A Novel Approach to Stock Price Direction and Price Prediction Based on News Sentiments

Harmanjeet Singh[1*] and Manisha Malhotra[2†]

[1*]University Institute of Computing, Chandigarh University, Sahibzada Ajit Singh Nagar (Mohali), 140413, Punjab, India.
[2*]University Institute of Computing, Chandigarh University, Sahibzada Ajit Singh Nagar (Mohali), 140413, Punjab, India.

*Corresponding author(s). E-mail(s): harmangahir@outlook.com; harmanjeetsingh.uic@cumail.in;
Contributing authors: mmanishamalhotra@gmail.com;
†These authors contributed equally to this work.

Abstract

Forecasting stock trends guide investment management, financial policy, and the country's economic growth. Investor-generated textual information has impacted stock movements across media channels in recent years. Most sentiment index studies weigh linguistic content equally. Such studies ignore that the sentiment index's impact on the stock market decreases over time. This study analyses stock indices using dual classifier coupling and sentiment analysis. A dual classifier is created by combining two popular classifiers, Decision Tree (DT) with Convolution Bi-Directional Gated Recurrent Unit (GRU). The proposed model is tested using Reliance Industries shares. The adjusted sentiment index improved overall accuracy in the Reliance Industries stock news sentiment analysis case study by 84.12 percent. The investor sentiment indicator improves stock index trend prediction, as shown by a 3.16 RMSE (Root Mean Squared Error) and 0.97 R2(Coefficient of determination) reduction. The adjusted sentiment index improves predicted accuracy considerably. The investors' sentiments improve the overall results in Reliance Industries' stock price prediction with our fusion of proposed VADER (Valence Aware Dictionary and sEntiment Reasoner) and CNN + BDGRU models compared to benchmark models.

Keywords: Economics, Stock Price forecast, Reliance Industries, Sentiment Analysis, Stock News

1 Introduction

Big financial data lets risk managers assess assets beyond their portfolios. This utility incorporates financial market systematic risk to improve risk predictions. Thus, precise risk prediction capabilities for market players can help financial regulators stabilize financial markets and reduce systemic risk. Stock market forecasting has a deep academic history. The efficient market hypothesis, proposed by Fama [1], states that past information is fully absorbed into stock prices in optimal conditions, leaving stock prices only vulnerable to new information. However, rigorous postulates make other academics challenge the idea. Active investment strategies use fundamental, technical, and quantitative analysis. Behavioral finance has raised awareness of market irrationality. The herd effect can induce stock market movements in response to the news. Network public opinion data can be analyzed statistically to predict stock market values. Our keyword augmentation technique uses Bidirectional Encoder Representation from Transformers (BERT) to give financial institutions more recent time series data from web search indexes. This method can improve risk management and help them adapt to market changes. Fig. 1 shows the research plan used to combine sentiment analysis into market predictions in the next part. Sentiment analysis and autoregression were considered with an exogenous factor model (ARX) to achieve goal. Section 3 explained the suggested model's design, using *"EconomicTimes"* for news data and NSE for stock history data. The fourth section describes a process experiment. Section 5 describes the model's performance parameters. The results begin with graphical representations of each stock's sentiment scores in Section 6. A Box plot shows the sentiment score's link with the stock's starting price. The results and pertinent studies are discussed in this section. The study's main findings follow.

2 Literature Review

This section briefly discusses how news items can predict stock values. Next, study stock market prediction technology literature.

2.1 The use of sentiment analysis in the prediction of stock market trends

Scholars and practitioners have long studied stock trend forecasting. Investor sentiment is critical in determining stock changes. The study by [2] showed that investor sentiment broadly affects stock prices. According to the author, the quality of news and social media information determines financial market predictability. Thus, using the information source makes sentiment analysis for stock prediction easier. Text mining allows researchers to compile textual information gleaned from online sources such as online media, and internet searches [3] and pre-trained models on massive data-set. Investor sentiment affects the stock market. Online platforms enable securities market commentary. Thus, investors' investing decisions affect the stock market's performance [4].

Using the least squares method, [5] created a novel investor sentiment index that outperforms many existing sentiment indicators. The study examined the 2008 subprime crisis and China's 2015–2016 stock market upheaval before and after the financial crisis and the impact of the investor sentiment index on the stock market. Most of the sample's mood indicators were shown to be effective only before and after the financial crisis. . These sentiment measures worked in crisis and non-crisis situations.

A multidimensional investor sentiment index covers macro, meso, and micro dimensions. The micro-sentiment index was created by crawling Oriental Fortune Stock Bar Forum comments from 2015 to 2018 [6]. This article calculates the investor sentiment index by counting bullish and bearish articles over time. It does not account for post-readership.

