NEXT-DAY TO NEAR-TERM CARDAMOM PRICE PREDICTION IN KERALA VIA MULTI-MODEL FUSION AND XGBOOST

ANOOP P. S., BIJU M.K. *

ABSTRACT: *Fluctuations in agricultural crop prices pose challenges for economic planning, making accurate forecasting essential for effective crop production and market strategies. This study focused on predicting future cardamom prices in Kerala using data from the Spices Board of Kerala (2014–2024). An ensemble model incorporating the XG Boost machine learning algorithm was proposed to enhance predictive accuracy. Analysis showed that daily average price and date are sufficient predictors for forecasting cardamom prices. The results demonstrated that the hybrid ensemble model, particularly with XG Boost, outperformed traditional methods, emphasizing the importance of tailored machine learning approaches for complex agricultural markets. The findings indicate a generally stable price structure for cardamom in Kerala, highlighting the need for strategic planning to mitigate potential risks and support farmers' livelihoods.*

KEY WORDS: *Cardamom pricing, Price forecasting, Ensemble model, Machine learning, XG boost.*

JEL CLASSIFICATIONS: *Q13, C53, Q11.*

1. INTRODUCTION

Kerala is the largest producer of cardamom in India. It is one of the most demanding export-valued spices in Kerala, spread over about 38882 hectares, predominantly in Idukki, Wayanad, Palakkad, and Pathanamthitta districts in Kerala. The productivity of cardamom has increased by 49% from 2014 to 2023 (Government of Kerala, Department of Economics and Statistics, 2023). Cardamom cultivation contributes significantly to the livelihood of farmers; however, the volatility of the

* *Ph.D. Student, School of Management and Business Studies, Mahatma Gandhi University Kottayam, anoopps888@gmail.com*
Assistant Professor, Dr., School of Management and Business Studies, Mahatma Gandhi University Kottayam, bijubijumk@gmail.com

prices of cardamom affects the economic stability of these farmers. They are suffering from financial hardships due to the stringent price fixation imposed by businessmen (Raman, 2022).

There was inefficiency in price forecasting for cardamom amidst adverse climatic events. The price forecasting of cardamom was found inaccurate due to problems such as drought-induced production loss, low-quality stock oversupply, and insufficient market demand (Vellaram, 2024). Previous literature also mentioned that cardamom prices are enduring wide price instability (Narayana et al., 1985; Varghese, 2004). Anticipating market dynamics and predicting future prices allows farmers and other stakeholders to optimise the situation of losses and uncertainties. Past literature on price forecasting of cardamom was employed SARIMA model (Thomas & Menon, 2022) and ARIMA model (Harini et al., 2018).

There are certain limitations to these techniques. The model's reliability depends upon forecaster's skill and experience. Other disadvantages are the subjectivity of its progress evaluation, and the existence of several limitations to the parameters and classes of possible models (Spyrou et al., 2022; Kontopoulou et al., 2023). Machine learning models are advantageous in terms of simplicity of workflow, quick and authentic prediction for a typical declining production curve (Kontopoulou et al., 2023).

The study by Menculini et al. (2021) compared ARIMA model with other machine learning models such as Prohet, LSTM and CNN for price forecasting. They found that the Prophet model is expedient to train and preprocessing of data is not required. Nonetheless, it does not approach the performance of alternative models. The usage of this model is suggested primarily when simplicity and rapidity of the prediction are considered as the primary objective of the study. Reviewing the literature, it is found that there is a lack of literature with regard to the prediction of cardamom prices using machine learning models. Most of the existing literature employed ARIMA and SARIMA models for predicting cardamom prices (Myneedi et al., 2023; Thomas & Menon, 2022). Naveenkumar et al. (2024) used for predictive modelling and yield production in Tamilnadu district and pointed out that machine learning methods are crop prediction dilemmas. The machine learning methods such as Support Vector Machine (SVM), Artificial Neural Networks (ANN), Random Forest (RF), K-Nearest Neighbors (KNN), and Decision Tree (DT) along with XG Boost for developing a hybrid model to predict cardamom prices.

The study focus more on XG Boost model for getting accurate result. Past literature indicated that the XG Boost tool is distinguished with high accuracy by lower cost and complexity and rapid processing times. It has higher ability in terms of learning speed especially when data normalization is applied and performance compared to other machine learning tools (Tissaoui et al., 2022; Chen & Guestrin, 2016). Therefore, the study use of ensemble model incorporating XGBoost (Extreme Gradient Boosting) model rather than univariate ARIMA and SARIMA models when precise forecast and brief extraction periods are necessary in a multivariate data set. An ensemble model is newly developed for this study to enhance prediction accuracy. The research highlights the potential socioeconomic benefits of predictive analytics for cardamom farmers, enabling better market preparedness and economic stability.

2. REVIEW OF LITERATURE

The price forecasting of agriculture products refers to the estimation and judging of agriculture product price changes over a period of time in the future by using some scientific methods based on historical data and current information (Sun et al., 2023). There have been considerable research achievements in the method of agricultural price forecasting after more than a hundred years of development, which includes both qualitative and quantitative methods (Wang et al., 2020).

