML (SVM)	80	Ouahilal et al. [105]	TECH	ML (SVM)
9	Zhang et al. [15]	TECH	NN	45	Huang and Tsai [70]	TECH	ML (SVM)	81	Rout et al. [106]	TECH	ML and EA
10	Bisoi et al. [16]	TECH	ML(SVM) and EA	46	Kara et al. [71]	TECH	ML (SVM)	82	Castelli et al. [107]	TECH	Genetic Alg.
11	Ding and Qin [17]	TECH	Deep RNN	47	Groth and Muntermann [72]	SENT	STAT	83	Tao et al. [108]	TECH	STAT
12	Qiu et al. [18]	TECH	LSTM	48	Schumaker et al. [73]	SENT	STAT	84	Chong et al. [109]	TECH	ML (ARIMA)
13	Jin et al. [19]	SENT	NN (LSTM)	49	Yolcu et al. [74]	TECH	NN	85	Weng et al. [110]	TECH	ML (RF)
14	Vijh et al. [20]	TECH	ML-NN	50	Hagenau et al. [75]	SENT	STAT	86	Zhugue et al. [111]	SENT	NN
15	Lu et al. [21]	TECH	CNN-LSTM	51	Umoh and Udosen [76]	TECH	STAT (FUZZY)	87	Kraus and Feuerriegel [112]	SENT	NN
16	Pang et al. [35]	TECH	LSTM	52	Nayak et al. [77]	TECH	NN and EA	88	Jeon et al. [113]	TECH	NN and STAT
17	Bommareddy et al. [36]	TECH	STAT	53	Tsai and Hsiao [78]	FUND	NN and EA	89	Agustini et al. [114]	TECH	STAT
18	Hiransha et al. [37]	TECH	NN	54	Li et al. [79]	T/S	NN and STAT	90	Matsubara et al. [115]	SENT	NN
19	Nguyen et al. [39]	TECH	SVM and LSTM	55	Sun et al. [80]	TECH	NN	91	Ebadati and Mortazavi [116]	TECH	NN and EA
20	Kumar et al. [40]	TECH	LSTM	56	Dash et al. [81]	TECH	EA	92	Kooli et al. [117]	T/F	NN
21	Mittal et al. [41]	TECH	LSTM-SVM	57	Adebiyi et al. [82]	TECH	NN and STAT	93	Lahmiri [118]	TECH	ML (SVM)
22	Perdana and Rokhim [42]	TECH	LSTM	58	Lee et al. [83]	SENT	ML (RF)	94	Zhou et al. [119]	TECH	NN
23	Shirata and Harada [43]	TECH	LSTM	59	Junqué de Fortuny et al. [84]	SENT	ML (SVM)	95	Shah et al. [120]	TECH	NN and EA
24				60	Bisoi and Dash [85]	T/F	ML (SVM)	96	Gocken et al. [121]	TECH	NN and EA
25	Liu et al. [44]	T/S	LSTM-CNN	61	Mondal et al. [86]	TECH	STAT (ARIMA)	97	Vantstone et al. [122]	SENT	NN
26	Ammer and Aldhyani [45]	TECH	LSTM	62	Geva and Zahavi [87]	T/S	NN	98	Zheng et al. [123]	TECH	ANN-SVR
27	Yu and Yan [49]	TECH	NN	63	Jiang et al. [88]	SENT	STAT	99	Mehtab et al. [126]	TECH	LSTM
28	Cheng and Yang [50]	TECH	STAT (FUZZY)	64	Hafezi et al. [89]	TECH	NN and EA	100	Chen et al. [124]	TECH	ML (VAR)
29	Rundo et al. [52]	TECH	NN	65	Ballings et al. [90]	TECH	ML (RF)	101	Jiang et al. [125]	TECH	ML (DT, RF)
30	Khairi et al. [53]	T/F/S	ML and NN	66	Nguyen et al. [91]	SENT	ML (SVM)	102	Antad et al. [126]	TECH	ML (LR)
31	Ghorbani and Chong [54]	TECH	STAT (PCA)	67	Wang et al. [92]	FUND	STAT	103	Khan et al. [127]	SENT	Hybrid NNs
32	Zhong and Hitchcock [55]	SENT	LSTM	68	Sun et al. [93]	TECH	STAT (FUZZY)	104	Shaban et al. [128]	SENT	Hybrid
33	Gupta and Chen [58]	SENT	ML (TF-IDF)	69	Gocken et al. [94]	TECH	NN and EA	105	Belcastro et al. [129]	T/S	LSTM
34	Jing et al. [59]	SENT	CNN-LSTM	70	Dash and Dash [95]	TECH	NN and EA	106	Al-Nefaie et al. [130]	TECH	MLP and GRU
35	Wu et al. [60]	SENT	CNN-LSTM	71	Zhou et al. [96]	TECH	NN	107	Current Study	T/S	LSTM and MLP
36	Wang [61]	TECH	STAT (FUZZY)	72	Shynkevich et al. [97]	SENT	ML (KNN)				

### 2.1. ML Techniques in Stock-Price Prediction

Several ML techniques have been employed in stock-price prediction. Among these techniques, the K nearest neighbor (KNN), random forests (RF), support vector machines (SVM), regression, support vector regression (SVR), and ARIMA are the most popular. The KNN algorithm is one of the initial ML algorithms for stock-price prediction and is a supervised algorithm used exclusively for classification and regression. The primary operation involves identifying the closest neighbors to the queried data point. If the task is classification, the most-occurring neighbor value is returned and used as the output value. However, if regression is the objective, then the average of all neighbor values is returned and used as the output value. Chen and Hao [131] applied the regression format of this algorithm for stock-price prediction. Nevertheless, its application in stock-price prediction has decreased in recent years due to its simplicity, as it results in poor performance when compared to newer ML approaches. For instance, Shynkevich et al. [97] utilized the KNN and SVM algorithms for stock-price prediction and compared their performance to show the superiority of the SVM algorithm. Shynkevich et al. [104] also used ML algorithms to investigate the impact of the forecasting window length on stock-price prediction. They concluded that KNN-SVM results in appropriate prediction accuracy; however, the KNN leads to poor performance.

Regression is another supervised ML technique used for stock-price prediction. For instance, Jiang et al. [88] incorporated this algorithm to predict stock-price movements. To this end, they divided users into stakeholder groups and then analyzed how stakeholder group postings correlated with events in the company. They then used this information to predict movements in stock prices. Furthermore, Gupta and Chen [58] investigated the sentiments extracted from a vast repository of tweets sourced from StockTwits. They employed three distinct machine-learning algorithms (i.e., Naïve Bayes, SVM, and logistic regression) and additionally explored five different featurization techniques (i.e., bag of words, bigram, trigram, term frequency-inverse document frequency (TF-IDF), and latent semantic analysis (LSA)) in an attempt to comprehensively understand the nuanced relationship between sentiment and stock-market dynamics. The work of Zheng et al. [123] also utilized ML for stock-price prediction. However, they introduced the bat algorithm to optimize the three free parameters of the SVR model to create the BA-SVR hybrid model. This model was then employed to forecast the closing prices of 18 stock indexes in the Chinese stock market. The empirical results demonstrated that the BA-SVR model surpassed both the polynomial kernel SVR model and the sigmoid kernel SVR model, particularly when compared to the latter models without the optimized initial parameters.

Moreover, random forests (RF) are another ML algorithm for stock-price prediction (e.g., [15,83,110,132]). This algorithm aggregates the power of a large number of individual decision-tree algorithms to improve the prediction performance [133]. More specifically, it is based on the simple principle that many relatively simple models that exhibit low correlation will outperform any of the individual models when they operate as a joint group [134]. Lee et al., [83] examined whether textual data improves stock-price prediction when RFs are employed in training their models. The authors' proposed RF model consisted of 2000 trees, and all the models tested were successfully trained. Their results demonstrated that the incorporation of textual data improved next-day price-movement prediction by 10%. Similarly, Weng et al. [110] examined the effectiveness of incorporating both textual data and technical data in stock-price prediction. Applying decision trees was one of their tested methods alongside SVM and NN. The authors also highlighted that incorporating textual data can improve prediction results. Patel et al. [132] compared RF to NN and SVM and concluded that RF outperforms the others. However, it is worth noting that the applied NN was very simple, with only one hidden layer containing two neurons. Later, Zhang et al. [15] utilized RF as a critical part of training in a proprietary stock prediction model. They concluded that the relatively effective performance of their prediction model (in terms of accuracy and returns) is due to the incorporation of RF as one of the integrated models used as a learning method.

Support vector machines (SVM) is another promising model in the domain of ML. This type of model separates data into distinct classes via a decision boundary and then seeks to maximize the margin. Due to the nonlinearity of stock-price data, SVM is an appropriate technique for prediction, as it can project the data points into a higher dimensional space by performing a function on the data that makes the classes linearly separable. As a result, many studies (e.g., [100,135]) have utilized SVM for stock-price prediction. Additionally, several studies have successfully incorporated textual data into their SVMs for stock prediction [75,91]. Nevertheless, the SVM algorithm is computationally time-consuming. Thus, some studies (e.g., [16]) utilized a kernel function to improve the efficiency of mapping data into a higher dimensional space. Similar to SVM, support vector regression (SVR) employs the same components and techniques; however, rather than classifying the mathematical operations, these operations are tasked with regressing the data points. Li et al. [79] utilized a multiple-kernel SVR to test whether or not the inclusion of news articles alongside technical indicators improved the predictive power of the SVR model. The results revealed that the multiple-kernel SVR outperformed a normal SVR model. The work of Schumaker et al. [73] employed the AZfintext qualitative stock-price predictor and regressed stock quotes and financial news article data as inputs into an SVR algorithm for stock-price prediction. The authors examined whether incorporating sentiment-based analysis into the AZfintext system would improve the stock direction prediction accuracy. The results showed that incorporating sentiment analysis into the AZfintext system did not improve the overall prediction accuracy.

Another widely employed method for stock-price prediction involves utilizing the auto-regressive integrated moving average (ARIMA) model, which is also commonly known as the Box–Jenkins model in finance. ARIMA is specifically designed for training in forecasting time-series data [136]. Functioning as a generalized random walk model, ARIMA is finely tuned to eliminate residual autocorrelation, which is a statistical measure of the correlation between variables based on past values. As a generalized exponential smoothing model, ARIMA can incorporate long-term trends and seasonality into its predictions [137]. In recent stock-price prediction research, ARIMA has frequently been integrated either into other ML algorithms or used as a benchmark for comparison. An early example is the work of Pai and Lin [66], who developed and tested a hybrid ARIMA and SVM model. They recognized ARIMA's declining popularity and demonstrated its utility in enhancing ML models. The results revealed that the proposed hybrid ARIMA-SVM model outperformed both the standalone ARIMA and SVM models. A subsequent study conducted by Adebisi et al. [82] compared ARIMA with a three-layer NN and found that the NN consistently outperformed ARIMA in most cases. The graphical representation of their ARIMA predictions indicated a linear pattern, thereby emphasizing its limitation in providing value-based forecasting. Similarly, Chong et al. [109] discovered that NNs significantly outperformed the benchmark autoregressive model in stock-price prediction.

## 2.2. NN Techniques in Stock-Price Prediction

Neural networks (NN) are an ML technique that has recently gained popularity in the domain of stock-price prediction. The earliest application of NNs in predicting stock-market trends dates back to the early 1990s [12,15]. More recently, NNs have become widely adopted, given their proven practical and powerful attributes in stock-price prediction [18–23,35,37,48,49,55,59,60,117,138].

Sun et al. [80] utilized data on stock movements to analyze whether trading behavior could be mapped and used to predict stock prices. For each individual stock, they first analyzed stock trading activities and mapped a network. Then, they classified the trading relationships and grouped them into appropriate categories. Next, they employed Granger causality analysis to prove that the stock prices were indeed correlated with the different trading categories. Moreover, they used a simple three-layer feed-forward NN to test the trading predictability power. The NN incorporated technical indicators as well as trading indicators. The results revealed that the NN performed well overall. Lastly, it is worth

emphasizing that this positive result can be considered relatively intuitive, as it is generally well-known that the activities of one group of traders influence another.

Furthermore, Geva and Zahavi [87] investigated whether market data, simple news item counts, business events, and sentiment scores could improve various ML algorithms in stock-price prediction. The authors considered NNs, decision trees, and basic regression. The results demonstrated that among the algorithms tested, only the NN could fully exploit the more intricate nature of the proposed sentiment/news inputs. The other models could not take advantage of these inputs, given the complicated relationship between price and sentiment/news indicators. Zhuge et al. [111] utilized Shanghai Composite Index data and emotional data. Emotional data, in this case, involved sentiment analysis from the news and microblogs that were related to a specific company. The authors demonstrated that 15 input variables, comprised of sentiment and technical indicators, could successfully predict a Chinese company's stock opening prices.

