A Solution to Improve the Detection of the Nominal Value of the Financial Market: A Case Study of the Alphabet Stocks

Zhaohua Li[1]*, Xinyue Chang[2]
School of Management, Henan Institute of Technology, Xinxiang Henan, 453003, China[1]
School of Art and Design, Henan Institute of Technology, Xinxiang Henan, 453003, China[2]

***Abstract*—Given the regular occurrence of non-stationarity, non-linearity, and high levels of noise in time series data, predicting the value of stocks is a considerable difficulty. Traditional methods have the potential to enhance the precision of forecasting, although they concurrently introduce computational complexity, hence augmenting the probability of prediction inaccuracies. To effectively tackle a range of concerns, the existing body of research proposes a novel approach that combines a light gradient boosting machine, a machine learning methodology, with artificial bee colony optimization. In the context of the examined dynamic stock market, the proposed model demonstrated better efficiency and performance compared to alternative models. The recommended model exhibited optimal performance, characterized by a low error rate and high efficacy. The analysis utilized data about the stock of Alphabet over the period spanning from January 2, 2015, to June 29, 2023. The outcomes of the study provide evidence of the predictive accuracy of the proposed model in determining stock prices. The study's findings demonstrate how well the suggested model performs when it comes to correctly predicting stock prices. The proposed model presents a pragmatic methodology for evaluating and forecasting time series data about stock prices. The research's findings show that, in terms of forecast accuracy, the suggested model performs better than the methods currently in use.**

***Keywords*—Alphabet stock; machine learning; light gradient boosting machine; optimization; artificial bee colony algorithm**

I. Introduction

Retail and institutional investors can buy and sell shares of publicly listed companies on stock exchanges to make money. It acts as a vital measure of a country's financial condition in general since it reflects firm performance and the business environment. The marketplaces where stocks are traded are the physical and online markets where both buyers and sellers come along. [1][2]. Investors, as well as traders, utilize a range of techniques to evaluate stocks and identify profitable opportunities. The investigation of stock price behavior has long piqued the curiosity of both academics and investors.

Consequently, several models have been created and tested to comprehend the underlying variables that affect the behavior of stock prices. Fama carried out one such study in 1965 [3]. This subject has been essential to developing the current knowledge of stock price behavior and is still an important area of study in the field of finance. The stock market is a complex system that fluctuates and is unpredictable. Given that pricing changes tend to be chaotic, noisy, nonparametric, nonlinear, nonstationary, and nonlinear, it is challenging for analysts to analyze and anticipate price changes accurately [4]. These characteristics imply that traditional statistical methods may not be sufficient for effective market analysis.

Therefore, a range of artificial intelligence and machine learning methods have been developed by researchers to circumvent these problems and improve the accuracy of stock market predictions. In contrast to traditional time series methods, reports based on machine learning can handle the stock market's complex, noisy, nonlinear, and unstable data to provide forecasts that tend to be more accurate [5]. Consequently, it has become the preferred technique for time series analysis in several areas [6][7]. Ensemble learning combines a variety of algorithms for machine learning to improve effectiveness, minimize mistakes, and increase accuracy [8]. This approach generates results that are more accurate and reliable than any one model could provide by integrating the results of several models that use various approaches and sets of features. The "boosting" machine learning method involves training many models in succession. Every new design aims to make the shortcomings of the prior one worse. Boosting is especially beneficial for classification and prediction issues since it may improve the precision of mediocre models. Due to their effectiveness, ensemble learning techniques like light gradient boosting machines (LGBM) and extreme gradient boosting (XGBoost) are often utilized. LGBM separates trees by the leaf and is quicker than XGBoost, which separates trees by depth [9][10]. Microsoft introduced LGBM, a machine-learning technique based on decision trees, by the end of 2017 [11]. In the machine learning field, especially in data science, LGBM has become quite well-liked because of its benefits of quick convergence time and less memory usage. It is commonly employed and discovered to be the successful answer in data mining contests [12]. Dhungana et al. [13] used LGBM to conduct an empirical investigation of S&P 500 price forecasting.

The utilization of artificial intelligence techniques in the stock market has gained significant traction due to its capacity to analyze vast quantities of data and discern patterns that are hard for humans to perceive. It is crucial to remember that these strategies' performance largely depends on how their initial parameter settings are created. Incorrect initialization will likely lead to forecasts and results that are untrustworthy [14][15]. Therefore, it is essential to thoroughly examine the

*Corresponding Author. Email: lzh104@126.com

initial configuration of the parameters when implementing artificial intelligence in stock trading. Consequently, a variety of optimization strategies have been developed to circumvent these restrictions, Like Aquila optimizer (AO) [16], Biogeography-based optimization (BBO) [17], Grey wolf optimization (GWO) [18], and Artificial bee colony (ABC) [19] and more can be used. The ABC optimizer is an optimization technique that exhibits excellent effectiveness, drawing inspiration from the behavioral patterns observed in honeybee colonies. The technique in question is classified under the swarm intelligence family and is extensively employed across diverse domains such as engineering, machine learning, and operations research. The aforementioned metaheuristic algorithm, which draws inspiration from nature, has a wide range of applications and possesses robust optimization capabilities [19].

This study tackles the formidable challenge of forecasting stock prices in time series data containing high levels of noise, non-stationary, and non-linearity. Driven by this undertaking, this research develops fundamental inquiries: In what ways can conventional forecasting techniques be refined to achieve higher levels of accuracy while preserving computational efficiency? Could the integration of an artificial bee colony optimization system with a light gradient enhancing machine present an innovative resolution to these obstacles? The primary aims of this research are to improve the accuracy of stock price predictions, reduce the complexity of computational tasks involved, and exhibit exceptional performance in this domain by utilizing the proposed fusion of methodologies. This work is significant due to the efficiency and efficacy of this model that has been demonstrated in dynamic stock markets. By employing Alphabet's stock data, the model's ideal performance, minimal error rate, and significant effectiveness, underscoring its superiority over current approaches have been demonstrated and its feasibility in practical scenarios.

The study contributed to forecasting a multivariate time series by outlining a brand-new training method based on metaheuristics for artificial bee colonies. This essay explores the field of ABC-LGBM hybrid stock price forecasting. The model was contrasted with several other models, including LGBM, GWO-LGBM, BBO-LGBM, and ABC-LGBM, to ascertain its correctness. The evaluation's findings offer insightful information on the effectiveness of the ABC-LGBM hybrid model and its potential as a stock price prediction tool. One of the analytical processes used in the inquiry was a thorough assessment of the data source and all of its pertinent features in the second part. The data was analyzed using several procedures, such as optimizer methods, evaluation metrics, and the LGBM model. The analysis results are provided and compared with results from other methodologies in the third part. The discussion part is provided in the fourth part. The investigation's results and recommendations are briefly described in the epilogue.