The quantitative analysis methods are the main method analysis used for agriculture price forecasting in the contemporary period. They are primarily classified as regression analysis, time series analysis, machine learning models and combined models (Sun et al., 2023). Zhao (2020) emphasized that the machine learning models are pertinent for price forecasting. The present study initially uses machine learning techniques such as Support Vector Machine (SVM), Artificial Neural Network (ANN), Random Forest (RF), K-Nearest Neighbors (KNN) and Decision Tree (DT) for cardamom price forecasting and later employs a new robust ensemble model by integrating XG Boost model along with these initially employed models.

The regression variant of support vector machine (SVM) is being utilized in the scientific literature of time series forecasting. In the domain of time series forecasting the regression variant of Support Vector Machine (SVM). This model can handle complicated nonlineral relationship between past and future values (Kontopoulou et al., 2023). Artificial Neural Network is another machine learning method used for future predictions. Zou et al. (2007) mentioned in their study that the ability to model complex relationship without a priori assumptions of the nature of relationship is the the greatest advantage of ANN analysis. It performs a nonlinear functional mapping from the past observations performs a nonlinear functional mapping from the past observations (Xt 1, Xt 2,y,Xt p) to the future value Xt, i.e., Xt ¼fXt 1;Xt 2;...;Xt p;w ð Þþ et, where w is a vector of all parameters and f is a function determined by the network structure and connection weights. So, this analysis is similar to a nonlinear auto regreesive model.

The next model of price forecasting used in this article is Random Forest which was developed by Breiman (2001) as a way of gaining more precise predictions without overfitting the data. RF makes use of a randomised subset of predictors for each split of each tree (González et al., 2015). KNN algorithm is also a common machine learning method for price estimation. It summarises and synthesise information obtained by humans through computer-based programs. It is uncomplicated and intuitive and also suitable for different types of data structures (Lin et al., 2020).

Another important machine learning method incorporated in the present study is decision tree learning. It is usually used in data mining. This algorithm works by splitting the data set in order to train a model through a recursive partitioning process and predict the value of a determined variable based on the independent variables (Gordon et al., 1984). Along with these models, Extreme Gradient Boosting model, short-termed as XG Boost model was integrated to form a comprehensive ensemble model for obtaining more accurate future price discovery for cardamom.

XG-Boost Regression Algorithm is developed by Chen and Guestrin (2016), is an algorithm that incorporates the boosting model proposed by Friedman (2001). It is an ensembled machine learning algorithm that is similar to a random forest with slight differences (Jabeur et al., 2021; Sonkavde et al., 2023). It constitutes a combination of weak learners such as decision trees. It is considered a satisfactory stock forecasting prediction model as it functions on a sequential model that takes into account the gradient for each iteration so that the weights are refurbished for every iteration of the decision tree (Zhu and He 2022).

XG Boost is a scalable tree-boosting system that is broadly used by data scientists and provides cutting-edge results on various problems. Chen and Guestrin (2016) found that to build a scalable end-to-end system for tree boosting cache access patterns, compression of data and sharding are inevitable. These can be applied in other machine learning models. Combining these observations, XG boost becomes efficient in unravelling real-world scale issues using the minimum amount of resources. XG Boost model is widely used in stock price prediction (Somkunwar et al., 2024; Almaafi et al., 2023; Dezhkam & Manzuri, 2022) and crude oil price prediction (Simsek et al., 2024; Tissaoui et al., 2022).

Wang et al. (2022) used XG Boost model along with variational mode decomposition (VMD), and gradient boosting decision tree (GBDT) to predict stock index future prices. The study emphasised that the XG Boost model was used to perform nonlinear integration on the modelling outcomes to generate the final results. XG Boost model have many parameters. The setting of these parameters affects the accuracy of the model. It is very difficult to find an accurate parameter by artificially setting parameters.

Therefore, the study by Wu et al. (2022) used PSO (Particle swarm optimization) to optimise the main parameters of XG Boost for power price forecasting. Jabeur et al. (2021) underlined that XGBoost is the best model for gold price prediction when compared with other machine learning models such as linear regression, neural networks, random forest, and boost algorithm. XG Boost outperforms all other established benchmark models yielding the superior results.The study also incorporated SHAP (SHapley Additive exPlanation) approach to interpret the outputs of XGBoost for the gold price prediction.Vuong et al. (2021) applied XG Boost model in stock price forecasting as a feature-selection technique to draw out dominant features from high-dimensional time series data and eliminate reiterative attributes. Gumus and Kiran (2017) used XG Boost model for crude oil price forecasting and again proved it to be an efficient prediction model for both classification and regression tasks. And also exerted that component-wise gradient boosting fosters the attractiveness of boosting by adding the selection of automatic variables during the process of fitting. Y. Wang and Guo (2020).

The present study developed a hybrid ensemble model with Support Vector Machine (SVM), Artificial Neural Networks (ANN), Random Forest (RF), K-Nearest Neighbors (KNN), Decision Tree and XG Boost to predict cardamom prices by using average daily prices.

3. METHODOLOGY

The data for the analysis were collected from Trade Information Department of Spices Board of Kerala from 2014 - 2024. Due to the unpredictability and fluctuations, using daily average price instead of monthly or annual data to obtain regular and short-term swings and extracts relevant insights within the data.

