



Review

Editor's Choice

Stock Market Prediction Using Machine Learning Techniques: A Decade Survey on Methodologies, Recent Developments, and Future Directions

Nusrat Rouf, Majid Bashir Malik, Tasleem Arif, Sparsh Sharma, Saurabh Singh, Satyabrata Aich and Hee-Cheol Kim

Special Issue

Recent Advanced Technologies and Applications of Smart Computing and Cyber Security


Edited by

Dr. Satyabrata Aich, Prof. Dr. Madhusanka Liyanage and Dr. Bruce Ndibanje



Review

Stock Market Prediction Using Machine Learning Techniques: A Decade Survey on Methodologies, Recent Developments, and Future Directions

Nusrat Rouf ¹, Majid Bashir Malik ², Tasleem Arif ³, Sparsh Sharma ⁴, Saurabh Singh ⁵ , Satyabrata Aich ^{6,*}  and Hee-Cheol Kim ^{7,*}

¹ Research Lab., Department of Computer Sciences, BGSB University, Rajouri 185234, India; nusratrouf@bgsbu.ac.in

² Department of Computer Sciences, BGSB University, Rajouri 185234, India; majidbashirmalik@bgsbu.ac.in

³ Department of Information Technology, BGSB University, Rajouri 185234, India; t.arif@bgsbu.ac.in

⁴ Department of Computer Science and Engineering, NIT Srinagar 190001, India; sparsh.sharma@nitsri.net

⁵ Department of Industrial and System Engineering, Dongguk University, Seoul 04620, Korea; saurabh89@dongguk.edu

⁶ Department of Computer Engineering, Institute of Digital Anti-Aging Healthcare, Inje University, Gimhae 50834, Korea

⁷ College of AI Convergence, Institute of Digital Anti-Aging Healthcare, u-AHRC, Inje University, Gimhae 50834, Korea

* Correspondence: satyabrataaich@gmail.com (S.A.); heeki@inje.ac.kr (H.-C.K.); Tel.: +82-55-320-3720 (H.-C.K.)



Citation: Rouf, N.; Malik, M.B.; Arif, T.; Sharma, S.; Singh, S.; Aich, S.; Kim, H.-C. Stock Market Prediction Using Machine Learning Techniques: A Decade Survey on Methodologies, Recent Developments, and Future Directions. *Electronics* **2021**, *10*, 2717. <https://doi.org/10.3390/electronics10212717>

Academic Editor: George Angelos Papadopoulos

Received: 17 October 2021

Accepted: 5 November 2021

Published: 8 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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/>).

Abstract: With the advent of technological marvels like global digitization, the prediction of the stock market has entered a technologically advanced era, revamping the old model of trading. With the ceaseless increase in market capitalization, stock trading has become a center of investment for many financial investors. Many analysts and researchers have developed tools and techniques that predict stock price movements and help investors in proper decision-making. Advanced trading models enable researchers to predict the market using non-traditional textual data from social platforms. The application of advanced machine learning approaches such as text data analytics and ensemble methods have greatly increased the prediction accuracies. Meanwhile, the analysis and prediction of stock markets continue to be one of the most challenging research areas due to dynamic, erratic, and chaotic data. This study explains the systematics of machine learning-based approaches for stock market prediction based on the deployment of a generic framework. Findings from the last decade (2011–2021) were critically analyzed, having been retrieved from online digital libraries and databases like ACM digital library and Scopus. Furthermore, an extensive comparative analysis was carried out to identify the direction of significance. The study would be helpful for emerging researchers to understand the basics and advancements of this emerging area, and thus carry-on further research in promising directions.

Keywords: generic review; machine learning; stock market prediction; support vector machine

1. Introduction

An advancement in the fundamental aspects of information technology over the last few decades has altered the route of businesses. As one of the most captivating inventions, financial markets have a pointed effect on the nation's economy [1]. The World Bank reported in 2018 that the stock market capitalization worldwide has surpassed 68.654 trillion US\$ [2]. Over the last few years, stock trading has become a center of attention, which can largely be attributed to technological advances. Investors search for tools and techniques that would increase profit and reduce the risk [3]. However, Stock Market Prediction (SMP) is not a simple task due to its non-linear, dynamic, stochastic, and unreliable nature [4]. SMP is an example of time-series forecasting that promptly

examines previous data and estimates future data values. Financial market prediction has been a matter of worry for analysts in different disciplines, including economics, mathematics, material science, and computer science. Driving profits from the trading of stocks is an important factor for the prediction of the stock market [5]. The stock market is dependent on various parameters, such as the market value of a share, the company's performance, government policies, the country's Gross Domestic Product (GDP), the inflation rate, natural calamities, and so on [6]. The Efficient Market Hypothesis explains that stock market costs are significantly determined by new information, and follow a random walk pattern, such that they cannot be predicted solely based on past information [7]. This was a widely accepted theory in the past. With the advent of technology, researchers demonstrated that stock market prices could be predicted to a certain extent. Historical market data, combined with the data extracted from social media platforms, can be analyzed to predict the changes in the economic and business sectors [8]. The performance of stock market prediction systems relies intensely on the quality of the features it is using [9]. While researchers have used some strategies for enhancing the stock-explicit features, more attention needs to be paid to feature extraction and selection mechanisms. Figure 1 presents the outline of this article.

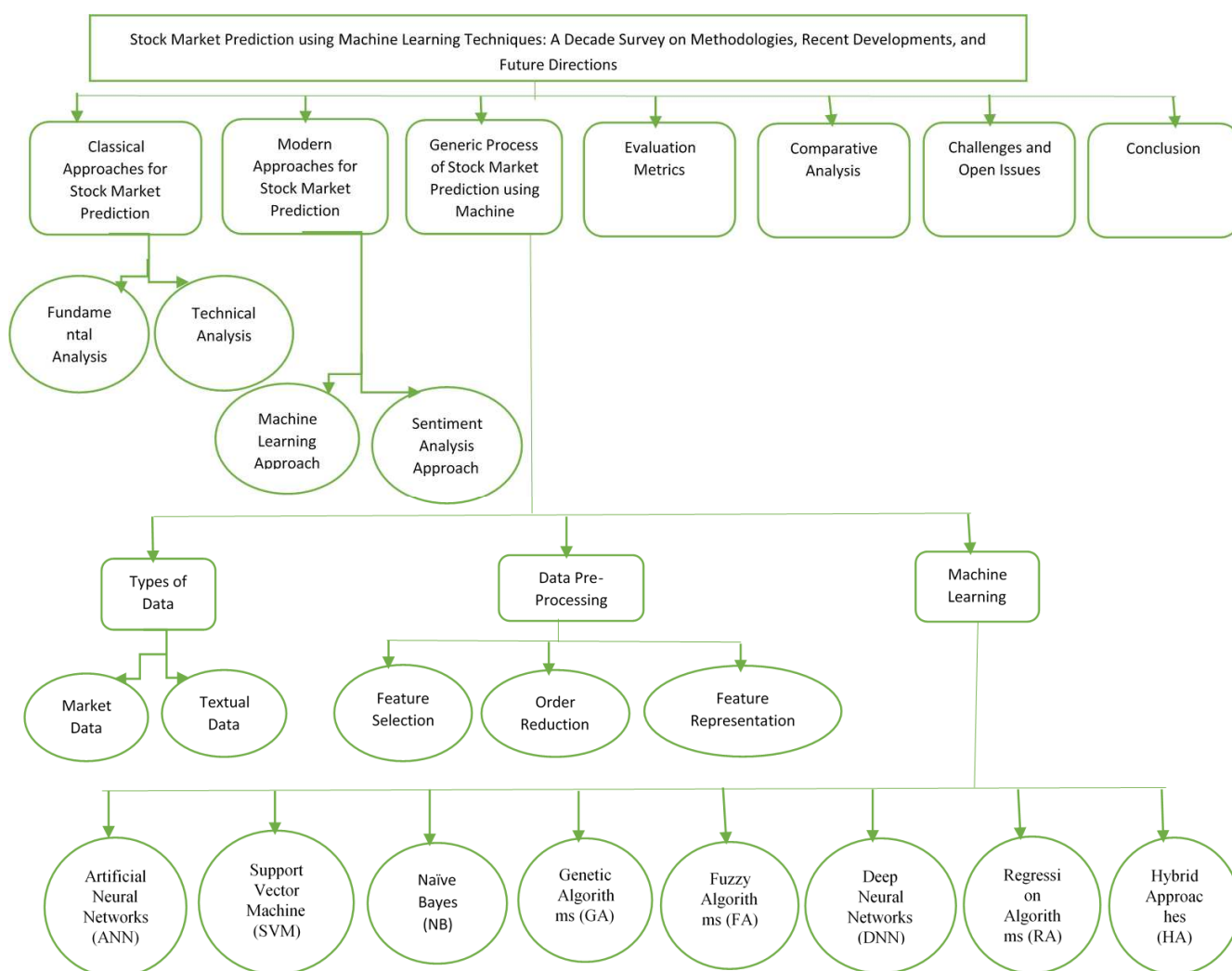


Figure 1. Article outline.

1.1. Classical Approaches for SMP

According to [10], there exist two main traditional approaches to the analysis of the stock markets: (1) fundamental analysis and (2) technical analysis.

1.1.1. Fundamental Analysis

Fundamental analysis calculates a genuine value of a sector/company and determines the amount that one share of that company should cost. A supposition is made that, if given sufficient time, the company will move to a cost agreeing with the prediction. If a sector/company is undervalued, then the market value of that company should rise, and conversely, if a company is overvalued, then the market price should fall [11]. The analysis is performed considering various factors, such as yearly fiscal summaries and reports, balance sheets, a future prospectus, and the company's work environment [12]. If stocks are overvalued, then the market price will fall [13], e.g., the Dotcom bubble burst in the year 2000 [14]. The two most common metrics used to predict long-term price movements yearly for fundamental analysis are (a) the Price to Earnings ratio (P/E) and (b) the Price by Book ratio (P/B). The P/E ratio is used as a predictor. The companies with a lower P/E ratio yield higher returns than companies with a high P/E ratio [15]. Financial analysts also use this to prove their stock recommendations [16]. Fundamental analysis can be used for the consideration of financial ratios to distinguish poor stocks from quality stocks [17]. The P/B ratio compares the company value specified by the market to the company value specified on paper. If the ratio is high, the company may be overvalued, and the company's value might fall with time. Conversely, if the ratio is low, the company may be underestimated, and the price may rise with time. Of course, fundamental analysis is a powerful method. Still, it has some drawbacks. Fundamental analysis, firstly, lacks adequate knowledge of the rules governing the workings of the system, and secondly, there is non-linearity in the system [18].

1.1.2. Technical Analysis

Technical analysis is the study of stock prices to make a profit, or to make better investment decisions [19]. Technical analysis predicts the direction of the future price movements of stocks based on their historical data, and helps to analyze financial time series data using technical indicators to forecast stock prices. Meanwhile, it is assumed that the price moves in a trend and has momentum [20]. Technical analysis uses price charts and certain formulae, and studies patterns to predict future stock prices; it is mainly used by short-term investors. The price would be considered high, low or open, or the closing price of the stock, where the time points would be daily, weekly, monthly, or yearly. Dow theory puts forward the main principles for technical analysis, which are that the market price discounts everything, prices move in trends, and historic trends usually repeat the same patterns [21]. There are several technical indicators, such as the Moving Average (MA), Moving Average Convergence/Divergence (MACD), the Aroon indicator, and the money flow index, etc. The evident flaws of technical analysis as per [18] are that expert's opinions define rules in technical analysis, which are fixed and are reluctant to change. Various parameters that affect stock prices are ignored.

The prerequisite is to overcome the deficiencies of fundamental and technical analysis, and the evident advancement in the modelling techniques has motivated various researchers to study new methods for stock price prediction. A new form of collective intelligence has emerged, and new innovative methods are being employed for stock value forecasting. The methodologies incorporate the work of machine learning algorithms for stock market analysis and prediction.

1.2. Modern Approaches for SMP

There are some modern approaches that can be functional and fruitful for SMP that would enhance prediction accuracies. In this review, we will highlight some modern functional approaches.

