International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(20s) 2026 pp. 814-828
https://doi.org/10.70917/ijcisim-2026-5220
Article

Ensemble Machine Learning Approach for Performance Evaluation of Mutual Funds (Index) with news-based Sentiment

Sapna Jain[1], Prof. Saurabh Mukherjee[2]

[1]PhD Scholar, Banasthali Vidyapith, Rajasthan, India
Email: rite2sapna@gmail.com
[2]Professor, Computer Science Department, Banasthali Vidyapith, Rajasthan, India
Email: msaurabh@banasthali.in

Abstract: This study proposes a hybrid machine learning framework for predicting equity index fund performance based on ensemble regression. It aims to amalgamate the forecasting accuracy by combining the machine learning algorithms viz., SVR, KNN and Random Forest Regression under a Voting Regressor model. The hybrid algorithm combines the strength of individual models to reduce the prediction errors and improve the generalizations. The models were evaluated using metrics such as R2, MSE, RMSE and MAPE The study showed that the hybrid Voting Regressor had higher prediction accuracy and stability on several index funds than the individual models. The ensemble method yielded higher R^2 values and lower forecast errors, suggesting its capacity to capture complex market dynamics and reduce bias. The research showed that hybrid approaches to machine learning can substantially improve index fund predictions, offering insights for investors and financial experts in turbulent market environments.
Keywords: Mutual Funds performance, Index Funds, price prediction, SVM, KNN, Random Forest, Majority Voting algorithm

1. Introduction

News sentiment analysis is a crucial task to interpret market dynamics and forecast price volatility. However, the creation of accurate financial news datasets, especially in terms of proper classification and sourcing, remains a major challenge. The equity and commodity markets are both subject to dynamics that are influenced by global and regional sentiment. Hence, the market's sentiment depends on major world events like changes in the monetary policy of the United States and the geopolitical situation. These transformations have a direct impact on market volatility and investor confidence [1].

The financial press reports that the numerous messages posted on the internet stock message boards have the potential to sway the market. An increase in the volume of messages on a daily basis is correlated with a subsequent rise in trading activity and volatility, with a pronounced surge in smaller-sized trades [2].

With the evolution of capital market, equity investments in particular have emerged as an especially attractive avenue for achieving superior returns. But there is a built-in risk factor in the uncertainty of which stocks to choose and what price changes will occur. These considerations have established mutual funds as an exemplary alternative for the allocation of savings, given their reduced risk profile and the complete investment responsibility entrusted to professional investment managers [3]. Consequently, the projection of the market value of both stocks and mutual funds has garnered significant interest among individuals participating in the stock market independently, particularly in light of the substantial fees charged by these investment advisors for their guidance [3].

The sentiment analysis helps to determine the customer inclination and whereby also affect the market trends. These trends often signify the drift in prices of stocks based upon certain major announcements and the big news. With the aid of machine learning algorithms by suitably training and testing the data this effect has been studied and significant price movements have been seen [4]. Machine Learning (ML) and Deep Learning (DL) are endowed with the capability to assume a crucial function within the financial industry. Through the application of diverse ML and DL algorithms, it becomes feasible to systematically gather and visualize extensive datasets pertaining to the stock market and manipulate this information skillfully to precisely anticipate stock market trajectories [5].

In recent years, financial economists have manifested a keen interest in the impact of news and events on financial assets. A considerable number of studies have been devoted to the investigation of the influence of different kinds of information, e.g. political events, macroeconomic indicators and earnings announcements, on the valuation of financial assets. Performance of diversified equity mutual funds is highly dependent on stock market conditions [6]. Moreover, the performance of mutual funds, as measured by changes in Net Asset Value (NAV), is correlated with news or events in the economic environment.

1.1 Index Funds

A stock market index or simply a stock market index helps people understand how prices are changing in the financial, commodity or other markets. Stock market indices are made to show how the whole equity market is doing. A stock market index is made up of a group of stocks that represent the market or a specific part of it. Each stock market index has its way of being calculated and is explained in terms of how much it has changed from a starting value.

A stock market index is a collection of stocks that are carefully chosen to track how a specific market, sector, commodity, bond or other type of investment is doing. The stock market index is a tool that investors and financial managers use to get a picture of the market and to compare how well different investments are doing. Stock market indices give people a sense of how the stock market has done in the past which helps investors make decisions about their money [9].

A stock market index is basically an idea so people cannot invest in it directly. However, many index funds try to do well as the stock market index. An index mutual fund is a kind of mutual fund that is made to match a specific stock market index. Index mutual funds take money from investors. Use it to buy the stocks that are, in the index trying to match the performance of the stock market index [9].

On a global scale, index funds have exhibited exceptional performance by providing entirely new opportunities for both retail and institutional investors.

Key Benefits of Index Mutual Funds:

Low Cost: The cost of index funds is really low compared to other ways of investing.

Simplicity: Managing index funds is pretty simple and easy to understand. Once the index funds metric is known it is easy to figure out what is going on. An investor can check on their index funds a few times a year like every six months or every year and that is manageable. Index mutual funds are simple to manage.

Turnovers: When securities are bought and sold then shares turnover is an important thing to think about. There are factors that affect how index mutual funds work and they all play a role in how the market works. Because index mutual funds are a way of investing, they do not buy and sell as much so turnovers are lower. This means index mutual funds have turnovers.

No style Drift: Index mutual funds are a way of investing. This approach can help an investor diversify the investments, which's a good thing for index mutual funds. Index mutual funds do not change their strategy so it can be counted upon to stay the same. This is good, for people who invest in index funds [10].

2. Literature Review

With the rapid development of online social media, large-scale tweets are spreading on platforms such as Twitter and Weibo. These tweets carry not only the factual information but also the emotional sentiments of their authors.

For years there has been debate over whether online social media platforms like Twitter and its Chinese equivalent Weibo can be used as a reliable proxy to predict stock market movements, especially for China. However, based on the conventional theories in behavioural finance, the emotional state of an investor can have a significant impact on the decision-making processes. Thus, it is advisable to further investigate this controversial issue by using the idea of online emotional expressions that are widely propagated through numerous tweets posted on social media [7].

The results suggest that the number of tweets per day is correlated with stock market indicators. Twitter data seems to be really useful for making predictions about the stock market. It can help to figure out what is going to happen with stock market indicators. Twitter is a tool for forecasting stock market trends. For example, when the stock market is seen Twitter data can be used to predict if the S&P 500 closing price will go up or down. This makes the predictions more accurate [6].

