International Journal of Computers and Informatics

Journal Homepage: https://www.ijci.zu.edu.eg

Int. j. Comp. Info. Vol. 7 (2025) 69–78

Paper Type: Original Article

Predictive Analytics for Product Prices

Salma Elsayed [1,*], Ahmed Salah [1] and Marwa Abdellah [1]

[1] Faculty of Computer and Informatics, Zagazig University, Zagazig 44519, Egypt.
Emails: salmaelsaied492@gmail.com; ahmad@zu.edu.eg; marwaabdella2@gmail.com.

Received: 30 Jan 2025 **Revised**: 26 Mar 2025 **Accepted**: 18 Jun 2025 **Published**: 23 Jun 2025

Abstract

Product prices are affected by company stock prices. If the stock price rises, so will the product price. Otherwise,when the stock value falls, so does the product price. If a company's stock falls owing to financial difficulties, it may reduce product prices to increase revenues or liquidity. The aim is to estimate whether the product price will rise or fall according to the stock market price. Stock price prediction has risen in prominence due to its function in estimating the future value of company shares. There are various ways to predicting stock prices, including machine learning, deep learning, and ensemble learning. To estimate stock values, the information was collected for a variety of popular companies, such as Amazon. The datasets used are divided into two parts: the first comprises a set of tweets for the stocks under consideration in this study, acquired from the X social media site, and the second includes numerical stock price values. Sentimental characteristics from tweets were retrieved in two various manners to produce polarity. Vader was used to calculate the sentimental score and to generate the percentage of positive, negative, and neutral. Numerical data is also used to generate the sentimental score. All the columns of the two files were merged to obtain one dataset. Then, the problem was framed as a regression task. The evaluation difference between the proposed models was investigated for forecasting stock values according to tweets. In this regard, many ensemble learning models were developed to forecast the price changes of each stock. Furthermore, many machine learning and deep learning models were employed for evaluation. Several assessment indicators were used to assess the effectiveness of the presented techniques. The findings showed that the stacking regressor technique outscored the other approaches, achieving the lowest MAPE in the majority of the datasets.

Keywords: Ensemble Learning; Deep Learning; Machine Learning; Stock Price; Prediction Regression; CatBoost Classifier; CatBoost Regressor; Bagging Regressor; Stacking Regressor; Gradient Boosting Regressor.

1 | Introduction

Several research findings propose that company stock prices may act as early warning signs for product pricing Strategies. Rising costs repel clients or lead to negative news, therefore the brand could incur long-term damage, lowering stock price. A rise in costs may enable competitors to beat out others and gain market share. In contrast, lowering prices to increase market share might erode profit margins, which may concern shareholders. For example, Tesla has regularly altered its EV pricing. Similarly, Apple's stock rise in late 2022 was accompanied by continued higher prices despite globally discounting pressures, but Netflix's stock rebound in mid-2023 following earlier falls was followed by subscription price hikes in most regions by October. Rising costs repel clients or lead to negative news, therefore the brand could incur long-term damage, lowering stock price. A rise in costs may enable competitors to beat out others and gain market share. Every time: Reduced prices frequently resulted in stock declines owing to concerns about decreased demand or loss of margins. Price increases indicated high demand or growing costs, which occasionally boosted the stock, according to the market environment.

Corresponding Author: salmaelsaied492@gmail.com

Market participants are constantly on the lookout for a competitive advantage in order to profit from market moves before they become widely known. To profit from the running stock markets, investors constantly watch for trends and try to make accurate forecasts. Stock market projections are critical in this scenario. The stock market prediction [1] is useful for analyzing current stock market trends. This enables traders and investors to time their deals and valuation models. It allows them to obtain vital insights about the economy, securities, and stock market. Enhanced predictions enable more precise estimates of possible damages and market risks. Techniques can help find optimum property pairings and automatically rebalance investments based on market trends. Organizations may employ predictive techniques to identify gaps in the market quicker than rivals, providing them a strategic advantage. Early alert mechanisms have helped minimize losses and fluctuations. Although predictive marketing models have enormous possibilities in real-world settings, their effective application in computerized platforms requires considerably more than model precision. It needs a strong infrastructure, a mindful of danger approach, and ongoing evaluation of results. Those who successfully integrate these components can acquire a major competitive advantage in today's financial markets [2, 3].

Based on the market, price prediction can precisely forecast the most competitive price that an organization should select. Predicting prices has become more and more crucial in the power of a company to stay competitive. It is the process of forecasting a good, product, or service's price by examining a variety of variables, including requests, seasonal patterns, the prices of similar goods, offers from many suppliers, and many more.