Kooli et al. [117] proposed a simple NN to examine whether the inclusion of accounting variables (generated from the release of accounting disclosures) improved the prediction accuracy of the NNs. The results showed that combining 48,204 daily stock closing prices of 39 companies with the respective accounting disclosure variables improved the NNs' prediction quality. However, this level of improvement drastically dropped when the NN predicted prices in 2011, a time of civil unrest in Tunisia. This extreme example is noteworthy, as it portrays how an observed variable (i.e., one that was able to consistently improve the model accuracy) could lose its impact when emotional events occurred.

Vantstone et al. [122] investigated if the prediction of the price of 20 Australian stocks by a neural network autoregressive (NNAR) model could be improved with the inclusion of inputs in the form of counts of both news articles and tweets. The sentiment-based indicators used in this study were generated by Bloomberg. These types of sentiment-based indicators are increasingly becoming available. Additionally, due to the overall improvement in data-mining techniques, these indicators should theoretically be more reliable than ever. Their study found that the NNAR that incorporated the Bloomberg-generated news and Twitter-based sentiment indicators had a higher quality of stock-price predictions. Due to the indicators being created and readily available by Bloomberg, no text/data-mining models had to be utilized by the authors. The resulting ease of incorporating news and Twitter indicators into NNs was equal to any other technical indicator.

As shown in Table 1, LSTM stands out as one of the most popular NN techniques in the price prediction literature. This technique was introduced by Olah [139] in 2015 and has since been utilized in numerous studies. For example, Jin et al. [19] introduced an LSTM-based model that incorporated sentiment analysis and employed empirical modal decomposition to break down stock-price sequences. The authors' approach enhanced prediction accuracy by leveraging LSTM's capacity to analyze relationships among time-series data through its memory function. Furthermore, Lu et al. [21] introduced the CNN-BiLSTM-AM method, which combined convolutional neural networks, bidirectional long short-term memory, and an attention mechanism. The authors' model aimed to predict the following-day stock closing prices by extracting features using CNN, using BiLSTM for prediction, and employing an attention mechanism to capture feature influences at different times. A comparative analysis against seven other methods for predicting stock closing prices on the Shanghai Composite Index revealed the superior performance of the CNN-BiLSTM-AM method in terms of MAE and RMSE.

Vijh et al. [20] employed artificial neural network and random-forest techniques to predict next-day closing prices for stocks across various sectors. By utilizing financial data, such as open, high, low, and closed prices, they created new variables as inputs in the model. The evaluation based on standard indicators RMSE and MAPE highlighted the efficiency of their models for predicting stock closing prices. Wu et al. [60] explored LSTM for stock-price prediction by introducing the S\_I\_LSTM method in incorporating multiple data sources and investor sentiment. The authors' approach leveraged sentiment analysis based on convolutional NNs for calculating investors' sentiment index and combined this

with technical indicators and historical transaction data as features for LSTM prediction. The results indicated that the predicted stock closing prices aligned more closely with the accurate closing prices compared to the traditional LSTM methods. Lastly, Kurani et al. [23] presented a comprehensive study on the use of artificial neural networks (ANN) and support vector machines (SVM) for stock forecasting to provide further insights into the application of machine-learning techniques in the financial domain.

Khan et al. [127] applied algorithms to social media and financial news data to assess their impact on stock-market prediction accuracy over a span of ten subsequent days. They conducted feature selection and minimized spam tweets in the datasets to enhance prediction quality. Furthermore, the study involved experiments to identify stock markets that were challenging to predict and those heavily influenced by social media and financial news. The researchers also compared the outcomes of various algorithms to determine a reliable classifier. The results recommended random forest for stock-trend prediction due to its consistent results in all the cases. Finally, deep-learning techniques were employed, and classifiers were combined to maximize prediction accuracy. The experimental findings revealed the highest prediction accuracies of 80.53% and 75.16% using social media and financial news data, respectively. Shaban et al. [128] introduced a new system based on deep learning to predict the stock price. They combined LSTM and bidirectional gated recurrent unit (BiGRU) to predict the closing price of the stock market. Then, they applied the proposed method to some stocks and predicted their close price 10 and 30 min before the actual time. Liu et al. [44] have another study that considered news in the market price prediction. They developed a model based on TrellisNet and a sentiment attention mechanism (SA-TrellisNet) to predict stock-market prices. They integrated the LSTM and CNN models for sentiment analysis while employing a sentiment attention mechanism to allocate weights and a trellis network for stock prediction. The hybrid model includes three components: sentiment analysis, sentiment attention mechanism, and the prediction model. Finally, they compared the proposed model with general methods to demonstrate its performance.

Recently, the substantial impact of cryptocurrencies on the global financial markets has led to an increasing number of price prediction studies in academic research. Ammer and Aldhyani [45] proposed an LSTM algorithm to forecast the values of four types of cryptocurrencies: AMP, Ethereum, Electro-Optical System, and XRP. To overcome the problem of price-fluctuation prediction, they proposed an LSTM that captures the time dependency aspects of the prices of cryptocurrencies and proposed an embedding network to capture the hidden representations from linked cryptocurrencies. They then employed these two networks in conjunction with each other to predict price. In addition, Belcastro et al. [129] introduced a methodology aimed at optimizing cryptocurrency trading decisions to enhance profit margins. Their approach integrates various statistical, text analytics, and deep-learning methodologies to support a recommendation trading algorithm. Notably, the study leverages supplementary data points, such as the correlation between social media activity and price movements, causal relationships within price trends, and the sentiment analysis of cryptocurrency-related social media, to generate both buy and sell signals. Finally, the researchers conducted numerous experiments utilizing historical data to evaluate the efficacy of the trading algorithm, achieving an average gain of 194% without factoring in transaction fees and 117% when accounting for fees. Lastly, Al-Nefaie et al. [130] employed AI algorithms, including the gated recurrent unit (GRU) and MLP, for forecasting Bitcoin prices. They evaluated their models using various metrics, such as mean square error (MSE), root mean square error (RMSE), Pearson correlation (R), and R-squared (R<sup>2</sup>), to assess performance. Their findings indicated that the MLP method outperformed the GRU approach. Given these studies, the primary contributions of the current study to the extant literature are as follows.

- We are the first to use Bloomberg Twitter and news publication count variables as critical inputs for stock-price prediction within the North American context;



- We use a novel approach in employing Twitter and news publication count variables as inputs into multi-layer perceptron (MLP) and long short-term memory (LSTM) NNs. This novel approach seeks to assess the influence of these variables on various NN architectures, allowing us to concurrently evaluate and contrast the stock-price prediction performance of both models;
- We focus on examining the existence of a potential notable decline in model performance during periods rife with market panic (e.g., the COVID-19 pandemic). Therefore, we seek to provide insights into the robustness of the proposed models under stressful conditions in financial markets.

### 3. Problem Definition and Formulation

Given the significant influence of social media and news on investor behavior in the stock market [140,141], we employ machine-learning (ML) techniques for leveraging such information to predict stock prices in the North American context. Predicting stock prices accurately, especially during periods of heightened market uncertainty, presents a unique challenge for existing prediction models. Notably, stock-price volatility experiences a substantial fluctuation during economic downturns [142] which, in turn, adds complexity to accurate forecasting. Trade volume serves as a more effective predictor of panic-induced volatility than the traditional inputs commonly used in the literature (e.g., [143]). Furthermore, some studies (e.g., [32,144,145]) revealed that investors' reactions to news differ significantly during times of market panic. Therefore, beyond the development of an ML-based prediction model, our study investigates the public panic's impact on the price prediction. Hence, we seek to compare the predictive efficacy of the developed model in periods of public panic with its performance in more stable market conditions. This dual focus aims to provide an understanding of the model's reliability under varying market conditions, which enhances its practical applicability in real-world scenarios.

Both the accuracy and applicability of NN models in predicting stock prices based on technical variables have been explored extensively in recent years (e.g., [21,23,38–41,49,55,59,60,89,124,146]). Building on this foundation, our study leverages NN techniques to incorporate market-sentiment information into the stock-price prediction process. According to the literature, NN techniques fall into five main categories: (i) multiple-layer perceptron (MLP) (e.g., [15,37,94,102,113,116,121,138]), (ii) long short-term memory (LSTM) (e.g., [52,147]), (iii) convolutional NN (CNN) (e.g., [37,119]), (iv) recurrent NN (RNN) (e.g., [17,19]), and (v) heuristic (e.g., [89,96,138]).

Multiple-layer perceptron (MLP) and long short-term memory (LSTM) have emerged as the most commonly applied techniques in stock-price prediction. Consequently, we adopt these two prevalent network types for formulating an advanced stock-price prediction model. To this end, several crucial design decisions on NN configuration should be made. These decisions encompass determining the optimal number of layers, nodes, and training epochs and selecting the appropriate cost functions and optimizers. It is worth noting that, within the extensive body of research utilizing NN as the primary algorithm, there is a notable absence of standardization concerning their design. Moreover, the predominant focus in NN design research lies in hyper-parameter optimization through algorithms rather than in establishing empirical standards. This lack of proven baselines hinders researchers from effectively comparing and advancing various NN architectures [148].

#### 3.1. Multiple-Layer Perceptron (MLP) Network

A multiple-layer perceptron (MLP) is a specific feed-forward NN designed to analyze non-linear data effectively. It comprises multiple layers of perceptrons, each functioning as an algorithm for supervised learning. MLP's capacity to handle chaotic and non-linear data is enhanced by increasing the number of perceptrons. A concise representation of the

input variable processing within an MLP is depicted through the following mathematical expression:

$$Y_n(t) = \Phi \left( \sum_n w_n^o \cdot \Phi \left( \sum_k w_{k_n}^h X_k(t) + b_n^h \right) + b_n^o \right) \quad (1)$$

where  $X_k$  denotes the value of the  $k$ th variable input into the perceptron,  $w_k$  represents the corresponding weight,  $b$  is the bias, and  $\Phi$  is the activation function, including the *sigmoid* and *tanh* functions ([23,49]). Furthermore,  $K_n$  denoted the number of neurons in the  $n$ th layer. These layers fall into three categories: (i) input, (ii) output, and (iii) hidden layers. Finally, “ $o$ ” represents values associated with the output neurons, and “ $h$ ” represents values associated with the hidden neurons. While it has been established that a single hidden layer in a neural network can approximate any univariate function [149], stock-price prediction inherently involves multi-variate complexities. Given the successful approximation of multi-variate functions with just two hidden layers in a simple feed-forward network [150,151], we adopt an NN configuration with two hidden layers for the MLP network.

Although various cost functions have been explored for the MLP network in the literature, we specifically opt for the mean squared error (MSE). The effectiveness of MSE in optimizing NNs has previously been demonstrated [152], thereby showcasing its ability to handle an extensive magnitude of training data [153]. Additionally, we employ the Adam (Adaptive Moment) optimization algorithm. This algorithm was first introduced by Hasan et al. [154] and has been tested by researchers at OpenAI and Google Deepmind [154,155]. Adam has been found to be successful in handling non-stationary data while also being able to handle both sparse and noisy gradients.

The number of hidden nodes is another parameter that should be determined for NN configuration. Unlike epoch iterative testing, a generally agreed-upon formula can guide the establishment of a testing range for the optimal number of hidden nodes. This formula is presented as follows:

$$N_h = \frac{N_s}{\alpha(N_i + N_o)} \quad (2)$$

where  $N_h$  denotes the number of hidden nodes,  $N_i$  denotes the number of input neurons, and  $N_o$  and  $N_s$  denote the number of output neurons and samples in the training dataset, respectively. Finally,  $\alpha$  is an arbitrary scaling factor, usually 2–10. With a training dataset comprising 1000 samples, an input layer of six neurons, and an output layer of one neuron, we have opted for 13 hidden neurons in the MLP. Another critical aspect of the design process involves determining the number of epochs. Since the optimal number of epochs should be established on a case-by-case basis, we have conducted iterative testing to identify the point at which the loss function ceases to decrease. It is worth noting that, despite the iterative approach, we have set a maximum limit of 5000 epochs for the MLP for the sake of simplicity.

### 3.2. Long Short-Term Memory (LSTM) Networks

LSTM networks represent a prominent class of NN extensively employed in stock-price prediction research. LSTMs, categorized as recurrent NNs, distinguish themselves from their feed-forward counterparts by incorporating the previous output as an input for the subsequent timestamp. Unlike feed-forward NNs like MLP, which treat the first and 1000th inputs similarly, recurrent networks process data sequentially. This characteristic makes LSTMs particularly suitable for addressing the inherent linearity of MLPs. Despite the resemblance of basic linear transformations within an LSTM to those in an MLP, the pivotal feature contributing to the widespread adoption of LSTMs is the integration of gates and states, which fundamentally alter the nature of the NN.

To construct the LSTM network, similar to the MLP, we specify two hidden layers with the objective of minimizing the mean squared error (MSE) as the cost function. Additionally, based on the demonstrated effectiveness in prior studies [156,157], we have chosen to combine LSTM and the Adam optimization algorithm for this LSTM NN. Moreover, con-

sidering the network's complexity, we set the number of hidden neurons and the maximum number of epochs to 60 and 1000, respectively, when the training sample size is 1000.

