Artificial Intelligence Review (2020) 53:3007–3057
https://doi.org/10.1007/s10462-019-09754-z

A systematic review of fundamental and technical analysis of stock market predictions

Isaac Kofi Nti[1,2] · Adebayo Felix Adekoya[2] · Benjamin Asubam Weyori[2]

Published online: 20 August 2019

Abstract

The stock market is a key pivot in every growing and thriving economy, and every investment in the market is aimed at maximising profit and minimising associated risk. As a result, numerous studies have been conducted on the stock-market prediction using technical or fundamental analysis through various soft-computing techniques and algorithms. This study attempted to undertake a systematic and critical review of about one hundred and twenty-two (122) pertinent research works reported in academic journals over 11 years (2007–2018) in the area of stock market prediction using machine learning. The various techniques identified from these reports were clustered into three categories, namely technical, fundamental, and combined analyses. The grouping was done based on the following criteria: the nature of a dataset and the number of data sources used, the data timeframe, the machine learning algorithms used, machine learning task, used accuracy and error metrics and software packages used for modelling. The results revealed that 66% of documents reviewed were based on technical analysis; whiles 23% and 11% were based on fundamental analysis and combined analyses, respectively. Concerning the number of data source, 89.34% of documents reviewed, used single sources; whiles 8.2% and 2.46% used two and three sources respectively. Support vector machine and artificial neural network were found to be the most used machine learning algorithms for stock market prediction.

Keywords Machine-learning · Ensemble · Stock-prediction · Artificial intelligence · Technical-analysis · Fundamental-analysis

1 Introduction

The well-being of every growing economy, country or societies in this twenty first century mainly hinges on their market economies and stock-price, with the financial market being the pivot (Nassirtoussi et al. 2014; Göçken et al. 2016). Thus, it is essential and vital to study and learn about the financial market extensively. Due to a number of

✉ Isaac Kofi Nti
Ntious1@gmail.com

[1] Department of Computer Science, Sunyani Technical University, Sunyani, Ghana

[2] Department of Computer Science and Informatics, University of Energy and Natural Resources, Sunyani, Ghana

uncertainties (such as general economic conditions, social factors and political events at both homegrown and international levels), it is difficult to predict financial markets (Adebiyi et al. 2012, 2014b; Bisoi and Dash 2014; Ding et al. 2014; Rajashree et al. 2014; Rather et al. 2014; Lin 2018).

A stock market is a place for trading stocks (equity) and other financial instruments of public listed companies, where the price of shares is termed "share" or "stock price" (Wanjawa and Muchemi 2014). In reality, the stock price level of a firm, to a large extent reflects how it "cuts its pie" (Chan et al. 2017). Investments in the stock markets are often guided by some form of prediction (Wanjawa and Muchemi 2014; Ghaznavi et al. 2016). Three main approaches for stock market prediction, namely: fundamental analysis, technical analysis (charting) and technology (Machine learning) methods (Dunne 2015). Conversely, some scholars do categorise these three into 2, thus technical analysis and fundamental analysis (Nassirtoussi et al. 2014; Dunne 2015; Gyan 2015; Prem Sankar et al. 2015; Ahmadi et al. 2018).

The fundamental analysts approach concerned with the company that underlies the stock itself instead of the actual stock (Anbalagan and Maheswari 2014; Ghaznavi et al. 2016; Agarwal et al. 2017). The data used by the fundamental analyst usually are unstructured, which poses a difficult challenge. However, occasionally been proven to be a good predictor of stock price movement in the works of Zhang et al. (2011), Li et al. (2014a, b, d), Rather et al. (2014), Ballings et al. (2015), Liu et al. (2015), Sun et al. (2016), Kumar et al. (2016), Checkley et al. (2017), Tsai and Wang (2017), Zhang et al. (2017), Coyne et al. (2017), Pimprikar et al. (2017) and Sorto et al. (2017).

On the other hand, in the technical analysis, the analyst predicts the future price of stocks by studying the trends in the past and present stock price (Anbalagan and Maheswari 2014; Agarwal et al. 2017; Ahmadi et al. 2018). The following studies (Akinwale Adio et al. 2009; Guresen et al. 2011; Ju-Jie et al. 2012) and present (Rather et al. 2014; Laboissiere et al. 2015; Adebayo et al. 2017; Thanh et al. 2018; Umoru and Nwokoye 2018; Zhou et al. 2018) predicted future stock-price movement based on technical analysis. Figure 2 shows the general format for a technical analysis approach of stock market prediction.

Globally, billions of dollars are traded on the stock market daily, to make a profit (Dunne 2015). Thus, making stock-market prediction an attractive research area for researchers, investors and financial analysts, despite its difficulty (Ticknor 2013; Chen et al. 2014; Metghalchi et al. 2014; Rather et al. 2014; Prem Sankar et al. 2015; Wanjawa 2016; Agarwal et al. 2017; Tsai and Wang 2017; Lin 2018; Zhou et al. 2018). Hence, resulting in the application of machine learning and computational intelligence techniques in analysing the stock market trend. These include hidden Markov model, neural network, neuro-fuzzy inference system, genetic algorithm, time series analysis, regression, mining association rules, support vector machine (SVM), principal component analysis (PCA) and rough set theory among others (Chen et al. 2014; Lin 2018; Thanh et al. 2018).

This research seeks to perform a comprehensive systematic review of previous studies on stock market predictions based on the fundamental and technical analyst point of view, leading to the clarification of the current state-of-the-art and its possible future directions. Succinctly, this work contributes to the body of knowledge as summarised:

1. A well-organised review of pertinent literature with an emphasis on the different factors, which affects the movement of stock prices.
2. Acknowledgement of the distinguishing factors among existing works in the literature and the comparative analysis of the methods used in the predictive models.

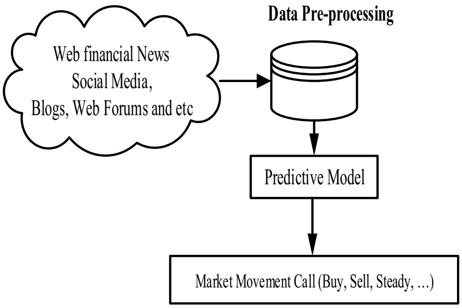

Fig. 1 Fundamental analysts approach

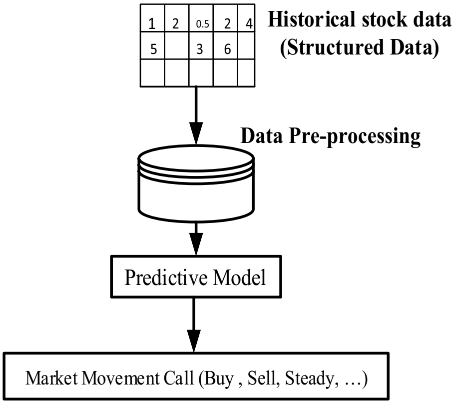

Fig. 2 Technical analysts approach

3. Established justifications for future research directions that could alleviate the deficiencies identified in the existing techniques and proffer solutions for improvement.

The remaining sections of this paper are organised as follows. Section 2 provides insight to markets predictability, machine-learning algorithms, technical and fundamental analysis, predictive model evaluation, and a review of related works. Section 3 presents the research design, research framework, and sample techniques. Section 4 presents the results and discussions of findings. Section 5 gives a summary of the findings. Besides, Sect. 6 concludes this work and outline the directions for future research.

2 The categorisation of stock market decision techniques

This section discussed in brief the fundamental and technical analysis as decision-making tools for stock market decisions. The fundamental and technical analysis approach to stock market prediction is, as shown in Figs. 1 and 2, respectively.

Technical analysis The technical analyst tries to predict the stock market through the learning of charts that portray the historical market-prices and technical indicators (Sureshkumar and Elango 2011; Wei et al. 2011; Suthar et al. 2012; de Oliveira et al. 2013; Ballings et al. 2015; Gaius 2015; Su and Cheng 2016). As shown in Fig. 2, the historical stock prices are preprocessed, and appropriate indicators are calculated and fed into the predictive model. Some of the technical indicators used in technical analysis discussed in Anbalagan and Maheswari (2014), Bisoi and Dash (2014) and Rajashree et al. (2014) are simple-moving average (SMA), exponential moving average (EMA), moving average convergence/divergence rules (MACD), relative-strength index (RSI) and on-balance-volume (OBV) as shown in Fig. 3.

SMA The SMA is ascertained by totalling the most recent closing prices of a stock and then dividing that by the number "n" of periods in the calculation average (Anbalagan and Maheswari 2014).

EMA The EMA is similar to the SMA line except the given day's EMA determination depends on the EMA calculations for all the days preceding that day.

$$EMA = Price(t) \times k + EMA(y) \times (1 - k) \tag{1}$$

where t and y represents today and yesterday respectively, N is the number of days in EMA and k (smoothing) = 2/(N + 1).

MACD The MACD indicator was a momentum indicator; it tries to predict stock market trends by a comparison between short and long-term trends. To ascertain MACD, find the difference between a 26-day EMA and a 12-day.

$$MACD = \sum_{i=1}^{n} EMA_k - \sum_{i=1}^{n} EMA_d \tag{2}$$

where k = 12 (number of days) and d = 26 reflect the number of days in EMA.

OBV The OBV indicator is also a momentum indicator that employs volume-flow to predict movements in stock price. An indication of a fall in stock price is a falling OBV line, whiles a future rising in stock price is indicated by growing OBV line.

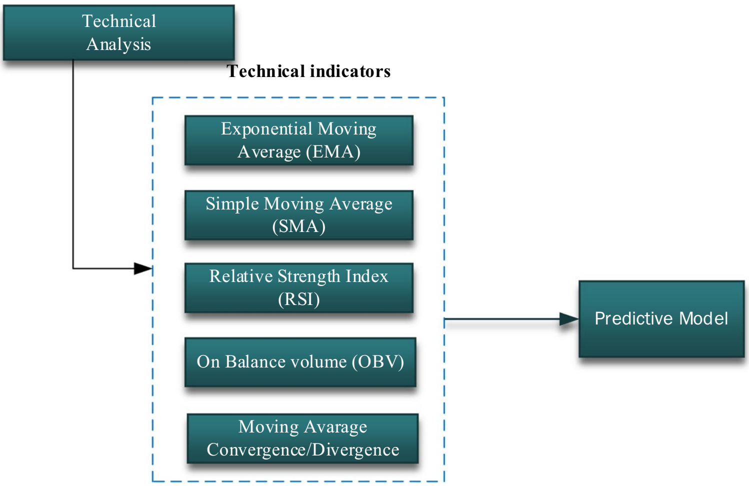

Fig. 3 Technical analysis

RSI This is an indicator that measures whether a stock bought is oversold or overboug. The following equation shows how is obtained.

$$RSI = 100 - \left(\frac{100}{1} + RS\right) \tag{3}$$

where RS = average gain/average loss

Conventionally most stock market prediction methods usually employ technical analysis techniques to predict future trends in stock values. However, Li et al. (2015) argue that quantified data cannot wholly convey the wide variety of firms' financial status. Hence, the quality of information concealed in traditional news and social network sites (unstructured data) can serve as complementary to quantitative data to enhance prediction models, specifically in this age of social media.

Fundamental analysis The fundamental analysis uses the economic standing of the firm, employees, the board of directors, financial status, firm's yearly report, balance-sheets, income-reports, terrestrial and climatic circumstances like unnatural or natural disasters and political data to predict future stock price (Tsai and Hsiao 2010; Anbalagan and Maheswari 2014; Ghaznavi et al. 2016; Agarwal et al. 2017). Due to the unstructured nature of fundamental factors, automation of fundamental analysis is difficult. On the other hand, the emergence of machine learning has enabled researchers to automate stock market prediction based on unstructured data, which in some cases has reported higher prediction accuracy. Nonetheless, fundamental analysis is useful for long-term stock-price movement, but not suitable for short-term stock-price change (Khan et al. 2011).

The fundamental analyst uses the openly accessible facts about the stock to perform analysis of stock price movement in three dimensions, concerning the economy, its industry, and the firm, as shown in Fig. 4. Again, the fundamental analyst also considers different financial ratios of the firm. Few of these important ratios discussed in Renu and Christie (2018) include:

Return on equity (ROE) This ratio offers an overview of how well the shareholder's funds were used and the gain made out of its investment. When ROE is low, it implies that the shareholder's funds were not used properly. The formula computes ROE:

$$ROE = \frac{PTP}{SE} \tag{4}$$

where PTP = post-tax profit and SE = shareholder equity

Debt/equity ratio (D/E) Reveals the power of the available capital as opposed to the capital engaged. A low value of D/E means the credit accessible was not used. D/E is computed as follows:

$$\frac{D}{E} = \frac{\text{Overall liabilities}}{\text{Shareholders}' \text{ equity}} \tag{5}$$

Market capitalization (MC) MC measures the total stocks transacted in the market. Concerning MC, stocks can be categorised into three groups, namely: small-cap, medium-cap, and large-cap. The formula can compute MC:

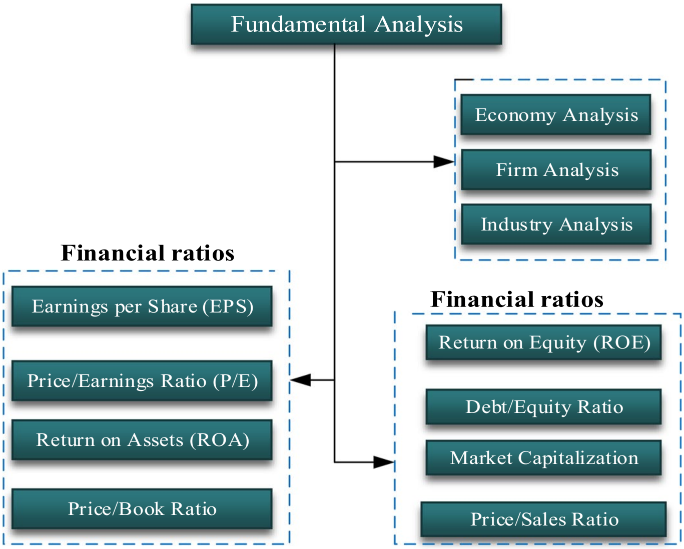

Fig. 4 Fundamental analysis

$$MC = \text{Total shares} \times \text{Price per share} \tag{6}$$

Price/sales ratio (P/S) This ratio ascertains if a share price of a stock depicts stock's value. The formula calculates P/S:

$$\frac{P}{S} = \frac{Share\ Price}{(\text{Returns over a 12 month time frame})} \tag{7}$$

Price/book ratio (P/B) Is a comparison of the stock's fundamental value with the share price. P/B is an indication of underestimate or overestimate of the stock. The formula computes P/B:

$$\frac{P}{B} = \frac{\text{Stock Price}}{(\text{Total assets} - \text{Intangible assets and liabilities})} \tag{8}$$

Earnings per share (EPS) EPS provides the profitability indication of a firm, and can be determined by dividing the firm's net income with its whole number of remaining stocks. EPS can be computed in two ways, as illustrated in Eqs. (9) and (10).

$$EPS = \frac{\text{Net Income after Tax}}{\text{Total Number of Outstanding Shares}} \tag{9}$$

$$\text{Weighted earnings per share} = \frac{(\text{Net Income after Tax} - \text{Total Dividends})}{\text{Total Number of Remaining Stocks}} \tag{10}$$

Price/earnings ratio (P/E) This ratio a very valuable evaluation metric for estimating the relative attractiveness of a firm's current stock price compared to the firm's per-share earnings. It is calculated as:

$$P/E = \frac{\text{Market value per share}}{EPS} \tag{11}$$

Return on assets (ROA) This ratio signifies the proportion of earnings a firm earns about the firm's overall assets or resources. Thus, an indication of how profitable a firm is relative to the firm's total resources or assets.

$$\text{ROA} = \frac{\text{Net Income}}{\text{Total Assets}} \tag{12}$$

When the cost of debt is ignored, then ROA is given by the formula:

$$\text{ROA} = \frac{(\text{Net Income} + \text{Interest Expense})}{\text{Average Total Assets}} \tag{13}$$

This method of stock market analysis has become common in recent years, with the introduction of text mining techniques. Many studies have used the fundamental analysis for stock prediction, but (Talib et al. 2016) in their work titled "Text Mining-Techniques Applications and Issues", they argue that quite several problems are associated with the text mining process which turns to affect the effectiveness and efficacy of decision making. Another critical issue is a multi-lingual text minor change dependency that creates problems (Henriksson et al. 2016) and only a hand-full of tools are available that has the support for multiple languages (Solanki 2013). Despite the increase in stock-market prediction from both technical and fundamental analysis point of view, some scholars (Fama 1965, 1970; Malkiel 1999) holds the belief that the stock market is unpredictable.

