Predicting Future Trends with an Ensemble Approach in Time Series Forecasting

Mohamed EL Mahjouby, Mohamed Taj Bennani, Mohamed EL FAR
Department of Computer Science, Laboratory (LPAIS), Faculty of Science Dhar EL Mahraz,
University Sidi Mohamed Ben Abdellah, Fez, Morocco
mohamed.elmahjouby@usmba.ac.ma, taj.bennani@usmba.ac.ma, mohamed.elfar@usmba.ac.ma

***Abstract*— Predicting economic and financial time series has consistently presented challenges due to their susceptibility to influences stemming from economic, political, and social variables. Consequently, individuals involved in speculative activities within financial markets often seek strong models. Many advanced algorithms exist for forecasting the behavior of financial markets. This article introduces an approach that combines adaptive boosting regression with linear regression as the foundational estimator. The objective is to employ this amalgamation technique to forecast future closing prices of NASDAQ, gold commodities, the British pound sterling relative to the US dollar, and the euro against the US dollar. Our methodology integrates seven technical indicators in training adaptive boosting regression to enhance precision in predicting future closing prices. The assessment incorporates four metrics— coefficient of determination, mean absolute percentage error, root mean squared error, and mean squared error—to compare different machine-learning models. The analysis of experiments indicates that our technique attains superior accuracy compared to adaptive gradient-boosting regression, extreme gradient-boosting, and linear regression. This research will enable investors to make well-informed choices regarding future closing prices for four datasets, enabling them to determine the optimal timing for entering or exiting the market.**

***Index Terms* — Ensemble model, Time Series, Linear Regression, Adaptive Boosting Regression, Root Mean Squared Error**

I. INTRODUCTION

The global marketplace for currency exchange, commonly known as the Forex market, facilitates the trading of currencies around the clock, excluding weekends. Utilizing time series forecasting is beneficial for traders in enhancing their decision-making processes. With different improvements in forecasting time series, there is a genuine requirement for additional research and analysis in this field.

Artificial intelligence is vital, given the growing complexity of time series forecasting. It has proven its ability to predict financial markets effectively by integrating diverse analyses, thereby enhancing prediction accuracy and overall quality.

Some scholars have suggested various approaches for forecasting time series data. Another method aims to enhance the prediction of exchange rates in the financial domain. The paper presented in reference [1] suggests an approach for forecasting volatility based on financial data. In [2], the authors combined ARIMA and MLP for stock prediction. Additionally, a technique introduced in [3] aims to enhance the predictive capabilities of stock price trends. In reference [4], the authors employed linear regression for forecasting time series. This model utilizes the EUR_USD currency pair for making predictions.

Our investigation significantly contributes to this domain, and additionally, we will delve into a captivating scenario forecasting upcoming closing prices. The recommended ensemble model was employed to forecast the future closing prices of significant foreign currency pairs, along with the prices of gold and Nasdaq. This model is used to predict the prices of Gold, NASDAQ, and various currency pairs. Analysis of the experimental outcomes involves four employed performance metrics: RMSE, MAPE, MAE, and the coefficient of determination. The proposed model evaluates against linear regression, XGBoost, and AdaBoost model performance metrics.

The format is as follows: The next section produces a review of the literature review. In the third, we furnish the data and methodology used. Fourth, we describe the architecture of the proposed model. The fifth section examines performance measures within this field. Transitioning to the sixth, we describe the results obtained. Lastly, the concluding section offers final remarks.

II. LITERATURE REVIEW

Most researchers in this field utilize time series data alongside deep learning approaches to forecast prices. In [5], Scientists commonly employ the Decision Tree algorithm for forecasting financial time series due to its capacity to produce straightforward and analyzable rules. In reference [6], the authors employed an LSTM model to predict the closing price of the Dow Jones index for the following day. They compared the results of the LSTM model with those of a feed-forward neural network (FFNN), revealing that the LSTM model surpassed the FFNN in terms of predictive accuracy.