1.2.1. Machine Learning Approach

Because of global digitization, SMP has entered a technological era. Machine learning in stock price prediction is used to discover patterns in data [22]. Usually, a tremendous amount of structured and unstructured heterogeneous data is generated from stock markets. Using machine learning algorithms, it is possible to quickly analyze more complex heterogeneous data and generate more accurate results. Various machine learning methods have been used for SMP [23]. The machine learning approaches are mainly categorized into supervised and unsupervised approaches. In the supervised learning approach, named input data and the desired output are given to the learning algorithms. Meanwhile, in the unsupervised learning approach, unlabeled input data is provided to the learning algorithm, and the algorithm identifies the patterns and generates the output accordingly. Furthermore, different algorithmic approaches have been used in SMP, such as the Support Vector Machine (SVM), k Nearest Neighbors (kNN), Artificial Neural Networks (ANN), Decision Trees, Fuzzy Time-Series, and Evolutionary Algorithms. The SVM is a supervised machine learning technique that limits error and augments geometric margins, and is a pattern classification algorithm [24]. In terms of accuracy, the SVM is an important machine learning algorithm compared to the other classifiers [25]. In the kNN, stock prediction is mapped into a classification based on closeness. Using Euclidean distance, the kNN classifies the “k” nearest neighbors in the training set. The ANN is a nonlinear computational structure for various machine learning algorithms to analyze and process complex input data together. The FIS (Fuzzy Inference Systems) apply rules to fuzzy sets and then apply de-fuzzification to give crisp outputs for decision making [26]. The evolutionary algorithms include gene-inspired neuro-fuzzy and neuro-genetic algorithms, mimic the natural selection theory of species, and can give an optimal output.

1.2.2. Sentiment Analysis Approach

One of the phenomena of current times that is changing the world is the global availability of the internet. The most-used platforms on the internet are social media. It is estimated that social media users all over the world will number around 3.07 billion [27]. There is a high association between stock prices and events related to stocks on the web. The event information is extracted from the internet to predict stock prices; such an approach is known as event-driven stock prediction [28]. Through social networks, people generate tremendous amounts of data that is filled with emotions. Much of this data is related to user perceptions and concerns [29]. Sentiment analysis is a field of study that deals with the people's concerns, beliefs, emotions, perceptions, and sentiments towards some entity [30,31]. It is the process of analyzing text corpora, e.g., news feeds or stock market-specific tweets, for stock trend prediction. The Stock Twits, Twitter, Yahoo Finance, and so on are well-known platforms used for the extraction of sentiments. There is a significant importance of using sentimental data for enhancing the prediction of volatility in the stock market. The ‘Wisdom of Crowds’ and sentiment analysis generate more insights that can be used to increase the performance in various fields, such as box office sales, election outcomes, SMP, and so on [32]. This suggests that a good decision can be made by taking the opinions and insights of large groups of people with varied types of information [33]. The information generated through social media allows us to explore vast and diverse opinions. Exploring sentiments from social media in addition to numeric time-series stock data would enhance the accuracy of the prediction. Using time-series data as well as social media data would intensify the prediction accuracy. Different approaches and techniques have been proposed over time to anticipate stock prices through numerous methodologies, thanks to the dynamic and challenging panorama of stock markets [34].

2. Research Methodology

This section explains the overall process of the literature collection on SMP using machine learning. Initially, the phrase “stock market prediction using machine learning” was keyed to various search engines, digital libraries and databases, including ‘google

scholar', 'research gate', 'ACM digital library', 'IEEE Explore', 'Scopus', and so on. During the process of literature collection, various phrases like "stock market prediction methods", "impact of sentiments on stock market prediction", and "machine learning-based approach for stock market prediction" were keyed. The OR and AND operators were used for the keyword searches in single and multiple classes, respectively. As a result, some of the fundamental papers in the field of stock market prediction were retrieved. By the careful analysis of a few basic papers, a primary insight into the domain was obtained. The search criteria were further modified to collect the literature of the last decade, in order to enhance and improve the domain. In addition, the literature selected was screened by applying quality criteria, where metrics such as indexing, quartiles, impact factors and publishers were observed. Figure 2 presents the steps followed in the literature collection.

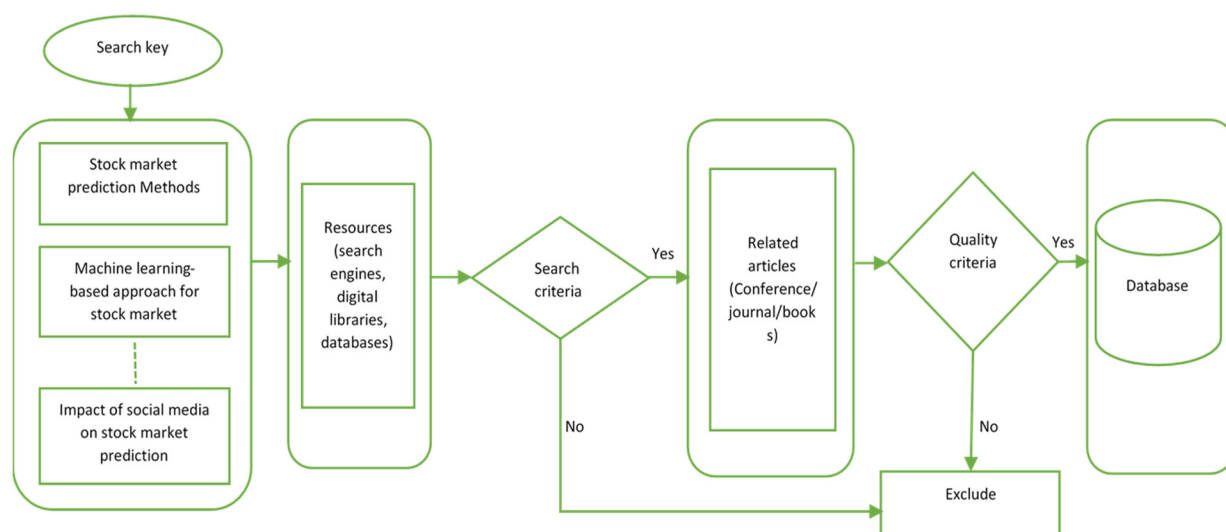


Figure 2. Literature collection process.

3. Generic Scheme for SMP

Figure 3 describes the generic process involved in SMP. The process starts with the collection of the data, and then pre-processing that data so that it can be fed to a machine learning model. The prediction models generally use two types of data: market and textual data. The literature of both types is discussed in the following section. The next section classifies the previous studies based on the type of data used. Furthermore, the next section surveys the previous studies based on the various data-preprocessing approaches applied. Moreover, the literature is further surveyed based on the machine learning algorithms used by different systems.

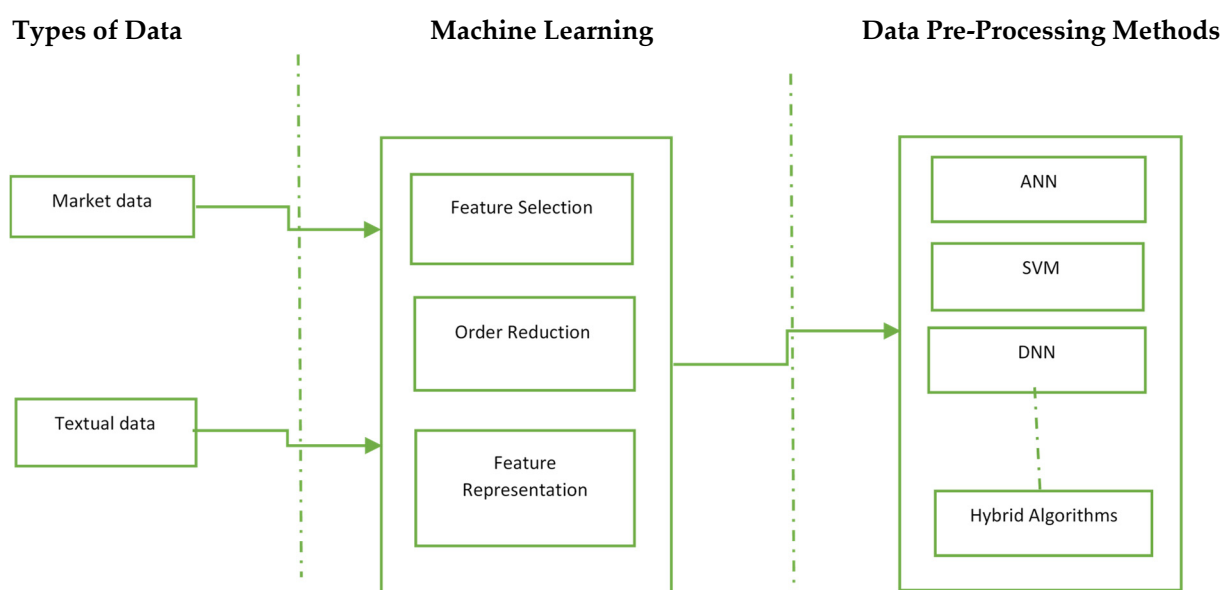


Figure 3. Generic Scheme for SMP (Stock Market Prediction).

4. Types of Data

SMP systems can be classified according to the type of data they use as the input. Most of the studies used market data for their analysis. Recent studies have considered textual data from online sources as well. In this section, the studies are classified based on the type of data they use for prediction purposes. At the end of this section, Table 1 points out the comparison of the data sources, type of input and prediction duration used in the studies so far.

4.1. Market Data

Market data are the temporal historical price-related numerical data of financial markets. Analysts and traders use the data to analyze the historical trend and the latest stock prices in the market. They reflect the information needed for the understanding of market behavior. The market data are usually free, and can be directly downloaded from the market websites. Various researchers have used this data for the prediction of price movements using machine learning algorithms. The previous studies have focused on two types of predictions. Some studies have used stock index predictions like the Dow Jones Industrial Average (DJIA) [35], Nifty [36], Standard and Poor's (S&P) 500 [37], National Association of Securities Dealers Automated Quotations (NASDAQ) [38], the Deutscher Aktien Index (DAX) index [39], and multiple indices [40,41]. Other studies have used individual stock prediction based on some specific companies like Apple [42], Google [43], or groups of companies [12,44].

Furthermore, the studies focused on time-specific predictions like intraday [45], daily [20], weekly [46], and monthly predictions [47], and so on. Moreover, most of the previous research is based on categorical prediction, where predictions are categorized into discrete classes like up, down, positive, or negative [32,48]. Technical indicators have been widely used for SMP due to their summative representation of trends in time series data. Some studies considered different types of technical indicators, e.g., trend indicators, momentum indicators, volatility indicators and volume indicators [32,49,50]. Furthermore, numerous studies have used an amalgam of different types of technical indicators for SMP [42,51].

4.2. Textual Data

Textual data is used to analyze the effect of sentiments on the stock market. Public sentiments have been proven to affect the market considerably. The most challenging part is

to convert the textual information into numerical values so that it can be fed to a prediction model. Furthermore, the extraction of textual data is a challenging task. The textual data has many sources, such as financial news websites, general news, and social platforms [52]. Most of the studies were carried out on textual data try to predict whether the sentiment towards a particular stock is positive or negative. The previous studies considered several textual sources for SMP, such as the Wall Street Journal [53], Bloomberg [22], CNBC and Reuters [54], Google Finance [55], and Yahoo Finance [56]. The extracted news may be either generalized news or some specific financial news, but the majority of the researchers use financial news, as it is deemed to be less susceptible to noise [57]. Some researchers have used less formal textual data, such as message boards [58,59].

Meanwhile, the textual data from microblogging websites and social networking websites are comparably less explored than other textual data forms for SMP. Besides this, one challenge faced for the processing of the textual data are that the information generated on these platforms is enormous, increasing the computational complexities [60,61]. For example, the researchers in [62] processed 1,00,000 tweets, and the researchers in [63] processed around 2,500,000 tweets, which was a complex task. Moreover, for the textual data, no proper standard format is followed while posting on social media, which increases the processing complexities. In addition, the detection of shorthand spellings, emoticons and sarcastic statements is yet another challenge. Machine learning algorithms come to the forefront to deal with all kinds of challenges faced while processing textual data. Previous studies have mostly considered the sentiment of textual data as positive or negative [35,48,64]. A few studies have considered mood words to determine the mood of a tweet, such as [8,58,65].