2.1 The unpredictability of the stock market

In Fama (1965, 1970) and Malkiel (1999), the authors holds a view that the stock market is stochastic, and hence, it is not predictable. This lead to the two famous hypotheses, namely, The random-walk hypothesis (RWH) and the efficient market hypothesis (EMH).

2.1.1 The random-walk hypothesis (RWH)

The Random-walk hypothesis reveals the unpleasant view of the predictability of the stock market. The assumption holds the belief that the stock price is fundamentally stochastic; hence, any initiative or effort to forecast or predict the future stock price will unavoidably fail (Dunne 2015). If indeed the market is stochastic, then there is a little chance of continuing.

2.1.2 The efficient market hypothesis (EMH)

The second hypothesis that the market is random, hence not predictable is the famous EMH by Fama (1965), which says the stock market is "informationally-efficient." It hypothesised that the market is efficient at discovering the correct price for the stock market. On the other hand, the credibility of this hypothesis is challengeable since the hypothesizer Fama revised it and categorised it into three levels of efficacy as Weak-form, semi-strong, and robust (Fama 1970). However, the EMH is susceptible to debate on which one, if any, is correct.

Carefully studying these two hypotheses, there is a chance to predict the stock market when one has fundamental and technical knowledge about the stock market. That is, knowing and understanding of the historical stock data and fundamental or financial data of a firm can lead to a successful prediction of the firm's future stock price.

2.2 Markets' predictability

Despite the stands of the EMH (Fama 1965) against stock-market forecast established on historical publicly accessible data and information, a considerable amount of research advocates that more or fewer markets, particularly markets emerging, are not entirely and thoroughly well-organized, and prediction of future stock-prices and stock-returns possibly will yield better outcomes than random selection (Zhang et al. 2014). Again, Chen et al. (2014) argues that the stock market is predictable to an extent when looking from behavioural economics and socioeconomic theory of finance viewpoint.

2.3 Machine learning

Machine learning is a branch of Artificial Intelligent (AI), and it is a learning process, that starts with the identification of the learning-domain and concludes with testing and employing the obtained results of learning in solving a problem (Perwej and Perwej 2012). Many machine learning algorithms have been developed and applied to stock market prediction (Dunne 2015; Paik and Kumari 2017).

2.4 The general overview of predictive models

Figure 5 shows the general overview of stock market predictive models, with fundamental data (unstructured data) or technical data (historical market data) serving, as input datasets and the output are some predictive market values.

2.5 Input data

In literature, it is revealed that, for one to make an effective economic prediction, it is essential to detect which variables help or contributes to predicting other economic variables (de Oliveira et al. 2013). Generally, financial data can be characterised by quantitative data (technical analysis) and qualitative reports of companies and investors sentiments (fundamental analysis) (Li et al. 2015).

2.5.1 Quantitative (structured) data

Historical stock price (HSP) Past stock price, which includes previous closing price, opening price, the current closing price, price change, closing bid price, volume, and closing offer price.

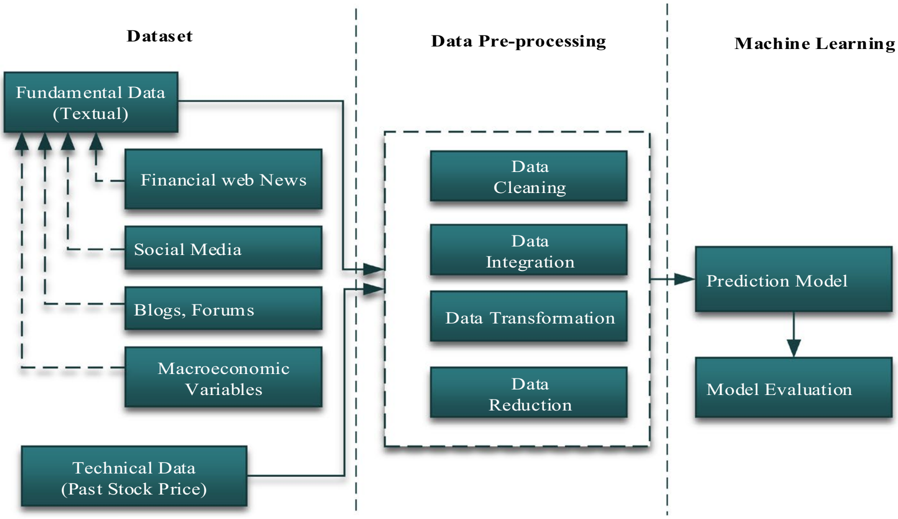

Fig. 5 General overview of predictive models

2.5.2 Qualitative (unstructured) data

These are textual information concerning the stock owners and can be categorised as, financial Web news data (FWN), social media sentiment data (SMS), and macroeconomic variables (MVs). Lahmiri argues that these indicators are better predictors of stock price movement than technical indicators (Lahmiri 2011).

2.5.3 Data pre-processing

The obtained dataset is preprocessed to remove noise, data inconsistency, and finally, feature selection data transformation and normalisation for better performance and accuracy (Uysal and Gunal 2014).

2.6 Model evaluation

Every prediction model needs evaluation to ascertain the accuracy of the model. Some of the most commonly used accuracy metrics in literature include: the mean absolute percentage error (MAPE), mean square error (MSE), mean absolute error (MAE) and root mean squared error (RMSE), which are defined in Javed et al. (2014), Rajashree et al. (2014) and Nayak et al. (2015) as follows.

1. The correlation coefficient (R): performance index unveils the degree of associations between predicted values and actual values, it ranges from 0 to 1, and the bigger the Correlation coefficient, the better model performance.

$$R = \frac{\sum_{i=1}^{n}\left(t_i - \bar{t}\right)\left(y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(t_i - \bar{t}\right)^2 . \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}} \tag{14}$$

where $\bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i \, and \, \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ are the average values of t_i and y_i, respectively

2. Root mean squared error (RMSE): This performance index will show an estimation of the residual between the actual (t_i) value and predicted value (y_i) as given in Eq. (15).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(t_i - y_i\right)} \tag{15}$$

where y_i is the predicted value produced by the model, t_i is the actual value and n = total number of testing data.

3. The next performance metric is the mean absolute percentage error (MAPE): this metric is an indicator of an average absolute percentage error; lower MAPE is better than higher MAPE.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{t_i - y_i}{t_i}\right| \tag{16}$$

4. Mean Absolute Error (MAE): This index measures the average magnitude of the errors in a set of predictions, without considering their direction.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left(t_i - y_i\right). \tag{17}$$

Volatility A comparison of volatility prediction for (1 day, 1 week, 1 month) ahead horizon in terms of root mean squared prediction error (RMSPE), mean squared prediction error (MSPE), and mean absolute prediction error (MAPE) defined in Minxia and Zhang (2014) and Nayak et al. (2015).

$$RMSPE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\vartheta_i - \hat{\vartheta}_i\right)^2} \tag{18}$$

$$MSPE = \frac{1}{n}\sum_{i=1}^{n}\left(\vartheta_i - \hat{\vartheta}_i\right)^2 \tag{19}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\vartheta_i - \hat{\vartheta}_i\right| \tag{20}$$

where ϑ_i is the realised volatility and $\hat{\vartheta}_i$ is the predicted volatility.

Momentum An Assessment for energy on (1-day, 1-week, 1-month) ahead horizon in terms of MSPE, RMSPE and MAPE defined in Nayak et al. (2015)

$$MSPE = \frac{1}{n}\sum_{i=1}^{n}\left(m_i - \hat{m}_i\right)^2 \tag{21}$$

$$RMSPE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(m_i - \hat{m}_i\right)^2} \tag{22}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|m_i - \hat{m}_i\right| \tag{23}$$

where m_i being the actual momentum and momentum predicted is $\hat{m}_i$

The accuracy, precision, recall, and F-score of the models are evaluated as proposed in Kumar et al. (2016).

Accuracy The accuracy of a prediction model is calculated as:

$$Accuracy(\%) = \frac{TR + TF}{TR + FR + FF + TF} \times 100 \tag{24}$$

Precision The precision signifies the portion of predicted rise (or fall) in stock, which are genuinely rising (or fall)

$$Precision\,for\,stock\,Rise\,(P^R) = \frac{TR}{TR + FR} \times 100 \tag{25}$$

$$Precision\,for\,stock\,Fall\left(P^F\right) = \frac{TF}{TF + FF} \times 100 \tag{26}$$

Recall This represents the fraction of the rise (or fall) in stock those predicted by the proposed model.

$$Recall\,for\,stock\,Rise\left(R^R\right) = \frac{TR}{TR + FF} \times 100 \tag{27}$$

$$Recall\,for\,stock\,Fall\left(R^F\right) = \frac{TF}{TF + FR} \times 100 \tag{28}$$

where TR = number of the correctly predicted rise in stock price. TF = number of correctly predicted fall in stock price.

FR = number of the incorrectly predicted rise in stock price. FF = number of incorrectly predicted fall in stock price.

F-score The association that exists between right stock (rise/fall) and that given by a predictor, if there is equality between precision and recall. F-score is represented as F_{sc}

$$Score\,for\,stock\,Rise\left(F^R_{sc}\right) = \frac{2 \times P^R \times R^R}{P^R + R^R} \tag{29}$$

$$Score\,for\,stock\,Fall\left(F^F_{sc}\right) = \frac{2 \times P^F \times R^F}{P^F + R^F} \tag{30}$$

The normalized mean squared error (NMSE) is a way to assess a model regarding the random walk (RW) impasse for financial time series model defined in de Araújo (2010) as:

$$NMSE = \frac{\sum_{j=1}^{N-1}\left(target_j - output_j\right)^2}{\sum_{j=1}^{N-1}\left(output_j - output_{j+1}\right)^2} \tag{31}$$

Prediction of change in direction (POCID) this indicator measures the ability of a predictive model to predict the future series value (target of prediction model) will see a decrease or an increase concerning the past value (de Araújo and Ferreira 2013).

$$POCID = \frac{100}{N}\sum_{j=1}^{N} D_j \tag{32}$$

where

$$D_j = \begin{cases} 1, & if\ (target_j - target_{j-1})(output_j - output_{j-1}) > 0 \\ 0, & otherwise \end{cases}$$

2.7 Relevant systematic reviews on stock market prediction

Though there exist a number studies on stock-market prediction we have not found any dedicated and complete comparative review and analysis of the available systems based on the type of input data, the number of data-source and the technique used, percentage of training and testing dataset.

In Baker and Wurgler (2007) a review of the sentiments of investors and its effects on the stock-market was conducted. An extensive discussion of literature and classification scheme for categorising previous studies on market prediction into theoretical work, description, law, and politics and applications was carried out by Tziralis and Tatsiopoulos (2007). In Demyanyk and Hasan (2010) a summary of the methodologies used and the experimental results achieved in various operations research and economics papers to explain, forecast, or propose solutions for fiscal crises or banking-defaults; was also outlined. In Krollner et al. (2010b), the authors gave a summary of the machine learning algorithm, input variables, and performance metrics. A brief review of Text-mining methodologies for stock-market prediction was performed by Nikfarjam et al. (2010).

Review of artificial neural network (ANN) in stock market prediction has been carried out by Dase and Pawar (2010), Soni (2011), Neelima et al. (2012), Chang et al. (2013), Goel et al. (2016) and Murekachiro (2016). These works concluded that ANN dominates in stock-market predictions globally. A comparative summary of predictive models for financial stock-market projections was carried out by Suthar et al. (2012). An overview of the techniques employed in predicting the stock market and enhancement made on these techniques in India was presented by Agrawal et al. (2013). In another study, a comprehensive, systematic reveal of fundamental analysis techniques for stock market prediction was undertaken in Nassirtoussi et al. (2014). Again, Kearney and Liu (2014) perform a survey of literature on the textual sentiment, contrasting and relating the various data sources, content-analysis methods, and experimental prototypes that have been used to date.

An analysis of the present and new (fundamental analysis, technical analysis, and machine learning) techniques in stock-market prediction were carried to verify if there is sufficient evidence to support weak-form EMH (Dunne 2015). The history and components involved in fundamental and technical analysis for decision making in the stock market were examined by Renu and Christie (2018). A review of technical analysis on stock-markets to categorise and code published articles, to offer a summary of research works that have added up to the development in stock-market predictions was performed by Nazário et al. (2017). The authors concluded that ANNs are best effected with backpropagation (BP). In the report of Shobana and Umamakeswari (2016), the authors gave a brief review of some data mining techniques employed in 16 articles for stock-market prediction.

From the above reviews, it was evident that none of the previous studies considered

1. The number of data-sources employed in stock prediction and how it influences the predictive models and methodology used.
2. A comparison of self-stated accuracy among research works of same soft-computing approach.
3. The software package for building predictive models and the approach (technical analysis or fundamental analysis) used.

This paper fills in the gap by reviewing past, and current state-of-the-art stock market prediction works based on the type of input data; the number of data-source and the soft-computing technique used; and a comparison of accuracy, time frame, software packages used for modelling.

3 Research design

This section discusses the methods adopted in selecting literature and the systematic review criteria.

3.1 Research framework

One hundred and twenty-two (122) related essays were collected using random sampling technique. These include published journal articles, conference proceedings papers, doctoral dissertations or supplementary unpublished academic working papers and reports between 2007 and 2018.

First, the selected papers were grouped into three broad categories based on the input data used, namely textual data, historical market data, and a combination of textual and historical market data. Secondly, each group was examined based on their exclusive features, as illustrated in the conceptual framework in Fig. 6.

4 Results and discussions

This section presents the results and the discussion of the study based on the conceptual framework in Fig. 6.