In reference [7], authors applied a deep learning technique called stacked LSTM to forecast the closing price of the NASDAQ Composite on the American Stock Exchange. Their study showed an improvement in predictive accuracy when predicting stock prices.

In reference [8], the researchers utilized an LSTM model to forecast the closing price using historical stock data from the Taiwan Stock Exchange (TWSE). The main objective was to replace the RNN model with the LSTM model due to its superior ability to retain long-term memory, addressing the RNN's difficulty in maintaining data over extended periods.

In [9], the authors introduced a novel strategy for forecasting stock market trends, employing a combination of LSTM and GA. Their research indicated that this hybrid approach surpassed the performance of the benchmark model. Meanwhile, in [10], the authors fine-tuned four ensemble models specifically crafted for assessing macroeconomic factors to improve the precision of predicting stock market movements one month in advance. These models encompass boosting, bagging, neural network ensemble regressor, and random forest. Another aim was to integrate LSTM into a hybrid framework, demonstrating that macroeconomic variables offer the most reliable predictions for the stock market.

In reference [11], the authors introduced an innovative stock forecasting technique employing Support Vector Regression (SVR) with a refined version. The process involves utilizing grid search for the selection and optimization of the kernel function, resulting in decreased time and memory demands and improved accuracy. This proposed method will be employed to evaluate diverse metrics related to stock market performance, demonstrating superior accuracy and faster computation when compared to similar approaches in terms of RMSE and MAPE.

In reference [12], the authors used RNN and CNN algorithms. Evaluating the precision of these models involves comparing their effectiveness against actual stock market data in real-world scenarios.

III. MATERIALS AND METHODS

In this portion, we have introduced the specified information: the data source, the different technical indicators, and the methods to predict time series.

A. Dataset

This research employed machine learning technology to forecast the prices of GBP_USD, EUR_USD, Gold commodities, and NASDAQ. The dataset used for this analysis spans from December 1, 2003, to February 2, 2023, encompassing daily prices without any gaps. Consequently, the cumulative observation period spans 20 years, covering High, Open, Low, and Close values. The historical closing price trends from 2003 to 2023 are in Figure 1.