#### 4. Model Construction

The current study aims to examine the specific impact of Twitter and news count variables on stock-price prediction, with a primary emphasis on the North American context. To this end, as shown in Figure 1, we first selected a group of stocks for investigation. A subset of the chosen stocks is collectively known as "FAANG", an acronym representing Facebook Inc. (FB), Apple Inc. (AAPL), Amazon.com Inc. (AMZN), Netflix Inc. (NFLX), and Alphabet Inc. (GOOG). Coined by former Goldman Sachs fund manager Jim Cramer [158], FAANG stocks are particularly significant for North American investors and the overall stock market and are publicly traded on the NASDAQ. As of July 2020, these five companies boasted a combined market capitalization of USD 4.1 trillion, which constituted 16.64% of the total S&P 500 market capitalization. The S&P 500, an index comprised of 500 companies, historically represents 70% to 80% of the total US stock-market capitalization. The substantial contribution of FAANG stocks to the S&P 500 highlights their broader importance in shaping the North American stock market. The movements of FAANG stocks directly influence North American investors' perceptions of the overall market and thereby impact trading decisions.

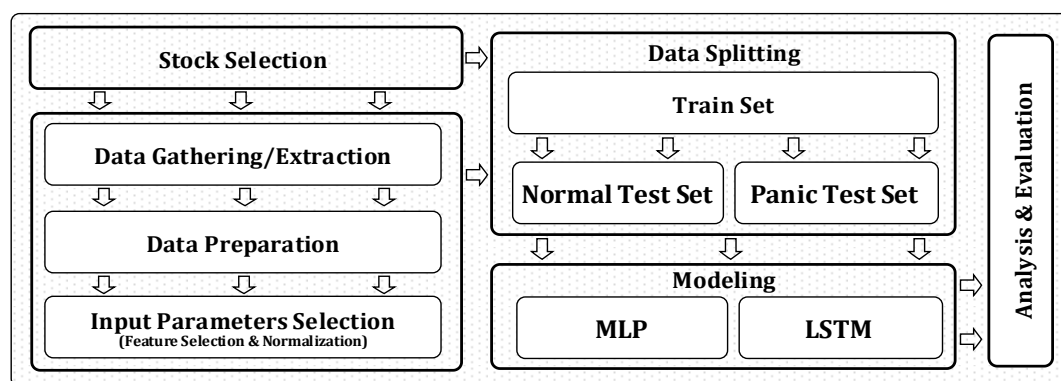


Figure 1. The methodology used for analysis.

The choice of Walmart as a focal point stems from the unique characteristics of the Bloomberg Twitter and news count variables. These variables gauge the frequency of mentions a specific company receives on Twitter or on digital news platforms. Given the recent surge in the popularity of digital and social media, particularly amongst a younger demographic, there is potential variation in the significance of Twitter and News count variables for companies adhering to traditional business models versus those embracing non-traditional models. To investigate this distinction, we opted to conduct a comparative analysis between Walmart and Amazon. Our aim was to discern any disparities in how Twitter and news count variables manifested for a conventional brick-and-mortar retailer like Walmart versus a more contemporary retailer like Amazon. Similarly, the selection of Ford and Tesla was motivated by the desire to contrast a more traditional automobile manufacturer with a purely electric one. Ford adheres to the traditional car dealership model while Tesla's showrooms are commonly situated in malls and have all transactions occurring online [159,160]. Exploring the potential impact of this dichotomy in business models on both the Twitter and news count variables, as well as their effectiveness as inputs in stock-price prediction models, adds an intriguing dimension to our analysis.

##### 4.1. Input Parameters Selection

To establish the input parameters for our analysis, we reviewed the parameters commonly explored in other studies. As depicted in Table 2, the most frequently examined inputs include open price, high price, low price, close price, moving average, and trade

volume. Among these, the four most prevalent ones are open, high, low, and close. Consequently, the moving average (representing the price averaged over a specified number of periods) and trade volume (indicating the number of trades executed in a day) are incorporated into only half of the models. In contrast, the remaining half incorporates the four most common inputs alongside Bloomberg-generated Twitter and news count data. Several reasons justify the exclusion of the least utilized variables. Primarily, this choice facilitates a more direct comparison of the individual impact of these variables on enhancing stock-price prediction. Likewise, since the total number of variables remains constant, any observed performance improvement cannot be attributed to an increase in data volume. Additionally, in the context of NN, adjustments to design decisions, such as the number of neurons, are influenced by changes in the input variable count. Hence, for a methodologically sound comparison, we maintain consistency in the NN’s design, irrespective of the variable set in use. It is worth noting that all variables are sourced from Bloomberg and are formatted in a comma-separated value structure. The dataset we used spanned from January 2015 to May 2020. We chose this time period in order to consider a more comprehensive coverage, that is, to include both normal market conditions (i.e., pre-COVID-19) as well as market-panic conditions (i.e., a few months post-COVID-19 outbreak). The definitions for each variable can be found in Table 3.

**Table 2.** The indicators that have been investigated in the literature.

No	Reference	OP	CP	HP	LP	TV	30MA	BB	TB	MACD
1	Zhang et al. [15]	■	■	■	■	■				
2	Ding and Qin [17]	■	■	■						
	Hiransha et al. [37]		■							
3	Rundo et al. [52]	■	■	■	■					
4	Wu et al. [60]	■	■	■	■	■				
5	Hafezi et al. [89]	■		■	■					
6	Gocken et al. [94]	■	■		■					
7	Zhou et al. [96]		■							
8	Qiu and Song [102]					■	■			
9	Jeon et al. [113]	■		■	■	■				
10	Ebadati and Mortazavi [116]			■						
11	Zhou et al. [119]	■		■	■	■	■	■	■	■
12										
13	Gocken et al. [121]	■	■	■	■		■	■		
14	Moghaddam et al. [138]		■							
15	Nelson et al. [147]	■	■	■	■					
16	Dey et al. [161]	■	■	■	■	■				

OP: open price, CP: close price, HP: high price, LP: low price, TV: trade volume, MA: moving avg, BB: Bollinger band, TB: turnover bias, MACD: MACD index.

**Table 3.** Definitions of input parameters.

Variable Name	Definition
Open Price	The dollar value of the first trade since the market opened
High Price	The highest dollar value trade of the day
Low Price	The lowest dollar value trade of the day
Close Price	The dollar value of the last trade before the market closed
30-Day Moving Average	The average dollar value of one share in the last 30 days
Trade Volume	The total quantity of shares traded all day.
Twitter Count (TC)	Represents the difference between the number of tweets expressing positive sentiment and those expressing negative sentiment towards the parent company over a 24-h period.
News Publication Count (NC)	Represents the total number of news publications mentioning the parent company over a 24-h period.

As mentioned, the data utilized in our study was extracted from Bloomberg, a reputable financial data provider widely used in academic and industry research. To ensure

accuracy and reliability, we accessed Bloomberg's database and retrieved the required information using their data-export functionality. Bloomberg's database is renowned for its accuracy, timeliness, and depth of coverage, making it a preferred choice for researchers and practitioners in the financial industry. Specifically, we employed Bloomberg's Excel API to extract the data directly into an Excel format. This API allowed us to access various financial data, including stocks' open prices, high prices, low prices, and close prices daily from 1 January 2015 to 31 May 2020. It also includes trading volume, the 30-day moving average, the number of tweets with positive sentiment, the number of tweets with negative sentiment, the number of tweets with neutral sentiment, and the news publication count. To capture the sentiments of tweets, we consider the difference between positive and negative sentiment tweets as the number of Twitter counts (TC). However, we do not involve the sentiment in the news publication count (NC) parameter and consider the total number of news publications mentioning the parent company over a 24 h period. The dataset includes 1412 data items for each stock, and the attributes were selected based on their potential significance for analyzing stock-market dynamics and sentiment and their availability within the Bloomberg database. The daily average number of tweets and news publications for the selected companies are Apple (4952 tweets, 2506 news), Amazon (3009 tweets, 757 news), Facebook (3912 tweets, 701 news), Netflix (1556 tweets, 637 news), Google (3930 tweets, 1196 news), Walmart (743 tweets, 375 news), Tesla (2218 tweets, 594 news), and Ford (192 tweets, 271 news). Furthermore, by utilizing the Excel API, we could integrate the data into our analysis workflow, facilitating further processing and analysis. Additional details regarding the collected data can be found in Table A1 in Appendix A.

#### 4.2. Data Splitting, Modeling, and Analysis

Following the collection of daily stock data from Bloomberg, the dataset is divided into two subsets: (i) a technical set (T) and (ii) a technical-plus set incorporating TC and NC (T+). To address skewness, a log transformation is applied to the TC and NC. After this transformation, all TC and NC variables exhibit an acceptable level of skewness. Min-max normalization is then applied to all the remaining variables. Subsequently, the dataset is further categorized into three sets: (i) a training set, (ii) a normal test set, and (iii) a panic test set. The training set encompasses daily variables from January 2015 to 2019, the normal test set comprises variables from January 2019 to November 2019, and the panic test set includes variables from January 2019 to May 2020. For each stock, four price prediction models are constructed. Each model undergoes training and execution of the NN five times and the predictions are averaged to obtain the final model prediction. This approach addresses the stochastic nature of NNs, which can result in slight performance variances. The first two models utilize LSTM, where one model incorporates OP, HP, LP, CP, TV, and 30MA as inputs and the other uses OP, HP, LP, CP, TC, and NC as inputs. The next day's CP is defined as the target variable in these models. The third and fourth models are built using MLP with the respective input sets.

Each defined model undergoes testing on both the normal test set and the panic test set. The accuracy measure for model evaluation is the root mean squared error (RMSE), where lower RMSE values indicate better stock-price predictions. RMSE is selected over other error measures due to its expression in the original unit being measured. This characteristic makes RMSE useful for analyzing the error gap between the expected and predicted values [162]. Moreover, RMSE assigns a higher weight to larger errors compared to other measures, making it particularly suitable for domains where significant errors in accuracy are undesirable [161]. Incorporating the magnitude of error into RMSE is pertinent in stock-price prediction research, as larger errors in prediction can potentially lead to greater losses in buy or sell decisions. Recent stock-price prediction research adopted RMSE as the error measure for comparative analysis of different companies' stock-price predictions [163].

## 5. Results and Discussion

In this section, we investigate the potential enhancement of stock-price prediction by including Twitter and news variables and explore whether their influence varies significantly during periods of market panic. In this regard, we employ the constructed models to forecast the selected stock prices and incorporate the identified input parameters. To distinguish the distinct impact amongst various parameters that might influence the prediction accuracy, we apply both MLP and LSTM models with a technical set (T) and a technical-plus set (T+) separately. We calculated the root mean squared error (RMSE) as the accuracy index for each model and test-set configuration. The results of the designed experiments are presented in Table A2 of Appendix A.

To evaluate the improvement in stock-price prediction between T+ and T models' performance, we assessed the impact of Twitter and news variables by subtracting the former's RMSE from the latter's. Furthermore, we examined whether this impact differed in a panic market by comparing the results obtained in a normal market with those achieved during a panic market. These results were then subjected to a statistical test for model comparison. Additionally, when exploring the impact of input-data type, test-data type, and NN selection, our analysis emphasized group-level analysis rather than a focus on individual performances. Concentrating on the average relative difference between T+ input data and T input data for the same test set as a group aimed to provide insights independent of specific decisions (e.g., individual stock selection) that may influence stock-price prediction performance. This approach holds the potential for increased practical replicability and utility. Therefore, the impact of T+ variables is analyzed based on the RMSE difference compared to models utilizing exclusively technical variables.

### 5.1. Comparing Predictive Models under Panic and Normal Circumstances

To investigate the models' performance under panic and normal circumstances, we created two separate test sets, the normal set (nor-set) and the panic set (pan-set), for each stock. The reason behind creating two sets for each stock is due to the nature of Bloomberg, which generates TC and NC variables relating to the number of times the public mentions a company. Therefore, interesting insights can be gained by analyzing the performance of all models in times of increased panic in the stock market. It is a general consensus that, in the first half of 2020, North America experienced widespread panic related to the performance of the stock market and the economy due to the impact of the global COVID-19 pandemic outbreak [164]. The average RMSE for models tested on the pan-set was 0.0670, while the average RMSE for the nor-set was 0.0354. There is an apparent decrease in performance when testing models on the nor-set compared to the pan-set. To test the statistical significance of this difference, a Wilcoxon signed-rank test was conducted between the RMSEs of the models tested on these sets. The results showed that this difference in performance is significant, even at a 1% level ( $p$ -value~0.001).

To better understand this discrepancy in performance, an analysis of how the variables of interest differed between the different test sets can be found in Appendix. This analysis provides insights into the changes in mean, standard deviation, and coefficient of variation (CV) of the T+ variable (including TC, NC, and close price) across the overall dataset, training dataset, normal test sets, and panic test set. It is worth noting that the normal test set spans from January 2019 to November 2019 (including 218 observations), whereas the panic test set includes 151 observations, covering December 2019 to May 2020 (this is mainly due to the date of a report issued by the World Health Organization which resulted in panic in the US stock market [34]). Additionally, the training data encompasses 1043 observations, covering the period from January 2015 to December 2018. By splitting our analysis into these three distinct sections, we allow ourselves a starting point from which to theorize what may have caused an average decrease in performance for the tests done on the pan-set versus the nor-set.