5. Data Pre-Processing

Once the data is available, it needs some pre-processing so that it can be fed to a machine learning model. The significance of the output depends on the pre-processing of the data [66,67]. The textual data must be transformed into a structured format that can be used in a machine learning model. The previous studies revealed that there are three significant pre-processing steps, i.e., feature selection, order reduction and the representation of features. Table 1 presents the comparison of the data sources, type of input and prediction duration. Table 2 presents the comparison of the data pre-processing techniques used in the studies so far.

Table 1. Comparison of the data sources, type of input and prediction duration.

References	Data	Type of Input	Prediction Duration
[37]	S&P 500	Market data	Few days ahead
[38]	NASDAQ index	Market data	Few days ahead
[39]	DAX 30	broker house newsletters, RSS market feeds, and stock exchange data	Intraday
[56]	Yahoo Finance	Financial News	Intraday
[44]	DGAP, Euro-Adhoc	Corporate announcements financial new	Daily
[58]	Yahoo finance (18 Stock Companies data)	Market data, yahoo finance message board data	Daily
[35]	DJIA	Market data and Twitter	Daily
[32]	BSE and NSE stocks	Market data, technical indicators, Twitter data	Intraday
[36]	Nifty and Sensex	Market data and news	Intraday

Table 1. *Cont.*

References	Data	Type of Input	Prediction Duration
[47]	Yahoo Finance	Market data, Twitter data, and news data	Daily and monthly
[49]	S&P, NYSE, DJIA	Market data, Technical Indicators, Social media data	Daily weekly
[51]	Apple, yahoo	Market data, technical indicators	60 day and 90-day prediction
[63]	Microsoft company	Twitter	Daily
[42]	NASDAQ, DJIA, Apple Stock (AAPL)	Market data, technical indicators, news.	One-day ahead
[43]	Google stock	Market data	Five days horizon
[45]	Taiwan Stock Exchange CWI	Market data	High-frequency trading
[68]	S&P 500	Market data	Daily
[50]	Columbia Stock Market	Market data, Technical indicators	Next day
[69]	S&P 500	Financial news from Noodle, Reuters	Intraday
[70]	Enron Corpus	Sentiment data	Daily, weekly
[46]	BSE, Tech Mahindra	Market data	Daily and weekly
[71]	Apple stock data	Market data	Daily
[72]	United States stock exchange	Market data, technical indicator	Daily
[73]	KSE, LSE, Nasdaq, NYSE	Twitter, yahoo finance, Wikipedia	Weekly
[74]	Google stock	Market data	Daily

Table 2. Comparison of various pre-processing techniques.

References	Feature Selection	Order Reduction	Feature Representation
[39]	Bag of Words	Stemming	Sentiment value
[56]	Opinion Finder overall tone and polarity	Minimum Occurrence per document	Boolean
[44]	Bag-of-words, noun phrases, word combinations, n-gram	Frequency for news, Chi2-approach and bi-normal separation (BNS) for exogenous-feedback-based feature selection, dictionary	TF-IDF
[75]	Bag-of-words	WordNet to replace words	TF-IDF
[76]	N-grams	Document frequency	Boolean
[32]	Context based approach	SentiWordNet	Sentiment value

Table 2. Cont.

References	Feature Selection	Order Reduction	Feature Representation
[58]	Bag of words, LDA, JST, Aspect Based	-	TF-IDF
[47]	Correlation	Lemmatization	Boolean
[69]	Bag of Words	Chi2, Information Gain, Document Frequency, Occurrence	TF-IDF
[77]	Bag-of-word, Word2vec		TF-IDF
[78]	GA	PCA, FA, FO	-
[79]	N-grams	SVM based Recursive Feature Elimination, PCA, KPCA, and XGB	-
[73]	Bag-of-words	Occurrence	TF-IDF
[80]	GA, Feature Ranking	PCA-SVM, DA-RNN	-

5.1. Feature Selection

Feature selection is a crucial step in textual data processing. Most of the studies on SMP have used basic feature extraction techniques such as Bag of Words, where the text is broken into words and each word is converted into a numeric feature. The feature selection depends on the number of occurrences of a word. Table 2 points out the various feature selection methods used in the literature so far.

As in [67,70], most of the previous literature used feature selection techniques where the order of words is discarded, causing the loss of context. Another feature selection method is Word2Vec, as proposed by [32,81]. It is a word embedding technique based on a multi-layer perceptron. This technique takes into consideration the order and co-occurrence of words, and hence retains the context. Word2Vec has been used in some of the works, such as [63,77].

Moreover, a few studies have used Latent Dirichlet Allocation (LDA) [82–84], where words are viewed as a probabilistic collection of concepts, and the concepts are used as features [82–84]. Some works [44,76,79] used the N-grams technique. N-grams is the contiguous collection of N words from a given sequence of text. Other methods like genetic algorithms and particle swarm optimization have also been used for feature selection, like in [80,85].

5.2. Order Reduction

The feature selection process for the textual data leads to an increase in the number of features. High dimensional features are extremely difficult to process, and leads to the poor efficiency of most of the learning algorithms [86]. This phenomenon is known as the curse of dimensionality [87]. Lower numbers of features will decrease the training complexity of the algorithms. Table 2 points out the order reduction techniques used in previous studies. A well-known form of multi-variant analysis, Principal Component Analysis (PCA), is used to select the most relevant features, reducing the dimensionality of the features. In [68], the daily direction of the S&P 500 index is predicted using 60 features. The authors used three variants of PCA, and concluded that the inclusion of PCA not only reduced the overall training complexity but also increased the accuracy of the predictions. In [78], numerous feature reduction techniques, e.g., PCA, Factor Analysis (FA), Genetic Algorithms (GA), and Firefly Optimization (FO), were used to solve the data complexity.

5.3. Feature Representation

Feature representation is one of the important factors for the efficient training of machine learning algorithms. Once the number of required features is determined, the input data is converted to a numeric representation so that machine learning algorithms can readily process it. Table 2 presents the type of representation or the weighing used in the literature so far. Boolean representation is one of the most basic techniques of feature representation, in which the presence and absence of the feature (word) are represented by 1 and 0, respectively, for Bag of words [56]. Another technique, Term Frequency-Inverse Document Frequency (TF-IDF), has been used in numerous studies [44,69,73]. Generally, the text pre-processing phase is considered to be a crucial phase, and may significantly impact the model's accuracy [88,89].

6. Machine Learning Methods

This section attempts to summarize the machine learning models used in previous studies for stock prediction and forecasting. After the data is pre-processed and transformed to a standard representation, it is fed to machine learning models for further processing. Table 3 presents the distribution of various machine learning techniques used in literature so far. The following section briefly summarizes the different machine learning approaches presented:

- Artificial Neural Networks (ANN)
- Support Vector Machine (SVM)
- Naïve Bayes (NB)
- Genetic Algorithms (GA)
- Fuzzy Algorithms (FA)
- Deep Neural Networks (DNN)
- Regression Algorithms (RA)
- Hybrid Approaches (HA)

6.1. Artificial Neural Networks (ANN)

The ANN is a biological brain-inspired technique in which a large number of artificial neurons are strongly interconnected in order to solve complex problems [90]. These models understand the context of a problem by creating multiple transformations on the feature space, followed by non-linearity, to create its simplified representations [91]. Numerous studies have employed ANN models for SMP [38,40,92–95]. For example, the authors in [68] employed ANN for daily trend prediction of the S&P 500 index. Three-dimensional reduction techniques—e.g., PCA, Fuzzy Robust Principal Component Analysis (FRPCA), and Kernel-based Principal Component Analysis—were applied to streamline the dataset. The results suggested that combining the ANNs with the PCAs is more efficient. Furthermore, the selection of an appropriate kernel function directly affects the performance of KPCA [68].

Multilayer perceptron (MLP) is a frequently used technique for SMP [42,43,96,97]. MLP is an ANN with one input and output layer, and one or more intermediate layers. Generally, the MLP uses the backpropagation method for training, in which predicted errors are back propagated from the output layer to the input layer to minimize the errors [98,99].

A study compared three ANN models—MLP, dynamic artificial neural network (DAN2) and autoregressive conditional heteroscedasticity (GARCH)—for NASDAQ price prediction [38]. All three models were evaluated using the Mean Absolute Deviate (MAD) and Mean Square Error (MSE). The results demonstrated that the MLP outperformed DAN2 and GARCH-MLP. Furthermore, it provides a future direction for researchers by suggesting that they focus on finding out whether GARCH has a remedying impact on forecasts or other correlated variables that have a remedying impact on forecasts.

Another study used Generalized Feed Forward (GFF) and MLP models for the prediction of the Istanbul Stock Exchange (ISE) market index [100], where the data were taken from the Central Bank of Turkey. A total of eight sets of predictions (six ANN and two

MA) were performed by changing the number of hidden layers. Two sets of predictions were based on MA. The accuracy of the prediction was calculated using the coefficient of determination, and the highest accuracy for both MLP and GFF was achieved using one hidden layer.

In addition, often-used ANN technique for SMP is the Radial Basis Function network (RBF). It is a layered network where hidden layers use a radial activation function [101,102]. For example, the authors in [91] used the RBF neural network to predict the Shanghai and NASDAQ index by using an extension of LPP (Locality Preserving Projection) known as two-dimensional LPP for the selection of most relevant features for the prediction. The proposed method performed well on both of the market indices.

6.2. Support Vector Machine (SVM)

The SVM is a supervised machine learning technique that limits error and augments geometric margins. It is a pattern classification and regression algorithm that was given by [24]. In terms of accuracy, the SVM is an important linear separation algorithm compared to other classifiers [25]. As presented in Table 4, it is the most popular method used for SMP [39–41,44,50,51,72,103,104].

The authors in [47] developed a daily and monthly SMP model using historical and sentimental data for the bank, mining, and oil sectors. The historical prices were obtained from yahoo finance, and a sentiment dataset was created by using news and tweets for one year. PCA with multiple factors was applied to the sparse dataset considered for the sentiment analysis. In this study, three algorithms—i.e., Decision-Boosted Tree, SVM, and Logistic Regression—were compared, and the accuracy was used as a performance metric. The Decision-Boosted Tree outflanked the Logistic Regression and SVM. The Decision-Boosted Tree achieved accuracies of 54.8%, 76%, and 76.9% for the bank, mining, and oil sectors, respectively. The Logistic regression attained accuracies of 65.4%, 61%, and 44.2%, respectively, and the SVM achieved accuracies of 51%, 59%, and 44.2% for the respective sectors. The study finally suggested the consideration of the impact of intra-day price movement for the next-day stock price to improve the accuracy.

6.3. Naïve Bayes (NB)

NB is a classification method that classifies the data points based on the Bayesian Theorem of probability. This classification method is extremely fast, and can scale over large datasets. This classification approach has been used widely for SMP [49,69,103,105–107]. For example, the authors in [108] employed the Naïve Bayes algorithm for the sentiment analysis of textual data from multiple sources. The authors compared the effect of conventional and social media data sources on different companies and their interrelatedness.

6.4. Genetic Algorithms (GA)

GA are a heuristic approach to problem-solving that mimic the natural evolution process. The algorithms apply the concept of natural selection to select the optimal possible solution. In SMP, GA is used to fine-tune the parameters for the generation of the best trading rule. Numerous studies have used GA to enhance SMP accuracies [4,26,78,80,97,109–111]. For example, the authors in [112] developed an intelligent decision support system for stock trading. This study employed rough sets and GA for non-linear and complex stock data to find the features that can be used to generate the optimal trading rules. These rules are applied to generate optimal buy or sell strategies.

6.5. Fuzzy Algorithms (FA)

Fuzzy logic is a human reasoning-based method where all of the intermediate possibilities between 0 and 1 are used for decision making. It is a powerful approach in which the degree of belongingness to a certain category is considered. The adaptive neuro-fuzzy inference system (ANFIS) is the most popular fuzzy algorithm that is used for SMP. Some example studies in which ANFIS was employed for SMP include [72,113,114]. The authors

in [29] developed a fuzzy logic approach to analyze the sentiments on social media for SMP. Furthermore, several studies have used hybrid fuzzy approaches for SMP [115–117]. As an example of a hybrid fuzzy approach, Sedighi et al. (2019) proposed a novel model for prediction using an Artificial Bee Colony (ABC), SVM, and ANFIS. In this study, data from 50 companies were taken from the US Stock Exchange from 2008–2018. The model used 20 technical indicators as the input. The criteria for the performance measures were accuracy and quality. Furthermore, the model had a more exact forecasting accuracy than the others.