This reseach applies five ensemble learning techniques [4] which are CatBoostClassifier, CatBoostRegressor BaggingRegressor, GradientBoostingRegressor, and Stacking Regressor. The study also used many machine learning methods. Furthermore, it applied a deep learning model, namely CNN-LSTM model[6]. All models was tested using many measurements[5]. The research of the suggested techniques demonstrated that the ensemble learning techniques[7], e.g., Stacking Regressor [8], CatBoostClassifer, and BaggingRegressor [9] operate superior for the majority of the various stocks e.g. Stacking Regressor achieves the lowest MAPE in 10 stocks and the lowset MAE in 6 stocks and in the Whole Stock, TSLA, CatBoostRegressor achieves the lowest MAPE, MSE, RMSE. CatBoostRegressor outperform the other models in META, AMZN, and VZ as it achieves the lowest MSE and in NFLX the lowest RMSE.

Reliable simulations enable superior projections of valuations and market movements, assisting shareholders to identify assets that are underrated or overpriced and the time access and departures are more effective and enables to select categories of investments, fields, or regions with more certainty. Financial institutions and mutual funds achieve an advantage over others because their predictions exceed those of competitors and can capitalize on flaws or arbitrage possibilities more quickly than others and can react faster to sudden events.

The stock market plays an important role in the rapid economic growth of developing countries like India. It is critical in today's financial and social situation. The stock market's ups and downs affect investors' profits. Stock price prediction has long piqued the attention of stakeholders and scholars because of its difficulty, inherent inconsistency, and constantly shifting nature. This is accomplished by taking into account both past stock values and price fluctuations over the preceding days. Because of fluctuations in markets, predicting stock price is extremely difficult, demanding a reliable predicting model. They applied the deep learning approaches that are currently employed to forecast changes in stock markets.

The Efficient Market Hypothesis states that stock prices completely represent every piece of data accessible at any given time. As a result, it is difficult to continually outperform the market average in terms of risk-adjusted earnings. In Weak Form Efficiency, prices represent all previous trade data (such as historical prices and quantities). Technical assessment is inefficient. Semi-Strong Form Efficiency, prices include all information that is freely available (such as financial reports and news). Fundamental assessment cannot produce ongoing excessive profits. Powerful Form Efficiency, prices represent all publicly available and private information. But exclusive knowledge cannot provide a shareholder an advantage.

2 | Data Collection

The dataset of various companies was collected from the Kaggle website (https://www.kaggle.com/datasets/equinxx/stock-tweets-for-sentiment-analysis-and-prediction?select=stock_tweets.csv), recently visited on March 23, 2024. The dataset has two components. The first half comprises 80,793 tweets with four attributes: date, tweet, stock name, and business name. From September 30, 2021 to September 29, 2022, it includes 22 firms such as TSLA, Amazon, META, and Microsoft. The collected tweets include extraneous data such as unique symbols, URLs, emoticons, hashtags (#), and @. Thus, we cleaned these tweets to extract only the basic phrases. The remaining portion of the collection comprises 6,300 market data points with eight features. The Arithmetic operations were applied to this numerical data in the feature extraction process. The collected tweets included superfluous data such as distinctive characters, URLs, emojis hashtags (#), and @. These tweets were edited to generate only intelligible contents. Preprocessing [10] often incorporates a variety of systematic procedures, which may be summarized as follows:

- Data Cleaning [11-17]: Determine missing values in columns/features using statistical analysis or visualizations. NaN (Not a Number) or NULL values are frequently employed to indicate missing values.
- Handling records: Duplicate records can distort analysis and model training by magnifying particular trends or assumptions. Use the duplicated() function in the pandas package like to find duplicate rows based on specified columns or the complete row. If duplicate entries are redundant and contain no extra data, you can delete them by calling the drop duplicates() function in pandas.
- Data Normalization: Normalization is a data preparation procedure that scales and standardizes the quantities of features in a dataset. Normalization's primary purpose is to put all feature values into a comparable range while preserving value ranging disparities. Normalization is done using the MinMaxScaler() function.
- Data Reduction: Approaches used to minimize data size without missing substantial data. It assists in the elimination of superfluous or unnecessary information, this enhancing the general performance of forecasting algorithms.
- Data Integration: The operation of merging data from several sources into a single dataset.