II. Materials and Methods

A. Artificial Bee Colony Algorithm

The development of the Artificial Bee Colony (ABC) technique arose from an examination and emulation of the foraging habits of wild bees to address optimization problems through a mathematical model. The hired bees, observers, and scouts are the three different categories of honey bee ABC agents in the colony. Two groups of bees with an equal number of members each comprise the ABC algorithm's bee population. Onlooker bees are the other half, whereas employed bees comprise the first half. Within the context of the ABC algorithm, the determination of the bee's food supply location is considered an optimization problem, including several variables that require optimization. The assessment of the solution's fitness can be employed to characterize the excellence of the food supply about the objective function of this issue. In other words, finding the best answer is similar to the process that bees go through while searching for a suitable food source and the process of the ABC method shown in Fig. 1.

These are the specifics of the ABC algorithm: The initial solutions are produced at random and used as the locations of the bee agents' food sources. The bee agents are put through three primary cycles of repetition after startup. The process of selecting the most optimal and viable alternatives while actively avoiding subpar choices and continually revising and enhancing the workable solutions. The worker bees collectively select a novel potential food source location to enhance the efficacy of their solutions. They make their decision depending on the area surrounding the previously chosen food source. Utilizing Eq. (1), the location of the new food supply is determined.

$$v_{ij} = x_{ij} + \phi_{ij}\left(x_{ij} - x_{kj}\right), \tag{1}$$

where, $k \in \{1,2,3,..,SN\}$ and $k \neq i, j \in \{1,2,3,..,D\}$ are chosen at random indexes, v_{ij} is a new feasible solution that is modified from its previous solution value $\left(x_{ij}\right)$ based on a comparison with a randomly selected position from its neighboring solution $\left(x_{kj}\right)$, and ϕ_{ij} is a random number between [-1,1] that is used to randomly adjust the previous solution to become an alternate solution in the next iteration. There is a positional difference between x_{ij} and x_{kj} in one specific dimension.

If a bee that is now employed encounters a new food source location with a higher fitness value, it will replace the previous food source position in its memory with the new one. Employed bees will share the nutritional advantages of their new food sources with the other bees when they return to their hive. The next step is for each observation bee to select one of the recommended food sources based on the fitness value calculated by the bees in use. The probability that a recommended food source will be picked is given in Eq. (2).

$$P_i = \frac{fit_i}{\sum_{i=1}^{SN} fit_i} \tag{2}$$

In which Fit $_i$ represents the fitness value of the food source i. SN stands for the number of practical food sources.

The likelihood of an observer bee choosing a given food source is positively correlated with the fitness value of that food item. Once a food source has been chosen, the observer bees will proceed to the chosen food source and identify another potential food source location within the vicinity of the earlier picked food source. The calculation and expression of the novel candidate food source can be determined using Eq. (1).

During the third stage, food-supplying positions lacking an enhanced fitness value will be discarded and substituted with a newly determined standing, assigned at random by a scout bee. This strategy aids in the prevention of suboptimal solutions. The calculation for determining the initial random place selected by the scout bee is expressed by Eq. (3):

$$x_{ij} = x_j^{min} + \mathrm{rand}[0,1]\left(x_j^{max} - x_j^{min}\right), \quad (3)$$

The lower limit and higher limit of the food supply location in dimension j are represented by the x_j^{min} and x_j^{max}, respectively.

Fig. 1. The diagram of the ABC optimizer.

The termination criteria, in this case, are based on the count of functional evaluation. The three major processes outlined above will be conducted iteratively until the predetermined number of function evaluations is reached. The cycle of bees can be shown in Fig. 2.

B. Grey Wolf Optimization (GWO)

GWO method was initially created by Mirjalili et al. [20], which mimics the hunting habits and leadership hierarchy of the Grey Wolf (Canis lupus) in the wild. The magnificent grey wolves, who are members of the Canidae family, are extremely accomplished apex predators. They are powerful in their native environments due to their amazing hunting prowess and clever group strategies. These animals prefer to live in groups and adhere to a strict social hierarchy. Depending on their characteristics, wolves may be divided into four different groups: alpha ($\boldsymbol{\alpha}$), beta ($\boldsymbol{\beta}$), delta ($\boldsymbol{\delta}$), and omega ($\boldsymbol{\omega}$). There is a common belief that within a pack of wolves, the alpha wolf is the member who possesses the most successful and efficient solution to any given problem or challenge that the pack may face. There is a common belief among some individuals that the alpha wolf, or the highest-ranking member in a wolf pack, possesses the most optimal solution to a given situation or problem and has authority over the whole pack of grey wolves. The beta and delta wolves come in second and third, respectively. The other wolves are from the omega group, which has a low rank. The alpha, beta, and delta wolves, who are the strongest, considerably help in hunting. The responsibility of tracking, chasing, encircling, and attacking the target falls on the three wolves. Following are the three main steps of the grey wolf hunting process:

- The target is pursued and tracked down
- The prey is pursued, surrounded, and harassed until it stops moving
- The prey is aimed and attacked

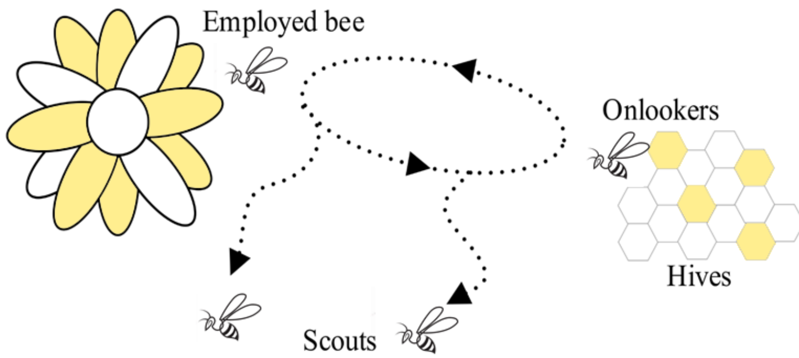

Fig. 2. The illustration of the cycle of bees.

