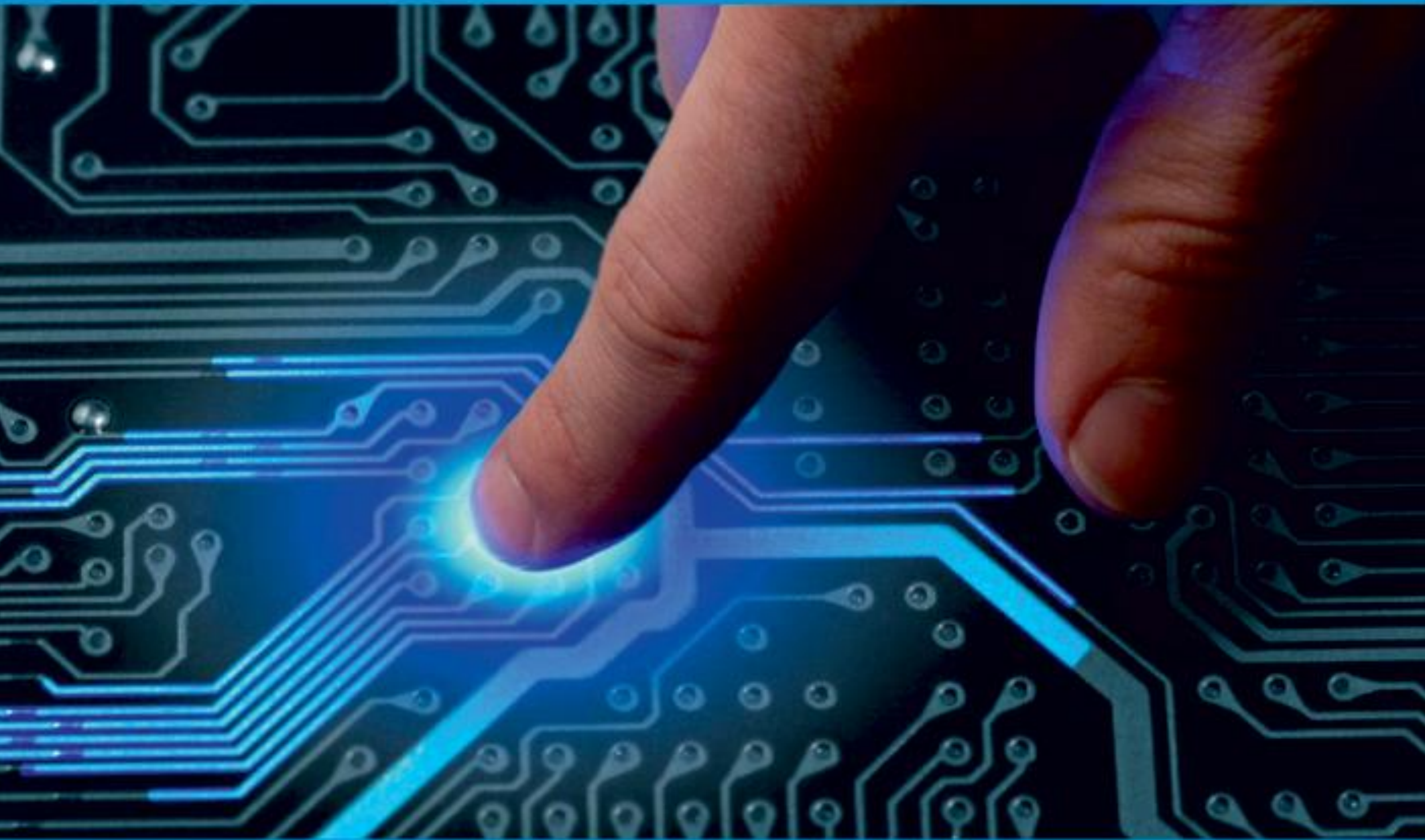
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 8, August 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.625

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

A Stock Price Prediction Model Based onInvestor Sentiment and Optimized Deep Learning

Deepa Konnur[1], Raj Ranjan[2], Patra Chinmay Mochiram [3], Raj Aryan Jasrotia [4]

Assistant Professor, Department of Computer Science & Engineering, Rajarajeswari College of Engineering, Bangalore, Karnataka, India

U.G. Student, Department of Computer Science & Engineering, Rajarajeswari College of Engineering, Bangalore, Karnataka, India

U.G. Student, Department of Computer Science & Engineering, Rajarajeswari College of Engineering, Bangalore, Karnataka, India

U.G Student, Department of Computer Science & Engineering, Rajarajeswari College of Engineering, Bangalore, Karnataka, India

ABSTRACT: Predictive sentiment analysis offers a powerful tool to understand and analyze sentiment trends related to Stockport, providing valuable insights for decision- makers, businesses, and the community.