Five machine learning models such SVM, ANN, KNN, RF, and DT were employed initially to predict the price. Ensemble methods were tested for improved accuracy. Evaluation metrics included Mean Squared Error (MSE), R^2, and Mean Absolute Error (MAE). A training and test split of 70%-30% was used for all models.

4. RESULTS

4.1 Price analysis

The graph depicts the fluctuations in average price of cardamom over the years from 2014 to 2024.The prices of cardamom remain relatively stable with a gradual increase from 2014 to 2017.Then a sharp spike occurred in the year 2019 and reached the highest value of 2500 rupees. A notable decline followed with prices dipping significantly by 2021. After the year 2021 the prices started to recover and showed a steady upward trend towards the year 2024.

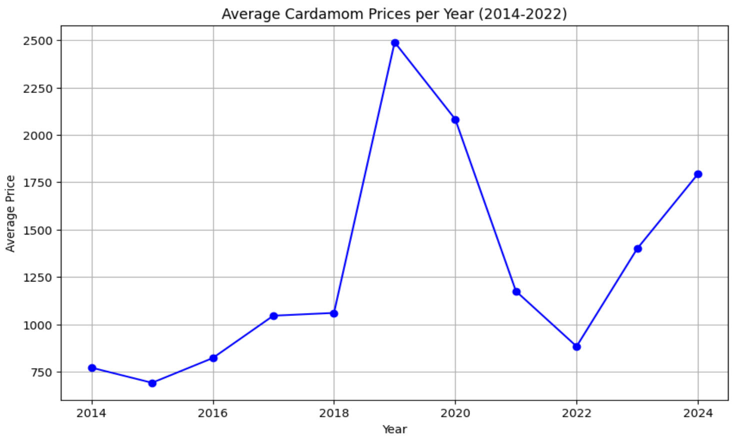

Figure 1. Trend Analysis of the daily average price in rupees of cardamom from 2014 to 2024

The Figure 2 diagram shows the average cardamom prices which depicts significant fluctuations. There is an increased level in prices the years 2019-2020 compared to other years. After 2020 the prices show some stabilization though they are significantly higher compared to earlier years.

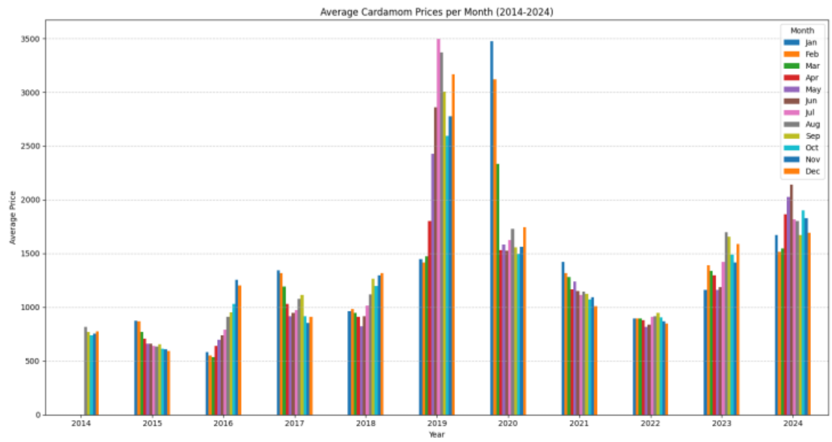

Figure 2. Average Cardamom Prices per Month (2014-2024)

The Figure 3 diagram shows the prices during the monsoon period such as June, July and August during the years taken, highlighting seasonal price trends. Prices show a similar pattern but have some peaks obtained during 2019 and 2020. After maintaining steadiness the price trend showed a rising tendency during the year 2023-2024 consistently over the three months.

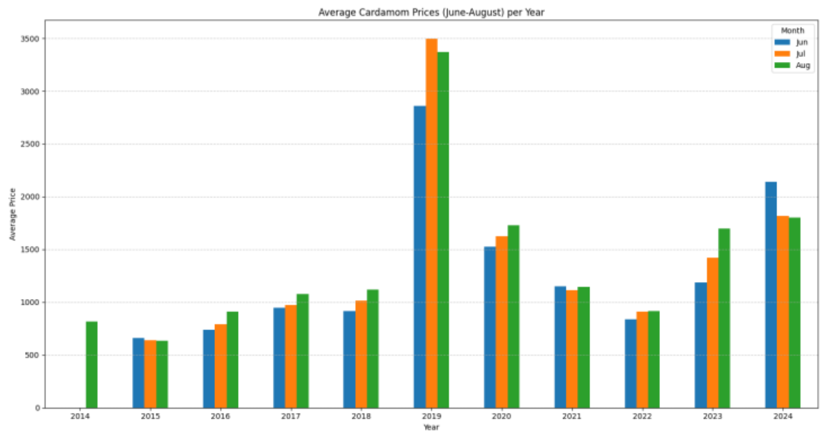

Figure 3. Average Cardamom Prices (June-August) per Year

The Figure 4 shows the average prices of cardamom during the summer season. The average price of cardamom has shown variability from 2015 to 2024. Prices have shown variability from 2015 to 2024. There is a significant spike in 2019, followed by a slight decline during the subsequent years, and a strong recovery by 2024. For most years, May appears to have the highest prices compared to March and April. There's a significant price increase in 2024, with May prices reaching new highs.