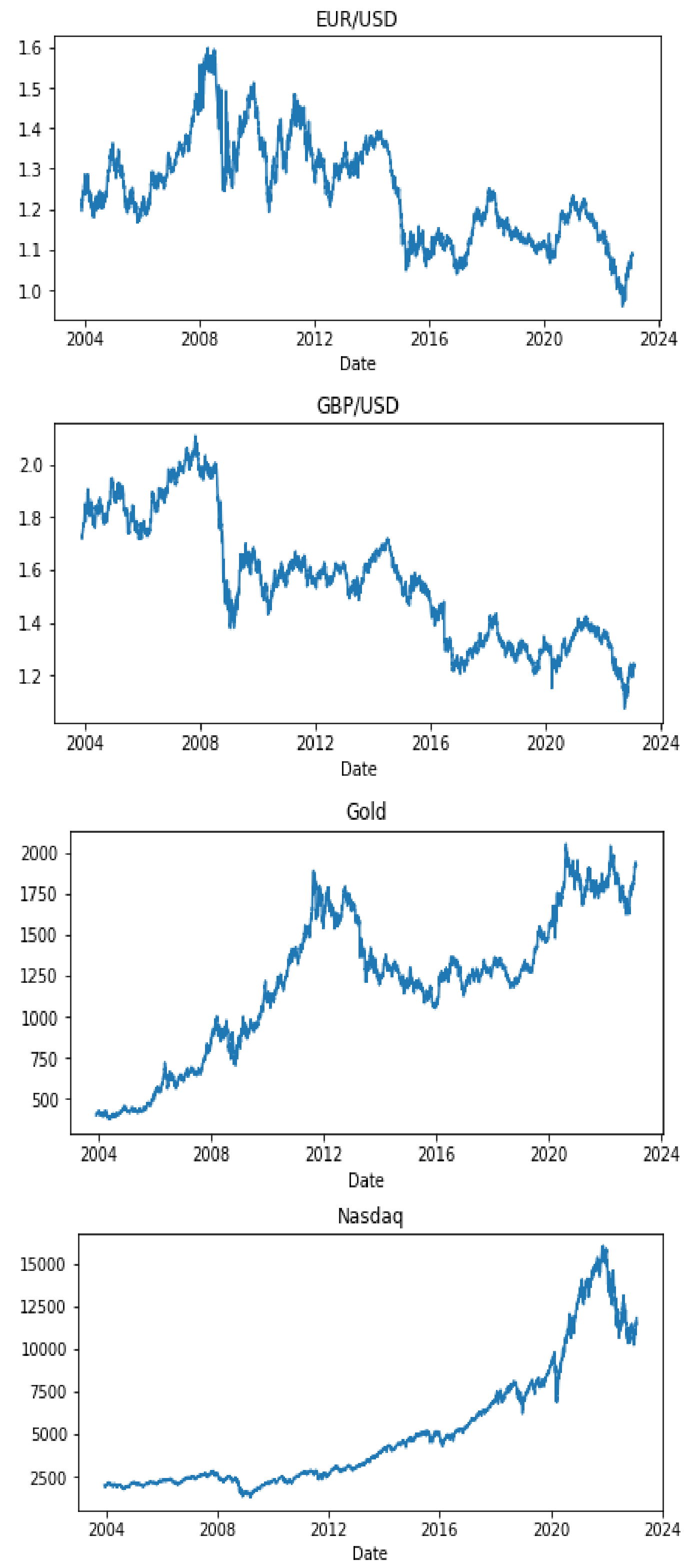

Fig. 1. The Closing prices.

B. Technical Indicators

The provided technical indicators are derived from past data and for predicting stock prices in financial markets employed. Reference [13] delves into various technical indicators such as the momentum indicator, MACD, stochastic, RSI, and William's to explore their effectiveness in price prediction.

In this study, we made use of seven distinct technical indicators:

1) The 20-day exponential moving average.
2) Index of relative strength.
3) An exponential moving average spanning 100 days.
4) An exponential moving average spanning 150 days.
5) The future close price is the target.
6) TargetClass is verifying whether the price is increasing or decreasing.

This paper uses a target price called TargetNextClose, the next closest price.

C. *Linear Regression (LR)*

We use LR to predict time series. In this case, Equation (1) represented mathematical linear regression. The independent variable is by the letter y, and the dependent variable is by the letter W. The interceptor is by the symbol c, and the slope is called b.

$$w = by + c \tag{1}$$

D. *XGBoost*

XGBoost, as outlined in [14], is a widely employed model across different machine learning methodologies, proving highly efficient. Moreover, other algorithms like stochastic gradient boosting and tree boosting are available as alternatives.

E. *AdaBoost (Adaptive Gradient Boosting)*

AdaBoost, outlined in [15], is the pioneering boosting technique created by Freund and Schapire. Over time, it has evolved into one of the most widely used methods in diverse fields owing to its utility and adaptability.

AdaBoost, also known as Adaptive Boosting, is a method in ensemble learning designed to enhance the effectiveness of weak learners, usually basic models, by amalgamating them to form a robust learner. It belongs to the boosting family of algorithms, and its idea is to assign different weights to the observations in the dataset based on their performance. Misclassified instances receive higher weights, and the weaker students that follow concentrate more on fixing these errors.

F. *Proposed Method*

Our research centers around a Machine Learning framework, with our primary focus being the prediction model we've developed. This model comprises several stages: (1) Downloading data, (2) feature engineering,(3) preprocessing the data, (4) building the model, and (5) forecasting future closing prices. Figure 2 illustrates the architecture of our model.