By leveraging advanced machine learning models and analyzing diverse textual data, sentiment analysis enables real-time monitoring of public sentiment, empowering stakeholders To make well-informed choices and implement proactive actions.

The application of predictive Conducting analysis on sentiment Stockport facilitates data-driven decision-making, allowing local authorities to align policies, enhance community engagement, manage crises effectively, and foster economic development.

Overall, predictive sentiment analysis provides a valuable framework for understanding and responding to sentiment towards Stockport, driving evidence-based strategies, enhancing brand reputation, and fostering a stronger connection between local authorities, businesses, and the community.

KEYWORDS: Deep learning, LSTM model, stock price prediction, sentiment analysis, sentiment dictionary, sparrow search algorithm

I. INTRODUCTION

This study applies predictive sentiment analysis to examine sentiment trends related to Stockport, a town in Greater Manchester, England.

By analyzing online content such as examining content from social media posts as well as news articles, and online reviews, sentiment towards Stockport is predicted and evaluated.

The research aims to identify shifts in sentiment and potential factors contributing to these changes, providing insights for decision-makers, local authorities, businesses, and the community.

The study utilizes advanced machine learning models, preprocessing techniques, and feature extraction to classify sentiment and track sentiment trends over time.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The findings of this research have practical applications for local authorities, businesses, and the community, enabling informed decision-making and fostering open dialogue and understanding.

This paper introduces a novel Forecasting stock prices model, termed MS-SSA-LSTM, that integrates LSTM neural networks with the Sparrow Search Algorithm to effectively handle multi-source data characteristics. The MS-SSA-LSTM a model for predicting stock prices is capable of making forecasts regarding stock price movements in advance and help investors and traders make more informed investment decisions. Investors and traders gather specific stock data, such as historical transaction records and market shareholder comments, for input into the MS-SSA-LSTM model. Subsequently, the model autonomously generates a stock price trend chart and predicts the following day's stock price.

II. RELATED WORK

In this section, we provide a concise overview of machine learning and examine the latest studies that investigate the relationship between financial markets and data obtained from social networking platforms.

A. Machine Learning (ML)

Machine learning (ML) involves a robot or computer software acquiring knowledge (Kw) related to specific tasks (Tk) and evaluation metric (Ep), where improvement in Ep in a To address the challenges posed by unstable, unstructured, disorderly, and nonlinear time series datasets, various learning- based algorithms such as Support Vector Machines (SVM), Decision Trees (DT), and Artificial Neural Networks (ANN) are commonly utilized in stock market predictions. In this study, ANN was chosen due to its superior learning capability in solving classification, prediction, and regression problems.

RELATED WORKS:

The literature reveals that investment choices are impacted by factors beyond rationality. As a result, many studies have endeavored to comprehend the influence on investors when making investment decisions. Several sentiment analysis tools are evident in the literature, generally classified into machine learning and word count analysis methods [13]. Word count techniques employ dictionaries to ascertain sentiment (positive or negative) for individual words, aggregating their sentiment scores [13].

Primarily, negative words are tallied and assigned varying weights based on their negativity, similarly for positive words, and the side with the highest score prevails. Among word count techniques, Loughran and McDonald's financial lexicon and Harvard-IV dictionary are commonly utilized in stock market predictions. In the approach, commonly employed techniques include classification algorithms such as Support Vector Classifier (SVC), Naive Bayes, and Neural Network (NN) [5]. A primary limitation of this approach is the time-intensive manual labeling required for the training dataset. Regardless of the chosen technique, text data for sentiment analyses are typically extracted from the web, with the most prominent sources being search engine queries, financial news websites, and Twitter tweets. Consequently, our review of previous studies is categorized based on these three data sources.

A. SEARCH ENGINE QUERIES AND STOCK PRICEMOVEMENT

A examination into the relationship between Google search trends and trading volume and volatility in the stock market was conducted in [27]. The study aimed to determine if Google search queries could elucidate current stock prices and forecast future abnormal returns in stock market trends. Findings from the study revealed that Google searches were unable to forecast future abnormal returns. However, an increase in Google searchqueries were linked to increased trading volume and volatility.

