Int J Syst Assur Eng Manag (June 2024) 15(6):2011–2018
https://doi.org/10.1007/s13198-023-02214-6

REVIEW PAPERS

A comprehensive review on sentiment analysis of social/web media big data for stock market prediction

Pratham Shah[1] · Kush Desai[2] · Mrudani Hada[3] · Parth Parikh[3] · Malav Champaneria[4] · Dhyani Panchal[4] · Mansi Tanna[6] · Manan Shah[5]

Received: 22 July 2022 / Revised: 29 August 2023 / Accepted: 17 November 2023 / Published online: 26 January 2024

Abstract It is generally known that public opinion and stock market dynamics are inextricably linked. With the growth of social and web-based media, online platforms have emerged as a key gauge of public mood. This digital environment produces a lot of data quickly. This extensive dataset's analysis offers priceless insights into the general public's perception, which in turn might influence market performance. The vast array of approaches for efficiently processing the sizable amount of data originating from social and web-based media are reviewed in detail in this study. Additionally, it looks at studies exploring the integration of big data analytics and sentiment insights for more accurate market predictions, as well as studies studying the prediction of stock market trends using sentiment analysis.

Keywords Sentiment · Stock market · Big data

✉ Manan Shah
manan.shah@spt.pdpu.ac.in

[1] Department of Computer Science and Engineering, Indus University, Ahmedabad, Gujarat 382115, India

[2] Department of Computer Science and Engineering, The LNM Institute of Information and Technology, Jaipur, Rajasthan, India

[3] Department of Electronics and Communication Engineering, Nirma University, Ahmedabad, Gujarat, India

[4] Department of Information and Technology, Chandubhai S Patel Institute of Technology, CHARUSAT, Anand, Gujarat, India

[5] Department of Chemical Engineering, School of Energy Technology, Pandit Deendayal Energy University, Gandhinagar, Gujarat, India

[6] Department of Computer Science Engineering, Devangpatel Institute of Technology Research (DEPSTAR) Chandubhai S Patel Institute of Technology, CHARUSAT, Anand, Gujarat, India

1 Introduction

The stock market is a crucial component of a free-market economy because it gives businesses a wonderful way to expand while attracting the public as investors. When we consider things from the viewpoint of an investor, there are numerous subtleties that are at play. In essence, the role of an investor, a stock market participant who actively participates, is to attempt to forecast the market's future direction in order to profit from it. When it comes to analyzing the market, there are many different methods. According to (Otoo 1999), the stock market and consumer confidence are closely related. A rise in equity value raises public attitude, and individuals view changes in share values as a good indication of economic activity and prospective increases in labor income. One of the most important stock market prediction strategies is predicting stock prices by gauging public opinion. The relationship between these two variables may be described by behavioral economics, which integrates the influence of market and economic psychology (Mullainathan and Thaler 2000). According to studies (Marg, 1995; Nofsinger 2005; Smailovi et al. 2013), emotions affect social behavior and logical thought, and the stock market itself may be viewed as a gauge of the general mood. Today's users spend a large portion of their time on social media; as a result, a study of these social media data may provide a more thorough knowledge of user behavior that can be used to a variety of fields, including finance (Oliveira et al. 2017).

Big data is the analysis of big, complicated data collections to extract useful information from them. These vast, complex data sets can be structured or unstructured. Big data is also defined as a sizable dynamic data set that is produced by or generated from human activities, communications, movements, and behaviors (Tsou, 2015); in this context, we focus on human behavior as it is expressed online.

(Dobre and Xhafa, 2014) claimed that 2.5 quintillion bytes of data were generated per day from sources such as posts to social media sites, digital photos and videos, buy transaction records, cell phone signals, and sensors used to gather climate data. Big data analytics is the processing of structured and unstructured data from external sources as opposed to conventional models, which primarily use internal data sources, such as an index or a database. It also encompasses more information as it also covers social network data and web media data, which can be used for tactical benefit or achieving the 'alpha' in the stock market (Kwon et al. 2014).