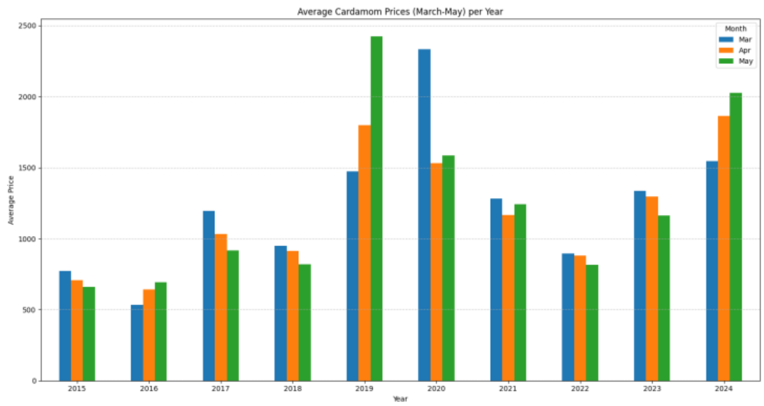

Figure 4. Average Cardamom Prices (March-May) per Year

The trend analysis depicted in the diagram shows that the fluctuations in the prices are significant. There was a price decline effected in the year 2020, and a sharp increase was affected towards the end of 2016 and 2019. The recent trends in the years 2023 and 2024 show that prices exerted a combination of ups and downs. In the year 2024, there is an upward trend in the middle of the year, followed by a slight decrease at the end of the year.

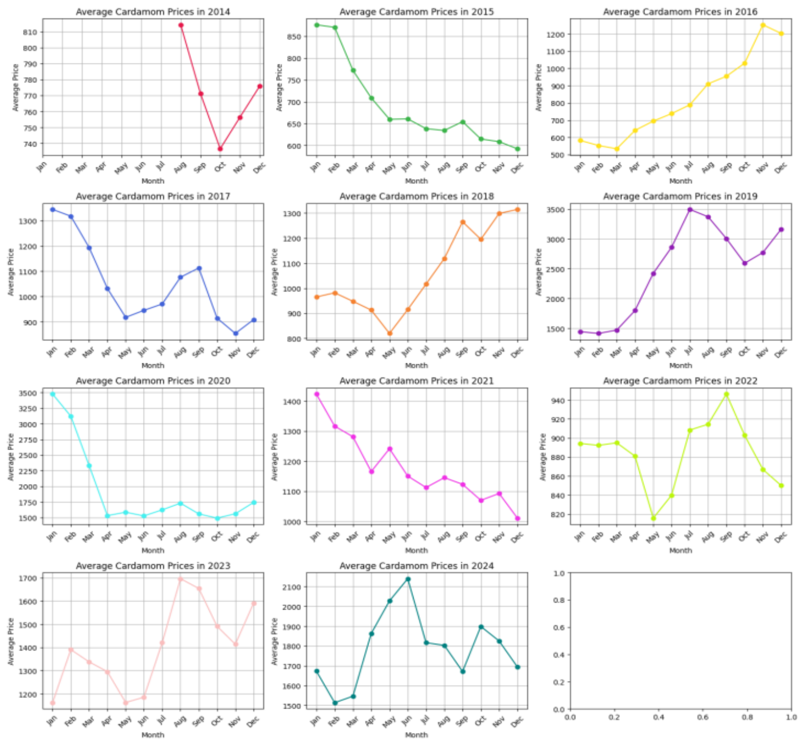

Figure 5. Yearly Price Trends from 2014 – 2024

4.2 Prediction Results

4.2.1 SVM Model (Support Vector Machine Model)

The figure 6 portrays the price prediction of cardamom using SVM. The actual prices show significant fluctuations, with values ranging from 500 to 4000 above. The forecasted prices were determined to be more stable and confined within a limit range of approximately 1000 - 1200. This signifies that the model fails to account for the excessive volatility and significant variability in the actual price data. The predicted values are consistently misaligned with actual values, suggesting that the SVM model inaccurately represents the relationship between features and cardamom prices. This causes substantial discrepancies and weak alignments especially during the ups and downs of prices.

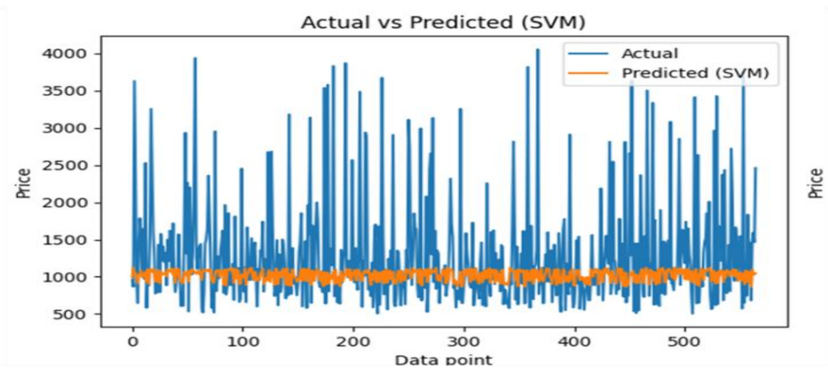

Figure 6. Price prediction of cardamom using SVM

4.2.2 ANN analysis (Artificial Neural Networks)

From the ANN analysis Figure 7 portrayed that there was significant variability in the actual prices, ranging from 500 to 4000 above.