3 | Feature Extraction

Feature extraction is the technique of converting raw data into useful representations that machine learning algorithms can understand. It aims to discover the data's most relevant features while removing noise and unnecessary information. Two different approaches(as shown in Figure1) were employed to check the polarity score's correcteness. Two distinct methods were used to confirm the accuracy of the polarity value. The VADER technique [18] was used to assess the polarity. It exists in the NLTK package [19] and may be applied efficiently with pure textual information. The emotional results ranges from -1 to 1. In the second technique, the emotional score was calculated by calculating the greatest result of positive, negative, and neutral percentages. To determine the polarity polarity score using financial data, the percentage change of the closing price was computed , with the identical conditions as in Eq. (1).

$$Polarity = \begin{cases} Positive & if\ x \geq 0.5 \\ Neutral & if\ 0.5 > x \geq -0.5 \\ Negative & Otherwise \end{cases} \quad (1)$$

,where x is the sentimental score value.

All the fields from the two datasets (tweets part and financial data) were combined according to the date and stock name to generate a single dataset.

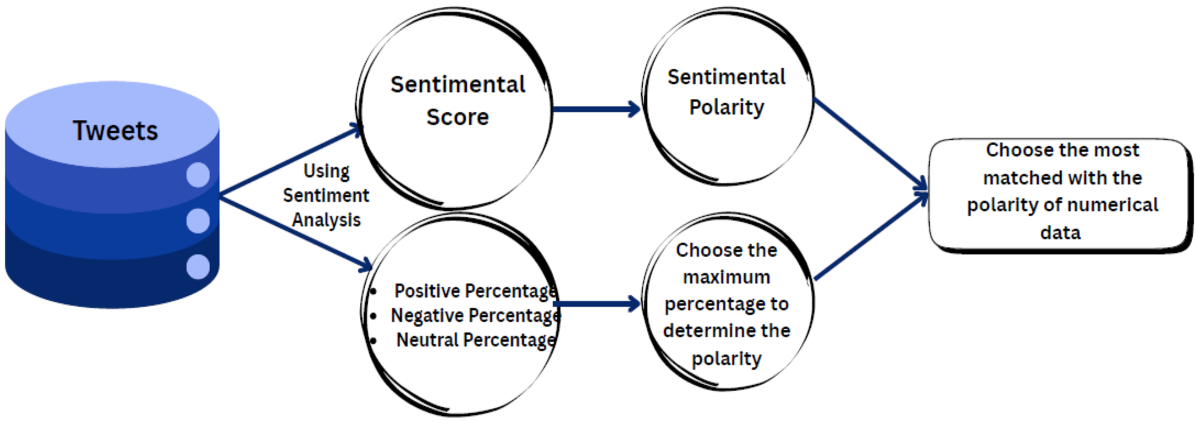

Figure 1. Two ways for Extracting Features.

4 | Proposed Models

4.1 | Ensemble Learning Models

Ensemble learning integrates several learners (e.g., neural networks or regression models) to enhance predictions. Ensemble Learning [20] is a machine learning technique that trains multiple learners to comprehend a single problem. Ensemble learning is based on the promise that a group of learners may get more precise results than a single learner. The most often employed ensemble learning strategies[25] are bagging, boosting, and stacking. Stacking is a sophistacted ensemble learning approach that layers individual model predictions and uses them as feed to train the meta-model. The training dataset is split into n sections. The basic model is trained for each n – 1 component. Stacking [21] involves training several base learners using different methodologies on the same dataset as shown in Figure 2. Training both basic and meta-learner models on the same dataset can lead to overfitting. To use test set data as a meta-learner training data, it may be necessary to remove instances from the base learner's training data.

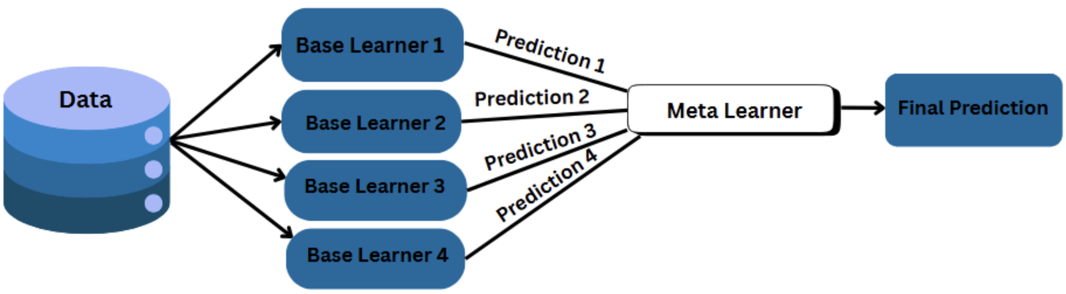

Figure 2. Stacking approach.