C. Light Gradient Boosting Machine Algorithm

The construction of the model can be represented as follows since the LGBM approach is developed from decision trees. Given the data of training set $S = \{(x_i, y\}; f = 1,2,\cdots,n; x_i \in \kappa^*, y \in R\}$ where n is the number of samples containing features of m. To get an estimate, the forecasts generated by the decision trees are aggregated in the following manner:

$$y_i^{2a} = \sum_{p=1}^{p} f_p(x_0) \tag{4}$$

With f_p trees included as the trees, and there are p total trees. To get f_p, the objective function below must be minimized.

$$f_p = \arg\min_f \sum_{i=1}^{n} L(y_i y^{\text{mop}}) + \Omega(f_p) \tag{5}$$

The reduction function is marked as L, and the Ω parameter is for regularization. The specified value is provided by.

$$\hat{\Omega}(f_p) = a\,\mathrm{T} + \frac{1}{2} i \sum_{j=1}^{T} \mathrm{w}_j^2 \tag{6}$$

where, the penalty parameters for T leaves and the weight of the leaves, w, are α and λ, respectively. Assuming that, L represents the loss function and that it is a squared error. Then

$$L\left(y_i, \hat{y}_i^{LG(p-1)} + f_p(x)\right)$$

$$= \left(y_i - \hat{y}_i^{LG(p-1)} - f_p(x)\right) = \left(r - f_p(x)\right)^{2^2} \tag{7}$$

The residual, denoted as r, is utilized in the fitting process to generate the function f_p. Using a quadratic approximation function, the goal function at iteration p is minimized.

$$\begin{aligned} f_P &\simeq \arg\min_{f_p} \sum_{l=1}^{n} \left[g f_P(x_i) + \frac{1}{2} h_i f_P^2(x_i)\right] + \Omega(f_P), \\ g_{\mathrm{I}} &= \partial_j \omega_{(p-1)} L\left(y_l, \hat{y}_l^{LC(p-1)}\right), \\ h_{\mathrm{I}} &= \partial_{y^2{}^2_{(p-1)}} L\left(y_l, \hat{y}_l^{LC(p-1)}\right). \end{aligned} \tag{8}$$

A new tree by minimizing the goal function, denoted as f_p, is produced. The decision tree partitions each node based on the criterion of the greatest data gain.

The variance gains for a node at position s that divides feature j are provided by:

$$Z_{j|O}(s) = \frac{1}{n_O}\left\{\frac{\left(\sum_{\{x_i \in O: x_{ij} \le s\}} g_i\right)^2}{n_{l|O}^j(s)} + \frac{\left(\sum_{\{x_i \in O: x_{ij} > s\}} g_i\right)^2}{n_{r|O}^j(s)}\right\}, \tag{9}$$

D. Data Preparation and Collection

To perform a comprehensive analysis, it is essential to include the trade volume as well as the Open, High, Low, and Close (OHLC) prices within a designated time frame. The dataset utilized in this research aims to enable the prediction of Alphabet stock market prices throughout an extensive temporal span, spanning from January 1, 2015, to the middle of 2023. Precise forecasting of stock prices has paramount importance for shareholders, financial professionals, and managers operating within the finance industry. This dataset contains the necessary historical stock price information and its related characteristics for conducting prediction analyses. The key sources of financial market information for the dataset are the stock exchanges and financial news outlets. Historical daily market share prices for Alphabet were collected for the chosen period. Between January 1, 2015, and mid-2023, this paper's dataset includes various variables about Alphabet stock shares for each trading day. The essential components encompassed within the context of stock market data are the specific date, the initial price at the commencement of the session of trading, the final price after the trading session, the maximum price attained by the shares throughout the day, the minimum price reached by the shares during the day, and the trading volume denoting the aggregate number of shares exchanged within the day. To ensure the quality and consistency of the data, rigorous data preparation procedures were implemented before conducting any predictive analysis. Data standardization was performed to facilitate precise modeling and prediction. Normalizing data is the transformation of numerical parameters to a standardized range, often ranging from 0 to 1 or with a mean of 0 and a standard deviation of 1. When working on analytical or modeling projects, it's important to treat variables with different units or magnitudes equally. The use of diverse techniques, including scaling and normalization, is vital in the process of data cleaning. These approaches play a crucial role in mitigating gradient mistakes and ensuring consistent training outcomes. Based on the information shown in Fig. 3, the dataset was divided into two parts: 80% of the data was allocated for training, while the remaining 20% was reserved for testing.

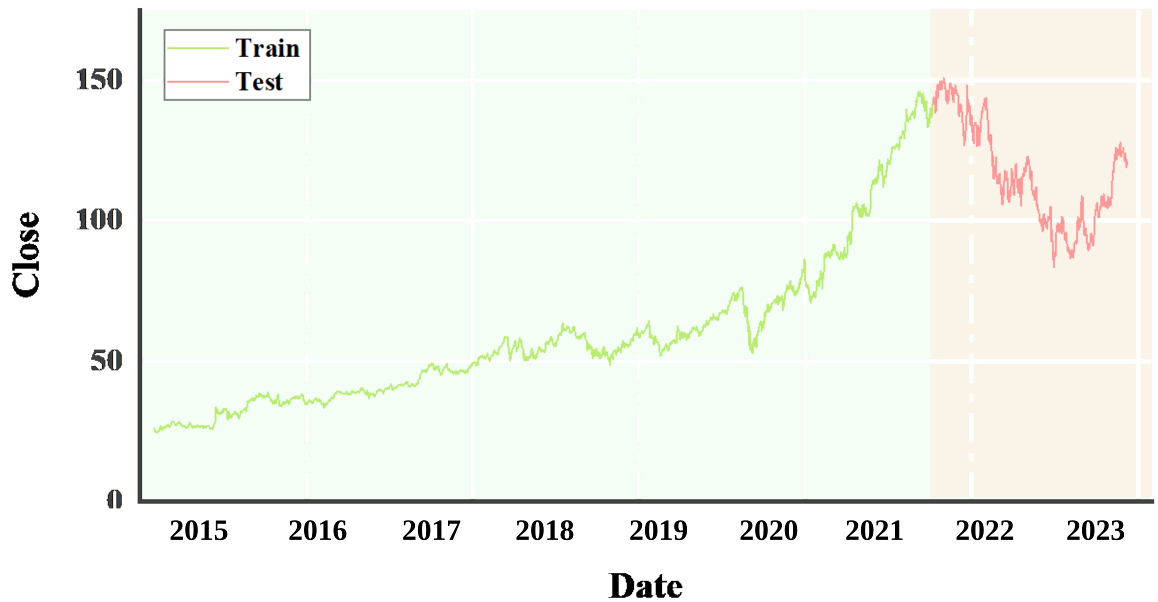

Fig. 3. Dividing data for both training and testing.