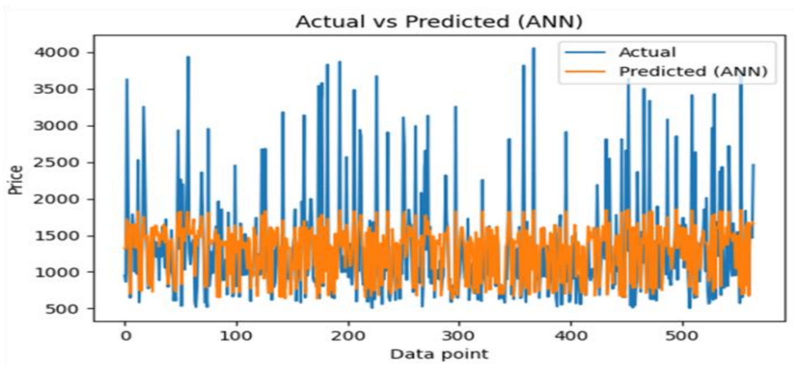

Figure 7. ANN Predictions in Capturing Actual Price Variability

The ANN predictions appear to have a smaller range clustering between approximately 1000 and 2000 and fail to capture the full range of the actual prices. The model consistently under predicts the actual prices, especially for the higher values.

4.2.3 KNN model (K-Nearest Neighbors model)

The Figure 8 KNN model illustrated a high variability in actual prices. The actual prices shown a a wide range from about 500 to 4000 above which is similar to ANN model. As like the same The KNN model predictions span a wider range compared to the ANN model but still fail to capture the extreme peaks and valleys in the actual data. At some points, the predicted prices appear closer to the actual values compared to the ANN predictions. However, the overall alignment is still weak. Similar to the ANN model, the KNN predictions consistently under predict the higher actual price values. The KNN predictions show rapid oscillations, potentially reflecting sensitivity to local data points.

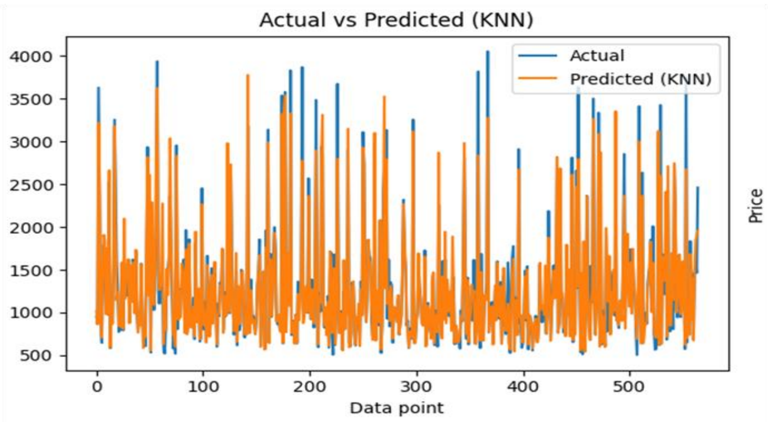

Figure 8. Challenges of KNN Predictions in Accurately Reflecting Actual Price Fluctuations

4.2.4 Random Forest Model

When comparing with the models ANN and KNN, (Figure 9) the Random Forest Model predictions align more closely with the actual prices, capturing some of the variability in the data. The peak and troughs of price levels are better represented, although extreme values are still not perfectly captured. While Random Forest predictions are closer to the actual values, there are still noticeable gaps, especially for higher prices. The model struggles with outliers or extreme values in the dataset, as ensemble methods like Random Forest often average predictions, resulting in conservative estimates for high or low extremes.

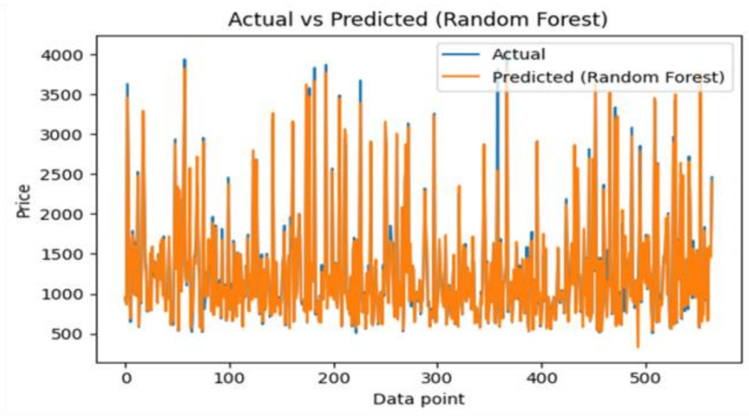

Figure 9. Strengths and Limitations of Random Forest in Capturing Actual Price Dynamics

4.2.5 Decision Tree Model

The price prediction through the decision tree model (Figure 10) was found reasonably accurate in capturing price trends. Both actual and predicted prices exhibit high variability, with frequent spikes and drops. The Decision Tree model appears to have captured the general price trends but struggles with some of the sharp spikes, potentially due to over fitting or the inability of the model to generalise in certain scenarios. The decision tree provides a good approximation; it might not be the best model for such volatile data.