Turn forum postings into an investor sentiment index to study how investor sentiment affects market movements. However, many literary works ignore how audience size affects the market, particularly the difference between small and large audiences exposed to the news. Investors also forget news over time. Thus, investor sentiment will gradually lose its impact on the stock market. Later, analyzed the above data and created a weighted investor sentiment index to fill the research gap. This index will test stock index prediction accuracy.

2.2 Techniques in the prediction of stock market trends

Long Short-Term Memory (LSTM) RNNs can learn and understand datasets' long-term dependencies. They are suitable for analyzing and predicting major patterns with extended time intervals and delays in time series. The study discusses the creation and usage of Long Short-Term Memory (LSTM) to predict stock market patterns. Recurrent neural networks (RNNs) are widely used for index forecasting [7–9].

The vanishing gradient issue makes this methodology untrustworthy for prediction outcomes. Long Short-Term Memory (LSTM) networks can solve RNNs' disappearing gradients. LSTM networks improve stock price prediction by replacing RNN hidden layer units with memory cells [10]. According to multiple studies, LSTM networks surpass other neural network topologies in prediction accuracy. LSTM is used in many fields, including natural language recognition [11], time series prediction (particularly stock price prediction) [12], water desalination [13], [14], material sciences [15], laser technology [16], and material processing [17].

Early stock market predictions use mathematical statistics. Mathematical statistics studies random phenomena using probability theory and statistics. The autoregressive integrated moving average (ARIMA), and ARMA models are utilized in this discipline. These statistical models can anticipate stationary random processes but not non-stationary time series. Nonlinear properties allow several machine learning systems to anticipate stock prices. Artificial intelligence has improved stock market forecasts. RF and SVM are the main machine-learning algorithms. The framework also includes advanced deep-learning architectures like LSTM and CNN. Scholars use several methods to choose machine learning and deep learning algorithms for stock market analysis.

The authors used AI-based hybrid models to predict stock prices in [18]. The model was trained using 14-year ADANI stock price data. The stacked Long Short-Term Memory (LSTM) model outperforms conventional stock price prediction models in various conditions. Predicting stock values has become a practical economics topic. This study uses a convolutional network and a stacked Long Short Term Memory (LSTM) learning model.

In [19], researchers used stock market technical index data to test deep learning's ability to predict stock market developments. The CNN model predicts individual stocks with 69.89% accuracy.

Technical and fundamental studies benefit from stock market trend prediction methods in the literature. Machine learning algorithms include SVM, KNN, LR, NB, DT, and RFC. Fuzzy logic algorithms are also notable. Several studies have used online stock news sentiment analysis and historical stock data.

3 The Proposed Model

The schematic representation of the proposed model's comprehensive system architecture is depicted in Figure 1. The methodology involves eight key procedures for predicting stock prices:

- For separate news reports, the NSE and the Economic Times websites were consulted to gather historical data on stocks traded on those exchanges.
- pre-processing of stock historical data on the basis of its adjusted close price index to obtain stationary.
- pre-processing of news data related to stocks obtained from the Economic Times categorizing
- price changes based on the news publication
- merging the news data frame with the stock data frame
- classifying whether to increase or decrease in stock price
- classifying whether to buy or sell stocks
- adjusting the learning rate schedule and optimizing the output units

Preprocessing removes unnecessary data and makes EconomicTimes news data compatible with the prediction method. The classification data frame concerning stock price direction is also merged with historical stock data from the NSE website. Next, machine learning algorithms classify the data frame into increasing and decreasing price change categories. Hyperparameter tuning was used to classify whether to purchase or sell Reliance Industries stock.

Prediction of the price of a stock using the Convolutional Neural Network(CNN) + Bi-Directional Gated Recurrent Unit(GRU) model. Optimizers based on Adam perform adaptive learning rate calculations for method parameters. These learning rates were stored as average square gradients with an exponentially decaying function.

Determine the appropriate GRU output to improve the model's predictive power. Lemmatization uses semantic meaning to determine a word's lemma algorithmically.

