



Design and Development of Artificial Intelligence Framework to Forecast the Security Index Direction and Value in Fusion with Sentiment Analysis of Financial News

Harmanjeet Singh¹ · Manisha Malhotra² · Supreet Singh³ · Preeti Sharma⁴ · Chander Prabha⁴

Received: 17 February 2024 / Accepted: 17 July 2024

© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd. 2024

Abstract

The domain of stock price prediction is extensively researched owing to its complex data structure and numerous influential factors. In the current epoch, many modern financial applications demonstrate non-linear and uncertain characteristics that exhibit temporal variability. As a result, there has been a notable rise in the need for resolutions to highly non-linear and time-variant issues. External factors, including public sentiment and political events, can influence stock market trends. The primary aim of this research is to propose a novel framework for predicting the blue-chip stock price in the future by combining historical stock data corresponding to the financial news published regarding stock in the newspaper. The initial step involves integrating sentiment and situational features into a machine-learning model. This integration aims to evaluate public sentiments' influence on algorithms' predictive precision thirty days ahead. Furthermore, regression models are utilized to analyze the inter-dependencies that exist among companies. To facilitate an experimental analysis, the researchers acquired historical data from the stock market through reputable sources such as the National Stock Exchange-India and collected news data about stock from the financial newspaper "EconomicTimes". A series of machine learning models were tested to facilitate the classification of news sentiments of the stocks, followed by the fusion of resultant polarities with the stock's historical data on the standard column "DATE" and then tested series of deep learning models to observe the impact of news on stock price and by taking reference of this, the authors successfully predicted the stock price on a random day. The findings unequivocally indicate that the proposed research design produces better results than other state-of-the-art models.

Keywords Comparative analysis · Economics · Sentiment analysis · Stock news · Stock price forecast

✉ Preeti Sharma
pspreeti916@gmail.com
Harmanjeet Singh
harmangahir@outlook.com
Manisha Malhotra
mmanishamalhotra@gmail.com
Supreet Singh
supreet.singh@ddn.upes.ac.in
Chander Prabha
prabhanice@gmail.com

¹ University School of Computing, Rayat Bahra University, SAS Nagar 140104, Punjab, India

² University Institute of Computing, Chandigarh University, SAS Nagar 140413, Punjab, India

³ School of Computer Science, UPES, Dehradun, 248007, Uttarakhand, India

⁴ Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India

Abbreviations

CNN	Convolutional neural network
GRU	Gated recurrent unit
RF	Random forest
SVM	Support vector machine
EMD	Empirical mode decomposition
LSTM	Long-short term memory
CNN	Convolutional neural network
ALSTM	Attention-based long-short term memory
MAE	Mean absolute error
RMSE	Root mean squared error
SVR	Support vector regression
BOW	Bag of words
CPU	Central processing unit
LDA	Linear discriminant analysis
TLBO	Teaching and learning based optimization
GAN	Generative adversarial networks
PCA	Principal component analysis
LASSO	Least absolute shrinkage and selection operator

VADER	Valence aware dictionary for sentiment reasoning
BERT	Bidirectional encoder representations from transformers
NSE	National stock exchange
NB	Naive Bayes
SGD	Stochastic gradient descent
KNN	K-nearest neighbor
GPC	Gaussian process classification
RFC	Random forest classifier
MLP	Multi-layer perceptron
CISI	Chartered Institute for Securities and Investment

Introduction

Stock market forecasting is a subject that fascinates finance and statistics professionals. One rationale for stock market prediction is to identify and acquire securities likely to experience significant price increases while simultaneously selling stocks that are anticipated to decline in value [1]. Two methodologies have been implemented to predict stock prices. The core information of a company, such as its rate of growth, position in the market, earnings, and expenditures, are the subject of essential fundamental analysis [2]. Technical analysis, the following method, employs historical stock prices to forecast market trends [3]. Using historical price charts and trends, technical analysts make predictions. Fundamental analysis concentrates on the financial statements, company assets, public sentiment, and local and global economic trends [4]. Political conditions and global business associations are also taken into account. Technical analysis employs statistical analysis to evaluate historical stock price fluctuations [5]. Technical indicators, including moving averages, dead crosses, and golden crosses, are employed by stock traders to make strategic decisions. Market analysis continues to be a complex and ongoing process despite the existence of existing methodologies [6]. Financial experts previously predicted stock markets. Computer scientists have addressed prediction challenges using novel data collection methods [7–9]. Data mining has emerged as a field of research that is dedicated to the extraction of insights from data as a result of the vast amount of data that is available for examination. Algorithms for machine learning are employed by data scientists to predict stock prices, resulting in the emergence of numerous stock prediction methods. Nevertheless, only a small number of models have taken into account the political conditions and popular sentiment. Current stock prediction models identify the critical mood factor influencing stock prices or establish correlations between Twitter sentiment and stock prices [10, 11]. This research aims to develop a deep-learning model that is devoid of emotion. This method predicts stock fluctuations

by analysing public sentiment towards a company. This novel approach evaluates individuals' perspectives regarding specific organisations. The researchers anticipate that their discoveries will enhance the performance of machine learning algorithms that predict stock prices. The research concludes with developing a hybrid automated system that integrates sentiment, historical, mixed, and final analysis. The system predicts the movements of firm share prices. The following is a summary of the paper's contributions to this study.

- Covers the groundwork for understanding and using sentiment analysis for stock forecasting.
- Using sentiment analysis of news articles, this project hopes to build a reliable module for anticipating movements in stock markets. Machine learning classifiers will be used on live news data in this component.
- Aims to present efficient modules for predicting the historical future trends of stocks. The modules provided for this purpose include:
 - The objective is to utilise deep learning classifiers to analyse past data in financial markets.
 - The study utilises streamlined selecting features along with the information normalisation procedure to improve the precision and dependability of the analysis.
- Proposes an advanced hybrid system combining Convolutional Neural Networks (CNN) with Bi-Directional Gated Recurrent Unit (GRU) and Sentiment techniques to effectively predict stock trends and movements. The utilization of the output from preceding modules is received by this particular function, which then proceeds to employ deep learning classifiers.
- Using historical stock data, the proposed method will be rigorously tested. The proposed system outperformed current methods, according to the results.

The following sections of this work are structured in the following manner: “[Preliminary knowledge](#)” presents a comprehensive review of the extant scholarly investigations undertaken in stock prediction. In this study, “[Proposed framework](#)” outlines the proposed stock prediction model. The experimental results are reported and analyzed in “[Evaluation parameter](#)”, whereas the conclusion and future directions of this paper are outlined in “[Experimental study](#)”.