The internet is an excellent site for the assessment of public mood since it is populated with opinions and the number of individuals with those opinions is constantly growing online (Shayaa et al. 2018). This assessment cannot be done through a personal network. Web and social media are channels that constantly generate enormous volumes of data, most of which reflects the opinions of the public. The emergence of big data from social media and web media has given the field of big data analytics a fresh perspective. Big data from social media can be analyzed using both more modern machine learning techniques and different older data mining methods (Ghani et al. 2019). Data mining techniques can be used to perform this sentiment analysis on social media and online media. The researchers have a significant problem when trying to derive relevant insights from the social media data due to its size, noise, and dynamic nature. Using data mining approaches, researchers are able to solve these issues and gain valuable knowledge that would not have been possible otherwise (Barber et al. 2012).

By comparing the content of social media posts with a dictionary of mood-related words, sentiment can be evaluated. This sentiment metric has the tendency to provide a general sense of the market's mood and is used to forecast not only the market's direction but also its volatility and trading volume (Checkley et al. 2017). This is reinforced by behavioral finance, which demonstrates that mood and emotion play a significant part in financial decision making (Bollen et al. 2011) (Nofsinger 2005). While news affects the stock market, people's emotions may also play a significant impact in the movement of the stock market.

2 Various stock market prediction techniques and the role of big data

Big Data was defined more broadly by (Demchenko et al. 2013) using the 5Vs: volume, velocity, variety, value, and veracity. The volume keeps rising, especially with the rise of social media. Financial markets can provide alpha if velocity, value, and veracity are combined properly because the markets are frequently about obtaining reliable information faster than the competition.

Since both economic and non-economic factors have an impact on the stock market, accurately predicting it is a difficult undertaking (Gandhmal and Kumar 2019). Big data facilitates the use of social media and numerical data to make precise forecasts in real time (Attigeri et al. 2015). Investors, consumers, and others are influenced by their social mood, which makes social media a useful tool for measuring social mood (Nofsinger 2005). Big data analytics employ the data generated from these media as an input. The introduction of machine learning-based analytics is the next significant development in big data analytics. These analytics use large amounts of data as input and, using that input, build a model that acts as the machine's brain to perform automated or semi-automated analysis that can result in real-time analytical decisions (Ramesh et al. 2019). Big data techniques are utilized to construct prediction models since the structured and unstructured data that is continually being generated on the internet is important and requires tremendous technical help to parse (Bollen et al. 2011).

3 Advanced techniques for stock market prediction

Traditional stock market forecasting methods like fundamental analysis and technical analysis are unable to accommodate the huge amounts of data being produced every second. Big data can be handled and used using a variety of contemporary ways. Artificial Neural Network is a bio-inspired algorithm that has achieved success in the realm of artificial intelligence by effectively utilizing the relationship between stock performance and its determining elements (Vui et al. 2013; Gandhmal and Kumar 2019). To forecast the movement of stock values, an ANN model that combines Bayesian regularization and the Levenberg Marquardt algorithm was put forth (Ticknor 2013). Kim and Lee (2004) created a hybrid model integrating ANN and GA (Genetic algorithms) that was used to forecast the future course of the KOSPI (Korea Composite Stock Price Index). According to Zhou et al. (2018), CNN (Convoluted Neural Network) is a sort of feed-forward ANN that comprises of input, output, and several hidden layers. One of the most well-known statistical stock market forecasting methods is the Hidden Markov Model. The experimental results of the HMM with a graph-based large data optimization strategy proposed by (Sassi et al. 2021) are notable. The model in (Bollen et al. 2011) that used 20,000 bits of clickstream data gathered as an entire sequence of user data revealed that HMM may even be used for online data analytics with the aid of Clustering. (Verma et al. 2019) revealed a pattern identification method based on data mining that used historical data to forecast future patterns. This method proved to be more accurate than conventional ANN and SVM (Support Vector Machine) methods. SVM is yet another well-liked method for big data analysis. (Wang et al. 2013) presented a hybrid DT (Decision Tree)-SVM strategy that outperforms conventional DT and SVM methods. We can get insightful knowledge regarding huge data that can be used to get a feel of the market's direction.

4 Methodology

The following is the flow of methodology for sentiment analysis and using it for stock market prediction (Fig. 1).

5 Sentiment analysis from big data and stock market prediction

Awan et al. (2021) used big data gathered from news and social media to forecast stock movement. Utilising sentiment analysis, information gathered from message boards and social media was examined. They used PySpark to create several machine learning models with Spark MLlib, which is more effective and scalable than standard models and so produces better results. Data from Twitter, blogs, and message boards were combined with historical datasets to create a third type of dataset. On news and messages, a logistic regression classifier and naive Bayes classifier were used to forecast market values. The logistics regression model was shown to be superior with 60–80% accuracy, while the naive bayes model only reached 60–70% accuracy.