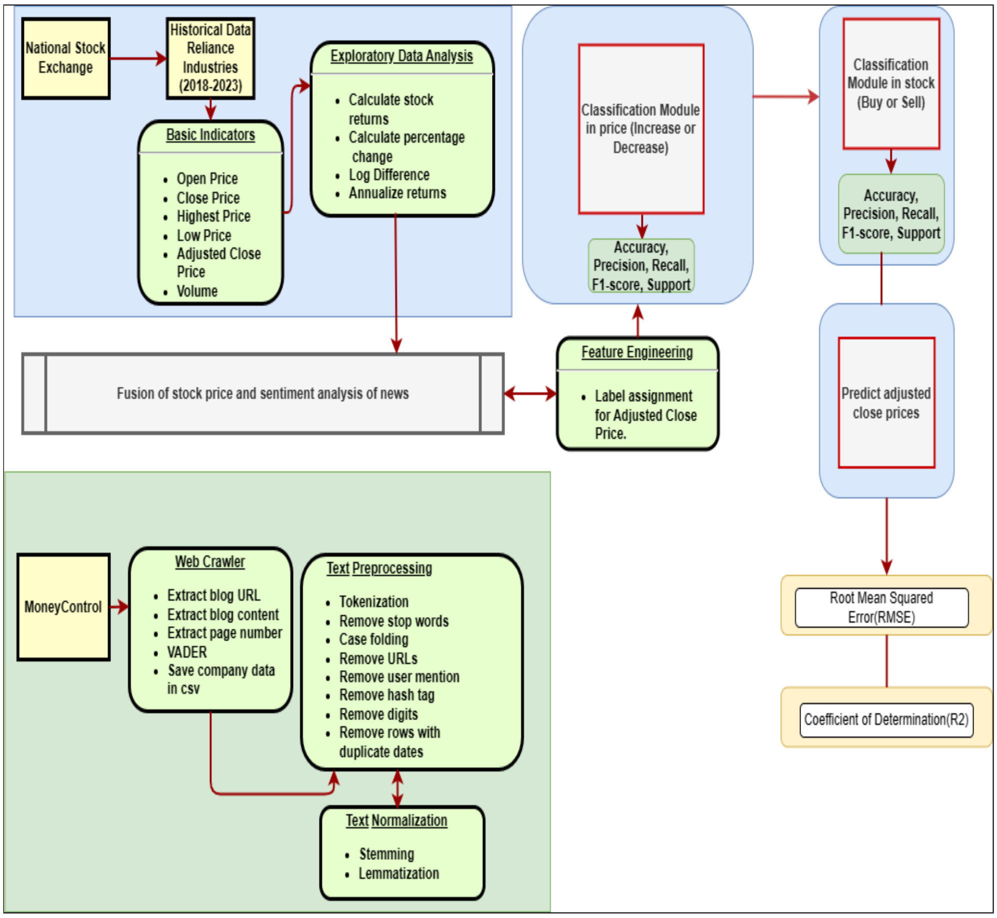


Fig. 1 The proposed architecture combines historical stock data with sentiment analysis of internet stock news.

4 Experimental Study

4.1 Data Collection and Preprocessing

The National Stock Market (NSE) in India, a significant stock market, provides experimental data for this study. Fig. 2 shows the Nifty50 closing prices, with red bars highlighting instances when the index fluctuated due to the 2008 global recession and Covid-19.

NSE indexes include the Nifty50. It includes 50 well-known stock indexes that large companies use. This article uses fundamental indicators such as stock prices' maximum, open, volume, lowest, and close. This research focuses on Reliance Industries' broad investor base in India and globally, then predicts this company's adjusted close price index. Reliance Industries Ltd. has the highest weightage of 10.41% in the 2023 Nifty Companies ranking. HDFC Bank is second with 9.06%, and ICICI Bank is third

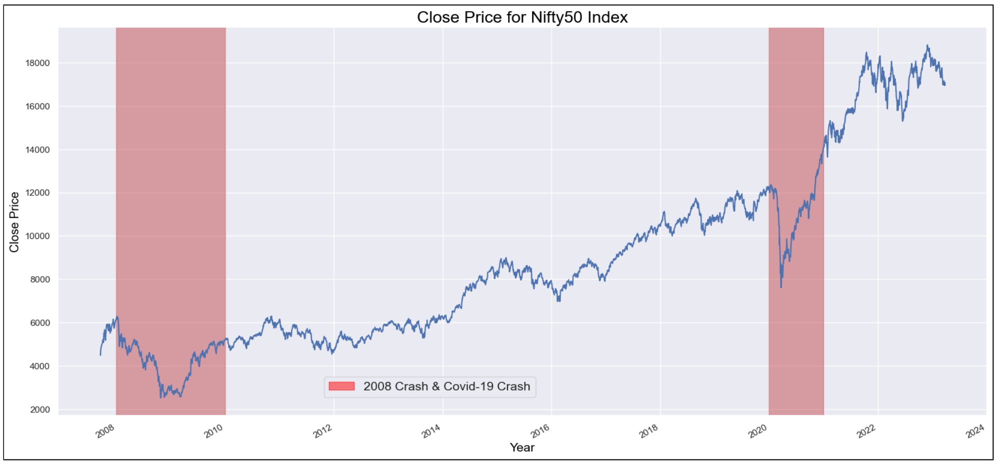


Fig. 2 Nifty50 Index data-set from April 1, 2011 to March 31, 2023

with 7.44%. Reliance Industries, historical stock index data, was taken from NSE Website from April 1, 2011, to March 31, 2023. Fig. 3 shows the close price index. 2011 was chosen because Reliance Industry stock recovered from the worldwide slump. It aids in price prediction.

Fig. 3 Reliance Industries stock data from April 1, 2011 to March 31, 2023

4.2 Sentiment Analysis

4.2.1 News Collection

EconomicTimes.com is a popular Indian online forum noted for its high-quality news material used in this study for crawling Reliance Industries news published on this platform from 2011 to 2023. The news dataset is split into sentences and stored in.csv format.