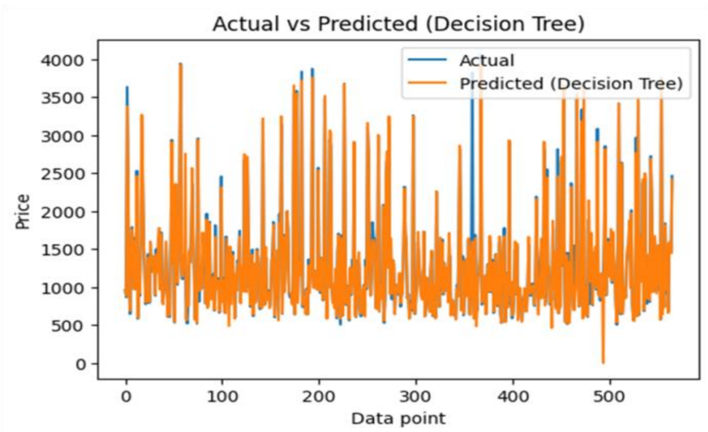

Figure 10. Trend Capture and Challenges with Volatile Price Data

Table 1. Comparison between SVM, ANN, KNN, Random Forest and Decision Tree Models

Model	MSE	R-squared	MAE
SVM	507569.06	-0.056047	434.11
ANN	255709.32	0.467972	338.73
KNN	39188.82	0.918464	110.57
Random Forest	17258.02	0.964093	51.17
Decision Tree	28787.65	0.940104	57.15

The comparison of various models (SVM, ANN, KNN, Random Forest, and Decision Tree) highlights significant differences in their ability to predict cardamom prices. The SVM model performs poorly, with the highest MSE (507,569.06), a negative R2R^2 (-0.056), and a high MAE (434.11), indicating its unsuitability for capturing the nonlinear relationships in the data. ANN shows moderate improvement with an R2R^2 of 0.467, but its MSE (255,709.32) and MAE (338.73) suggest further tuning is required. KNN performs well, achieving a high R2R^2 of 0.918, significantly lower MSE (39,188.82), and a moderate MAE (110.57), demonstrating its ability to capture patterns effectively.

Random Forest emerges as the best-performing model with the lowest MSE (17,258.02), the highest R2R^2 (0.964), and the smallest MAE (51.17), indicating its robustness in handling the data's variability. Decision Tree, while performing slightly worse than Random Forest, still exhibits excellent forecasting capability with an R2R^2 of 0.940, an MSE of 28,787.65, and an MAE of 57.15. However, these models face limitations, including sensitivity to noise, risks of over fitting (especially Decision Tree and Random Forest), poor scalability and interpretability (ANN and Random Forest), and dependency on the quality of input features.

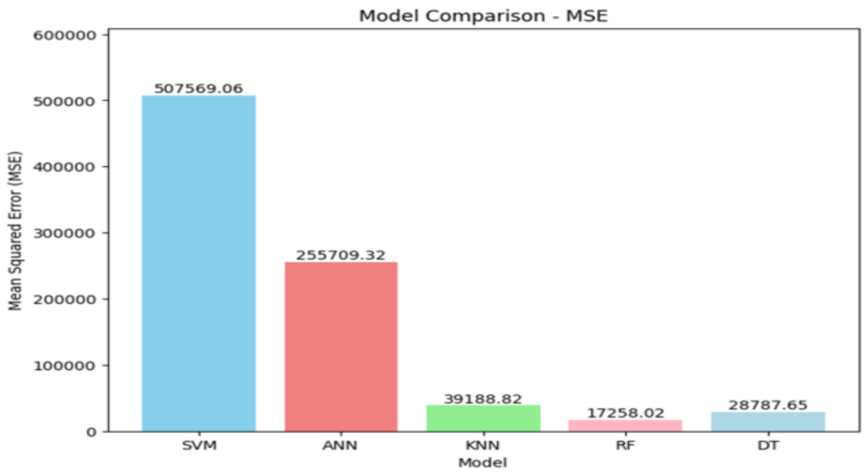

Figure 11. Model evaluation based on Mean Squared Error (MSE)

The Figure 11 bar chart compares the Mean Squared Error (MSE) of various models such as SVM, ANN, KNN, Random Forest (RF), and Decision Tree (DT) for predicting cardamom prices. SVM exhibits the highest MSE (507,569.06), indicating its poor predictive capability for this dataset. ANN follows with a significantly lower MSE of 255,709.32, but it still shows limited accuracy. KNN achieves a much lower MSE of 39,188.82, indicating improved performance and the ability to capture patterns in the data more effectively. Random Forest emerges as the best-performing model, with the lowest MSE of 17,258.02, closely followed by Decision Tree at 28,787.65. The substantial reduction in MSE for Random Forest and Decision Tree compared to the other models highlights their robustness in handling the nonlinearity and variability of cardamom prices. This comparison demonstrates that models like Random Forest outperform simpler models, while SVM struggles to generalise in this context.

