A survey on fundamental and technical analysis used in stock market prediction

[1]Shobhita Singh, [2]Dr. Divya Khanna

RIMT University, Mandi Gobindgarh, Punjab, India

Abstract : In any rising and prospering economy, every stock market investment aims to increase profit while reducing associated investment risk. As a consequence, multiple studies utilizing different soft-computing methodologies and algorithms on stock-market prediction using technical or fundamental analysis or both have been conducted. Fundamental analysis is the process of determining a company's fair market value by examining all the company's components, as well as the industry, market, and local and global environment, with the goal of long-term investment. Technical analysis looks for patterns in data, such as historical returns and price movements, that may be utilized to anticipate future price movement for securities and the market for short-term active traders. When evaluating a company's growth and profitability potential, investors and analysts often combine fundamental, technical, and quantitative analysis. The purpose of this study is to conduct a systematic and critical examination of about fifty relevant research articles published in academic journals over a seven-year period (2015–2021) around machine and deep learning-based stock market prediction. Three categories of techniques are identified in these publications: technical, fundamental, and combination analysis through input data types utilized. 50% of the reviewed literature used quantitative data (structured data) generally contain historical stock data (open, close, high, low, volume) employed in technical analysis. 24% is qualitative data (unstructured data) normally includes sentimental analysis, tweets, blogs, social media etc. The combination analysis 26% reviewed literature uses both quantitative and qualitative data.

Keywords: stock-market prediction, fundamental analysis, technical analysis, machine-learning, quantitative data, qualitative data.

1. Introduction

Individuals may purchase and sell publicly listed business shares on the stock market. It provides an effective stock exchange platform. These financial transactions take happen on regulated over-the-counter (OTC) markets or on established formal exchanges (physical or electronic). The National Stock Exchange (NSE) and the Bombay Stock Exchange (BSE) are India's two major stock exchanges, whereas the New York Stock Exchange (NYSE) and Nasdaq are the United States' two largest stock markets. Investments in the stock market are typically guided by some kind of forecasting approach [1]. The main two approaches used in investment strategies using various machine learning techniques for stock market prediction are: Fundamental analysis and Technical analysis.

The majority of investors interested in making long-term investment decisions begin by doing a fundamental analysis of a company, a particular stock, or the market as a whole. Basic analysis is the process of evaluating all aspects of a business or market in order to establish the fundamental value of an asset. Intangible assets like as trademarks, patents, branding, and intellectual property are evaluated alongside tangible assets such as land, equipment, and buildings owned by a corporation. Fundamental analysis can be categorized into: Quantitative analysis (which includes Balance sheets, Profit and loss analysis, Cash flows etc.) and Qualitative analysis (which includes Management analysis strategies). The fundamental analyst's data is frequently unstructured, which presents a difficult problem. However, it has been shown to be a good predictor of stock price change occasionally [2]. **Figure 1** demonstrates the basic structure of fundamental analysis method to stock market prediction.

Alternatively, Technical analysis is the study of securities via the use of statistics. Market activity data, such as historical returns, stock prices, and trading volume, is used by analysts and investors to determine trends in the movement of assets. Technical analysts and investors think that patterns derived from previous performance data may be used to forecast future performance. Due of the short length of time required to obtain data for technical analysis, investors prefer to implement this method for short-term trading [3]. Technical analysis, when combined with fundamental research, may be an excellent tool to analyze long-term investments. **Figure 2** demonstrates the basic structure of technical analysis method to stock market prediction.

Because the stock market is extremely volatile as a result of a variety of social, economic, and political reasons, it is critical to forecast stock market trends in order to benefit. For stock market prediction, a variety of machine learning and computational intelligence techniques are used, including Artificial Neural Networks (ANN), Support Vector Machines (SVM), Decision Trees (DT), genetic algorithm and regression [4]. These strategies, when combined with fundamental and technical trading rules, can be used to create a trading system capable of forecasting the future direction of securities prices based on quantitative and qualitative analysis, historical price, and volume data [5].

The purpose of this study is to conduct a complete, systematic evaluation of prior research studies on stock market forecasting from both a fundamental and technical analyst perspective. This results in the clarification of the existing state-of-the-art and probable future directions. To summarize, this effort adds to the following corpus of knowledge:

1. Detailed study and analysis of various commonly used fundamental ratios and technical indicators for stock market prediction.
2. A well-structured assessment and review of significant work done, along with a focus on the many variables and factors that affect stock price movement.

3. Established considerations for future research areas that might address the shortcomings of present methodologies and propose solutions for improvement.