E. Assessment Criteria

The purpose is to evaluate the ability of combined models to make accurate predictions. This category comprises the root mean square error (RMSE), mean absolute percentage error (MAPE), mean squared error (MSE), and coefficient of determination (R^2).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{10}$$

$$MAPE = \left(\frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|\right) \times 100 \tag{11}$$

$$\mathrm{MSE} = \frac{1}{N}\sum_{\mathrm{k}=0}^{\mathrm{n}}\binom{\mathrm{n}}{\mathrm{k}}(\mathrm{Fi} - \mathrm{Yi})\mathrm{b}^2 \tag{12}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{13}$$

III. Results

A. Statistical Results

Table I presents an extensive compilation of statistical information about the dataset under consideration. The OHLC price and volume figures shown in this table provide a more clear depiction of the factual information. To conduct a more comprehensive and precise evaluation of the data, it is advisable to use statistical metrics such as the mean, amount, mean, std., min, median, max, and variance numbers.

B. Results of the Models

Finding and evaluating the best hybrid algorithm for predicting stock prices is the main goal of this study. Forecasting models have been developed, and complex variables that affect stock market trends have been investigated. The primary objective was to provide investors and analysts with valuable information to facilitate informed investment decision-making. Table II, Fig. 4 and Fig. 5 present a comprehensive assessment of the performance of each model, accompanied by an in-depth examination of its effectiveness.

A comprehensive data analysis assessment was conducted using four commonly employed metrics: RMSE, MSE, MAPE, R^2. These metrics are widely recognized for their ability to offer a precise assessment of the dependability, precision, and overall usefulness of the analysis. The efficacy evaluation of the LGBM model was conducted using the RMSE, MSE, MAPE, and R^2 metrics, comparing the model's performance with and without an optimizer. Enhanced understanding of the model's performance facilitates the ability to make well-supported judgments based on the obtained outcomes.

C. Comparison with Other Works

In order to assess the accuracy of the ABC-LGBM model, the primary objective of this research was to examine its performance. To accomplish this, we conducted a comparative analysis of the R^2 values of our framework and other recent works. Upon meticulous examination, we determined that the ABC-LGBM exhibited superior performance to the other works, thereby substantiating the effectiveness of this framework. The specifics of this comparison are provided in Table III.

TABLE I. DATASETS STATISTICAL SUMMARIES

	Open	High	Low	Volume	Close
amount	2137	2137	2137	2137	2137
mean	70.05219	70.81457	69.3428	32.59751	70.09629
Std.	34.54605	34.97686	34.14654	15.6062	34.55914
min	24.66478	24.7309	24.31125	6.936	24.56007
median	58.4235	58.9	57.871	28.734	58.4095
max	151.8635	152.1	149.8875	223.298	150.709
variance	1193.43	1223.381	1165.986	243.5536	1194.334

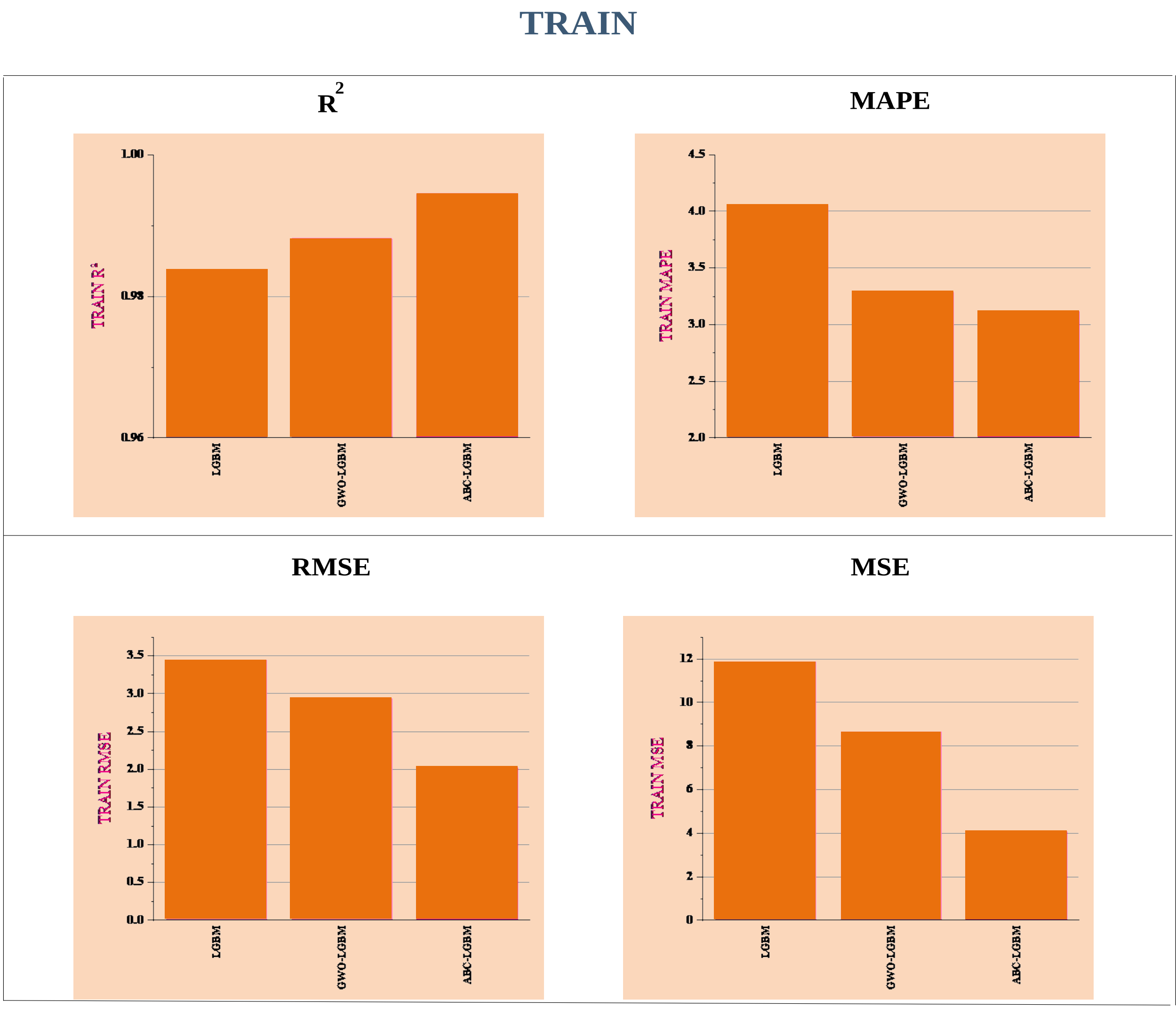

Fig. 4. The values to evaluate each model during training.