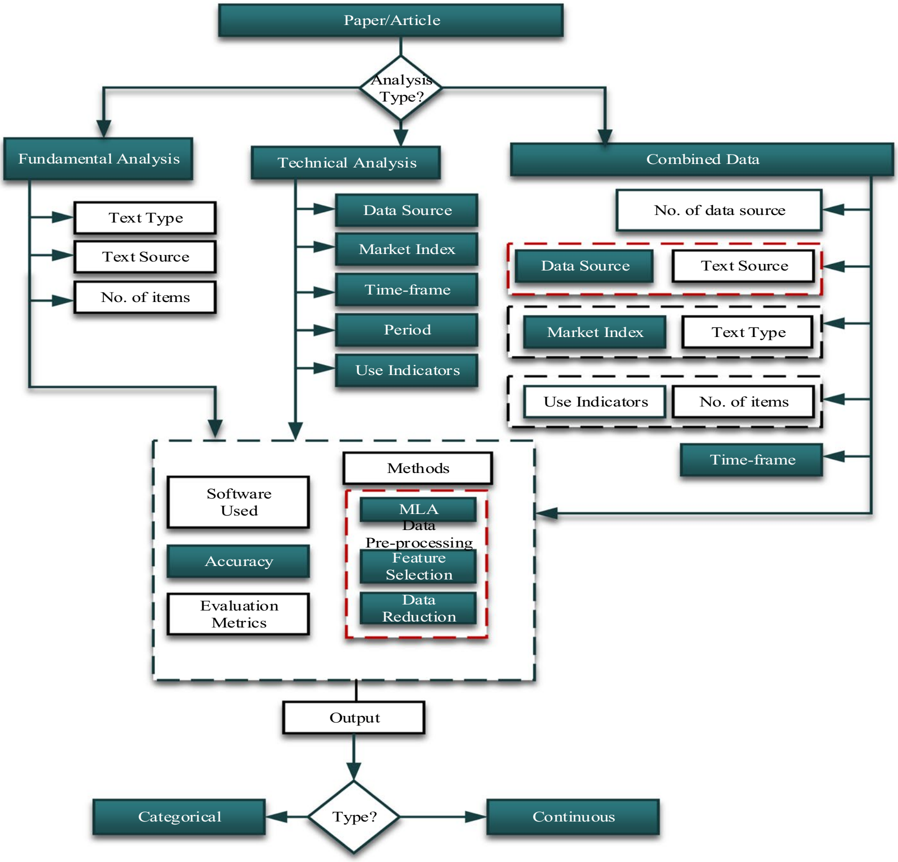

Fig. 6 The conceptual framework for categorizing stocks predictive techniques

4.1 Distribution of literature

Table 1 shows the distribution of surveyed works based on categorisation into the fundamental analysis, technical analysis, and combined. Eighty-one (81) was based on technical analysis (historical stock prices), twenty-eight (28) based on fundamental (web news, social media sentiments and macroeconomic variable) and thirteen (13) based on the combined analysis.

Table 1 Distribution of the literature based on categorisation

Category	No. of papers	%
Technical analysis	81	66
Fundamental analysis	28	23
Combined analysis	13	11
Total	122	100

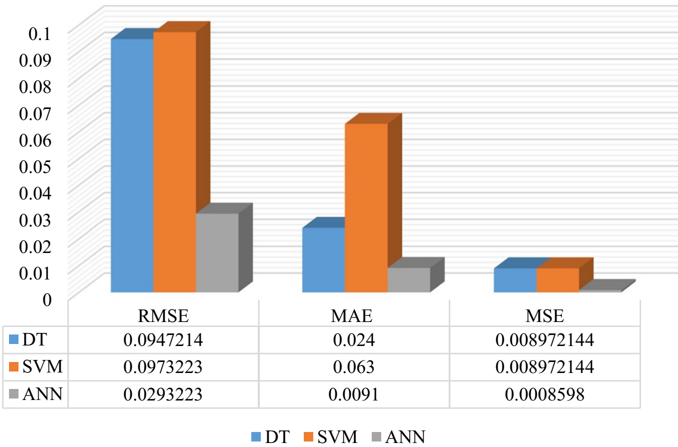

Fig. 7 Experimental results

4.2 Technical analysis (quantitative stocks-market data)

The study reveals that 66% of surveyed works were based on technical analysis, as shown in Table 1, due to its structured nature of historical stock price. Again, 99.9% of the reviewed works predicted the stock market as compared to the foreign exchange market (FOREX), as shown in Table 2 (Appendix). The time frame and the period of data collection were also examined, and the minimum period for data collection was a month by Khan et al. (2011) and 38 years being the maximum by Anthony et al. (2011). The results also revealed that the most utilised technical indicator was the simple moving average (SMA). Besides, it was observed that a very high percentage of studies that employed SMA are likely to use EMA, MACD, and RSI, as shown in Table 2 (Appendix). The outcome contradicts (Krollner et al. 2010a) assertion that 75% of their reviewed papers depended on some form on lagged-index data. An interesting observation from the study is that as little as 3.28% of the 122 articles focused on the African market. Also, the predictive timeframe of most studies was a day ahead prediction followed by Intraday in the ranges of to 1, 2, 3, 5 or 6 h and 1-week to 1-month as shown in Table 2 (Appendix).

4.3 Fundamental analysis (qualitative data)

The results revealed that twenty-eight (28) out of 122 reviewed papers used fundamental analysis for stock market prediction, and 98% of these works used sentiment analysis of social network sites (SSNs), as predictors of market movement as shown in Table 3 (Appendix). This result confirms (Bollen et al. 2011) report that the analysis of daily content of twitter feeds had the ability and capacity to cause an increase of DJIA prediction accuracy up to 87.6%. Out of this twenty-eight (28), social media accounted for 54%, financial web-news accounted for 29%, while search engine queries and macroeconomic variables accounted for 7% each. Again, it is revealed that macroeconomic variables as the only data source to stock-market prediction have not seen much attention, as shown in Table 3 (Appendix). Despite the following authors (Adam and Tweneboah

2008; Kuwornu and Victor 2011; Adusei 2014; Boachie et al. 2016; Suhaibu et al. 2017; Ayub 2018; Kwofie and Ansah 2018; Pervaiz et al. 2018; Tsaurai 2018; Umoru and Nwokoye 2018) sees a positive correlation between macroeconomic variables and stock-market returns. This gap in literature creates the need for future research into the stock market prediction based on macroeconomic variables. Furthermore, the study revealed that 89% out of the twenty-eight (28) works based on social media sentiments, were all on stock markets outside African. Thus, there is a need for studies measuring social media sentiments influence on the Africa stock markets.

4.4 Combined (qualitative and quantitative data) analysis

As discussed in Sect. 2.6, data for stock-market prediction is either quantitative data or qualitative data or both. Some researchers sought to harness the power in both data sources by formulating a joint input data of fundamental and technical indicators to improve the accuracy of stock-price predictive models. The study revealed that thirteen (13) out one hundred and twenty-two (122) of reviewed works was based on this approach, as shown in Table 4 (Appendix). The study revealed that 77% of these works used two (2) data sources, and 23% used three (3) data sources. To the best of our knowledge, none of the previous study as at the time of this paper have used four (past stock-data, social media, financial news, and macroeconomic variable) or more data source for stock-market prediction. Another opportunity for future studies based on four (4) or more data sources for stock market prediction.

4.5 Methods used for modelling and analysis

A summary of all the machine-learning algorithms used in the reviewed works is presented in this section. Hence, the main objective here was to give a report of what has been used and obtain a clearer understanding of what is lacking, which could be a pointer for future research. The results reveal that 92% of the algorithms used were classification machine-learning algorithms as tabulated in Table 5 (Appendix). This revelation implies that most of the reviewed work predicted stock price movement. Few of these works predicted the actual price of future stock. Hence, further studies can look at the difficulty in predicting the exact cost as compared to the movement.

The study again reveals that DTs, SVM, and ANN are the most used machine learning algorithms in stock market predictions, with ANN and SVM topping the list, as shown in Table 5 (Appendix). This outcome confirms (Almeida et al. 2010; Adebiyi et al. 2014a) report that ANN and SVM achieve higher generalisation potential than their counterparts. Again, more than 50% of the works reviewed, used hybrid algorithms as a way of compensating the flaws in individual algorithms, and this is evident in the accuracy reported in some hybrid models compared to different models of the same kind, as shown in Table 5 (Appendix). Hence, investigation of such hybrid algorithms in the environment of stock-market prediction may lead to novel insights that can lead to curiosity for future researchers.

About 5% of the works reviewed, showed that ensemble learning techniques were used for stock-market prediction in Europe, the Bovespa Index, and Shanghai, as shown in

Table 5 (Appendix). However, Ballings et al. (2015) in their works concludes that ensemble techniques should be benchmarked against other technology, with market-data from different continents. Their reason is that the accuracy of ensemble methods might differ over different dataset from different continents. Again, an opportunity for future studies. Term Frequency-Inverse Document Frequency (TF-IDF) was among the most common feature-representation technique for the textual data. However, 99% of the works reviewed implements feature selection algorithms that depend on correlation analysis. The most used metrics identified in the literature were MSE, RMSE, and MAPE. The high use of MSE and RMSE can be attributed to their effectiveness in measuring predictive model performance for short-term prediction. Furthermore, it was observed that MATLAB is the most used modelling tool for stock market prediction, as shown in Table 5 (Appendix). For prediction accuracy of stock-price movement, previous studies reported accuracy within 36.55–97.8%, as shown in Table 5 (Appendix). The outcome confirms that the stock market is highly predictable.

4.6 Training verse testing data volume

Every predictive model receives training and testing datasets, and Table 5 (Appendix) gives how most research works on stock market predictions partitioning their dataset. A higher percentage of the paper reviewed divides that dataset between (70–80%) for training and (30–20%) for testing. Except for a few cases (de Araújo 2010; Nhu et al. 2013; Sun et al. 2016) where the data were divided into three, that is training, testing, and validation.

4.7 Empirical setup

To verify the key findings in the literature, three (3) of the most used machine learning algorithms (DTs, SVM, and ANN) identified in Sect. 4.5 of this study for stock market prediction were selected and modelled against same data. Publicly available stock market data on the Ghana stock exchange (GSE) official website (https://gse.com.gh) was downloaded from January 2009 to February 2019. The downloaded dataset includes year high, year low, previous closing price, opening price, closing price, price change, closing bid price and closing offer. The multi-layer perceptron (MLP) was adopted. The performance of the selected algorithms was then compared based on the three most-used metrics identified in the literature (MSE, RMSE, and MAPE) in Sect. 4.5 of this paper. The dataset was cleaned from missing values (i.e. every missing value was replaced with the average value). We then normalised the dataset for efficiency using Eq. (33). Where: $a^{'}$ is the normalisation value; a = the value to be normalised, a_{min} *and* a_{max} are the minimum and maximum value of the dataset. Six (6) most-used technical indicators (SMA, EMA, MACD, RSI, OBV and volume ratio (VR)) identified under Sect. 4.2 of this paper were calculated from the dataset (using Microsoft Excel 2013) and use as input parameters to predict a 30-day ahead rise or fall of a stock price.

Based on the data portioning identified under Sect. 4.6 of this study, our dataset was portioned into two 80% for training and 20% for testing. The MLP implemented has three hidden layer (HL), HL_1 and HL_2 (with five (5) nodes), and HL_3 (with ten (10) nodes), the maximum iteration was set to 5000, optimizer = Limited-memory BFGS (lbfgs), activation = relu. For SVM, the Radial Basis Function (RBF) kernel was used, and the regularisation (C) = 100. The DT setting were, criterion = entropy, max_depth = 4. The experiments were conducted using scikit-learn library in Python (where the MLP, DT and SVM are already implemented) on the Anaconda framework.

$$a' = \frac{a - a_{(min)}}{a_{(\max)} - a_{(\min)}} \tag{33}$$

Figure 7 shows the outcome of the experiment. The error metrics (RMSE, MAE and MSE) values of ANN model (0.093, 0.009 and 0.00086) compared with DT (0.0947, 0.024 and 0.00897) and SVM (0.0973, 0.063 and 0.00947) reveals a better model fit of the ANN on stock market prediction than the SVM and DT models. However, the DT model offered less error margin between the predicted values and actual values compared to that of the SVM model. Consequently, the experimental outcome in this study confirms the high percentage of stock market predictions using ANN, as shown in Table 5 (Appendix). Despite, the DT performance is better than the SVM, which explains why it was among the top most used machine learning algorithm for building predictive models in stock market prediction.

5 Summary of findings

The extensive literature survey done in this paper was embarked on, to identify and assess all stock-market price and movement prediction related to academic articles from all possible sources of stock-market prediction research. Hence, resulted in the identification and assessment of one hundred and twenty-two (122) relevant literature on stock-market prediction between 2007 and 2018. However, we do not claim that this review is exhaustive, in that this paper does not give detailed practical understandings into the state of the predictive model research. There is a high bias in the use of technical indicators as input variables in the above-reviewed experiments; this leaves a gap for future research that combines behavioural and fundamental input variables.

Again the commonest and used technical indicators for stock market prediction were found to be SMA, EMA, MACD, RSI, and rate of change (ROC) which confirms (Krollner et al. 2010a; Renu and Christie 2018). Also, none of the one hundred and twenty-two (122) reviewed works has incorporate variable from past stock data, financial news, macroeconomic data, and social media sentiment as the input dataset. If all these data sources serve as input to a predictive model, a better and higher prediction accuracy results might be obtained as argued by Geva and Zahavi (2014).

More than 87% of the papers reviewed reported that their model beat their benchmark model. On the other hand, a percentage of previous studies did not cover real-world constraints like slippage and trading costs. A high percentage of stock-market prediction studies were carried out on the Asian and European stock markets, but Kumar and Thenmozhi (2006) and Ballings et al. (2015), argues that benchmarking ensemble machine learning algorithms for different continents against other techniques is of a higher necessity. In that, some learning techniques tend to perform better and of high accuracy in some parts of the globe than others.

Finally, an experimental setup with stock data from the Ghana Stock Exchange shows and affirms that artificial neural networks fit very well for stock market prediction as compared with support vector machines and decision trees, based on RMSE, MAE and MSE error metrics.

6 Conclusion

Previous works have also undertaken a review of the literature on fundamental and technical analysis (Nazário et al. 2017; Renu and Christie 2018), and machine learning algorithms applied in stock prediction by Dase and Pawar (2010), Soni (2011), Neelima et al. (2012), Chang et al. (2013) and Murekachiro (2016).

However, this study reviewed the pertinent literature on fundamental and technical analyses used in stock market predictions. Succinctly, the current study focused mainly on:

1. The nature of a dataset and the number of data sources used.
2. The data timeframe, the machine learning algorithms and task used.
3. A comparison of self-stated accuracy, error metrics, and software packages used for modelling in previous studies.
4. An experimental setup to verify finding in the literature.

The results revealed that ANN and SVM are usually used machine-learning algorithms for stock prediction. However, a lot of research work to improve stock prediction accuracy are ongoing using hybrid ensemble machine–learning method. It was noticed that, considering internal and more external factors could provide a more precise and accurate prediction. Besides, a deficient percentage of market prediction has been made on the African market, despite the volume of articles on stock-prediction.

6.1 Direction for future research

Many gaps were identified as opportunities for future studies in Sect. 4; however, our future work will focus on the performance of ensemble techniques over diverse stock-data from different continents.

Funding The declare that they have not received any funding or Grant for this work.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Appendix

See Tables 2, 3, 4, 5, 6 and 7.