Blending [22] is a strategy that, like stacking, makes forecasts based on validation set from the training set. The training set is sepaerated into two sets: training and validation.

Bagging[23] (as shown in Figure 3) is a method for combining the results of various techniques to produce a more generic result. Individual approahes, however, do not receive the identical data set. Whereas, bootstrapping is employed to create alternative portions of the initial dataset.

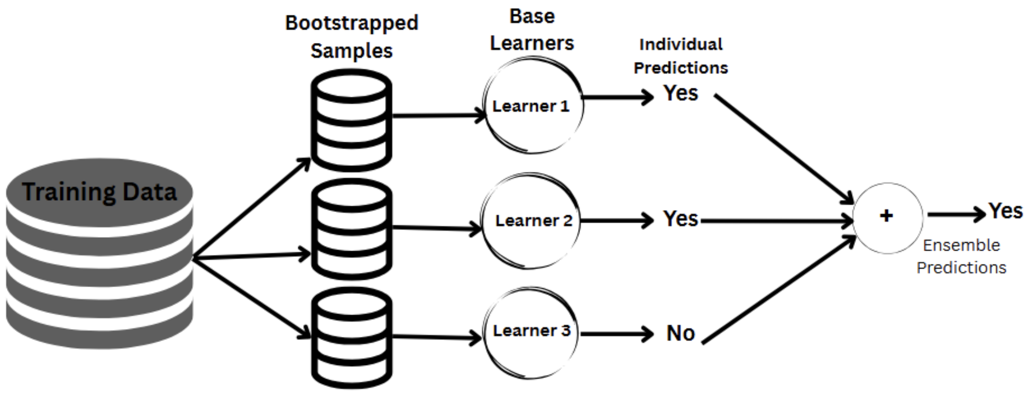

Figure 3. Bagging approach.

Boosting is a technique for training a learner with an initial dataset. Un like bagging, boosting, prioritizes incorrectly classified input from the original model or learner. In order to make final predictions, boosting aggreagate and weights all of the learners. The pool can accelerate the training cycle. The training dataset was then put into the model to fit it. The real values were then passed to the assessment mechanism. The Bagging Regressor and Gradient Boosting Regressor [24] classes were applied with default parameter values to fit the model. The StackingRegressor was applied with the estimators' parameters, which are DecisionTreeRegressor, LinearRegression, and LinearSVR. Then the model was fitted with X train and Y train parameters.

4.2 | Machine Learning Models

The K-nearest neighbor approach [26] is a machine learning algorithm that is deemed easy to apply. The previous stock data and test data are transformed into a set of variables. Each vector represents N dimensions for each stock characteristic. The determination is then made based on a similarity measure. KNN is regarded as a lazy learning that doesn't create a model or function in advance, but returns the closest k values of the training data set with the greatest degree of similarity to the test. The classification label is then determined by the overall decision of the selected k records and applied to the query record.

Support Vector Regressor [27] is a machine learning technique used in regression analysis. It is an expansion of the widely used Support Vector Machine technique for classification issues. SVR attempts to fit a line or curve around the data points in such a way that the difference between predicted and real values is reduced. The method accomplishes this by detecting the support vectors, or data points closest to the line or curve being fitted, and uses them to enhance the margin.

Linear regression [28] is a statistical technique for modeling the association between a dependent parameter (target) and one or more independent variables (features). The aim is to select the line that most closely fits the data points. Linear regression fits a line to the data points with the goal of minimizing the total of the squared differences (errors) between the actual and forecasted data points. The regression line may then be utilized to create predictions. Linear regression is useful for stock market forecasting because it can detect linear movements and trends in historical data. While it fails to manage for all of the complexity of the stock market, it serves as an adequate basis for developing more advanced models.

4.3 | Deep Learning Models

They presented a hybrid architecture that combines long short-term memory with a convolutional neural network, named CNN-LSTM. The proposed approach combines using of a convolution layer properties for obtaining significant characteristics, as well as the LSTM [31] the architecture's capacity to accumulate patterns over time. The model's first layer is a 1D convolutional layer that scans segments and uses numerous criteria to gather characteristics or predictions of the input series. Then, a ReLU activate function is utilized to boost the technique's potential to identify complicated frameworks. Following a convolutional layer, a pooling layer is typically used to reduce the consistency of the final attribute pattern. The suggested LSTM model consists of three layers, each with 40 units. A dropout layer is a normalization methodology that enables networks to be trained with several designs at once by arbitrarily eliminating part of layer's results throughout training; it is regarded as an effective method for eliminating overfitting. The final level is an entirely compicated stratum that serves as the model's outcome and comprises one neuron. The framework fits the data employing the Adam optimizer. The Adam optimizer is an adaptive optimization strategy that has been shown to be useful in addressing actual deep learning difficulties. Next step consists of approach training and evaluation. The technique forecast can currently be carried out. In the final step, the forecasted results would be inverted from the distinct data to recover the actual outcomes.

