Babylonian Journal of Artificial Intelligence
Vol. 2024, **pp**. 118–127
DOI: https://doi.org/10.58496/BJAI/2024/012 ISSN: 2958-6453
https://mesopotamian.press/journals/index.php/BJAI

Research Article

Enhancement of The Performance of Machine Learning Algorithms to Rival Deep Learning Algorithms in Predicting Stock Prices.

Rusul Mansoor Al-Amri [1,*], Ahmed Adnan Hadi [2], Mayameen S. Kadhim [3], Ayad Hameed Mousa [4], Ali Z.K Matloob [2], Hasanain Flayyih Hasan [5]

1 College of Nursing, University of Al-Ameed, Karbala, PO No: 198, Iraq .
2 Computer Techniques Engineering Department, College of Engineering and Technologies, Al-Mustaqbal University, Babil, Iraq .
3 Department of Medical Instruments Techniques Engineering, Technical College of Engineering, Al-Bayan University, Baghdad, Iraq.
4 College of computer science and Information technology University of Kerbala, Iraq.
5 College of Computer Science and Information Technology, University of Wasit, Wasit, Iraq .

ARTICLE INFO

Article History
Received 02 Jun 2024
Revised: 03 Aug 2024
Accepted 03 Sep 2024
Published 25 Sep 2024

Keywords
Gradient Boosting machine (GBM)
Natural Language Processing (NLP)
Stock Market
Sentiment Analysis

ABSTRACT

This paper objectives to improve stock market prediction accuracy by training data on sentiment analysis of tweets, overcoming volatility and complexity. Utilizing the Use of natural language processing (NLP) algorithms, the tweet's sentiments were classified into (negative - neutral - and positive). The stock value price was predicted by implementing Machine learning algorithms (KNN, SVM, GBM, LR, DT, RF, EL). Among the techniques of ML, (GBM) achieved the greatest results in terms of accuracy (96%). Its results were compared with the results of a deep learning algorithm that uses the same data where GBM got better results, and other algorithms showed results (KNN = 55%, SVM=90%, LR=82%, DT=90%, RF=90%, EL = 88%). The results obtained were superior to previous studies.

1. INTRODUCTION

Stock exchanges are crucial for promoting financial and capital gains in the economy. They involve economic transactions where shares are bought and sold, reflecting ownership claims[1-2]. These transactions shift money from small investors to large trading investors but are considered high-risk due to their erratic behavior [3]. Stock trading is a crucial area of interest for researchers and investors, as it affects the rise and fall of stock value prices. Accurate projections help investors obtain profits and avoid dangers [4]. Moreover, stock prices serve as a technical index for companies and contribute to economic development research [5]. Consequently, understanding the ingrained value and stock price prediction holds philosophical significance and wide application potential [6][7].

Stock price market forecasting has developed from statistical analysis and time-series modeling to machine learning techniques[8]. Machine learning algorithms, designed to learn from data, can predict patterns and relationships, revealing complex patterns that older approaches may not. This shift in forecasting has led to increased interest in machine learning [9-10].

In this research, a data set was chosen from Twitter, in which these tweets were classified using algorithms Natural Language Process (NLP) into (positive) (negative) (natural). After the process of predicting the type of tweet, seven machine learning algorithms (KNN, SVM, GBM, LR, DT, RF, EL) were used, and the data was trained and tested on them. Evaluative standards were applied to measure the accuracy of these algorithms in the process of predicting stock prices for these

**Corresponding author. Email: sosoalhaji28@gmail.com*

companies and measuring the quality of performance according to the metrics (MSE, RMSE, Recall, precision, F-score). The results obtained when applying the criteria to machine learning algorithms were excellent compared to the results of previous studies.

In this study, the performance of machine learning algorithms in predicting stock prices was improved, as the results were better and closer to the results obtained when applying deep learning algorithms, especially when these algorithms were applied to the same dataset.

The following is the organization of the research paper. In section 2, previous work on predicting stock prices was reviewed. In section 3, the method of work was demonstrated and the work of each algorithm in this research was explained. section 4, the results reached were presented, discussed, and compared with the results of the remaining research. In section 5, the sources were added that help behind this research paper.

2. RELATED WORK

This section discusses earlier research on algorithms that make use of deep learning and machine learning.

The aim of the research study for Measure the quality of stock price predictions made utilizing stock value by Gradient Boosting Machines(GBM) rather than the Naive Bayes Algorithm(NBA) The sample size for the (GBM) Technique is 20. and the (NBA) is iterated several intervals to estimate computing the accuracy of price for stocks. The (GBM) Algorithm's current Loss Function, which is based on the previous stock price, helps to decrease the total forecasting error. the GBM approach accurate (92.3%) [11].