4.2.2 News Text Preprocessing

Python's extensive standard library includes BeautifulSoup, a software tool that extracts and analyses internet data [20]. After identifying the data source, Python script using BeautifulSoup used to extract Economic Times headlines. This study preprocesses news data to remove superfluous information. This stage also removes noise from the supplied dataset.

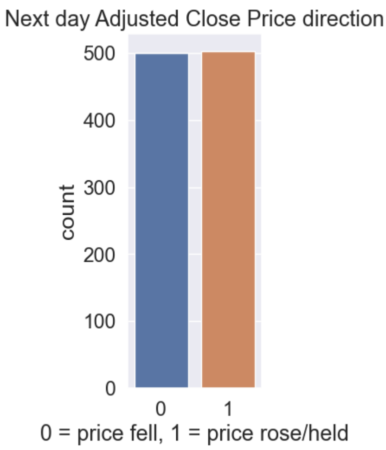
- Regular expressions can be used to remove URLs and other similar entities.
- Negations should be written in full rather than truncated. "Isn't" can be expanded to "is not," "can't" to "cannot," "couldn't" to "could not," "hasn't" to "has not," "hadn't" to "had not," and "won't" to "will not," among others.
- The punctuation marks, such as the full stop, comma, hyphen, parenthesis, and forward slash, should be removed.
- To replace the occurrence of "Username" with "usr" using regular expression equivalent, the following code can be implemented.
- The utilization of hashtags (#) provides valuable information, thereby rendering the removal of the hashtag symbol unnecessary. The replacement of "Especially '#Lee'" with simply "Lee" has been implemented.
- Eliminate the digits, email, mentions, and HTML tags from news text.
- The objective is to remove all stop words, such as "the," "as," and "is," from the news text.
- Replace the white spaces within the text with individual white spaces.
- Stemming, lemmatization, and tokenization improve natural language processing tasks. Stemming reduces words to their roots. However, several English words' roots were not English after stemming. NLTK tasks use lemmatization to achieve great precision.

VADER sentiment analysis followed. Pandas, a Python-based data analysis tool, was used to analyze and retrieve financial article headline sentiment analysis scores. Sentiment analysis requires a model. The VADER model, short for "Valence Aware Dictionary for Sentiment Reasoning," is a simple rule-based method for large-scale sentiment analysis. The model above is sensitive to emotion polarity and intensity, making it acceptable for unannotated textual data. VADER helps Python programmers analyze human language data [21]. VADER performed similarly to eleven well-known sentiment analysis techniques. The VADER technique used to analyze financial news items shared on social media yielded a negative, neutral, positive, and compound probability. The Economic Times relied on major financial news publications, which sometimes required preprocessing for research:

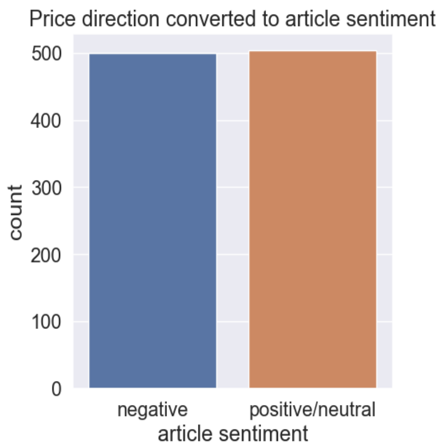
- Eliminate rows either without publishing date or incorrect date format
- Eliminate rows with duplicate titles.
- Eliminate rows with duplicate dates.
- Combine articles that were published on the same date

Subjectivity and polarity scores for processed and consolidated news datasets were obtained using TextBlob. ”Polarity” is between ’-1’ and ’1’. Positive statements are ’1’ while negative statements are ’-1’. Subjective statements express feelings and judgments, while objective words state facts. ”Subjectivity” is a variable in the interval [0,1], with ’0’ signifying objectivity and ’1’ denoting subjectivity. The Label column was formed by assessing whether the Adjusted Close price increased, stayed the same, or decreased the next day.

After merging the appropriate groups, ’1’ means prices climbed or remained the same, whereas ’0’ means prices declined. The study studied how news story sentiment affected Reliance Industry stock prices the next day. It was achieved by developing the BERT model for sentiment analysis, preparing and formatting text, building a Sentiment Classifier using the Hugging Face transformers library, and testing the model.



(a) Reliance Industry stock’s next day adjusted close price direction



(b) Sentiments obtained from adjusted close price direction

Fig. 4 Data visualization for price direction and sentiment label counts

In Fig. 4, adding the class where prices decreased to the class where prices were held yields a balanced distribution. To comprehend it better, we will categorize the data into negative (a price fall suggests pessimism) and neutral (a price increase indicates optimism).