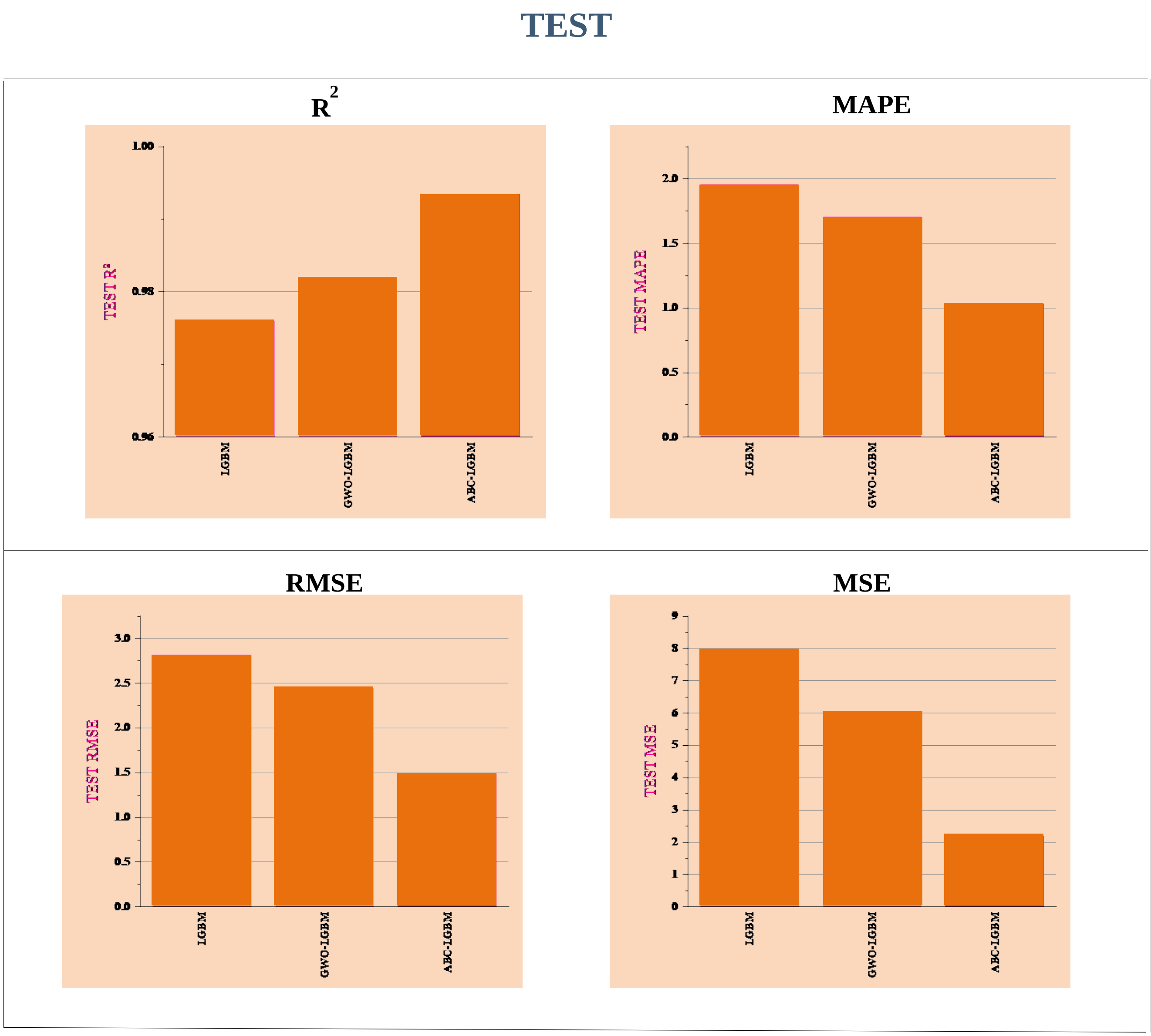

Fig. 5. The values to evaluate each model during testing.

TABLE II. THE RESULTS OF FORECASTING EVALUATION FOR THE BENCHMARKING METHODS

MODEL / Metrics	TRAIN SET				TEST SET			
	R^2	*RMSE*	*MAPE*	*MSE*	R^2	*RMSE*	*MAPE*	*MSE*
LGBM	0.984	3.446	4.054	11.878	0.976	2.821	1.955	7.958
GWO-LGBM	0.988	2.940	3.286	8.646	0.982	2.449	1.709	5.997
ABC-LGBM	0.994	2.022	3.118	4.089	0.993	1.479	1.029	2.188

TABLE III. COMPARISON OF THE MODEL IN COMPARISON TO THE OTHER RECENT WORKS

References	*Model*	R^2
[21]	LSTM	0.977
[22]	DNN LSTM	0.972
Present study		0.993

IV. DISCUSSION

This study aims to find and evaluate the best hybrid stock price prediction system. Complex stock market variables have been studied and forecasting algorithms constructed. Information for investors and analysts to make informed investment decisions was the main goal. Table II, Fig. 4 and Fig. 5 evaluate each model's performance and efficacy. A thorough data analysis was performed utilizing four popular metrics: RMSE, MSE, MAPE, R^2 . These criteria are extensively used to evaluate the analysis's dependability, precision, and usefulness. The LGBM model was evaluated for efficacy using RMSE, MSE, MAPE, and R^2 metrics, comparing performance with and without an optimizer.

Upon meticulous analysis of the training and test datasets, it was ascertained that the LGBM model, when implemented without the optimizer, yielded R^2 coefficients of 0.984 and 0.976 for the respective training and testing datasets. Furthermore, MAPE and MSE values obtained for the test dataset were notably lower at 1.955 and 7.958, respectively, in comparison to the corresponding values obtained during the training phase. Furthermore, RMSE values for the training and testing sets were 3.446 and 2.821, respectively. The integration of optimization techniques has significantly enhanced the efficacy of the LGBM model's performance. The usage of the GWO has yielded significant improvements, resulting in a rise in the R^2 value to 0.988 during the training phase and 0.982 during the testing phase. Furthermore, the MAPE and MSE values displayed a decrease in both the testing and training datasets.