Thus, the study suggests that Google searches provide better indications of future rather than recent trading activity. Similarly, an investigation into the impact of economic uncertainty and investor attention on gold price fluctuations and volatility was carried out in [28]. Interestingly, the study discovered that Google searches originating from India were more correlated with the gold market than searches from the US or other countries. Conversely, the study established that an increase in Google search volume was linked to declines in gold prices and heightened volatility. The relationship, as per the paper, was not merely correlational but also exhibited high predictive power in both directions.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

III. METHODOLOGY

MATERIALS AND METHODS

Drawing from the preceding discussion in Section II of this paper highlights the scarcity of research dedicated to examining how investor sentiment affects stock markets in developing economies. Additionally, as suggested in [37], a sentiment analysis model relying solely on public opinions from a single nation might not be universally applicable. Hence, comprehending or analyzing global sentiments may appear intricate owing to the multitude of factors significantly shaping them. Primarily, cultural beliefs and practices, variations in policies and regulatory structures across nations, and discrepancies in the sophistication of stock markets stand out. Consequently, sentiment analysis may prove more straightforward within countries or regions that share cultural similarities or proximity compared to a global scale. In consideration of this perspective, we opted to explore the impact of investor sentiments on the Ghanaian Stock Market is also a focus. Likewise, as highlighted in [23], [25], although the accuracy of stock market predictions has significantly improved, there is still room for further enhancements. enhancement by seeking out additional internet-based information sources. However, the majority of prior research outlined within Section II, researchers relied on data from various sources like Google, Twitter, Internet stock message boards, or web news separately, instead of integrating them comprehensively. Furthermore, a notable portion of earlier investigations concentrated on enhancing stock market return predictability while minimizing volatility. Hence, unlike the aforementioned studies delineated in Section II, we propose a novel approach that amalgamates diverse unstructured data sources from social networking sites to address limitations associated with single-source data. Our methodology combines public opinion data from tweets, web news, Google trends, and forum discussions into a singular input for more precise and accurate stock price trend predictions. Specifically, our study seeks to address the following research question: Can investor sentiment genuinely aid in predicting stock price movements, and to what extent? To the best of our knowledge, this study represents the first attempt to utilize social networking site data in exploring the relationship between investor sentiment and stock market volatility on the GSE. In the subsequent section, we present a detailed overview of the materials and methods employed to achieve the objectives of this study.

A. RESEARCH FRAMEWORK

Figure 1 showcases the data-pipeline framework utilized in this study. The framework consists of five separate stages: dataset description, dataset preprocessing, dataset integration, development of predictive models, and criteria for evaluating performance. Below, we provide detailed explanations of the role of each stage.
I

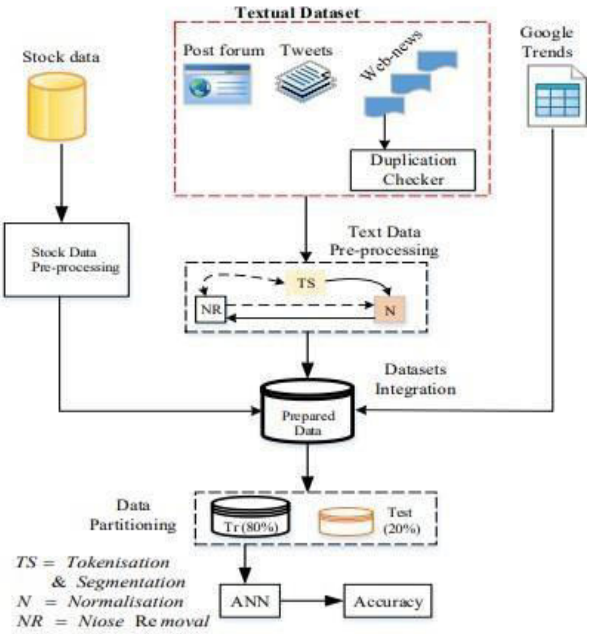


Fig. 1. Data pipeline of the study.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