Accordingly, using a combination of machine learning algorithms is a good way to forecast mutual fund prices. Altogether three algorithms were used viz., Support Vector Regression, K-Nearest Neighbor Regression and Random Forest Regression. The predictions were taken from each of these models. Combined them using a Voting Regressor strategy. This makes the overall predictions more accurate and stable.

Using models together also known as ensemble learning is a good way to make predictions about financial things. This is because it helps reduce errors and makes the models work better in general. Twitter data and ensemble learning can be really helpful, for predicting stock market trends and mutual fund prices [7][8].

3. Mathematical Formulation

Let the temporal series of mutual fund prices be given as in Equation 1.

P={P1,P2,………,PT} Eq 1

where

Pt indicates the closing price of the mutual funds at time t

T is the aggregate number of observations.

The primary objective is to estimate the forthcoming price is given as Equation 2.

$\widehat{\mathrm{P}}_{\mathrm{t}} = \mathrm{f}(\mathrm{P}_{\mathrm{t}-1} - \mathrm{P}_{\mathrm{t}-2})$ Eq 2

where $\mathrm{P}_{\mathrm{t}-1}$ and $\mathrm{P}_{\mathrm{t}-2}$ represent the lagged price values employed as predictor variables[11].

3.1 Feature Construction

In order to encapsulate temporal dependencies inherent in the financial time series, lagged features are formulated as given in Equation 3.

$x_t = [\mathrm{P}_{\mathrm{t}-1}, \mathrm{P}_{\mathrm{t}-2}]$ Eq 3

Consequently, the dataset can be represented as given in equation 4.

D={(xt,Pt)}$_{t=3}^{T}$ Eq 4

Where $x_t \in R^2$ denotes the input feature vector and Pt signifies the target variable.

Feature engineering using lagged variables enables machine learning algorithms to capture temporal relationships and sequential dependencies in financial datasets [12].

I. Support Vector Regression Model

Support Vector Regression formulates a nonlinear regression function within a high-dimensional feature space. SVR was introduced by Vapnik in year 1995 [13][14] and later extended for regression applications by Cortes and Vapnik. [15].

The regression function is articulated as given in Equation 5.

$$f_{SVR}(x) = wT\phi(x) + b \qquad \text{Eq 5}$$

where

- w is the weight vector
- b is the bias term
- $\phi(x)$ is a nonlinear transformation.

SVR aims to resolve the optimization problem as per Equation 6:

$$\min_{w,b,\xi i,\xi i^*} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi i + \xi i^*) \qquad \text{Eq 6}$$

subject to the constraints, as per equation 8 and equation 9.

$$y_{i-(w^T\phi(x_i)} + b \leq \epsilon + \xi_i \quad \text{Eq 7}$$

$$(w^T\phi(x_i) - y_i \leq \epsilon + \xi_i^* \quad \text{Eq 8}$$

where

$$\xi_i, \xi_i^* \geq 0$$

Through the application of kernel functions, the prediction can be expressed as given in Equation 9.

$$f_{SVR}(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(x_i, x) + b \quad \text{Eq 9}$$

where $K(x_i, x)$ represents the kernel function. Kernel-based learning allows SVR to efficiently model nonlinear relationships present in mutual fund price movements and financial market data[16].

II. K-Nearest Neighbor Regression

KNN regression estimates the target value by calculating the average of the k nearest training samples within the feature space. KNN is a non-parametric instance-based learning algorithm that performs prediction based on similarity measures among observations [17].

Let $N_k(x_t)$denote the set of k nearest neighbors of x_t. Then the predicted price is given as equation 10.

$$\widehat{\mathrm{P}}_{\mathrm{t}}^{KNN} = \frac{1}{k}\sum_{i\in N_{k(x_t)}R^2} P_i \quad \text{Eq 10}$$

The distance between two feature vectors is determined through the computation of Euclidean distance [18] given in equation 11.

$$d(x_i, x_j) = \sqrt{\sum_{m=1}^{2}(x_{im} - x_{jm}\ )^2} \qquad \text{Eq 11}$$

KNN regression effectively captures local data structures and short-term fluctuations in financial time-series datasets [19].

III. Random Forest Regression

Random Forest constitutes an ensemble learning paradigm that amalgamates multiple decision trees. Each tree Tj(x) is cultivated on a bootstrapped subset of the data [7].

The prediction derived from Random Forest is acquired as given in equation 12.

$$\widehat{P}_t^{RF} = \frac{1}{M}\sum_{j=1}^{M} T_j(x_t) \quad \text{Eq 12}$$

where,

M denotes the total number of decision trees. Each tree segments the feature space through recursive binary divisions, and the prediction at a leaf node is delineated as given in equation 13.

$$T_j(x) = \frac{1}{|L|}\sum_{i \in L} P_i \quad \text{Eq 13}$$

where L constitutes the set of samples that belong to that particular leaf node.

Random Forest regression is highly effective in modeling nonlinear interactions, noisy financial data, and complex variable relationships [7].

To improve prediction robustness, the outputs of the individual regression models are combined using an ensemble Voting Regressor.

IV.Ensemble Voting Regressor

To improve prediction robustness and forecasting stability, the outputs of the individual regression models are combined using an ensemble Voting Regressor. Ensemble methods combine multiple learners to achieve better predictive performance than individual models [20].

Let the predictions of individual models be given as Equation 14, Equation 15, Equation 16

Let the predictions of individual models be:

$$f_1(x_t) = f_{SVR}(x_t) \quad \text{Eq 14}$$

$$f_2(x_t) = f_{KNN}(x_t) \quad \text{Eq 15}$$

$$f_3(x_t) = f_{RF}(x_t) \quad \text{Eq 16}$$

The ensemble prediction is calculated as the average of all model predictions as per equation14, equation 15 and equation 16.

$$\widehat{P}_t^{VR} = \frac{1}{3}[f_1(x_t) + f_2(x_t) + f_3(x_t)] \quad \text{Eq 17}$$

In the weighted voting scheme, model contributions can be adjusted using weights w_i given in equation 18.

$$\boldsymbol{\widehat{P}_t^{VR} = w_1 f_{SVR}(x_t) + w_2 f_{KNN}(x_t) + w_3 f_{RF}(x_t)} \quad \textbf{Eq 18}$$

subject to

$$w_1 + w_2 + w_3 = 1$$

The proposed ensemble framework integrates the nonlinear learning capability of SVR, the local approximation property of KNN, and the robustness of Random Forest to improve forecasting accuracy and prediction stability in mutual fund price prediction.