The work in this study aims to Determine the accuracy of stock price forecasting based on stock price values. Used (gradient Boosting Machine) GBM and (Support Vector Machine) SVM methods with 20 sample sizes are to be repeated at different times to forecast stock price value accuracy. appear result GBM Algorithm has better accuracy (92.3 %) comparing to the SVM Algorithm accuracy (76%) The GBM algorithm, equipped with a novel loss function, enhances the accuracy of predicting the percentage of stock values. [12]

The study develops a stock price forecast model using Social analysis (SA) on StockTwits and data Twitter. The use of SA was applied to tweets, and seven machine-learning models were implemented(KNN, SVM, LR, NB, DT, RF, and MP). This work integrates multiple ML methods and SA, focusing on retrieving extra features of social media, to enhance stock prediction accuracy. The analysis of tweets using Valence Aware Dictionary, sEntiment Reasoner (VADER), and SVM yielded the best results, with an F-score of 76.3% and an AUC of 67%.[13]

This paper's main objective is to train the data to predict price movements in stocks based on the sentiment analysis of tweets, which will assist in tackling some of these challenges in the stock market. The tweet sentiments were divided into three categories utilizing the natural language processing (NLP) technology: positive, neutral, and negative. With the use of algorithms developed using deep learning, the stock price was forecast (RNNs, CNNs, BiLSTMs, LSTMs). Out of all the algorithms, (BiLSTM) produced the best accuracy results (94%). [14]

In this study, the researchers proposed a deep generative adversarial network to solve the problem of predicting stock prices. This algorithm was applied to data for a company in Italy ("the Financial Times Stock Exchange Milan N-Day di Borsa") and through the training process, a one-step prediction or a multi-step prediction is made. The results of this prediction using this algorithm showed superiority to the commonly used algorithms and conformity to industry standards. The results of this study showed that the accuracy reached 95%, indicating that using time series is very useful for financial forecasting. In this research, it was found that the proposed (GAN) has the shortest implementation periods [15].

The study proposes an inventory stock price forecasting model based on generative key Adversarial networks (GANs). The model comprises a generator and discriminator to use on the dataset. However, it doesn't provide complete Stock exchange development algorithms were used to analyze sentiment (such as positive, natural, or negative opinions) in this mentioned study, The model is trained using GAN and learns to generate new data samples. The model's effectiveness is tested on the dataset and results show good forecasting despite challenging macroeconomic climates, enhancing market stability [16].

3. METHODOLOGY

3.1 Dataset

The data used in this study are those used for the testing and training phases, which were both carried out on Twitter. The size of the dataset used is 80793 which was acquired from Twitter.

3.2 NLP Utilizing for Analyzing Data

The analysis of the tweets that are from the companies that engage with the price shares, where the information for 22 companies was obtained from Twitter and a total of 80,793 tweets were collected. In this study, the work was programmed in the Python environment. Four companies were selected from the substantial data set, including (AMZN, AAPL, AMD, and BA), where the sum of tweets for each of them was (11,771). The sentiments of these companies' tweets were analyzed using natural language processing algorithms to determine whether they were positive, negative, or natural sentiments. illustrated the following figure.

	Date	Tweet	Stock Name	Company Name	sentiment_score	Negative	Neutral	Positive
0	2022-09-29 22:40:47+00:00	A group of lawmakers led by Sen. Elizabeth War...	AMZN	Amazon.com, Inc.	-0.0772	0.084	0.841	0.075
1	2022-09-29 22:23:54+00:00	$NIO just because I'm down money doesn't mean ...	AMZN	Amazon.com, Inc.	0.25	0.158	0.684	0.158
2	2022-09-29 18:34:51+00:00	Today's drop in $SPX is a perfect example of w...	AMZN	Amazon.com, Inc.	-0.3182	0.164	0.728	0.108
3	2022-09-29 15:57:59+00:00	Druckenmiller owned $CVNA this year \nMunger b...	AMZN	Amazon.com, Inc.	0.2382	0.065	0.851	0.083
4	2022-09-29 15:10:30+00:00	Top 10 $QQQ Holdings \n\nAnd Credit Rating\n\n...	AMZN	Amazon.com, Inc.	0.7783	0.0	0.799	0.201

Fig .1. illustrate tweets classified by using NLP

3.3 Machine Learning Algorithm

Machine learning can be utilized to anticipate stock market trends and predict stock prices using various algorithms, including k-nearest neighbors (KNN), logistic regression (LR), decision trees (DT), random forests (RF), Support Vector Machines (SVMs), Gradient Boosting and Ensemble learning.

1- **The k-Nearest Neighbors (KNN**) algorithm A popular classification and regression technique in machine learning, the k-Nearest Neighbors (KNN) algorithm is utilized for both supervised and unsupervised applications. It is a non-parametric, lazy learning technique that makes predictions based on the similarity of data points. Despite being computationally expensive and sensitive to unimportant features, anomaly detection, recommendation systems, and picture and text categorization all benefit from its versatility [17-18].