As opposed to conventional frameworks, which can't process large data in real time, van Dieijen et al. (2020) developed new models to stream and analyze huge data streams. Here, a real-time processing system was developed to examine tweets and determine whether there was any connection between them and the movement of the stock price. The data collection comprised tweets and news that were collected via the Twitter API, which produced 30 tweets per second and processed streams in 0.4 s. High accuracy in tweet classification was achieved by applying machine learning techniques. The Naive Bayes method, which has been shown to perform better for brief tweets, was employed in place of the Random Forest algorithm. On a scale of 0 to 1, Twitter data classification accuracy was 0.77; the strongest correlation, at 0.5, was found between entertainment news and stock prices. When it comes to making predictions about the stock market, volatility is one of the most crucial elements. Big data analysis was used to investigate the relationship between user-generated volatility and stock market volatility by (van Dieijen et al. 2020). The data was generated by users and was gathered from blog postings, Google ticker searches, and Twitter tweets, which were divided into positive and negative tweets. Multivariate GARCH BEKK (Baba, Engle, Kraft and Kroner) model along with VAR (Vector Autoregressive) was utilized to analyze the relationship between user-generated content and the rate of stock returns, and the Graer Test was used to establish its causality. Future stock returns were most significantly impacted by Google ticker searches, and as these searches increased, so did stock returns.

Even web media data, in addition to social media data, can be used to predict stock market patterns. One such method was used by (Moat et al. 2014) in their attempt to forecast future stock market patterns by taking the number of Wikipedia page views into consideration. Before stock market drops, (Li et al. 2018) discovered a statistically significant correlation between the page views of financial companies or themes.

One of the most widely used methods for Big Data analytics is machine learning. To measure the sentiment for stock market prediction, (Qasem et al. 2015) employed machine learning algorithms over the large data generated on twitter. They compiled a dataset of 42,000 tweets about the financial

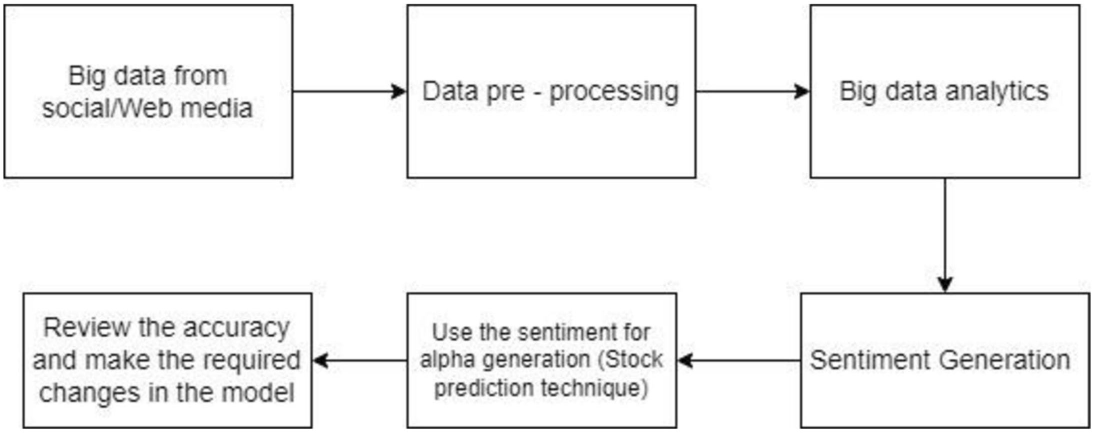
Big data from social/Web media
Data pre - processing
Big data analytics
Sentiment Generation
Use the sentiment for alpha generation (Stock prediction technique)
Review the accuracy and make the required changes in the model

Fig. 1 Methodology for market prediction using big data

sector and divided them into three categories: good, negative, and neutral. Both Bigram TF and Unigram TF-IDF were used to classify the tweets, with Unigram TF-IDF performing better overall with a 58% accuracy rate compared to Bigram TF. For the training of the test dataset, they used multi-class logistic regression and neural network classification models. When it came to predicting instances of the negative class, neural networks surpassed logistic regression (83%). The neutral class has the lowest accuracy since the training did a poor job of identifying it. Overall accuracy was from 50 to 60% for logistic regression and neural networks, respectively. Because clustering approaches don't require training the dataset, the authors advised employing them for improved accuracy.