Table 4 shows a summary of the analysis of two T+ variables, the TC and the NC, and the price differences between the pan-set and the training data. The companies are

displayed in descending order by mean RMSE for the pan-set. As shown in Table 4, when comparing the bottom 50% to the top 50% of the list, the CV for the TC and NC variables exhibit a 19% and 27% absolute difference, respectively. In addition, the percentage change of mean TC and NC variables shows an 18% and 19% absolute difference, respectively. The difference in variance between the data the models are learning from and the additional period added to the nor-set to make it the pan-set are critical to our analysis. The average percentage change in mean for the price data, as well as the average CV of the price data, exhibits a similar difference between the panic data and training data for all the stocks being predicted. This discrepancy in variable variance between the training and test sets can be the primary cause of the drop in performance when it comes to the pan-set compared to the nor-set.

**Table 4.** T+ variables' mean and coefficient of variation in the panic test set and training data.

		TC	NC	Price
		Train vs. Pan-Set	Train vs. Pan-Set	Train vs. Pan-Set
Google	CV	−7%	−15%	−14%
	Mean	−46%	23%	−49%
Amazon	CV	−19%	−7%	−37%
	Mean	−55%	−17%	114%
Walmart	CV	−27%	−35%	6%
	Mean	−26%	132%	129%
Facebook	CV	−48%	−46%	−16%
	Mean	−10%	−71%	47%
Netflix	CV	22%	38%	−45%
	Mean	−69%	−53%	117%
Apple	CV	−59%	−20%	−16%
	Mean	−71%	0%	105%
Ford	CV	−103%	−2%	12%
	Mean	−14%	34%	−44%
Tesla	CV	−35%	−13%	−10%
	Mean	−57%	13%	52%
CV	Top 50%	−25%	−26%	−15%
	Bottom 50%	−44%	1%	−15%
Mean	Top 50%	−35%	17%	60%
	Bottom 50%	−53%	−2%	57%

### 5.2. Comparing the Impact of T and T+ Variables

To determine whether using T+ variables leads to a significant improvement in price prediction, we analyzed instances where the utilization of T+ variables, as opposed to T variables, resulted in improvements in terms of RMSE. As shown in Table 5, there are 25 instances where using T+ variables led to an RMSE reduction. Among those instances that do not lead to improvement, four are associated with Walmart, two are related to Tesla, and one is concerned with Facebook. Furthermore, among these seven instances, four are associated with the LSTM model, and three are related to the MLP model. It is also worth noting that Tesla had a model that improved by 42% and another that lowered by 89%. In essence, this signifies an ample need for researchers and traders to test across a wide variety of scenarios before deciding to use a variable as an input in a stock-price prediction model.

**Table 5.** The RMSE change when utilizing T vs. T+ variables.

NN	Test Set	Company	T RMSE	T+ RMSE	RMSE ±	RMSE% ±
LSTM	PANIC	Amazon	0.06246	0.06018	0.00228	4.0%
		Apple	0.12860	0.12711	0.00149	1.0%
		Facebook	0.07699	0.07392	0.00306	4.0%
		Ford	0.13641	0.07553	0.06087	45.0%
		Google	0.06271	0.04938	0.01333	21.0%
		Netflix	0.11226	0.09216	0.02010	18.0%
		Tesla	0.17632	0.18758	−0.01126	−6.0%
	Walmart	0.05845	0.07278	−0.01434	−25.0%	
	NORMAL	Amazon	0.04672	0.04210	0.00462	10.0%
		Apple	0.05848	0.05299	0.00549	9.0%
		Facebook	0.04404	0.03668	0.00737	17.0%
		Ford	0.04024	0.03007	0.01017	25.0%
		Google	0.03487	0.02874	0.00613	18.0%
		Netflix	0.04211	0.03706	0.00505	12.0%
Tesla		0.05344	0.10091	−0.04748	−89.0%	
Walmart	0.04021	0.04251	−0.00231	−6.0%		
MLP	PANIC	Amazon	0.01950	0.01928	0.00022	1.0%
		Apple	0.02188	0.02141	0.00047	2.0%
		Facebook	0.02746	0.02407	0.00339	12.0%
		Ford	0.16027	0.03661	0.12366	77.0%
		Google	0.02388	0.02069	0.00320	13.0%
		Netflix	0.02553	0.02399	0.00154	6.0%
		Tesla	0.08409	0.04863	0.03546	42.0%
	Walmart	0.01753	0.01814	−0.00062	−4.0%	
	NORMAL	Amazon	0.02602	0.02426	0.00176	7.0%
		Apple	0.02002	0.01995	0.00007	0.0%
		Facebook	0.02070	0.02193	−0.00123	−6.0%
		Ford	0.04886	0.02047	0.02839	58.0%
		Google	0.02355	0.02147	0.00209	9.0%
		Netflix	0.02954	0.02842	0.00113	4.0%
Tesla		0.03610	0.03531	0.00079	2.0%	
Walmart	0.01278	0.01362	−0.00085	−7.0%		

A summary of these impacts is presented in Table 6. As highlighted in Table 6, the most considerable improvement belongs to Ford, which has a 36% improvement over the stock prediction using T+ variables. Notably, both MLP and LSTM configurations for Ford exhibit superior performance with the pan-test set compared to the normal test set. This ranking will guide the analysis of input variables, test data, and NN selection in future assessments. Notably, no clear trends are apparent concerning the magnitude of mean TC and NC variables and their relationship to RMSE% improvement. Table 7, however, reveals insights when comparing the average TC to the average NC ratios. The top 50% of the list in terms of RMSE% reduction was associated with the data featuring an average ratio of 1.66 TC/NC. In contrast, the bottom 50% of the list exhibited an almost 57% increase in the average TC to average NC ratio, reaching 2.92. A potential interpretation of these findings suggests that neither the TC nor NC alone necessarily fortifies the variables, emphasizing the nuanced interaction of factors in predictive modeling.



**Table 6.** The average RMSE% improvement across all configurations.

Rank	Company	Average Improvement via T+
1	Ford	51%
2	Google	15%
3	Netflix	10%
4	Facebook	7%
5	Amazon	5%
6	Apple	3%
7	Walmart	−10%
8	Tesla	−13%
	Mean	9%

**Table 7.** The mean Twitter count to mean news publication count ratio.

	Company	Avg TC	Avg NC	TC/NC
Top 50%	Ford	195.8240	274.7395	0.7128
	Google	4005.3400	1213.2782	3.3013
	Netflix	1587.4800	647.8841	2.4503
	Facebook	710.2050	3983.5188	0.1783
Bottom 50%	Amazon	3058.1700	769.4627	3.9744
	Apple	5060.5100	2545.8502	1.9877
	Walmart	754.8970	380.7387	1.9827
	Tesla	2254.9500	600.7312	3.7537
	Top 50% Avg	1624.7100	1529.8552	1.6606
	Bottom 50% Avg	2782.1300	1074.1957	2.9246

In general, within ML, a distinctive requirement involves training with a distinct dataset. It is necessary to thoroughly examine input variables across various stages of both the training and testing processes [165]. The objective is to assess the potential impact of the TC and NC variables on enhancing stock-price prediction. To this end, we conducted comparisons between these variables within the training data and across both test sets. The rationale behind this comparative analysis is to ascertain whether the pre-emptive examination of input data (prior to actual testing) could reveal the possibility of improving the accuracy of stock-price prediction by incorporating T+ variables into the model. Table 8 provides an analysis of the change in the CV between the training and the test sets. The CV for the variables exhibits relative stability across both the training and the test data for all models. Although CV is always non-negative, the change percentage can be positive or negative.

Table 9 presents an analysis of the mean T+ variables across the training data, normal test set (nor-set), and panic test set (pan-set). Notably, there is a consistent improvement in RMSE across all eight tests with the inclusion of T+ variables. Among the top 50% of performers, there is an average 25% reduction in the mean TC between the training and test sets, suggesting that, on average, the TC is 25% lower in both test sets compared to the training set. Conversely, the bottom 50% of performers exhibit an almost 60% decrease in the mean TC between the training data and the average of both test sets. The difference in the mean NC between the training data and test data for the top 50% of models indicates a 40% decrease. In contrast, the bottom 50% of companies (based on the RMSE percentage improvement due to the addition of T+ variables) show an average 25% increase in the mean NC between the training and testing phases.

**Table 8.** The variable CV change between the training and test sets.

Rank by RMSE% +/-	Company	% ± in CV of Training TC vs Nor-Set TC	% ± in CV of Training TC vs Pan-Set TC	% ± in CV of Training NC vs Nor-Set NC	% ± in CV of Training NC vs Pan-Set NC
1	Ford	−90%	−93%	−6%	−4%
2	Google	22%	21%	−2%	−5%
3	Netflix	−9%	2%	−14%	9%
4	Facebook	−42%	−41%	−37%	−36%
5	Amazon	−19%	−18%	−14%	−7%
6	Apple	−18%	−31%	0%	−5%
7	Walmart	−31%	−32%	7%	2%
8	Tesla	−39%	−31%	−29%	−29%
Top 50% Avg		−30%	−28%	−15%	−9%
Bottom 50% Avg		−27%	−28%	−9%	−10%

**Table 9.** The mean TC and NC in the training data, the normal test set, and the panic test set.

Company	Mean TC			Mean NC		
	Train	Nor-Test	Pan-Test	Train	Nor-Test	Pan-Test
1 Ford	212	131	149	253	334	336
2 Google	4320	3560	3140	1206	1109	1237
3 Netflix	1823	1077	907	620	956	730
4 Facebook	686	866	780	4758	1914	1741
5 Amazon	3515	1816	1734	727	1035	890
6 Apple	6229	1612	1675	2479	2864	2735
7 Walmart	881	394	388	364	439	429
8 Tesla	2499	1395	1548	488	823	930
Top 50%	1760	1408	1244	1709	1078	1011
Bottom 50%	3281	1304	1337	1015	1290	1246
Top 50%	Train vs. Avg Test	75%		Top 50%	Train vs. Avg Test	61%
Bottom 50%	Train vs. Avg Test	40%		Bottom 50%	Train vs. Avg Test	125%

### 5.3. Comparing the MLP and LSTM Models

To thoroughly analyze the MLP and LSTM models’ performance, we evaluate the mean and variance of the RMSE across various groupings for each model. The corresponding outcomes are presented in Table 10. Table 10 illustrates that the MLP consistently outperformed the LSTM across all subsets, showcasing lower RMSE values. Additionally, the MLP models demonstrated a lower CV, averaging 0.1693. Despite the LSTM being more complex than the MLP, it failed to surpass the MLP in price prediction accuracy. This finding is in line with the literature, such as the work of Hiransha et al. [37], who also found that their LSTM model did not outperform their MLP model for stock-price prediction over a 400-day period. However, when tested over a 10-year period, the LSTM exhibited improved accuracy and outperformed the MLP model—this was mainly attributed to its memory feature. In other words, the memory feature is a key advantage of LSTMs that becomes more beneficial in larger test sets. The intuition behind this phenomenon lies in the increased prediction period, which allows for more opportunities for memory utilization. Consequently, the enhanced complexity of the LSTM proves advantageous in longer prediction periods; however, the same is not true in shorter testing periods.

**Table 10.** The RMSE analysis for T vs. T+ variables and the panic test set vs. the normal test set.

		T Vars	T+ Vars	Nor-Set	Pan-Set	
Mean	MLP	0.073395	0.069357	0.045699	0.097053	
Range		0.141446	0.158834	0.072171	0.138198	
ST.Dev		0.040658	0.040633	0.016259	0.041585	
CV		0.554	0.586	0.356	0.428	
Mean	LSTM	0.037357	0.024891	0.025188	0.037060	
Range		0.147494	0.035011	0.036088	0.142744	
ST.Dev		0.035671	0.008374	0.008748	0.035683	
CV		0.955	0.336	0.347	0.963	
		MLP vs. LSTM				Average
+/-Mean		-0.036038	-0.044466	-0.020511	-0.059993	-0.0403
+/-CV		-0.401	0.249	0.008	-0.534	-0.1693

As discussed regarding leveraging MLP and LSTM techniques, we examined factors influencing stock-price prediction accuracy, particularly non-technical variables such as Twitter and news-related data. In other words, we investigated how incorporating these variables impacts prediction performance in normal and panic market conditions. By calculating RMSE, a widely used metric in predictive modeling, we quantitatively assessed the accuracy of these predictions. By comparing the RMSEs of models utilizing only technical variables (T) with those incorporating additional Twitter and news variables (T+), we analyzed how these supplementary factors contribute to improved prediction accuracy. In addition, we conducted several statistical tests to validate observed differences in the MLP and LSTM models' performance. Despite the LSTM's reputation for handling sequential data and its inherent complexity, the study finds that MLP consistently outperforms LSTM across various subsets. One possible reason for the superiority of MLP is the volume of data. More specifically, the memory feature is a key advantage of LSTMs that becomes more beneficial in larger test sets. One plausible explanation for MLP's superiority lies in the volume of data. Specifically, while the memory feature of LSTMs becomes more advantageous in larger test sets due to increased prediction periods [166,167], it may not include the same benefits in smaller datasets. This result highlights the sophisticated relationship between model complexity and dataset size in predictive modeling.