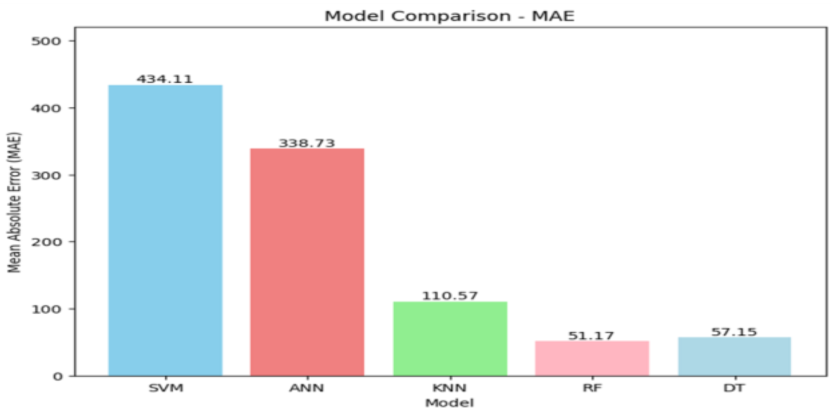

Figure 12. Evaluation of models based on Mean Absolute Error

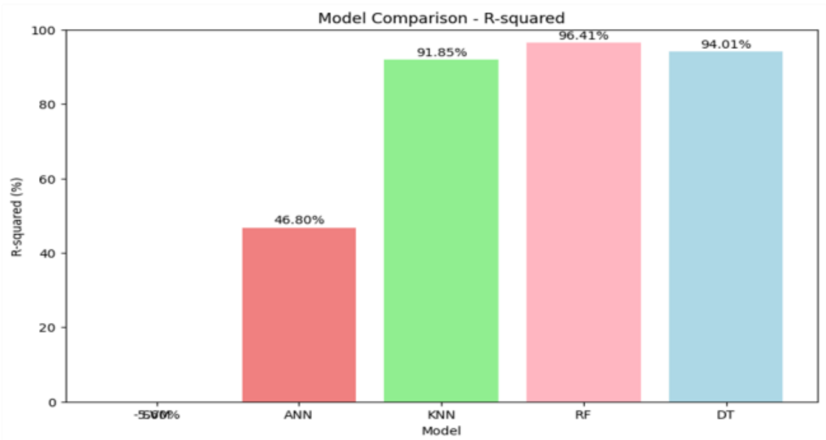

Figure 13. Evaluation of models based on R Square

The bar chart illustrates the R-squared values of various models ANN, KNN, Random Forest (RF), and Decision Tree (DT) in predicting cardamom prices. Random Forest achieved the highest R-squared value of 96.41%, signifying its superior capacity to explicate the variance in the data and make accurate predictions. Decision Tree followed closely with an R-squared value of 94.01%, also demonstrating excellent predictive performance. KNN performed well, with an R-squared value of 91.85%, indicating a strong fit to the data but slightly less robust than the ensemble methods. ANN had a lower R squared value of 46.80%, showing limited accuracy and an inadequate capability for accurately the underlying data structure. The R squared value of SVM is negative, demonstrating the inadequate fit to the data. The results reaffirm the superiority of Random Forest and Decision Tree as stable models for cardamom price forecasting while ANN and SVM do not possess similar efficacy. The results obtained independently from each of these models failed to produce reliable predictions. The present research recommends the application and use of ensemble models for precise future price forecasting as an advanced alternative to independent methodologies.

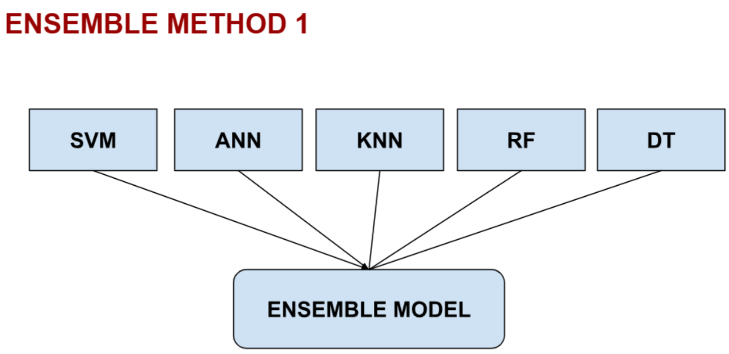

Figure 14. Ensemble Model: 1

Table 2. Ensemble Model Enhances Accuracy in Cardamom Price Prediction

MODEL	MSE	R^2	MAE
Ensemble	87651.882457	81 %	178.193389

The first ensemble model constitutes all the individually analysed models such as SVM, ANN, KNN, RF and DT. The performance measures of this ensemble model indicate efficiency in forecasting cardamom prices. The MSE of 87651.88 signifies the average squared variance between actual and predicted values and the MAE value 178.19 denotes the average absolute difference between these values. Moreover, the coefficient of determination (R^2) of 81% underlines the model's efficacy in predicting cardamom prices. These result underscores the precision of the proposed ensemble method which incorporates the predictions from various models to improve prediction accuracy.