Preliminary Knowledge

The primary components of the literary work encompass the following elements:

- Scientists have utilised algorithms for machine learning and many data types, such as historical, community, and journalistic data, to forecast financial patterns. The precision of these forecasts has demonstrated fluctuation.
- Researchers have proven that naive Bayes regression is the optimal predictive technique for recognising texts.
- Recent research has demonstrated that sentiments and viewpoints conveyed by the general public through Twitter posts possess the potential to serve as a valuable tool for predicting stock market trends.
- The enhancement of prediction accuracy can be achieved by using additional data sources.

Stock Price Prediction with Stock's Historical Data

Scientists have used machine learning algorithms to gather data from past records, internet posts, and news items to create prediction models for stock markets [12–16]. In the past, forecasting the stock market relied mainly on historical data since the widespread usage of networking sites and internet news sources was not as prevalent as it is today. In study [4], a range of machine learning techniques were utilized to predict the fluctuations in the Stock Market index price by leveraging past price data. The authors demonstrated that Random-Forest (RF) and Least-Square Support Vector Machine (SVM) exhibited superior forecasting performance compared to alternative predictive models.

In study [17], in the preliminary phase, the authors aim to conduct a comprehensive sentiment analysis that integrates users' sentiment preferences towards the stock, in conjunction with the stock's history data, as a determinant in predicting the stock's closing price. To calculate the sentiment index, the authors try to ensure the data's representative by picking comments that have had a higher number of likes. This study aims to analyze and authenticate the notable advancements, including the sentiment index, Empirical Mode Decomposition (EMD), and Long Short-Term Memory (LSTM) model augmented with an attention mechanism.

In [1], the authors make stock price forecasts using hybrid models powered by artificial intelligence. Compared to more traditional methods of predicting stock prices, the stacked Long Short-Term Memory (LSTM) model performs better.

The study [18] presents a comprehensive sentiment indicator for Germany. This study examines the persistence of investor sentiment within a stock market characterized by a limited involvement of retail investors. Finter et al. demonstrate that the sentiment indicator exhibits limited efficacy in forecasting future return spreads. The authors show that this indicator can explain the difference in returns between stocks sensitive to sentiment and those not. The research team posits that although the prevalence

of retail investors in the German stock market is typically limited, there may exist certain disparities between distinct segments. The Automated Trading Programme, which enabled incremental trading, was implemented by Deutsche Börse in December 2007.

In study [19], The researchers gathered analyst recommendations published in the *Frankfurter Allgemeine Zeitung*, a prominent German newspaper published daily, excluding Sundays. The distribution of buy, sell, and neutral recommendations stands at 54%, 18%, and 28%, respectively, mirroring the proportions reported in traditional print media. The authors compile and analyze weekly stock recommendations from various sources, subsequently determining the conventional bull-bear spread. Incorporating national and international sentiment indicators holds significant significance in the investment decision-making process for numerous market participants. This research thoroughly examines the relationship between stock performance and the sentiment that professional analysts convey via their published stock recommendations.

In study [20], based on examining the correlation between real-world data and network data, the authors offer a new model for dealing with financial time series. The authors plan to enhance the emotion extraction technique and incorporate additional informative variables into the model training process. They discover negative emotions dominate the network public opinion area during the trading session. The social phenomenon known as the "spiral of silence" is primarily to blame. In study [21], the authors utilized a computational approach to investigate the correlation between the sentiment expressed on Guba, a prominent financial social media platform in China, and the stock market's trajectory. Guba is widely used in the country to exchange stock-related information, making it a significant platform in this context. A positive correlation exists between the level of bullish sentiment and the number of posts and the extent of agreement. Sina Guba and Eastmoney Guba are widely acknowledged as the primary Guba platforms in China, and they are renowned for their significant user base of highly engaged participants. The researchers have devised a distributed web crawler to extract stock-related posts from Sina Guba autonomously. The study's results revealed the existence of specific equities that displayed a noteworthy positive association between their Guba feelings and stock prices.

In a recent study [22], the researchers propose an enhanced sentiment classification model based on BERT for sentiment analysis. The survey encompassed four information technology enterprises. This study employs a predictive model to estimate the closing prices on day $t + 1$ by including stock prices, technical indicators, and GuBa attitudes.

In conclusion, it can be inferred that the aforementioned prediction models employed historical data from the stock market to make forecasts regarding future stock market trends.

Social Media Data for Stock Market Forecasting

Social media lets individuals chat immediately. These systems are helpful for the financial analysis of financial securities [23, 24]. This created massive volumes of social media data that academics mined for profit. Social media investor attitudes have been studied to anticipate stock markets. In study [25], Twitter data showed varied and organized favourable, negative, and neutral attitudes. Researchers examined lexicon-, rule-, and machine learning-based Twitter sentiment categorization methods. Lexicon-based strategies employed word count and feature grading. The machine learning techniques used in this study included maximum entropy, Naïve Bayes (NB), and Support Vector Machine(SVM). A comparison was conducted between the bag-of-words (BOW) model and n-gram and part-of-speech annotations. The Bag-of-Words (BOW) approach was uncomplicated, efficient, and showed superior performance.

Researchers in study [26] used social media attitudes about company-related issues to forecast stock prices. Model accuracy was 54.41%. Sentiment analysis data improved the prediction model by 2.07%.

This analysis of leveraging social media data to forecast stock market movements demonstrates a clear association between the opinions expressed on various online networking sites and the stock market's performance.

Using Political News Stories as a Leading Indicator for the Stock Market

Publicly accessible online news articles offer a compelling data source for mining and analyzing to derive valuable insights. The news can be categorised into three primary classifications: news in general, economic news, and news related to politics. Such specific dataset can also be employed to predict stock market patterns. Political factors, much like electronic media, also have an impact on the fluctuations of stock values. Consequently, scholars have also researched examining political news about predicting stock market trends.

In study [27], the author's objective is to evaluate the efficacy of utilizing emotion factors derived from the aggregate count of news articles and Twitter tweets in making predictions about stock prices. This research compares the predictive accuracy of the two models. The algorithms employed to ascertain the emotional states of individual participants are maintained as proprietary

information. A paired sample t-test can determine the average root mean square errors (RMSEs) achieved by the two models, as they generate predictions for each day of the out-of-sample period. According to the hypothesis, the underlying probability of each result in this particular scenario is 0.5. In other words, both designs are equally reliable and effective.

In study [28], introduces a new deep-learning stock price prediction architecture. The hold-out test data-set LDA can exceed 67%. GRU outperforms LSTM in Central Processing Unit(CPU) convergence, parameter updating, and generalization. The authors calculate MSE and MAP using the expected and actual stock prices. The blended ensemble model reduces MSE by 57.55%. Tables 3 and 4 provide a synopsis of the cited relevant work. This table summarises the relevant literature regarding this study question, the many contributions made to the field, the variables we examined, and the conclusions drawn.