Specifically, the MAPE values for the training and testing sets were 3.286 and 1.709, while the MSE values were 8.646 and 5.997, respectively. The RMSE values have decreased to 2.449 for the testing set and 2.940 for the training set. The findings of this research demonstrate the efficacy of the optimization techniques employed in enhancing the efficacy of the LGBM model. The ABC-LGBM model demonstrated superior performance in comparison to the GWO-LGBM model. The values of R^2 obtained for the training phase and testing phases of the ABC-LGBM model were 0.994 and 0.993, respectively. Significantly, the training MAPE and MSE values exhibited a decrease to 3.118 and 4.089, respectively. Similarly, the testing MAPE and MSE values experienced a decrease to 1.029 and 2.188, respectively, thereby suggesting a significant enhancement in precision. Moreover, it can be observed that the RMSE values showed a decrease to 2.022 and 1.479 for the test and training datasets, accordingly. This reduction in error further substantiates the superior performance of the ABC-LGBM model in comparison to the GWO-LGBM model, as it demonstrates the ABC-LGBM's ability to provide accurate forecasts. The investigation demonstrates that the ABC-LGBM model exhibits superior effectiveness and efficiency compared to the GWO-LGBM model. The ABC-LGBM model exhibits a high level of efficacy, as seen in its outstanding R^2 scores of 0.994 for training and 0.993 for testing. The model had outstanding efficiency, as evidenced by its lowest MAPE and MSE testing values of 1.029 and 2.188, respectively. Findings suggest that the ABC-LGBM model exhibits a high level of precision and reliability.

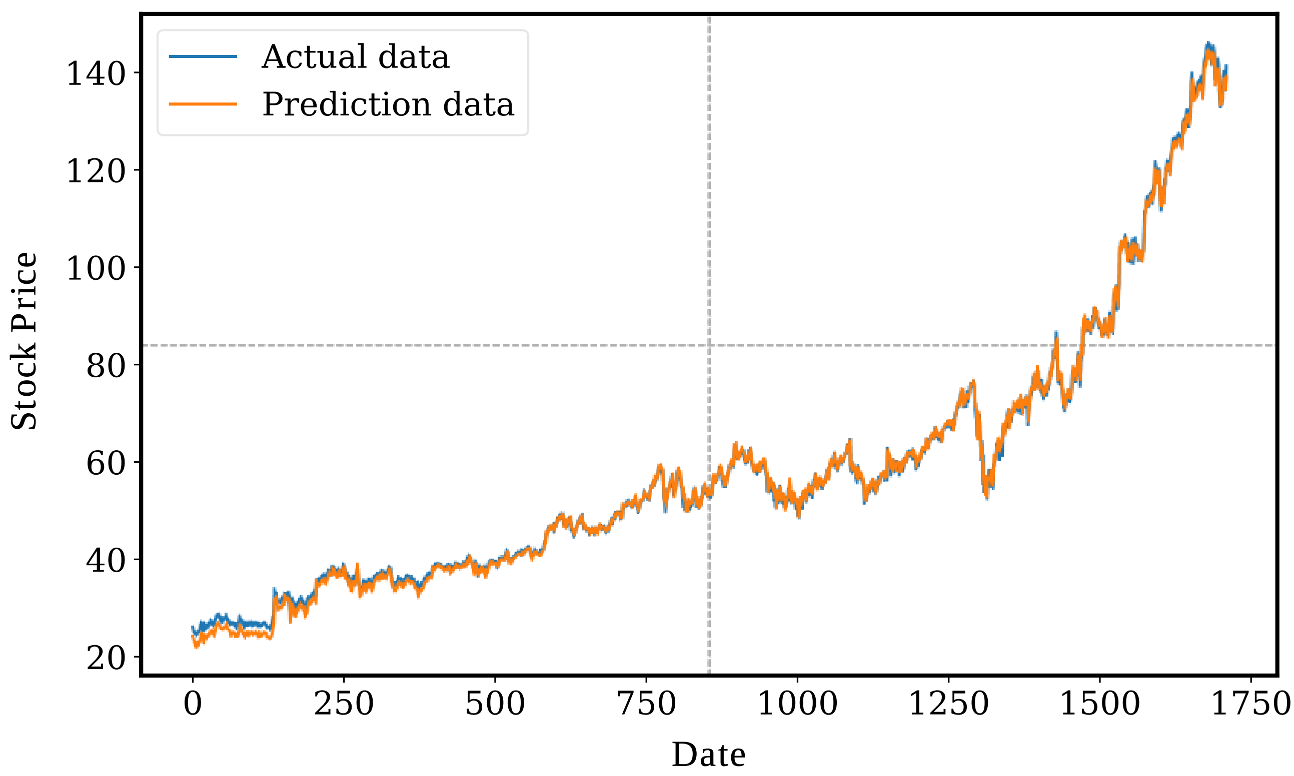

Fig. 6. Forecasting curve training created with ABC-LGBM.