Table 2 References based on quantitative data

Reference	Data source	Market index	Time frame	Period	Used indicators
Esfahanipour and Aghamiri (2010)	Stock	TSE and TEPIX	6 days	July 18, 2003–December 31, 2005	MA and BIAS
Vaisla and Bhatt (2010)	Stock and FOREX	NSE and exchange rate	Daily	1 April, 2005–30 March 2007	ER, SCP and SOP
Almeida et al. (2010)	Stock	15 stock from the SPSE	Daily	April 1, 1998–March 9, 2009	SMA, RSI, V, SCP, stop loss and stop gain
Nair et al. (2010b)	Stock	BSE-SENSEX	Daily	September 3, 200–March 7, 2010	RSI, daily open, high, low, close, trading volumes SMA, MACD
de Araújo (2010)	Stock	6 stock from NASDAQ	Daily	June 2000–June 22, 2007	Not stated
Luo et al. (2010)	Stock	SSE	Daily	January 2007–July 10 2008	SMA, MV, MVR, EMA, Disparity,
Naeini et al. (2010)	Stock	TSM	Daily	2000–2005	lowest, highest, and average stock prices
Hadavandi et al. (2010)	Stock	IBM corporation DELL Inc. British airlines Ryanair airlines	Daily	February 2003–Sep. 2004	SCP, SOP, High price and Low Price
Agrawal et al. (2010)	Stock	NASDAQ	Daily	2007–2008	RSI, SMA, MFI, MACD
Nair et al. (2010a)	Stock	BSE-SENSEX	Intraday (1 day)	February 2006–March 2010	MACD, SMA, V, RSI, SOP, SCP and OBV
Ansari et al. (2010)	Stock	NASDAQ	Not stated	December 2006–December 2008	Not stated
Boyacioglu and Avci (2010)	Stock	ISE	Daily	January 1990–December 2008	Republic gold selling price, exchange rates Treasury bill, etc.
Kannan et al. (2010)	Stock	Not stated	Daily	Not stated	TP, CMI, SMI, RSI, (BB), SMA and Bollienger Signal
Sureshkumar and Elango (2011)	Stock	NSE	Intraday	October 2007–October 2011	SCP, SOP, high price and low price

Table 2 (continued)

Reference	Data source	Market index	Time frame	Period	Used indicators
Atsalakis et al. (2011)	Stock	NBG	Daily	April 2007–November 2008,	SOP, SCP and SMA
Khan et al. (2011)	Stock	The stock of ACI pharma-ceutical	Next 8 Days	1-9-2010–31-10-2010	Net Asset Value (NAV), P/E, EPS
Enke et al. (2011)	Stock	S&P 500	Daily	January 1, 1980–January 1, 2009	T-bill-rate, Certificate of Deposit rate, Money Supply,
Anthony et al. (2011)	Stock	U.S. treasury bonds and treasury bills	Daily	April 1969–November 2007	U.S. treasury bonds and treasury bills
Kara et al. (2011)	stock	ISE	Daily	January 1997–December 31, 2007.	SMA, RSI, MACD and momentum
Wang and Qiang (2011)	Stock	SSEC	Daily	18/11/1991–10/02/2009	Not stated
Olaniyi et al. (2011)	Stock	NSE	Monthly	January 2007 and June 2008	SMA
Yeh et al. (2011)	Stock	TAIEX	Daily	October 2002–December 2005	EMA
Wei et al. (2011)	Stock	TAIEX	Daily	2000–2005	SMA, RSI volume Ratio, psychology line (PSY), V
Guresen et al. (2011)	Stock	NASDAQ	Daily	October 7, 2008–June 26, 2009	Not stated
Ju-Jie et al. (2012)	stock	SZII and DJIAI	Not stated	January 1991–December 2010	Not stated
Argiddi and Apte (2012)	Stock	BSE	Daily	Not stated	Not stated
Mohapatra and Raj (2012)	Stock	INSE (NSE NIFTY, INFY and BSE)	1-day, 1-week, 2-weeks, and 1-month	March 2000–March 2012	SMA and EMA
Shom and Padhy (2012)	Stock	INSE	Not stated	January 2007–December 2010	EMA and difference in the percentage of the price (RDP)
Shen et al. (2012)	Stock	NASDAQ, S&P500 and DJIA	Next-day	January 2000–October 2012	Not stated

Table 2 (continued)

Reference	Data source	Market index	Time frame	Period	Used indicators
Abhishek et al. (2012)	Stock	Microsoft Corporation	Daily	January 2011–December 31, 2011	open, high, low, adj. close and volume
Asadi et al. (2012)	Stock	TSE, TEPIX, NASDAQ, DJIA	Daily	March 7, 2001–January 30, 2009	KD, RSI, MACD, MA, BIAS
Wensheng et al. (2012)	Stock	ASMI	Daily	December 14, 2004–February 23, 2009	Not stated
Dutta et al. (2012)	Stock	INSE	Not stated	2005–2008	Percentage increase in net sales cash, earnings per share, book value
Yu and Liu (2012)	Stock	SSEC	Daily	2008–2010	Not stated
Chakravarty and Dash (2012)	Stock	S&P 500 BSE and DJIA	1 day, 1 week and 1 month	February 1992–October 2008	Not stated
Fajiang and Wang (2012)	Stock	SSE	Daily	January 4, 2000–April 30, 2010	SOP, SCP, daily highest price, and daily trade volume
Enke and Mehdiyev (2013)	Stock	S&P 500	Daily	Not stated	Treasury Bill, Producer Price Index, SOP and SCP
Kumar and Murugan (2013)	Stock	BSE-SENSEX and NIFTY MIDCAP50	Daily	1979 until 2012	Not stated
Ticknor (2013)	Stock	Goldman Sachs Group, Inc. (GS) and Micro-soft Corp. (MSFT)		4 January 2010–31 December 2012	low price, high price, opening price, close price, six technical indicators
Lin et al. (2013)	Stock	TSE	Daily	July 2008–January 2012	53 technical indicators (On balance volume (OBV), Price-earnings ratio (PER), ROI
Kazem et al. (2013)	Stock	NASDAQ	Daily	9/12/2007–11/11/2010	Daily closing (last)
de Araújo and Ferreira (2013)	Stock	Stock prices of (Directv Group Inc Microsoft Corporation and Yahoo Inc)	Daily	March 28th 2005–March 16th 2009	Stock Prices values

Table 2 (continued)

Reference	Data source	Market index	Time frame	Period	Used indicators
de Oliveira et al. (2013)	Stock	BM&FBOV-ESPA	Not stated	December 2000–November 2012	SCP, SOP, MACID, SMA, RSI, momentum and etc.
Hassan et al. (2013)	Stock	DJI, NASDAQ, S&P500, Germany (DAX), Japan (NIKKEI)	Not stated	January 1956–August 1995	Not stated
Hegazy et al. (2013)	Stock	S&P500	Not stated	January 2009–January 2012	RSI, Money Flow Index (MFI), EMA, MACD
Nhu et al. (2013)	Stock	HSE	Daily	2006–2013	Not stated
Sheta et al. (2013)	Stock	S&P500	Daily	Not stated	Not stated
Wang (2013)	Stock	KOSPI and HSI	Not stated	January 2002–January 2012	Not stated
Adebiyi et al. (2014b)	Stock	NYSE and NSE	Daily	25th April 1995–25th February 2011	open price, low price, high price and close price
Adebiyi et al. (2014a)	Stock	NYSE	Daily	August 17, 1988–February 25, 2011	open price, low price, high price, close price, and volume traded
Bhagwant et al. (2014)	Stock	Not stated	Next-5 days	Not stated	Not stated
Bisoi and Dash (2014)	Stock	BSE-SENSEX	1-day-ahead	3rd January 2005–13th August, 2008	SMA, stochastic oscillators (K) and %R (William indicator), etc.
Rajashree et al. (2014)	Stock	BSE-SENSEX Tokyo stock exchange, S&P500	1-day-ahead	4th January 2010–30th April 2013 and	SMA, SOP and SCP
Fang et al. (2014)	Stock	Shenzhen Composite Index	Not stated	September 11, 2006–September 18, 2007	Not stated
Yetis et al. (2014)	Stock	NASDAQ	Daily	Between 2012 and 2013	SOP, SCP, V and high
Gupta and Sharma (2014)	Stock	SSE	Not stated	5 years	Not stated
Pulido et al. (2014)	Stock	Mexican stock exchange	Not stated	11/09/05–01/15/09	Not stated

Table 2 (continued)

Reference	Data source	Market index	Time frame	Period	Used indicators
Stanković et al. (2015)	Stock	S&P and Morgan Stanley capital international (MSCI)	Daily	2009–1 October, 2013	SMA, RSI, MACDB and EMA
Wanjawa and Muchemi (2014)	Stock	NYSE	Daily	2008–2012	Not stated
Xi et al. (2014)	Stock	Chinese Shanghai stock market	Daily	4 January 2012–8 October 2012	SCP
Zhang et al. (2014)	Stock	SSE	Daily	1999–2011	25 (E/P, B/P, S/P, ROA, gross)
Rather et al. (2014)	Stock	INSE	Daily	January 2007–March 2010	
Patel et al. (2015b)	Stock	CNX Nifty and S&P BSE SENSEX	Intraday 1–10, 15 and 30 days	January 2003–Decembe 2012	SMA, EMA, MACD, RSI, Momentum
Patel et al. (2015a)	Stock	CNX nifty and S&P BSE sensex	Daily	January 2003–Dec 2012	SMA, RSI, Momentum, stochastic, MACD
Nayak et al. (2015)	Stock	BSE sensex and CNX nifty	1 day, 1 week and 1 month	January 1990–August 2014	SMA, RSI, EMA, triangular moving average (TMA)
Göçken et al. (2016)	Stock	BIST100	Intraday	08/06/2005–27/05/2013	45 technical indicators (SOP, SCP, RSI, SMA and more.)
Wei (2016)	Stock	TAIEX and HSI	Daily	2000–2006	Not stated
Ertuna (2016)	Stock	Apple Inc. (AAPL)	Daily	January 2000–January 2016	SOP
Dash and Dash (2016)	Stock	BSE SENSEX and (S&P500)	Daily	02 July 2012–06 August 2014	SCP, SMA, Williams %R, RSI
Wanjawa (2016)	Stock	SSE	Daily	2016–2016	Not stated
Ghaznavi et al. (2016)	Stock	Tehran Artmis Company	Daily	2008–2015	Not stated
Kumar et al.(2016)	Stock	CNX NIFTY, S&P BSE SENSEX	Not stated	January 2008–December 2013	RSI, MTM, CCI, ROC, TSI %R, MACD, EMA and SMA

Table 2 (continued)

Reference	Data source	Market index	Time frame	Period	Used indicators
Chen and Hao (2017)	Stock	SSE & SZSE COMP SUB IND	Intraday [5, 10, 15, 20, 30] days	October 31st 2008–Dec 31st 2014	SMA, EMA, MACD Volume Ratio (VR), RSI, OBV and MTM
Ibrahim (2017)	Stock	NSE		1985–2014	MA, AR
Sasan et al. (2017)	Stock	TSE	Daily	2002–2012	44 Financial ratios, fundamental index (P/E, P/S, EPS, and etc.)
Chong et al.(2017)	Stock	Korea KOSPI	Daily	January 2010–December 2014	Not stated
Adebayo et al.(2017)	Stock	NSE	Daily	January–May 2015	number of deals, closing price and value traded
García et al.(2018)	Stock	German Dax-30 index	1 day ahead	1999–2017	SCP, SOP, MACD, SMA, MTM, OBV, RSI and PSY
Zhou et al. (2018)	Stock	China stock market	Not stated	January–December 2016	SOP, SCP, MACD, EMA, SMA, RSI
Dosdoğru et al. (2018)	Stock	S&P 500	Daily	Not stated	36 technical indicators (SCP, SOP, MTM, RSI, MACD
Thanh et al. (2018)	Stock	Vietnamese stock market index	Daily	January 2010–April 2016	CPI and EXP

Table 3 Reference on fundamental analysis (textual data) for stock prediction

Reference	Text form	Text source	Period	No. of items
Bollen et al. (2011)	Tweets	Twitter	February–Dec 2008.	9,853,498
Zhang et al. (2011)	Tweets	Twitter	March 2009–September 2009	7,747,200
Perwej and Perwej (2012)	Inflation rate	BSE	Not stated	Not Stated
Vu et al. (2012)	Tweets	Twitter	April 2011–May 2011	5,001,460
Bordino et al. (2012)	Web search queries	Yahoo! search engine	2010–2011	Not stated
Chen and Lazer (2013)	Tweets	Twitter	Not stated	Not stated
Makrehchi et al. (2013)	Tweets	Twitter	March–13th July 2012	30 million
Hagenau et al. (2013)	Web news and corporate announcements	Deutsche Gesellschaft für Adhoc-Publizität (DGAP)	1997–2011	14,3 48
Dondio (2013)	Web-traffic	Online search on US SP500	2007–May 2012	552,016
Si et al. (2013)	Tweets	Twitter	November 2012–February 2013	624,782
Ding et al. (2014)	Public financial news	Reuters news and Bloomberg news	October 2006–November 2013	Reuters: 106,521 Bloomberg: 447,145
Li et al. (2014d)	Financial news articles	FINET	January 2003–March 2008	Not stated
Jianfeng et al. (2014)	Tweets	Twitter	November 2 2012–April 3 2013	629,977
Yoosin et al. (2014)	News articles	Naver.com	1 year (2011)	78,216
Chen et al. (2014)	blogging	Sina Weibo	September 29th, 2011–Match 29th, 2013	256,691
Li et al. (2014b)	Web news and financial discussion board	www.sina.com www.eastmoney.com	01/01/2011–12/31/2011	124,470
Li et al. (2014a)	Web news and public moods	www.sina.com www.eastmoney.com	January 1, 2011–December 31, 2011	124,470
Liu et al. (2015)	Firm-specific social media metrics (tweets)	Twitter	June 30, 2013–July 4, 2013	Not stated
Sun et al. (2016)	Social media	StockTwits.com.	January 2011–August 2015	45 million
Boachie et al. (2016)	Macroeconomic variables	Bank of Ghana	2010–2013	
Yifan et al. (2017)	financial discussion board	East Money Forum	September 2015–September 2016	96,000
Checkley et al. (2017)	Tweets and public moods	Twitter and StockTwits,	17th February 2012–17th October 2014	Not stated
Pimprikar et al. (2017)	Tweets	Twitter	2010	Not stated

Table 3 (continued)

Reference	Text form	Text source	Period	No. of items
Sorto et al. (2017)	News articles	Not stated	30 days	Not stated
Coyne et al. (2017)	Twits	StockTwits	May 2016–April 2017	1,013,794
Zhou et al. (2017)	Chinese tweets	Weibo	December 2014–December 2015.	10,550,525
Maknickiene et al. (2018)	Tweets	Twitter	2010–2016	Not stated
Nisar and Yeung (2018)	Tweets	Twitter	May 4th 2016eMay 9th 2016	60,944