2- **Logistic Regression (LR)** A popular approach for classification jobs is logistic regression. Based on a set of input data, it can be used to predict whether the value of a specific stock will rise or fall. By applying a logistic curve to the data, logistic regression may provide probabilistic forecasts about how an event will turn out [19-20].

3- **Decision Trees (DT)** are A supervised learning approach called decision trees can be used for both classification and regression applications. Based on the values of the input variables, they operate by splitting the data into smaller groupings. Decision trees may thus capture intricate correlations between the input and output variables, which makes them ideal for forecasting stock prices [21-22].

4- **Random Forests (RF)** an ensemble learning system mixes different decision trees to provide predictions that are more accurate. A huge number of decision trees are built using random subsets of the data by random forests, and their predictions are then averaged. By doing so, overfitting may be reduced and prediction accuracy may be improved [23-24].

5- **Support Vector Machines (SVMs)** are flexible supervised machine learning algorithms that are used for classification and regression tasks, especially for challenges involving the classification of input points into two or more classes. Support Vector Machines (SVMs) can be used to predict stock prices, but it's important to understand their constraints and methodology because they might not always produce reliable outcomes on their own [25-26].

6- **Gradient Boosting** A potent ensemble machine-learning method called gradient boosting is employed for both classification and regression tasks. Each tree in the ensemble corrects the mistakes caused by the previous ones as it is constructed progressively. Due to its strong prediction accuracy and resistance against overfitting, gradient boosting has gained popularity. For predictive tasks like stock price prediction, regression, and categorical label categorization, it is a flexible machine-learning technique [27-28].

7- **Ensemble learning** is Machine learning techniques like ensemble learning, which are especially useful for complicated, noisy, or high-dimensional data, improve performance by aggregating predictions from numerous models. Despite its complexity and susceptibility to many conditions, ensemble learning improves stock price prediction by integrating the predictive capacity of various models [29-30].

3.4 Training Processing Using Machine Learning Technique

1- **Data Gathering**: - For predicting, collect historical stock price information on opening, closing, highest and lowest prices, trading volume, and Stack Name.

	Date	Open	High	Low	Close	Adj Close	Volume	Stock Name
0	2021-09-30	165.800003	166.392502	163.699493	164.251999	164.251999	56848000	AMZN
1	2021-10-01	164.450500	165.458496	162.796997	164.162994	164.162994	56712000	AMZN
2	2021-10-04	163.969498	163.999496	158.812500	159.488998	159.488998	90462000	AMZN
3	2021-10-05	160.225006	163.036499	160.123001	161.050003	161.050003	65384000	AMZN
4	2021-10-06	160.676498	163.216995	159.931000	163.100494	163.100494	50660000	AMZN
...	...	...	...	...	...	...	...	...
1003	2022-09-23	135.649994	136.190002	129.500000	131.259995	131.259995	8927900	BA
1004	2022-09-26	129.770004	132.449997	126.879997	127.339996	127.339996	7095100	BA
1005	2022-09-27	129.320007	130.449997	125.599998	127.510002	127.510002	6682300	BA
1006	2022-09-28	128.039993	133.889999	127.400002	133.440002	133.440002	10257200	BA
1007	2022-09-29	131.199997	131.610001	123.800003	125.330002	125.330002	8905400	BA

Sample Add sentiment score

	Date	Open	High	Low	Close	Adj Close	Volume	Stock Name	sentiment_score
0	2021-09-30	165.800003	166.392502	163.699493	164.251999	164.251999	56848000	AMZN	0.246194
1	2021-10-01	164.450500	165.458496	162.796997	164.162994	164.162994	56712000	AMZN	0.331286
2	2021-10-04	163.969498	163.999496	158.812500	159.488998	159.488998	90462000	AMZN	0.215558
3	2021-10-05	160.225006	163.036499	160.123001	161.050003	161.050003	65384000	AMZN	0.146800
4	2021-10-06	160.676498	163.216995	159.931000	163.100494	163.100494	50660000	AMZN	0.193359

Fig. 2. Explain Data Gathering for prediction

2- **Preprocessing Data**: - Using machine learning models, data preprocessing entails cleaning, addressing values that are missing, eliminating outliers, normalizing, scaling, and dividing data into training and testing sets.

3- **Engineering Features**: - Engineering new features using domain knowledge or assumptions entails discovering relevant elements influencing stock prices, such as technical signals, fundamentals, and external variables.

4- **Data splitting**: - For the purpose of training and performance evaluating machine learning models, split preprocessed data into training (70%) and testing sets (30%).

5- **Selecting a Model**: - Choose for prediction Logistic Regression (LR), k-Nearest Neighbors (KNN), random forests, and Decision Trees (DT). the machine learning model for forecasting stock prices depending on the size, complexity, and desired accuracy of the dataset.