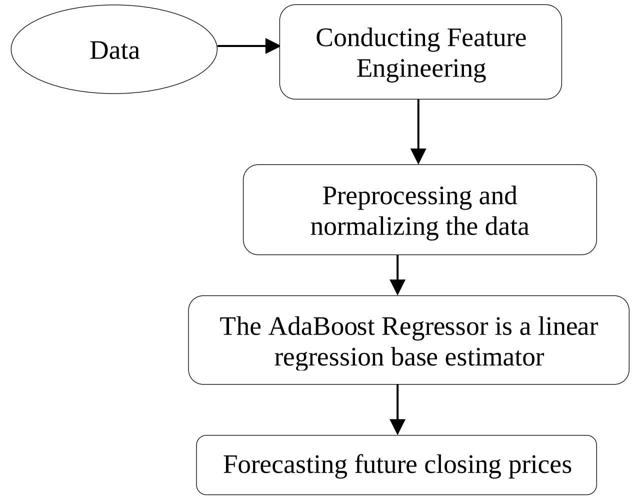

Fig. 2. The structure of our model

IV. EVALUATION MEASURES

To assess how well the models utilized in this study perform. Metrics such as MAE, RMSE, MAPE, and R^2 are employed. A decrease in these metrics indicates a more effective model.

The RMSE is a commonly employed measure to assess the performance of regression models. It computes the square root of the mean of the squared differences between predicted and observed values, acting as a gauge of the model's precision[16]. This calculation is by Equation (2).

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

MAPE, an acronym for Mean Absolute Percentage Error, is a measurement used to evaluate the accuracy of forecasting and prediction. The formula for MAPE is in Equation (3) [17].

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\text{actual} - \text{forecast}}{\text{actual}}\right| \tag{3}$$

MAE is a metric employed in conjunction with regression models. The Mean Absolute Error (MAE) is a metric commonly used to assess the accuracy of a regression model. It calculates the average absolute differences between the predicted and actual values, providing a measure of the model's performance without considering the direction of the errors. The MAE formula is the average of the absolute value variances between predicted and actual values. Equation (4) illustrates the formula for calculating this metric [18].

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y - y_i| \tag{4}$$

The coefficient R^2, also known as R-squared, serves as a scoring metric for evaluating the performance of a linear regression model. It is a numeric measure ranging from 0 to 1, providing insight into the precision of a statistical model in predicting outcomes. Equation (5) expresses the formula for calculating R^2.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y_i})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \quad (5)$$

$\overline{y}$: The mean value of the measurement.
$\hat{y}_i$: The anticipated outcome.

V. RESULTATS AND DISCUSION

The models underwent training on four datasets, namely GBP_USD, EUR_USD, NASDAQ, and Gold. Subsequently, we normalized the data using the Sklearn library in Python, computed seven indicators, and eliminated any missing data. The division of the four datasets into training and testing sets followed, with the training set comprising 80% and the test set 20%. We employed all seven indicators as input in this study.

This research employed adaptive gradient boosting regression, Linear Regression, our model, and extreme gradient boosting regression on four datasets (GBP_USD, EUR_USD, NASDAQ, and Gold). The outcomes obtained offer an all-encompassing prediction and comparative analysis. Tables 1 through 4 illustrate the performance of these models.

TABLE I.
DISPLAYING PERFORMANCE FOR GOLD

Model	Squared Error of the Root Mean	Average Absolute Error	Average Absolute Percentage Error	R^2
Extreme gradient boosting	69.4333	40.6500	0.0218	0.8508
Adaptive gradient boosting	56.0504	37.8887	0.0209	0.9028
Linear regression	8.7729	5.9208	0.0034	0.9976
Our model	8.7768	5.9968	0.0034	0.9976

TABLE II.
DISPLAYING PERFORMANCE FOR GBP_USD

Model	Squared Error of the Root Mean	Average Absolute Error	Average Absolute Percentag e Error	R^2
Extreme gradient boosting	0.0226	0.0113	0.0093	0.9038
Adaptive gradient boosting	0.0294	0.0159	0.0131	0.8368
Linear regression	0.0058	0.0041	0.0032	0.9935
Our model	0.0056	0.0040	0.0031	0.9940