5 | Evaluation Matrics

Five assessment metrices [32] are utilized to verify the reliability and precision of the forecasting algorithms, namely, mean absolute error (MAE) [33], mean squared error (MSE), root mean squared error(RMSE), mean absolute percentage error(MAPE), and R squared(R^2).

- MAE estimates the average difference between the actual and anticipated values. As a consequence, we can estimate how well the forecasts match the actual data.

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y - \bar{y}| \quad (1)$$

- MSE computes the average of the squares of the errors. It is applied to assess the accuracy of regression issies.

$$\mathrm{MSE} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \left(\mathrm{y} - \bar{y}\right)^2 \quad (2)$$

- RMSE [34] calculates the square root of the average difference between the original and predicted values.

$$RMSE = \sqrt{\frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \left(\mathrm{y} - \bar{y}\right)^2} \quad (3)$$

- MAPE [35] measures forecasting technique accuracy. It calculates how close the forecasted outcome was to the real outcome by averaged the absolute percentage errors of all items in the data.

$$\mathrm{MAPE} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \left(\mathrm{y} - \bar{y}\right)^2 \quad (4)$$

- R^2 is denoted in Eq. (5). It compares the residual sum of squares (SSres), to the total sum of squares(SStot) .

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (5)$$

$$SS_{res} = \mathrm{sum}\left(\mathrm{y} - \bar{y}\right)^2 \quad (6)$$

$$SS_{tot} = \sum_{i=1}^{N}\left(\mathrm{y} - \bar{y}\right)^2 \quad (7)$$

6 | Results and Discussion

All models were tested using many measurements, including MAPE, MAE, MSE, RMS, and R^2. Evaluation of the suggested ensemble learning models revealed that the Stacking Regressor achieved the least MAPE for most of the datasets as shown in Table 1. The META, AMD, TSM, and DIS datasets achieved the lowest MAPE in the Stacking Regressor technique. META, AMD, TSM, and DIS accomplish the lowest MSE in Stacking Regressor algorithm.

Table 1. MAPE and MSE metrices for different models.

Stock Name	Matric	Stacking Regressor	CatBoostClassifier	CatBoostRegressor	BaggingRegressor	GradientBoosting
META	MAPE MSE	0.158 2948.285	0.263 11786.673	0.28 9424.84	0.26 10066.16	0.265 9982.078
AMD	MAPE MSE	0.147 766.059	0.230 1630.742	0.180 1046.482	0.20987 1347.666	0.199 1226.36
AAPL	MAPE MSE	0.137 453.43	0.097 290.256	0.07 158.645	0.087 232.723	0.085 211.66
TSM	MAPE MSE	0.124 236.18	0.168 483.26	0.135 257.17	:0.15433 388.726	0.146 327.23
DIS	MAPE MSE	0.06 193.61	0.25 2482.86	0.21 1435.33	0.21 1578.35	0.21 1638.55

Table 2. MAPE metric for different models.

Stock Name	Stacking Regressor	CatBoostRegressor	BaggingRegressor	GradientBoosting	LSTM
TSLA	0.169	0.155	0.183	0.170	0.222
MSFT	0.053	0.105	0.132	0.115	0.227
AMZN	0.187	0.184	0.174	0.178	0.293
GOOG	0.042	0.119	0.135	0.137	0.208
AMD	0.147	0.180	0.209	0.199	0.265
NFLX	0.213	0.439	0.459	0.492	2.814

All models' efficiency was tested using a variety of measures. The Stacking Regressor produced the lowest MAPE for the majority of the datasets, as shown in Table 2. The analysis of the suggested models [6] demonstrated that the Stacking Regressor, CatBoostClassifer, and BaggingRegressor, work more effectively for the majority of the various stocks. The TSLA dataset with 251 rows achieved the lowest MAPE in the CatBoostRegressor technique as it achieves 0.155, lowest MAE in the Stacking Regressor, the lowest R^2 in the CatBoostClassifier. TSLA is a very volatile stock and notorious for its extreme volatility and continual price fluctuations. This indicates that its pattern is not clear. Models such as CatBoost Regressor and Gradient Boosting Regressor have low R^2 values (negative), showing difficulty with volatility. Stacking Regressor and LSTM models works reasonably well. Deep learning (LSTM) [6] may be more suited to capturing nonlinear patterns and sequences in high-volatility environments. Stacking is the best model as it achieves 49.13 for the MAE metric.