6.6. Deep Neural Networks (DNN)

DNN are an improvement over conventional neural networks where more hidden layers and neurons are employed for automatic feature extraction and transformation. The increase in the number of hidden layers with non-linear processing units improves the efficiency of learning from raw data [118]. DNN have been used frequently for financial predictions using textual and numeric data [74,119]. Different studies have used DNN algorithms such as Convolutional Neural Networks (CNNs) [120–122], Long-Short Term Memory (LSTM) [123,124], and Deep Belief Networks (DBNs) [106,125–127]. For example, a recent study by [128] made a comparison of four prediction models for stock market price prediction, including an Auto-Regressive Integrated Moving Average (ARIMA), Vector Auto Regression (VAR), LSTM, and Nonlinear Auto-Regressive with exogenous inputs (NARX). The model performance was evaluated using an accuracy metric. The data used for the analysis were the closing price of the NASDAQ. The results revealed that NARX made accurate predictions for the short term but failed in long-term predictions. It also concluded that models that integrate machine learning and technical indicators could predict more accurately. LSTM networks are able to learn long-term dependencies, such that they have a vigilant effect on time series prediction. Moreover, the authors in [43] compared three Recurrent Neural Network (RNN) models on Google stock price data, namely basic RNN, Gated Recurrent Unit (GRU) and the LSTM. The results revealed that LSTM outperformed other techniques and achieved an accuracy of 72% on a 5-day horizon. Furthermore, the authors in [129] applied the dynamic LSTM network to predict Nifty prices using Open, High, Low, and Close as features, and achieved a Root Mean Square Error (RMSE) of 0.00859 in terms of daily percentage changes.

6.7. Regression Algorithms (RA)

Regression is a predictive approach that models the relationship between a dependent variable and independent variables [130]. Different regression approaches have been used in previous studies: simple linear regression [131–133], multiple regression [134,135], decision tree regression [17,136], logistic regression [137], support vector regression (SVR) [56,138], and ensemble regression [41,69,139]. For example, the authors in [140] developed a model that predicts the stock price of a user-specified company a few days ahead. Regression analysis and candlestick pattern detection were applied to the data, which were collected from multiple sources. The model predicted the market movement to a satisfactory level of efficiency. Furthermore, different machine learning algorithms were used, and an improved accuracy of 85% was achieved.

6.8. Hybrid Approaches (HA)

The hybrid approach is the amalgam of various techniques used for the enhancement of performance in prediction models. Hybrid algorithms increase the efficiency of prediction models, as suggested by [141]. A few studies have used various hybrid approaches [9,59,71,72,114,142–144]. For example, the authors in [145] proposed a novel, intelligent, hybrid model for stock prediction by combining the predictions of the linear and non-linear models. The authors used an exponential smoothing model as a linear model and an autoregressive moving reference neural network as a non-linear model. The initial predictions were performed by a linear model, and the prediction errors were calculated

and then fed to an autoregressive moving reference neural network. This model minimized the errors due to non-linear processing. Summation and multiplication methods were used for the generation of predictions from the prediction errors. The NSE data were used for the testing of the model, and the results indicated that the model could be a promising and a new approach for the prediction of stock returns.

Table 3. Comparison of the different techniques applied in the studies. Artificial Neural Networks (ANN), Support Vector Machine (SVM, Naïve Bayes (NB), Fuzzy Logic (FL), Hybrid Algorithms (HA), Genetic Algorithms (GA), Regression Algorithms (RA), Ensemble Algorithms (EA).

References	ANN	SVM	NB	DNN	FL	HA	GA	RA	EA
[134]								✓	
[38]	✓								
[44]		✓							
[39]		✓							
[56]								✓	
[92]	✓								
[40]	✓	✓							
[143]					✓	✓			
[58]		✓							
[63]		✓							✓
[103]		✓	✓						✓
[51]		✓							✓
[68]	✓								
[43]	✓			✓					
[102]	✓								
[114]						✓			
[42]	✓	✓						✓	
[45]							✓	✓	
[69]		✓	✓					✓	
[77]				✓					
[133]		✓						✓	
[106]			✓	✓	✓				
[71]					✓	✓			
[72]		✓			✓	✓			
[124]	✓			✓					
[41]		✓						✓	✓
[74]		✓		✓					
[73]	✓		✓						✓
[80]				✓			✓		

In terms of price predictions, the model outperformed the RNN and achieved a lower MSE and Mean Absolute Error (MAE) compared to the constituent models.

7. Evaluation Metrics

Generally, two approaches are used for SMP: classification and regression. The former approach classifies the market trend into categories like Up and Down. For the latter, output is a numerical value predicting the ups and downs of the price. Figure 4 presents the taxonomy of the evaluation metrics used in the studies so far. Table 4 points out the different evaluation parameters used in the reviewed studies, as well as the time frame of the prediction. For the most part, the studies used accuracy as an evaluation metric, which is the percentage ratio of correct predictions over the total number of test instances [146].

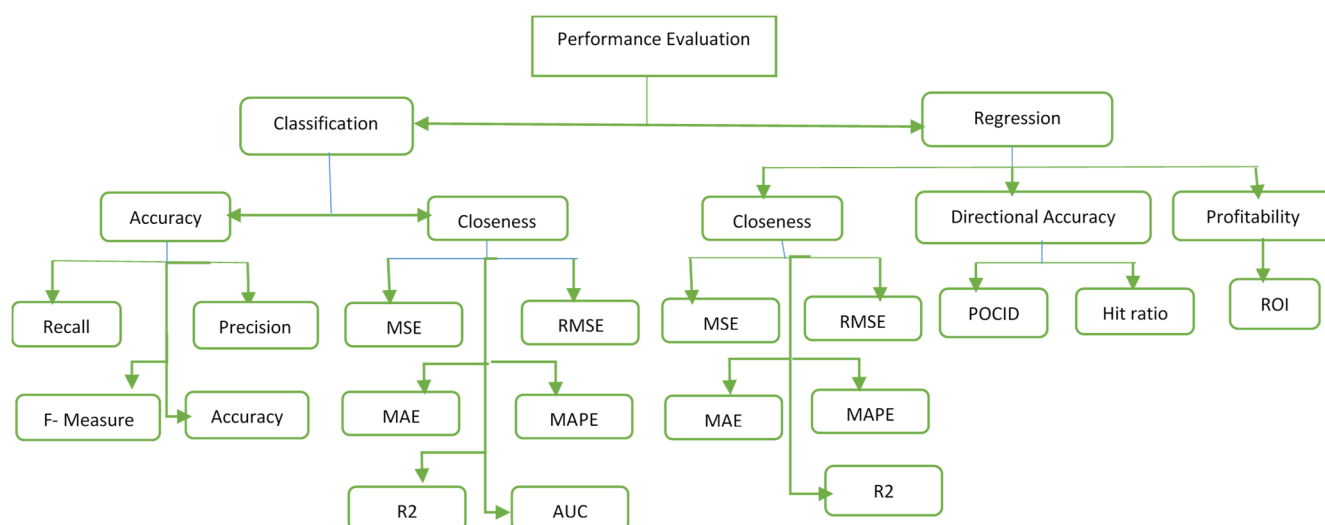


Figure 4. Taxonomy of the performance metrics.

Moreover, other metrics like MSE, the Area Under Curve (AUC), Akaike Information Criterion (AIC), R-Squared (R2), Precision, Recall, F-measure, MAE and Mean Absolute Percentage Error (MAPE) are used as well. The accuracy of almost all of the studies lies in a range of 50–90%, as presented in Table 4. The accuracy metric is handy to use and has less computation complexity than the other metrics. However, it doesn't consider the importance of Type 1 and Type 2 errors in the case of skewed data distributions [147]. The MSE measures the mean difference between the predicted and actual output. It is an important metric in regression analysis because it measures how close the predicted value is to the actual value. An area under curve measures the degree of separability between classes. It is an important metric for classification problems. The higher the AUC, the better the model's predictability [148]. Precision is the ratio of correctly predicted positives to the total positives predicted [42]. Recall is the ratio of correctly predicted positives to the total number of positives [73]. The F-measure is the harmonic mean of the precision and recall, and indicates the importance of false positives and false negatives in the confusion matrix [42,49,69]. The R2, or the coefficient of determination, is the measure of the closeness of the data to the predicted regression line [41,133]. The MAE measures the average difference between the predicted and actual data [133].

Table 4. Comparison of the evaluation metrics applied to different SMP methods.

Reference	Performance Measure	Prediction Type	Output
[38]	MSE, MAD%	Few days ahead	MAD% (2.32) MLP
[56]	Accuracy, trading return	Intraday	59.0%, 3.30%
[44]	Accuracy	Daily	65.1%
Kalyanaraman, V., 2014	Accuracy	Daily	81.82%
	Accuracy	Daily	Average accuracy of 54.4%
	Accuracy, RMSE	Daily	59.6%
[47]	Accuracy	Daily and monthly	DBT achieved better accuracy (76.9%) than SVM and LR
[63]	Accuracy and correlation	Daily	Accuracy of around 70%
[51]	Accuracy, RMSE	Long-term	99% accuracy for yahoo data (XGBoost)
[49]	Error Rate, F-measure	Next Month, Next Week	0.85
[42]	Accuracy, f-measure, precision, AUC	One day ahead	85%
[43]	Log loss and accuracy	Daily, weekly	72% accuracy (LSTM)
[68]	Accuracy, Return	Daily	58.1%

Table 4. Cont.

Reference	Performance Measure	Prediction Type	Output
[123]	Accuracy, MSE	Long short-term	56.7% (LSTM), 57.2% (ELSTM)
[70]	Accuracy	Daily, weekly	80%
[69]	Accuracy, f-measure		0.84
[50]	Accuracy	Daily	72%
[133]	MSE, MAE, MAPE and R2	Daily	LR 0.73 SVM 0.93
[106]	MAPE	Daily	2.03–2.17
[71]	training error, testing error	Daily	0.03, 0.072
[41]	RMSE, Accuracy, AUC, R2, MAE	Monthly	>90% (Ensemble)
[74]	Accuracy	Monthly	87.32
Seethalakshmi, R., 2020	R2, AIC	0.992 (R2)
	Accuracy	Next few days	90–96% (KNN Regression)
[140]	Precision, recall, f-measure, accuracy	Weekly	76.5%
[73]	accuracy	Weekly	76.5%
[80]	MSE	Daily	0.0039 (GA-LSTM)

Moreover, the MAPE is used as a performance indicator in a few studies that measure the mean of absolute error percentages in predictions [38,106]. Furthermore, a few of the reviewed studies used trading return or return on investment (ROI) as an evaluation metric, where the trading technique was tested to measure the profitability of predictions [56,68]. Other studies have used Prediction of Change in Direction (POCID) [149] and hit ratios [144].

8. Overfitting

One of the most well-known and challenging issues in machine learning models is overfitting. In this phenomenon, the model tries too hard to learn from training data. This means that the model picks up on noise or random fluctuations in the training data and learns them as ideas. These ideas don't apply to the new data that is to be predicted, thereby resulting in poor model generalization. Because stock market data is highly stochastic, it is imperative to explain the methods used to resolve this issue. The most common approach to mitigate the issue of overfitting is cross validation. A few studies have applied this approach, like in [38,42,150–152]. In a typical k-fold cross-validation, the data is partitioned into k subsets, or folds. The model is trained iteratively on k-1 folds, and the remaining fold—also known as the hold-out fold—is treated as a test set. Numerous studies have used the early stopping method to overcome overfitting [153]. Another method is to remove irrelevant features and noise from the data, which greatly increases the model's generalizability. A few studies have implemented these procedures to avoid overfitting, such as [42,44,68,121]. The most important preventive measure against overfitting is regularization. This technique removes the extra weights from the selected features and redistributes them uniformly. It discourages the learning of models that are complex or more flexible, hence avoiding the risk of overfitting. The majority of the reviewed studies applied regularization approaches to prevent overfitting [23,25,83]. A few recent studies applied the procedure of data augmentation to prevent overfitting [154,155].