TABLE III.
DISPLAYING PERFORMANCE FOR NASDAQ

Model	Squared Error of the Root Mean	Average Absolute Error	Average Absolute Percentage Error	R^2
Extreme gradient boosting	4224.6189	3432.2672	0.2658	-1.8439
Adaptive gradient boosting	4400.9686	3637.3728	0.2843	-2.0863
Linear regression	108.4760	74.6958	0.0068	0.9981
Our model	118.9768	83.1882	0.0075	0.9977

TABLE IV.
DISPLAYING PERFORMANCE FOR EUR_USD

Model	Squared Error of the Root Mean	Average Absolute Error	Average Absolute Percentage Error	R^2
Extreme gradient boosting	0.0179	0.0083	0.0080	0.9165
Adaptive gradient boosting	0.0288	0.0169	0.0160	0.7859
Linear regression	0.0290	0.0181	0.0169	0.7829
Our model	0.0077	0.0060	0.0053	0.9844

The outcomes depicted in the four tables showcase a comparative evaluation of the performance of distinct machine learning methods. Notably, the ensemble model distinguishes itself by attaining lower values for the Squared Error of the Root Mean, MAE, and MAPE, along with higher values for R^2. This superiority of the ensemble model comparison to methods validates the effectiveness of integrating adaptive gradient-boosting regression as a foundational estimator in conjunction with linear regression.

The comparison between the foreseen price by the ensemble model and the actual is in Figures 3, 4, 5, and 6.