7 | Limitations

Despite its thorough approach, the study faces multiple important limitations that should be taken into consideration. The primary limitation is the intrinsic volatility and unpredictable nature of stock markets, which are impacted by a wide range of external variables beyond social media sentiment and historical price data, such as worldwide economic conditions political developments, and company-specific news that may not be included in the dataset. The negative R^2 values [36] for all models point to a primary challenge in reliably predicting stock behavior, suggesting that the current strategy does not adequately account for complex relationships influencing changes in stock prices. Furthermore, the study mainly uses data from the X platform, which could not fully capture market sentiment because it leaves out other important social media sites and conventional news sources. The time duration of the data collection may also present limitations, as market circumstances and social media activity patterns might fluctuate greatly over time. Additionally, the study's concentration on 22 particular stocks, although significant, may restrict the findings'applicability to other stocks or market groups and may not accurately represent the dynamics of the larger market. The feature extraction techniques and preprocessing stages, although comprehensive, may unintentionally exclude potentially useful information, and sentiment analysis's binary character may oversimplify the complex nature of market sentiment. The computing resources needed to handle huge datasets and train several ensemble models are examples of technical constraints that might impact the practical implementation of the technique in real-time trading settings.

8 | Conclusion

Stock price movement affects the product price. A very high stock price may indicate great company trust, allowing the corporation to boost pricing. So the goal is to forecast whether the product price would rise or fall in response to the stock market price. Profits, margins, demands, and fluctuations in markets all influence the association between product and stock prices. Traders consider whether price movements will increase long-term value or reveal fundamental problems. Various ensemble learning techniques were applied for the predicting task. It was suggested to combine tweets and stock closing price changes into a single dataset to anticipate stock price changes according to tweets. First, the received tweets were preprocessed to remove unstructured data. The polarity of the tweets was then retrieved using two separate ways to confirm it was proper. The Vader was used to get the polarity result. The tweets whose polarity followed the overall pattern of the numerical information were chosen, while the others were eliminated. The dataset's size was decreased. The suggested ensemble learning techniques were compared to some machine learning and deep learning approahees. Five alternative assessment measures were used. The research findings showed that the suggested ensemble learning techniques outperformed latest models and ML models on average for most stocks. In addition, the assessment measures revealed that the Stacking Regressor model beat all techniques. StackingRegressor routinely outperforms or is equivalent to most stock markets (lower MAPE, MAE, MSE, and higher R2, particularly for TSLA, MSFT, and GOOG). While CNN-LSTM being more complicated and computationally costly, often performs worse than standard models. For example, MAPE for NFLX equal 2.814, MAE: 468.603, and R2: -728.341. This indicates that it may be overfitting or not expanding adequately to specific datasets. Bar charts in Figures 2 show that StackingRegressor frequently has the lowest MSE and RMSE, implying greater accuracy. CatBoostClassifier, while complicated, has extremely high error values, particularly for NFLX, suggesting low accuracy when applied to regression tasks. Complicated models do not guarantee greater results. In this situation, StackingRegressor beats more computationally intensive models such as CNN-LSTM. Model selection should be determined by data characteristics; utilizing a regression classifier (such as CatBoostClassifier) might have a negative impact on performance. Bagging, GradientBoosting, and StackingRegressor strike a balance between accuracy and computational resources. Based on the findings, some prospective avenues for further research are proposed including extending into Additional Financial Products such as Global products which Include gold, oil, and crops, which are frequently impacted by economic trends, forecasting exchange couple fluctuations (for example, EUR/USD) based on sentiment and macroeconomic factors, using this strategy to un certain investments such as the digital currency Bitcoin, and alternative currencies which have distinct fluctuations in the market, and merging Computational Intelligence(CI) with real-time data streams (such as online communities).

Funding

This research has no funding source.

Data Availability

The data of various companies was collected from the Kaggle website (https://www.kaggle.com/datasets/equinxx/stock-tweets-for-sentiment-analysis-and-prediction?select=stock_tweets.csv), last accessed on 23-March-2024).

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] Balasubramanian P, P. C, Badarudeen S, Sriraman H. 2024. "A systematic literature survey on recent trends in stock market prediction."PeerJ Computer Science, 10:e1700, doi:10.7717/peerj-cs.1700.