In our supervised work, we fine-tune a pre-trained BERT model to classify sentiment as positive (1) or negative (0) using the BertForSequenceClassification class. We add an untrained linear layer to the pre-trained BERT model to classify article sentiment. As we add data, this linear layer is added to the pooled output and trained alongside the pre-trained BERT model. Unprocessed text data cannot be used to train BERT. Hence the dataset must be converted. To input our text into BERT, we must partition it into tokens and assign them to respective tokenizer vocabulary indices. A pre-trained BertTokenizer will conduct tokenization. We used ”bert-base-uncased” tokenization. To train this model, the WordPiece sub-word segmentation algorithm was used to tokenize lowercase English text. The tokenization vocabulary is 30,000 words—table 1 random filtered tokenization.

Table 1 Tokenizer to one random article

Filtered:	Aditya Birla Money bearish Reliance Industries recommended sell rating stock target Rs April research report
Tokenized:	['adi', '##tya', 'bi', '##rl', '##a', 'money', 'bear', '##ish', 'reliance', 'industries', 'recommended', 'sell', 'rating', 'stock', 'target', 'rs', 'april', 'research', 'report']
Token IDs:	27133, 21426, 12170, 12190, 2050, 2769, 4562, 4509, 17975, 6088, 6749, 5271, 5790, 4518, 4539, 12667, 2258, 2470, 3189

The printed version removes non-alphabetic characters. The BertTokenizer deconstructs Out-Of-Vocabulary (OOV) phrases like "Aditya" and "RCOM" into sub-words using "##." Adding tokens to the BertTokenizer and retraining the model fixed this issue.BERT requires formatted inputs after retraining. BERT requires "special tokens" at the beginning and end of each successive span of text, which is done by attaching a "special token" to each sentence to create separation and prepending [CLS] at the beginning.

Padding and truncating all phrases to 512 tokens ensures consistency. Padding handles sentences under the maximum length. Adding a BERT vocabulary padding token at index 0. This padding completes the sentence. This attention mask instructs BERT's "Self-Attention" mechanism to ignore padding tokens. The dataset's maximum length is 3687 after tokenization, exceeding the model's 512 token limit. Splitting the input into 200-token chunks improves processing. Each segment was entered into the base model for analysis, overlapping by 50 tokens.

4.2.3 Similarity Selection

Based on pretraining parameters and BERT vectorization, the multilayer stacked encoder mechanism vectorizes stock price predictor variables. Next, semantically similar words were selected. This study uses word vector cosine values to compare words. Cosine similarity for each seed term about stock price prediction and its candidate words. 18,267 stock index prediction keywords were found at 0.95. Initial screening yielded these keywords and text context in Table 2.

Using the BERT vectorization model, the preliminary screening process involves the computation of similarity scores. This process enables the model to eliminate words that exhibit a minimal correlation with the seed vocabulary, which is predicted based on the stock index.

4.2.4 Fusion of Sentiment Score with BERT articles

This analysis combined contextual information of candidate keywords and applied downstream finetuning tasks to initially screened words. This method lets us carefully choose contextual keywords. This study uses "www.economictimes.com" news text

Table 2 BERT similarity findings

Seed Keyword	Candidate words	Cosine similarity	Results
Stock market	Liquid funds	0.9514	Keep
	Equity	0.9617	Keep
	Financial trading	0.9564	Keep
	Bank policy	0.7964	Remove
	Amendments	0.6370	Remove
	Market capital	0.9514	keep
Insurance Plan	Premium	0.9754	Keep
	Loan	0.8257	Remove
	Long term	0.9589	Keep
	Discount rate	0.7492	Remove
	Maturity amount	0.9453	Keep
	Nominee	0.9767	Keep
Intra-day trading	Entry Point	0.9766	Keep
	Exit Point	0.9643	Keep
	Stop-loss	0.9712	Keep
	Compounding	0.8190	Remove
	Lock-in period	0.7956	Remove
	Candle charts	0.9512	Keep

dataset. To balance the training sample, a corresponding number of pseudo keywords are randomly picked from the text and aligned with the manually labeled keywords. The standard data set is then created, consisting of tuples with the text, keyword, or pseudo keyword and a binary value of 0 or 1 mentioned in Table 3.

Table 3 Sentiment score and BERT classification

BERT filtered news text	Predicted Prob.	Subjective	Polarity	BERT class
Portfolio Manager, PN Vijay is of the views that ONGC far better than Reliance Industries. Vijay told CNBC-TV18,	neg:0 neu:0.805 pos:0.195 compound:0.4404	0.355	0.125	0
Reliance Industries has support at around Rs 780-760, says Rakesh Gandhi, Senior Tech Analyst at LKP.	neg:0.058 neu:0.763 pos:0.179 compound:0.8271	0.371	0.083	1

4.3 Feature Engineering

Our analysis has focused on assessing the sentiment polarity of Reliance Industries news stories from April 1, 2011, to March 31, 2023, using the BERT pre-trained model "BertForSequenceClassification." This methodology predicts sentiment polarity classes and BERT subjectivity and polarities. We also incorporated Reliance Industries stock data from April 1, 2011, through March 31, 2023. Merging data uses both data frames' date columns. Date-based inner joins integrated the data frames. The BERT classification model's sentiment polarity of released news regarding Reliance Industries on a random day increased the stock price. We also determined the adjustment close price change of the following day, categorizing it as an increase, no change, or a drop. These observations were utilized to build and train classifier models to predict sentiment-based price direction.