A. Algorithm of the Proposed Hybrid Model

Step 1: Collect historical mutual fund price data.

Step 2: Generate lagged features $P_{t-1}\ and\ P_{t-2}$

Step 3: Normalize features using standard scaling.

Step 4: Split dataset into training and testing sets.

Step 5: Train individual regression models:

SVR

KNN

Random Forest.

Step 6: Generate predictions from each model.

Step 7: Combine predictions using Voting Regressor.

Step 8: Evaluate model performance using MSE, R2, RMSE and MAPE.

Step 9: Forecast future mutual fund prices.

B. Flowchart of the Proposed Ensemble Model

To understand the proposed hybrid algorithm working the flow chart is given in Figure 1:

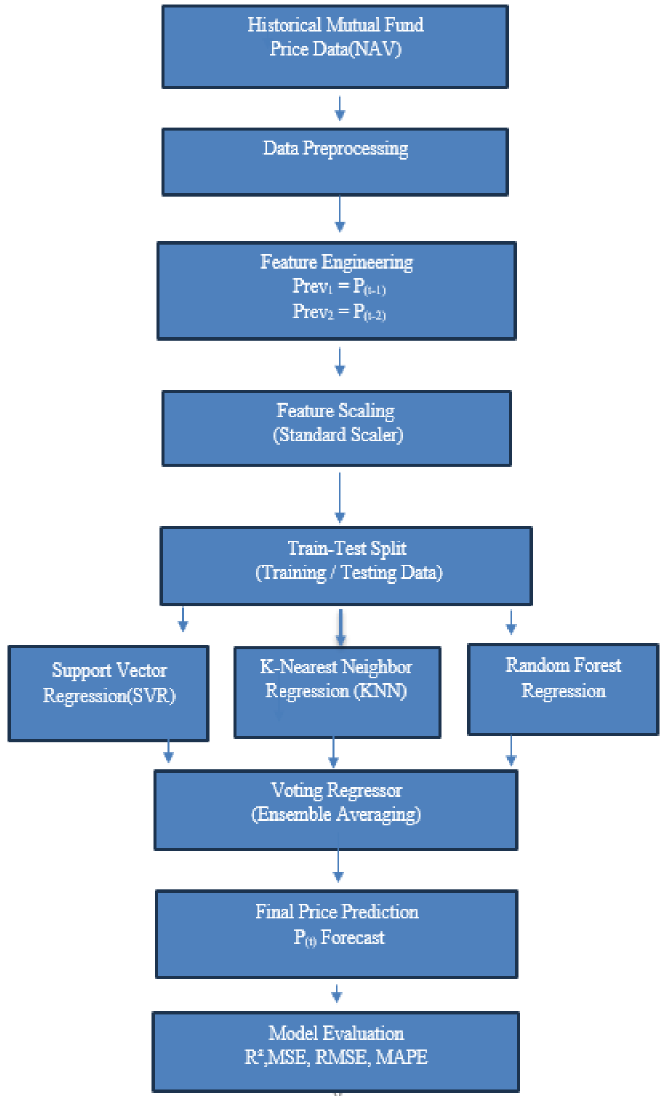

Figure 1: Flowchart of proposed Hybrid Algorithm

The proposed framework first preprocesses the mutual fund price data and constructs lagged features. Three machine learning regression models are trained independently. Their predictions are then aggregated using a Voting Regressor ensemble to produce the final forecast.

4. Data Description

The dataset taken for analysis is the listed in the index equity funds in India available on www.investing.com.Nearly, seven-year (2nd July 2018 to 31st Mar 2025) data is taken for analysis. The funds selected are given as below as below in Table 1.

Table 1: List of funds

S.no	Code Name	Name of Fund	Asset Class
1	0P0000OPMN	ICICI Prudential Nifty Next 50 Index Fund Growth	Equity
2	0P0000OPMO	ICICI Prudential Nifty Next 50 Index Fund Payout of Income Dist cum Cap Wdrl	Equity

3	0P0000XUYR	ICICI Prudential Nifty Next 50 Index Fund Direct Plan Payout of Income Dist cum Cap Wdrl	Equity
4	0P0000XUYS	ICICI Prudential Nifty Next 50 Index Fund Direct Plan Growth	Equity
5	0P0000XUZX	LIC MF Nifty Next 50 Index Fund Direct Payout of Income Distribution cum Cap Wdr	Equity
6	0P0000XV00	LIC MF Nifty Next 50 Index Fund Direct Growth	Equity
7	0P0000PUUH	LIC MF Nifty Next 50 Index Payout of Income Distribution cum Cap Wdrl	Equity
8	0P0000PUUG	LIC MF Nifty Next 50 Index Growth	Equity

A. Data Preprocessing

The dataset obtained from www.investing.com necessitates preprocessing to render it suitable for our analytical models and to enhance the precision of the results; consequently, we implement the following methodologies.

The information extracted from the historical dataset undergoes examination for the presence of absent or null values. Data visualization is performed through graphical representations using the Matplotlib library. This visualization elucidates the characteristics of the dataset, indicating any missing or null entries. The records that contain absent values are either removed from the dataset. Linear Regression is employed to estimate the missing and null values. Subsequently, this refined data is archived in a distinct data frame.

Test-Train Split – A total of 1650 rows were evaluated for each fund. The acquired dataset was partitioned into two segments, specifically the Test dataset, which constitutes 80% of the total data, allocated for the purpose of training the model, while the remaining 20% is designated as Test data, representing unseen data utilized to evaluate the performance of the trained model.