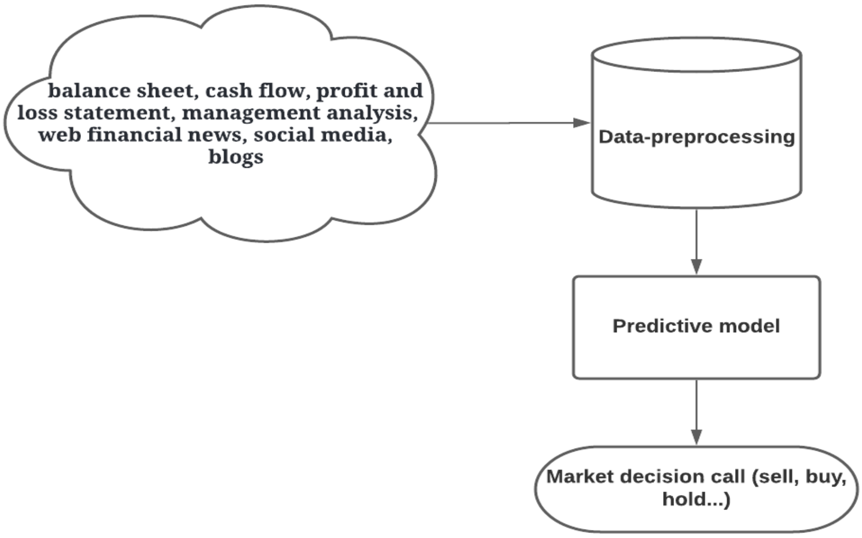

Figure 1: Approach used by Fundamental analyst

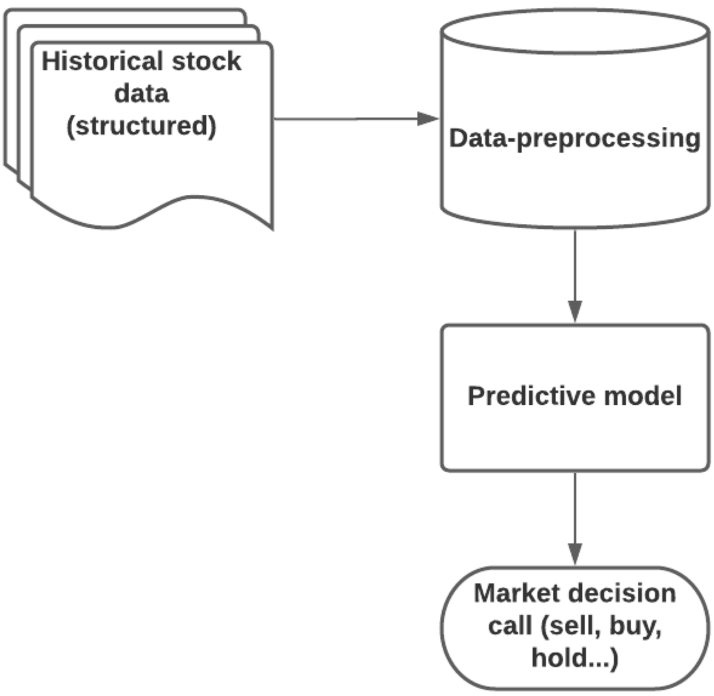

Figure 2: Approach used by Technical analyst

The remainder of this work is divided into the following sections: **Section 2** discusses market forecasting, technical and fundamental analysis, and the assessment of predictive models based on input data. **Section 3** discusses the systematic review of related research work. **Section 4** build an evaluation table of the study. **Section 5** summarizes the findings and discusses them. Additionally, **Section 6** closes this study and outlines future research directions.

2. Classification of stock market decision-making strategies

This section discusses fundamental and technical analysis in short as decision-making tools in the stock market. As shown in **Figure 1** and **Figure 2**, the fundamental and technical analysis approaches to stock market forecasting.

Fundamental analysis: Fundamental analysis is the process of determining the inherent worth of a security by weighing economic, financial, qualitative, and quantitative aspects. Macroeconomic and microeconomic variables are considered to have an effect on the value of a security [6]. Economic circumstances, industry conditions, financial conditions, and managerial competency are all examples of such variables. The primary objective of fundamental analysis should be to determine a security's intrinsic worth and compare it to the security's current stock price, so evaluating whether the asset is under- or over-valued. Basic analysis is difficult to automate due to the unstructured nature of fundamental elements. On the other side, the advent of machine learning has allowed academics to automate stock market prediction using unstructured data, resulting in increased forecast accuracy in certain circumstances. Nonetheless, fundamental research is beneficial for predicting long-term stock price movement but is ineffective for predicting short-term stock price movement [7].

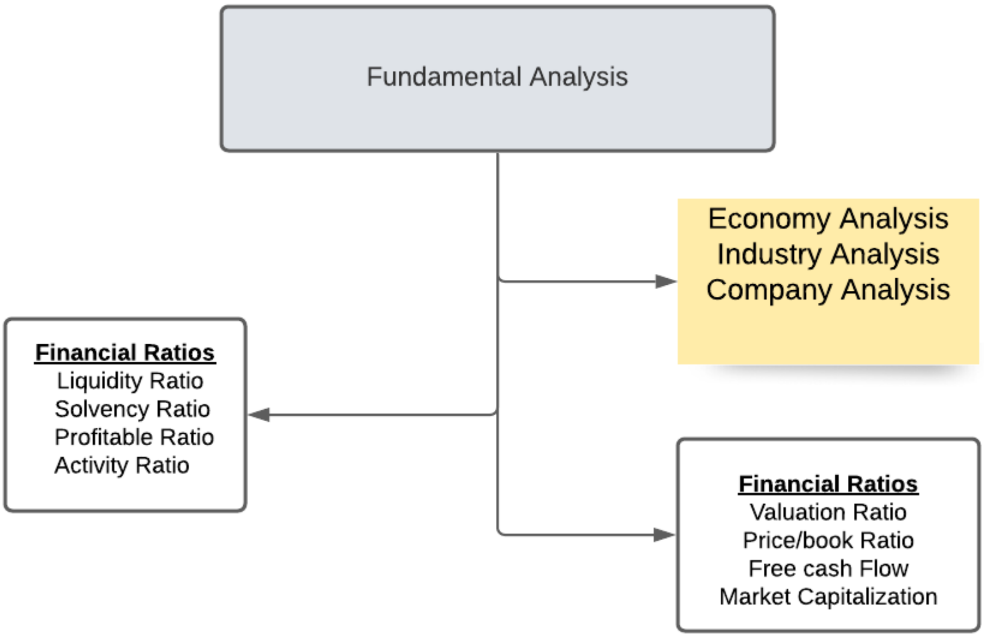

Figure 3: Various aspects of Fundamental Analysis

As seen in **Figure 3**, the fundamental analyst analyses publicly available information about the company to conduct analysis of the stock's price movement in three dimensions: the economy, its industry, and the company. Furthermore, the fundamental analyst takes into account the company's various financial ratios [8].

Liquidity Ratio: This ratio discusses the liquidity position of the company. It is also known as current ratio. An ideal current ratio is 2:1. A ratio below 1:1, where current assets are lower than current liabilities is dangerous. The formula for liquidity ratio:

Liquidity Ratio= Current assets / Current liabilities

Solvency Ratio: This ratio analysis the capital structure (loan funds VS own funds). It should be maximum of 2:1. The formula computes solvency ratio:

Solvency Ratio= Long term (non-current) debt / Shareholder's funds

Profitability Ratio: Profitability means Return On Capital Employed (ROCE). It states the proper usage of the money invested. If the firm is existing for 3-5 years on stock market index, the ROCE should be atleast 10%. The formula for profitability ratio (ROCE):

ROCE= (EBIT / Capital Employed) *100

EBIT= Earnings Before Interest and Tax. It is also called operating profit.

Capital Employed= share funds + non-current liabilities.

Activity Ratio: This ratio determines the number of times goods (inventories) were manufactured and sold. Better inventory turn over leads to better liquidity. It measures the operational efficiency. Higher the ratio, better the operational performance of the company.

Inventory Turn Over= COGS / Average inventory

COGS= Cost Of Goods Sold.

Valuation Ratio: this ratio tells how much times one is ready to pay to get a certain return. It is termed as P/E ratio. The formula computed is:

P/E Ratio= Market price per share / Earnings per share

Earnings per share= PAT (Profit After Tax) / Total number of shares.

If P/E ratio is low when compared with Industry P/E, it means the share is cheap.

Another valuation ratio is P/B ratio i.e., market price per share / book value per share.

Book value= Assets – External liabilities.

Free Cash Flows: It measures the free cash available with the company for the future operations of the company. It should always be in positive. Negative free cash flow is a bad sign. Its formula is:

Free Cash Flow= Cash from operation – Purchase of fixed assets + Sales of fixed assets.

Higher the cash flow, better is the company's position.

Market Capitalization (MC): MC quantifies the entire volume of stock traded in the market. Stocks in MC may be classified into three categories: small-cap, medium-cap, and large-cap. The formula for market capitalization is:

Market Capitalization= total number of shares * price of each share.

The above eight pointers ratios have become an important criterion for the evaluation of any company through fundamental analysis, with the increased usage of text mining techniques [9].