To sum up, we learn from the literature that stock news, social media, such as Twitter tweets, and political events affect stock market prices, trends, and returns and that these factors may be optimized through specific novel strategies. Three key areas need further investigation: these studies' correctness, performance, and connection between market movements and public opinion. The authors came to the following conclusions after doing a thorough analysis of the prior research that had been conducted in the relevant field: (1) Support Vector Machine, or SVM, is the algorithm that has proven to be the most successful for text categorization, and (2) Long Short-Term Memory (LSTM) demonstrates the prospective ability to generate accurate predictions. The authors arrived at this conclusion after doing some study on other subjects that were related to this one. Also, the authors found that using news articles as data is another technique to help boost precision and reached this conclusion after performing this research. The earlier research served as a foundation for the methodology used in this latest investigation, which directed how this study selected the data sources, tools for machine learning, and independent variables.

Proposed Framework

Figure 1 describes the brief idea about the framework design required to predict stock price direction and value in the timeline of 1-day or 1-week. The framework is divided into different phases, as explained in Fig. 2, which illustrates the proposed model's entire system architecture.

The following are the five primary steps included in the process of determining future stock prices:

Fig. 1 The suggested framework integrates long-term stock historical data and sentiment analysis of stock news published in EconomicTimes

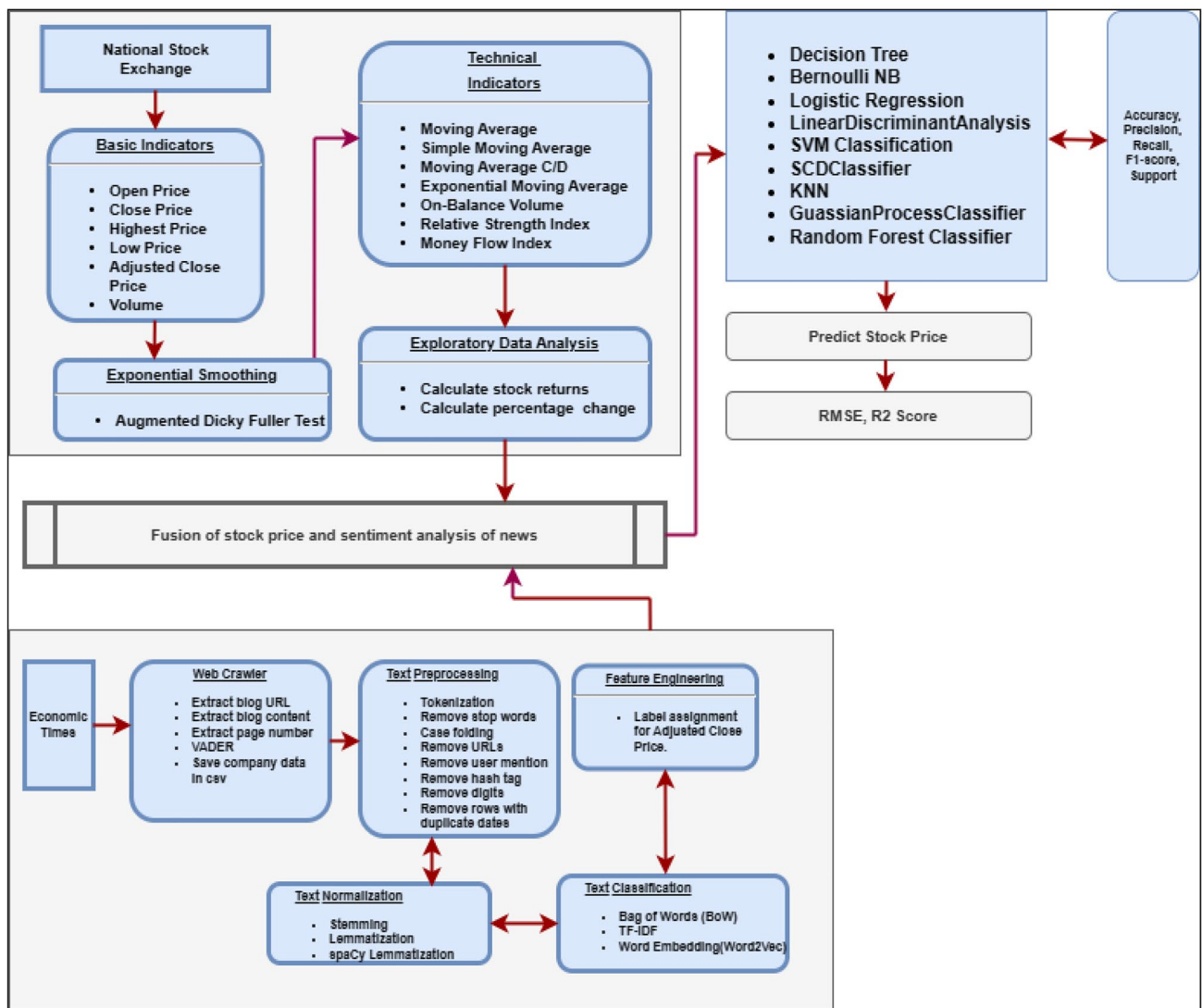
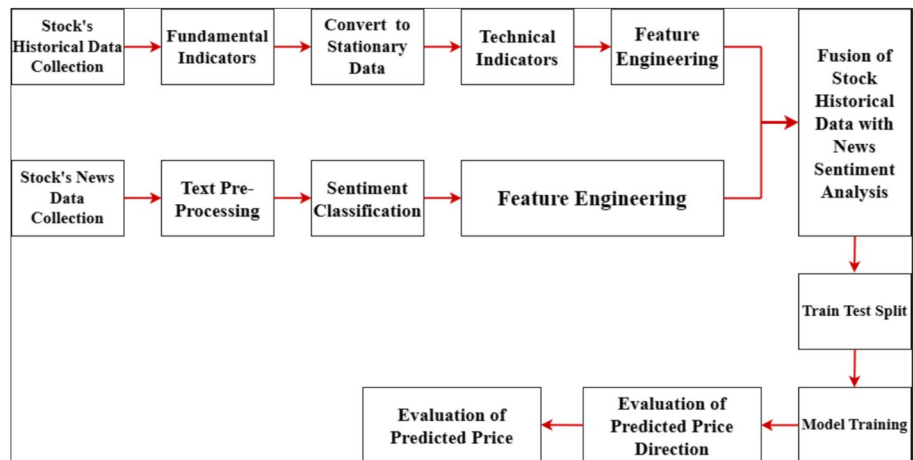