B. MODEL CONSTRUCTION

HISTORICAL TRADING INDICATORS

This study centers on individual stocks as its research subject and chooses indices that comprehensively mirror fluctuations in stock prices. Key stock trading indicators encompass the preceding day's closing price, initial price, final price, highest price, lowest price, adjusted close price, daily range, volume of trades, and rate of turnover. The opening and closing prices signify the initial and ultimate prices of a stock on a given day. Trading volume indicates the quantity of shares traded. Typically, investment profits and daily losses are determined by the stock's closing price. Thus, the closing price serves as the target for prediction. The adjusted close price represents the stock price after adjusting transaction prices. Daily range illustrates the extent of fluctuation between the lowest and highest points of the stock's price during the day. Turnover rate indicates the proportion of a stock's daily turnover within its circulation, reflecting its trading frequency. The precise calculation formula is depicted below: Complex Price = $C * \beta$ (15) Daily Amplitude n = $(Hn - Ln)/Cn-1$ (16) where C represents the final price. β is the cumulative stock price adjustment multiplier. $Cn-1$ represents the final price on the $n - 1$st day. Ln represents the lowest price at the nth day. Hn represents the highest price at the nth day. The above historical indicators can directly reflect the fluctuations in stock prices and are indispensable for predicting stock prices. Because the deep learning method can mine the stock price rule from the fundamental indicator data [45], the paper only considers historical indicators of the stock.

C. SENTIMENT INDEX BUILDING

Investors' decisions are influenced by not only the technical indicators of the supplies but also the sentiments generated by the comments. Therefore, we integrate the sentiment index with fundamental trading data to forecast the stock price. This paper chooses East Money Net as the source of the text information. We employ the Octoparse tool for web data scraping. Octoparse is a robust web data extraction tool built on Visual Windows, offering comprehensive search capabilities and the ability to mimic human interactions with web pages. East Money Net serves as a valuable source of sentiment data for research, boasting the longest history among stock forum communities in China, the highest average engagement, and the most extensive user activity. On the East Money Net (http://guba.eastmoney.com/), we have Octoparse tool to mine comment data related to stock within a specific time range. The collected target fields include stock code, stock name, post title, and time. The title can get more knowledge than the content [44]. Therefore, this paper's research object is the post's title.

Firstly, we preprocess the collected text information. Empty titles, identical text content, announcements, and forwarded articles are all noisy data and must be deleted. Then, we process the obtained non-noise text data. Jieba in the Python library is used for word segmentation, and the stop word is removed to facilitate subsequent text analysis [45]. Part of the sentiment dictionary constructed in this paper comes from GooSeeker software, whose content is rich and comprehensive, including 22,215 sentiment words. The dictionary provided by the software is used as the essential sentiment dictionary. Another aspect involves acquiring financial sentiment words through statistical analysis of literature related to a stock-specific emotional lexicon. However, both parts of the sentiment dictionary require refinement for stock sentiment analysis. Comments about stocks in online forums are derived from users' opinions and often contain numerous internet slang terms. Consequently, this study constructs a stock-specific lexicon based on crawled text data to facilitate more precise sentiment analysis. Initially, word frequency analysis is conducted on segmented text, extracting the most frequently occurring words. These words are manually sorted and given sentiment values. Furthermore, sentiment words and specific expressions from stock forums are incorporated into the sentiment dictionary of the GooSeeker software to establish a novel and more accurate stock-specific sentiment lexicon. Finally, The sentiment indicator for each comment is acquired using equation (1).

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

TABLE

TABLE I
1-DAY AHEAD PREDICTION OF STOCK PRICES BASED ON PUBLIC SENTIMENTS

Data sources	Specificity	Sensitivity	RMSE	MAPE	Accuracy (%)
Google trends	0.31	0.41	0.0276	2.7948	49.40
Tweets	0.42	0.49	0.0278	2.8238	55.50
Forum post	0.29	0.40	0.0521	3.3181	41.52
Web financial news	0.41	0.45	0.0495	3.1564	50.43
Combined data	0.51	0.69	**0.0251**	**2.5193**	70.66

TABLE II
7-DAY AHEAD PREDICTION OF STOCK PRICES BASED ON PUBLIC SENTIMENTS

Data sources	Specificity	Sensitivity	RMSE	MAPE	Accuracy (%)
Google trends	0.36	0.44	0.0240	2.4267	49.89
Tweets	0.45	0.50	0.0231	2.2924	56.30
Forum post	0.30	0.41	0.0441	2.7593	41.52
Web financial news	0.46	0.49	0.0405	2.5272	51.89
Combined data	0.59	0.71	0.0200	1.9468	73.69