9. Comparative Analysis

The distribution of the number of papers published in recent years is presented in Figure 5. The number of publications increased from 2009, and was at its peak in 2019, but over the previous two years, the publication number was low. The distribution of machine learning algorithms used for SMP is shown in Figure 6, where the SVM was the most popular technique used. However, the ANN and DNN have attracted the research community's attention for the last few years. Traditional neural network approaches may not make accurate SMPs as initially; the weight of the randomly selected problems may suffer from the local optimal, and results in incorrect predictions [123]. The deep learning approaches are used to analyze complicated patterns in the stock data, and provide much

faster results. Furthermore, there is no such single technique that can promise to give the optimum results. The comparative analysis between the type of data used and the performance of the models is represented in Figure 7. Data alone from social media do not perform better than using market data and technical indicators. However, if data from textual sources is combined with them, then the model performance increases.

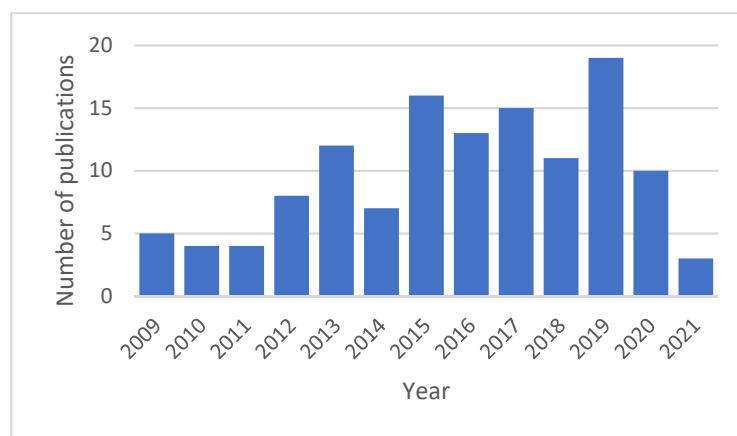


Figure 5. Number of publications per year.

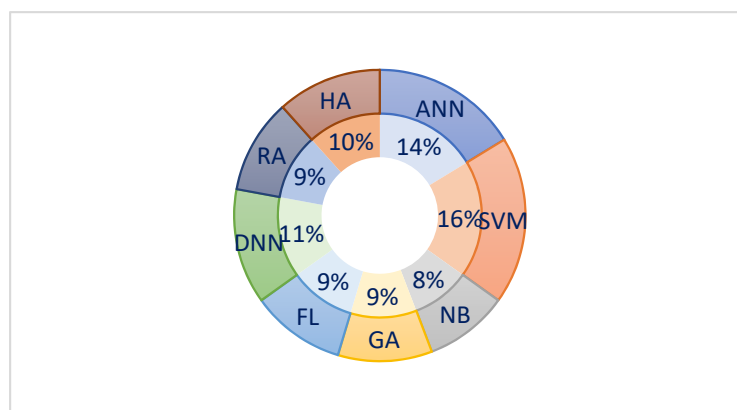


Figure 6. Distribution of the SMP techniques.

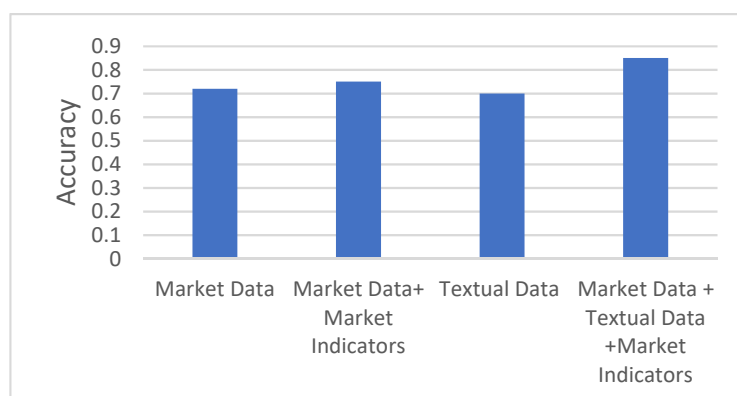


Figure 7. Comparison of the accuracies with different types of data.

10. Challenges and Open Issues

Financial market analysis and prediction continues to be a fascinating and challenging problem. Nowadays, data access is becoming easier, but difficulties are increasing in the acquisition and processing of data to extract valuable insights and analyze their impact on

stock prices. Feature extraction from the financial data is a challenging task, as it is essential to observe the diversity of the variables that are used for the prediction. The Financial datasets are usually noisy [156]. The quality of the data significantly affects SMPs.

Most literature on stock prediction regarding live testing affirms that the previously proposed methodologies can be utilized in real time. However, these methods may work in controlled circumstances. Still, a big challenge will be the live testing for the prediction. The live testing comes up with challenging factors, such as variations in prices, noise, and unpredicted events. One such example is the Knight Capital Tragedy, in which the loss of 440 million dollars was endured by the company [157].

Market volatility is the severity with which the market price of an investment fluctuates. The main reasons for the volatility are uncertainty and inflation, and the risk increases when the market is volatile. The influence of volatility on our emotions is ceaseless. The prediction of stock prices is challenging when the market is volatile. One of the reasons for market volatility is algorithmic trading. One such example is the flash crash, which expunged \$860 billion within 30 min from US stock markets [158]. International politics also plays a dramatic role in stock market volatility [159].

Events in which panic selling is triggered are nowadays becoming more common, and they result in market overreaction. Panic selling is the reaction to fear and loss, which leads to the wide-scale selling of investments. The leading causes which result in panic selling are high speculation in the market, political issues, and economic instability [46,160]. It becomes more difficult for a researcher to evaluate market behavior in such situations.

New algorithms are proceeding to flood the markets consistently at a pace, and it is challenging to compare the adequacy and exactness of these algorithms. A fascinating part of this research area is its self-defeating nature. In simple terms, sharing the methodologies that generate high profits with market competitors will render the methodologies useless. In this way, best-class algorithm exchanging in the markets is restricted, and is private. The procedure or strategy behind such algorithms is never published.

The data on social media platforms can either be generated by humans or bots. The sentiments of bots can sometimes result in inaccurate predictions. As such, there arises a need for social bot detection to obtain better predictions [161]. Investigators, analysts, and researchers are continuously reporting the potential dangers brought about by social bots. Market investors actively participate in and react to social media sentiments. As such, it can be said that the data from social platforms play a significant role in stock prediction. One example is of 23 April 2013, when the Syrian Electronic Army hacked the Twitter account of the Associated Press, and they posted fake news of a terror attack on the White House in which President Obama was allegedly injured. This provoked an immediate crash in the stock markets [162–164]. Due to the rising impact of online networking on numerous aspects of our lives, more attention is paid to sentiment analysis based on data generated from social media. This data can be temperamental and hard to process due to various factors, such as fake news and the bot data published on the web by numerous sources. It is challenging to identify the quality data and draw valuable insights from it. A decent option or an extra asset that can be used is quarterly or yearly reports documented by the organizations for the prediction of stocks. These records, when decoded accurately, give a significant knowledge of an organization's status, which can help with the understanding of the future stock trend.

11. Conclusions

Financial markets provide an excellent platform for investors and traders, who can trade from any gadget that connects to the internet. Over the last few years, people have become more attracted to stock trading. Like any other walk of life, the stock market has also changed due to the advent of technology. Now, people can make their investments grow. Online trading has only changed the way individuals purchase and sell stocks. The budgetary markets have advanced rapidly, and have formed an interconnected global marketplace. These advancements pave the way to new opportunities.

In contrast to conventional frameworks, SMP is currently performed using machine learning, big data analytics, and deep learning, which provide more optimal decision making. Stock markets, nowadays, are vulnerable to social media sentiments and cyber-attacks. Researchers can play a significant role and flourish in these areas by developing the frameworks for better and more secure trading.

This article reviewed studies based on a generic framework of SMP, as presented in Figure 2. It mainly focused on the studies from last decade (2011–2021). The studies were analyzed and compared based on the type of data used as the input, the data pre-processing approaches, and the machine learning techniques used for the predictions. Furthermore, it reviewed the different evaluation metrics used for performance measurement by different studies, as presented in Section 7. Moreover, an extensive comparative analysis was performed, and it was concluded that SVM is the most popular technique used for SMP. However, techniques like ANN and DNN are mostly used, as they provide more accurate and faster predictions. Furthermore, the inclusion of both market data and textual data from online sources improve the prediction accuracies. Section 9 discussed the generic challenges and open issues in SMP systems.

Author Contributions: Conceptualization, N.R., M.B.M., T.A., S.S. (Sparsh Sharma) and S.S. (Saurabh Singh); methodology, N.R., M.B.M. and T.A.; validation, S.A., S.S. (Sparsh Sharma) and H.-C.K.; formal analysis, S.A. and S.S. (Saurabh Singh); data curation, N.R., M.B.M. and T.A.; writing—original draft preparation, N.R., M.B.M. and T.A.; writing—review and editing, N.R., M.B.M., T.A. and S.S. (Sparsh Sharma); supervision, S.A. and H.-C.K.; project administration, S.A. and H.-C.K.; funding acquisition, H.-C.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Commercializations Promotion Agency for R&D Outcomes (COMPA) grant funded by the Korean Government (Ministry of Science and ICT) (R&Dproject No.1711139492).

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

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

Abbreviations

SMP	Stock Market Prediction
SVM	Support Vector Machine
ANN	Artificial Neural Network
DNN	Deep Neural Network
RA	Regression Analysis
FA	Fuzzy Algorithm
NB	Naïve Bayes
GA	Genetic Algorithm
HA	Hybrid Approach
kNN	k- Nearest Neighbors
LDA	Latent Dirichlet Allocation
PCA	Principle Component Analysis
XGB	eXtreme Gradient Boost
FO	Firefly Optimization
TF-IDF	Term Frequency- Inverse Document Frequency
GARCH	Generalized Auto-regressive Conditional Heteroskedasticity
DAN	Deep Attention Neural Network
MLP	Multi-linear Perceptron
GFF	Generalized Feed Forward
NARX	Non-linear Auto-regressive Network with exogenous inputs
RBF	Radial Basis
MA	Moving Average
LPP	Locality Preserving Projection

FRPCA	Fast Robust Principle Component Analysis
KPCA	Kernel Principle Component Analysis
GRU	Gated Recurrent Unit
LSTM	Long Short Term Memory
ANFIS	Adaptive Neuro-Fuzzy Inference System
ABC	Ant Bee Colony
RNN	Recurrent Neural Network
RMSE	Root Mean Square Error
SVR	Support Vector Regression
CNN	Convolution Neural Network
DBN	Deep Belief Network
ARIMA	Auto Regressive Integrated Moving Average
VAR	Vector Auto-regression
AUC	Area Under Curve
MSE	Mean Square Error
MAE	Mean Absolute Error
R2	R-Squared
MAPE	Mean Absolute Percentage Error
POCID	Prediction of Change in Direction
DJIA	Dow Jones Industrial Average
S&P	Standard and Poor's
GDP	Gross Domestic Product
NASDAQ	National Association of Securities Dealers Automated Quotations
DAX	Deutscher Aktien Index
KSE	Karachi Stock Exchange
LSE	London Stock Exchange
NYSE	New York Stock Exchange
BSE	Bombay Stock Exchange
AIC	Akaike Information Criterion