(Kanavos et al. 2020) illustrated a two-step process to predict the movement of the starting price of a certain company's shares. The first component of this approach involved gathering data from Twitter using the Apache Flume streaming engine. The system then performed a sentiment analysis procedure to predict the movement of stock prices in relation to the sentiment of the tweets. The second part of the process involved data analysis, which included storing the data in Cassandra, a NoSQL database where the data goes through various models of pre-processing to get an accurate estimate of the sentiment. The proposed methodology's weakness, according to the authors, is that the existing tool cannot reliably identify non-literal expressions like sarcasm.

(Jiang et al. 2018) suggested an approach to determine the sentiment of news-based events utilizing massive amounts of social media data. Using words and the rules that go along with them, a word emotion association network (WEAN) was constructed for this technique. After some adjustments, their experiment, which comprised 48,396 microblogs on the Malaysia Airlines MH370, produced an average accuracy of 78%.

Financial rumors are essential to be aware of when forecasting market movements, since they have a substantial impact on price changes. With the development of social media, rumors have become much more common. In order to automatically detect financial rumors in the massive amounts of data generated by social and web media, researchers uses NLP to classify textual context that is it genuine or just a false rumor, and related to this (Majumdar and Bose 2018) proposed a cutting-edge methodology that blends Natural Language Processing with Predictive Analytics.

The sentiment data obtained from news items might be used to forecast the stock market. When putting this idea into practice, (Shah et al. 2018) began by using the Beautiful Soup library to fetch new articles. In an effort to predict pharmaceutical stock prices, they focused on publications with the terms "US" and "USFDA" in the headlines, as well as those with the words "Q1," "Q2," and so on. The sentences were converted into numerical vectors using the "patterns" package, and bigrams and trigrams were eventually generated from them. The accuracy of their sentiment analysis algorithm's predictions for the direction of daily stock market movement for a subset of pharmaceutical stocks traded on the Indian stock market was 70.59%, and it generated three buy, sell, and hold signals.

A logit model and fuzzy-set qualitative comparative analysis (fsQCA) were used by (Pieiro-Chousa et al. 2017) to show the effect of social media activity on the stock market after the Chicago Board Options Exchange Market Volatility Index (VIX). In the QCA version called fsQCA, a direct calibration procedure is applied. It makes a distinction between technical investors and non-technical investors by constructing relationships in terms that are both necessary and sufficient to produce a result.

Data from the microblogging service StockTwits.Com were gathered using the website's API. The study came to the conclusion that the effect of social media on the stock market depends on the type of investor. Rejoicing and mood are two essential components in reducing amplified threat for investors lacking technical understanding. However, when focusing on technical buyers, experience and holding term become significant factors. For the goal of textual content mining, (Sun et al. 2016) gathered a dataset of 45 million messages from StockTwits over the course of four years. They started by compiling a lexicon by researching the most popular terms for each year and combining them with the tickers of 420 stocks. They increased prediction accuracy by diagnosing the most important parameters when using a Sparse Matrix Factorization (SMF) version. For daily predictions, the SMF version had an accuracy rate of 51.37%. Another interesting conclusion showed that better prediction frequency no longer significantly increases prediction accuracy. The authors proposed that the content-sharing feature of the website, which frequently eclipses original content, may be responsible for this phenomenon.

Instead of forecasting positive or negative movement, (Ge et al. 2020) article investigates the connection between social media and stock market crashes. This is accomplished by their proposed cognition-based ECM model (Emotion-Cognition-Market), which was used for sentiment analysis of the emotion, a logistic regression was used to establish a relationship between ESM (Emotion in social media) and market cognition, and a Hidden Markov Model (HMM) was used to mine market cognition. The dataset included more than 280,000 Weibo posts for 34 stocks that are SSE 50-listed.

Using Twitter data to forecast stock market movements based on sentiment conveyed in tweets was explored by (Bollen et al. 2011) in his paper. They gathered massive amounts of tweets about the financial markets and used natural language processing (NLP) methods to analyze sentiment. To explore the connection between sentiment and stock market movements, they used correlation analysis. Building predictive models that would foresee stock market moves based on sentiment trends included feature engineering. indicators like accuracy, precision, recall, as well as financial-specific indicators like return on investment (ROI), were used to evaluate the models. Advanced NLP approaches, multimodal analysis, and domain-specific sentiment analysis are recent advancements in the sentiment analysis of Twitter moods on the stock market. (Table 1).