### 6. Conclusions and Future Research

This study investigated the influence of Twitter count (TC) and news count (NC) variables on stock-price prediction within both normal and market conditions. We incorporate Bloomberg Twitter and news publication count variables into MLP and LSTM neural networks to assess their predictive influence. Additionally, we analyze these effects during the market panic to evaluate their stability. Our methodology integrates MLP and LSTM neural networks with technical variables (T variables), TC, and NC (creating T+ variables) for price prediction. The models are trained on data from January 2015 to December 2018 and tested on normal (January 2019 to November 2019) and panic periods (December 2019 to May 2020). We applied statistical analyses to these results, which revealed a notable enhancement in stock-price prediction accuracy across various model types when these additional variables were incorporated. Furthermore, the comparison between the T and T+ variables indicated that both traders and researchers could derive substantial benefits from including TC and NC variables as inputs in neural network-based stock-price prediction models. This integration not only enhanced prediction accuracy and provided significant value to traders and investors, but it also facilitated the seamless incorporation of public opinion into prediction models. Given the escalating impact of social media on societal perspectives, the inclusion of TC and NC variables allows traders to consider the public's perception of corporations and products in their analyses. This strategic utilization of social media data empowers traders to make more informed decisions, reflecting a nuanced understanding of market sentiment and public opinion.

While the proposed models aimed to analyze the impact of news and tweets on stock-price prediction accuracy, it is important to acknowledge certain limitations for further research. First, this study focused primarily on the COVID-19 pandemic as a period of market distress. The impact identified in this case might not capture the full range of possible market behaviors under different types of panic conditions, crises, financial crashes, or geopolitical events. In addition, this study used LSTM and MLP models, which, while established, may not represent the cutting edge in predictive modeling. More advanced techniques, like hybrid approaches, might provide better performance or additional insights. Furthermore, the study used sentiment analysis tools to quantify investor sentiment from news and tweets. The accuracy of these tools can vary, and errors in sentiment classification could affect the overall findings. Finally, the impact of news and tweets on stock prices might vary across different regions and industry sectors. The study does not explicitly address whether the findings are consistent across various markets and industries or if specific segments primarily drive the results.

There are several avenues to develop future studies in this area. The current configuration of TC and NC variables in T+ entails countable measures and shares similarities with other variables. In contrast to more generalized Twitter and news-based indicators, Bloomberg's generated Twitter and news count variables employ clearly defined and replicable terms linked to objective measures. Moreover, certain indicator providers, such as Yahoo Finance, may yield identical values for TC and NC, which can be integrated into price prediction models. However, further analysis is imperative to ascertain why these variables demonstrate efficacy across the majority of scenarios but exhibit sub-optimal performance in specific instances. Thus, exploring these cases and identifying their underlying sources of impact may be a promising avenue for future research. Another prospective area for future investigation involves considering whether the influence of TC and NC variables correlates with the size of the company. More specifically, future studies are encouraged to explore whether the inclusion of TC and NC variables equally enhances stock-price prediction accuracy for small-, medium-, and large-sized companies.

Furthermore, the observed variations in prediction performance between MLP and LSTM models can be attributed to the test-period length and the extent of data preprocessing. Traders and researchers should be conscious of these factors when selecting stock-price prediction models. To enhance model applicability across different scenarios, it is recommended to standardize data-preparation techniques to ensure the optimal performance of various model types. This standardization should be replicable for each use of a specific model and thereby promote consistency in the analyses related to stock-price prediction. Likewise, it should advance research aimed at refining neural networks. Additionally, exploring alternative neural network architectures beyond MLP and LSTM is advisable, in an attempt to identify simpler models that may outperform stock-price prediction. Moreover, the current study focused solely on the quantity of tweets and news (TC and NC) while neglecting their sentiment and diverse impacts. Future research could cover the varying impact weights of distinct tweets or news items for each stock and involve their dominant sentiment. Such considerations can be promising areas for future research, as they may refine prediction accuracy in the dynamic landscape of stock-market forecasting.

**Author Contributions:** Conceptualization, H.Z. and S.R.; methodology, H.Z., M.N. and S.R.; software, S.R.; validation, M.N. and S.R.; formal analysis, M.N. and S.R.; investigation, H.Z. and S.R.; resources, H.Z. and A.H.; data curation, S.R.; writing—original draft preparation, M.N.; writing—review and editing, H.Z., S.R. and A.H.; visualization, M.N. and S.R.; supervision, H.Z.; project administration, A.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Data can be accessed on Bloomberg Market News website at <https://www.bloomberg.com/>. The dataset we used was daily and spanned from January 2015 to May 2020.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

**Table A1.** Comparing variables in the overall, training, normal test, pan-test, and panic-period test set.

	Ford			Google			Netflix			Facebook			Amazon			Apple			Tesla			Walmart		
	TC	NC	Price	TC	NC	Price	TC	NC	Price	TC	NC	Price	TC	NC	Price	TC	NC	Price	TC	NC	Price	TC	NC	Price
<b>Training Set</b>																								
Mean	212	253	12.3	4320	1206	1061	1823	620	170	686	4758	133	3515	727	947	6229	2479	141	2499	488	268	881	364	78
Variance	159,345	28,773	3.52	12,068,276	378,508	34,340	3,092,910	161,330	9125	497,886	21,219,182	1281	8,044,743	190,822	207,865	50,038,010	1,775,513	1228	5,862,998	203,445	3227	913,234	58,531	134
St Dev	399.18	169.63	1.88	3473.94	615.23	185.31	1758.67	401.66	95.52	705.61	4606.43	35.79	2836.33	436.83	455.92	7073.76	1332.48	35.04	2421.36	451.05	56.81	955.63	241.93	11.58
CV	1.88	0.67	0.15	0.80	0.51	0.17	0.96	0.65	0.56	1.03	0.97	0.27	0.81	0.60	0.48	1.14	0.54	0.25	0.97	0.92	0.21	1.08	0.66	0.15
<b>Normal Set</b>																								
Mean	131	334	9	3560	1109	708	1077	956	330	866	1914	180	1816	1035	1790	1612	2864	202	1395	823	264	394	439	107
Variance	16,797	41,977	0.3966	13,358,463	295,951	3622	894,611	239,512	1255	275,603	1,315,295	238	1,266,840	222,560	11,013	2,371,030	2,337,762	825	643,150	273,027	1873	93,790	104,811	74
St Dev	129.60	204.88	0.63	3654.92	544.01	60.18	945.84	489.40	35.43	524.98	1146.86	15.43	1125.54	471.76	104.94	1539.81	1528.97	28.72	801.97	522.52	43.28	306.25	323.75	8.60
CV	0.99	0.61	0.07	1.03	0.49	0.09	0.88	0.51	0.11	0.61	0.60	0.09	0.62	0.46	0.06	0.96	0.53	0.14	0.57	0.63	0.16	0.78	0.74	0.08
<b>Pan Set</b>																								
Mean	149	336	8	3140	1237	653	907	730	343	780	1741	185	1734	890	1869	1675	2735	232	1548	930	382	388	429	111
Variance	19,947	44,343	2.31	10,181,305	323,122	8425	804,327	288,309	1773	235,199	1,103,842	376	1,171,956	223,752	37,428	1,895,009	1,812,073	2471	1,036,664	343,690	37,976	88,037	85,835	89
St Dev	141.23	210.58	1.52	3190.82	568.44	91.79	896.84	536.94	42.11	484.97	1050.64	19.39	1082.57	473.02	193.46	1376.59	1346.13	49.71	1018.17	586.25	194.87	296.71	292.98	9.43
CV	0.95	0.63	0.19	1.02	0.46	0.14	0.99	0.74	0.12	0.62	0.60	0.10	0.62	0.53	0.10	0.82	0.49	0.21	0.66	0.63	0.51	0.76	0.68	0.08
<b>Overall</b>																								
Mean	196	275	11	4005	1213	956	1587	648	214	710	3984	147	3058	769	1184	5061	2546	164	2255	601	297	755	381	87
Variance	124,300	34,101	6	11,703,649	363,891	59,469	2,666,375	196,237	12,979	432,228	17,787,737	1562	6,884,086	204,387	327,469	41,627,970	1,797,736	3117	4,799,559	275,483	14,664	747,508	66,393	331
St Dev	352.56	184.66	2.45	3421.06	603.23	243.86	1632.90	442.99	113.93	657.44	4217.55	39.52	2623.75	452.09	572.25	6451.97	1340.80	55.83	2190.79	524.86	121.10	864.59	257.67	18.19
CV	1.80	0.67	0.22	0.85	0.50	0.26	1.03	0.68	0.53	0.93	1.06	0.27	0.86	0.59	0.48	1.27	0.53	0.34	0.97	0.87	0.41	1.15	0.68	0.21

**Table A1.** *Cont.*

	Ford			Google			Netflix			Facebook			Amazon			Apple			Tesla			Walmart		
	TC	NC	Price	TC	NC	Price	TC	NC	Price	TC	NC	Price	TC	NC	Price	TC	NC	Price	TC	NC	Price	TC	NC	Price
<b>Pan Set</b>																								
Mean	149	336	8	3140	1237	653	907	730	343	780	1741	185	1734	890	1869	1675	2735	232	1548	930	382	388	429	111
Variance	19947	44343	2.31	10181305	323122	8425	804327	288309	1773	235199	1103842	376	1171956	223752	37428	1895009	1812073	2471	1036664	343690	37976	88037	85835	89
St Dev	141.23	210.58	1.52	3190.82	568.44	91.79	896.84	536.94	42.11	484.97	1050.64	19.39	1082.57	473.02	193.46	1376.59	1346.13	49.71	1018.17	586.25	194.87	296.71	292.98	9.43
CV	0.95	0.63	0.19	1.02	0.46	0.14	0.99	0.74	0.12	0.62	0.60	0.10	0.62	0.53	0.10	0.82	0.49	0.21	0.66	0.63	0.51	0.76	0.68	0.08
<b>Overall</b>																								
Mean	196	275	11	4005	1213	956	1587	648	214	710	3984	147	3058	769	1184	5061	2546	164	2255	601	297	755	381	87
Variance	124300	34101	6	11703649	363891	59469	2666375	196237	12979	432228	17787737	1562	6884086	204387	327469	41627970	1797736	3117	4799559	275483	14664	747508	66393	331
St Dev	352.56	184.66	2.45	3421.06	603.23	243.86	1632.90	442.99	113.93	657.44	4217.55	39.52	2623.75	452.09	572.25	6451.97	1340.80	55.83	2190.79	524.86	121.10	864.59	257.67	18.19
CV	1.80	0.67	0.22	0.85	0.50	0.26	1.03	0.68	0.53	0.93	1.06	0.27	0.86	0.59	0.48	1.27	0.53	0.34	0.97	0.87	0.41	1.15	0.68	0.21

This table provides insights into the changes in mean, standard deviation, and coefficient of variation (CV) of the T+ variable (including TC, NC, and close price) across the overall dataset, training dataset, normal test sets, and panic test set. It is worth noting that the normal test set spans from January 2019 to November 2019, whereas the panic test set covers December 2019 to May 2020 (this is mainly due to the date of a report issued by the World Health Organization, which resulted in a panic in the US stock market [34]). Additionally, the training data encompasses the period from January 2015 to December 2018.

Table A2. Experimental results.