6- **Training**: Use the training dataset to train the chosen machine learning model. In order to learn the underlying patterns and linkages among the input features and the target stock prices, it is necessary to feed the model's previous stock price data and related features.

7- **Hyperparameter Tuning:** To enhance the performance of the machine learning model, optimize the hyperparameters.

8- **Evaluating the Model :** To evaluate the accuracy and dependability of stock price predictions, evaluate the trained model using the testing dataset and performance measures such accuracy, F1 score, MAE, RMSE, and MPE.

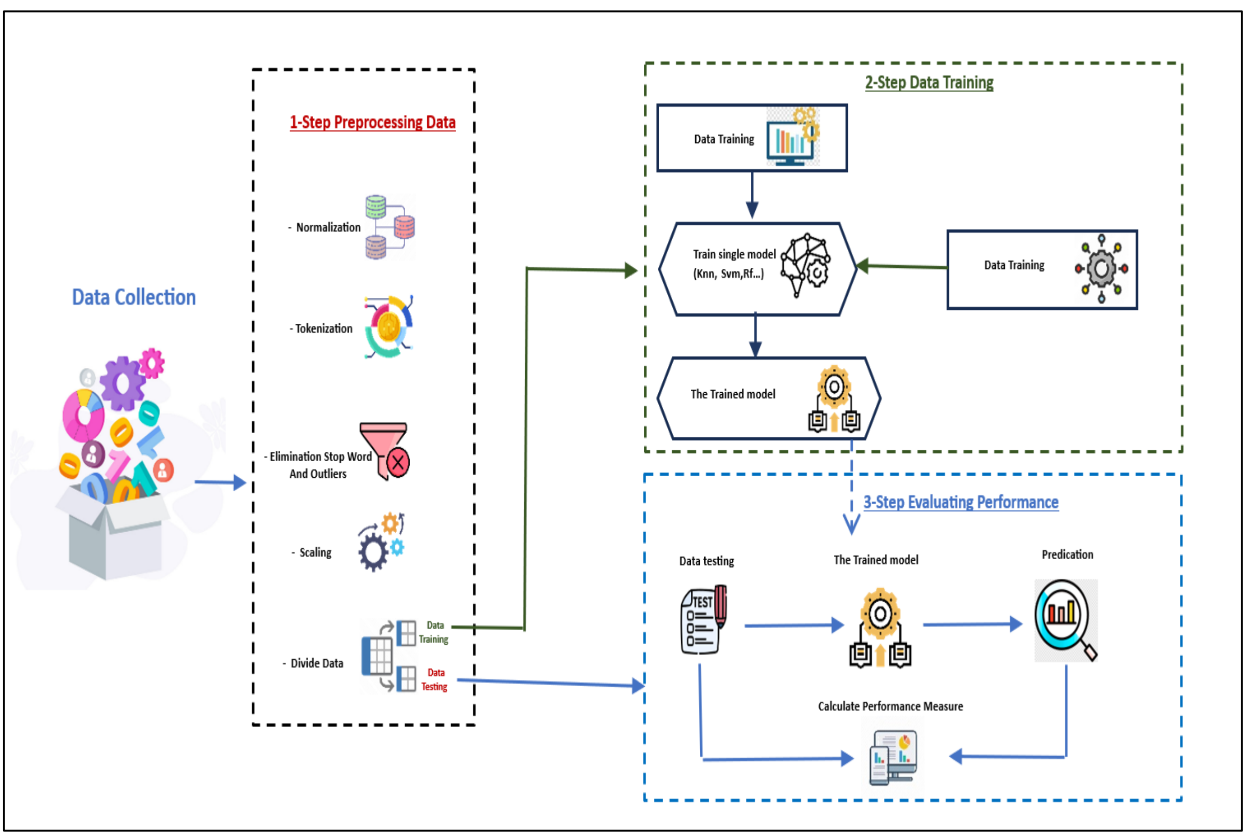

Fig. 3. Training Processing and Testing Processing Using Machine Learning Technique

4. RESULTS AND DISCUSSION

after the procedure of training and testing the machine learning techniques suggested in this study on data, each algorithm's accuracy was established and measurement criteria were utilized for evaluating each algorithm's level of performance, where the criteria that were applied (MSE, RMSE, F-Score, Recall score).

Note that the level of accuracy of the computation is calculated by Eq (1)[31-32].

$$\text{Accuracy } = \frac{NCP}{TNP} \ldots\ldots (Eq\ 1)$$, where NCP=Number Correct Predication, TNP=Total Number Predication.