6 Challenges

Sarcasm management presents a significant challenge for sentiment analysis. Sarcasm includes saying the opposite of what is intended, which makes it difficult to maintain forecast accuracy. Despite this effort, researchers have fully confirmed their capacity to recognize sarcasm in text. For instance, (Bouazizi and Ohtsuki 2015); used 3 models—SVM, Naive Bayes, and Max Entropy to combine 4 units of capabilities—Sentiment Related, Punctuation Related, Syntactic Related, and Pattern Related. Their findings showed that when the sarcasm identification system was improved, version accuracy significantly improved. Another issue results from the use of figurative language, or language whose significance goes beyond the immediate context, rather than literal language.

Managing texts from many languages can be difficult. One text can contain multiple languages on its own. Sometimes the machine translation data can obscure a text's meaning. In the studies conducted by (Vilares et al. 2017), it was discovered that it was simple to teach a model a second language without suffering much from performance loss. However, neither the monolingual nor the multilingual

Table 1 Stock market prediction using sentiment analysis

Method	Data used	Models applied over	Result	Reference
LDA and JST based model	Yahoo Finance message board	American Stocks	Over 60% average accuracy	(Nguyen et al. 2015)
Novel autoregressive model	Twitter	S&P 500	Beats S&P 500 by 20% over 4 months	(Makrehchi et al. 2013)
"LDA-POS" method (latent Dirichlet allocation)-(Part of speech)	Yahoo Finance message board	Iranian market	Around 56% average accuracy	(Derakhshan & Beigy 2019)
SVM algorithm	Twitter	S&P 500	Average 64% accuracy	(Porshnev et al. 2013)
Naive classifier and OLS regression	Twitter	NASDAQ 100	Over 80% directional accuracy	(Rao & Srivastava 2012)
Naive Bayes and Logistic regression classifier	Social media, blogs and Message boards	American Stocks	Between 60 and 80% accuracy	(Javed Awan et al. 2021)
Random forest and Naive Bayes	Twitter	NASDAQ	Highest 50% accuracy	(Lee & Paik 2017)
Logistic regression and LibSVM	Twitter	Microsoft stock	Around 70% accuracy	(Pagolu et al. 2016)
Logistic regression and TF-IDF	Twitter	American stocks	Accuracy between 58.3% to 65.6%	(Gupta & Chen 2020)
Part-of-speech tagger, dictionary based approach and sentence sentiment score (SSS) algorithm	RSS news feed	Amman Stock Exchange	Accuracy of 78.5%	(Bharathi & Geetha 2017)
BERT (bidirectional encoder representations from transformers)	Economic news from CNBC, Forbes, New York Times, Washington Post, etc	Dow Jones Industrial Index	Hit rate of 69% in predicting the movement direction	(Sousa et al. 2019)
Sentiment analysis with SVM	Sina stock forum and East money stock forum	SSE 50 Index	Highest accuracy of 89.93%	(Ren et al. 2019)
Naive Bayes and KNN	News from Reuters, Nasdaq.com, Wall Street Journal, etc	Yahoo Inc, Microsoft Corporation and Facebook Inc	Highest accuracy of 89.90%	(Khedr et al. 2017)
GARCH SVM model	Sina Finance	50 Military sector stocks	Accuracy of 73.8%(value stocks) and 60.9%(growth stocks)	(Wu et al. 2014)

methodologies used to predict code switching texts are ideal. For code switching text, the best accuracy was 59.3% for English–Spanish text.

Another problem is when bots inflate social media posts. When this happens, the algorithm would only be evaluating the fake sentiment produced by the social media bots rather than the sentiment of actual people. To effectively forecast the sentiment, it becomes essential to remove information produced by bots. Natural Language Processing was used in an implementation by (Heidari and Jones 2020), and they were able to discriminate between a genuine user and a bot with 94% accuracy. The findings demonstrate that, during the in-crash and post-crash periods, respectively, an increase in ESM increases the chance of cognition of a crash state by 9.99 and 17.41%. The emotion of "fear" was the greatest risk factor, increasing cognition during the accident and decreasing it in the aftermath.