References

1. Krishna, V. ScienceDirect ScienceDirect NSE Stock Stock Market Market Prediction Prediction Using Using Deep-Learning Deep-Learning Models Models. *Procedia Comput. Sci.* **2018**, *132*, 1351–1362.
2. Market Capitalization of Listed Domestic Companies (Current US\$) Data. Available online: <https://data.worldbank.org/indicator/CM.MKT.LCAP.CD> (accessed on 19 May 2021).
3. Upadhyay, A.; Bandyopadhyay, G. Forecasting Stock Performance in Indian Market using Multinomial Logistic Regression. *J. Bus. Stud. Q.* **2012**, *3*, 16–39.
4. Tan, T.Z.; Quek, C.; Ng, G.S. Biological Brain-Inspired Genetic Complementary Learning for Stock Market and Bank Failure Prediction. *Comput. Intell.* **2007**, *23*, 236–261. [[CrossRef](#)]
5. Ali Khan, J. Predicting Trend in Stock Market Exchange Using Machine Learning Classifiers. *Sci. Int.* **2016**, *28*, 1363–1367.
6. Gupta, R.; Garg, N.; Singh, S. Stock Market Prediction Accuracy Analysis Using Kappa Measure. In Proceedings of the 2013 International Conference on Communication Systems and Network Technologies, Gwalior, India, 6–8 April 2013; pp. 635–639.
7. Fama, E.F. Random walks in stock-market prices. *Financ. Anal. J.* **1995**, *51*, 75–80. [[CrossRef](#)]
8. Bujari, A.; Furini, M.; Laina, N. On using cashtags to predict companies stock trends. In Proceedings of the 2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 8–11 January 2017; pp. 25–28.
9. Inthachot, M.; Boonjing, V.; Intakosum, S. Artificial Neural Network and Genetic Algorithm Hybrid Intelligence for Predicting Thai Stock Price Index Trend. *Comput. Intell. Neurosci.* **2016**, *2016*, 3045254. [[CrossRef](#)]
10. Park, C.-H.; Irwin, S.H. What Do We Know about the Profitability of Technical Analysis? *J. Econ. Surv.* **2007**, *21*, 786–826. [[CrossRef](#)]
11. Venkatesh, C.K.; Tyagi, M. Fundamental Analysis as a Method of Share Valuation in Comparison with Technical Analysis. *Bangladesh Res. Publ. J.* **2011**, *1*, 167–174.
12. Nair, B.B.; Mohandas, V. An intelligent recommender system for stock trading. *Intell. Decis. Technol.* **2015**, *9*, 243–269. [[CrossRef](#)]
13. Shiller, R. Measuring bubble expectations and investor confidence RJ Shiller. *J. Psychol. Financ. Mark.* **2000**, *1*, 49–60. [[CrossRef](#)]
14. Aharon, D.Y.; Gavious, I.; Yosef, R. Stock market bubble effects on mergers and acquisitions. *Q. Rev. Econ. Financ.* **2010**, *50*, 456–470. [[CrossRef](#)]
15. Molodovsky, N. A Theory of Price-Earnings Ratios. *Financ. Anal. J.* **1953**, *9*, 65–80. [[CrossRef](#)]
16. Kurach, R.; Słoński, T. The PE Ratio and the Predicted Earnings Growth—The Case of Poland. *Folia Oecon. Stetin.* **2015**, *15*, 127–138. [[CrossRef](#)]

17. Dutta, A. Prediction of Stock Performance in the Indian Stock Market Using Logistic Regression. *Intern. J. Bus. Inf.* **2012**, *7*, 105–136.
18. Deboeck, G. *Trading on the Edge: Neural, Genetic, and Fuzzy Systems for Chaotic Financial Markets*; Wiley: New York, NY, USA, 1994.
19. Zhu, Y.; Zhou, G. Technical analysis: An asset allocation perspective on the use of moving averages. *J. Financ. Econ.* **2009**, *92*, 519–544. [CrossRef]
20. Peachavanish, R. Stock selection and trading based on cluster analysis of trend and momentum indicators. *Lect. Notes Eng. Comput. Sci.* **2016**, *1*, 317–321.
21. Hulbert, M. Viewpoint: More Proof for the Dow Theory. Available online: <https://www.nytimes.com/1998/09/06/business/viewpoint-more-proof-for-the-dow-theory.html> (accessed on 17 October 2021).
22. Rahman, A.S.A.; Abdul-Rahman, S.; Mutalib, S. Mining Textual Terms for Stock Market Prediction Analysis Using Financial News. In *International Conference on Soft Computing in Data Science*; Springer: Singapore, 2017; pp. 293–305. [CrossRef]
23. Ballings, M.; Poel, D.V.D.; Hespels, N.; Gryp, R. Evaluating multiple classifiers for stock price direction prediction. *Expert Syst. Appl.* **2015**, *42*, 7046–7056. [CrossRef]
24. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
25. Srivastava, D.K.; Bhambhu, L. Data classification using support vector machine. *J. Theor. Appl. Inf. Technol.* **2010**, *12*, 1–7.
26. Venugopal, K.R.; Srinivasa, K.G.; Patnaik, L.M. Fuzzy based neuro—Genetic algorithm for stock market prediction. *Stud. Comput. Intell.* **2009**, *190*, 139–166.
27. Number of Social Network Users Worldwide from 2017 to 2025. Available online: <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/> (accessed on 30 May 2021).
28. Ding, X.; Zhang, Y.; Liu, T.; Duan, J. Using Structured Events to Predict Stock Price Movement: An Empirical Investigation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, Doha, Qatar, 25–29 October 2014; pp. 1415–1425.
29. Howells, K.; Ertugan, A. Applying fuzzy logic for sentiment analysis of social media network data in marketing. *Procedia Comput. Sci.* **2017**, *120*, 664–670. [CrossRef]
30. Liu, B. *Sentiment Analysis and Opinion Mining*; Morgan & Claypool Publishers: San Rafael, CA, USA, 2012; p. 167.
31. Li, W. Improvement of Stochastic Competitive Learning for Social Network. *Comput. Mater. Contin.* **2020**, *63*, 755–768.
32. Devi, K.N.; Bhaskaran, V.M. Semantic Enhanced Social Media Sentiments for Stock Market Prediction. *Int. J. Econ. Manag. Eng.* **2015**, *9*, 684–688.
33. Hill, S.; Ready-Campbell, N. Expert Stock Picker: The Wisdom of (Experts in) Crowds. *Int. J. Electron. Commer.* **2011**, *15*, 73–102. [CrossRef]
34. Chen, T.-L.; Chen, F.-Y. An intelligent pattern recognition model for supporting investment decisions in stock market. *Inf. Sci.* **2016**, *346–347*, 261–274. [CrossRef]
35. Ranco, G.; Aleksovski, D.; Caldarelli, G.; Grčar, M.; Mozetic, I. The effects of Twitter sentiment on stock price returns. *PLoS ONE* **2015**, *10*, e0138441. [CrossRef] [PubMed]
36. Bhardwaj, A.; Narayan, Y.; Dutta, M. Sentiment Analysis for Indian Stock Market Prediction Using Sensex and Nifty. *Procedia Comput. Sci.* **2015**, *70*, 85–91. [CrossRef]
37. Zhang, Y.; Wu, L. Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. *Expert Syst. Appl.* **2009**, *36*, 8849–8854.
38. Guresen, E.; Kayakutlu, G.; Daim, T.U. Using artificial neural network models in stock market index prediction. *Expert Syst. Appl.* **2011**, *38*, 10389–10397. [CrossRef]
39. Lugmayr, A.; Gossen, G. Evaluation of methods and techniques for language based sentiment analysis for dax 30 stock exchange—A first concept of a ‘LUGO’ sentiment indicator. In *Proceedings of the 5th International Workshop on Semantic Ambient Media Experience (SAME)*, Newcastle, UK, 18 June 2012; pp. 69–76.
40. Porshnev, A.; Redkin, I.; Karpov, N. Modelling Movement of Stock Market Indexes with Data from Emoticons of Twitter Users. *Commun. Comput. Inf. Sci.* **2015**, *205*, 297–306. [CrossRef]
41. Nti, I.K.; Adekoya, A.F.; Weyori, B.A. A comprehensive evaluation of ensemble learning for stock-market prediction. *J. Big Data* **2020**, *7*, 1–40. [CrossRef]
42. Weng, B.; Ahmed, M.A.; Megahed, F. Stock market one-day ahead movement prediction using disparate data sources. *Expert Syst. Appl.* **2017**, *79*, 153–163. [CrossRef]
43. Di Persio, L.; Honchar, O. Recurrent neural networks approach to the financial forecast of Google assets. *Int. J. Math. Comput. Simul.* **2017**, *11*, 7–13.
44. Hagenau, M.; Liebmann, M.; Hedwig, M.; Neumann, D. Automated News Reading: Stock Price Prediction Based on Financial News Using Context-Specific Features. In *Proceedings of the 2012 45th Hawaii International Conference on System Sciences*, Maui, HI, USA, 4–9 January 2012; pp. 1040–1049.
45. Huang, C.-F.; Li, H.-C. An Evolutionary Method for Financial Forecasting in Microscopic High-Speed Trading Environment. *Comput. Intell. Neurosci.* **2017**, *2017*, 9580815. [CrossRef] [PubMed]
46. Shah, D.; Isah, H.; Zulkernine, F. Stock Market Analysis: A Review and Taxonomy of Prediction Techniques. *Int. J. Financ. Stud.* **2019**, *7*, 26. [CrossRef]
47. Nayak, A.; Pai, M.M.M.; Pai, R.M. Prediction Models for Indian Stock Market. *Procedia Comput. Sci.* **2016**, *89*, 441–449. [CrossRef]

48. Makrehchi, M.; Shah, S.; Liao, W. Stock Prediction Using Event-Based Sentiment Analysis. In Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), Atlanta, GA, USA, 17–20 November 2013; Volume 1, pp. 337–342.
49. Ghanavati, M.; Wong, R.K.; Chen, F.; Wang, Y.; Fong, S. A Generic Service Framework for Stock Market Prediction. In Proceedings of the 2016 IEEE International Conference on Services Computing (SCC), San Francisco, CA, USA, 27 June–2 July 2016; pp. 283–290.
50. Bustos, O.; Pomares, A.; Gonzalez, E. A comparison between SVM and multilayer perceptron in predicting an emerging financial market: Colombian stock market. In Proceedings of the 2017 Congreso Internacional de Innovación y Tendencias en Ingeniería (CONITI), Bogota, Colombia, 4–6 October 2017; pp. 1–6.
51. Dey, S.; Kumar, Y.; Saha, S.; Basak, S. Forecasting to Classification: Predicting the direction of stock market price using Xtreme Gradient Boosting Forecasting to Classification: Predicting the direction of stock market price using Xtreme Gradient Boosting. *PESIT South Campus* **2016**. [[CrossRef](#)]
52. Murshed, B.A.H.; Al-Ariki, H.D.E.; Mallappa, S. Semantic Analysis Techniques using Twitter Datasets on Big Data: Comparative Analysis Study. *Comput. Syst. Sci. Eng.* **2020**, *35*, 495–512. [[CrossRef](#)]
53. Xie, B.; Passonneau, R.; Wu, L.; Creamer, G.G. Semantic Frames to Predict Stock Price Movement. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria, 4–9 August 2013; pp. 873–883.
54. Ding, X.; Zhang, Y.; Liu, T.; Duan, J. Knowledge-Driven Event Embedding for Stock Prediction. In Proceedings of the COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, Osaka, Japan, 11–17 December 2016; pp. 2133–2142.
55. Sirimevan, N.; Mamalgaha, I.G.U.H.; Jayasekara, C.; Mayuran, Y.S.; Jayawardena, C. Stock Market Prediction Using Machine Learning Techniques. In Proceedings of the IEEE 2019 International Conference on Advancements in Computing (ICAC), Malabe, Sri Lanka, 5–7 December 2019; Volume 1, pp. 192–197.
56. Schumaker, R.P.; Zhang, Y.; Huang, C.-N.; Chen, H. Evaluating sentiment in financial news articles. *Decis. Support Syst.* **2012**, *53*, 458–464. [[CrossRef](#)]
57. Huang, C.-J.; Liao, J.-J.; Yang, D.-X.; Chang, T.-Y.; Luo, Y.-C. Realization of a news dissemination agent based on weighted association rules and text mining techniques. *Expert Syst. Appl.* **2010**, *37*, 6409–6413. [[CrossRef](#)]
58. Nguyen, T.H.; Shirai, K.; Velcin, J. Sentiment analysis on social media for stock movement prediction. *Expert Syst. Appl.* **2015**, *42*, 9603–9611. [[CrossRef](#)]
59. Rajput, V.; Bobde, S. Stock market prediction using hybrid approach. In Proceedings of the 2016 International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, India, 29–30 April 2016; pp. 82–86.
60. Asghar, M.Z.; Subhan, F.; Imran, M.; Kundi, F.M.; Khan, A.; Shamshirband, S.; Mosavi, A.; Koczy, A.R.V.; Csiba, P. Performance Evaluation of Supervised Machine Learning Techniques for Efficient Detection of Emotions from Online Content. *Comput. Mater. Contin.* **2020**, *63*, 1093–1118. [[CrossRef](#)]
61. Akhtar, M.J.; Ahmad, Z.; Amin, R.; Almotiri, S.H.; Al Ghamdi, M.A.; Aldabbas, H. An Efficient Mechanism for Product Data Extraction from E-Commerce Websites. *Comput. Mater. Contin.* **2020**, *65*, 2639–2663. [[CrossRef](#)]
62. Pandarachalil, R.; Sendhilkumar, S.; Mahalakshmi, G.S. Twitter Sentiment Analysis for Large-Scale Data: An Unsupervised Approach. *Cogn. Comput.* **2014**, *7*, 254–262. [[CrossRef](#)]
63. Pagolu, V.S.; Reddy, K.N.; Panda, G.; Majhi, B. Sentiment analysis of Twitter data for predicting stock market movements. In Proceedings of the 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPEs), Paralakhemundi, India, 3–5 October 2016; pp. 1345–1350.
64. Mittal, A.; Goel, A. *Stock Prediction Using Twitter Sentiment Analysis*; Stanford University: Stanford, CA, USA, 2009; Volume 1, pp. 337–342.
65. Zhang, X.; Fuehres, H.; Gloor, P.A. Gloor, Predicting Stock Market Indicators Through Twitter “I hope it is not as bad as I fear”. *Procedia-Soc. Behav. Sci.* **2011**, *26*, 55–62. [[CrossRef](#)]
66. Uysal, A.; Gunal, S. The impact of preprocessing on text classification. *Inf. Process. Manag.* **2014**, *50*, 104–112. [[CrossRef](#)]
67. Wang, J.; Wang, X.; Yang, Y.; Zhang, H.; Fang, B. A Review of Data Cleaning Methods for Web Information System. *Comput. Mater. Contin.* **2020**, *62*, 1053–1075. [[CrossRef](#)]
68. Zhong, X.; Enke, D. Forecasting daily stock market return using dimensionality reduction. *Expert Syst. Appl.* **2017**, *67*, 126–139. [[CrossRef](#)]
69. Ihlaiyyel, H.A.; Sharef, N.M.; Nazri, M.Z.A.; Abu Bakar, A. An enhanced feature representation based on linear regression model for stock market prediction. *Intell. Data Anal.* **2018**, *22*, 45–76. [[CrossRef](#)]
70. Zhou, P.-Y.; Chan, K.C.C.; Ou, C.X. Corporate Communication Network and Stock Price Movements: Insights from Data Mining. *IEEE Trans. Comput. Soc. Syst.* **2018**, *5*, 391–402. [[CrossRef](#)]
71. Chandar, S.K. Fusion model of wavelet transform and adaptive neuro fuzzy inference system for stock market prediction. *J. Ambient. Intell. Humaniz. Comput.* **2019**, 1–9. [[CrossRef](#)]
72. Sedighi, M.; Jahangirnia, H.; Gharakhani, M.; Fard, S.F. A Novel Hybrid Model for Stock Price Forecasting Based on Metaheuristics and Support Vector Machine. *Data* **2019**, *4*, 75. [[CrossRef](#)]
73. Khan, W.; Malik, U.; Ghazanfar, M.A.; Azam, M.A.; Alyoubi, K.H.; Alfakeeh, A. Predicting stock market trends using machine learning algorithms via public sentiment and political situation analysis. *Soft Comput.* **2019**, *24*, 11019–11043. [[CrossRef](#)]