Technical Analysis: It is the conventional way used to predict stock market trends and price through the study of chart patterns and statistical figures which depict historical market data and technical indicators [10]. The open, close, high, low, and volume of historical stock data is preprocessed, and relevant indicators are developed and incorporated in the prediction model, as seen in **Figure 2**. Several technical indicators used during technical analysis have been explored in [11] [12] are shown in **Figure 4**.

Simple Moving Average (SMA): The term "Simple Moving Average" measures the average of most recent closing price of a stock over a specific time period. The term "moving average" refers to the fact that the stock price varies frequently, and so the moving average moves as well. The SMA is a critical indicator in technical analysis since it is often the simplest moving average to calculate. It is calculated as the average of a security's closing prices over the previous "n" periods (number of days).

$$SMA= CP_1+CP_2+CP_3+\ldots +CP_n / n$$

Relative Strength Index (RSI): In this the closing prices (highs) are compared with closing prices (lows) of previous day. RSI observation is of 14 days. It is called as a momentum indicator that measures the magnitude of recent price change. It is an oscillator (it moves to and fro within the range of 0 and 100). If it is above 60, means RSI taking support at 60 therefore buy (uptrend) and if it is below 40, means RSI facing resistance at 40 therefore sell.

$$RSI= 100 – (100 / 1+RS) = 0 \text{ to } 100$$

RS= Average gain over a specified period / Average loss over the same period.

RSI is not a trend. It is an outcome of price movements.

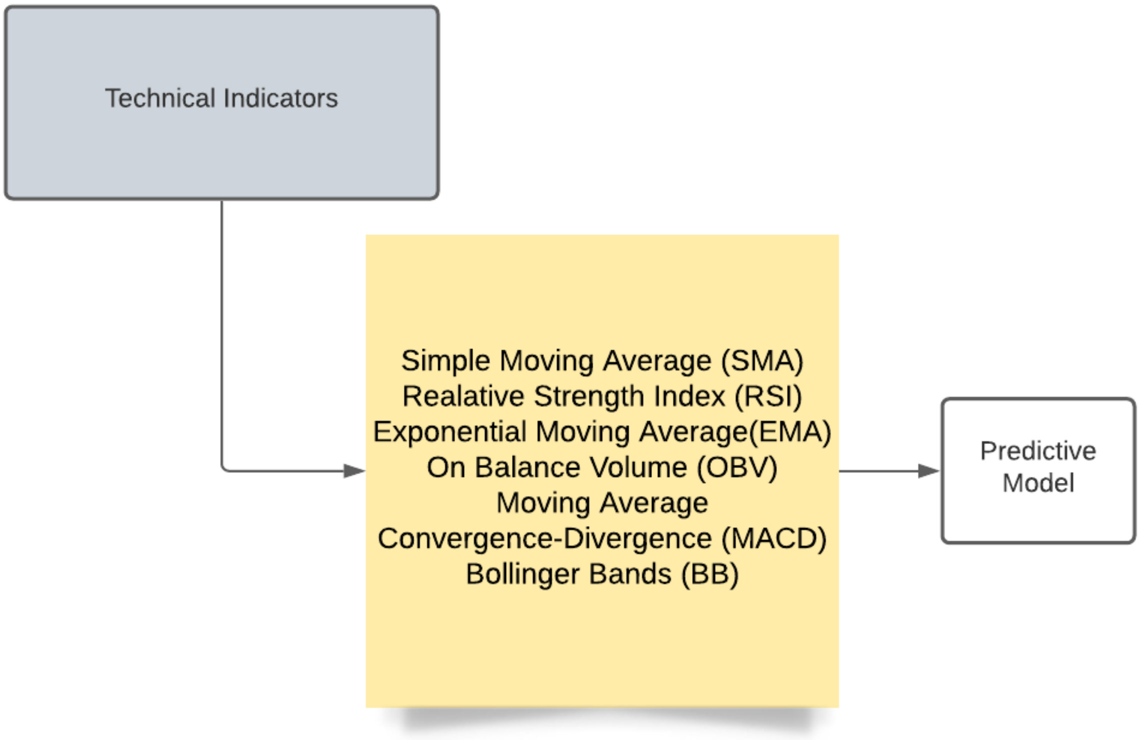

Figure 4: Technical Indicators used during Technical Analysis

Exponential Moving Average (EMA): The exponential moving average is a price computation average over a specified time period that prioritizes the most recent price data, leading it to respond more quickly to price changes. Day traders use this indicator to assist them assess the trend, direction, and strength of the market. Others depend on it as well to determine entrance and departure locations. The EMA is identical to the SMA line, except that the EMA for any particular day is determined by the EMA computations for all prior days.

EMA (current) = ((Price(current) – EMA (previous)) x Multiplier) + EMA (previous)

The time period always has a considerable impact on weighing multiplier.

On Balance Volume (OBV): The On Balance Volume (OBV) line is simply a running total of positive and negative volume. The volume of the period is positive when the closure surpasses the previous close, and it is negative when the close goes below the prior close. It was the first indication to detect whether a volume flow was positive or negative.

If the closing price is more than the previous close price, the following applies:

Current OBV is the sum of the previous OBV and the current volume.

If the closing price is less than the previous close price, the following applies:

OBV Current = OBV Previous − OBV Current

If the closing prices are identical to the previous close price, then the following applies:

OBV Current = OBV Previous (no change)

Moving Average Convergence-Divergence (MACD): This is a leading momentum indicator that attempts to forecast stock market movements by comparing short- and long-term patterns. There are three major components in MACD- MACD line (Fast Line), Signal line (Slow Line), Histogram (indicates the difference between MACD Line and Signal Line). MACD is calculated by subtracting 26 DEMA (Day Exponential Moving Average) from 12 DEMA (Day Exponential Moving Average).

Bollinger Bands (BB): It is based on standard deviation, can be taken to see support or resistance by drawing an upper and lower band along a 21-day simple moving average using two standard deviations. If the BB sequence is for a longer (atleast one month) duration and if the current candle breaks the upper BB; it's a buy call.

Quantifiable statistics cannot adequately portray the huge range of financial statuses among companies. Thus, the quality of information included in conventional news and social networking sites (unstructured data) may be used in conjunction with quantitative data to improve prediction models, particularly in current era of social media [13].

2.1 An overview of predictive models based on fundamental data and technical data

Knowing and comprehending a company's past stock data as well as its fundamental or financial data might result in a successful projection of the company's future stock price [14]. Due to developing machine learning and deep learning approaches, computers are being trained to identify the learning domain, perform tests, and finally forecast the stock market.

Figure 5 illustrates a high-level perspective of stock market prediction models, with fundamental data (unstructured data) or technical data (historical market data) serving as input datasets and specific anticipated market values serving as output.