Fig. 2 The suggested framework integrates long-term stock historical data and sentiment analysis of stock news published in EconomicTimes

Methodology Involved in Stock's Historical Data

- **Collection of stock's historical data:** This step follows the collection of stock's historical data from the NSE website from April 2011 to March 2023. The authors selected OHLCV indicators and adjusted the close price using NSE website data.
- **Fundamental Indicators:** This step involves determining a stock's actual market value. Indicators such as Profit Earnings ratio, ROE ratio, Revenue growth, and return on investment significantly determine stock's underlying value and potential for future growth.
- **Stationary Data:** Forecasting or simulating non-stationary data is impossible due to the inherent unpredictability of such data. Non-stationary time series can lead to incorrect conclusions by suggesting the existence of a link between two variables even when there is none. Non-stationary data must be converted to stationary data for consistent and reliable results. Fixing the issue requires converting time series data to stationary. Statisticians use one of two methods to determine time series stationarity. The Augmented Dickey–Fuller Test (ADF) and Kwiatkowski Phillips Schmidt Shin (KPSS) can decide if a time series is stationary around a mean, linear, or non-stationary due to a unit root.
- **Technical Indicators:** A more profound understanding of investment possibilities and trade prospects can be attained through technical analysis. This involves scrutinizing statistical patterns derived from trade data, such as variations in prices and trading volume. Unlike fundamental analysts who use economic or financial data, technical analysts use analytical charting techniques, including trading signals and price movement patterns, to evaluate a security's performance. Technical indicators fall into two categories: Superimposed elements: Prices with comparable technical indications on a stock chart are valuable. Moving averages are visualised. Technical indicators like oscillators show values from a local minimum to a maximum above or below a price chart. Example: RSI, MACD, oscillators.
- **Feature Engineering:** One common feature that can be engineered for stock market index movement forecasting is a “rolling average” or “moving average.” This feature captures the smoothed historical performance of the index over a specific time window. After calculating the rolling and moving average, the authors obtained an adjusted close price for stock as a significant feature.

Methodology Involves in Stock's Financial News Data

- **Collection of stock's news data:** This stage collects financial news of long-established NSE equities from

2013 to 2023 using “pandas” to render financial news from “Economicstimes.com”. Finding each news's URL and reading the news text allowed us to retrieve the data from the news blogs. This data frame includes ‘company name,’ ‘date published,’ ‘author,’ ‘headline,’ ‘description,’ ‘articleBody,’ ‘tags,’ and ‘URL’ for each news.

- **Text pre-processing:** Each news article passed through the process of “Tokenisation”, where the authors separated the news article into smaller units called tokens, most commonly using space as a delimiter. Types can be broadly classified as word, character and subword (n-gram) tokenisation. Later, the authors removed “stopwords” since the authors convey little or no relevant information and are filtered out during text preparation to remove noise so machine learning algorithms can focus on the signal or words that constitute the text. After removing “stopwords,” the authors used stemming and lemmatization to reduce inflected/derived words to their stem, base, or root. Authors employed Porter-Stemmer to locate the root or base words. Stemming removes the last few characters, while “lemmatization” converts correctly with a corpus. It returns a lemma, a language term's primary or dictionary form, by removing inflectional endings and considering the whole vocabulary.
- **Sentiment classification:** This stage entails assessing each word in the story to obtain VADER scores based on positive or negative sentiment and emotion intensity. Stock polarity indicates buy or sell signals. The classification module classifies stocks as Increase or Decrease based on the Intrinsic Value (IV) generated by the updated and redesigned fundamental models with NSE variables and sentimental module compound polarity.
- **Feature Engineering:** The authors chose RIL based on its long-standing presence on the NSE. The authors employed TextBlob to calculate polarity and subjectivity scores for aggregated and refined news data. Polarity, unlike the discrete values of 1 and -1, representing positive and negative, respectively, is a continuous variable between these two extremes. Subjective statements express emotions, views, or judgements, while objective statements present factual information. Subjectivity, like objectivity, can be quantified on a scale from 0 to 1. On the other hand, the measurement of subjectivity is done in reverse, with 0 signifying total objectivity and 1 representing total subjectivity. The authors built and trained ML classifier models to forecast sentiment-based price fluctuations. To describe the likelihood of the Adjusted Close price going up, staying unchanged, or down the following day, these models were constructed to assign a value of 1 or 0.

Fusion of Stock Historical Data with Stock News

Fusion of features generated from stock’s historical and news data based on the standard “Date” field. Splitting whole data into train and test data concerning the 70:30 ratio. This phase involves training a model on 70% train-data to predict future stock prices and evaluating its precision, recall, F1 score and accuracy. 30% test data is examined and assessed after model training. Once the predicted stock price direction is evaluated, the model is trained on 70 per cent training data to predict the share price and evaluated with state-of-the-art parameters such as RMSE and R2. After model training, 30% test data is tested with a trained model to determine future stock prices.

Evaluation Parameter

The study demonstrated its influence by conducting a comparison analysis using categorised evaluation criteria. This analysis monitored changes in stock prices within the framework of the classification task. Metrics such as precision, recall, precision, f1-score, R-squared, and Root Mean Squared Error are illustrative examples of these measures. The summary of evaluation parameters is shown in Table 1.

Experimental Study

Stock Historical Dataset

The empirical data for this inquiry is obtained from India’s significant national stock market, the NSE. Figure 3 exhibits

Table 1 Overview of model evaluation metrics

Parameter	Description	Formula
Accuracy	The metric represents the instances in which the model accomplished flawless recognition	$\frac{TR+TF}{TR+FR+FF+TF} * 100$ (1)
Precision	Precision in prediction refers to accurately forecasting the percentage of actual stock gains or losses	$= \frac{TR}{TR+FR} * 100$ (2)
Recall	This figure is the proportion the model predicts the stock price will increase or decrease	$= \frac{TR}{TR+FF} * 100$ (3)
F1	The F1-Score links the right stock movement (rise/fall) to the projected stock when the precision and recall of a predictor are equal	$= \frac{2 * P^R * R^R}{P^R + R^R}$ (4)
R2	The performance index measures the degree of correlation between predicted and observed values, ranging from 0 to 1	$= \frac{\sum_{i=1}^n (t_i - \bar{t})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (t_i - \bar{t})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}$ (5)
RMSE	This performance indicator will quantify the disparity between the observed (ti) and projected (yi) values	$= \sqrt{\frac{i}{n} \sum_{i=1}^n (t_i - y_i)}$ (6)



Fig. 3 The dataset includes the Nifty50 Index data from April 1, 2011 to March 31, 2023

the final values of the Nifty50. The red bars represent periods of volatility in the index caused by the worldwide economic downturn 2008 and the COVID-19 epidemic.