Performance was assessed using a standard approach that calculates the mean squared error and Root mean squared error variance and Lower MSE and RMSE values indicate a better fit between the real and expected values [33-34-35]

$$MSE = \frac{1}{n}\sum_i^n\left(Z_i - \hat{Z}_i\right)^2 \ldots.\ (Eq\ 2) \quad , \ RMSE = \sqrt{MSE} \ \ldots.(Eq\ 3)$$, Where n= Number of data ,

$$Z_i = represented\ real\ target\ value\ , \ \hat{Z}_i = \ represented\ prediction\ \ target\ value$$

Recall was measurement assessed the model's ability to recognize each and every positive instance among all of the real positive examples. It is often referred to as sensitivity or true positive rate. As shown the (Eq 4) [36-37].

$$Recall = \frac{TP}{TP + FN} \ldots.(Eq\ \ 4) \quad ,Tp = True\ Positives\ ,FN = \ \ False\ Negatives$$

Precision was measurement evaluates how well the model predicts positive outcomes. Out of all the projected positive instances, it shows the percentage of accurately predicted positive instances. As shown the (Eq 5)[38-39-40].

$$Precision\ = \frac{TP}{TP + FP} \ldots.(Eq\ \ 5) \quad ,Tp = True\ Positives\ ,FN = \ \ False\ Positives$$

F-Score was a measure employed to assess how well a categorization model performs. In order to produce one rating which balances both of these metrics, it takes into account both recall and precision. As shown the (Eq 6)[41-42-43].

$$F - Score = 2 \times \frac{Recall\ \times precision}{Recall + precision} \ldots.(Eq\ 6)$$

In the table a below shown accuracy score for each algorithm's output, and metrics (MSE ,RMSE, Recall, Precision , F-score) for each machine learning algorithms .

TABLE I. EXPLAIN THE METRICS AND ACCURACY FOR EACH MACHINE LEARNING ALGORITHM.

Metrics	KNN	GBC	LR	DT	RF	SVM	EL
Accuracy	55%	96%	82%	90%	93%	90%	88%
MSE	0.91	0.09	0.44	0.16	0.11	0.16	0.51
RMSE	0.95	0.29	0.66	0.39	0.33	0.39	0.71
Recall	0.56	0.95	0.82	0.90	0.93	0.90	0.88
precision	0.55	0.96	0.82	0.90	0.93	0.90	0.88
F-score	0.54	0.96	0.82	0.90	0.93	0.90	0.88

The table above found the best and highest accuracy score for GBC (gradient boosting classifier) where accuracy = 96 %, MSE=0.09 and RMSE =0.29, and following RF (Random Forests) have second best accuracy =93%, MSE=0.11, and RMSE=0.33 and KNN (The k-Nearest Neighbors) achieved lower of accuracy =55% form the machine learning techniques.

The figure (4) shows the accuracy ,MSE and RMSE of each machine learning algorithm and the difference between the algorithms that were used in the proposed research.

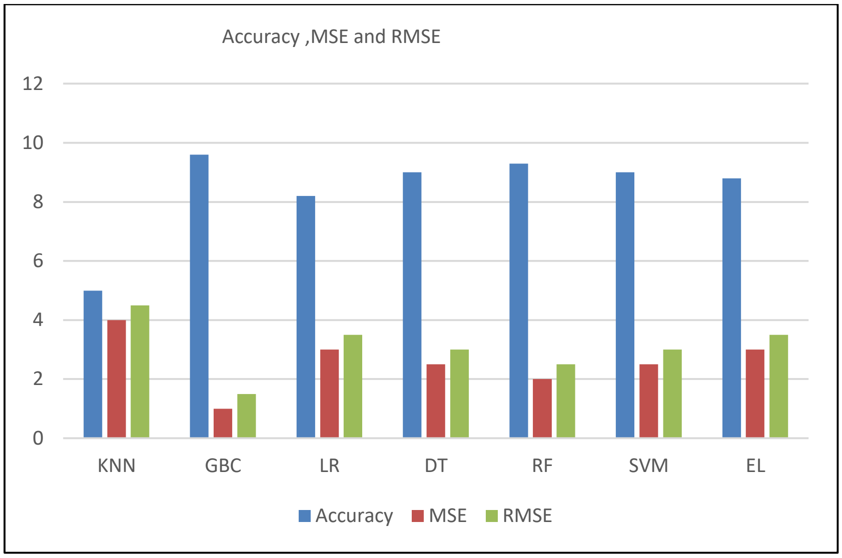

Fig .4 Describe the accuracy and criteria evaluated

From the chart above, find that the best performance algorithm is (GBM) from the ML algorithms, as it obtained the highest percentage of accuracy (96%) and the best values for the evaluation criteria in its application to predict stock prices, followed by the (RF) algorithm as the second-best value in terms of accuracy and performance (93%), and the lowest values for the evaluation criteria and accuracy was the (KNN) algorithm. Where its value was (55%) .

In the table (2), the results accuracy obtained in this research regarding the performance of the machine learning algorithms are compared to the results related to the machine learning algorithms of previous research.

TABLE II. COMPARED ACCURACY PROPOSED METHODS WITH PREVIOUS STUDY.