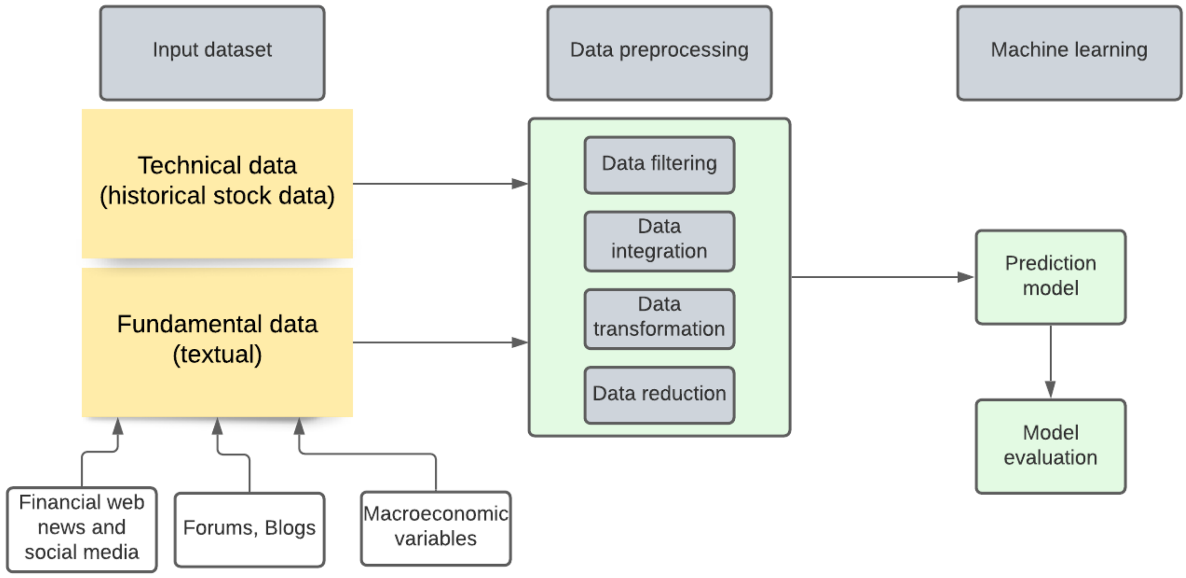

Figure 5: General overview of stock market predictive models

2.1.1 Input datasets: fundamental data and technical data

According to the research, it is critical to identify which factors aid or contribute to the forecast of other economic variables in order to produce an effective economic prediction. In general, past stock data may be classified as technical analysis dataset and qualitative data (company management, social media and investor sentiment assessments) as well as quantitative (balance sheets, profit and loss statements, free cash flows) both are treated as fundamental analysis dataset. Classification is based on the investment strategy used.

Quantitative dataset (structured data): Historical (past) stock market data used in technical analysis, such as closing price, opening price, volume traded, the day's high and low comes under quantitative dataset. It is a structured data purely based on numbers that appears (unbiased). Conventionally, only quantitative data were used for stock market prediction.

Qualitative dataset (unstructured data): These are textual records containing information on stockholders. They may be classified as financial Web news data (FWN), blogs, social media sentiment data (SMS), or macroeconomic factors (MVs), market situation, company profile, political and economic factors.

2.1.2 Data preprocessing

Data preprocessing includes data filtering (which removes unwanted noise and data inconsistency), data integration, data transformation and data reduction and normalization to achieve better accuracy or performance of predictive model.

2.1.3 Predictive models and its evaluation

It embraces a variety of machine learning models (such as DTs, SVM, ANN) that are used to predict stock market trends/price and evaluate the model through several performance metrics such as the mean absolute percentage error (MAPE), mean square error (MSE), mean absolute error (MAE) and root mean squared error (RMSE), precision and accuracy.

3. Systematic literature review

This section elaborates on critical review on literature on the basis of two major analysis: Technical analysis and Fundamental analysis utilized in various studies on stock market prediction. The section discusses fifty research articles and papers from the latest 7 years of time span and determines the conclusion about the input dataset depending on the analysis nature, the machine and deep learning techniques exercised, and performance metrics employed. The section is sub-divided into three parts constructed on the nature of analysis engaged: Technical analysis, Fundamental analysis, and Combination analysis.

3.1 Review based on Technical Analysis (Quantitative data)

Technical Indicators are critical criteria that are computed from time series stock data in order to anticipate the direction of the financial markets. They are extensively utilized by investors to determine if a market is bearish or bullish. Below are some related works based on technical analysis:

In 2019, Torres P. collected the data from Google Finance, which includes the Apple Inc. stock's Open, High, Low, and Close prices, as well as the volume of transactions for the previous 250 trading sessions were used to execute the machine learning algorithm: random forests and multilayer perceptron algorithms to produce the forecasts of closing prices [15]. In 2020, Mojtaba Nabipour [16] gathered from November 2009 to November 2019, 10 years of historical data for four stock market sectors (petroleum, diversified financials, basic metals, and non-metallic minerals) were used to implement various machine learning and deep learning techniques. The open, close, high, and low readings for each trading day are used to calculate certain technical indicators. In 2018, Elham Ahmadi et al. [3] using two hybrid methodologies, the daily stock prices, comprising the Low, High, Open, and Close values, are converted into 15 and 24 indications, respectively. In 2018, Salim Lahmiri proposed a method for assessing stock market price behavior using a variety of technical analysis metrics and a multiple predictive system in order to solve the issue of technical analysis information fusion in order to improve stock market index-level prediction [17]. In 2019, Xiao Zhong established the daily data based on the closing price of the SPDR S&P 500 on 2518 trading days between June 1, 2003, and May 31, 2013, that were employed on DNN and ANN techniques [18]. In 2020, Mehar Vijh collected 10 years financial data: Open, High, Low, and Close prices of stocks are utilized to create new variables that are then to use as inputs to a model that employs Artificial Neural Network and Random Forest methods to forecast the following day's closing price for five stock listed companies [19]. In 2020, Adebayo Felix Adekoya used the technical analysis by collecting the datasets that include daily stock data (year high, year low, previous closing price, opening price, closing price, price change, closing bid price, and closing offer price). To generalize the research, five (5) well-known technical indicators, including the simple moving average (SMA), the exponential moving average (EMA), the moving average convergence/divergence rules (MACD), the relative strength index (RSI), and the on-balance volume (OBV) were used [20]. In 2016, the data set utilized in this experiment spans the period August 19, 2004, to December 10, 2015. It contains daily data on the closing price, the highest price, the lowest price, and the opening price using deep leaning techniques [21]. In 2016, Luckyson Khaidem used the Relative Strength Index (RSI), stochastic oscillator, and other technical indicators to train the model. As a learning model, an ensemble of multiple decision trees was employed [22]. In 2020, Xiongwen Pang et al. filter historical data from the Shanghai A-share market for a single stock, including the opening, highest, lowest, closing, and volume. To the initial five indicators, they added daily, weekly, and monthly amplitudes, as well as volume amplitudes, for a total of nine indicators that anticipate a single stock's price and trend. Ten years of data from January 1, 2006, to October 19, 2016, were chosen [23].