5 Experimental Evaluation Parameters

We analyze stock price variations under the classification problem. Comparative analysis using a variety of categorized evaluation measures showed the study's impact. These metrics include accuracy, precision, recall, f1-score, coefficient of determination(R Square), and Root Mean Squared Error (RMSE).

6 Results

Multiple indicators predict Reliance Industries stock's modified closing stock price. Table 4 shows how well the Decision Tree model predicts stock price percentage change the following day. Identifying 87% of stock price increases, the algorithm had a classification accuracy of 84%. The SGD classifier correctly classified stock price fluctuations the following day at 86%. Due to its lower accuracy than the Decision Tree classifier, its 81% accuracy needed to be improved for future investigation. The Decision Tree model correctly predicts an 83% stock price drop the following day, the largest of any model.

Table 4 Confusion Matrix results on classifying Increase or Decrease in adjusted close price on subsequent day

	Increase[1]			Decrease[2]			
Model	Precision	Recall	f1-score	Precision	Recall	f1-score	Accuracy
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

[1]Indicator of increase in Reliance Industries stock price change on subsequent day

[2]Indicator of decrease in Reliance Industries stock price change on subsequent day

Table 5 shows that the Decision Tree model was most accurate in predicting whether to purchase or sell Reliance Industries stock the following day. In this study, we calculated the exponential moving average for 50, 21, 14, and 5 days. A prediction column classified signals as buy or sell. 80/20 split the data frame into training and test data. This section helped classify Reliance Industries shares for purchase or sale. Table 5 presents the three most accurate models of the ten. The Decision Tree model accurately predicted stock buy/sell decisions. Later, we estimated a stock's value if the Decision Tree correctly recognized an upward price trend and indicated "Buy"

Table 5 Confusion Matrix results on classifying Buy or Sell Reliance Industries stock on subsequent day

	Buy[1]			Sell[2]			
Model	Precision	Recall	f1-score	Precision	Recall	f1-score	Accuracy
Decision Tree	0.86	0.88	0.85	0.83	0.86	0.87	0.84
AdaBoost	0.80	0.94	0.81	0.84	0.89	0.84	0.82
Logistic Reg.	0.81	0.82	0.86	0.84	0.83	0.88	0.83

[1]Indicator to the investor to buy Reliance Industries stock

[2]Indicator to the investor to sell Reliance Industries stock

in the first two phases. We used 14 regression models to predict the 30th-day price. Fig. 5 shows that linear models and neural networks outperform. Analysis used Linear Regression and MLP Regressor models. Both linear and non-linear models use the Gradient Descent technique.

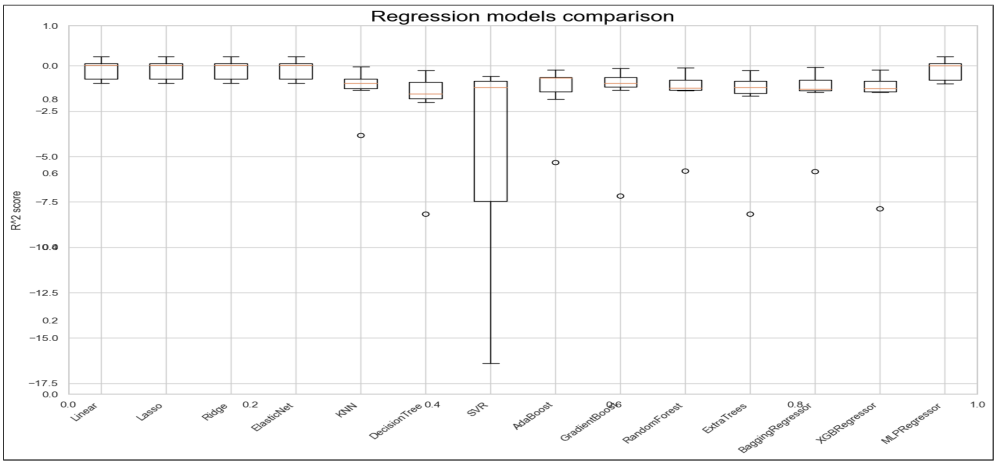


Fig. 5 Box plot comparison of regression models on R2 score

Table 6 presents the results of the linear regression analysis. Initially, the model produced a root mean square error (RMSE) 209.28. The calculation of RMSE in the train-test split came next, producing a result of 41.47. Afterward, The model was used to predict the adjusted close price of Reliance Industries stock, resulting in an RMSE of 41.47. In contrast, the multi-layer perceptron model demonstrated a Root Mean Square Error (RMSE) of 40.58 when predicting the adjusted close price, indicating superior performance compared to the Linear Regression model.