Methods	Proposed method	Relatedwork [11]	Relatedwork [12]	Relatedwork [13]
KNN	55%	---	-----	65.2%
GBC	96%	92%	92.3%	
LR	82%	---	-----	70.1%
DT	90%	---	-----	58%
RF	93%	---	-----	61.5%
SVM	90%	----	76%	76.3%
EL	88%	-----	-----	

From the above comparison, it was found that the results obtained are better than the results of previous studies.

It is possible to compare the best results obtained for the algorithms in this study with some previous studies on deep learning algorithms, especially those that were applied to predict price quotes on the same dataset.

Figure (5) shows the process of comparing the best results accuracy obtained in the proposed research with the results accuracy of deep learning algorithms .

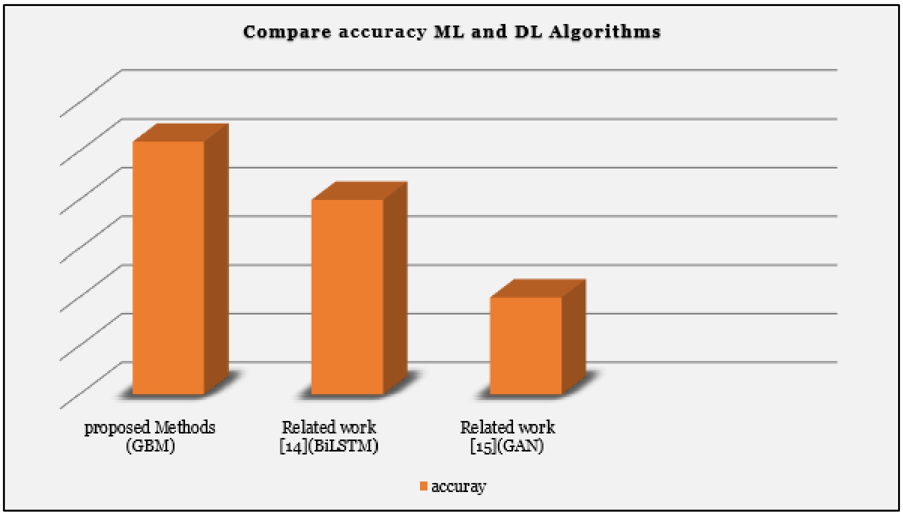

Fig. 5 Compare Between Accuracy ML And DL Algorithms

After the above comparisons, we show that the proposed work achieved the best results from previous studies, whether at the level of machine learning algorithms or deep learning algorithms that were applied to predict stock prices.

5. CONCLUSION

In this study, the performance of machine learning algorithms for predicting name prices for several companies was developed and improved (AMZN, AAPL, AMD, and BA). The tweets of these companies were analyzed using (NLP) and divided into (negative - positive - natural). The results obtained by the machine learning algorithms were excellent results compared to the results obtained by previous studies and other algorithms. These results were also compared with the results of deep learning. The results showed that these algorithms that were applied in this research were better Or equivalent to Deep learning algorithms. Thus, it is possible to say that machine learning algorithms have succeeded in achieving effective and strong results and can be used in the process of predicting stock prices at the present time, including their high accuracy, especially the (SVM=90%, DT=90%, RF=90%, GBM = 96%) algorithms which have achieved a high level of accuracy.

Conflicts Of Interest

The paper states that there are no personal, financial, or professional conflicts of interest.

Funding

The absence of any funding statements or disclosures in the paper suggests that the author had no institutional or sponsor backing.

Acknowledgment

The author extends appreciation to the institution for their unwavering support and encouragement during the course of this research.

References

[1] J. H. Moedjahedy et al., "Stock price forecasting on telecommunication sector companies in Indonesia Stock Exchange using machine learning algorithms," 2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS), IEEE, 2020, pp. 1-4, doi: 10.1109/ICORIS50180.2020.9320758.

[2] M. Diqi, M. E. Hiswati, and A. S. Nur, "StockGAN: robust stock price prediction using GAN algorithm," International Journal of Information Technology, vol. 14, no. 5, pp. 2309-2315, 2022, doi: 10.1007/s41870-022-00929-6.

[3] X. Tang, S. Xu, and H. Ye, "The Way to Invest: Trading Strategies Based on ARIMA and Investor Personality," Symmetry, vol. 14, no. 11, p. 2292, 2022, doi: 10.3390/sym14112292.

[4] J. H. Moedjahedy et al., "Stock price forecasting on telecommunication sector companies in Indonesia Stock Exchange using machine learning algorithms," 2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS), IEEE, 2020, pp. 1-4, doi: 10.1109/ICORIS50180.2020.9320758.

[5] H. T. Dao et al., "Macro-economic factors affecting the Vietnam stock price index: an application of the ARDL model," The Journal of Asian Finance, Economics and Business, vol. 9, no. 5, pp. 285-294, 2022, doi: 10.13106/jafeb.2022.vol9.no5.0285.