3.2 Review based on Fundamental Analysis (Qualitative data)

Along with quantitative data, social media (such as Twitter) sentiments analysis also plays a vital role in improving stock market prediction accuracy. Investor’s sentiment is a significant factor in the stock market. As a supplement to stock market data, user-generated textual material on the Internet is a valuable source for reflecting investor psychology and forecasting stock prices. The information comprises managerial quality, contentment with different stakeholders, ethics, and brand value, among other things, in order to assess the company's investment prospects. Below are the few related research works based on fundamental analysis.

Sidra Mehtab, in 2021 proposed a hybrid approach by collecting daily price movement of Nifty 50 index along with sentiment analysis module based on Twitter data to link public sentiment about company prices with market sentiment. This was accomplished by analyzing Twitter sentiment and the previous week's closing values in order to forecast stock price movement for the next week [24]. In 2019, Zhigang Jin applied deep learning (LSTM) considering investor’s emotional tendency and shown that investors' emotional tendencies may significantly enhance projected outcomes; the addition of EMD can significantly increase the prediction of inventory sequences [25]. In 2015, Thien Hai Nguyen instead of considering the overall mood and sentiments from social media, they considered sentiments from the specific topics of the company which are automatically extracted from the texts in a message board by using the proposed method Aspect-based sentiment [26]. In 2017, Shri Bharath proved that sentiments analysis of RSS news feed has improved the accuracy in predicting the stock price of the company ARBK from Amman Stock Exchange (ASE) [27]. In 2016, Venkata Sasank Pagolu et al. used the current work analyses public opinions in tweets using two distinct textual representations, Word2vec and Ngram by using sentiment analysis and supervised machine learning concepts to Twitter tweets and examine the association between a company's stock market performance and tweet sentiment [28]. In 2018, Rui Ren take into account the day-of-week impact and create more dependable and realistic sentiment indices employing SVM [29]. In 2018, Sahar Sohangir applied deep learning techniques (LSTM, CNN) to improve sentiment analysis performance of StockTwits so that an investor could get better, and accurate understanding of the sentiment related to that stock [30]. In 2020, Wasiat Khan [13] proposed a model by collecting two types of stock data: social media and financial news and applying deep learning and hybrid techniques on them to predict stock market accurately for ten successive days. To enhance the performance and quality of predictions, data sets are subjected to feature selection and spam tweet reduction.

3.3 Review based on Combined Analysis (Technical and Fundamental Analysis)

As seen in the above research work both quantitative and qualitative stock data have their own important roles to play in stock market prediction strategy because stock market is highly volatile, and each factor makes an impact on its movement. The models (either using machine learning or deep learning) performs better when the stocks and indices are analyzed fundamentally and technically and hence, profitability of the investors also increases proportionally. In the recent years numerous experiments have been performed based on combined analysis. Below are few related research work utilizing both quantitative data (historical stock data or financial ratios) as well as qualitative (social media sentiments, management quality) data.

In 2019, Saloni Mohan et al. did the experiment on two distinct datasets, from February 2013 to March 2017, the daily stock price dataset containing the closing stock prices of the Standard and Poor's 500 firms and news stories on S&P 500 firms from February 2013 to March 2017 from the websites of worldwide daily newspaper publishers. The techniques used were ARIMA, Facebook

Prophet and RNN-LSTM [31]. In November 2019, Andrea Picasso et al. proposed a model where inputs include technical analysis indicators and the sentiment of news stories (fundamental). The result is a powerful predictive model capable of forecasting the trajectory of a portfolio comprised of the NASDAQ100 index's twenty largest businesses [10]. In February 2018, [32] Qili WANG provided a groundbreaking paradigm for predicting financial market movements by integrating the wisdom of crowds with technical analysis and recommended an approach in which both social media sentiment and market technical indicators are taken into consideration in stock market forecasting. They create a data mining strategy called deep random subspace ensembles (DRSE), which integrates deep learning methods with ensemble learning algorithms for more successful mining of stock market fluctuations by 14.2 percent, based on the characteristics of the prediction assignment. In the year 2020, an experiment was completed on more than five years of Hong Kong Stock Exchange data, employing numerical price data through technical indicators and textual news items via sentiment vectors via sentiment analysis. Xiaodong Li stated that models that include both pricing and sentiments beat models that rely only on technical indicators or news moods alone [33]. In 2022 [34], the experiment using Genetic Algorithm on 6 years of data from January 2015 until 31st of December 2020 on 26 companies for technical (data from Yahoo finance) as well as sentimental (from articles, news, tweets) analysis separately was performed and Eva Christodoulaki concluded that sentiment analysis features SA statistically produced better results as compared to technical indicators TA sharpe ratio and risk.

4. Assessment analysis on various factors based on the study

This section is sub-divided into three parts according to the investment strategy applied i.e., either fundamental analysis or technical analysis or both combined. Each sub-section delivers the evaluation of various factors that have a significant impact on the stock market movement.

4.1 Evaluation based on technical analysis

This sub-section appraises the research work (some discussed in Section 3) on stock market prediction applying technical analysis (quantitative data). Several factors under technical analysis that influence the trends in stock market are data source used, technical indicators applied, time frame and period of collected data and performance metrics are discussed in **Table 1**.

Table 1: Evaluation table based on technical analysis (structured data)

Reference	Data source	Time Frame	Dataset Time Period	Technical Indicators used	Performance metrics
[15]	NASDAQ: AAPL from Google Finance	Daily closing price	Last 250 trading sessions	Date, open, high, low, close and volume	Correlation coefficient, MAE, RMSE, RAE, RRSE
[16]	Tehran stock exchange from tsetmc website	Not stated	November 2009 to November 2019 -10 years	SMA, WMA, MOM, STCK, STCD, RSI, SIG, LWR, ADO, CCI.	F1-Score, Accuracy and Receiver Operating Characteristics-Area Under the Curve (ROC-AUC) metrics.
[3]	Japanese Candlestick	1-6 days	48 datasets (each having different training period between 2000 to 2004)	Daily stock prices including Open, high, low, close prices	Hit rate
[17]	DAX (Germany), CAC (France) and FTSE (UK) Also, Apple, AT&T, and (GE)	Not stated	Daily stock data from 2 January 2007 to 26 March 2012 for IEM and from 1 August 2010 to 1 August 2017 for active firms	BB, RSI, PVI, NVI, accumulation/distribution line, highest high, lowest low, median price, PVT, typical price, volume rate of change, OBV, price rate of change, weighted close, and William's accumulation/distribution	MAE, RMSE, MARE, MSRE, RMSRE, MAPE, MSPE, RMSPE
[18]	Closing price of SPDR S&P 500 ETF (ticker symbol: SPY)	Daily return direction	2518 trading days between June 1, 2003, and May 31, 2013	EMA, T-bill rates, certificate of deposit rate	MSE, mean of daily return, std. of daily return, sharpe ratio