[2] Chopra, Ritika and Sharma, and Gagan, "Application of artificial intelligence in stock market forecasting: a critique, review, and research agenda," International Journal of Scientific Research in Science and Technology, 2021, vol.7, no.4, doi: 10.3390/jrfm14110526

[3] Padmanayana, Varsha and Bhavya, K, "Stock market prediction using twitter sentiment analysis," International Journal of Scientific Research in Science and Technology, 2021, vol. 7, Issue. 4, pp. 265-270, doi: 10.32628/CSEIT217475

[4] MIENYE, Ibomoiye Domor; SUN, Yanxia. "A survey of ensemble learning: Concepts, algorithms, " applications, and prospects. Ieee Access,,2022, 10: 99129-99149, doi:10.1109/ACCESS.2022.3207287.

[5] arunasingha, Dulakshi Santhusitha Kumari. "Root mean square error or mean absolute error? Use their ratio as well." Information Sciences. 585 (2022): 609-629. doi:10.1016/j.ins.2021.11.036

[6] Walid Abdullah, Ahmad Salah, "A novel hybrid deep learning model for price prediction," International Journal of Electrical and Computer Engineering (IJECE), vol. 13, no.3, 2023.

[7] ADHY, Neelamadhab, et al. " Stock market prediction performance analysis by using machine learning regressor techniques." International Conference on Computing, Communication and Learning. Cham: Springer Nature Switzerland. (2023): p. 39-50. doi:10.1007/978-3-031-56998-2-4

[8] A. Chatzimparmpas, R. M. Martins, K. Kucher and A. Kerren, "StackGenVis: Alignment of Data, Algorithms, and Models for Stacking Ensemble Learning Using Performance Metrics," IEEE Transactions on Visualization and Computer Graphics , vol. 27, no. 2, pp. 1547-1557, Feb. 2021, doi: 10.1109/TVCG.2020.3030352.

[9] Krouska, C. Troussas and M. Virvou, "The effect of preprocessing techniques on Twitter sentiment analysis," International Conference on Information, Intelligence, Systems & Applications (IISA), 2016, pp. 1-5, doi:10.1109/IISA.2016.7785373.

[10] U, Shengting, et al. " S I LSTM: stock price prediction based on multiple data sources and sentiment analysis. ", Connection Science., 2022, 34.1: 44-62, doi: 10.1080/09540091-2021-1940101.

[11] A. Kumar, K. Yun, T. Gebregzabiher, B. Y. Tesfay and S. G. Adane, "COVID19 Tweeter Dataset Sentiment Analysis," 2021 Fourth International Conference on Computational Intelligence and Communication Technologies (CCICT), Sonepat, India, 2021, pp. 110-115, doi: 10.1109/CCICT53244.2021.00032.

[12] . Krouska, C. Troussas and M. Virvou, "The effect of preprocessing techniques on Twitter sentiment analysis," International Conference on Information, Intelligence, Systems & Applications (IISA), 2016, pp. 1-5, doi: 10.1109/IISA.2016.7785373.

[13] AHUJA, Ravinder, " The impact of features extraction on the sentiment analysis, " Procedia Computer Science , 2019, 152: 341-348, doi:10.1016/j.procs.2019.05.008

[14] LI, Yang; PAN, Yi. "A novel ensemble deep learning model for stock prediction based on stock prices and news. " International Journal of Data Science and Analytics, 2022, 13.2: 139-149, doi:0.1007/s41060-021-00279-9

[15] ZHENG, Haotian, et al. "Predicting financial enterprise stocks and economic data trends using machine learning time series analysis. " Applied and Computational Engineering, 2024, 87: 26-32., doi:10.54254/2755-