[6] T. Lu et al., "The integration of technological advancements in fossil fuels towards environmental sustainability: Insights from big data analytics," Resources Policy, vol. 86, 2023, Art. no. 104196, doi: 10.1016/j.resourpol.2023.104196.

[7] Y. Kong et al., "The moderating role of technological innovation on environment, social, and governance (ESG) performance and firm value: Evidence from developing and least-developed countries," Sustainability, vol. 15, no. 19, p. 14240, 2023, doi: 10.3390/su151914240.

[8] Y. Tang et al., "A survey on machine learning models for financial time series forecasting," Neurocomputing, vol. 512, pp. 363-380, 2022, doi: 10.1016/j.neucom.2022.09.003.

[9] X. Shu and Y. Ye, "Knowledge Discovery: Methods from data mining and machine learning," Social Science Research, vol. 110, Art. no. 102817, 2023, doi: 10.1016/j.ssresearch.2022.102817.

[10] R. J. Woodman and A. A. Mangoni, "A comprehensive review of machine learning algorithms and their application in geriatric medicine: present and future," Aging Clinical and Experimental Research, vol. 35, no. 11, pp. 2363-2397, 2023, doi: 10.1007/s40520-023-02552-2.

[11] P. V. Reddy and S. M. Kumar, "A Novel Approach to Improve Accuracy in Stock Price Prediction using Gradient Boosting Machines Algorithm compared with Naive Bayes Algorithm," 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), IEEE, 2022, pp. 695-699, doi: 10.1109/ICAC3N56670.2022.10074387.

[12] P. V. Reddy and S. M. Kumar, "A Method for Determining the Accuracy of Stock Prices using Gradient Boosting and the Support Vector Machines Algorithm," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), doi: 10.1109/ICOSEC54921.2022.9952143.

[13] P. Koukaras, C. Nousi, and C. Tjortjis, "Stock market prediction using microblogging sentiment analysis and machine learning," Telecom, MDPI, 2022, pp. 358-378, doi: 10.3390/telecom3020019.

[14] R. M. Al-Amri et al., "The Development of a Deep Learning Model for Predicting Stock Prices," Journal of Advanced Research in Applied Sciences and Engineering Technology, vol. 31, no. 3, pp. 208-219, 2023, doi: 10.37934/araset.31.3.208219.

[15] A. Staffini, "Stock price forecasting by a deep convolutional generative adversarial network," Frontiers in Artificial Intelligence, vol. 5, Art. no. 837596, 2022, doi: 10.3389/frai.2022.837596.

[16] H. Yukhymenko and I. Lazarenko, "Forecasting The Stock Prices Using Generative Adversarial Networks And Sentiment Analysis Of Social Networks," Modeling The Development Of The Economic Systems, vol. 10, 2022.

[17] S. U. Hassan, J. Ahamed, and K. Ahmad, "Analytics of machine learning-based algorithms for text classification," Sustainable Operations and Computers, vol. 3, pp. 238-248, 2022, doi: 10.1016/j.susoc.2022.03.001.

[18] H. Li, "K-nearest neighbor," Machine Learning Methods, Singapore: Springer Nature Singapore, 2023, pp. 55-66, doi: 10.1007/978-981-99-3917-6_3.

[19] W.-Y. Loh, "Logistic regression tree analysis," Springer handbook of engineering statistics, London: Springer London, 2023, pp. 593-604, doi: 10.1007/978-1-4471-7503-2_30.

[20] H. Ming and H. Yang, "L0 regularized logistic regression for large-scale data," Pattern Recognition, vol. 146, Art. no. 110024, 2024, doi: 10.1016/j.patcog.2023.110024.COSTA, Vinícius G.; PEDREIRA, Carlos E. Recent advances in decision trees: An updated survey. Artificial Intelligence Review, 2023, 56.5: 4765-4800 , doi.org/10.1007/s10462-022-10275-5

[21] [1] A. G. C. Menezes et al., "Induction of decision trees to diagnose incipient faults in power transformers," IEEE Trans. Dielectr. Electr. Insul., vol. 29, no. 1, pp. 279-286, 2022, doi: 10.1109/TDEI.2022.3148453.

[22] [2] A. Saha, S. Basu, and A. Datta, "Random forests for spatially dependent data," J. Am. Stat. Assoc., vol. 118, no. 541, pp. 665-683, 2023, doi: 10.1080/01621459.2021.1950003.

[23] [3] D. Borup et al., "Targeting predictors in random forest regression," Int. J. Forecast., vol. 39, no. 2, pp. 841-868, 2023, doi: 10.1016/j.ijforecast.2022.02.010.

[24] [4] A. Roy and S. Chakraborty, "Support vector machine in structural reliability analysis: A review," Reliab. Eng. Syst. Saf., vol. 232, p. 109126, 2023, doi: 10.1016/j.ress.2023.109126.