B. Performance Metrics:

Four metrics were mainly used to evaluate the performance of the funds using the various models. These are given as below:

1.Mean Squared Error (MSE) - The Mean Squared Error (MSE) represents the arithmetic mean of the squared discrepancies between the actual and predicted values [21]. A lower MSE signifies a model of superior performance, while an MSE value of zero indicates flawless predictions without any errors, an occurrence that is exceedingly rare yet theoretically possible. It is calculated using the formula in equation 19:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \qquad \text{Eq19}$$

2.Coefficient of determination(R^2) - The coefficient of determination, denoted as R^2, quantifies the degree to which the model has effectively fitted the data. It represents the proportion of variance in the target variable that can be elucidated by the independent variables. R^2 values are constrained within the interval of 0 to 1, where a value of 1 denotes a perfect model fit, and values approaching 1 are deemed desirable [21]. It is calculated using the formula in equation 20.

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(Xi - Yi)^2}{\sum_{i=1}^{m}(\bar{Y} - Yi)^2} \qquad \text{Eq20}$$

3.Root Mean Square Error (RMSE)- RMSE quantifies the magnitude of prediction errors, attributing greater significance to larger errors due to the squaring operation. It is particularly beneficial when substantial errors are undesirable [21]. It is calculated using equation 21.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_l)^2} \quad \text{Eq21}$$

RMSE quantifies the magnitude of prediction errors, attributing greater significance to larger errors due to the squaring operation. It is particularly beneficial when substantial errors are undesirable.

4.Mean Absolute Percentage Error (MAPE): MAPE assesses prediction errors as a percentage, rendering it scale-independent [21]. It is calculated using the Equation 22.

$$\text{MAPE}= \frac{1}{n}\sqrt{\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_l}{y_i}\right|} \times 100 \quad \text{Eq22}$$

Practical Applications:

The choice of statistical metrics depends on the specific goals of the predictive model. For instance:

▪ In regression problems-RMSE are commonly used to evaluate the magnitude of prediction errors.

▪ In classification tasks, accuracy, is critical for assessing the reliability of the predictions.

Combining multiple metrics often provides a comprehensive view of model performance. For example, RMSE gives insight into large errors. Similarly, precision perspectives on classification performance delivers a balanced evaluation.

Statistical metrics are indispensable for assessing predictive model. By selecting appropriate metrics tailored to the problem, analysts and data scientists can ensure robust evaluations and derive actionable insights to refine their models.

5. Results and Discussions

The proposed ensemble model, comprises Support Vector Regression (SVR), K-Nearest Neighbor Regression (KNN), Random Forest Regression (RF), and Voting Regressor (VR), was subjected to rigorous evaluation on various mutual fund datasets. The performance of the models was carefully checked using four evaluation metrics, specifically the Coefficient of Determination Mean Squared Error, Root Mean Squared Error and Mean Absolute Percentage Error.

Table 2: Results Sheet

S.No	**Name of fund**	**SVR**				**KNN**				**Random Forest**				**Voting Regressor (Ensemble)**			
		R^2	**MSE**	**RMSE**	**MAPE (%)**	**R^2**	**MSE**	**RMSE**	**MAPE (%)**	**R^2**	**MSE**	**RMSE**	**MAPE (%)**	**R^2**	**MSE**	**RMSE**	**MAPE (%)**
1	0P0000OPMN	0.9981	0.299	0.5468	1.04%	0.9984	0.2431	0.4931	0.93%	0.9985	0.2289	0.4784	0.92%	0.9986	0.2194	0.4684	0.89%
2	0P0000OPMO	0.9981	0.299	0.5468	1.04%	0.9984	0.2425	0.4925	0.93%	0.9985	0.2266	0.4761	0.91%	0.9986	0.2182	0.4671	0.89%
3	0P0000XUYR	0.9981	0.33638	0.57998	1.05%	0.9985	0.26496	0.51474	0.92%	0.9986	0.24854	0.49854	0.91%	0.9986	0.24017	0.49007	0.88%

S.No	Name of fund	SVR				KNN				Random Forest				Voting Regressor (Ensemble)			
4	0P0000XUYS	0.9981	0.33627	0.57989	1.05%	0.9985	0.26629	0.51603	0.92%	0.9986	0.25086	0.50086	0.91%	0.9986	0.24104	0.49096	0.88%
5	0P0000XUZX	0.9974	0.3503	0.5918	1.24%	0.9983	0.2244	0.4737	1.03%	0.9983	0.2288	0.4783	1.04%	0.9984	0.2123	0.4608	1.01%
6	0P0000XV00	0.9974	0.34441	0.58686	1.20%	0.9983	0.22788	0.47737	0.98%	0.9982	0.23809	0.48795	1.02%	0.9984	0.2163	0.46509	0.96%
7	0P0000PUUH	0.9972	0.2998	0.5475	1.21%	0.9983	0.1833	0.4281	1.01%	0.9984	0.1756	0.419	1.02%	0.9983	0.1854	0.4306	1.01%
8	0P0000PUUG	0.9967	0.3577	0.5981	1.31%	0.9978	0.242	0.492	1.07%	0.9975	0.2689	0.5185	1.13%	0.9976	0.2538	0.5038	1.09%

The models were checked using the Coefficient of Determination and the other metrics. High Coefficient of Determination values. Low Mean Squared Error, Root Mean Squared Error and Mean Absolute Percentage Error values show that the models are performing well [12] as given in Table 2.

The empirical results indicate that the proposed Voting Regressor ensemble persistently outperforms the individual regression models in terms of predictive accuracy across the majority of mutual fund datasets.

The results show that the Voting Regressor ensemble is doing better than the regression models in terms of predictive accuracy across most of the mutual fund datasets. Among the models Random Forest is usually doing better than SVR and KNN because it is good at modeling nonlinear relationships and handling noisy financial data.[22]. KNN is also doing well because it is good at approximation within time-series forecasting. The Voting Regressor is making forecasting by combining the advantages of all three models, which reduces prediction variance and makes the model more stable through ensemble averaging [20].

Across all eight mutual fund datasets:

* The Voting Regressor got the co-highest Coefficient of Determination values.

* The Voting Regressor consistently got the lowest Root Mean Squared Error values.

* The Voting Regressor got the minimum Mean Absolute Percentage Error values across most of the datasets.

The forecasting accuracy of the models was above 98 % that even surpassed 99 % for some ensemble predictions. These results show that ensemble learning is really helping to improve mutual fund forecasting performance.

The individual models like Random Forest and KNN are all part of the models. The Voting Regressor and the individual models are doing well in terms of accuracy. The Voting Regressor is doing well because it is combining the advantages of the models. The individual models like Random Forest and KNN are also doing well because they are good, at handling financial data and doing local approximation within time-series forecasting.