The Nifty50 index is a constituent of the NSE indices. The index consists of fifty key stock indices that significant organisations use. This article utilises fundamental indicators, including stock price highs, lows, trading volumes, and closing prices. This methodology predicts the corrected closing price index of Reliance Industries by using data from diverse Indian and overseas sources.

RIL has the most significant load among Nifty Companies for 2023, amounting to 10.41%. ICICI Bank ranks as the third largest bank, holding a market share of 7.34%, while HDFC Bank follows closely with a market share of 9.06%. The stock index data for the Reliance Industries company was available on the NSE website from April 1, 2011, to March 31, 2023. Figure 4 depicts the closing price index of Reliance Industry Limited. Reliance Industry shares underwent a resurgence in 2011, so it was chosen as the year after the 2008 global crash.

Feature Engineering

The authors utilised the TA-lib module to derive a variety of financial signs that have been demonstrated to have a substantial influence on stock prices based on previous research. RSI, SMA_5, SMA_20, EMA, MACD, signal, Stochastic RSI_fastk, Stochastic RSI_fastd, Stochastic Oscillator Index_slowk, stochastic oscillator index_slowd, WilliamR, Momentum, and ROC were selected as the following attributes: OHLCV. The maximum point of a stock should be considered to be the price at the end of a trading period, while the low point should be regarded

as the price at the beginning of the trading period. The volume of shares or contracts exchanged within a specific time frame is a metric of market activity known as trading volume. The RSI is a momentum oscillator that uses the rate and change of price movements to identify overbought and oversold conditions. The SMA sum all closing values over a specific period. Consider the 5-day moving average (SMA_5) and the 20-day moving average (SMA_20) as examples. The EMA can more effectively respond to recent price movements by designating them as having a higher weight. The MACD is a momentum indicator illustrating the interaction between two price-moving averages to monitor trends. Signal lines, the moving averages derived from MACD lines, are the critical buy-and-sell signals that traders and investors rely on. To improve the accuracy of forecasts, stochastic RSI_fastk and stochastic RSI_fastd, calculated using the stochastic oscillator and RSI, respectively, effectively identify potential price reversal sites. To guarantee that the stochastic oscillator was smooth, it was assumed that it contained supplementary components, specifically the stochastic oscillator indices_slowk and _slowd. Williams R, or William's %R, was an additional critical investigation element. To thoroughly comprehend the market's sentiment, this momentum indicator determines whether the market is overbought or oversold. Momentum is the subsequent indicator that is implemented. Momentum may be employed to determine the rate of price fluctuations. Momentum quantifies the rate of change in stock prices, thereby supplying valuable information about the rate of price change. Lastly, the ROC is a momentum-like statistic that establishes the extent of price fluctuations by evaluating price changes over a predetermined period.

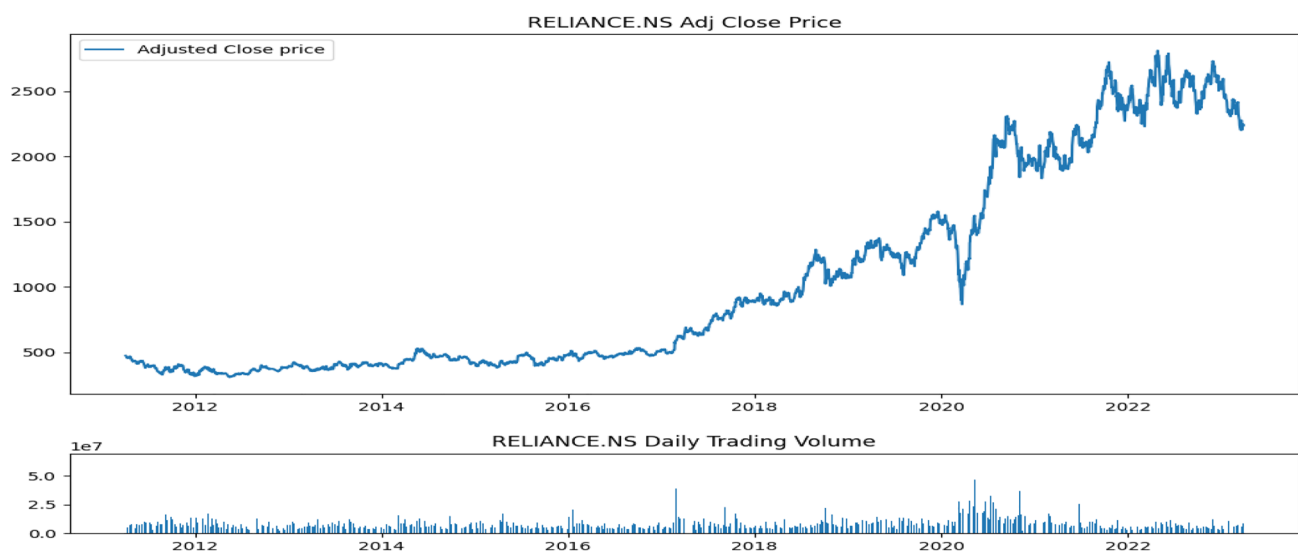


Fig. 4 Stock data for Reliance Industries is available from April 1, 2011 to March 31, 2023

FinBERT Method for Calculating Sentiment Indices Using Financial Bidirectional Encoder Representations

Before FinBERT sentiment analysis, news data was stop-word removed and lemmatized. FinBERT, based on BERT, encodes text bi-directionally and is a practical NLP and comprehending language model. FinBERT specialises in domain knowledge by retraining BERT’s model with financial data. FinBERT predicts positive, negative, or neutral financial text sentiment. Texts may include financial news, reports, and websites. Equation (1) gave negative sentiments a Te score of 0 and positive sentiments 1.

$$sentimentscore = \frac{M_{tpos} - M_{meg}}{M_{tpos} + M_{meg}} \tag{7}$$

A study [29] determined sentiment measures by counting positive and negative posts in a dataset. On day *t*, *M_{tpos}* represents positive news stories and *M_{meg}* represents negative articles. The sentiment index varied from -1 to 125. A sentiment index value near -1 indicates unfavourable news for that date. However, if it’s near to 1, news is generally good. Before inputting features into the framework, a min-max scaler standardised their values from 0 to 1.