[25] [5] H. Li, "Support vector machine," in Machine Learning Methods, Singapore: Springer Nature Singapore, 2023, pp. 127-177, doi: 10.1007/978-981-99-3917-6_7.

[26] [6] A. Manoharan et al., "Artificial neural networks, gradient boosting, and support vector machines for electric vehicle battery state estimation: A review," J. Energy Storage, vol. 55, p. 105384, 2022, doi: 10.1016/j.est.2022.105384.

[27] [7] H. Tyralis and G. Papacharalampous, "Boosting algorithms in energy research: A systematic review," Neural Comput. Appl., vol. 33, no. 21, pp. 14101-14117, 2021, doi: 10.1007/s00521-021-05995-8.

[28] [8] M. A. Ganaie et al., "Ensemble deep learning: A review," Eng. Appl. Artif. Intell., vol. 115, p. 105151, 2022, doi: 10.1016/j.engappai.2022.105151.

[29] [9] Y. Yang, H. Lv, and N. Chen, "A survey on ensemble learning under the era of deep learning," Artif. Intell. Rev., vol. 56, no. 6, pp. 5545-5589, 2023, doi: 10.1007/s10462-022-10283-5.

[30] [10] N. Karmous et al., "IoT real-time attacks classification framework using machine learning," in Proc. 2022 IEEE 9th Int. Conf. Commun. Netw. (ComNet), 2022, pp. 1-5, doi: 10.1109/ComNet55492.2022.9998441.

[31] [11] A. Raza et al., "Predicting employee attrition using machine learning approaches," Appl. Sci., vol. 12, no. 13, p. 6424, 2022, doi: 10.3390/app12136424.

[32] [12] P.-P. Phyo and C. Jeenanunta, "Advanced ML-based ensemble and deep learning models for short-term load forecasting: Comparative analysis using feature engineering," Appl. Sci., vol. 12, no. 10, p. 4882, 2022, doi: 10.3390/app12104882.

[33] [13] S. Nazari, M. Najafzadeh, and R. Daghigh, "Techno-economic estimation of a non-cover box solar still with thermoelectric and antiseptic nanofluid using machine learning models," Appl. Therm. Eng., vol. 212, p. 118584, 2022, doi: 10.1016/j.applthermaleng.2022.118584.

[34] [14] R. Mansoor, D. N. Hamood, and A. K. Farhan, "Image steganography based on chaos function and randomize function," Iraqi J. Comput. Sci. Math., vol. 4, no. 1, pp. 71-86, 2023, doi: 10.52866/ijcsm.2023.01.01.007.

[35] [15] S. Li, Y. Lu, and J. Li, "CAD-IDS: A cooperative adaptive distributed intrusion detection system with fog computing," in Proc. 2022 IEEE 25th Int. Conf. Comput. Support. Coop. Work Des. (CSCWD), 2022, pp. 635-640, doi: 10.1109/CSCWD54268.2022.9776147.

[36] [16] N. Karmous et al., "Enhanced machine learning-based SDN controller framework for securing IoT networks," in Int. Conf. Adv. Inf. Netw. Appl., Cham, Switzerland: Springer International Publishing, 2023, pp. 60-69, doi: 10.1007/978-3-031-28694-0_6.

[37] [17] E. J. Michaud, Z. Liu, and M. Tegmark, "Precision machine learning," Entropy, vol. 25, no. 1, p. 175, 2023, doi: 10.3390/e25010175.

[38] [18] A. H. Shamman et al., "The artificial intelligence (AI) role for tackling against COVID-19 pandemic," Mater. Today: Proc., vol. 80, pp. 3663-3667, 2023, doi: 10.1016/j.matpr.2021.07.357.

[39] [19] S. S. Sumit et al., "ReSTiNet: An efficient deep learning approach to improve human detection accuracy," MethodsX, vol. 10, p. 101936, 2023, doi: 10.1016/j.mex.2022.101936.

[40] [20] M. Makino, K. Shimizu, and K. Kadota, "Enhanced clustering-based differential expression analysis method for RNA-seq data," MethodsX, p. 102518, 2023, doi: 10.1016/j.mex.2023.102518.

[41] [21] B. Gedela and P. R. Karthikeyan, "Credit card fraud detection using AdaBoost algorithm in comparison with various machine learning algorithms to measure accuracy, sensitivity, specificity, precision and F-score," in Proc. 2022 Int. Conf. Bus. Anal. Technol. Secur. (ICBATS), 2022, pp. 1-6, doi: 10.1109/ICBATS54253.2022.9759022.

[42] [22] V. Vichianchai and S. Kasemvilas, "Text classification using improving term frequency with Gaussian and multilayer perceptron techniques," in Proc. 2023 20th Int. Joint Conf. Comput. Sci. Softw. Eng. (JCSSE), 2023, pp. 13-18, doi: 10.1109/JCSSE58229.2023.10202049.