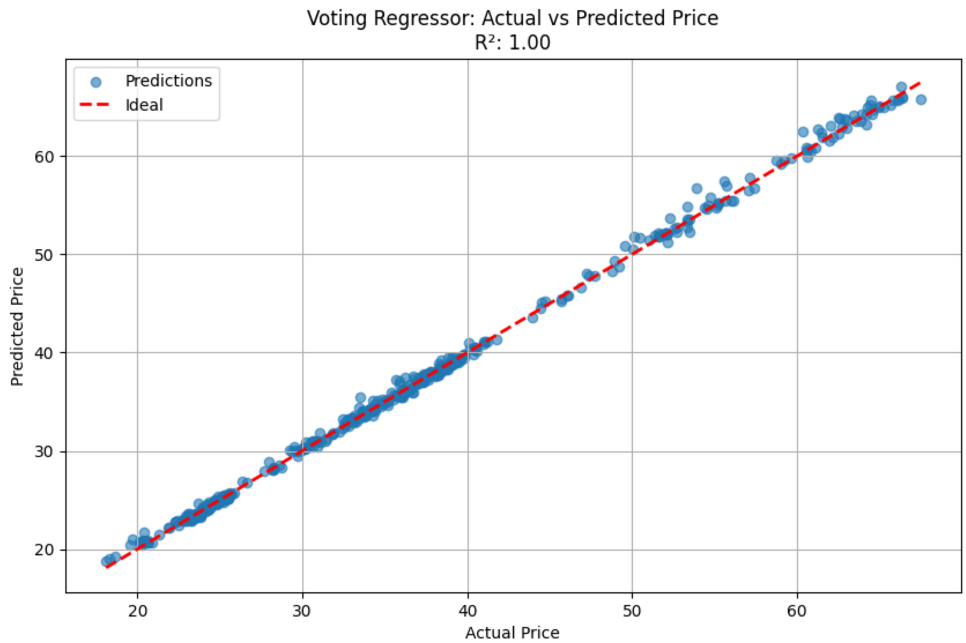

Figure 2: Prediction for 0P0000OPMN fund

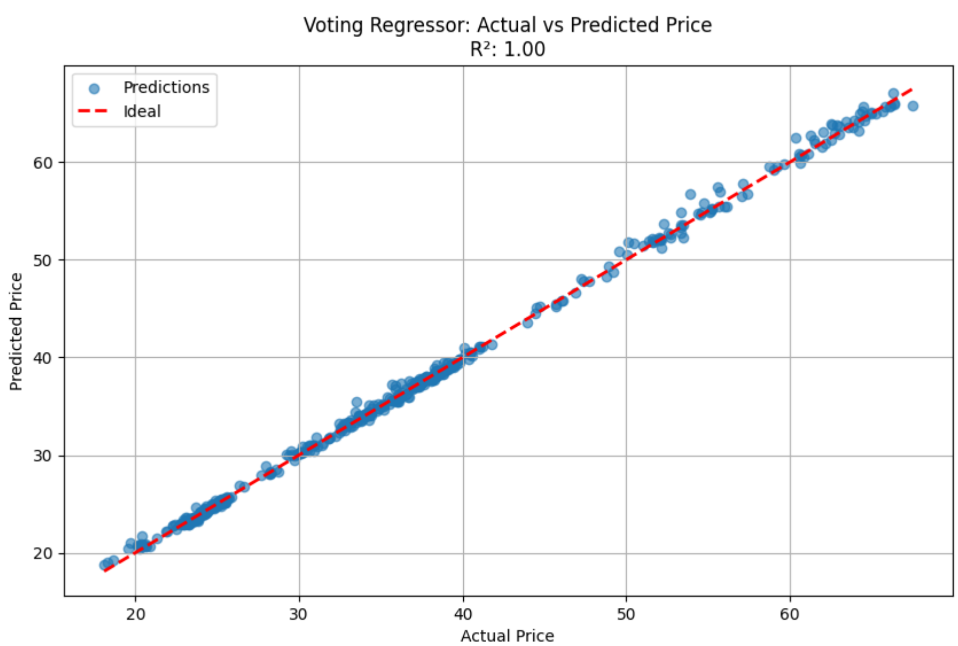

Figure 3: Prediction for 0P0000OPMO fund

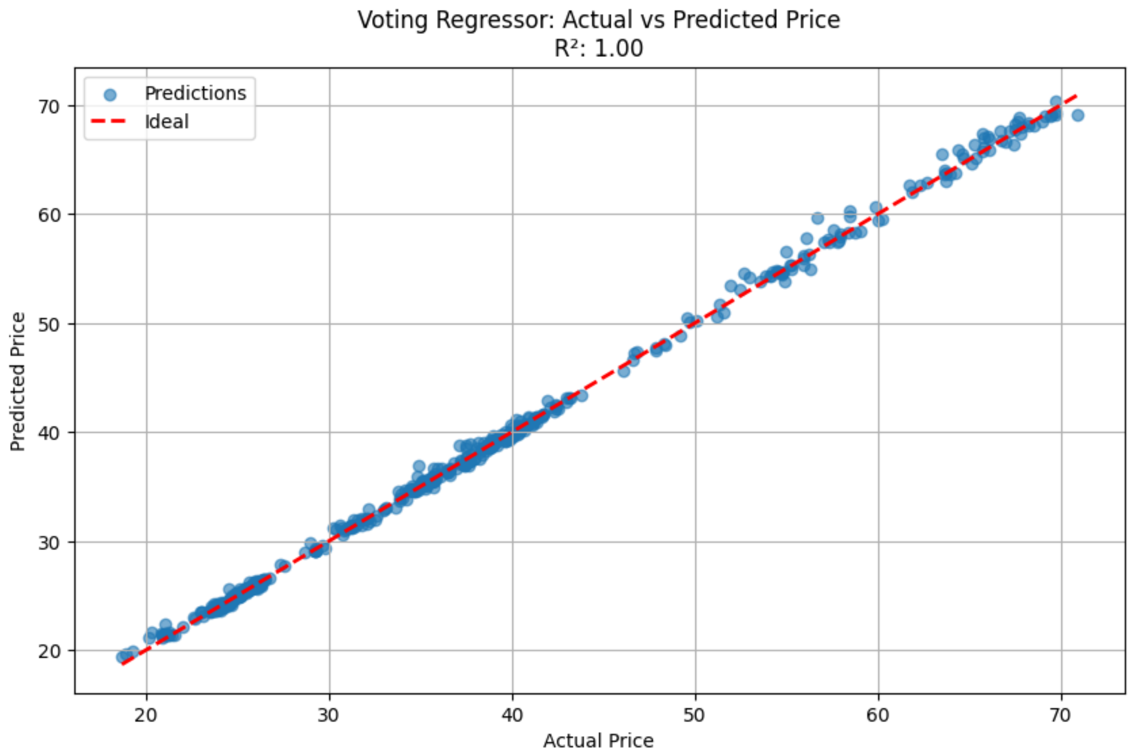

Figure 4: Prediction for 0P0000XUYR fund

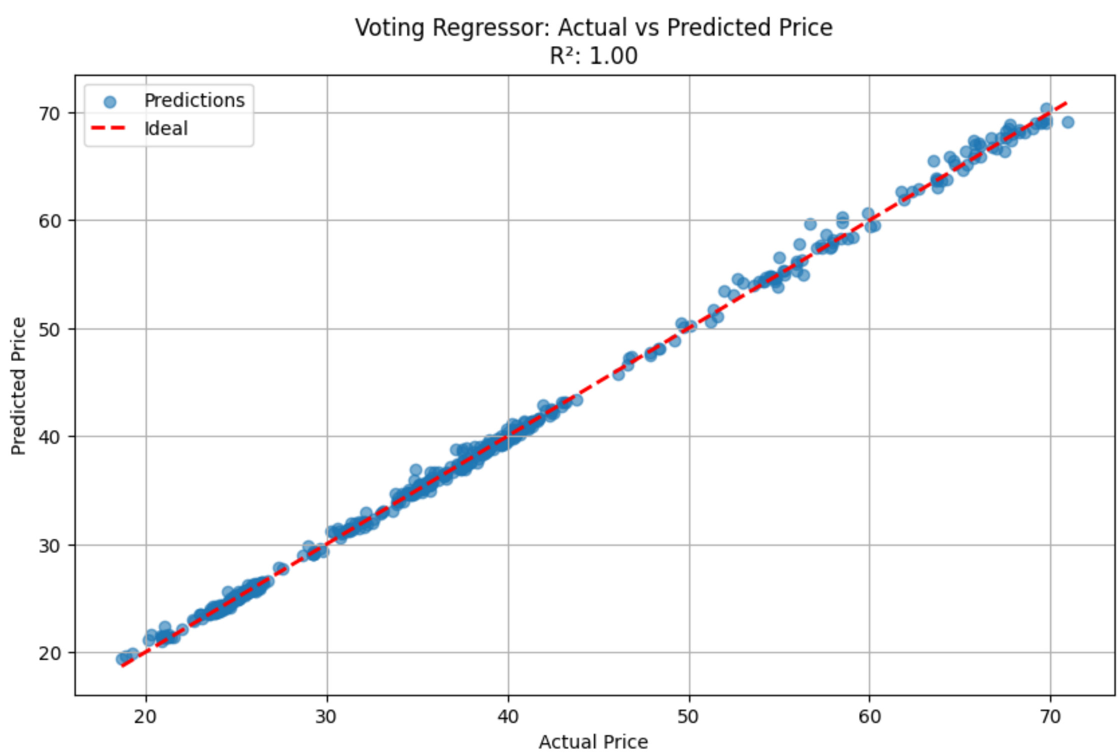

Figure 5: Prediction for 0P0000XUYS fund

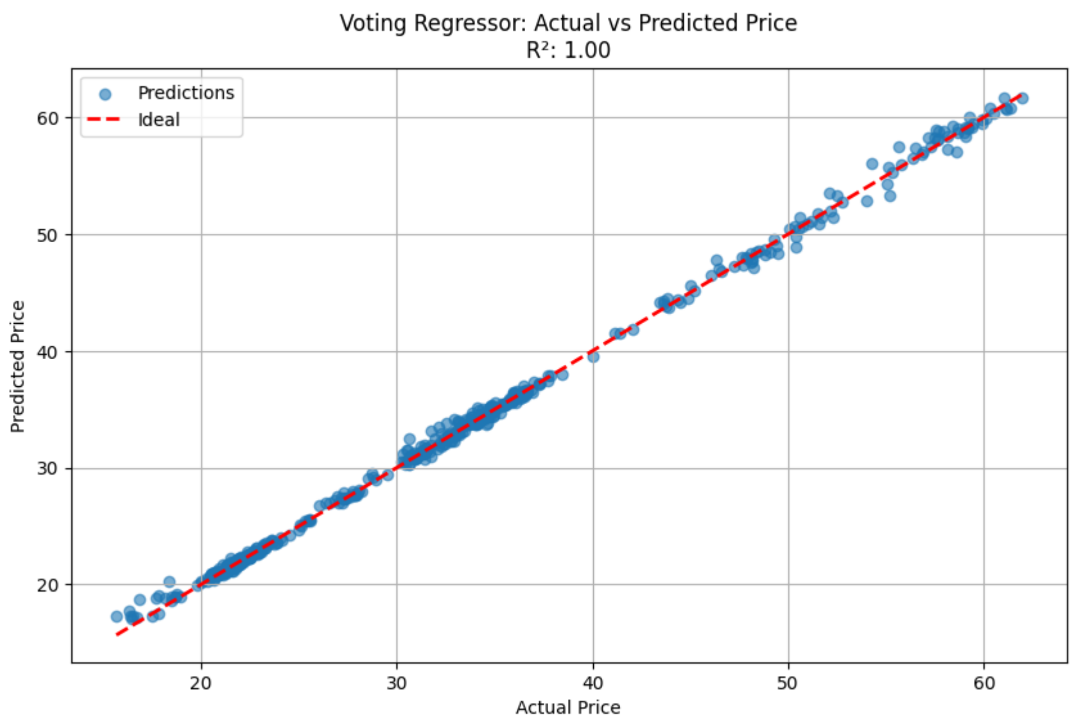

Figure 6: Prediction for 0P0000XUZX fund

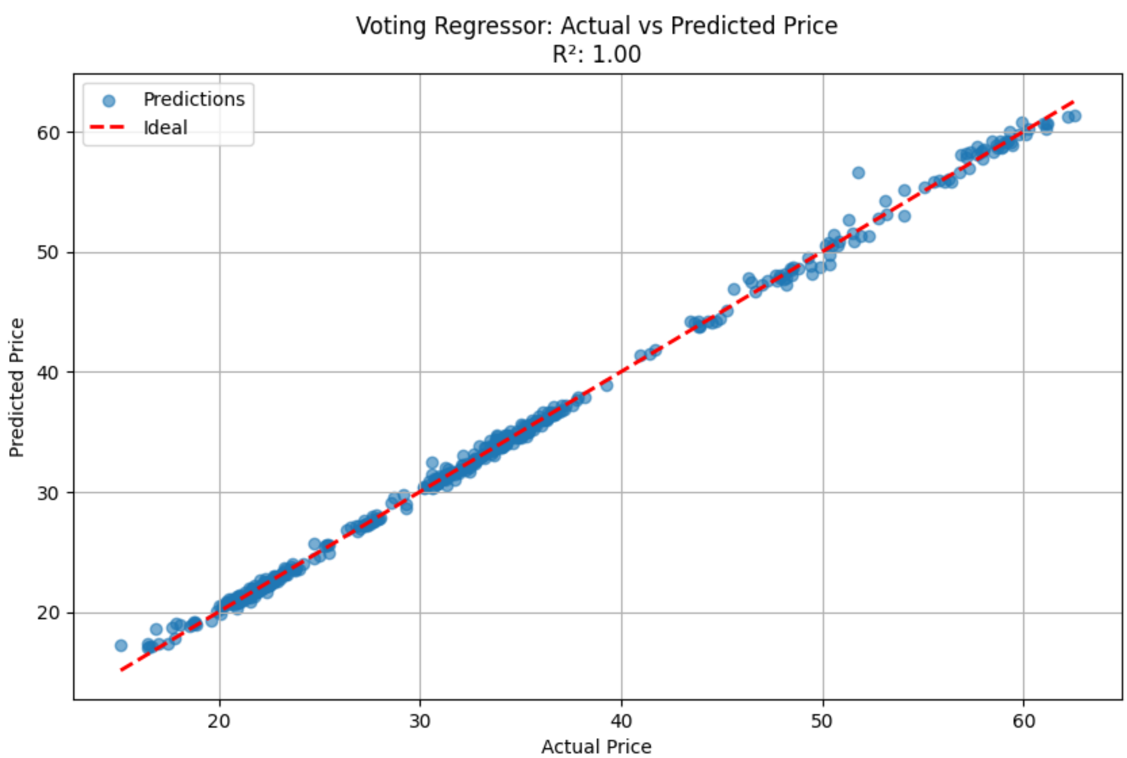

Figure 7: Prediction for 0P0000XV00 fund

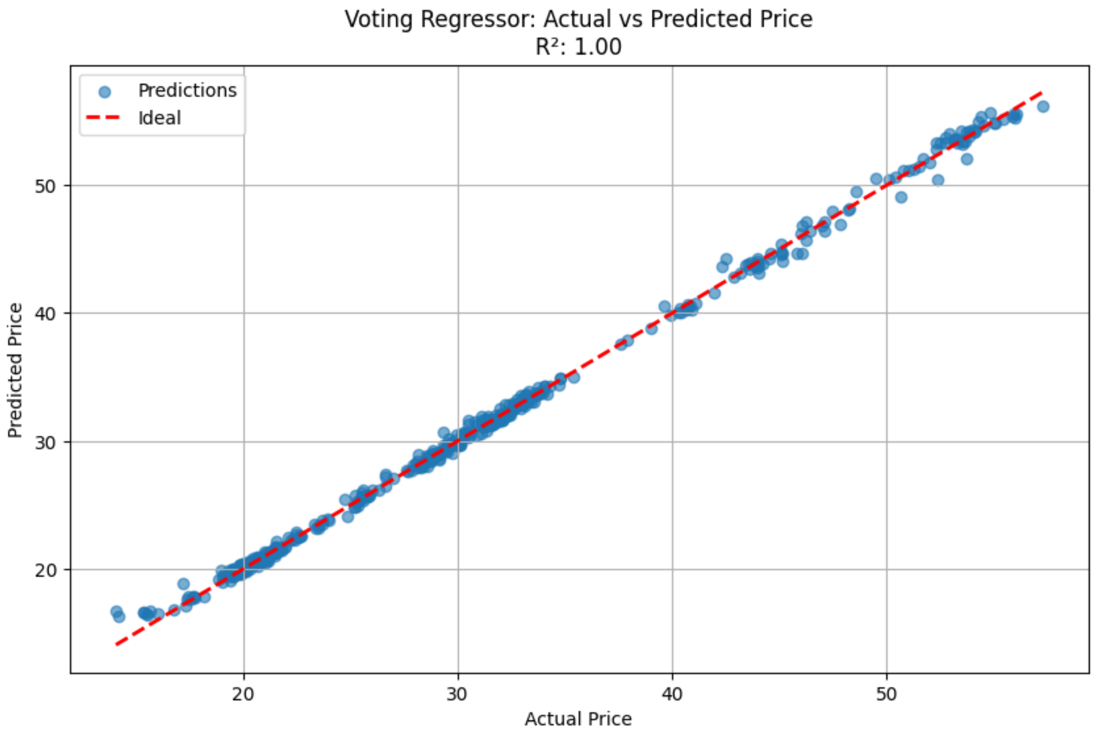

Figure 8: Prediction for 0P0000PUUH fund

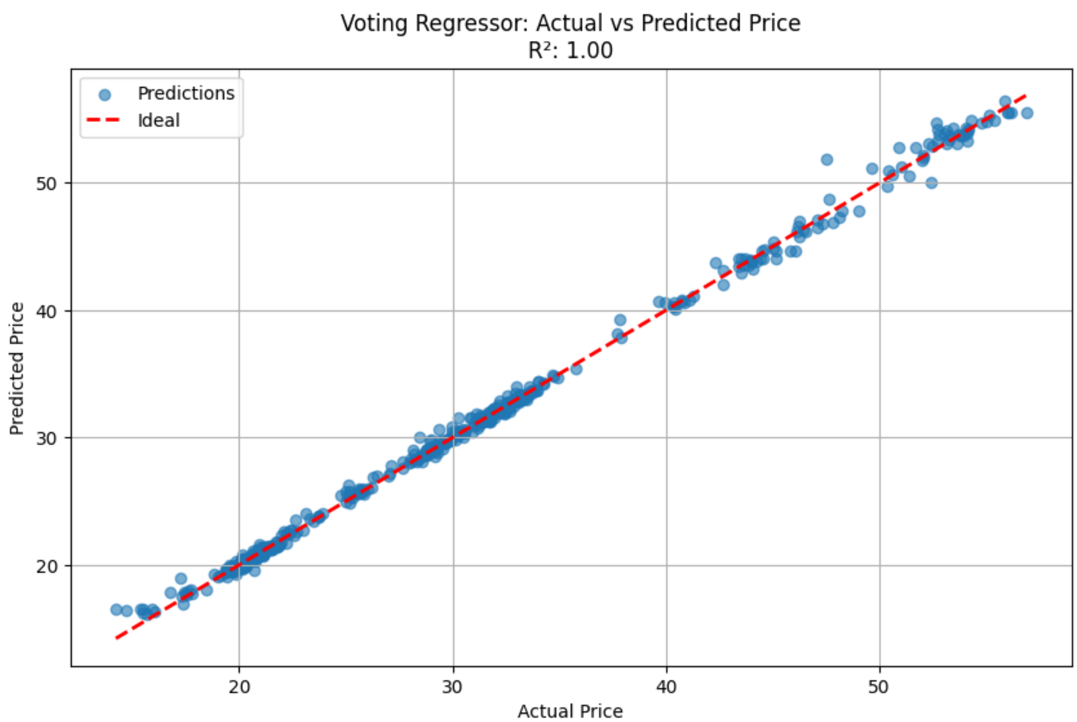

Figure 9: Prediction for 0P0000PUUG fund

As per Figure 2, Figure 3, Figure 4, Figure 5, Figure 6, Figure 7, Figure 8 and Figure 9 clearly indicate that Voting Regressor algorithm shows very high accuracy and small prediction errors. Most values align with the ideal line and show only negligible deviation.

The algorithm was implemented in the category of Index funds and it suitably performed good as per values obtained in Table 2. These results show that ensemble learning is really helping to improve mutual fund forecasting performance.

6. Conclusions and Future Works