Table 4 Combined (two or more data source) analysis

Reference	No. of data source	Data source	Market index	Text form	No. of items	Time frame	Used indicators
Tsai and Hsiao (2010)	3	Stock data Macroeconomic Variables Financial variables	Taiwan Stock Exchange Financial Price	Not stated	Not stated	March 2000–June 2007	Not stated
Lahmiri (2011)	2	Stock data and Macroeconomic Variables	S&P500	Saint-Louis Federal Bank	Not stated	January 2000–January 2008	Treasury Bill Rate, SMA, Exchange Rate
Babu et al. (2012)	2	Financial Reports and financial ratios	BSE30 index	Yahoo, BSE, NSE	Not stated	January 1, 1996–December 31, 2009	operating margin, return on equity, etc.
Adebiyi et al. (2012)	2	Not stated	Not stated	Not stated	Not stated	Not stated	SOP, SCP, EPS, high-price,
Mittal and Goel (2012)	2	Stock data and Social media sentiments	Dow Jones Industrial Average (DJIA)	Twitter	476 million	June 2009–December 2009	SOP, SCP, Low, High
Porshnev et al. (2013)	2	Historical stock data and social media sentiments	S&P500, DJIA	Twitter	755,000,101	13/02/2013–29/09/2013	SOP and SCP
Li et al. (2014c)	2	Market news and stock prices	HKSE	Bloomberg	28,885	1 year (2001)	RSI, Raw Stochastic Value
Ming et al. (2014)	2	News articles and past stock prices	S&P 500 DJIA NASDAQ	Wall Street Journal	Not stated	1/1/2008–9/30/2013	SOP, SCP
Geva and Zahavi (2014)	2	News articles and past stock prices	NYSE and Quotes (TAQ) database	Reuters 3000	51,263	15th, 2006–August 31st, 2007	Not stated
Ballings et al. (2015)	2	Stock data and macroeconomic data	Not stated	Not stated	Not stated	2009–2010	the current ratio, public debt nad GDP.
Li et al. (2015)	3	Stock data News articles Discussion boards	China Securities Index (CSI 100)	Sina.com and East-Money.com	124,470	January1, 2011 and December31, 2011	Not stated

Table 4 (continued)

Reference	No. of data source	Data source	Market index	Text form	No. of items	Time frame	Used indicators
Kraus and Feuerriegel (2017)	2	Financial new Past stock data	Germany (DAX)	www.dgap.de	13,135	Not stated	Not stated
Zhang et al. (2017)	3	Past stock data Social media Web news	China A-share HK	Finance.sina.com.cn	Web news = 83,729 Social media = 6,166,247	January 1, 2015–December 31, 2017	P/E, P/B and PCF

Table 5 Methods, reported accuracy, evaluation metrics and output type

Reference	Method				Accuracy	Evaluation metrics	Output type
	Software	MLA	Feature selection	Data reduction			
Esfahanipour and Aghamiri (2010)	MATLAB and SPSS	TSK fuzzy	Stepwise regression analysis		97.8%	MAPE	Cat: buy, sell
Vaisla and Bhatt (2010)	NEURALWARE and neural works predict	Neural network and statistical techniques	Not stated	Neural log	Not stated	MAPE, MSE and RMSE	Not stated
Almeida et al. (2010)	Not stated	k-NN	Not stated		36.55–50%	Profit obtained	Cat: buy, sell and keep
Tsai and Hsiao (2010)	Not stated	ANN	PCA, GA and CART	Not stated	62.22–78.98%	Not stated	Not stated
de Araújo (2010)	Not stated	ANN	Not stated			MSE, MAPE, NMSE, POCID	Not stated
Luo et al. (2010)	MATLAB	SVM Regression (SVR–CM)	Stepwise (SRA), Multi–level Recursive		75%	MAPE, RMSE, trend accuracy and R^2	Not stated
Naeini et al. (2010)	Not stated	NNN	Not stated	Not stated	Not Stated	MSE, MAPE, MAD RMSE	Not stated
Hadavandi et al. (2010)	Not stated	Genetic Fuzzy Systems (GFS) and ANN	SRA	Not stated	Not stated	MAPE and MSE	Not stated
Ansari et al. (2010)	MATLAB	Adaptive Neural Fuzzy Interface System (ANFIS)	Statistical tests and Analysis and tests and analysis		74%	Not stated	Not stated
Boyacioglu and Avci (2010)	MATLAB	Adaptive-Network-Based Fuzzy Inference System (ANFIS)	Not stated	Not stated	98.3%	RMS, R^2 and coefficient of variation (Cov)	Not stated
Kannan et al. (2010)	Not Stated	Data Mining Techniques	Not stated	Not stated	50%	Not stated	Not stated

Table 5 (continued)

Reference	Method				Accuracy	Evaluation metrics	Output type
	Software	MLA	Feature selection	Data reduction			
Sureshkumar and Elango (2011)	Weka 3.6.3	Gaussian processes, isotonic regression, least mean square	Not stated	Not stated	Not stated	MAE, RMSE, RAE and RRSE	Cat: up or down
Bollen et al. (2011)	Not stated	Fuzzy neural network	By OpinionFinder	By OpinionFinder	86.7%	MAPE	Catl: up and down
Atsalakis et al. (2011)		Neuro-Fuzzy	Not stated	Not stated		Not stated	Cat: Buy and sell
Khan et al. (2011)	Not stated	ANN	Not stated	Not stated	Not stated	Not stated	Con
Enke et al. (2011)	Not stated	Fuzzy type-2 neural network	Multiple regression analysis (MRA)	Multiple regression	Not stated	RMSE	Not stated
Anthony et al. (2011)	NeuroSolutions version 5.0.	Time-delay neural network (TDNN)	Not stated	Not stated	Not stated	MSE and R-square	Not stated
Lahmiri (2011)	Not stated	Probabilistic neural networks (PNN) and SVM	Not stated	Not stated	Not stated	Not stated	Cat: up and down
Kara et al. (2011)	MATLAB	ANN and SVM	Not stated	ANN = 75.74% SVM = 71.52%		Not stated	Not stated
Wang and Qiang (2011)	Not stated	ANN	Not stated	Not stated	Not stated	Not stated	Not stated
Olaniyi et al. (2011)	Not stated	Regression analysis	Not stated	Not stated	Not stated	Not stated	Not stated
Yeh et al. (2011)	Not stated	Multiple-kernel SVR	Not stated	Not stated	Not stated	RMSE	Not stated
Wei et al. (2011)	Not stated	ANFIS	Pearson correlation with	Subtractive clustering method		RMSE	Not stated
Guresen et al. (2011)	NeuroSolutions 5.06 software	ANN	Not stated	Not stated	Not stated	MSE, and Mean Absolute Deviate	Not stated

Table 5 (continued)

Reference	Method				Accuracy	Evaluation metrics	Output type
	Software	MLA	Feature selection	Data reduction			
Babu et al. (2012)	Not stated	Hierarchical agglomerative and Recursive K-means clustering	Not stated	stemming algorithm (Bag-of-words)	Not stated	Not stated	Cat: rise, and drop
Ju-Jie et al. (2012)	MATLAB	ESM, ARIMA and BPNN	Not stated	Not stated	Not stated	MAE, RMSE, MAPE,	Not stated
Argiddi and Apte (2012)	Not stated	Association rule mining	Not stated	Not stated	Not stated	Not stated	Not stated
Adebiyi et al. (2012)	MATLAB	ANN	Not stated	Not stated	Not stated	Not stated	Con
Perwej and Perwej (2012)	Not stated	ANN and genetic algorithm (GA)	Not stated	Not stated	Not stated	Not stated	Not stated
Mohapatra and Raj (2012)	Not stated	Differential evolutionary neural network	Not stated	Not stated	Not stated	RMSE and MAPE	Not stated
Mittal and Goel (2012)	MATLAB	Linear regression, logistic regression, SVMs	Not stated	Not stated	Not stated	MAPE	Not stated
Vu et al. (2012)	MATLAB	Decision trees	Daily total	Pre-defined company	75.00–82.93%		Cat: up and down
Shom and Padhy (2012)	MATLAB	SVM and back propagation technique (BP)	Not stated	Not stated	Not stated	Normalised mean squared error (NMSE), MAE	Not stated
Shen et al. (2012)	Not stated	SVM	Autocorrelation and cross-correlation technique	NASDAQ = 74.4% S&P500 = 76% DJIA = 77.6%		RMSE	Not stated

Table 5 (continued)

Reference	Method				Accuracy	Evaluation metrics	Output type
	Software	MLA	Feature selection	Data reduction			
Abhishek et al. (2012)	MATLAB	ANN	Not stated	Not stated	Not stated	MSE	Not stated
Asadi et al. (2012)	Not stated	ANN (GA and Lev-enbergMarquardt (LM) algorithm)	SRA	Not stated	Not stated	RMSE, MAPE	Not stated
Wensheng et al. (2012)	Not stated	Nonlinear independent component analysis and NN	Nonlinear independent component analysis	Not stated	Not stated	RMSE, MAD, MAPE,	Not stated
Dutta et al. (2012)	Not stated	Logistic regression (LR)	Not stated	Not stated	Not stated	Not stated	Cat: good or poor
Yu and Liu (2012)	Not stated	SVM and empirical mode decomposition	Not stated	Not stated	Not stated	Absolute and comparative errors	Not stated
Chakravarty and Dash (2012)	Not stated	PSO, type-2 fuzzy and NN	Not stated	Not stated	Not stated	MAPE and RMSE	Not stated
Fajiang and Wang (2012)	Not stated	NN	Not stated	Not stated	Not stated	MAPE	Not stated
Enke and Mehdiyev (2013)	Not stated	Fuzzy inference neural network	SRA	Not stated	Not stated	RMSE	Not stated
Kumar and Murugan (2013)	MATLAB	ANN	Not stated	Not stated	Not stated	MAE, MAPE	Not stated
Chen and Lazer (2013)	Not stated	Linear regression (LR)	Not stated	Not stated	Not stated	Not stated	Not stated
Ticknor (2013)	MATLAB	ANN	Not stated	Not stated	Not stated	MAPE	Con

Table 5 (continued)

Reference	Method				Accuracy	Evaluation metrics	Output type
	Software	MLA	Feature selection	Data reduction			
Hagenau et al. (2013)	Not stated	SVM	Bag-of-words, Noun Phrases, word-combinations, N-grams	The frequency for news, Chi^2-approach	76%	Not stated	Cat: positive or negative
Lin et al. (2013)	Not stated	SVM	Correlation-based SVM filter	Not stated	Not stated	Not stated	Con
Kazem et al. (2013)	MATLAB	SVR	Not stated	Not stated	Not stated	MSE and MAPE	Not stated
de Araújo and Ferreira (2013)	Not stated	Morphological-rank-linear filter and modified genetic algorithm	Not stated	Not stated	Not stated	MSE, NMSE, POCID and MAPE	Not stated
de Oliveira et al. (2013)	Not stated	ANN	Not stated	Not stated	93.62–87.50%	MAPE, NMSE and RMSE	Not stated
Hassan et al. (2013)	Not stated	Fuzzy	Not stated	Not stated	Not stated	MAPE, NRMSE and *t* test	Not stated
Hegazy et al. (2013)	Not stated	PSO and least square SVM	Not stated	Not stated	Not stated	MSE	Not stated
Nhu et al. (2013)	MATLAB	Neuro-fuzzy	Not stated	Not stated	Not stated	RMSE	Not stated
Porshnev et al. (2013)	Not stated	SVM	Dictionary-based approach	Not stated	68.63%	Not stated	Not stated
Sheta et al. (2013)	HeuristicLab framework	Genetic programming (GP)	Not stated	Not stated	Not stated	Variance-accounted-for (VAF) and RMSE	Not stated
Wang (2013)	Not stated	PCA and SVM	PCA	Not stated	Not stated	Not stated	Not stated
Adebiyi et al. (2014b)	Eviews version 5	ARIMA	Not stated	Not stated	Not stated	R-square	Con

Table 5 (continued)

Reference	Method				Accuracy	Evaluation metrics	Output type
	Software	MLA	Feature selection	Data reduction			
Adebiyi et al. (2014a)	EViews and MATLAB	ANN and ARIMA	Not stated	Not stated	Not stated	MSE, R-square	Con
Bhagwant et al. (2014)	Visual studio (C#)	ANN	Not stated	Not stated	Not stated	MSE	Con
Rajashree et al. (2014)	Not stated	Hybrid NN	Not stated	Not stated	Not stated	RMSE, MAPE and AMAPE	Cat: up, down and no trend
Ding et al. (2014)	Not stated	Deep NN and SVM	Bag-of-words features (TFIDF) and events-based	Not stated	Not stated	Matthews Correlation Coefficient(MCC)	Con
Fang et al. (2014)	Not stated	GA and wavelet neural networks (WNN)	Not stated	Not stated	Not stated	MSE, MAE and MAPE	Not stated
Yetis et al. (2014)	MATLAB	ANN	Not stated	Not stated	Not stated	MSE	Not stated
Gupta and Sharma (2014)	Not stated	K-means and decision trees	Not stated	Not stated	Not stated	Not stated	Con
X. Li, Huang, et al. (2014)	MATLAB	SVR	Not stated	Not stated	Not stated	Precision, MAE and MSE	Not stated
X. Li, Xie, et al. (2014)	Not stated	SVM	Bag-of-words	Not stated	Not stated	Not stated	Cat
Ming et al. (2014)	Not stated	Autoregressive (AR) models	Bag-of-word	Not stated	Not stated	Not stated	Not stated
Pulido et al. (2014)	Not stated	PSO, NN and Fuzzy	Not stated	Not stated	Not stated	Not stated	Con
Stanković et al. (2015)	Not stated	Least square SVM (LS-SVM)	Not stated	Not stated	Not stated	Not stated	Con
Wanjawa and Muchemi (2014)	C#	ANN	Not stated	Not stated	Not stated	MAPE and RMSE	Con
Xi et al. (2014)	MATLAB	RBF NN	Not stated	Not stated	Not stated	RMSE	Con

Table 5 (continued)

Reference	Method				Accuracy	Evaluation metrics	Output type
	Software	MLA	Feature selection	Data reduction			
Yoosin et al. (2014)	Not stated	Natural LANGUAGE PROCESSIng	Bag-of-words	Not stated	Not stated	Recall, Precision, F1-score	Cat: buy, sell, hold
Zhang et al. (2014)	Weka	LR, NaiveBayes (NB), BayesNet (BN), DT	PCA, CART, LASSO, CFS	Not stated	Not stated	Precision	Cat
Geva and Zahavi (2014)	GainSmarts software package	Feed-forward NN, decision trees, GA	Bag-of-words, stepwise logistic regression	Not stated	Not stated	Not stated	Cat
Chen et al. (2014)	Not stated	SVM, generalized linear model (GLM)	Bag of words, noun phrases, verb phrases	Factorisation machine (FM)	81%	Accuracy	Cat: up or down
Q. Li, Wang, Li, et al. (2014)	Not stated	Granger causality SVR	Not stated	Not stated	Not stated	RMSE	Con
Q. Li, Wang, Gong, et al. (2014)	Not stated	SVR	Not stated	Not stated	Not stated	RMSEs	Con
Rather et al. (2014)	Not stated	ARM model, exponential smoothing model and RNN.	Not stated	Not stated	Not stated	MSE and MAE	Cat
Patel et al. (2015b)	Not stated	SVR, ANN and RF	Not stated	Not stated	Not stated	MAPE, MAE, rRMSE	Con
Ballings et al. (2015)	Not stated	RF, SVM Kernel Factory, AdaBoost, NN, K-NN and LR	Not stated	Not stated	Not stated	The area under the receiver operating characteristic curve (AUC)	Cat
Patel et al. (2015a)	Not stated	ANN, SVM, RF and Naive-Bayes	Not stated	Not stated	86.69–89.33%, 89.98–90.19%	Accuracy F-measure, precision	Con

Table 5 (continued)