[19]	Five companies collected from Yahoo Finance	Next day closing price for 2 years from 2017-2019	04/05/2009 – 04/05/2019	Open, high, low and close price, adjacent close and volume used as: H-L, C-O, 7 DAYS MA, 14 DAYS MA, 21 DAYS MA, 7 DAYS STD DEV.	RMSE, MAPE, MBE
[20]	GSE, JSE, NYSE, BSE-SENSEX	Daily	January 2012 to December 2018	SMA, EMA, MACD, RSI, OBV.	Accuracy, Mean, STD, RMSE, MAE, R^2, precision, recall.
[21]	NASDAQ stock exchange	Next day closing price	August 19, 2004, to December 10, 2015	BB indication, closing price, highest price, lowest price, and opening price	PCD, SMAPE, MAPE, RMSE, HR, r_1, R^2, r_2
[22]	AAPL, GE dataset (NASDAQ) and Samsung Electronics Co. Ltd. (Korean Stock Exchange)	One month, 2 months and 3 months	Not stated	RSI, Stochastic Oscillator, Williams %R, MACD, PRC, OBV.	Accuracy, precision, recall, specificity
[23]	Shanghai A-shares composite index and Sinopec	Daily	January 1, 2006–October 19, 2016	daily amplitude, 5-day amplitude, 10-day amplitude, and amplitude of volume fluctuation based on open, close, highest and lowest price, and volume traded	Accuracy, MSE,
[35]	Tehran Securities Exchange Technology Management Co (TSETMC)	1, 2, 5, 10, 15, 20 and 30 days	November 2009 to November 2019	Momentum, Stochastic K percent, Stochastic D percent, RSI, A/D oscillator, CCI, Signal, Simple n-day moving average, weighted 14-day moving average	MAPE, MAE, RRMSE, MSE
[36]	NSE and NYSE from Yahoo finance	Daily	1 January 1996 to 30 June 2015 (dataset for NSE) 3rd January 2011 to 30th December 2016 (dataset for NYSE)	Day-wise closing price	MAPE
[37]	S&P 500 index, NASDAQ Composite, Dow Jones Industrial Average, NYSE Composite, and RUSSELL 2000	Daily closing price	Jan 2010 to Nov 2017	Open, close, high, low price, volume traded.	Macro-Averaged-F-Measure

[38]	S&P 500 from0 Yahoo finance	Daily closing price	01/01/2010 to 30/11/2017	Open, close, high, low price, volume	MAPE, RMSE, R^2
[39]	Bloomberg	8 January 2014	4 November 2008 to 7 January 2014	MA, RSI, Williams R, BB, ROC, Stochastic Oscillator, Ichimoku, LRS, Aroon Oscillator	Accuracy
[40]	GOOGL and NKE from Yahoo finance	Daily	8/19/2004 to 12/19/2019 for GOOGL and 1/4/2010 to 12/19/2019 for NKE.	Daily opening price	MSE
[41]	Dhaka Stock Exchange (DSE)	Not stated	January 2013 to April 2015	Opening and closing prices, highest and lowest prices, and total number of shares traded daily (no.)	RMSE, R^2,
[42]	Wind Information Inc. for CSI 300	Daily	January 1, 2016, to December 31, 2016	Opening price, Maximum price, Minimum price, Trading volume, Turnover, Bias, Bollinger bands, Directional movement index, Exponential moving averages, Stochastic index, Moving averages, MACD, Relative strength index	RMSRE, DPA
Total	**25 research articles**				

4.2 Evaluation based on fundamental analysis

This sub-section talks about the research work related to stock prediction by means of textual data i.e., qualitative dataset. There are number of variables that have an impact on stock market price movement are discussed below in **Table 2**.

Table 2: Evaluation table based on fundamental analysis (unstructured textual data)

Reference	Data source	Dataset Time period	Time frame	Number of entries	Performance metrics
[24]	Twitter on Nifty 50 (NSE)	three years (2015- 2017)	Closing price with window of one week for the period January 2,2018 till June 28,2019	Not stated	MAPE, correlation, GRANGER TEST P-VALUES AT DIIFERENT LAGS
[25]	Sentiment analysis on AAPL using stocktwits	1219 days	daily closing price	96,903	MAE, RMSE, MAPE, R^2, t, accuracy
[26]	Yahoo finance message board	July 23, 2012, to July 19, 2013	Daily from April 01, 2013, to July 19, 2013	171 transaction dates for 18 stocks (23,200 max. entries for 18 stocks per transaction date)	Accuracy
[27]	Investing.einnews.com Website	2005-2007	5-day, 10-day, 15-day of the	Not stated	Precision

			month April 2006.		
[28]	Twitter API	August 31st, 2015, to August 25th,2016	next day with best results of window size of 3 days	2,50,000	Accuracy, precision, recall, F-measure.
[29]	Web news and financial news	June 17th, 2014, and June 7th, 2016	Day-of-week	1,930,592	Accuracy
[30]	Stocktwits website	Messages posed First six months of 2015	Daily	Not stated	Correlation coefficient, window size, accuracy, precision, recall, F-measures.
[13]	Twitter and news	July 1, 2016, to June 30, 2018	Daily for 10 days	1,055,846 tweets, 9586 news counts	Accuracy, precision, recall, F-measure.
[39]	News and Twitter feed (KSE-100 index)	September 2015 to January 2016	Not stated	100	Accuracy, correlation
[43]	guba.eastmoney.com for CSI300 index	January 1, 2009, and October 31, 2014	Daily closing price	18 million	Precision, recall, F-measure, accuracy
Total	**12 research articles**				

4.3 Evaluation based on Combined (technical and fundamental) Analysis

This sub-section examines the research work based on combined analysis applying both technical indicators (quantitative data) and textual data (qualitative data) and discusses different factors predicting stock market movement shown in **Table 3**. It has been noticed through related research work that using combined analysis accuracy has been increased as compared to earlier single type of analysis evaluated because of several data sources employed.