The experimental analysis shows that the proposed hybrid ensemble framework really works well to improve the accuracy of mutual fund price predictions. The framework combines SVR, KNN and Random Forest using Voting Regression so it uses learning, local approximation and ensemble tree-based modeling all at the same time. This means the hybrid ensemble framework and the mutual fund price predictions get results.

The results obtained validate that the Voting Regressor model offers-enhanced prediction accuracy, reduced forecasting error, increased robustness and superior generalization capability. Consequently, the proposed ensemble framework has the potential to function as a highly effective intelligent forecasting model for financial time-series prediction and mutual fund investment analysis.

This study can be replicated to other categories of the mutual funds across the varied timelines and can be generalized to predict the funds prices effectively.

REFERENCES

1. A. Ehsan and S. Habib, “Financial News Sentiment Analysis Using NLP and Machine Learning for Asset Price Prediction : A Systematic Review,” pp. 279–308, 2025, doi: 10.21015/vtse.v13i3.2165.
2. W. Antweiler and M. Z. Frank, “Is All That Talk Just Noise ? The Information Content of Internet Stock Message Boards Is All That Talk Just Noise ? The Information Content of Internet Stock Message Boards,” no. March 2001, 2002.
3. A. Panigrahi, P. Karwa, and P. Joshi, “Impact of Macroeconomic Variables on the Performance of Mutual Funds: A Selective Study,” J. Econ. Policy Res., vol. 15, no. 1, pp. 29–43, 2020.
4. S. Jain1* and S. Mukherjee, “Effects of News on the Stock Prices using Machine Learning: A Comparative Study,” Bentham Books.
5. J. Guo and B. Tuckfield, “News-based Machine Learning and Deep Learning Methods for Stock Prediction,” J. Phys. Conf. Ser., vol. 1642, no. 1, 2020, doi: 10.1088/1742-6596/1642/1/012014.
6. B. Wang, “Correlating S & P 500 Stocks with Twitter Data,” 2012.
7. L. E. O. Breiman, “Bagging Predictors,” vol. 140, pp. 123–140, 1996.
8. H. Crc, Ensemble Methods.
9. P. Dhabolkar and Y. V Reddy, “Evaluating the Tracking Performance of Index Mutual Funds and Exchange Traded Funds in India,” 2019.
10. P. V. Deshapande, “‘ A study on performance evaluation of index mutual funds ,’” vol. 7, no. 9, pp. 168–182, 2022.
11. G. E. P. Box and G. M.Jenkins, Time Series Analysis.