Reference	Method				Accuracy	Evaluation metrics	Output type
	Software	MLA	Feature selection	Data reduction			
Liu et al. (2015)	Not stated	K-means clustering	Pairwise correlation	Not stated	Not stated	Not stated	Cat
Li et al. (2015)	Not stated	SVR, PCA, Tensor	ISOMAP	Not stated	Not stated	Not stated	Cat
Nayak et al. (2015)	Not stated	SVM and KNN	Not stated	Not stated	Not stated	MAPE, RMSE, MSFE	Cat
Sun et al. (2016)	MATLAB	Sparse matrix factorization	Not stated	Not stated	Not stated	Precision Recall	Cat
Göçken et al. (2016)	Not stated	ANN	Harmony Search (HS) and GA	Not stated	Not stated	RMSE, MARE, MSRE, RMSRE,	Cat
Wei (2016)	Not stated	ANFIS	Not stated	Not stated	Not stated	RMSE	Cat
Boachie et al. (2016)	STATA version 11	Linear regression (LR)	Not stated	Not stated	Not stated	R^2 p value	
Ertuna (2016)	Encog machine learning framework	NN	Not stated	Not stated	Not stated	Not stated	Cat
Dash and Dash (2016)		Recurrent neuro-fuzzy inference	Not stated	Not stated	Not stated	RMSE, MAPE and MAE	Cat
Wanjawa (2016)	C#	ANN	Not stated	Not stated	Not stated	MAPE and RMSE	Con
Ghaznavi et al. (2016)	Not stated	Radial basis function neural networks (RBFNN)	Not stated	Not stated	Not stated	MSE	Not stated
Kumar et al. (2016)	Not stated	Proximal SVM	Linear correlation (LC), rank correlation (RC), regression relief (RR)	Not stated	Not stated	Joint Prediction Error (JPE), Accuracy, Precision, F-score and Recall	Cat: rise, fall

Table 5 (continued)

Reference	Method				Accuracy	Evaluation metrics	Output type
	Software	MLA	Feature selection	Data reduction			
Chen and Hao (2017)	Not stated	FWSVM FWSVM-FWKNN SVM KNN		Not stated	Not stated	MAPE, RMSE	Cat
Yifan et al. (2017)	Not stated	RNN		Not stated	Not stated	Accuracy	Cat
Checkley et al. (2017)	Not stated	Naïve model	Direct Granger method	Not stated	Not stated	MAPE RMSE	Con
Ibrahim (2017)	Not stated	(ARCH, GARCH, TGARCH, EGARCH), ARIMA	Autocorrelation functions (ACF) and Partial autocorrelation function (PACF)	Not stated	Not stated	*P* value	Con
Pimprikar et al. (2017)	Not stated	LR, SVM and LSTM-NN Naïve Bayes		LR = 82% SMV = 60%		RMSE	Not stated
Sasan et al. (2017)	Not stated	DT, ANN SVM K-NN and RF	Wrapper-GA		83.6–88.2%	Not stated	Not stated
Chong et al. (2017)	Not stated	Deep NN	PCA, autoencoder, and restricted Boltzmann machine	Not stated	Not stated	NMSE, RMSE, MAE and MI	Con
Coyne et al. (2017)	Python and Scikit-Learn	Naïve Bayes, MLP	bag-of-words, linear regression,	Not stated	65%	Not stated	Not stated
Kraus and Feuer-riegel (2017)	Python, Scikit-Learn TensorFlow and Theano	Naïve Bayes, RNN and LSTM	bag-of-words	Not stated	Not stated	Accuracy, AUC RMSE, MSE and MAE	Cat: Up and down
García et al. (2018)	Not stated	Fuzzy-Neuron	Not stated	Not stated	Not stated	Not stated	Cat
Zhang et al. (2017)	Not stated	SVM, PCA Tensor-based (TeSIA)	PCA	Not stated	61.73–62.50%	ACC and MCC	Cat

Table 5 (continued)

Reference	Method				Accuracy	Evaluation metrics	Output type
	Software	MLA	Feature selection	Data reduction			
Adebayo et al. (2017)	R	Takagi-Sugueno-Kang -Fuzzy Rule-Based System	Not stated	Not stated	Not stated	RMSE, MSE and SMAPE	Cat: buy, sell, or hold
Zhou et al. (2018)	Not stated	LSTM and convolutional neural network (CNN)	Not stated	Not stated	Not stated	RMSRE and direction prediction accuracy (DPA)	Cat
Maknickiene et al. (2018)	Not stated	RNN	Not stated	Not stated	73%	Accuracy	Cat
Dosdoğru et al. (2018)	Not stated	ANN	Ant Lion Optimization (ALO), Bird Swarm Optimization	Not stated	Not stated	MAE, RMSE, MARE, MSRE, RMSRE,	Cat
Thanh et al. (2018)	Not stated	Multiple regression	PCA	PCA	Not stated	% absolute error, R^2 Adjusted R^2	Con

Cat categorical, *Con* continuous

Table 6 Data partitioning of reviewed work

Reference	Total dataset	Training dataset	Testing dataset
Esfahanipour and Aghamiri (2010)	614	494 (80.5%)	120 (19.5%)
Vaisla and Bhatt (2010)	500	Not stated	Not stated
Almeida et al. (2010)	Not stated		
Tsai and Hsiao (2010)	Not stated	Not stated	Not stated
de Araújo (2010)	1758	879 (50%)	Validation (440) 25% and testing (436) 25%
Luo et al. (2010)	440	400 (90.9%)	40 (9.1%)
Naeini et al. (2010)	Not stated	Not stated	Not stated
Hadavandi et al. (2010)	IBM and DELL 491	400 (81.5%)	91 (18.5%)
	British Airlines 594	503 (84.7%)	91 (15.3%)
	Ryanair airlines 471	400 (85%)	71 (15%)
Agrawal et al. (2010)	371	318 (85.7%)	53 (14.3%)
Nair et al. (2010a)	1097	1040 (94.8%)	57 (5.2%)
Ansari et al. (2010)	652	529 (81%)	123 (19%)
Boyacioglu and Avci (2010)	228	122 (54%)	106 (46%)
Kannan et al. (2010)	Not stated	Not stated	Not stated
Sureshkumar and Elango (2011)	1000	Not stated	Not stated
Bollen et al. (2011)	9,853,498	9,661,490 (98%)	192,008 (2%)
Atsalakis et al. (2011)	2060	2000 (97%)	60 (3%)
Khan et al. (2011)	Not stated	Not stated	Not stated
Enke et al. (2011)	1800	1440 (80%)	360 (20%)
Anthony et al. (2011)	947	905 (95.56%)	42 (4.44%)
Lahmiri (2011)	Not stated	Not stated	Not stated
Kara et al. (2011)	2733	Not stated	Not stated
Wang and Qiang (2011)	Not stated	Not stated	Not stated
Olaniyi et al. (2011)	Not stated	Not stated	Not stated
Wei et al. (2011)	2000–2005	10-month period of the stock data, from January to October	November and December,

Table 6 (continued)

Reference	Total dataset	Training dataset	Testing dataset
Guresen et al. (2011)	182 days	146 days (80%)	36 days (20%)
Babu et al. (2012)	26,255	20,884 (79.54%)	5371 (20.45%)
Ju-Jie et al. (2012)	SZII: 216	168 (78%)	48 (22%)
	DJIAI: 240	180 (75%)	60 (25%)
Argiddi and Apte (2012)	731	Not stated	Not stated
Mohapatra and Raj (2012)	INSE NIFTY: 1600 INFY: 2400 BSE: 2000	1200 (75%) 2000 (83%) 1600 (80%)	400 (25%) 400 (17%) 400 (20%)
Wensheng et al. (2012)	1000	800 (80%)	200 (20%)
Chakravarty and Dash (2012)	(S&P 500): 3228	2000 (62%)	450 (14%)
	BSE: 4000	2000 (50%)	450 (11.25%)
	DJIA: 2302	1000 (43.44%)	450 (19.54%)
Fajiang and Wang (2012)	January 4, 2000–April 30, 2010	January 2000–March 2009	April 2009–April 2010
Kumar and Murugan (2013)	3698	2588 (70%)	1110 (30%)
Ticknor (2013)	734	80%	20%
Hagenau et al. (2013)	14,3 48	50%	50%
Lin et al. (2013)	880	95%	5%
Kazem et al. (2013)	800	80%	20%
Hegazy et al. (2013)	Not stated	70%	30%
Nhu et al. (2013)	4860	60%	Testing: 20% Validation: 20%
Porshnev et al. (2013)	755,000,101	84%	16%
Sheta et al. (2013)		50%	50%
Adebiyi et al. (2014a)	5680	Not stated	Not stated
Bisoi and Dash (2014)	905	500 (55.25%)	400 (44.2%)
Ding et al. (2014)	67,826	54,776	13,050
Fang et al. (2014)	238	200 (84%)	38 (16%)

Table 6 (continued)

Reference	Total dataset	Training dataset	Testing dataset
Yetis et al. (2014)	Not stated	70%	30%
Ming et al. (2014)	1500 trading days	67.2% (1008 trading days)	12.53% (188 trading days)
Pulido et al. (2014)		70%	30%
Wanjawa and Muchemi (2014)		80% (January 2, 2008–December 31, 2011)	20% (January 2, 2012–March 31, 2012.)
Yoosin et al. (2014)	78,216	41,978 (54%)	36,238 (46%)
Chen et al. (2014)	361 day	261 days (72%)	100 days (28%)
Li et al. (2014b)	11 months	9 months (81.8%)	3 months (18.2%)
Rather et al. (2014)		50%	50%
Sun et al. (2016)		43%	Testing: 36% Validation: 21%
Göçken et al. (2016)	4160	4000 (96%)	160 (4%)
Dash and Dash (2016)	1003	668 (66.6%)	335 (33.4%)
Kumar et al. (2016)		80%	20%
Chong et al. (2017)	73,041	80%	20%
Kraus and Feuerriegel (2017)		80%	20%
Zhou et al. (2017)	December 1st 2014–December 7th 2015	December 1st 2014–September 16th 2015 (80%)	September 17th–December 7th in 2015 (20%)
García et al. (2018)	4345	December 1999–November 2014 (87%)	December 2014–30th January 2017 (13%)
Dosdoğru et al. (2018)	300	240 (80%)	60 (20%)

Table 7 Abbreviated technical indicators and stock market

S/N	Technical indicators	Abbreviation	Stock exchange	Abbreviation
1.	Simple moving average	MA	Taiwan stock exchange index	TSE
2.	Days bias	BIAS	Tehran stock exchange indexes	TEPIX
3.	Relative strength index	RSI	Korea Composite Stock Price Index	KOSPI
4.	Days stochastic line	KD	Reserve Bank of India	RBI
5.	Moving average convergence and divergence	MACD	Sao Paulo stock exchange	SPSE
6.	Days psychological line	PSY	Bombay stock exchange	BSE-SENSEX
7.	Volume	V	Shanghai stock exchange index	SSE
8.	Exchange rate	ER	Shanghai stock exchange composite index	SSEC
9.	Stock closing price	SCP	Shenzhen stock exchange component index	SZSC
10.	Stock opening price	SOP	Istanbul stock exchange	ISE
11.	Typical price	TP	National Bank of Greece	NBG
12.	Chaikin money flow indicator	CMI	Nigerian stock exchange	NSE
13.	Stochastic momentum index	SMI	New York stock exchange	NYSE
14.	Bollienger bands	BB	Taiwan capitalization weighted stock index	TAIEX
15.	Percent mean absolute deviation	PMAD	Shenzhen integrated index	SZII
16.			Dow Jones industrial average index	DJIAI
17.			India national stock exchange	INSE
18.			Asian stock market indexes	ASMI
19.			Hanoi stock exchange	HSE
20.			Hong Kong stock exchange	HKSE

References

Abhishek K et al (2012) A stock market prediction model using artificial neural network. In: Third international conference on computing communication & networking technologies (ICCCNT), pp 1–5. https://doi.org/10.1109/icccnt.2012.6396089

Adam AM, Tweneboah G (2008) Macroeconomic factors and stock market movement: evidence from Ghana. University of Leicester, Leicester. https://doi.org/10.2139/ssrn.1289842

Adebayo AD, Adekoya AF, Rahman TM (2017) Predicting stock trends using Tsk-fuzzy rule based system. JENRM 4(7):48–55

Adebiyi AA et al (2012) Stock price prediction using neural network with hybridized market indicators. J Emerg Trends Comput Inf Sci 3(1):1–9

Adebiyi AA, Adewumi AO, Ayo CK (2014a) Comparison of ARIMA and artificial neural networks models for stock price prediction. J Appl Math 2014:9–11. https://doi.org/10.1155/2014/614342

Adebiyi AA, Adewumi AO, Ayo CK (2014) Stock price prediction using the ARIMA model. In: Proceedings—UKSim-AMSS 16th international conference on computer modelling and simulation, UKSim 2014, pp 106–112. https://doi.org/10.1109/uksim.2014.67

Adusei M (2014) The inflation-stock market returns nexus: evidence from the Ghana stock exchange. J Econ Int Finance 6(2):38–46. https://doi.org/10.5958/2321-5763.2016.00010.X

Agarwal P et al (2017) Stock market price trend forecasting using machine learning. Int J Res Appl Sci Eng Technol: IJRASET 5(IV):1673–1676

Agrawal S, Jindal M, Pillai GN (2010) Momentum analysis based stock market prediction using adaptive neuro-fuzzy inference system (ANFIS). In: International multiconference of engineers and computer scientists (IMECS). Hong Kong

Agrawal JG, Chourasia VS, Mittra AK (2013) State-of-the-art in stock prediction techniques. Int J Adv Res Electr Electron Instrum Eng 2(4):1360–1366

Ahmadi E et al (2018) New efficient hybrid candlestick technical analysis model for stock market timing on the basis of the support vector machine and heuristic algorithms of imperialist competition and genetic. Expert Syst Appl 94(April):21–31. https://doi.org/10.1016/j.eswa.2017.10.023

Akinwale Adio T, Arogundade OT, Adekoya AF (2009) Translated Nigeria stock market prices using artificial neural network for effective prediction. J Theor Appl Inf Technol. pp 36–43. http://jatit.org/volumes/research-papers/Vol9No1/6Vol9No1.pdf

Almeida L, Lorena A, De Oliveira I (2010) Expert systems with applications a method for automatic stock trading combining technical analysis and nearest neighbor classification. Expert Syst Appl 37(10):6885–6890. https://doi.org/10.1016/j.eswa.2010.03.033

Anbalagan T, Maheswari SU (2014) Classification and prediction of stock market index based on fuzzy metagraph. Procedia Comput Sci 47(C):214–221. https://doi.org/10.1016/j.procs.2015.03.200

Ansari T et al (2010) Sequential combination of statistics, econometrics and adaptive neural-fuzzy interface for stock market prediction. Expert Syst Appl 37(7):5116–5125. https://doi.org/10.1016/j.eswa.2009.12.083

Anthony J, Maurice L, Eshwar S (2011) Predictive ability of the interest rate spread using neural networks. Procedia Comput Sci 6:207–212. https://doi.org/10.1016/j.procs.2011.08.039

Argiddi VR, Apte SS (2012) Future trend prediction of Indian IT stock market using association rule mining of transaction data. Int J Comput Appl 39(10):30–34. https://doi.org/10.5120/4858-7132

Asadi S et al (2012) Hybridization of evolutionary Levenberg–Marquardt neural networks and data pre-processing for stock market prediction. Knowl Based Syst 35:245–258. https://doi.org/10.1016/j.knosys.2012.05.003