The stock index for Reliance Industries was accessible on the NSE Website for the period spanning from April 1, 2011, to March 31, 2023. Due to the widely recognized reputation of “<https://www.economictimes.com>” as a prominent financial news platform renowned for its high-quality news material, we opted to utilize it as a source for our stock news data-set analysis. Specifically, we focused on scanning articles linked to Reliance Industries that were published over the timeframe spanning from 2011 to

2023. The TextBlob tool determined the “Subjectivity” & “Polarity” grades of the processed and summarised news.

The “Valence Aware Dictionary for Sentiment Reasoning” model was used to analyse social media financial news sentiment. We created “price increase” and “price decrease” indicators using the Bidirectional Encoder Representations from Transformers (BERT) sentiment analysis approach. Following the stock news sentiment analysis, we used date-based inner joins to combine the news data set with the stock’s history data on the column field “DATE”. Analysis of Decision Trees The programme accurately classified 84% of stock price changes, outperforming competing models, identifying 87% of stock price increases and 83% of stock price reductions.

Deep Learning Models

In addition to daily price predictions for the Reliance Industry, the authors evaluated the system using deep-learning models. Bidirectional RNNs can consider the sequence’s past and subsequent contexts. RNNs with recurrent layers can be utilised as time series models since they capture short-term factors that affect stock prices. Bi-RNN’s customisable structure makes it useful for pattern processing with many types of time series data. Bi-LSTMs are better than RNNs with LSTM cells. The models excel in learning long-range dependencies for time-series forecasting and other sequential data tasks. Table 2 compares deep learning models on modeling-relationship, train-test, and actual-predicted datasets using R2 and RMSE.

Table 2 The findings about classifying the upcoming day’s gain or reduction in a stock’s price

Model	Increase ^a			Decrease ^b			Accuracy
	Precision	Recall	f1-score	Precision	Recall	f1-score	
Decision Tree	0.87	0.83	0.88	0.83	0.80	0.86	0.84
Bernoulli NB	0.72	0.74	0.71	0.71	0.70	0.73	0.77
Logistic Reg	0.79	0.71	0.72	0.77	0.71	0.73	0.78
LDA	0.67	0.73	0.77	0.73	0.79	0.75	0.72
SVM	0.73	0.72	0.73	0.78	0.79	0.74	0.76
SGD	0.86	0.82	0.86	0.82	0.80	0.82	0.81
KNN	0.80	0.79	0.72	0.72	0.74	0.74	0.79
GPC	0.78	0.77	0.76	0.72	0.74	0.73	0.74
RFC	0.72	0.71	0.73	0.74	0.73	0.78	0.75
Adaboost	0.72	0.71	0.71	0.68	0.69	0.67	0.70
XGBoost	0.70	0.70	0.69	0.69	0.68	0.68	0.71

^a Predictor of a subsequent day’s rise in the share price of Reliance Industries

^b Predictor of a subsequent day’s fall in the share price of Reliance Industries

Results and Discussion

Reliance Industries, Adani Ports, and Bharati Airtel are three stocks traded on the National Stock Exchange (NSE) of the Indian stock market. These stocks are utilized as a data set for testing and analyzing the enhanced recurrent rider LSTM presented and approaches already in use. The open, high, low, adjusted close, and close values, as well as volume, subjectivity, polarity, label classification, and various other technical indicators, are all included in the dataset attributes. The Coefficient of determination (R2) is a direct measure of how well the model performs in terms of other metrics like accuracy, precision, and recall. Put another way, the coefficient of determination quantifies the variation in the response variable y , which can be accounted for by changes in the predictor variable x . It is widely used as a proxy for the model's reliability. The root mean squared error (RMSE) shows the average squared difference between a dataset's actual and predicted values. Both R2 and RMSE are considered to be metrics for assessment. If a model's RMSE number is lower and the R2 number is high, the model will have greater accuracy.

The strategy to perform a comparative analysis of the proposed model is described in three ways as follows:

- The outcomes of the proposed model were assessed about R2 and RMSE metrics and were compared to those of other advanced deep learning models.
- Adani Stock and Bharti Airtel are two well-established companies whose performance the suggested models matched favourably with.
- In terms of R2 and RMSE, the suggested model's outcomes were compared to those of other researchers' models.

The initial step was to forecast Reliance Industries' stock price on an arbitrary day using the suggested model and compare it to other state-of-the-art algorithms. Table 3 shows that when compared to other models, the Decision-Tree model reliably predicted that the stock price of Reliance Industries would go up or down the next day.

Table 3 shows that the Decision Tree model performed best when predicting whether to buy or sell Reliance Industries stock the next day. Then, for 50 days, 21 days, 14 days, and 5 days the exponential moving average was calculated. The forecast column does an excellent job of dividing signals into two groups: purchase and sell.

Table 4 shows the results of machine learning and deep learning models in terms of their respective coefficients of determination and root mean squared errors in modelling relationship, train-test split, and the actual-predicted price of a stock. Illustrates the results of these models in terms of their respective coefficients of determination and root mean squared errors. Because the proposed model had the highest R2 value (97%) and the lowest RMSE value (3.16), the performance of the Convolutional Gated Recurrent Unit was superior to that of other models. This was because it had the best R2 value.

Table 5 compares ML and DL models to determine the discrepancy between the actual and anticipated RIL values as of June 5, 2023.

The Convolutional Gated Recurrent Unit yielded a root mean square error (RMSE) of 3.16 and a coefficient of determination of 97%. When conducting a comparison of the stock price of Reliance Industries on a randomly selected day. This RMSE score is lower than the 41.479 reported by the Linear Regression model and the 40.580 written by the Multi-Layer Perceptron model. On June 5, 2023, the stock's

Table 3 The outcomes about the classification of whether to buy or sell Reliance Industries stock on a later day

Model	Buy ^a			Sell ^b			Accuracy
	Precision	Recall	f1-score	Precision	Recall	f1-score	
Decision Tree	0.86	0.88	0.85	0.83	0.86	0.87	0.84
Logistic Reg	0.81	0.82	0.86	0.84	0.83	0.88	0.83
Bernoulli NB	0.81	0.83	0.83	0.81	0.84	0.81	0.82
LDA	0.84	0.81	0.77	0.82	0.81	0.77	0.79
SVM	0.85	0.74	0.76	0.80	0.82	0.78	0.79
SGD	0.82	0.82	0.84	0.83	0.83	0.81	0.82
KNN	0.81	0.78	0.78	0.71	0.81	0.82	0.80
GPC	0.83	0.82	0.78	0.77	0.83	0.82	0.81
RFC	0.81	0.81	0.79	0.80	0.77	0.76	0.76
AdaBoost	0.81	0.81	0.78	0.78	0.79	0.81	0.81
XGBoost	0.82	0.83	0.80	0.81	0.80	0.82	0.79

^aAn indicator suggesting that investors may consider purchasing shares of Reliance Industries

^bAn indicator suggesting that investors may consider selling shares of Reliance Industries

actual value, 2477.25, was relatively close to the predicted value of 2471.13.