12. T. Hastie, R. Tibshirani, and J. Friedman, “Springer Series in Statistics The Elements of”.
13. C. CORTES and V. VAPNIK, “Support-Vector Networks,” vol. 297, pp. 273–297, 1995.
14. “vladimir-vapnik-the-nature-of-statistical-learning-springer-2010 (1).pdf.”
15. L. J. Cao and F. E. H. Tay, “Financial Time Series Forecasting,” vol. 14, no. 6, pp. 1506–1518, 2003.
16. A. J. Smola and B. S. C. H. Olkopf, “A tutorial on support vector regression ∗ ¨,” pp. 199–222, 2004.
17. F. Glover, “Heuristics for Integer Programming Using Surrogate Constraints,” Decis. Sci., vol. 8, no. 1, pp. 156–166, 1977, doi: 10.1111/j.1540-5915.1977.tb01074.x.
18. J. Sapna and S. Anu, “MINIMIZATION OF INTRA-CLUSTERERROR CRITERION FUNCTION USING IMPROVED K-MEANS Jain Sapna , Sharma Anu,” vol. 2, no. 5, pp. 1568–1571, 2013.
19. N. S. Altman, “An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression,” vol. 46, no. 3, pp. 175–185, 2007.
20. T. G. Dietterich, “Ensemble Methods in Machine Learning,” 1990.
21. D. Chicco, M. J. Warrens, and G. Jurman, “The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation,” PeerJ Comput. Sci., vol. 7, pp. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.
22. L. E. O. Breiman, “Random Forests,” pp. 5–32, 2001.