Atsalakis GS, Dimitrakakis EM, Zopounidis CD (2011) Elliott wave theory and neuro-fuzzy systems, in stock market prediction: the WASP system. Expert Syst Appl 38(8):9196–9206. https://doi.org/10.1016/j.eswa.2011.01.068

Ayub A (2018) Volatility transmission from oil prices to agriculture commodity and stock market in Pakistan. Capital University of Science and Technology, Islamabad

Babu MS, Geethanjali N, Satyanarayana PB (2012) Clustering approach to stock market prediction. Int J Adv Netw Appl 03(04):1281–1291

Baker M, Wurgler J (2007) Investor sentiment in the stock market. http://www.nber.org/papers/w13189

Ballings M et al (2015) Evaluating multiple classifiers for stock price direction prediction. Expert Syst Appl 42(20):7046–7056. https://doi.org/10.1016/j.eswa.2015.05.013

Bhagwant C et al (2014) Stock market prediction using artificial neural networks. Int J Comput Sci Inf Technol 5(1):904–907. https://doi.org/10.4028/www.scientific.net/AEF.6-7.1055

Bisoi R, Dash PK (2014) A hybrid evolutionary dynamic neural network for stock market trend analysis and prediction using unscented Kalman filter. Appl Soft Comput J 19:41–56. https://doi.org/10.1016/j.asoc.2014.01.039

Boachie MK et al (2016) Interest rate, liquidity and stock market performance in Ghana. Int J Account Econ Stud 4(1):46. https://doi.org/10.14419/ijaes.v4i1.5990

Bollen J, Mao H, Zeng X-J (2011) Twitter mood predicts the stock market. J Comput Sci 2(1):1–8. https://doi.org/10.1016/j.jocs.2010.12.007

Bordino I et al (2012) Web search queries can predict stock market volumes. PLoS ONE. https://doi.org/10.1371/journal.pone.0040014

Boyacioglu MA, Avci D (2010) Adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: the case of the Istanbul stock exchange. Expert Syst Appl 37(12):7908–7912. https://doi.org/10.1016/j.eswa.2010.04.045

Chakravarty S, Dash PK (2012) A PSO based integrated functional link net and interval type-2 fuzzy logic system for predicting stock market indices. Appl Soft Comput J 12(2):931–941. https://doi.org/10.1016/j.asoc.2011.09.013

Chan K et al (2017) What do stock price levels tell us about the firms? J Corp Finance 46:34–50. https://doi.org/10.1016/j.jcorpfin.2017.06.013

Chang SV et al (2013) A review of stock market prediction with artificial neural network (ANN). In: 2013 IEEE international conference on control system, computing and engineering, pp 477–482. https://doi.org/10.1109/iccsce.2013.6720012

Checkley MS, Higón DA, Alles H (2017) The hasty wisdom of the mob: how market sentiment predicts stock market behavior. Expert Syst Appl 77:256–263. https://doi.org/10.1016/j.eswa.2017.01.029

Chen C et al (2014) Exploiting social media for stock market prediction with factorization machine. In: 2014 IEEE/WIC/ACM international joint conference on web intelligence and intelligent agent technology—workshops, WI-IAT 2014, pp 49–56. https://doi.org/10.1109/wi-iat.2014.91

Chen Y, Hao Y (2017) A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction. Expert Syst Appl 80:340–355. https://doi.org/10.1016/j.eswa.2017.02.044

Chen R, Lazer M (2013) Sentiment analysis of Twitter feeds for the prediction of stock market movement. Stanf Educ 25:1–5. https://doi.org/10.1016/j.ufug.2017.05.003

Chong E, Han C, Park FC (2017) Deep learning networks for stock market analysis and prediction: methodology, data representations, and case studies. Expert Syst Appl 83:187–205. https://doi.org/10.1016/j.eswa.2017.04.030

Coyne S, Madiraju P, Coelho J (2017) Forecasting stock prices using social media analysis. In: IEEE 15th international conference on big data intelligence and computing and cyber science and technology congress. IEEE Computer Society, pp 1031–1038. https://doi.org/10.1109/dasc-picom-datacom-cyberscitec.2017.169

Dase RK, Pawar DD (2010) Application of artificial neural network for stock market predictions: a review of literature. Int J Mach Intell 2(2):14–17

Dash R, Dash PK (2016) Efficient stock price prediction using a self evolving recurrent neuro-fuzzy inference system optimized through a modified technique. Expert Syst Appl 52:75–90. https://doi.org/10.1016/j.eswa.2016.01.016

de Araújo RA (2010) A quantum-inspired evolutionary hybrid intelligent approach for stock market prediction. Int J Intell Comput Cybern 3(1):24–54

de Araújo RA, Ferreira TAE (2013) A morphological-rank-linear evolutionary method for stock market prediction. Inf Sci 237:3–17. https://doi.org/10.1016/j.ins.2009.07.007

de Oliveira FA, Nobre CN, Zárate LE (2013) 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 40(18):7596–7606. https://doi.org/10.1016/j.eswa.2013.06.071

Demyanyk Y, Hasan I (2010) Financial crises and bank failures: a review of prediction methods. Omega. https://doi.org/10.1016/j.omega.2009.09.007

Ding X et al (2014) Using structured events to predict stock price movement: an empirical investigation. In: The 2014 conference on empirical methods in natural language processing (EMNLP). Association for Computational Linguistics, Doha, pp 1415–1425. https://doi.org/10.3115/v1/d14-1148

Dondio P (2013) Stock market prediction without sentiment analysis: using a web-traffic based classifier and user-level analysis. In: Proceedings of the annual hawaii international conference on system sciences, pp 3137–3146. https://doi.org/10.1109/hicss.2013.498

Dosdoğru AT et al (2018) Assessment of hybrid artificial neural networks and metaheuristics for stock market forecasting. Ç. Ü. Sosyal Bilimler Enstitüsü Dergisi 24(1):63–78

Dunne M (2015) Stock market prediction. University College Cork, Cork

Dutta A, Bandopadhyay G, Sengupta S (2012) Prediction of stock performance in the indian stock market using logistic regression. Int J Bus Inf 7(1):105–136

Enke D, Mehdiyev N (2013) Stock market prediction using a combination of stepwise regression analysis, differential evolution-based fuzzy clustering, and a fuzzy inference neural network. Intell Autom Soft Comput 19(4):636–648. https://doi.org/10.1080/10798587.2013.839287

Enke D, Grauer M, Mehdiyev N (2011) Stock market prediction with multiple regression, fuzzy type-2 clustering and neural networks. Procedia Comput Sci 6:201–206. https://doi.org/10.1016/j.procs.2011.08.038

Ertuna L (2016) Stock market prediction using neural network time series forecasting (May). https://doi.org/10.13140/rg.2.1.1954.1368

Esfahanipour A, Aghamiri W (2010) Adapted neuro-fuzzy inference system on indirect approach TSK fuzzy rule base for stock market analysis. Expert Syst Appl 37(7):4742–4748. https://doi.org/10.1016/j.eswa.2009.11.020

Fajiang L, Wang J (2012) Fluctuation prediction of stock market index by Legendre neural network with random time strength function. Neurocomputing 83:12–21. https://doi.org/10.1016/j.neucom.2011.09.033

Fama EF (1965) Random walks in stock market prices. Financ Anal J 21:55–59

Fama EF (1970) Efficient capital markets: a review of theory and empirical work. J Finance 25:383–417

Fang Y et al (2014) Improving the genetic-algorithm-optimized wavelet neural network for stock market prediction. In: International joint conference on neural networks. IEEE, Beijing, pp 3038–3042. https://doi.org/10.1109/ijcnn.2014.6889969

Gaius KD (2015) Assessing the performance of active and passive trading on the Ghana stock exchange. University of Ghana, Accra

García F, Guijarro F, Oliver J (2018) Hybrid fuzzy neural network to predict price direction in the German DAX-30 index. Technol Econ Dev Econ 24(6):2161–2178

Geva T, Zahavi J (2014) Empirical evaluation of an automated intraday stock recommendation system incorporating both market data and textual news. Decis Support Syst 57(1):212–223. https://doi.org/10.1016/j.dss.2013.09.013

Ghaznavi A, Aliyari M, Mohammadi MR (2016) Predicting stock price changes of tehran artmis company using radial basis function neural networks. Int Res J Appl Basic Sci 10(8):972–978

Göçken M et al (2016) Integrating metaheuristics and artificial neural networks for improved stock price prediction. Expert Syst Appl 44:320–331. https://doi.org/10.1016/j.eswa.2015.09.029

Goel SK, Poovathingal B, Kumari N (2016) Applications of neural networks to stock market prediction. Int Res J Eng Technol: IRJET 03(05):2192–2197

Gupta A, Sharma SD (2014) Clustering-classification based prediction of stock market future prediction. Int J Comput Sci Inf Technol 5(3):2806–2809

Guresen E, Kayakutlu G, Daim TU (2011) Using artificial neural network models in stock market index prediction. Expert Syst Appl 38(8):10389–10397. https://doi.org/10.1016/j.eswa.2011.02.068

Gyan MK (2015) Factors influencing the patronage of stocks, Knu. Kwame Nkrumah University of Science & Technology (KNUST), Kumasi

Hadavandi E, Shavandi H, Ghanbari A (2010) Knowledge-based systems integration of genetic fuzzy systems and artificial neural networks for stock price forecasting. Knowl Based Syst 23(8):800–808. https://doi.org/10.1016/j.knosys.2010.05.004

Hagenau M, Liebmann M, Neumann D (2013) Automated news reading: stock price prediction based on financial news using context-capturing features. Decis Support Syst 55(3):685–697. https://doi.org/10.1016/j.dss.2013.02.006

Hassan MR et al (2013) A HMM-based adaptive fuzzy inference system for stock market forecasting. Neurocomputing 104:10–25. https://doi.org/10.1016/j.neucom.2012.09.017

Hegazy O, Soliman OS, Salam MA (2013) A machine learning model for stock market prediction. Int J Comput Sci Telecommun 4(12):17–23

Henriksson A et al (2016) Ensembles of randomized trees using diverse distributed representation of clinical events. BMC Med Inf Decis Mak 16(2):69

Ibrahim SO (2017) Forecasting the volatilities of the Nigeria stock market prices. CBN J Appl Stat 8(2):23–45

Javed K, Gouriveau R, Zerhouni N (2014) SW-ELM: a summation wavelet extreme learning machine algorithm with a priori parameter initialization. Neurocomputing 123:299–307. https://doi.org/10.1016/j.neucom.2013.07.021

Jianfeng S et al (2014) Exploiting social relations and sentiment for stock prediction. In: Conference on empirical methods in natural language processing (EMNLP). Association for Computational Linguistics, Doha, pp 1139–1145. https://doi.org/10.1080/00378941.1956.10837773

Ju-Jie W et al (2012) Stock index forecasting based on a hybrid model. Omega 40(6):758–766. https://doi.org/10.1016/j.omega.2011.07.008
Kannan KS et al (2010) Financial stock market forecast using data mining techniques. In: International multiconference of engineers and computer scientists (IMECS)
Kara Y, Acar Boyacioglu M, Baykan ÖK (2011) 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 38(5):5311–5319. https://doi.org/10.1016/j.eswa.2010.10.027
Kazem A et al (2013) Support vector regression with chaos-based firefly algorithm for stock market price forecasting. Appl Soft Comput J 13(2):947–958. https://doi.org/10.1016/j.asoc.2012.09.024
Kearney C, Liu S (2014) Textual sentiment in finance: a survey of methods and models. Int Rev Financ Anal 33(Cc):171–185. https://doi.org/10.1016/j.irfa.2014.02.006
Khan HZ, Alin ST, Hussain A (2011) Price prediction of share market using artificial neural network "ANN". Int J Comput Appl 22(2):42–47. https://doi.org/10.5120/2552-3497
Kraus M, Feuerriegel S (2017) Decision support from financial disclosures with deep neural networks and transfer learning. Decis Support Syst 104:38–48. https://doi.org/10.1016/j.dss.2017.10.001
Krollner B, Vanstone B, Finnie G (2010a) Financial time series forecasting with machine learning techniques: a survey. In: European symposium on artificial neural networks: computational and machine learning. Bond University, Bruges, pp 25–30
Krollner B, Vanstone B, Finnie G (2010b) Financial time series forecasting with machine learning techniques: a survey. http://epublications.bond.edu.au/infotech_pubs/110
Kumar DA, Murugan S (2013) Performance analysis of Indian stock market index using neural network time series model. In: Proceedings of the 2013 international conference on pattern recognition, informatics and mobile engineering, PRIME 2013, pp 72–78. https://doi.org/10.1109/icprime.2013.6496450
Kumar M, Thenmozhi M (2006) Forecasting stock index movement: a comparison of support vector machines and random forest. In Indian Institute of capital markets 9th capital markets conference paper.
Kumar D, Meghwani SS, Thakur M (2016) Proximal support vector machine based hybrid prediction models for trend forecasting in financial markets. J Comput Sci 17:1–13. https://doi.org/10.1016/j.jocs.2016.07.006
Kuwornu JKM, Victor O-N (2011) Macroeconomic variables and stock market returns: full information maximum likelihood estimation. Res J Finance Account 2(4):49–64
Kwofie C, Ansah RK (2018) A study of the effect of inflation and exchange rate on stock market returns in Ghana. Int J Math Math Sci. https://doi.org/10.1155/2018/7016792
Laboissiere LA, Fernandes RAS, Lage GG (2015) Maximum and minimum stock price forecasting of Brazilian power distribution companies based on artificial neural networks. Appl Soft Comput J 35:66–74. https://doi.org/10.1016/j.asoc.2015.06.005
Lahmiri S (2011) A Comparison of PNN and SVM for stock market trend prediction using economic and technical information. Int J Comput Appl 29(3):975–8887
Li Q et al (2015) Tensor-based learning for predicting stock movements. In: Twenty-ninth AAAI conference on artificial intelligence-2015, pp 1784–1790. https://doi.org/10.1073/pnas.0601853103
Li Q, Wang T, Gong Q et al (2014a) Media-aware quantitative trading based on public Web information. Decis Support Syst 61(1):93–105. https://doi.org/10.1016/j.dss.2014.01.013
Li Q, Wang T, Li P et al (2014b) The effect of news and public mood on stock movements. Inf Sci 278:826–840. https://doi.org/10.1016/j.ins.2014.03.096
Li X, Huang X et al (2014c) Enhancing quantitative intra-day stock return prediction by integrating both market news and stock prices information. Neurocomputing 142:228–238. https://doi.org/10.1016/j.neucom.2014.04.043
Li X, Xie H et al (2014d) News impact on stock price return via sentiment analysis. Knowl-Based Syst 69(1):14–23. https://doi.org/10.1016/j.knosys.2014.04.022
Lin Z (2018) Modelling and forecasting the stock market volatility of SSE composite index using GARCH models. Future Gener Comput Syst 79:960–972. https://doi.org/10.1016/j.future.2017.08.033
Lin Y, Guo H, Hu J (2013) An SVM-based approach for stock market trend prediction. In: Proceedings of the international joint conference on neural networks. https://doi.org/10.1109/ijcnn.2013.6706743
Liu L et al (2015) A social-media-based approach to predicting stock comovement. Expert Syst Appl 42(8):3893–3901. https://doi.org/10.1016/j.eswa.2014.12.049
Luo F, Wu J, Yan K (2010) A novel nonlinear combination model based on support vector machine for stock market prediction. In: Jinan C (ed) World congress on intelligent control and automation. IEEE, Piscataway, pp 5048–5053