In the second stage of comparative analysis, the proposed model trained on historical data-sets of Adani Ports and Bharti Airtel, where historical data of the stocks generated

from the NSE website and news data for both stocks crawled from Economics Times from April 2011 to March 2023. In Table 6, the results of the Convolution Gated Recurrent Unit model compared with the above-mentioned industrial stocks are shown:

Table 4 Reliance industries stock modified close price prediction statistical comparison

Applied model	Modelling relationship		Train-test		Actual-predicted	
	R2	RMSE	R2	RMSE	R2	RMSE
LR(Linear Regression)	0.77	209.28	0.79	34.90	0.71	41.47
Multi-Layer Perceptron	0.73	182.28	0.80	32.18	0.72	40.58
VADER+Conv. RNN	0.82	93.23	0.80	36.40	0.78	42.45
VADER+Conv. LSTM	0.98	26.98	0.98	4.67	0.97	5.95
VADER+Conv. GRU	0.98	73.59	0.98	36.12	0.97	30.26
VADER+Conv. Bi-directional RNN	0.97	65.34	0.98	17.23	0.96	19.01
VADER+Conv. Bi-directional LSTM	0.98	58.34	0.96	12.38	0.96	14.34
VADER+Conv. Bi-directional GRU	0.98	18.34	0.98	4.72	0.97	3.16

Table 5 Comparison of the statistical accuracy of the models used to forecast the price of Reliance Industries stock on June 5, 2023

Model	Actual-Predicted ^a		Actual-predicted	
	R2	RMSE	Predicted price	Actual price
Linear regression	0.71	41.47	1768.34	2477.25
MLP	0.72	40.58	1745.69	2477.25
VADER+Conv. RNN	0.78	42.45	1695.23	2477.25
VADER+Conv. LSTM	0.97	5.95	2453.87	2477.25
VADER+Conv. GRU	0.97	30.26	1889.51	2477.25
VADER+Conv. Bi-directional RNN	0.96	19.01	1956.23	2477.25
VADER+Conv. Bi-directional LSTM	0.96	14.34	2109.73	2477.25
VADER+Conv. Bi-directional GRU	0.97	3.16	2471.13	2477.25

^aIndicators of actual and predicted values of determination of coefficient and RMSE

Table 6 A statistical analysis was conducted to compare the outcomes of the proposed model across several equities to anticipate price fluctuations on June 05, 2023

Stock	Methodology ^a		Evaluation ^b		Stock Price ^c	
	Historical data (in years)	Technique	R2	RMSE	Predicted price	Actual price
Adani ports	12	VADER + CBGRU	0.96	4.23	728.46	738.80
Bharti airtel	12	VADER + CBGRU	0.94	5.12	821.62	834.50
Bajaj finance	10	VADER + CBGRU	0.94	4.03	1444.54.52	1458.90
Asian paints	10	VADER + CBGRU	0.95	4.99	3175.34.62	3199.75
Infosys	12	VADER + CBGRU	0.93	3.94	1296.83	1304.50
Wipro	12	VADER + CBGRU	0.97	3.84	411.72	404.20
Maruti	12	VADER + CBGRU	0.96	3.98	9574.82	9582.10
Tata motors	12	VADER + CBGRU	0.95	4.93	552.53	546.25
Axis Bank	10	VADER + CBGRU	0.94	4.17	963.12	950.85
HDFC Bank	10	VADER + CBGRU	0.95	4.06	1619.62	1606.50

^aIndicator of the methodology adopted in investigation of individual stock price prediction process

^bIndicator of state-of-the-art parameters for evaluation with respect to R2 and RMSE

^cIndicator of predicted price generated by proposed model on date June 5, 2023, compared with the actual stock value on same day

It reveals the degree to which the data clusters around the best-fitting difference. Lower MSE and higher R2 values indicate a better match. The root mean square error (RMSE) is a reliable metric for evaluating the accuracy of the model's predictions about the response variable. If making predictions is the main reason for building the model, then this fit criterion is essential.

The proposed model is evaluated on other blue-chip stocks based on historical data from twelve years, as shown in Table 6, and the difference between actual and predicted stock values was observed on random days on June 05, 2023.

In the end, The outcomes of the suggested model were juxtaposed with the outcomes of models formulated by other researchers, utilizing the dataset specific to the proposed model, encompassing the price fluctuations of Reliance Industries stock from April 2011 to March 2023. Table 7 presents the data sources chosen by the other researchers in 2020, 2021, and 2022.

Table 8 presents the outcomes of the suggested model, which demonstrate a substantial coefficient of determination and the lowest root mean square error (RMSE) in predicting the stock price on a randomly selected day. This scenario examines outcomes from forecasting the stock price on a randomly selected day.

According to the findings of other researchers, multi-models generally perform better than individual models, and this pattern is consistent across the many kinds of data sets examined. Even though there isn't a vast statistically significant difference between the hybrid models, It has been found that the utilization of these models yields superior outcomes in predicting reviews, as opposed to the conventional method of employing deep learning models. Despite no substantial and statistically significant disparity between the hybrid models, this phenomenon persists.

Conclusion

The current investigation has significantly contributed to comparing machine learning and deep learning models within the specific context of forecasting stock prices

Table 8 Statistical comparison of the proposed model with other researchers' models to predict reliance industries stock price

References	Technique	Evaluation ^a	
		R2	RMSE
[17]	EMD+LSTM	0.97	3.19
[30]	CNN+LSTM	0.91	3.17
[31]	LSTM+GRU	–	27.42
Proposed model	VADER+CNN+Bi-Directional GRU	0.97	3.16

^aIndicator of evaluation parameters to evaluate model

and market indices. The main objective of this study is to investigate the potential influence of public opinion on the stock market performance of particular corporations. The study used NSE and the "EconomicsTimes" as sources of information about public opinion and financial statistics. This study comprehensively evaluated eleven models to ascertain their effectiveness in forecasting the future trajectory of stock values, specifically in determining whether they would have an upward or downward movement in the subsequent trading day. The Decision Tree model demonstrated superior accuracy, including a signal indicating whether investors should engage in purchasing or selling the investment. Although the proposed model underwent a comparative analysis with other deep learning models, subsequent findings revealed its ability to achieve a low root mean square error (RMSE) and a high coefficient of determination (R2) in forecasting stock prices on arbitrary days. Given the limited scope of this study, which exclusively focuses on English-language news, it is essential to acknowledge that the proposed models and conclusions may lack generalisation when applied to a more heterogeneous context. Subsequent investigations about non-English languages could potentially derive advantages from adapting the currently proposed framework to incorporate these linguistic variations effectively. Additionally, our research reveals that machine learning algorithms performed significantly inferior to the neural network models within an identical context. Nevertheless, deep-learning models exhibit

Table 7 Data source of the proposed model and other research models to predict stock's price

Reference	Published year	Data source ^a			
		Stock/market	Stock data source	News data source	Duration (in years)
[17]	2020	Apple	Yahoo finance	Stockwits	5
[30]	2022	Shanghai Stock exchange	Tushare finance finance	CISI	1
[31]	2021	Shanghai Composite Index	NetEase finance	CISI	14
Proposed model	–	Reliance industries	NSE	Economics times	12

^aIndicator of the source of stock historical data and its respective news publication data

high computational costs despite their widely acknowledged superiority over machine-learning alternatives. To enhance the predictive performance of machine learning, future research endeavours may focus on exploring optimization strategies and incorporating supplementary ensemble boosting techniques.