74. Ullah, K.; Qasim, M. Google Stock Prices Prediction Using Deep Learning. In Proceedings of the 2020 IEEE 10th International Conference on System Engineering and Technology (ICSET), Shah Alam, Malaysia, 9 November 2020; pp. 108–113.
75. Nassirtoussi, A.K.; Aghabozorgi, S.; Wah, T.Y.; Ngo, D.C.L. Text mining of news-headlines for FOREX market prediction: A Multi-layer Dimension Reduction Algorithm with semantics and sentiment. *Expert Syst. Appl.* **2015**, *42*, 306–324. [\[CrossRef\]](#)
76. Wu, D.D.; Olson, D.L. *Enterprise Risk Management in Finance*; Palgrave Macmillan: London, UK, 2015.
77. Chen, M.-Y.; Liao, C.-H.; Hsieh, R.-P. Modeling public mood and emotion: Stock market trend prediction with anticipatory computing approach. *Comput. Hum. Behav.* **2019**, *101*, 402–408. [\[CrossRef\]](#)
78. Das, S.R.; Mishra, D.; Rout, M. Stock market prediction using Firefly algorithm with evolutionary framework optimized feature reduction for OSELM method. *Expert Syst. Appl.* **2019**, *4*, 100016. [\[CrossRef\]](#)
79. Bouktif, S.; Fiaz, A.; Awad, M. Augmented Textual Features-Based Stock Market Prediction. *IEEE Access* **2020**, *8*, 40269–40282. [\[CrossRef\]](#)
80. Chen, S.; Zhou, C. Stock Prediction Based on Genetic Algorithm Feature Selection and Long Short-Term Memory Neural Network. *IEEE Access* **2020**, *9*, 9066–9072. [\[CrossRef\]](#)
81. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*; The MIT Press: Cambridge, MA, USA, 2013; pp. 3111–3119.
82. Si, J.; Mukherjee, A.; Liu, B.; Li, Q.; Li, H.; Deng, X. Exploiting Topic based Twitter Sentiment for Stock Prediction. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria, 4–9 August 2013; pp. 24–29.
83. Zhang, X.; Zhang, Y.; Wang, S.; Yao, Y.; Fang, B.; Yu, P.S. Improving stock market prediction via heterogeneous information fusion. *Knowl.-Based Syst.* **2018**, *143*, 236–247. [\[CrossRef\]](#)
84. Zhou, Z.; Qin, J.; Xiang, X.; Tan, Y.; Liu, Q.; Xiong, N.N. News Text Topic Clustering Optimized Method Based on TF-IDF Algorithm on Spark. *Comput. Mater. Contin.* **2020**, *62*, 217–231. [\[CrossRef\]](#)
85. El Seidy, E.; Ibrahim, B.; Jamous, R.A.; Bayoum, B.I. A Novel Efficient Forecasting of Stock Market Using Particle Swarm Optimization with Center of Mass Based Technique. *Int. J. Adv. Comput. Sci. Appl.* **2016**, *7*. [\[CrossRef\]](#)
86. He, S.; Li, Z.; Tang, Y.; Liao, Z.; Li, F.; Lim, S.-J. Parameters Compressing in Deep Learning. *Comput. Mater. Contin.* **2020**, *62*, 321–336. [\[CrossRef\]](#)
87. Pestov, V. Is the k-NN classifier in high dimensions affected by the curse of dimensionality? *Comput. Math. Appl.* **2013**, *65*, 1427–1437. [\[CrossRef\]](#)
88. Kalra, V.; Aggarwal, R. Importance of Text Data Preprocessing & Implementation in RapidMiner. *Proc. First Int. Conf. Inf. Technol. Knowl. Manag.* **2018**, *14*, 71–75. [\[CrossRef\]](#)
89. Zhang, C.; Cheng, J.; Tang, X.; Sheng, V.S.; Dong, Z.; Li, J. Novel DDoS Feature Representation Model Combining Deep Belief Network and Canonical Correlation Analysis. *Comput. Mater. Contin.* **2019**, *61*, 657–675. [\[CrossRef\]](#)
90. Sharma, S.; Ahmed, S.; Naseem, M.; Alnumay, W.S.; Singh, S.; Cho, G.H. A Survey on Applications of Artificial Intelligence for Pre-Parametric Project Cost and Soil Shear-Strength Estimation in Construction and Geotechnical Engineering. *Sensors* **2021**, *21*, 463. [\[CrossRef\]](#) [\[PubMed\]](#)
91. Ganser, A.; Hollaus, B.; Stabinger, S. Classification of Tennis Shots with a Neural Network Approach. *Sensors* **2021**, *21*, 5703. [\[CrossRef\]](#)
92. Ticknor, J.L. A Bayesian regularized artificial neural network for stock market forecasting. *Expert Syst. Appl.* **2013**, *40*, 5501–5506. [\[CrossRef\]](#)
93. Adebisi, A.A.; Adewumi, A.; Ayo, C.K. Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction. *J. Appl. Math.* **2014**, *2014*, 614342. [\[CrossRef\]](#)
94. Chopra, S.; Yadav, D.; Chopra, A.N. Artificial Neural Networks Based Indian Stock Market Price Prediction: Before and After Demonetization. *Int. J. Swarm Intell. Evol. Comput.* **2019**, *8*, 174.
95. Seo, M.; Kim, G. Hybrid Forecasting Models Based on the Neural Networks for the Volatility of Bitcoin. *Appl. Sci.* **2020**, *10*, 4768. [\[CrossRef\]](#)
96. Vanstone, B.; Finnie, G.; Hahn, T. Creating trading systems with fundamental variables and neural networks: The Aby case study. *Math. Comput. Simul.* **2012**, *86*, 78–91. [\[CrossRef\]](#)
97. Khashei, M.; Hajirahimi, Z. Performance evaluation of series and parallel strategies for financial time series forecasting. *Financ. Innov.* **2017**, *3*, 24. [\[CrossRef\]](#)
98. Goh, A. Back-propagation neural networks for modeling complex systems. *Artif. Intell. Eng.* **1995**, *9*, 143–151. [\[CrossRef\]](#)
99. Zhang, F.; Li, J.; Wang, Y.; Guo, L.; Wu, D.; Wu, H.; Zhao, H. Ensemble Learning Based on Policy Optimization Neural Networks for Capability Assessment. *Sensors* **2021**, *21*, 5802. [\[CrossRef\]](#) [\[PubMed\]](#)
100. Bing, Y.; Hao, J.K.; Zhang, S.C. Stock Market Prediction Using Artificial Neural Networks. *Adv. Eng. Forum* **2012**, *6*, 1055–1060. [\[CrossRef\]](#)
101. Hota, H.S.; Handa, R.; Shrivastava, A.K. Time Series Data Prediction Using Sliding Window Based RBF Neural Network. *Int. J. Comput. Intell. Res.* **2017**, *13*, 1145–1156.
102. Guo, Z.; Ye, W.; Yang, J.; Zeng, Y. Financial index time series prediction based on bidirectional two dimensional locality preserving projection. In Proceedings of the 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), Beijing, China, 10–12 March 2017; pp. 934–938.
103. Milosevic, N. Equity forecast: Predicting long term stock price movement using machine learning. *arXiv* **2016**, arXiv:1603.00751.