2721/87/20241562
[16] Billah, M.M., Sultana, A., Bhuiyan, F. et al. "Stock price prediction: comparison of different moving average techniques using deep learning model. " Neural Comput & Applic, 2024, 36, 5861–5871., doi:10.1007/s00521-023-09369-0
[17] CHICCO, Davide; WARRENS, Matthijs J.; JURMAN, Giuseppe, "The coefficient of determination R squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. " Peerj computer science, 2021, 7: e623, doi:10.7717/peerj-cs.623/supp-1.
[18] biola, Odeyinka and Abayomi-Alli, Adebayo and Tale, Oluwasefunmi Arogundade and Misra, Sanjay and Abayomi-Alli, Olusola, "Sentiment analysis of COVID-19 tweets from selected hashtags in Nigeria using VADER and Text Blob analyser," Journal of Electrical Systems and Information Technology, vol. 10, no. 1, pp. 1–20, 2023, doi: 10.1186/s43067-023-00070-9.
[19] Abdelfattah, B.A.; Darwish, S.M.; Elkaffas, S.M. "Enhancing the Prediction of Stock Market Movement Using Neutrosophic-Logic-Based Sentiment Analysis. " Journal of Theoretical and Applied Electronic Commerce Research., 2024, 19(1):116-134. doi:10.3390/jtaer19010007
[20] MAHAJAN, Palak, et al. "Ensemble learning for disease prediction: A review.", Healthcare. MDPI, 2023. p. 1808, doi: 10.3390/healthcare11121808.
[21] GU, Jiayue, et al. "A stacking ensemble learning model for monthly rainfall prediction in the Taihu Basin, China.", Water, 2022, 14.3: 492, doi: 10.3390/w14030492.
[22] HASAN, Mahmudul, et al. "A blending ensemble learning model for crude oil price forecasting.", Annals of Operations Research, 2024, 1-31, doi: 10.1007/s10479-023-05810-8.
[23] XU, S. B., et al. "Prediction of the Dst index with bagging ensemble-learning algorithm.", The Astrophysical Journal Supplement Series, 2020, 248.1: 14, doi: 10.3847/1538-4365/ab880e.
[24] YANG, Senyan, et al. "Ensemble learning for short-term traffic prediction based on gradient boosting machine.", Journal of Sensors, 2017, 2017.1: 7074143, doi: 10.1155/2017/7074143.
[25] SHAHHOSSEINI, Mohsen; HU, Guiping; PHAM, Hieu. "Optimizing ensemble weights and hyperparameters of machine learning models for regression problems.", Machine Learning with Applications , 2022, 7: 100251, doi: 10.1016/j.mlwa.2022.100251
[26] CAO, Hongduo, et al. "Stock price pattern prediction based on complex network and machine learning", Complexity, 2019, 2019.1: 4132485, doi:10.1155/2019/4132485.
[27] P. Meesad and R. I. Rasel, "Predicting stock market price using support vector regression," 2013 Interna- tional Conference on Informatics, Electronics and Vision (ICIEV), Dhaka, Bangladesh, 2013, pp. 1-6, doi: 10.1109/ICIEV.2013.6572570.
[28] SU, Xiaogang; YAN, Xin; TSAI, Chih-Ling. "Linear regression." Wiley Interdisciplinary Reviews: Computa- tional Statistics, 2012, 4.3: 275-294, doi:10.1002/wics.1198.
[29] GRAW, J. H.; WOOD, W. T.; PHRAMPUS, B. J, "Predicting global marine sediment density using the random forest regressor machine learning algorithm. ", Journal of Geophysical Research: Solid Earth, 2021, 126.1: e2020JB020135, doi: 10.1029/2020JB020135.
[30] GUPTA, Aakash, et al. "Solar energy prediction using decision tree regressor.", 5th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2021. p. 489-495, doi: 10.1109/ICICCS51141.2021.9432322.
[31] SAK, Hasim, et al. "Long short-term memory recurrent neural network architectures for large scale acoustic modeling. ", Interspeech. 2014. p. 338-342.
[32] Ali M, Khan DM, Alshanbari HM, El-Bagoury AA-AH. "Prediction of Complex Stock Market Data Using an Improved Hybrid EMD-LSTM Model. " Applied Sciences., 2023, 13(3):1429, doi:10.3390/app13031429
[33] Sibindi, R., Mwangi, R. W., & Waititu, A. G. "A boosting ensemble learning based hybrid light gradient boosting machine and extreme gradient boosting model for predicting house prices." Engineering Reports., 2023, 5(4), e12599, doi:10.1002/eng2.12599
[34] Ferrouhi, E. M., & Bouabdallaoui, I. "A comparative study of ensemble learning algorithms for high-frequency trading. " Scientific African., 2024, e02161, doi:10.1016/j.sciaf.2024.e02161
[35] Haryono, A. T., Sarno, R., & Sungkono, K. R. "Stock price forecasting in Indonesia stock exchange using deep learning: A comparative study. " International Journal of Electrical and Computer Engineering (IJECE), 2024, 14(1), 861-869, doi:10.11591/ijece.v14i1.pp861-869
[36] Majumder, A., Rahman, M.M., Biswas, A.A., Zulfiker, M.S., Basak, S. (2022). Stock Market Prediction: A Time Series Analysis," Smart Innovation, Systems and Technologies, vol 235. Springer, doi:10.1007/978-981-16-2877-1-35