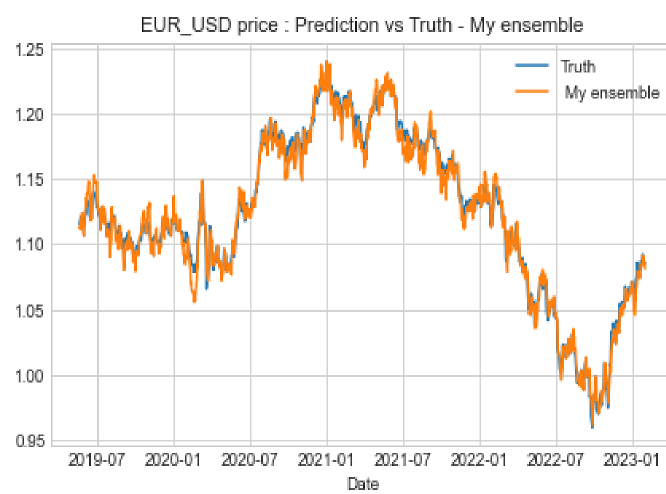

Fig. 3. Trends prices for EUR_USD

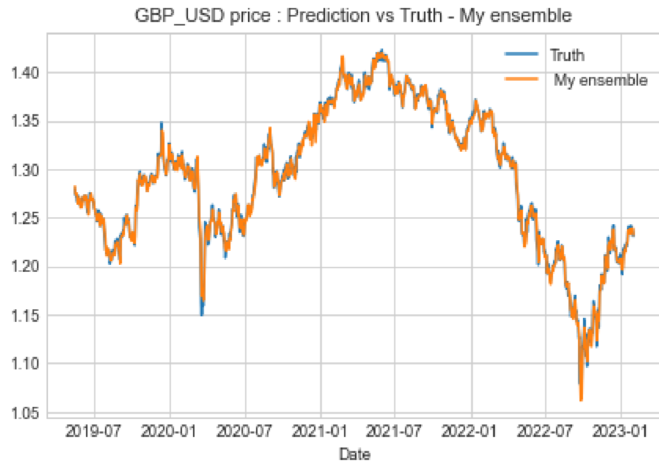

Fig. 4. Trends prices for GBP_USD

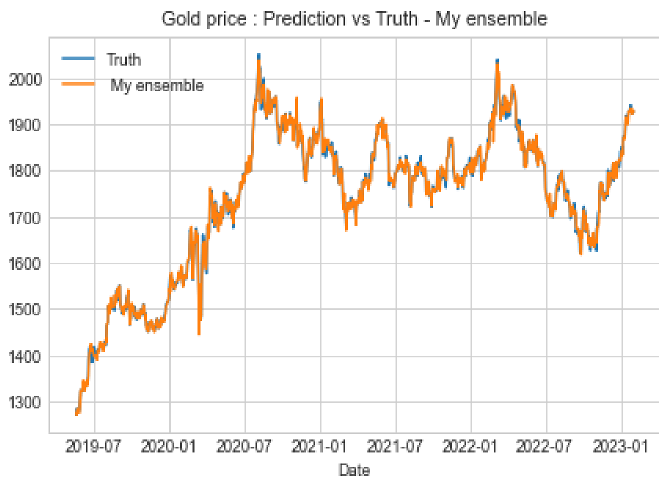

Fig. 5. Trends prices for gold

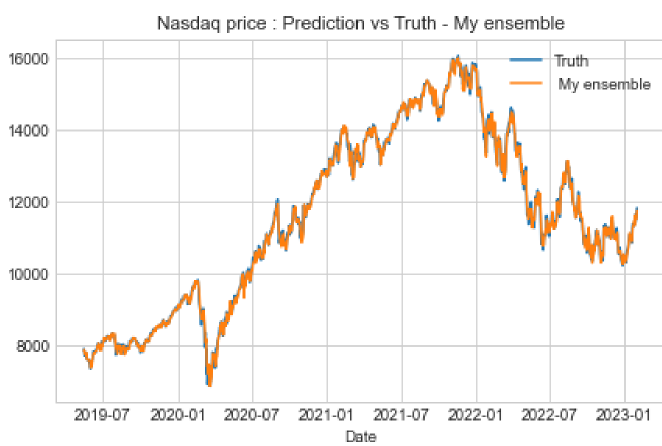

Fig. 6. Trends prices for NASDAQ

The utilization of the ensemble model in the four figures is depicted, with the predicted values shown by the orange line and the actual trend represented by the blue line. The forecast closely corresponds to the genuine trends, underscoring the effectiveness of the ensemble model. The strong correlation between these two lines indicates the effectiveness of the model.

This research will enable investors to make well-informed choices regarding future closing prices for four datasets, enabling them to determine the optimal timing for entering or exiting the market.

VI. Conclusion

We have developed a groundbreaking machine learning algorithm utilizing regression to predict forthcoming closing prices across four datasets (EUR_USD, GBP_USD, NASDAQ, and Gold). Each dataset incorporates novel technical indicators. The training process involved utilizing both ensemble model and linear regression learning approaches on the four datasets. This paper employed four metrics— coefficient of determination, mean absolute percentage error, root mean squared error, and mean squared error—to compare different machine-learning models. In our upcoming research endeavors, we'll employ deep learning techniques on the dataset to explore various financial markets. Our goal is to improve prediction results by leveraging the same indicators.

VII. References

[1] D. Pradeepkumar and V. Ravi, "Forecasting financial time series volatility using particle swarm optimization trained quantile regression neural network," *Appl. Soft Comput.*, vol. 58, pp. 35–52, 2017.

[2] M. Khashei and Z. Hajirahimi, "Performance evaluation of series and parallel strategies for financial time series forecasting," *Financ. Innov.*, vol. 3, no. 1, pp. 1–24, 2017.

[3] L. Lei, "Wavelet neural network prediction method of stock price trend based on rough set attribute reduction," *Appl. Soft Comput.*, vol. 62, pp. 923–932, 2018.

[4] M. S. Aziz Sarkar and U. A. M. Ehsan Ali, "EUR/USD Exchange Rate Prediction Using Machine Learning," *Int. J. Math. Sci. Comput.*, vol. 8, no. 1, pp. 44–48, 2022, doi: 10.5815/ijmsc.2022.01.05.