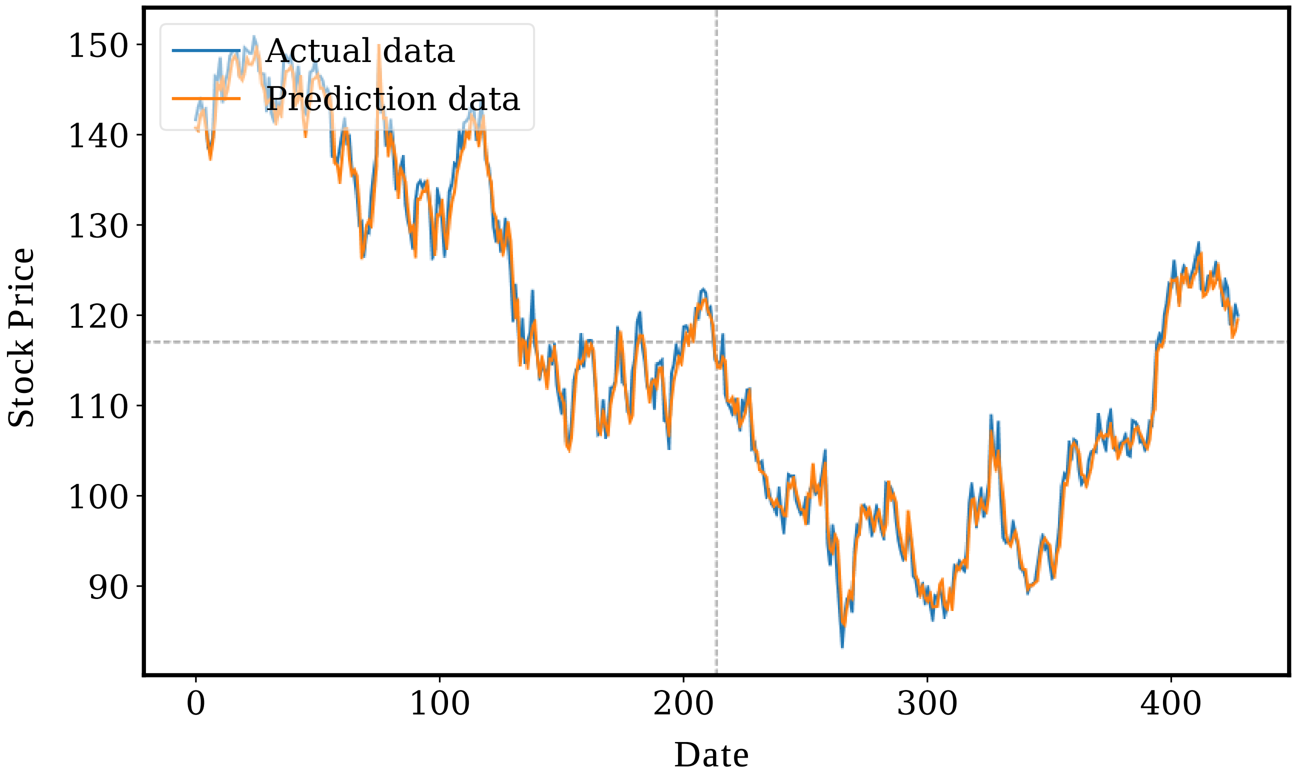

Fig. 7. Forecasting curve testing created with ABC-LGBM.

The ABC-LGBM model is widely regarded as a reliable and robust technique for generating highly precise market price forecasts. The stock share curves of Alphabet are presented in Table II, along with the corresponding figures, namely Fig. 6 and Fig. 7, which serve as evidence of the effectiveness of the model. In terms of precise prediction of stock prices, the ABC-LGBM model has superior performance compared to alternative models such as LGBM and GWO-LGBM. The ABC method is a highly effective technique that has been found to reduce price fluctuations significantly. By doing so, it has the added benefit of making trend prediction much easier and increasing overall model accuracy. This can be highly advantageous in a variety of industries where accurate forecasting is critical to success.

Additionally, the method is particularly useful in minimizing the impact of unforeseen events that might otherwise lead to significant market fluctuations. One of the distinctive features that sets the ABC-LGBM model apart from alternatives is its ability to learn from previous data effectively. In conclusion, the ABC-LGBM model has notable efficacy as a tool for predicting stock prices due to its high level of precision, accuracy, and capacity to assimilate information from previous datasets.

V. Conclusion and Recommendations

Forecasting stock prices is a complex endeavor characterized by a multitude of variables. The stock market is a complex and fluid process that is influenced by various factors, such as political events, societal dynamics, and economic conditions. To accurately assess the upcoming value of stocks, it is imperative to consider a diverse array of financial statements, market trends, as well as other pivotal factors. Furthermore, the behavior of stocks can be significantly influenced by economic factors such as interest rates, inflation, and worldwide market conditions. Developing trustworthy models for forecasting is a challenging task due to the intricate nature and extensive array of components involved. The ability to generate precise forecasts necessitates a comprehensive comprehension of the capricious and nonlinear characteristics inherent in the marketplace. Fortunately, the ABC-LGBM model is a viable solution to these issues and has been proven to be dependable and precise. There are numerous ramifications of this research for the community. The proposed model begins by tackling a significant obstacle in the prediction of stock prices, a matter that holds considerable importance for financial analysts, investors, and individuals engaged in stock market operations. Through adeptly managing the intricacies of non-linearity, non-stationarity, and elevated noise levels present in time series data, the model furnishes a more dependable and precise instrument for predicting stock prices. Furthermore, the integration of artificial bee colony optimization with light gradient boosting machine integration represents a novel strategy that not only improves accuracy but also addresses the issue of computational intricacy. This has ramifications that extend beyond the realm of academia, as it provides pragmatic resolutions to the practical obstacles encountered by practitioners in the financial industry. The utilization of Alphabet's stock data for analysis enhances the study's practicality, given that Alphabet is a significant participant in the stock market. The research findings, which illustrate the model's exceptional performance in comparison to alternative approaches, possess the capacity to impact the decision-making procedures of financial institutions and investors. In brief, this study not only enhances the theoretical comprehension of time series data analysis but also offers a practical and influential instrument for the field, enabling financial stakeholders to anticipate and navigate stock market trends more efficiently. The study examined the LGBM and GWO-LGBM models as part of its investigation into forecasting stock prices. However, the ABC optimizer technique, in combination with LGBM, produced the best outcomes. The study's dataset comprises Alphabet stock OHLC prices and volume from January 2, 2015, to June 29, 2023. Based on the findings of the inquiry, it has been determined that the ABC-LGBM model exhibits a high level of reliability and accuracy in predicting stock prices.

- Throughout the investigation, a comparative analysis was conducted to assess the accuracy and forecasting capacity of the ABC-LGBM model about other models. Based on the obtained data, it was consistently observed that the ABC-LGBM model exhibited superior performance compared to each of the other models. The R^2 value of 0.993 obtained after testing indicates a high level of accuracy, hence proving the precision of the predictions. The ABC-LGBM model consistently generated precise forecasts, as indicated by its minimal MAPE value of 1.029 and a low MSE value of 2.188. Overall, in terms of accuracy and efficacy, the ABC-LGBM model demonstrated outstanding results compared to the other approaches that were evaluated.

Recommendations:

Promotion of Hybrid Approach Adoption: Advocate for the implementation of the proposed hybrid approach, which integrates LGBM and ABC, on the grounds that it exhibits enhanced efficiency and performance in comparison to alternative models. Highlight the potential advantages that this model may offer in mitigating the difficulties presented by time series data's non-stationarity, non-linearity, and substantial levels of noise.

Application in Dynamic Stock Markets: It is recommended that the suggested model be implemented in dynamic stock markets as a means to improve the accuracy of predictions, particularly in situations where conventional approaches may introduce computational intricacy and possible errors. It is recommended to conduct additional validation and testing of the proposed model using a variety of stock datasets in order to evaluate its adaptability and efficacy across distinct market conditions. Investigate its functionality across a range of financial scenarios in order to ascertain its resilience.