Table 3: Evaluated table based on combined analysis

Reference	Data sources	Dataset time	Number of entries	Indicators	Performance metrics
[31]	S&P 500, news articles	February 2013 to March 2017	265,463 news articles	Daily Closing price	MAPE
[32]	Facebook, twitter, Sina Weibo and Chinese stock market records	2012-2015 (970 trading days)	84,034	MACD, KDJ, DMI, DMA, RSI, BB	F1, precision, recall, accuracy, AUC
[33]	stock data of Hong Kong Exchange and FINET news	January 2003 to March 2008	Not stated	MA, MACD, RSI, MFI	Accuracy, F1
[34]	Yahoo finance, python library Google search engine	January 2015 to 31st December 2020	Not stated	MA, the momentum, ROC, volatility	Sharpe ratio, risk
[10]	Google finance API (NASDAQ100 index), news from Intrinio API	03/07/2017 to 14/06/2018	Not stated	Open, close, mid, high, low price and volume	Accuracy, recall

[45]	Reuter's website and S&P 500 Yahoo finance	October 20, 2006, to November 21, 2013	106.494 news articles	Closing price of past day, week and month	Accuracy
[46]	Yahoo finance, news using crawler and Twitter	2003 to 2015	Not stated	open price, close price, low price, high price, adjusted close price and volume traded	Accuracy
[47]	Wind used for quantitative data as well as for news dataset for Shanghai Composite Index, Xueqiu used for social media feed	Jan. 1, 2015, to Dec. 31, 2016	38,727 and 39,465 news articles in 2015 and 2016 respectively and 6,163,056 postings from social media.	average prices, market index change and turnover rate of each trading day	F1 and Accuracy
[48]	Reddit World News Channel for news for DJIA and Yahoo finance for historical data and data from Guardian's restful news API	2000 to 2008 from Guardian's restful news and 2008 to 2016 from yahoo finance	Top 25 headlines from each day	Historical data of DJIA inex like close price, open price, high and low price of that day and the volume traded	Precision, recall and F1 score
Total	**13 research articles**				

The above is the evaluation of around 50 related research work from the year 2015 to 2021 demonstrating important feathers and some factors that have a significant impact on stock market movement are not mentioned in any previous survey papers studied. This survey paper identifies and fills in such gaps by taking into account key criteria such as the data sources used, the dataset time slot and frame, the indicators employed, and the performance measures used, among others and relished the opportunity to evaluate previous and present state-of-the-art stock market forecasting work.

5. Distribution of literature and Discussion

This section summarizes the distribution of literature (done in above section) based on various factors such as categorization of analysis (technical, fundamental, and combined analysis), categorization of different data sources applied, and different performance metrics employed, and analysis on indicators used in technical analysis.

Table 4 below depicts the categorization of the above reviewed research papers based on technical analysis (quantitative data), fundamental analysis (qualitative data) and combined analysis.

Table 4: Analysis Categorization

Category	Number of papers	Percentage
Technical analysis	25	50%
Fundamental analysis	12	24%
Combined analysis	13	26%
Total	50	100%

Technical Analysis

From the above **Table 4**, it is deduced that 50% of the related research work has been done through technical analysis i.e., on quantitative data extracting historical data of the company's stocks. Majorly used data source is Yahoo finance and Google finance and the commonly used technical indicators are SMA, RSI, BB, EMA, MACD and Williams R, calculated with the help of historical open, close, high and low price. It is also seen that the time frame is generally intraday, one day, one week or one month as shown in **Table 1**. The performance metrics employed are MAE, RMSE, MSE, accuracy, precision, recall, F1-measures etc.

Fundamental Analysis

In recent years, more studies are coming up with qualitative data (unstructured data) i.e., using social media feed, news feed, twits, blogs to do sentiment analysis of social network sites (SSNs). This have helped in improving the prediction of the movement in stocks. As seen in **Table 4**, 24% of the research work is accomplished using fundamental analysis. Multiple sources like Twitter, Yahoo finance message board, investment.einnews website, etc. are used to collect unstructured data (relevant messages and news).

Combined Analysis

Some researchers tried to utilize the power of both to attain the best results using both the data- historical (structured) and unstructured (financial news, twits) data. According to the reviewed literature 26% of the research work employ combined analysis using both technical indicators and number of collected relevant blogs/news and macro-economic variables.

6. Conclusion and future scope

In this paper, almost fifty research work is systematically reviewed under the analysis and dataset categorization that is technical, fundamental, and combined and various characteristics attached with them. It is observed that maximum research articles published used technical analysis for the stock market prediction. It has also been noted that in most of the reviewed work, 60% data is considered for training, 20% for validation and 20% for testing. Yahoo finance and Google finance are most employed data source for technical analysis and Twitter, news articles websites for fundamental analysis. The performance metrics like MAE, RMSE, MSE, accuracy, precision and F-measures are popularly used to measure the performance of the prediction accomplished. Also got knowledge of data collection time period and frame required to result in better performance.

Through this detailed review of literature, we encounter some gaps that could be filled in near future by using ensemble techniques on combining diverse stock data-sources together.

References

1. A. Ghaznavi, M. Aliyari and M. R. Mohammadi, "Predicting stock price changes of tehran artmis company using radial basis function neural networks," *Int. Res. J. App. Basic Sci,* vol. 10, p. 972, 2016.
2. M. S. Checkley, D. A. Higón and H. Alles, "The hasty wisdom of the mob: How market sentiment predicts stock market behavior," *Expert Systems with applications,* vol. 77, p. 256–263, 2017.
3. E. Ahmadi, M. Jasemi, L. Monplaisir, M. A. Nabavi, A. Mahmoodi and P. A. Jam, "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 Systems with Applications,* vol. 94, p. 21–31, 2018.
4. N. Maknickiene, I. Lapinskaite and A. Maknickas, "Application of ensemble of recurrent neural networks for forecasting of stock market sentiments," *Equilibrium. Quarterly Journal of Economics and Economic Policy,* vol. 13, p. 7–27, 2018.
5. A. Gorgulho, R. Neves and N. Horta, "Applying a GA kernel on optimizing technical analysis rules for stock picking and portfolio composition," *Expert systems with Applications,* vol. 38, p. 14072–14085, 2011.
6. N. Petrusheva and I. Jordanoski, "Comparative analysis between the fundamental and technical analysis of stocks," *Journal of Process Management. New Technologies,* vol. 4, p. 26–31, 2016.
7. P. M. Dechow, A. P. Hutton, L. Meulbroek and R. G. Sloan, "Short-sellers, fundamental analysis, and stock returns," *Journal of financial Economics,* vol. 61, p. 77–106, 2001.
8. D. Tarmidi, R. Pramukty and T. Akbar, "Fundamental Analysis of Financial Ratios on Stock Prices," *Saudi Journal of Economics and Finance,* vol. 4, p. 176–180, 2020.
9. R. Talib, M. K. Hanif, S. Ayesha and F. Fatima, "Text mining: techniques, applications and issues," *International Journal of Advanced Computer Science and Applications,* vol. 7, p. 414–418, 2016.
10. A. Picasso, S. Merello, Y. Ma, L. Oneto and E. Cambria, "Technical analysis and sentiment embeddings for market trend prediction," *Expert Systems with Applications,* vol. 135, p. 60–70, 2019.
11. M. R. Vargas, C. E. M. Dos Anjos, G. L. G. Bichara and A. G. Evsukoff, "Deep leaming for stock market prediction using technical indicators and financial news articles," in *2018 international joint conference on neural networks (IJCNN)*, 2018.
12. M. Agrawal, A. U. Khan and P. K. Shukla, "Stock price prediction using technical indicators: A predictive model using optimal deep learning," *Learning,* vol. 6, p. 7, 2019.
13. W. Khan, M. A. Ghazanfar, M. A. Azam, A. Karami, K. H. Alyoubi and A. S. Alfakeeh, "Stock market prediction using machine learning classifiers and social media, news," *Journal of Ambient Intelligence and Humanized Computing,* p. 1–24, 2020.
14. P. Jiao, A. Veiga and A. Walther, "Social media, news media and the stock market," *Journal of Economic Behavior & Organization,* vol. 176, p. 63–90, 2020.
15. E. P. Torres P, M. Hernández-Álvarez, E. A. Torres Hernández and S. G. Yoo, "Stock Market Data Prediction Using Machine Learning Techniques," in *International Conference on Information Technology & Systems*, 2019.
16. M. Nabipour, P. Nayyeri, H. Jabani, S. Shahab and A. Mosavi, "Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; a comparative analysis," *IEEE Access,* vol. p. 150199–150212, 2020.
17. S. Lahmiri, "A technical analysis information fusion approach for stock price analysis and modeling," *Fluctuation and Noise Letters,* vol. 17, p. 1850007, 2018.
18. X. Zhong and D. Enke, "Predicting the daily return direction of the stock market using hybrid machine learning algorithms," *Financial Innovation,* vol. 5, p. 1–20, 2019.
19. M. Vijh, D. Chandola, V. A. Tikkiwal and A. Kumar, "Stock closing price prediction using machine learning techniques," *Procedia computer science,* vol. 167, p. 599–606, 2020.