[5] C. F. Tsai, Y. C. Lin, D. C. Yen, and Y. M. Chen, "Predicting stock returns by classifier ensembles," *Appl. Soft Comput. J.*, vol. 11, no. 2, pp. 2452–2459, 2011, doi: 10.1016/j.asoc.2010.10.001.

[6] M. Fabbri and G. Moro, "Dow Jones Trading with Deep Learning: The Unreasonable Effectiveness of Recurrent Neural Networks," *DATA 2018 - Proc. 7th Int. Conf. Data Sci. Technol. Appl.*, no. Data, pp. 142–153, 2018, doi: 10.5220/0006922101420153.

[7] S. O. Ojo, P. A. Owolawi, M. Mphahlele, and J. A. Adisa, "Stock Market Behaviour Prediction using Stacked LSTM Networks∗," *Proc. - 2019 Int. Multidiscip. Inf. Technol. Eng. Conf. IMITEC 2019*, pp. 1–5, 2019, doi: 10.1109/IMITEC45504.2019.9015840.

[8] B. S. Lin, W. T. Chu, and C. M. Wang, "Application of Stock Analysis Using Deep Learning," *Proc. - 2018 7th Int. Congr. Adv. Appl. Informatics, IIAI-AAI 2018*, pp. 612–617, 2018, doi: 10.1109/IIAI-AAI.2018.00128.

[9] H. Chung and K. S. Shin, "Genetic algorithm-optimized long short-term memory network for stock market prediction," *Sustain.*, vol. 10, no. 10, 2018, doi: 10.3390/su10103765.

[10] B. Weng *et al.*, "Macroeconomic indicators alone can predict the monthly closing price of major U.S. indices: Insights from artificial intelligence, time-series analysis and hybrid models," *Appl. Soft Comput.*, vol. 71, no. October 2019, pp. 685–697, 2018, doi: 10.1016/j.asoc.2018.07.024.

[11] R. K. Dash, T. N. Nguyen, K. Cengiz, and A. Sharma, "Fine-tuned support vector regression model for stock predictions," *Neural Comput. Appl.*, vol. 35, no. 32, pp. 23295–23309, 2023, doi: 10.1007/s00521-021-05842-w.

[12] G. Rekha, B. D Sravanthi, S. Ramasubbareddy, and K. Govinda, "Prediction of stock market using neural network strategies," *J. Comput. Theor. Nanosci.*, vol. 16, no. 5–6, pp. 2333–2336, 2019, doi: 10.1166/jctn.2019.7895.

[13] P. Baser, J. R. Saini, and N. Baser, "INTELLIGENT SYSTEMS AND APPLICATIONS IN Gold Commodity Price Prediction Using Tree-

based Prediction Models," vol. 11, pp. 96–102, 2023.

[14] M. El Mahjouby and M. Elfar, "Time Series Forecasting Using Ensemble Model," in *2023 7th IEEE Congress on Information Science and Technology (CiSt)*, 2023, pp. 155–159. doi: 10.1109/CiSt56084.2023.10409946.

[15] R. E. Schapire, "Explaining adaboost," *Empir. Inference Festschrift Honor Vladimir N. Vapnik*, pp. 37–52, 2013, doi: 10.1007/978-3-642-41136-6_5.

[16] M. Tami and A. Y. Owda, "Efficient commodity price forecasting using long short-term memory model," vol. 14, no. 1, pp. 994–1004, 2024, doi: 10.11591/ijai.v13.i1.pp994-1004.

[17] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques," *Expert Syst. Appl.*, vol. 42, no. 1, pp. 259–268, 2015.

[18] K. Kumar, "Statistical Regression and Classification: from Linear Models to Machine Learning," *Journal of the Royal Statistical Society Series A: Statistics in Society*, vol. 181, no. 4. pp. 1263–1264, 2018. doi: 10.1111/rssa.12401.