Company	Test Set	Model	RMSE	Company	Test Set	Model	RMSE
Google	PANIC	LSTM (T)	0.06271	Amazon	PANIC	LSTM (T)	0.06246
		LSTM (TS)	0.04938			LSTM (T+)	0.06018
		MLP (T)	0.02388			MLP (T)	0.01950
		MLP (TS)	0.02069			MLP (T+)	0.01928
	NORMAL	LSTM (T)	0.03487		NORMAL	LSTM (T)	0.04672
		LSTM (TS)	0.02874			LSTM (T+)	0.04210
		MLP (T)	0.02355			MLP (T)	0.02602
		MLP (TS)	0.02147			MLP (T+)	0.02426
Facebook	PANIC	LSTM (T)	0.07699	Tesla	PANIC	LSTM (T)	0.17632
		LSTM (TS)	0.07392			LSTM (T+)	0.18758
		MLP (T)	0.02746			MLP (T)	0.08409
		MLP (TS)	0.02407			MLP (T+)	0.04863
	NORMAL	LSTM (T)	0.04404		NORMAL	LSTM (T)	0.05344
		LSTM (TS)	0.03668			LSTM (T+)	0.10091
		MLP (T)	0.02070			MLP (T)	0.03610
		MLP (TS)	0.02193			MLP (T+)	0.03531
Walmart	PANIC	LSTM (T)	0.05845	Apple	PANIC	LSTM (T)	0.12860
		LSTM (TS)	0.07278			LSTM (T+)	0.12711
		MLP (T)	0.01753			MLP (T)	0.02188
		MLP (TS)	0.01814			MLP (T+)	0.02141
	NORMAL	LSTM (T)	0.04021		NORMAL	LSTM (T)	0.05848
		LSTM (TS)	0.04251			LSTM (T+)	0.05299
		MLP (T)	0.01278			MLP (T)	0.02002
		MLP (TS)	0.01362			MLP (T+)	0.01995
Ford	PANIC	LSTM (T)	0.13641	Netflix	PANIC	LSTM (T)	0.11226
		LSTM (TS)	0.07553			LSTM (T+)	0.09216
		MLP (T)	0.16027			MLP (T)	0.02553
		MLP (T+)	0.03661			MLP (T+)	0.02399
	NORMAL	LSTM (T)	0.04024		NORMAL	LSTM (T)	0.04211
		LSTM (T+)	0.03007			LSTM (T+)	0.03706
		MLP (T)	0.04886			MLP (T)	0.02954
		MLP (T+)	0.02047			MLP (T+)	0.02842

## References

- Edwards, J. Global Market Cap Is Heading toward \$100 Trillion and Goldman Sachs Thinks the Only Way Is down. 2017. Available online: <https://www.businessinsider.de/global-market-cap-is-about-to-hit-100-trillion-2017-12?r=UK&IR=T> (accessed on 5 March 2023).
- SIFMA. Research Quarterly: Equities. Available online: <https://www.sifma.org/resources/research/research-quarterly-equities/> (accessed on 5 January 2024).
- NYSE: New York Stock Exchange, New York Stock Exchange. NYSE Total Market Cap. 2018. Available online: <https://www.nyse.com/market-cap> (accessed on 12 April 2023).
- FXCM. New York Stock Exchange (NYSE). 2016. Available online: <https://www.fxcm.com/uk/insights/new-york-stock-exchange-nyse/> (accessed on 13 April 2023).
- Chan, H.L.; Woo, K.Y. Studying the Dynamic Relationships between Residential Property Prices, Stock Prices, and GDP: Lessons from Hong Kong. *J. Hous. Res.* **2013**, *22*, 75–89. [CrossRef]
- Jones, J.U.S. Ownership down among All but Older Higher Income. 2017. Available online: <https://news.gallup.com/poll/211052/stock-ownership-down-among-older-higher-income.aspx> (accessed on 15 April 2023).
- Lusardi, A.; Mitchell, O.S. The Economic Importance of Financial Literacy: Theory and Evidence. *J. Econ. Lit.* **2014**, *52*, 5–44. [CrossRef]
- Armano, G.; Marchesi, M.; Murru, A. A hybrid genetic-neural architecture for stock indexes forecasting. *Inf. Sci.* **2005**, *170*, 3–33. [CrossRef]
- Wang, J.-J.; Wang, J.-Z.; Zhang, Z.-G.; Guo, S.-P. Stock index forecasting based on a hybrid model. *Omega* **2012**, *40*, 758–766. [CrossRef]
- Kao, L.-J.; Chiu, C.-C.; Lu, C.-J.; Chang, C.-H. A hybrid approach by integrating wavelet-based feature extraction with MARS and SVR for stock index forecasting. *Decis. Support Syst.* **2013**, *54*, 1228–1244. [CrossRef]
- Liu, H.; Lv, X.Y. Stock Price Prediction Model Based on IWO Neural Network and its Applications. *Adv. Mater. Res.* **2014**, *989*, 1635–1640. [CrossRef]

12. Guo, Z.; Wang, H.; Yang, J.; Miller, D.J. A Stock Market Forecasting Model Combining Two-Directional Two-Dimensional Principal Component Analysis and Radial Basis Function Neural Network. *PLoS ONE* **2015**, *10*, e0122385. [CrossRef] [PubMed]
13. Mahmud, M.S.; Meesad, P. An innovative recurrent error-based neuro-fuzzy system with momentum for stock price prediction. *Soft Comput.* **2016**, *20*, 4173–4191. [CrossRef]
14. Wang, S.; Wang, L.; Gao, S.; Bai, Z. Stock price prediction based on chaotic hybrid particle swarm optimisation-RBF neural network. *Int. J. Appl. Decis. Sci.* **2017**, *10*, 89. [CrossRef]
15. Zhang, L.; Wang, F.; Xu, B.; Chi, W.; Wang, Q.; Sun, T. Prediction of stock prices based on LM-BP neural network and the estimation of overfitting point by RDCI. *Neural Comput. Appl.* **2018**, *30*, 1425–1444. [CrossRef]
16. Bisoi, R.; Dash, P.K.; Parida, A.K. Hybrid Variational Mode Decomposition and evolutionary robust kernel extreme learning machine for stock price and movement prediction on daily basis. *Appl. Soft Comput.* **2019**, *74*, 652–678. [CrossRef]
17. Ding, G.; Qin, L. Study on the prediction of stock price based on the associated network model of LSTM. *Int. J. Mach. Learn. Cybern.* **2020**, *11*, 1307–1317. [CrossRef]
18. Qiu, J.; Wang, B.; Zhou, C. Forecasting stock prices with long-short term memory neural network based on attention mechanism. *PLoS ONE* **2020**, *15*, e0227222. [CrossRef] [PubMed]
19. Jin, Z.; Yang, Y.; Liu, Y. Stock closing price prediction based on sentiment analysis and LSTM. *Neural Comput. Appl.* **2020**, *32*, 9713–9729. [CrossRef]
20. Vijh, M.; Chandola, D.; Tikkiwal, V.A.; Kumar, A. Stock Closing Price Prediction using Machine Learning Techniques. *Procedia Comput. Sci.* **2020**, *167*, 599–606. [CrossRef]
21. Lu, W.; Li, J.; Wang, J.; Qin, L. A CNN-BiLSTM-AM method for stock price prediction. *Neural Comput. Appl.* **2021**, *33*, 4741–4753. [CrossRef]
22. Hu, Z.; Zhao, Y.; Khushi, M. A Survey of Forex and Stock Price Prediction Using Deep Learning. *Appl. Syst. Innov.* **2021**, *4*, 9. [CrossRef]
23. Kurani, A.; Doshi, P.; Vakharia, A.; Shah, M. A Comprehensive Comparative Study of Artificial Neural Network (ANN) and Support Vector Machines (SVM) on Stock Forecasting. *Ann. Data Sci.* **2023**, *10*, 183–208. [CrossRef]
24. Kim, K. *Electronic and Algorithmic Trading Technology: The Complete Guide*; Academic Press: New York, NY, USA, 2010.
25. Ritter, G. Machine Learning for Trading. *SSRN Electron. J.* **2017**. [CrossRef]
26. Chen, L.; Gao, Q. Application of Deep Reinforcement Learning on Automated Stock Trading. In Proceedings of the 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 18–20 October 2019. [CrossRef]
27. Zhang, Z.; Zohren, S.; Roberts, S.J. Deep Reinforcement Learning for Trading. *J. Financ. Data Sci.* **2020**, *2*, 25–40. [CrossRef]
28. Das, S.; Kadapakkam, P.-R. Machine over Mind? Stock price clustering in the era of algorithmic trading. *N. Am. J. Econ. Financ.* **2018**, *51*, 100831. [CrossRef]
29. Business Wire. Global Algorithmic Trading Market to Surpass US\$ 21,685.53 Million by 2026. 2019. Available online: <https://www.businesswire.com/news/home/20190205005634/en/Global-Algorithmic> (accessed on 10 September 2020).
30. Pricope, T.V. Deep Reinforcement Learning in Quantitative Algorithmic Trading: A Review. *arXiv* **2024**, arXiv:2106.00123v1.
31. Johnston, M. What Happened to Oil Prices in 2020. Investopedia. 2022. Available online: <https://www.investopedia.com/articles/investing/100615/will-oil-prices-go-2017.asp> (accessed on 5 March 2024).
32. Cevik, E.; Altinkeski, B.K.; Cevik, E.I.; Dibooglu, S. Investor sentiments and stock markets during the COVID-19 pandemic. *Financ. Innov.* **2022**, *8*, 69. [CrossRef] [PubMed]
33. Joseph, I.; Obini, N.; Sulaiman, A.; Loko, A. Comparative Model Profiles of COVID-19 Occurrence in Nigeria. *Int. J. Math. Trends Technol.* **2020**, *68*, 297–310. [CrossRef]
34. Baig, A.S.; Butt, H.A.; Haroon, O.; Rizvi, S.A.R. Deaths, panic, lockdowns and US equity markets: The case of COVID-19 pandemic. *Financ. Res. Lett.* **2020**, *38*, 101701. [CrossRef]
35. Pang, X.; Zhou, Y.; Wang, P.; Lin, W.; Chang, V. An innovative neural network approach for stock market prediction. *J. Supercomput.* **2020**, *76*, 2098–2118. [CrossRef]
36. Bommareddy, S.R.; Reddy, K.S.S.; Kaushik, P.; Kumar, V.; Hulipalled, V.R. Predicting the stock price using linear regression. *Int. J. Adv. Res. Comput. Sci.* **2018**, *9*, 81–85.
37. Hiransha, M.; Gopalakrishnan, E.A.; Menon, V.K.; Soman, K.P. NSE Stock Market Prediction Using Deep-Learning Models. *Procedia Comput. Sci.* **2018**, *132*, 1351–1362. [CrossRef]
38. Mehtab, S.; Sen, J.; Dutta, A. Stock Price Prediction Using Machine Learning and LSTM-Based Deep Learning Models. In *Machine Learning and Metaheuristics Algorithms, and Applications*; Springer: Berlin/Heidelberg, Germany, 2021. [CrossRef]
39. Nguyen, H.T.; Tran, T.B.; Bui, P.H.D. An effective way for Taiwanese stock price prediction: Boosting the performance with machine learning techniques. *Concurr. Comput. Pract. Exp.* **2021**, *35*, e6437. [CrossRef]
40. Kumar, A.; Hooda, S.; Gill, R.; Ahlawat, D.; Srivastva, D.; Kumar, R. Stock Price Prediction Using Machine Learning. In Proceedings of the 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India, 28–30 April 2023. [CrossRef]
41. Mittal, A.; Joshi, N.; Savani, V. Stock Price Prediction Using Machine Learning Algorithm with Web Interface (GUI). In *Artificial Intelligence and Communication Technologies*; SCRS: New Delhi, India, 2023; pp. 361–377. [CrossRef]
42. Perdana, I.L.; Rokhim, R. Stock price index prediction using machine learning. *AIP Conf. Proc.* **2023**, *2693*, 020031. [CrossRef]