Other than these, the main difficulty encountered while working with huge data is volume and velocity. Since all social media channels are used in real time and have immediate affects when the markets are open, velocity is more important than ever. Volume would also lengthen computation times, making it challenging to execute solutions involving shorter time frames. Another challenge with the increase in volume and velocity is the creation of new language elements that need to be updated in the model, such as acronyms, emoticons, idioms, and terminologies (Sharef et al. 2016).

7 Future scope

According to the report, numerous research has had great success forecasting changes in the inventory market. However, as social media usage increases, a wide range of structures that serve various demographic organizations are introduced. As a result, there is a need for a comprehensive model that combines opinions from various social media platforms and financial blogs. This would allow the model to represent a wider range of demographics and produce forecasts that would be more comprehensive. The capacity of this model to accurately handle the subtleties of sarcasm and metaphorical discourse is essential to improving its accuracy.

In the near future, sentiment analysis's trend suggests a change towards a greater focus on video content. Structures that primarily use visual media, such as Instagram's "Reels," TikTok, and Facebook's "Shorts," are examples of emerging developments. In order to improve forecast accuracy, future models should be modified to extract relevant textual records from graphical codecs. In this backdrop, the field of sentiment analysis for stock market forecasting using big data from social and internet media is primed for growth and innovation. The changing environment necessitates not only the improvement of current trends but also the development of fresh approaches to harness the power of many media sources and accommodate the evolving trends in data consumption.

8 Conclusion

In conclusion, the growth of social and web-based media has had a significant impact on the interaction between public opinion and stock market dynamics. With their ability to generate vast amounts of data in real time, these online structures have become priceless indicators of popular emotion. Utilizing this data through big data analytics and sentiment analysis holds the possibility of providing priceless insights into public opinion, consequently affecting market performance. Researchers have been able to identify the emotions portrayed in textual facts by combining machine learning techniques with Natural Language Processing (NLP), advanced models like Artificial Neural Networks (ANN), and Hidden Markov Models (HMM).

Innovative methods to improve sentiment analysis accuracy have been developed in response to the difficulties in sarcasm detection and dealing with multilingual information. Looking ahead, the growth of sentiment analysis is poised to change media consumption trends, notably the growing prominence of video content across many systems. The integration of sentiment research approaches with machine learning and artificial intelligence methodology may improve the precision and timeliness of stock market forecasts, opening up new opportunities for financial forecasting as these techniques continue to develop.

Acknowledgements The authors are grateful to Indus University, The LNM Institute of Information Technology, Nirma University, CHARUSAT University and the Department of Chemical Engineering, School of Energy Technology, Pandit Deendayal Energy University for the permission to publish this research.

Author contributions All the authors make a substantial contribution to this manuscript. PS, KD, MH, PP, MC, DP, MT, and MS participated in drafting the manuscript. PS, MC, DP and MS wrote the main manuscript. All the authors discussed the results and implications on the manuscript at all stages.

Funding Not Applicable.

Availability of data and material All relevant data and material are presented in the main paper.

Declarations

Conflict of interest The authors declare that they have no competing interests.

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

References

Attigeri GV, Manohara Pai MM, Pai RM, Nayak A (2015) Stock market prediction: a big data approach. In: IEEE region 10 conference on TENCON, pp 1–5

Awan MJ, Rahim MSM, Nobanee H, Munawar A, Yasin A, Zain AM (2021) Social media and stock market prediction: a big data approach. Comput Mater Contin 67(2):2569–2583

Barber BM, Lee YT, Liu YJ, Odean T (2012) The cross-section of speculator skill: evidence from day trading (December 31, 2012). Available at SSRN: https://ssrn.com/abstract=529063 or https://doi.org/10.2139/ssrn.529063

Bharathi S, Geetha A (2017) Sentiment analysis for effective stock market prediction. Int J Intell Eng Syst 10(3):146–154. https://doi.org/10.22266/ijies2017.0630.16

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

Bouazizi M, Ohtsuki T (2015) Opinion mining in twitter how to make use of sarcasm to enhance sentiment analysis. In: Proceedings of the 2015 IEEE/ACM international conference on advances in social networks analysis and mining 2015, pp 1594–1597. https://doi.org/10.1145/2808797.2809350

Checkley MS, Higon 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

Demchenko Y, Grosso P, De Laat C, Membrey P (2013) Addressing big data issues in scientific data infrastructure. In: IEEE international conference on collaboration technologies and systems (CTS), pp 48–55

Derakhshan A, Beigy H (2019) Sentiment analysis on stock social media for stock price movement prediction. Eng Appl Artif Intell 85:569–578. https://doi.org/10.1016/j.engappai.2019.07.002