104. Li, X.; Xie, H.; Wang, R.; Cai, Y.; Cao, J.; Wang, F.; Min, H.; Deng, X. Empirical analysis: Stock market prediction via extreme learning machine. *Neural Comput. Appl.* **2014**, *27*, 67–78. [\[CrossRef\]](#)
105. More, A.M.; Rathod, P.U.; Patil, R.H.; Sarode, D.R.; Student, B. Stock Market Prediction System using Hadoop. *Int. J. Eng. Sci. Comput.* **2018**, *8*, 16138–16140.
106. Mohan, S.; Mullapudi, S.; Sammeta, S.; Vijayvergia, P.; Anastasiu, D.C. Stock Price Prediction Using News Sentiment Analysis. In Proceedings of the 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService), Newark, CA, USA, 4–9 April 2019; pp. 205–208.
107. Xianya, J.; Mo, H.; Haifeng, L. Stock Classification Prediction Based on Spark. *Procedia Comput. Sci.* **2019**, *162*, 243–250. [\[CrossRef\]](#)
108. Yu, Y.; Duan, W.; Cao, Q. The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decis. Support Syst.* **2013**, *55*, 919–926. [\[CrossRef\]](#)
109. Lin, L.; Cao, L.; Wang, J.; Zhang, C. The applications of genetic algorithms in stock market data mining optimization. *Manag. Inf. Syst.* **2004**, *10*, 273–280.
110. Pimenta, A.; Nametala, C.; Guimarães, F.G.; Carrano, E.G. An Automated Investing Method for Stock Market Based on Multiobjective Genetic Programming. *Comput. Econ.* **2017**, *52*, 125–144. [\[CrossRef\]](#)
111. Strader, T.J.; Rozycki, J.J.; Root, T.H.; Huang, Y.H.J. Machine Learning Stock Market Prediction Studies: Review and Research Directions. *J. Int. Technol. Inf. Manag.* **2020**, *28*, 63–83.
112. Kim, Y.; Ahn, W.; Oh, K.J.; Enke, D. An intelligent hybrid trading system for discovering trading rules for the futures market using rough sets and genetic algorithms. *Appl. Soft Comput.* **2017**, *55*, 127–140. [\[CrossRef\]](#)
113. Nair, B.B.; Dharini, N.M.; Mohandas, V. A Stock Market Trend Prediction System Using a Hybrid Decision Tree-Neuro-Fuzzy System. In Proceedings of the 2010 International Conference on Advances in Recent Technologies in Communication and Computing, Kottayam, India, 16–17 October 2010; pp. 381–385.
114. Chandar, S.K. Stock market prediction using subtractive clustering for a neuro fuzzy hybrid approach. *Clust. Comput.* **2017**, *22*, 13159–13166. [\[CrossRef\]](#)
115. Chang, P.-C.; Wu, J.-L.; Lin, J.-J. A Takagi-Sugeno fuzzy model combined with a support vector regression for stock trading forecasting. *Appl. Soft Comput.* **2016**, *38*, 831–842. [\[CrossRef\]](#)
116. Yolcu, O.C.; Lam, H.-K. A combined robust fuzzy time series method for prediction of time series. *Neurocomputing* **2017**, *247*, 87–101. [\[CrossRef\]](#)
117. Rajab, S.; Sharma, V. An interpretable neuro-fuzzy approach to stock price forecasting. *Soft Comput.* **2019**, *23*, 921–936. [\[CrossRef\]](#)
118. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#) [\[PubMed\]](#)
119. Wu, H.; Liu, Y.; Wang, J. Review of Text Classification Methods on Deep Learning. *Comput. Mater. Contin.* **2020**, *63*, 1309–1321. [\[CrossRef\]](#)
120. Hoseinzade, E.; Haratizadeh, S. CNNpred: CNN-based stock market prediction using a diverse set of variables. *Expert Syst. Appl.* **2019**, *129*, 273–285. [\[CrossRef\]](#)
121. Gao, P.; Zhang, R.; Yang, X. The Application of Stock Index Price Prediction with Neural Network. *Math. Comput. Appl.* **2020**, *25*, 53. [\[CrossRef\]](#)
122. Sezer, O.; Ozbayoglu, A. Financial Trading Model with Stock Bar Chart Image Time Series with Deep Convolutional Neural Networks. *Intell. Autom. Soft Comput.* **2018**. [\[CrossRef\]](#)
123. Pang, X.; Zhou, Y.; Wang, P.; Lin, W.; Chang, V. An innovative neural network approach for stock market prediction. *J. Supercomput.* **2018**, *76*, 2098–2118. [\[CrossRef\]](#)
124. Shah, D.; Campbell, W.; Zulkernine, F.H. A Comparative Study of LSTM and DNN for Stock Market Forecasting. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 4148–4155.
125. Li, X.; Yang, L.; Xue, F.; Zhou, H. Time series prediction of stock price using deep belief networks with intrinsic plasticity. In Proceedings of the 2017 29th Chinese Control and Decision Conference (CCDC), Chongqing, China, 28–30 May 2017; pp. 1237–1242.
126. Zhang, J.; Teng, Y.-F.; Chen, W. Support vector regression with modified firefly algorithm for stock price forecasting. *Appl. Intell.* **2018**, *49*, 1658–1674. [\[CrossRef\]](#)
127. Zheng, J.; Fu, X.; Zhang, G. Research on exchange rate forecasting based on deep belief network. *Neural Comput. Appl.* **2017**, *31*, 573–582. [\[CrossRef\]](#)
128. Hushani, P. Using Autoregressive Modelling and Machine Learning for Stock Market Prediction and Trading. In *Third International Congress on Information and Communication Technology*; Springer: Singapore, 2018; pp. 767–774. [\[CrossRef\]](#)
129. Nguyen, D.H.D.; Tran, L.P.; Nguyen, V. Predicting Stock Prices Using Dynamic LSTM Models. *Int. Conf. Appl. Inform.* **2019**, *6*, 199–212. [\[CrossRef\]](#)
130. Zhang, L.; Zhang, L.; Teng, W.; Chen, Y. Based on Information Fusion Technique with Data Mining in the Application of Finance Early-Warning. *Procedia Comput. Sci.* **2013**, *17*, 695–703. [\[CrossRef\]](#)
131. Cakra, Y.E.; Trisedya, B.D. Stock price prediction using linear regression based on sentiment analysis. In Proceedings of the 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Depok, Indonesia, 10–11 October 2015; pp. 147–154.

132. Bhuriya, D.; Kaushal, G.; Sharma, A.; Singh, U. Stock market predication using a linear regression. In Proceedings of the 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 20–22 April 2017; pp. 510–513.
133. Gururaj, V.; Shriya, V.R.; Ashwini, K. Stock market prediction using linear regression and support vector machines. *Int. J. Appl. Eng. Res.* **2019**, *14*, 1931–1934.
134. Enke, D.; Grauer, M.; Mehdiyev, N. Stock market prediction with Multiple Regression, Fuzzy type-2 clustering and neural networks. *Procedia Comput. Sci.* **2011**, *6*, 201–206. [\[CrossRef\]](#)
135. Kamley, S.; Jaloree, S.; Thakur, R. Multiple Regression: A Data Mining Approach for Predicting the Stock Market Trends Based on Open, Close and High Price of The Month. *Int. J. Comput. Sci. Eng. Inf. Technol. Res.* **2013**, *3*, 173–180.
136. Yuan, J.; Luo, Y. Test on the Validity of Futures Market's High Frequency Volume and Price on Forecast. In Proceedings of the 2014 International Conference on Management of e-Commerce and e-Government, Shanghai, China, 31 October–2 November 2014; pp. 28–32.
137. Imran, K. Prediction of stock performance by using logistic regression model: Evidence from Pakistan Stock Exchange (PSX). *AJER* **2018**, *8*, 247–258. [\[CrossRef\]](#)
138. Meesad, P.; Rasel, R.I. Predicting stock market price using support vector regression. In Proceedings of the 2013 International Conference on Informatics, Electronics and Vision (ICIEV), Dhaka, Bangladesh, 17–18 May 2013.
139. Siew, H.L.; Nordin, M.J. Regression techniques for the prediction of stock price trend. In Proceedings of the 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE), Langkawi, Malaysia, 10–12 September 2012; pp. 99–103.
140. Ananthi, M.; Vijayakumar, K. Stock market analysis using candlestick regression and market trend prediction (CKRM). *J. Ambient. Intell. Humaniz. Comput.* **2020**, *12*, 4819–4826. [\[CrossRef\]](#)
141. Cheng, C.; Xu, W.; Wang, J. A Comparison of Ensemble Methods in Financial Market Prediction. In Proceedings of the 2012 Fifth International Joint Conference on Computational Sciences and Optimization, Harbin, China, 23–26 June 2012; pp. 755–759.
142. 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. [\[CrossRef\]](#)
143. Nair, B.B.; Mohandas, V.P.; Nayanar, N.; Teja, E.S.R.; Vigneshwari, S.; Teja, K.V.N.S. A Stock Trading Recommender System Based on Temporal Association Rule Mining. *SAGE Open* **2015**, *5*, 2158244015579941. [\[CrossRef\]](#)
144. Hu, H.; Tang, L.; Zhang, S.; Wang, H. Predicting the direction of stock markets using optimized neural networks with Google Trends. *Neurocomputing* **2018**, *285*, 188–195. [\[CrossRef\]](#)
145. Rather, A.M. A Hybrid Intelligent Method of Predicting Stock Returns. *Adv. Artif. Neural Syst.* **2014**, *2014*, 1–7. [\[CrossRef\]](#)
146. Kalaivaani, P.C.D.; Thangarajan, R. Enhancing the Classification Accuracy in Sentiment Analysis with Computational Intelligence Using Joint Sentiment Topic Detection with MEDLDA. *Intell. Autom. Soft Comput.* **2020**, *26*, 71–79. [\[CrossRef\]](#)
147. Hossin, M.; Sulaiman, M.N. A Review on Evaluation Metrics for Data Classification Evaluations. *Int. J. Data Min. Knowl. Manag. Process.* **2015**, *5*, 1–11. [\[CrossRef\]](#)
148. Huang, J.; Ling, C. Using AUC and accuracy in evaluating learning algorithms. *IEEE Trans. Knowl. Data Eng.* **2005**, *17*, 299–310. [\[CrossRef\]](#)
149. de Oliveira, F.A.; Nobre, C.N.; Zárate, L.E. Applying Artificial Neural Networks to prediction of stock price and improvement of the directional prediction index—Case study of PETR4, Petrobras, Brazil. *Expert Syst. Appl.* **2013**, *40*, 7596–7606. [\[CrossRef\]](#)
150. Khan, W.; Ghazanfar, M.A.; Azam, M.A.; Karami, A.; Aloubi, K.H.; Alfakeeh, A.S. Stock market prediction using machine learning classifiers and social media, news. *J. Ambient. Intell. Humaniz. Comput.* **2020**, 1–24. [\[CrossRef\]](#)
151. Bergmeir, C.; Hyndman, R.; Koo, B. A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Comput. Stat. Data Anal.* **2018**, *120*, 70–83. [\[CrossRef\]](#)
152. Ghiassi, M.; Saidane, H.; Zimbra, D. A dynamic artificial neural network model for forecasting time series events. *Int. J. Forecast.* **2005**, *21*, 341–362. [\[CrossRef\]](#)
153. Binkowski, M.; Marti, G.; Donnat, P. Autoregressive convolutional neural networks for asynchronous time series. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; Volume 2, pp. 933–945.
154. Baek, Y.; Kim, H.Y. ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module. *Expert Syst. Appl.* **2018**, *113*, 457–480. [\[CrossRef\]](#)
155. Zheng, H.; Zhou, Z.; Chen, J. RLSTM: A New Framework of Stock Prediction by Using Random Noise for Overfitting Prevention. *Comput. Intell. Neurosci.* **2021**, *2021*, 8865816. [\[CrossRef\]](#)
156. Robles-Granda, P.D.; Belik, I.V. A Comparison of Machine Learning Classifiers Applied to Financial Datasets. In Proceedings of the World Congress on Engineering and Computer Science 2010, San Francisco, CA, USA, 20–22 October 2010.
157. Popper, N. Knight Capital Says Trading Glitch Cost It \$440 Million—The New York Times. Available online: <https://dealbook.nytimes.com/2012/08/02/knight-capital-says-trading-mishap-cost-it-440-million/> (accessed on 20 July 2020).
158. Phillips, M. Nasdaq: Here's Our Timeline of the Flash Crash. *Wall Str. J.* **2010**. Available online: <https://www.wsj.com/articles/BL-MB-21942> (accessed on 30 May 2021).
159. Gul, S.; Khan, M.T.; Saif, N.; Rehman, S.U.; Roohullah, S. Stock Market Reaction to Political Events (Evidence from Pakistan). *J. Econ. Sustain. Dev.* **2013**, *4*, 165–175.
160. Suriani, N.S.; Hussain, A.; Zulkifley, M.A. Sudden Event Recognition: A Survey. *Sensors* **2013**, *13*, 9966–9998. [\[CrossRef\]](#) [\[PubMed\]](#)

-
161. Huang, S.Y.B.; Lee, C.-J.; Lee, S.-C. Toward a Unified Theory of Customer Continuance Model for Financial Technology Chatbots. *Sensors* **2021**, *21*, 5687. [[CrossRef](#)] [[PubMed](#)]
 162. Ferrara, E.; Varol, O.; Davis, C.; Menczer, F.; Flammini, A. The rise of social bots. *Commun. ACM* **2016**, *59*, 96–104. [[CrossRef](#)]
 163. Kalyanaraman, V.; Kazi, S.; Tondulkar, R.; Oswal, S. Sentiment Analysis on News Articles for Stocks. In Proceedings of the 2014 8th Asia Modelling Symposium, Taipei, Taiwan, 23–25 October 2014; pp. 10–15.
 164. Seethalakshmi, R. Analysis of Stock Market Predictor Variables using Linear Regression Analysis. *Int. J. Pure Appl. Math.* **2020**, *119*, 369–378.