43. Shirata, R.; Harada, T. A Proposal of a Method to Determine the Appropriate Learning Period in Stock Price Prediction Using Machine Learning. *IEEJ Trans. Electr. Electron. Eng.* **2024**, *19*, 726–732. [[CrossRef](#)]
44. Liu, W.-J.; Ge, Y.-B.; Gu, Y.-C. News-driven stock market index prediction based on trellis network and sentiment attention mechanism. *Expert Syst. Appl.* **2024**, *250*, 123966. [[CrossRef](#)]
45. Ammer, M.A.; Aldhyani, T.H.H. Deep Learning Algorithm to Predict Cryptocurrency Fluctuation Prices: Increasing Investment Awareness. *Electronics* **2022**, *11*, 2349. [[CrossRef](#)]
46. Wanjawa, B.W.; Muchemi, L. ANN model to predict stock prices at stock exchange markets. *arXiv* **2014**, arXiv:1502.06434.
47. Malkiel, B.G. *A Random Walk down Wall Street: Including a Life-Cycle Guide to Personal Investing*; WW Norton & Company: New York, NY, USA, 1999.
48. Moghar, A.; Hamiche, M. Stock Market Prediction Using LSTM Recurrent Neural Network. *Procedia Comput. Sci.* **2020**, *170*, 1168–1173. [[CrossRef](#)]
49. Yu, P.; Yan, X. Stock price prediction based on deep neural networks. *Neural Comput. Appl.* **2020**, *32*, 1609–1628. [[CrossRef](#)]
50. Cheng, C.-H.; Yang, J.-H. Fuzzy time-series model based on rough set rule induction for forecasting stock price. *Neurocomputing* **2018**, *302*, 33–45. [[CrossRef](#)]
51. Nguyen, N. Hidden Markov model for stock trading. *Int. J. Financ. Stud.* **2018**, *6*, 36. [[CrossRef](#)]
52. Rundo, F.; Trenta, F.; Di Stallo, A.L.; Battiato, S. Advanced Markov-Based Machine Learning Framework for Making Adaptive Trading System. *Computation* **2019**, *7*, 4. [[CrossRef](#)]
53. Khairi, T.W.A.; Zaki, R.M.; Mahmood, W.A. Stock Price Prediction using Technical, Fundamental and News based Approach. In Proceedings of the 2019 2nd Scientific Conference of Computer Sciences (SCCS), Baghdad, Iraq, 27–28 March 2019. [[CrossRef](#)]
54. Ghorbani, M.; Chong, E.K.P. Stock price prediction using principal components. *PLoS ONE* **2020**, *15*, e0230124. [[CrossRef](#)]
55. Zhong, S.; Hitchcock, D. S&P 500 Stock Price Prediction Using Technical, Fundamental and Text Data. *Stat. Optim. Inf. Comput.* **2021**, *9*, 769–788. [[CrossRef](#)]
56. Sorto, M.; Aasheim, C.; Wimmer, H. Feeling the stock market: A study in the prediction of financial markets based on news sentiment. In Proceedings of the Southern Association for Information Systems Conference, St. Simons Island, GA, USA, 25 March 2017; p. 19.
57. Wafi, A.S.; Hassan, H.; Mabrouk, A. Fundamental analysis models in financial markets—Review study. *Procedia Econ. Financ.* **2015**, *30*, 939–947. [[CrossRef](#)]
58. Gupta, R.; Chen, M. Sentiment Analysis for Stock Price Prediction. In Proceedings of the 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), Shenzhen, China, 6–8 August 2020. [[CrossRef](#)]
59. Jing, N.; Wu, Z.; Wang, H. A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. *Expert Syst. Appl.* **2021**, *178*, 115019. [[CrossRef](#)]
60. Wu, S.; Liu, Y.; Zou, Z.; Weng, T.-H. S\_I\_LSTM: Stock price prediction based on multiple data sources and sentiment analysis. *Connect. Sci.* **2022**, *34*, 44–62. [[CrossRef](#)]
61. Wang, Y.-F. Predicting stock price using fuzzy grey prediction system. *Expert Syst. Appl.* **2002**, *22*, 33–38. [[CrossRef](#)]
62. Leigh, W.; Purvis, R.; Ragusa, J.M. Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: A case study in romantic decision support. *Decis. Support Syst.* **2002**, *32*, 361–377. [[CrossRef](#)]
63. Kim, K.-J. Financial time series forecasting using support vector machines. *Neurocomputing* **2003**, *55*, 307–319. [[CrossRef](#)]
64. Wang, Y.-F. Mining stock price using fuzzy rough set system. *Expert Syst. Appl.* **2003**, *24*, 13–23. [[CrossRef](#)]
65. Chen, A.-S.; Leung, M.T. Regression neural network for error correction in foreign exchange forecasting and trading. *Comput. Oper. Res.* **2004**, *31*, 1049–1068. [[CrossRef](#)]
66. Pai, P.-F.; Lin, C.-S. A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega* **2005**, *33*, 497–505. [[CrossRef](#)]
67. Enke, D.; Thawornwong, S. The use of data mining and neural networks for forecasting stock market returns. *Expert Syst. Appl.* **2005**, *29*, 927–940. [[CrossRef](#)]
68. Kim, M.-J.; Min, S.-H.; Han, I. An evolutionary approach to the combination of multiple classifiers to predict a stock price index. *Expert Syst. Appl.* **2006**, *31*, 241–247. [[CrossRef](#)]
69. Schumaker, R.P.; Chen, H. Textual analysis of stock market prediction using breaking financial news. *ACM Trans. Inf. Syst.* **2009**, *27*, 1–19. [[CrossRef](#)]
70. Huang, C.-L.; Tsai, C.-Y. A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting. *Expert Syst. Appl.* **2009**, *36*, 1529–1539. [[CrossRef](#)]
71. Kara, Y.; Boyacioglu, M.A.; Baykan, K. Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert Syst. Appl.* **2011**, *38*, 5311–5319. [[CrossRef](#)]
72. Groth, S.S.; Muntermann, J. An intraday market risk management approach based on textual analysis. *Decis. Support Syst.* **2011**, *50*, 680–691. [[CrossRef](#)]
73. Schumaker, R.P.; Zhang, Y.; Huang, C.-N.; Chen, H. Evaluating sentiment in financial news articles. *Decis. Support Syst.* **2012**, *53*, 458–464. [[CrossRef](#)]
74. Yolcu, U.; Egrioglu, E.; Aladag, C.H. A new linear & nonlinear artificial neural network model for time series forecasting. *Decis. Support Syst.* **2013**, *54*, 1340–1347. [[CrossRef](#)]

75. Hagenau, M.; Liebmann, M.; Neumann, D. Automated news reading: Stock price prediction based on financial news using context-capturing features. *Decis. Support Syst.* **2013**, *55*, 685–697. [[CrossRef](#)]
76. Umoh, U.A.; Inyang, U.G. A FuzzFuzzy-Neural Intelligent Trading Model for Stock Price Prediction. *Int. J. Comput. Sci. Issues* **2014**, *12*, 36.
77. Nayak, S.C.; Misra, B.B.; Behera, H.S. Fluctuation prediction of stock market index by adaptive evolutionary higher order neural networks. *Int. J. Swarm Intell.* **2016**, *2*, 229. [[CrossRef](#)]
78. Tsai, C.-F.; Quan, Z.-Y. Stock Prediction by Searching for Similarities in Candlestick Charts. *ACM Trans. Manag. Inf. Syst.* **2014**, *5*, 9.
79. Li, X.; Xie, H.; Chen, L.; Wang, J.; Deng, X. News impact on stock price return via sentiment analysis. *Knowl.-Based Syst.* **2014**, *69*, 14–23. [[CrossRef](#)]
80. Sun, X.-Q.; Shen, H.-W.; Cheng, X.-Q. Trading Network Predicts Stock Price. *Sci. Rep.* **2014**, *4*, 3711. [[CrossRef](#)]
81. Dash, R.; Dash, P.; Bisoi, R. A self adaptive differential harmony search based optimized extreme learning machine for financial time series prediction. *Swarm Evol. Comput.* **2014**, *19*, 25–42. [[CrossRef](#)]
82. Adebisi, A.A.; Adewumi, A.O.; Ayo, C.K. Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction. *J. Appl. Math.* **2014**, *2014*, 614342. [[CrossRef](#)]
83. Lee, H.; Surdeanu, M.; MacCartney, B.; Jurafsky, D. On the Importance of Text Analysis for Stock Price Prediction. In Proceedings of the LREC 2014, Ninth International Conference on Language Resources and Evaluation, Reykjavik, Iceland, 26–31 May 2014; pp. 1170–1175.
84. de Fortuny, E.J.; De Smedt, T.; Martens, D.; Daelemans, W. Evaluating and understanding text-based stock price prediction models. *Inf. Process. Manag.* **2014**, *50*, 426–441. [[CrossRef](#)]
85. Bisoi, R.; Dash, P. A hybrid evolutionary dynamic neural network for stock market trend analysis and prediction using unscented Kalman filter. *Appl. Soft Comput.* **2014**, *19*, 41–56. [[CrossRef](#)]
86. Mondal, P.; Shit, L.; Goswami, S. Study of Effectiveness of Time Series Modeling (Arima) in Forecasting Stock Prices. *Int. J. Comput. Sci. Eng. Appl.* **2014**, *4*, 13–29. [[CrossRef](#)]
87. Geva, T.; Zahavi, J. Empirical evaluation of an automated intraday stock recommendation system incorporating both market data and textual news. *Decis. Support Syst.* **2014**, *57*, 212–223. [[CrossRef](#)]
88. Jiang, S.; Chen, H.; Nunamaker, J.F.; Zimbra, D. Analyzing firm-specific social media and market: A stakeholder-based event analysis framework. *Decis. Support Syst.* **2014**, *67*, 30–39. [[CrossRef](#)]
89. Hafezi, R.; Shahrabi, J.; Hadavandi, E. A bat-neural network multi-agent system (BNNMAS) for stock price prediction: Case study of DAX stock price. *Appl. Soft Comput.* **2015**, *29*, 196–210. [[CrossRef](#)]
90. Ballings, M.; Van den Poel, D.; Hespeels, N.; Gryp, R. Evaluating multiple classifiers for stock price direction prediction. *Expert Syst. Appl.* **2015**, *42*, 7046–7056. [[CrossRef](#)]
91. Nguyen, T.H.; Shirai, K.; Velcin, J. Sentiment analysis on social media for stock movement prediction. *Expert Syst. Appl.* **2015**, *42*, 9603–9611. [[CrossRef](#)]
92. Wang, L.; Wang, Z.; Zhao, S.; Tan, S. Stock market trend prediction using dynamical Bayesian factor graph. *Expert Syst. Appl.* **2015**, *42*, 6267–6275. [[CrossRef](#)]
93. Sun, B.; Guo, H.; Karimi, H.R.; Ge, Y.; Xiong, S. Prediction of stock index futures prices based on fuzzy sets and multivariate fuzzy time series. *Neurocomputing* **2015**, *151*, 1528–1536. [[CrossRef](#)]
94. Göçken, M.; Özçalıcı, M.; Boru, A.; Dosdoğru, A.T. Integrating metaheuristics and Artificial Neural Networks for improved stock price prediction. *Expert Syst. Appl.* **2016**, *44*, 320–331. [[CrossRef](#)]
95. Dash, R.; Dash, P. Efficient stock price prediction using a self evolving recurrent neuro-fuzzy inference system optimized through a modified differential harmony search technique. *Expert Syst. Appl.* **2016**, *52*, 75–90. [[CrossRef](#)]
96. Zhou, T.; Gao, S.; Wang, J.; Chu, C.; Todo, Y.; Tang, Z. Financial time series prediction using a dendritic neuron model. *Knowl. Based Syst.* **2016**, *105*, 214–224. [[CrossRef](#)]
97. Shynkevich, Y.; McGinnity, T.; Coleman, S.A.; Belatreche, A. Forecasting movements of health-care stock prices based on different categories of news articles using multiple kernel learning. *Decis. Support Syst.* **2016**, *85*, 74–83. [[CrossRef](#)]
98. Nie, C.-X.; Jin, X.-B. The Interval Slope Method for Long-Term Forecasting of Stock Price Trends. *Adv. Math. Phys.* **2016**, *2016*, 8045656. [[CrossRef](#)]
99. Chen, R.; Pan, B. Chinese Stock Index Futures Price Fluctuation Analysis and Prediction Based on Complementary Ensemble Empirical Mode Decomposition. *Math. Probl. Eng.* **2016**, *2016*, 3791504. [[CrossRef](#)]
100. Lahmiri, S. Intraday stock price forecasting based on variational mode decomposition. *J. Comput. Sci.* **2016**, *12*, 23–27. [[CrossRef](#)]
101. Qiu, M.; Song, Y.; Akagi, F. Application of artificial neural network for the prediction of stock market returns: The case of the Japanese stock market. *Chaos Solitons Fractals* **2016**, *85*, 1–7. [[CrossRef](#)]
102. Qiu, M.; Song, Y. Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model. *PLoS ONE* **2016**, *11*, e0155133. [[CrossRef](#)]
103. An, Y.; Chan, N.H. Short-Term Stock Price Prediction Based on Limit Order Book Dynamics. *J. Forecast.* **2017**, *36*, 541–556. [[CrossRef](#)]
104. Shynkevich, Y.; McGinnity, T.; Coleman, S.A.; Belatreche, A.; Li, Y. Forecasting price movements using technical indicators: Investigating the impact of varying input window length. *Neurocomputing* **2017**, *264*, 71–88. [[CrossRef](#)]