Gandhmal DP, Kumar K (2019) Systematic analysis and review of stock market prediction techniques. Comput Sci Rev 34:100190. https://doi.org/10.1016/j.cosrev.2019.08.001

Ge Y, Qiu J, Liu Z, Gu W, Xu L (2020) Beyond negative and positive: exploring the effects of emotions in social media during the stock market crash. Inf Process Manag 57(4):102218. https://doi.org/10.1016/j.ipm.2020.102218

Ghani NA, Hamid S, Hashem IAT, Ahmed E (2019) Social media big data analytics: a survey. Comput Hum Behav 119:417–428

Gupta R, Chen M (2020) Sentiment analysis for stock price prediction. In: 2020 IEEE conference on multimedia information processing and retrieval (MIPR), pp 213–218. https://doi.org/10.1109/MIPR49039.2020.00051

Heidari M, Jones JH (2020) Using BERT to extract topic-independent sentiment features for social media bot detection. In: 2020 11th IEEE annual ubiquitous computing, electronics & mobile communication conference (UEMCON), pp 0542–0547. https://doi.org/10.1109/UEMCON51285.2020.9298158

Javed Awan M, Mohd Rahim MS, Nobanee H, Munawar A, Yasin A, Zain AM (2021) Social media and stock market prediction: a big data approach. Comput Mater Contin 67(2):2569–2583. https://doi.org/10.32604/cmc.2021.014253

Jiang Y, Mo B, Nie H (2018) Does investor sentiment dynamically impact stock returns from different investor horizons? Evidence from the US stock market using a multi-scale method. Appl Econ Lett 25(7):472–476. https://doi.org/10.1080/13504851.2017.1340558

Kanavos P, Fontrier AM, Gill J et al (2020) Does external reference pricing deliver what it promises? Evidence on its impact at national level. Eur J Health Econ 21:129–151. https://doi.org/10.1007/s10198-019-01116-4

Khedr AE, Salama SE, Yaseen N (2017) Predicting stock market behavior using data mining technique and news sentiment analysis. Int J Intell Syst Appl 9(7):22–30. https://doi.org/10.5815/ijisa.2017.07.03

Kim KJ, Lee WB (2004) Stock market prediction using artificial neural networks with optimal feature transformation. Neural Comput Appl 13(3):255–260

Lee C, Paik I (2017) Stock market analysis from twitter and news based on streaming big data infrastructure. In: 2017 IEEE 8th international conference on awareness science and technology (ICAST), pp 312–317. https://doi.org/10.1109/ICAwST.2017.8256469

Li M, Yang C, Zhang J, Puthal D, Luo Y, Li J (2018) Stock market analysis using social networks. In: Proceedings of the Australasian computer science week multiconference (pp 1–10). Canberra, Australia: Association for Computing Machinery

Majumdar A, Bose I (2018) Detection of financial rumors using big data analytics: the case of the Bombay stock exchange. J Organ Comput Electron Commer 28(2):79–97. https://doi.org/10.1080/10919392.2018.1444337

Makrehchi M, Shah S, Liao W (2013) Stock prediction using event-based sentiment analysis. In: 2013 IEEE/WIC/ACM international joint conferences on web intelligence (WI) and intelligent agent technologies (IAT), pp 337–342. https://doi.org/10.1109/WI-IAT.2013.48

Moat HS, Curme C, Stanley HE, Preis T (2014) Anticipating stock market movements with google and wikipedia. In: Matrasulov D, Stanley H (eds) Nonlinear phenomena in complex systems: from nano to macro scale. NATO science for peace and security series C: environmental security. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-8704-8_4

Mullainathan S, Thaler (2000) RH Behavioral economics. NBER Working Paper No. w7948, Available at SSRN: https://ssrn.com/abstract=245733

Nguyen TH, Shirai K, Velcin J (2015) Sentiment analysis on social media for stock movement prediction. Expert Syst Appl 42(24):9603–9611. https://doi.org/10.1016/j.eswa.2015.07.052

Nofsinger JR (2005) Social mood and financial economics. J Behav Finance 6(3):144–160. https://doi.org/10.1207/s15427579jpfm0603_4

Otoo WM (1999) Consumer sentiment and the stock market. pp 1–20. https://doi.org/10.2139/ssrn.205028