Maknickiene N, Lapinskaite I, Maknickas A (2018) Application of ensemble of recurrent neural networks for forecasting of stock market sentiments. Equilib Q J Econ Econ Policy 13(1):7–27. https://doi.org/10.24136/eq.2018.001

Makrehchi M, Shah S, Liao W (2013) Stock prediction using event-based sentiment analysis. In: Proceedings—2013 IEEE/WIC/ACM international conference on web intelligence, WI 2013, 1, pp 337–342. https://doi.org/10.1109/wi-iat.2013.48

Malkiel BG (1999) A random walk down Wall Street: including a life-cycle guide to personal investing. WW Norton & Company

Metghalchi M, Kagochi J, Hayes LA (2014) Contrarian technical trading rules: evidence from Nairobi stock index. J Appl Bus Res 30(3):833–846

Ming F et al (2014) Stock market prediction from WSJ: text mining via sparse matrix factorization. In: EEE international conference on data mining, ICDM, pp 430–439. https://doi.org/10.1109/icdm.2014.116

Minxia L, Zhang K (2014) A hybrid approach combining extreme learning machine and sparse representation for image classification. Eng Appl Artif Intell 27:228–235. https://doi.org/10.1016/j.engappai.2013.05.012

Mittal A, Goel A (2012) Stock prediction using twitter sentiment analysis. Standford University, CS229, (June). https://doi.org/10.1109/wi-iat.2013.48

Mohapatra P, Raj A (2012) Indian stock market prediction using differential evolutionary neural network model. Int J Electron Commun Comput Technol: IJECCT 2(4):159–166

Murekachiro D (2016) A review of artificial neural networks application to stock market predictions. Netw Complex Syst 6(4):2010–2013

Naeini MP, Taremian H, Hashemi HB (2010) Stock market value prediction using neural networks. IEEE, Piscataway, pp 132–136

Nair BB et al (2010) Stock market prediction using a hybrid neuro-fuzzy system. In: International conference on advances in recent technologies in communication and computing, India, pp 243–247. https://doi.org/10.1109/artcom.2010.76

Nair BB, Mohandas VP, Sakthivel NR (2010) A decision tree-rough set hybrid system for stock market trend prediction. Int J Comput Appl 6(9):1–6

Nassirtoussi AK et al (2014) Text mining for market prediction: a systematic review. Expert Syst Appl 41(16):7653–7670. https://doi.org/10.1016/j.eswa.2014.06.009

Nayak RK, Mishra D, Rath AK (2015) A Naïve SVM-KNN based stock market trend reversal analysis for Indian benchmark indices. Appl Soft Comput J 35:670–680. https://doi.org/10.1016/j.asoc.2015.06.040

Nazário RTF et al (2017) A literature review of technical analysis on stock markets. Q Rev Econ Finance 66:115–126. https://doi.org/10.1016/j.qref.2017.01.014

Neelima B, Jha CK, Saneep BK (2012) Application of neural network in analysis of stock market prediction. Int J Comput Sci Technol: IJCSET 3(4):61–68

Nhu HN, Nitsuwat S, Sodanil M (2013) Prediction of stock price using an adaptive neuro-fuzzy inference system trained by firefly algorithm. In: 2013 international computer science and engineering conference, ICSEC 2013, pp 302–307. https://doi.org/10.1109/icsec.2013.6694798

Nikfarjam A, Emadzadeh E, Muthaiyah S (2010) Text mining approaches for stock market prediction. IEEE, vol 4, pp 256–260

Nisar TM, Yeung M (2018) Twitter as a tool for forecasting stock market movements: a short-window event study. J Finance Data Sci 4(February):1–19. https://doi.org/10.1016/j.jfds.2017.11.002

Olaniyi S, Adewole K, Jimoh R (2011) Stock trend prediction using regression analysis—a data mining approach. ARPN J Syst Softw 1(4):154–157

Paik P, Kumari B (2017) Stock market prediction using ANN, SVM, ELM: a review. Ijettcs 6(3):88–94. https://doi.org/10.1038/33071

Patel J et al (2015a) Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. Expert Syst Appl 42(1):259–268. https://doi.org/10.1016/j.eswa.2014.07.040

Patel J et al (2015b) Predicting stock market index using fusion of machine learning techniques. Expert Syst Appl 42(4):2162–2172. https://doi.org/10.1016/j.eswa.2014.10.031

Pervaiz J, Masih J, Jian-Zhou T (2018) Impact of macroeconomic variables on Karachi stock market returns. Int J Econ Finance 10(2):28. https://doi.org/10.5539/ijef.v10n2p28

Perwej Y, Perwej A (2012) Prediction of the Bombay stock exchange (BSE) market returns using artificial neural network and genetic algorithm. J Intell Learn Syst Appl 04(02):108–119. https://doi.org/10.4236/jilsa.2012.42010

Pimprikar R, Ramachadran S, Senthilkumar K (2017) Use of machine learning algorithms and Twitter sentiment analysis for stock market prediction. Int J Pure Appl Math 115(6):521–526
Porshnev A, Redkin I, Shevchenko A (2013) Improving prediction of stock market indices by analyzing the psychological states of Twitter users. Financ Econ. https://doi.org/10.2139/ssrn.2368151
Prem Sankar C, Vidyaraj R, Satheesh Kumar K (2015) Trust based stock recommendation system—a social network analysis approach. In: Procedia computer science: international conference on information and communication technologies (ICICT 2014). Elsevier Masson SAS, pp 299–305. https://doi.org/10.1016/j.procs.2015.02.024
Pulido M, Melin P, Castillo O (2014) Particle swarm optimization of ensemble neural networks with fuzzy aggregation for time series prediction of the Mexican stock exchange. Inf Sci 342(May):317–329. https://doi.org/10.1007/978-3-319-32229-2_23
Rajashree D, Dash PK, Bisoi R (2014) A self adaptive differential harmony search based optimized extreme learning machine for financial time series prediction. Swarm Evol Comput 19:25–42. https://doi.org/10.1016/j.swevo.2014.07.003
Rather AM, Agarwal A, Sastry VN (2014) Recurrent neural network and a hybrid model for prediction of stock returns. Expert Syst Appl 42(8):3234–3241. https://doi.org/10.1016/j.eswa.2016.05.033
Renu IR, Christie R (2018) Fundamental analysis versus technical analysis—a comparative review. Int J Recent Sci Res 9(1):23009–23013. https://doi.org/10.24327/IJRSR
Sasan B, Azadeh A, Ortobelli S (2017) Fusion of multiple diverse predictors in stock market. Inf Fusion 36:90–102. https://doi.org/10.1016/j.inffus.2016.11.006
Shen S, Jiang H, Zhang T (2012) Stock market forecasting using machine learning algorithms. Department of Electrical Engineering, Stanford University, Stanford, CA, pp 1–5
Sheta A, Farisy H, Alkasassbehz M (2013) A genetic programming model for S&P 500 stock market prediction. Int J Control Autom 6(6):303–314. https://doi.org/10.14257/ijca.2013.6.6.29
Shobana T, Umamakeswari A (2016) A review on prediction of stock market using various methods in the field of data mining. Indian J Sci Technol 9(48):9–14. https://doi.org/10.17485/ijst/2016/v9i48/107985
Shom P Das, Padhy S (2012) Support vector machines for prediction of futures prices in Indian stock market. Int J Comput Appl 41(3):22–26. https://doi.org/10.5120/5522-7555
Si J et al (2013) Exploiting topic based twitter sentiment for stock prediction. In: The 51st annual meeting of the association for computational linguistics, vol 2(2011), pp 24–29. http://www.scopus.com/inward/record.url?eid=2-s2.0-84907356594&partnerID=tZOtx3y1
Solanki H (2013) Comparative study of data mining tools and analysis with unified data mining theory. Int J Comput Appl 75(16):23–28
Soni S (2011) Applications of ANNs in stock market prediction: a survey. In: International conference on computer information systems and industrial management applications (CISIM), vol 2, no. 3, pp 132–136. https://doi.org/10.1177/1040638713493779
Sorto M, Aasheim C, Wimmer H (2017) Feeling the stock market: a study in the prediction of financial markets based on news sentiment. In: Hatzivassiloglou V, Klavans J, Eskin E (eds) Southern association for information systems conference. St. Simons Island, GA, USA, p. 19. http://aisel.aisnet.org/sais2017%0Ahttp://aisel.aisnet.org/sais2017/30%0Ahttp://aisel.aisnet.org/sais2017%0Ahttp://aisel.aisnet.org/sais2017/30
Stanković J, Marković I, Stojanović M (2015) Investment strategy optimization using technical analysis and predictive modeling in emerging markets. Procedia Econ Finance 19(15):51–62. https://doi.org/10.1016/S2212-5671(15)00007-6
Su CH, Cheng CH (2016) A hybrid fuzzy time series model based on ANFIS and integrated nonlinear feature selection method for forecasting stock. Neurocomputing 205:264–273. https://doi.org/10.1016/j.neucom.2016.03.068
Suhaibu I, Harvey SK, Amidu M (2017) The impact of monetary policy on stock market performance: evidence from twelve (12) African countries. Res Int Bus Finance 42(12):1372–1382. https://doi.org/10.1016/j.ribaf.2017.07.075
Sun A, Lachanski M, Fabozzi FJ (2016) Trade the tweet: social media text mining and sparse matrix factorization for stock market prediction. Int Rev Financ Anal 48:272–281. https://doi.org/10.1016/j.irfa.2016.10.009
Sureshkumar KK, Elango NM (2011) An efficient approach to forecast Indian stock market price and their performance analysis. Int J Comput Appl 34(5):44–49. https://doi.org/10.1196/annals.1364.016
Suthar BA, Patel RH, Parikh MS (2012) A comparative study on financial stock market prediction models. Int J Eng Sci: IJES 1(2):188–191. https://doi.org/10.1007/BF00629127

Talib R et al (2016) Text mining-techniques applications and issues. Int J Adv Comput Sci Appl 7(11):414–418

Thanh D Van, Minh Hai N, Hieu DD (2018) Building unconditional forecast model of stock market indexes using combined leading indicators and principal components: application to Vietnamese stock market. Indian J Sci Technol 11(2):1–13. https://doi.org/10.17485/ijst/2018/v11i2/104908

Ticknor JL (2013) A Bayesian regularized artificial neural network for stock market forecasting. Expert Syst Appl 40(14):5501–5506. https://doi.org/10.1016/j.eswa.2013.04.013

Tsai C-F, Hsiao Y-C (2010) Combining multiple feature selection methods for stock prediction: union, intersection, and multi-intersection approaches. Decis Support Syst 50(1):258–269. https://doi.org/10.1016/j.dss.2010.08.028

Tsai MF, Wang C-J (2017) On the risk prediction and analysis of soft information in finance reports. Eur J Oper Res 257(1):243–250. https://doi.org/10.1016/j.ejor.2016.06.069

Tsaurai K (2018) What are the determinants of stock market development in emerging markets? Acad Account Financ Stud J 22(2):1–11

Tziralis G, Tatsiopoulos I (2007) Prediction markets: an extended literature review. J Predict Mark 1:75–91

Umoru D, Nwokoye GA (2018) FAVAR analysis of foreign investment with capital market predictors: evidence on Nigerian and selected African stock exchanges. Acad J Econ Stud 4(1):12–20

Uysal AK, Gunal S (2014) The impact of preprocessing on text classification. Inf Process Manage 50:104–112

Vaisla SK, Bhatt KA (2010) An analysis of the performance of artificial neural network technique for stock market forecasting. Int J Comput Sci Eng 02(06):2104–2109

Vu T-T et al (2012) An experiment in integrating sentiment features for tech stock prediction in Twitter. In: Workshop on information extraction and entity analytics on social media data, pp 23–38. http://www.aclweb.org/anthology/W12-5503

Wang Y (2013) Stock price direction prediction by directly using prices data: an empirical study on the KOSPI and HSI, pp 1–13. https://doi.org/10.1504/ijbidm.2014.065091

Wang L, Qiang W (2011) Stock market prediction using artificial neural networks based on HLP. In: Proceedings—2011 3rd international conference on intelligent human-machine systems and cybernetics, IHMSC 2011, vol 1, pp 116–119. https://doi.org/10.1109/ihmsc.2011.34

Wanjawa BW (2016) Predicting future Shanghai stock market price using ANN in the period 21 Sept 2016 to 11 Oct 2016

Wanjawa BW, Muchemi L (2014) ANN model to predict stock prices at stock exchange markets. Nairobi

Wei LY (2016) A hybrid ANFIS model based on empirical mode decomposition for stock time series forecasting. Appl Soft Comput J 42:368–376. https://doi.org/10.1016/j.asoc.2016.01.027

Wei L-Y, Chen T-L, Ho T-H (2011) A hybrid model based on adaptive-network-based fuzzy inference system to forecast Taiwan stock market. Expert Syst Appl 38(11):13625–13631. https://doi.org/10.1016/j.eswa.2011.04.127

Wensheng D, Wu JY, Lu CJ (2012) Combining nonlinear independent component analysis and neural network for the prediction of Asian stock market indexes. Expert Syst Appl 39(4):4444–4452. https://doi.org/10.1016/j.eswa.2011.09.145

Xi L et al (2014) A new constructive neural network method for noise processing and its application on stock market prediction. Appl Soft Comput J 15:57–66. https://doi.org/10.4171/RLM/692

Yeh C-Y, Huang C-W, Lee S-J (2011) A multiple-kernel support vector regression approach for stock market price forecasting. Expert Syst Appl 38(3):2177–2186. https://doi.org/10.1016/j.eswa.2010.08.004

Yetis Y, Kaplan H, Jamshidi M (2014) Stock market prediction using artificial neural network. In: World Automation Congress. ISI Press, pp 1–5. https://doi.org/10.5120/17399-7959

Yifan L et al (2017) Stock volatility prediction using recurrent neural networks with sentiment analysis. https://doi.org/10.1007/978-3-319-60042-0_22

Yoosin K, Seung RJ, Ghani I (2014) Text opinion mining to analyze news for stock market prediction. Int J Adv Soft Comput Appl 6(1–13):44. https://doi.org/10.1016/S0399-077X(16)30365-1

Yu H, Liu H (2012) Improved stock market prediction by combining support vector machine and empirical mode decomposition. In: 2012 5th international symposium on computational intelligence and design, ISCID 2012, pp 531–534. https://doi.org/10.1109/iscid.2012.138

Zhang X, Fuehres H, Gloor PA (2011) Predicting stock market indicators through Twitter "I hope it is not as bad as I fear". Procedia Soc Behav Sci 26(2007):55–62. https://doi.org/10.1016/j.sbspro.2011.10.562

Zhang X et al (2014) A causal feature selection algorithm for stock prediction modeling. Neurocomputing 142:48–59. https://doi.org/10.1016/j.neucom.2014.01.057

Zhang X et al (2017) Improving stock market prediction via heterogeneous information fusion. Knowl Based Syst 143:236–247. https://doi.org/10.1016/j.knosys.2017.12.025

Zhou Z, Xu K, Zhao J (2017) Tales of emotion and stock in China: volatility, causality and prediction. https://doi.org/10.1007/s11280-017-0495-4
Zhou X et al (2018) Stock market prediction on high frequency data using generative adversarial nets. Math Probl Eng 2018:1–12. https://doi.org/10.1155/2018/4907423