Thorough Evaluation of Variables: Emphasize the significance of taking into account a wide range of financial statements, market trends, and critical factors in order to generate precise predictions regarding stock prices. Advocate for researchers and practitioners to adopt a comprehensive approach by incorporating a range of factors, such as economic conditions, political events, and societal dynamics.

Practical Significance: Elucidate the ways in which the ABC-LGBM model is effectively employed to tackle obstacles arising from the presence of noise, non-linearity, and non-stationarity within time series data. It is recommended to apply the model in practical financial contexts in order to assess its efficacy and influence on decision-making procedures.

Extension to Other Sectors: Advocate for an investigation into the feasibility of implementing the ABC-LGBM model in sectors other than finance, taking into account its capacity to offer practical solutions to computational complexities and forecasting obstacles. Promote collaborations across industries in order to capitalize on the capabilities of the model in various domains.

Sustained Comparative Analysis: Support the implementation of ongoing comparative analyses to evaluate the ABC-LGBM model's accuracy and predictive capability in comparison to emergent models and evolving methodologies. Motivate scholars to investigate its efficacy across diverse market environments and utilize an assortment of datasets.

References

[1] Y.-H. Wang, C.-H. Yeh, H.-W. V. Young, K. Hu, and M.-T. Lo, “On the computational complexity of the empirical mode decomposition algorithm,” Physica A: Statistical Mechanics and its Applications, vol. 400, pp. 159–167, 2014, doi: https://doi.org/10.1016/j.physa.2014.01.020.

[2] C. Zhang, J. Ding, J. Zhan, and D. Li, “Incomplete three-way multi-attribute group decision making based on adjustable multigranulation Pythagorean fuzzy probabilistic rough sets,” International Journal of Approximate Reasoning, vol. 147, pp. 40–59, 2022, doi: https://doi.org/10.1016/j.ijar.2022.05.004.

[3] E. F. Fama, “Random walks in stock market prices,” Financial analysts journal, vol. 51, no. 1, pp. 75–80, 1995.

[4] Y. S. Abu-Mostafa and A. F. Atiya, “Introduction to financial forecasting,” Applied Intelligence, vol. 6, no. 3, pp. 205–213, 1996, doi: 10.1007/BF00126626.

[5] Y. Chen and Y. Hao, “A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction,” Expert Syst Appl, vol. 80, pp. 340–355, 2017, doi: https://doi.org/10.1016/j.eswa.2017.02.044.

[6] M. Zounemat-kermani, O. Kisi, and T. Rajaee, “Performance of radial basis and LM-feed forward artificial neural networks for predicting daily watershed runoff,” Appl Soft Comput, vol. 13, no. 12, pp. 4633–4644, 2013, doi: https://doi.org/10.1016/j.asoc.2013.07.007.

[7] R. Bisoi, P. K. Dash, and A. K. Parida, “Hybrid Variational Mode Decomposition and evolutionary robust kernel extreme learning machine for stock price and movement prediction on daily basis,” Appl Soft Comput, vol. 74, pp. 652–678, 2019, doi: https://doi.org/10.1016/j.asoc.2018.11.008.

[8] L. G. Kabari and U. C. Onwuka, “Comparison of bagging and voting ensemble machine learning algorithm as a classifier,” International Journals of Advanced Research in Computer Science and Software Engineering, vol. 9, no. 3, pp. 19–23, 2019.

[9] T. R. Mahesh, V. Vinoth Kumar, V. Muthukumaran, H. K. Shashikala, B. Swapna, and S. Guluwadi, “Performance analysis of xgboost ensemble methods for survivability with the classification of breast cancer,” J Sens, vol. 2022, pp. 1–8, 2022.

[10] Y. Zhou, W. Wang, K. Wang, and J. Song, “Application of LightGBM Algorithm in the Initial Design of a Library in the Cold Area of China Based on Comprehensive Performance,” Buildings, vol. 12, no. 9, p. 1309, 2022.

[11] G. Ke et al., “Lightgbm: A highly efficient gradient boosting decision tree,” Adv Neural Inf Process Syst, vol. 30, 2017.

[12] M. Ustuner and F. Balik Sanli, “Polarimetric target decompositions and light gradient boosting machine for crop classification: A comparative evaluation,” ISPRS Int J Geoinf, vol. 8, no. 2, p. 97, 2019.

[13] S. Chakrabarty, P. Dhungana, and S. K. Sarada, “Application of Ensembles for Stock Index Price Prediction,” 2022.

[14] Y. Han, P. Pan, H. Lv, and G. Dai, “A hybrid optimization algorithm for water volume adjustment problem in district heating systems,” International Journal of Computational Intelligence Systems, vol. 15, no. 1, p. 39, 2022.

[15] M. Liang, S. Wu, X. Wang, and Q. Chen, “A stock time series forecasting approach incorporating candlestick patterns and sequence similarity,” Expert Syst Appl, vol. 205, p. 117595, 2022.

[16] L. Abualigah, D. Yousri, M. Abd Elaziz, A. A. Ewees, M. A. A. Al-Qaness, and A. H. Gandomi, “Aquila optimizer: a novel meta-heuristic optimization algorithm,” Comput Ind Eng, vol. 157, p. 107250, 2021.

[17] D. Simon, “Biogeography-based optimization,” IEEE transactions on evolutionary computation, vol. 12, no. 6, pp. 702–713, 2008.

[18] S. Mirjalili, S. M. Mirjalili, and A. Lewis, “Grey wolf optimizer,” Advances in engineering software, vol. 69, pp. 46–61, 2014.

[19] D. Karaboga, “Artificial bee colony algorithm,” scholarpedia, vol. 5, no. 3, p. 6915, 2010.

[20] V. Chandran and P. Mohapatra, “Enhanced opposition-based grey wolf optimizer for global optimization and engineering design problems,” Alexandria Engineering Journal, vol. 76, pp. 429–467, 2023, doi: https://doi.org/10.1016/j.aej.2023.06.048.

[21] Z. Jin, Y. Yang, and Y. Liu, “Stock closing price prediction based on sentiment analysis and LSTM,” Neural Comput Appl, vol. 32, pp. 9713–9729, 2020.

[22] A. C. Nayak and A. Sharma, PRICAI 2019: Trends in Artificial Intelligence: 16th Pacific Rim International Conference on Artificial Intelligence, Cuvu, Yanuca Island, Fiji, August 26–30, 2019, Proceedings, Part II, vol. 11671. Springer Nature, 2019