Pagolu VS, Reddy KN, Panda G, Majhi B (2016) Sentiment analysis of twitter data for predicting stock market movements. In: 2016 international conference on signal processing, communication, power and embedded system (SCOPES), pp 1345–1350. https://doi.org/10.1109/SCOPES.2016.7955659

Porshnev A, Redkin I, Shevchenko A (2013) Machine learning in prediction of stock market indicators based on historical data and data from twitter sentiment analysis. In: 2013 IEEE 13th international conference on data mining workshops, pp 440–444. https://doi.org/10.1109/ICDMW.2013.111

Qasem M, Thulasiram R, Thulasiram P (2015) Twitter sentiment classification using machine learning techniques for stock markets. IEEE international conference on ICACCI. Kochi, India, pp 834–840

Ramesh VP, Baskaran P, Krishnamoorthy A, Damodaran D, Sadasivam P (2019) Back propagation neural network based big data analytics for a stock market challenge. Commun Stat Theory Methods 48(14):3622–3642. https://doi.org/10.1080/03610926.2018.1478103

Rao T, Srivastava S (2012) Analyzing stock market movements using twitter sentiment analysis. https://doi.org/10.1109/ASONAM.2012.30

Ren R, Wu DD, Liu T (2019) Forecasting stock market movement direction using sentiment analysis and support vector machine. IEEE Syst J 13(1):760–770. https://doi.org/10.1109/JSYST.2018.2794462

Sassi WHO, Hussainey K (2021) The impact of mandatory adoption of XBRL on firm's stock liquidity: a cross-country study. J Financ Report Account 19(2):299–324

Shah D, Campbell W, Zulkernine F (2018) A comparative study of LSTM and DNN for stock market forecasting. Paper presented at the 2018 IEEE international conference on big data (Big Data), Seattle, WA, USA, December 10–13

Sharef NM, Zin HM, Nadali S (2016) Overview and future opportunities of sentiment analysis approaches for big data. J Comput Sci 12(3):153–168. https://doi.org/10.3844/jcssp.2016.153.168

Shayaa S, Jaafar NI, Bahri S, Sulaiman A, Wai PS, Chung YW, Piprani AZ, Al-Garadi MA (2018) Sentiment analysis of big data: methods, applications, and open challenges. IEEE Access 6:37807–37827. https://doi.org/10.1109/ACCESS.2018.2851311

Sousa MG, Sakiyama K, Rodrigues L de S, Moraes PH, Fernandes ER, Matsubara ET (2019) BERT for stock market sentiment analysis. In: 2019 IEEE 31st international conference on tools with artificial intelligence (ICTAI), pp 1597–1601. https://doi.org/10.1109/ICTAI.2019.00231

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

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

van Dieijen M, Borah A, Tellis GJ, Franses PH (2020) Big data analysis of volatility spillovers of brands across social media and stock markets. Ind Mark Manag 88:465–484. https://doi.org/10.1016/j.indmarman.2018.12.006

Verma JP, Tanwar S, Garg S, Gandhi I, Bachani NH (2019) Evaluation of pattern based customized approach for stock market trend prediction with big data and machine learning techniques. Int J Bus Anal 6(3):1–15. https://doi.org/10.4018/IJBAN.2019070101

Vilares D, Alonso MA, Gómez-Rodríguez C (2017) Supervised sentiment analysis in multilingual environments. Inf Process Manag 53(3):595–607. https://doi.org/10.1016/j.ipm.2017.01.004

Vui CS, Soon GK, On CK, Alfred R, Anthony, P (2013) A review of stock market prediction with artificial neural network (ANN). In: 2013 IEEE international conference on control system, computing and engineering, Penang, Malaysia, pp 477–482, https://doi.org/10.1109/ICCSCE.2013.6720012

Wang J, Fu G, Luo C (2013) Accounting information and stock price reaction of listed companies—empirical evidence from 60 listed companies in shanghai stock exchange. J Bus Manag 2:11–21

Wu DD, Zheng L, Olson DL (2014) A decision support approach for online stock forum sentiment analysis. IEEE Trans Syst Man, Cybern Syst 44(8):1077–1087. https://doi.org/10.1109/TSMC.2013.2295353

Zhou X, Pan Z, Hu G, Tang S, Zhao C (2018) Stock market prediction on high-frequency data using generative adversarial nets. Math Probl Eng. https://doi.org/10.1155/2018/4907423

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.