Author Contributions Harmanjeet Singh: conceptualization, methodology, writing—original draft. Manisha Malhotra: investigation, writing—review and editing, supervision. Supreet Singh: Writing—review and editing, software. Preeti Sharma: methodology, writing—review & editing. Chander Prabha: validation, writing—review & editing.

Funding The authors did not receive support from any organization for the submitted work.

Availability of data and materials The data that support the findings of this research are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics approval This material is the authors' own original work, which has not been previously published elsewhere.

References

- Singh H, Malhotra M. Artificial intelligence based hybrid models for prediction of stock prices. In: 2023 2nd International Conference for Innovation in Technology (INOCON); 2023. pp. 1–6. <https://doi.org/10.1109/INOCON57975.2023.10101297>.
- Murphy JJ. Technical analysis of the financial markets: a comprehensive guide to trading methods and applications. London: Penguin; 1999.
- Turner T. A beginner's guide to day trading online. 2nd ed. New York City: Simon and Schuster; 2007.
- Nti KO, Adekoya A, Weyori B. Random forest based feature selection of macroeconomic variables for stock market prediction. *Am J Appl Sci*. 2019;16(7):200–12.
- Li M, Li W, Wang F, Jia X, Rui G. Applying bert to analyze investor sentiment in stock market. *Neural Comput Appl*. 2021;33:4663–76.
- Nti IK, Adekoya AF, Weyori BA. A systematic review of fundamental and technical analysis of stock market predictions. *Artif Intell Rev*. 2020;53(4):3007–57.
- Singh H, Malhotra M. A novel approach of stock price direction and price prediction based on investor's sentiments. *SN Comput Sci*. 2023;4(6):823.
- Singh S, Mittal N, Nayyar A, Singh U, Singh S. A hybrid transient search naked mole-rat optimizer for image segmentation using multilevel thresholding. *Expert Syst Appl*. 2023;213: 119021.
- Singh H, Malhotra M. A time series analysis-based stock price prediction framework using artificial intelligence. In: International Conference on Artificial Intelligence of Things. Springer; 2023. pp. 280–89.
- Wang Y. Stock market forecasting with financial micro-blog based on sentiment and time series analysis. *J Shanghai Jiaotong Univ (Science)*. 2017;22:173–9.
- Singh H, Malhotra M. Stock market and securities index prediction using artificial intelligence: a systematic review. *Multidiscip Rev*. 2024;7(4):2024060–2024060.
- Wu J-L, Huang M-T, Yang C-S, Liu K-H. Sentiment analysis of stock markets using a novel dimensional valence-arousal approach. *Soft Comput*. 2021;25:4433–50.
- Singh G, Mantri A, Sharma O, Kaur R. Virtual reality learning environment for enhancing electronics engineering laboratory experience. *Comput Appl Eng Educ*. 2021;29(1):229–43.
- Li M, Chen L, Zhao J, Li Q. Sentiment analysis of Chinese stock reviews based on bert model. *Appl Intell*. 2021;51:5016–24.
- Sharma B, Mantri A. Assimilating disruptive technology: a new approach of learning science in engineering education. *Procedia Comput Sci*. 2020;172:915–21.
- Hájek P. Combining bag-of-words and sentiment features of annual reports to predict abnormal stock returns. *Neural Comput Appl*. 2018;29:343–58.
- Jin Z, Yang Y, Liu Y. Stock closing price prediction based on sentiment analysis and lstm. *Neural Comput Appl*. 2020;32:9713–29.
- Finter P, Niessen-Ruenzi A, Ruenzi S. The impact of investor sentiment on the German stock market. *Zeitschrift für Betriebswirtschaft*. 2012;82:133–63.
- Singer N, Laser S, Dreher F. Published stock recommendations as investor sentiment in the near-term stock market. *Empir Econ*. 2013;45:1233–49.
- Zhang G, Xu L, Xue Y. Model and forecast stock market behavior integrating investor sentiment analysis and transaction data. *Cluster Comput*. 2017;20:789–803.
- Sun Y, Fang M, Wang X. A novel stock recommendation system using guba sentiment analysis. *Pers Ubiquitous Comput*. 2018;22:575–87.
- Ji Z, Wu P, Ling C, Zhu P. Exploring the impact of investor's sentiment tendency in varying input window length for stock price prediction. *Multimed Tools Appl*. 2023;82:1–35.
- Lux T. Sentiment dynamics and stock returns: the case of the German stock market. *Empir Econ*. 2011;41:663–79.
- Balshetwar SV, Rs A. Fake news detection in social media based on sentiment analysis using classifier techniques. *Multimed Tools Appl*. 2023;82:1–31.
- Pagolu VS, Reddy KN, Panda G, Majhi B. Sentiment analysis of twitter data for predicting stock market movements. In: 2016 international conference on signal processing, communication, power and embedded system (SCOPE5). IEEE; 2016. pp. 1345–50.
- Nguyen TH, Shirai K, Velcin J. Sentiment analysis on social media for stock movement prediction. *Expert Syst Appl*. 2015;42(24):9603–11.
- Vanstone BJ, Gepp A, Harris G. Do news and sentiment play a role in stock price prediction? *Appl Intell*. 2019;49:3815–20.
- Li Y, Pan Y. A novel ensemble deep learning model for stock prediction based on stock prices and news. *Int J Data Sci Anal*. 2022;13:1–11.
- Wu S, Liu Y, Zou Z, Weng T-H. S_i_lstm: stock price prediction based on multiple data sources and sentiment analysis. *Connect Sci*. 2022;34(1):44–62.
- SUN Y. Prediction of shanghai stock index based on investor sentiment and cnn-lstm model. *J Syst Sci Inf*. 2022;10(6):620–32. <https://doi.org/10.21078/JSSI-2022-620-13>.
- Gao Y, Wang R, Zhou E. Stock prediction based on optimized lstm and gru models. *Sci Program*. 2021;2021:1–8.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the

author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.