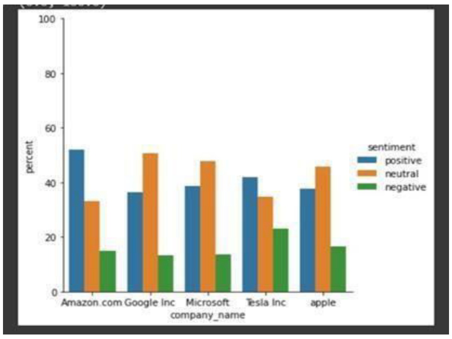
TABLE III
30-DAY AHEAD PREDICTION OF STOCK PRICES BASED ON PUBLIC SENTIMENTS

Data sources	Specificity	Sensitivity	RMSE	MAPE	Accuracy (%)
Google trends	0.38	0.49	0.0173	1.7753	50.02
Tweets	0.49	0.52	0.0162	1.6431	56.98
Forum post	0.40	0.50	0.0308	1.8717	41.58
Web financial news	0.50	0.61	0.0274	1.6495	52.94
Combined data	0.61	0.79	0.0131	1.3444	75.02

TABLE IV
60-DAY AHEAD PREDICTION OF STOCK PRICES BASED ON PUBLIC SENTIMENTS

Data sources	Specificity	Sensitivity	RMSE	MAPE	Accuracy (%)
Google trends	0.49	0.51	0.0116	1.2024	52.87
Tweets	0.51	0.69	0.0107	1.1013	58.98
Forum post	0.41	0.49	0.0205	1.2385	41.77
Web financial news	0.42	0.56	0.0178	1.0651	54.05
Combined data	0.64	0.82	0.0087	0.8951	76.57

Conversely, sentiment analysis from tweets exhibited reduced RMSE and MAPE values in specific cases. As a result, the combined data surpassed individual datasets, with Twitter tweets outperforming Google Trends, forum posts, and online financial news. This finding highlights that predicting future stock prices relies not solely on historical stock data and technical indicators but also to some extent on unstructured (fundamental) data. Thus, forecasting the stock market appears feasible through social networking platforms and sites. Further investigation using F-score can explore the impact of news dissemination (the number of shares), the quantity of comments on a news article, positive sentiment, and other factors. Analysis of sentiment, both positive and negative, on stock prices indicated that positive sentiment contributed to a 32% increase in stock prices compared to a 50% decrease attributed to negative sentiment. Additionally, the daily volume of traded stocks exhibited a stronger positive association with positive sentiments (63%) than with negative sentiments (50%). These findings underscore the significant influence of the dissemination of positive or negative sentiments across social networking sites on stock market dynamics. Once more, we noted that the dissemination of news headlines had a greater impact (64%) on stock volume compared to the quantity of comments generated regarding news coverage (62.2%), the objective was to examine how information dissemination across online platforms affects investor behavior. The results support the notion that the trading behaviors of Ghanaian investors are influenced to some extent by public news.



www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

IV. CONCLUSION

Predictive sentiment analysis offers a powerful tool to understand and analyze sentiment trends related to Stockport, providing valuable insights for decision-makers, businesses, and the community.

The application of predictive sentiment analysis in Stockport facilitates data-driven decision-making, allowing local authorities to align policies, enhance community engagement, manage crises effectively, and foster economic development.

Overall, predictive sentiment analysis provides a valuable framework for understanding and responding to sentiment towards Stockport, driving evidence-based strategies, enhancing brand reputation, and fostering a stronger connection between local authorities, businesses, and the community.

The improvement in predictive accuracy shown by the model proposed in forecasting future stock prices for 1 day, 7 days, 30 days, 60 days, and 90 days ahead (see Tables I–V) using both individual and combined datasets indicates that amalgamating stock

REFERENCES

1. Pang, B., & Lee, L. (2008). Sentiment analysis and opinion mining. Foundations and Trends in Information Retrieval, 2(1-2), 1-135.
2. Liu, B. (2012). Analyzing sentiment and mining opinions. This publication is part of the Synthesis Lectures on Human Language Technologies series., 5(1), 1-167.
3. Cambria, E., & Hussain, A. (2012). Applications of sentiment analysis. In Handbook of Natural Language Processing (pp. 653-682). Springer.
4. Liu, B. (2015). Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge University Press.
5. Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 1(12), 2009.
6. E. Khedr, S. E. Salama, and N. Yaseen, "Forecasting stock market movements through data minning techniques and sentiment analysis of news ," published in the International Journal of Intelligent Systems and Applications, volume 9, no. 7, pp. 22–30, Jul. 2017. (https://doi.org/10.5815/ijisa.2017.07.03)

related data can substantially enhance the accuracy ofstock prediction models. The findings of this research indicate.