20. I. K. Nti, A. F. Adekoya and B. A. Weyori, "A comprehensive evaluation of ensemble learning for stock-market prediction," *Journal of Big Data,* vol. 7, p. 1–40, 2020.
21. R. Singh and S. Srivastava, "Stock prediction using deep learning," *Multimedia Tools and Applications,* vol. 76, p. 18569–18584, 2017.
22. L. Khaidem, S. Saha and S. R. Dey, "Predicting the direction of stock market prices using random forest," *arXiv preprint arXiv:1605.00003,* 2016.
23. X. Pang, Y. Zhou, P. Wang, W. Lin and V. Chang, "An innovative neural network approach for stock market prediction," *The Journal of Supercomputing,* vol. 76, p. 2098–2118, 2020.
24. J. Sen and S. Mehtab, "A Robust Predictive Model for Stock Price Prediction Using Deep Learning and Natural Language Processing," 2021.
25. Z. Jin, Y. Yang and Y. Liu, "Stock closing price prediction based on sentiment analysis and LSTM," *Neural Computing and Applications,* vol. 32, p. 9713–9729, 2020.
26. T. H. Nguyen, K. Shirai and J. Velcin, "Sentiment analysis on social media for stock movement prediction," *Expert Systems with Applications,* vol. 42, p. 9603–9611, 2015.
27. S. Bharathi and A. Geetha, "Sentiment analysis for effective stock market prediction," *International Journal of Intelligent Engineering and Systems,* vol. 10, p. 146–154, 2017.
28. V. S. Pagolu, K. N. Reddy, G. Panda and B. Majhi, "Sentiment analysis of Twitter data for predicting stock market movements," in *2016 international conference on signal processing, communication, power and embedded system (SCOPES)*, 2016.
29. R. Ren, D. D. Wu and T. Liu, "Forecasting stock market movement direction using sentiment analysis and support vector machine," *IEEE Systems Journal,* vol. 13, p. 760–770, 2018.
30. S. Sohangir, D. Wang, A. Pomeranets and T. M. Khoshgoftaar, "Big Data: Deep Learning for financial sentiment analysis," *Journal of Big Data,* vol. 5, p. 1–25, 2018.
31. S. Mohan, S. Mullapudi, S. Sammeta, P. Vijayvergia and D. C. Anastasiu, "Stock price prediction using news sentiment analysis," in *2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)*, 2019.
32. Q. Wang, W. Xu and H. Zheng, "Combining the wisdom of crowds and technical analysis for financial market prediction using deep random subspace ensembles," *Neurocomputing,* vol. 299, p. 51–61, 2018.
33. X. Li, P. Wu and W. Wang, "Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong," *Information Processing & Management,* vol. 57, p. 102212, 2020.
34. E. Christodoulaki, M. Kampouridis and P. Kanellopoulos, "Technical and Sentiment Analysis in Financial Forecasting with Genetic Programming," 2022.
35. M. Nabipour, P. Nayyeri, H. Jabani, A. Mosavi, E. Salwana and others, "Deep learning for stock market prediction," *Entropy,* vol. 22, p. 840, 2020.
36. M. Hiransha, E. A. Gopalakrishnan, V. K. Menon and K. P. Soman, "NSE stock market prediction using deep-learning models," *Procedia computer science,* vol. 132, p. 1351–1362, 2018.
37. E. Hoseinzade and S. Haratizadeh, "CNNpred: CNN-based stock market prediction using a diverse set of variables," *Expert Systems with Applications,* vol. 129, p. 273–285, 2019.
38. K. A. Althelaya, E.-S. M. El-Alfy and S. Mohammed, "Evaluation of bidirectional LSTM for short-and long-term stock market prediction," in *2018 9th international conference on information and communication systems (ICICS)*, 2018.
39. R. Iacomin, "Stock market prediction," in *2015 19th International Conference on System Theory, Control and Computing (ICSTCC)*, 2015.
40. A. Moghar and M. Hamiche, "Stock market prediction using LSTM recurrent neural network," *Procedia Computer Science,* vol. 170, p. 1168–1173, 2020.
41. M. Billah, S. Waheed and A. Hanifa, "Stock market prediction using an improved training algorithm of neural network," in *2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE),* 2016.
42. X. Zhou, Z. Pan, G. Hu, S. Tang and C. Zhao, "Stock market prediction on high-frequency data using generative adversarial nets.," *Mathematical Problems in Engineering,* 2018.
43. M. Usmani, S. H. Adil, K. Raza and S. S. A. Ali, "Stock market prediction using machine learning techniques," in *2016 3rd international conference on computer and information sciences (ICCOINS)*, 2016.
44. J. Li, H. Bu and J. Wu, "Sentiment-aware stock market prediction: A deep learning method," in *2017 international conference on service systems and service management*, 2017.
45. M. R. Vargas, B. S. L. P. De Lima and A. G. Evsukoff, "Deep learning for stock market prediction from financial news articles," in *2017 IEEE international conference on computational intelligence and virtual environments for measurement systems and applications (CIVEMSA)*, 2017.

46. A. Nayak, M. M. Pai and R. M. Pai, "Prediction models for Indian stock market," *Procedia Computer Science*, vol. 89, p. 441–449, 2016.
47. X. Zhang, S. Qu, J. Huang, B. Fang and P. Yu, "Stock market prediction via multi-source multiple instance learning," *IEEE Access*, vol. 6, p. 50720–50728, 2018.
48. S. S. Maini and K. Govinda, "Stock market prediction using data mining techniques," in *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, 2017.