105. Ouahilal, M.; El Mohajir, M.; Chahhou, M.; El Mohajir, B.E. A novel hybrid model based on Hodrick–Prescott filter and support vector regression algorithm for optimizing stock market price prediction. *J. Big Data* **2017**, *4*, 31. [[CrossRef](#)]
106. Rout, A.K.; Dash, P.; Dash, R.; Bisoi, R. Forecasting financial time series using a low complexity recurrent neural network and evolutionary learning approach. *J. King Saud Univ. Comput. Inf. Sci.* **2017**, *29*, 536–552. [[CrossRef](#)]
107. Castelli, M.; Vanneschi, L.; Trujillo, L.; Popovič, A. Stock index return forecasting: Semantics-based genetic programming with local search optimiser. *Int. J. Bio-Inspired Comput.* **2017**, *10*, 159–171. [[CrossRef](#)]
108. Tao, L.; Hao, Y.; Yijie, H.; Chunfeng, S. K-Line Patterns’ Predictive Power Analysis Using the Methods of Similarity Match and Clustering. *Math. Probl. Eng.* **2017**, *2017*, 3096917. [[CrossRef](#)]
109. Chong, E.; Han, C.; Park, F.C. Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Syst. Appl.* **2017**, *83*, 187–205. [[CrossRef](#)]
110. Weng, B.; Ahmed, M.A.; Megahed, F.M. Stock market one-day ahead movement prediction using disparate data sources. *Expert Syst. Appl.* **2017**, *79*, 153–163. [[CrossRef](#)]
111. Zhuge, Q.; Xu, L.; Zhang, G. LSTM Neural Network with Emotional Analysis for Prediction of Stock Price. *Eng. Lett.* **2017**, *25*, 1–9.
112. Kraus, M.; Feuerriegel, S. Decision support from financial disclosures with deep neural networks and transfer learning. *Decis. Support Syst.* **2017**, *104*, 38–48. [[CrossRef](#)]
113. Jeon, S.; Hong, B.; Chang, V. Pattern graph tracking-based stock price prediction using big data. *Futur. Gener. Comput. Syst.* **2018**, *80*, 171–187. [[CrossRef](#)]
114. Agustini, W.F.; Affianti, I.R.; Putri, E.R. Stock price prediction using geometric Brownian motion. *J. Phys. Conf. Ser.* **2018**, *974*, 012047. [[CrossRef](#)]
115. Matsubara, T.; Akita, R.; Uehara, K. Stock Price Prediction by Deep Neural Generative Model of News Articles. *IEICE Trans. Inf. Syst.* **2018**, *101*, 901–908. [[CrossRef](#)]
116. Ebadati, E.O.M.; Mortazavi, T.M. An efficient hybrid machine learning method for time series stock market forecasting. *Neural Netw. World* **2018**, *28*, 41–55. [[CrossRef](#)]
117. Kooli, C.; Trabelsi, R.; Tlili, F. The impact of accounting disclosure on emerging stock market prediction in an unstable socio-political context. *J. Account. Manag. Inf. Syst.* **2018**, *17*, 313–329. [[CrossRef](#)]
118. Lahmiri, S. Minute-ahead stock price forecasting based on singular spectrum analysis and support vector regression. *Appl. Math. Comput.* **2018**, *320*, 444–451. [[CrossRef](#)]
119. Zhou, X.; Pan, Z.; Hu, G.; Tang, S.; Zhao, C. Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets. *Math. Probl. Eng.* **2018**, *2018*, 1–11. [[CrossRef](#)]
120. Shah, H.; Tairan, N.; Garg, H.; Ghazali, R. A Quick Gbest Guided Artificial Bee Colony Algorithm for Stock Market Prices Prediction. *Symmetry* **2018**, *10*, 292. [[CrossRef](#)]
121. Göçken, M.; Özçalıcı, M.; Boru, A.; Dosdoğru, A.T. Stock price prediction using hybrid soft computing models incorporating parameter tuning and input variable selection. *Neural Comput. Appl.* **2019**, *31*, 577–592. [[CrossRef](#)]
122. Vanstone, B.J.; Gepp, A.; Harris, G. The effect of sentiment on stock price prediction. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*; Springer: Cham, Switzerland, 2019; pp. 551–559.
123. Zheng, J.; Wang, Y.; Li, S.; Chen, H. The Stock Index Prediction Based on SVR Model with Bat Optimization Algorithm. *Algorithms* **2021**, *14*, 299. [[CrossRef](#)]
124. Chen, J.; Wen, Y.; Nanekaran, Y.; Suzaiddola; Chen, W.; Zhang, D. Machine learning techniques for stock price prediction and graphic signal recognition. *Eng. Appl. Artif. Intell.* **2023**, *121*, 106038. [[CrossRef](#)]
125. Jiang, Z.; Liu, J.; Yang, L. Comparison Analysis of Stock Price Prediction Based on Different Machine Learning Methods. In *Proceedings of the 2nd International Academic Conference on Blockchain, Information Technology and Smart Finance (ICBIS 2023)*, Hangzhou, China, 17–19 February 2023; pp. 59–67. [[CrossRef](#)]
126. Antad, S.; Khandelwal, S.; Khandelwal, A.; Khandare, R.; Khandave, P.; Khangar, D.; Khanke, R. Stock Price Prediction Website Using Linear Regression—A Machine Learning Algorithm. *ITM Web Conf.* **2023**, *56*, 05016. [[CrossRef](#)]
127. Khan, W.; Ghazanfar, M.A.; Azam, M.A.; Karami, A.; Alyoubi, K.H.; Alfakeeh, A.S. Stock market prediction using machine learning classifiers and social media, news. *J. Ambient. Intell. Humaniz. Comput.* **2022**, *13*, 3433–3456. [[CrossRef](#)]
128. Shaban, W.M.; Ashraf, E.; Slama, A.E. SMP-DL: A novel stock market prediction approach based on deep learning for effective trend forecasting. *Neural Comput. Appl.* **2023**, *36*, 1849–1873. [[CrossRef](#)]
129. Belcastro, L.; Carbone, D.; Cosentino, C.; Marozzo, F.; Trunfio, P. Enhancing Cryptocurrency Price Forecasting by Integrating Machine Learning with Social Media and Market Data. *Algorithms* **2023**, *16*, 542. [[CrossRef](#)]
130. Al-Nefae, A.H.; Aldhyani, T.H.H. Bitcoin Price Forecasting and Trading: Data Analytics Approaches. *Electronics* **2022**, *11*, 4088. [[CrossRef](#)]
131. Chen, Y.; Hao, Y. A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction. *Expert Syst. Appl.* **2017**, *80*, 340–355. [[CrossRef](#)]
132. Patel, J.; Shah, S.; Thakkar, P.; Kotecha, K. Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. *Expert Syst. Appl.* **2015**, *42*, 259–268. [[CrossRef](#)]
133. Athey, S.; Tibshirani, J.; Wager, S. Generalized random forests. *Ann. Stat.* **2019**, *47*, 1148–1178. [[CrossRef](#)]
134. Scornet, E.; Biau, G.; Vert, J.-P. Consistency of random forests. *Ann. Stat.* **2015**, *43*, 1716–1741. [[CrossRef](#)]

135. Fenghua, W.; Jihong, X.; Zhifang, H.; Xu, G. Stock Price Prediction based on SSA and SVM. *Procedia Comput. Sci.* **2014**, *31*, 625–631. [[CrossRef](#)]
136. Babu, C.N.; Reddy, B.E. A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data. *Appl. Soft Comput.* **2014**, *23*, 27–38. [[CrossRef](#)]
137. Kumar, M.; Anand, M. An application of time series ARIMA forecasting model for predicting sugarcane production in India. *Stud. Bus. Econ.* **2014**, *9*, 81–94.
138. Moghaddam, A.H.; Moghaddam, M.H.; Esfandyari, M. Stock market index prediction using artificial neural network. *J. Econ. Financ. Adm. Sci.* **2016**, *21*, 89–93. [[CrossRef](#)]
139. Olah, C. Understanding lstm networks–colah’s blog. Colah. github. io. 2015. Available online: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/> (accessed on 3 March 2023).
140. Bloomberg. Every Time Trump Tweets about the Stock Market. 2019. Available online: <https://www.bloomberg.com/features/trump-tweets-market> (accessed on 30 September 2023).
141. Investopedia. Can Tweets and Facebook Posts Predict Stock Behavior? 2019. Available online: <https://www.investopedia.com/articles/markets/031814/can-tweets-and-facebook-posts-predict-stock-behavior-and-rt-if-you-think-so.asp> (accessed on 15 September 2023).
142. Asteriou, D.; Pilbeam, K.; Sarantidis, A. The Behaviour of Banking Stocks During the Financial Crisis and Recessions. Evidence from Changes-in-Changes Panel Data Estimations. *Scott. J. Political Econ.* **2019**, *66*, 154–179. [[CrossRef](#)]
143. Erdogan, O.; Bennett, P.; Ozyildirim, C. Recession Prediction Using Yield Curve and Stock Market Liquidity Deviation Measures. *Rev. Financ.* **2014**, *19*, 407–422. [[CrossRef](#)]
144. Kleinnijenhuis, J.; Schultz, F.; Oegema, D.; van Atteveldt, W. Financial news and market panics in the age of high-frequency sentiment trading algorithms. *J. Theory Pract. Crit.* **2013**, *14*, 271–291. [[CrossRef](#)]
145. Angelovska, J. Investors’ behaviour in regard to company earnings announcements during the recession period: Evidence from the Macedonian stock exchange. *Econ. Res. Istraz.* **2017**, *30*, 647–660. [[CrossRef](#)]
146. Rath, S.; Das, N.R.; Pattanayak, B.K. An Analytic Review on Stock Market Price Prediction using Machine Learning and Deep Learning Techniques. *Recent Patents Eng.* **2024**, *18*, 88–104. [[CrossRef](#)]
147. Nelson DM, Q.; Pereira AC, M.; De Oliveira, R.A. Stock market’s price movement prediction with LSTM neural networks. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017. [[CrossRef](#)]
148. Yang, L.; Shami, A. On hyperparameter optimization of machine learning algorithms: Theory and practice. *arXiv* **2020**, arXiv:2007.15745. [[CrossRef](#)]
149. Guliyev, N.J.; Ismailov, V.E. A Single Hidden Layer Feedforward Network with Only One Neuron in the Hidden Layer Can Approximate Any Univariate Function. *Neural Comput.* **2016**, *28*, 1289–1304. [[CrossRef](#)]
150. Stathakis, D. How many hidden layers and nodes? *Int. J. Remote Sens.* **2009**, *30*, 2133–2147. [[CrossRef](#)]
151. Thomas, A.J.; Petridis, M.; Walters, S.D.; Gheyta, S.M.; Morgan, R.E. Two Hidden Layers are Usually Better than One. In *Engineering Applications of Neural Networks Communications in Computer and Information Science*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 279–290. [[CrossRef](#)]
152. Saleem, N.; Khatkhat, M.I. Deep Neural Networks for Speech Enhancement in Complex-Noisy Environments. *Int. J. Interact. Multimed. Artif. Intell.* **2020**, *6*, 84. [[CrossRef](#)]
153. Zhang, N.; Shen, S.-L.; Zhou, A.-N.; Xu, Y.-S. Investigation on Performance of Neural Networks Using Quadratic Relative Error Cost Function. *IEEE Access* **2019**, *7*, 106642–106652. [[CrossRef](#)]
154. Hasan, M.M.; Rahman, M.S.; Bell, A. Deep Reinforcement Learning for Optimization. In *Handbook of Research on Deep Learning Innovations and Trends*; Research Anthology on Artificial Intelligence Applications in Security; IGI Global: Hershey, PA, USA, 2021.
155. Fangasadha, E.F.; Soeroredjo, S.; Gunawan, A.A.S. Literature Review of OpenAI Five’s Mechanisms in Dota 2’s Bot Player. In Proceedings of the 2022 International Seminar on Application for Technology of Information and Communication (iSemantic), Semarang, Indonesia, 17–18 September 2022. [[CrossRef](#)]
156. Jiang, S.; Chen, Y. Hand Gesture Recognition by Using 3DCNN and LSTM with Adam Optimizer. In *Advances in Multimedia Information Processing—PCM 2017 Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 743–753. [[CrossRef](#)]
157. Chang, Z.; Zhang, Y.; Chen, W. Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform. *Energy* **2019**, *187*, 115804. [[CrossRef](#)]
158. Skehin, T.; Crane, M.; Bezbradica, M. Day ahead forecasting of FAANG stocks using ARIMA, LSTM networks and wavelets. In Proceedings of the 26th AIAI Irish Conference on Artificial Intelligence and Cognitive Science, Dublin, Ireland, 6–7 December 2018.
159. Johnson, A.; Reed, A. Tesla in Texas: A Showdown Over Showrooms. *SAM Adv. Manag. J.* **2019**, *84*, 47–56.
160. Cuofano, G. Tesla Distribution Strategy—FourWeekMBA. Tesla-Distribution-Strategy. 2024. Available online: <https://fourweekmba.com/tesla-distribution-strategy/> (accessed on 5 March 2024).
161. Dey, P.; Hossain, E.; Hossain, I.; Chowdhury, M.A.; Alam, S.; Hossain, M.S.; Andersson, K. Comparative Analysis of Recurrent Neural Networks in Stock Price Prediction for Different Frequency Domains. *Algorithms* **2021**, *14*, 251. [[CrossRef](#)]

162. Kumar, R.; Kumar, P.; Kumar, Y. Time Series Data Prediction using IoT and Machine Learning Technique. *Procedia Comput. Sci.* **2020**, *167*, 373–381. [[CrossRef](#)]
163. Thakkar, A.; Chaudhari, K. CREST: Cross-Reference to Exchange-based Stock Trend Prediction using Long Short-Term Memory. *Procedia Comput. Sci.* **2020**, *167*, 616–625. [[CrossRef](#)]
164. Pallavi, D.; Mourani, S.; Prosenjit, P.; Sufia, Z.; Abhijit, M. Use of Non-Linear Autoregressive Model to Forecast the Future Health of Shrimp Farm. *J. Mech. Contin. Math. Sci.* **2021**, *16*, 59–64.
165. Oyekale, J.; Oreko, B. Machine learning for design and optimization of organic Rankine cycle plants: A review of current status and future perspectives. *WIREs Energy Environ.* **2023**, *12*, e474. [[CrossRef](#)]
166. Greff, K.; Srivastava, R.K.; Koutník, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A Search Space Odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 2222–2232. [[CrossRef](#)]
167. Dong, J.; Chen, Y.; Guan, G. Cost Index Predictions for Construction Engineering Based on LSTM Neural Networks. *Adv. Civ. Eng.* **2020**, *2020*, 6518147. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.