Finally, deep learning is used to reduce MSE. This study used the extended recurrent neural network (RNN), specifically the LSTM network. The following two reasons explain why its investigational efficacy falls short. An RNN may need help finding the best historical observation window size, resulting in poor variational property retrieval. Gradient explosions can also occur while using gradient descent to process sequential data. A four-recurrent-layer, 50-neuron model was trained. 60-time steps

Table 6 Statistical comparison among models to predict Reliance Industries stock adjusted close price

	Modelling relationship[1]		Train-Test[2]		Actual-Predicted[2]	
Model	R2	RMSE	R2	RMSE	R2	RMSE
Linear Regression	0.77	209.28	0.79	34.90	0.71	41.47
MLP	0.73	182.12	0.81	31.18	0.72	40.58

[1]Indicator of dependency target variable on independent variables

[2]Indicator of dependency target variable on independent variables after train-test split

[3]Indicator of model evaluation to predict stock price on subsequent day

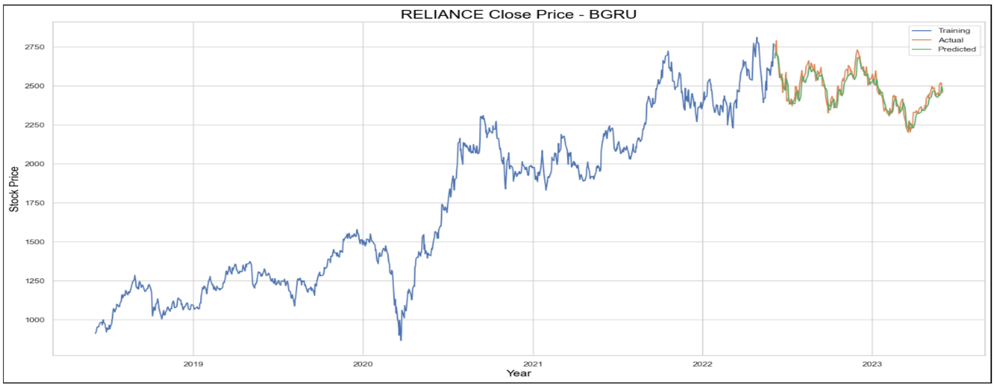


Fig. 6 Reliance Industry stock's 30th day adjusted close price direction

and 1-dimensionality characterized the input data. The output layer used tanh activation and had a one-time step. Dropout is used after each hidden layer to reduce overfitting. The Mean Squared Error loss function and Adam optimizer will compile the model. It will then be fitted on the training set for 200 epochs using 64 batches.

In Table 7, the proposed model generated a root mean square error (RMSE) 3.16 when comparing a model's predicted price to Reliance Industries' stock price on a random day. This RMSE number is better than the Linear Regression model's 41.47 and Multi-Layer Perceptron model's 40.58.

Table 7 Statistical results of the proposed model predict Reliance Industries stock price on June 05, 2023

	Modelling relationship[1]		Train-Test[2]		Actual-Predicted[2]	
Model	R2	RMSE	R2	RMSE	R2	RMSE
VADER+CBGRU	0.98	18.34	0.98	4.72	0.97	3.16

[1]Indicator of dependency target variable on independent variables

[2]Indicator of dependency target variable on independent variables after train-test split

[3]Indicator of model evaluation to predict stock price on subsequent day

Fig. 6 demonstrates that the LSTM model predicted the test data graph, matching the line. On June 5, 2023, the anticipated value of 2471.13 was close to the accurate stock price of 2477.25.

RNN and GRU were also trained using an identical manner. These hybrid models' RMSE worsened.

7 Conclusion

Data scientists and professionals find stock market index prediction fascinating and demanding. Predicting stock market trends accurately helps investors maximize profits. Since stock markets underpin national economies, such estimates also help governments. This section summarises the research findings. Scholars forecast stock values using a single time series or machine-learning algorithm. However, their low predictive power drew criticism. This paper proposes a new method for precisely and effectively predicting Reliance Industries' daily closing pricing. We combined Reliance Industries' stock data with Economic Times company news using an innovative way. Thus, the proposed technique predicts outcomes well in nonlinear and non-stationary datasets. Subcomponents after data segmentation into numerous sentiment components and a single adjusted closure price disqualify the proposed model as an ensemble model. The proposed model outperformed other hybrid or individual sentiment analysis models in Accuracy, Precision, f1-score, Recall, and three stock data statistical measures: coefficient of determination, R square, and RMSE. These results indicate that the proposed approach improves stock market prediction.

The recommended strategy can be applied to stock market data from multiple nations in future research to compare XGBoost, SVM, and ANN methods. Interest rates, political climate, and exchange rates can be inputs for DT and ANN models. We will also test our model on nonlinear and non-stationary time series data. We found that integrating sentiment data and the LSTM model improves Reliance Industries' daily adjusted closing price forecasts.

Declarations

- **Conflict of interest** The authors declare that they have no conflict of interest
- **Ethics approval** This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] Fama, E.F.: Efficient capital markets: A review of theory and empirical work. The Journal of Finance **25**(2), 383–417 (1970). Accessed 2023-07-01

[2] Nassirtoussi, A.K., Aghabozorgi, S., Wah, T.Y., Ngo, D.C.L.: Text mining for market prediction: A systematic review. Expert Systems with Applications **41**(16), 7653–7670 (2014)

[3] Ma, Y., Mao, R., Lin, Q., Wu, P., Cambria, E.: Multi-source aggregated classification for stock price movement prediction. Information Fusion **91**, 515–528 (2023)

[4] Wang, Q., Xu, W., Zheng, H.: Combining the wisdom of crowds and technical analysis for financial market prediction using deep random subspace ensembles. Neurocomputing **299**, 51–61 (2018)

[5] Gong, X., Zhang, W., Wang, J., Wang, C.: Investor sentiment and stock volatility: New evidence. International Review of Financial Analysis **80**, 102028 (2022) https://doi.org/10.1016/j.irfa.2022.102028

[6] Long, W., Zhao, M., Tang, Y.: Can the chinese volatility index reflect investor sentiment? International Review of Financial Analysis **73**, 101612 (2021) https://doi.org/10.1016/j.irfa.2020.101612

[7] Wang, Q., Xu, W., Huang, X., Yang, K.: Enhancing intraday stock price manipulation detection by leveraging recurrent neural networks with ensemble learning. Neurocomputing **347**, 46–58 (2019)

[8] Berradi, Z., Lazaar, M.: Integration of principal component analysis and recurrent neural network to forecast the stock price of casablanca stock exchange. Procedia computer science **148**, 55–61 (2019)

[9] Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation **9**(8), 1735–1780 (1997)

[10] Sundermeyer, M., Ney, H., Schlüter, R.: From feedforward to recurrent lstm neural networks for language modeling. IEEE/ACM Transactions on Audio, Speech, and Language Processing **23**(3), 517–529 (2015)

[11] Vo, N.N., He, X., Liu, S., Xu, G.: Deep learning for decision making and the optimization of socially responsible investments and portfolio. Decision Support Systems **124**, 113097 (2019)

[12] Ding, G., Qin, L.: Study on the prediction of stock price based on the associated network model of lstm. International Journal of Machine Learning and Cybernetics **11**, 1307–1317 (2020)

[13] Xu, P., Xie, F., Su, T., Wan, Z., Zhou, Z., Xin, X., Guan, Z.: Automatic evaluation of facial nerve paralysis by dual-path lstm with deep differentiated network. Neurocomputing **388**, 70–77 (2020)

[14] Elsheikh, A.: Bistable morphing composites for energy-harvesting applications. Polymers **14**(9), 1893 (2022)

[15] Moustafa, E.B., Hammad, A.H., Elsheikh, A.H.: A new optimized artificial neural

network model to predict thermal efficiency and water yield of tubular solar still. Case Studies in Thermal Engineering **30**, 101750 (2022)

[16] Elsheikh, A.H., Shehabeldeen, T.A., Zhou, J., Showaib, E., Abd Elaziz, M.: Prediction of laser cutting parameters for polymethylmethacrylate sheets using random vector functional link network integrated with equilibrium optimizer. Journal of Intelligent Manufacturing **32**, 1377–1388 (2021)

[17] Khoshaim, A.B., Moustafa, E.B., Bafakeeh, O.T., Elsheikh, A.H.: An optimized multilayer perceptrons model using grey wolf optimizer to predict mechanical and microstructural properties of friction stir processed aluminum alloy reinforced by nanoparticles. Coatings **11**(12), 1476 (2021)

[18] Singh, H., Malhotra, M.: Artificial intelligence based hybrid models for prediction of stock prices. In: 2023 2nd International Conference for Innovation in Technology (INOCON), pp. 1–6 (2023). https://doi.org/10.1109/INOCON57975.2023.10101297

[19] Ma, C., Yan, S.: Deep learning in the chinese stock market: The role of technical indicators. Finance Research Letters **49**, 103025 (2022) https://doi.org/10.1016/j.frl.2022.103025

[20] Breuss, M.: Beautiful soup: Build a web scraper with python. Preuzeto **30**, 2021 (2021)

[21] Tom Aarsen, J.N., Team: NLTK Documentation. https://www.nltk.org/ Accessed (accessed: 28.06.2023)