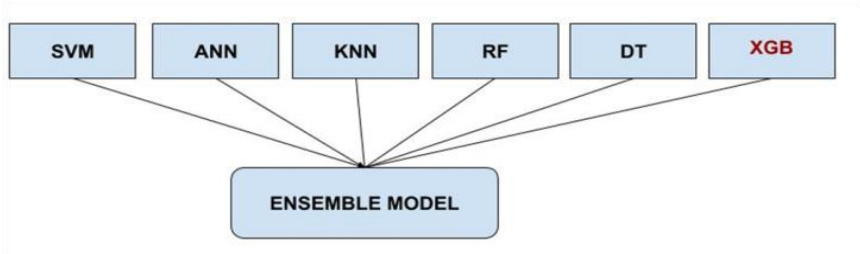

Figure 15. Ensemble Model – 2

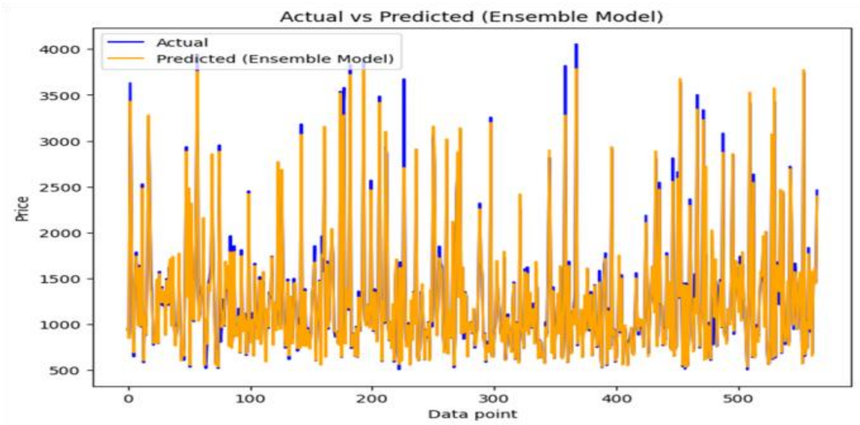

Figure 16. Actual vs Predicted Cardamom Prices Using Ensemble Model

Table 3. Ensemble Model Enhances Accuracy in Cardamom Price Prediction

MODEL	MSE	R^2	MAE
Ensemble	13202.080028	97.25%	54.19

This assessment measures indicate the efficiency of the ensemble model in forecasting cardamom prices. The model generated MSE value 13202.08, indicates a negligible divergence between actual and predicted values. The R squared value 97.25 denotes that the model explains 97.25% of the variance in the outcome variable, showing highest predictive capability. The value of MAE 54.19 underscores its accuracy by measuring the absolute average discrepancies between actual and predicted prices.

The illustration of actual and predictive prices signifies a strong correlation emphasising the highest predictive capacity of the proposed model which integrates the predictions of SVM, ANN, KNN, RF, and DT along with XG Boost model. The integration of XG Boost model provides the better prediction capacity and exhibit accurate prediction of the cardamom prices.

Table 4. Comparison of overall proposed models

Model	MSE	R-squared	MAE
SVM	507569.06	-0.056	434.11
ANN	255709.32	0.468	338.73
KNN	39188.82	0.918	110.57
Random Forest	17258.02	0.964	51.17
Decision Tree	28787.65	0.940	57.15
Ensemble 1	87651.88	0.818	178.19
Ensemble 2	13202.08	0.973	54.19

The measured values of all the models for forecasting cardamom prices indicate variances in predictive capability. Among the individually applied models, the SVM model exhibits least prediction power characterised by a high MSE value of 50759.06 and negative R squared value of 0.056 and the RF model shows more accurate prediction power with lowest MSE value of 17258.02, highest R squared value of 0.964 and a minimal MAE 51.17 which underlines strong correlation between actual and predicted values. The models such as KNN and Decision also showed comparatively better prediction results. The ensemble model approach of forecasting cardamom prices produced optimal results. The Ensemble 2 model with the integration of XG Boost model outperformed well with lowest MSE value (13202.08), highest R squared value 0.973 and minimal MAE (54.19) showcased the finest performance on cardamom price forecasting.

5. DISCUSSION

The prices of cardamom are exerting high volatility. The present study proposes a machine learning ensemble model to predict the prices of cardamom more accurately by using daily average prices. The study individually employed models such as SVM, ANN, KNN, Random Forest, and Decision tree to predict the prices. But none of the model can provide precise price prediction. The price prediction values of all the individual model analyses failed to align with actual prices. The shortcoming of SVM is that it might experience over fitting (Pathak, 2020). The SVM algorithm is not appropriate for large scale datasets.

The model exhibits inefficiency when the data set contains heavy amount of nosie.SVM exhibits inadequate performance when the number of features for each data point exceed the values of the training data samples. It has no possible probabilistic explanation for the classification it so carried out (Bansal et al., 2020). As mentioned above, here in this study the SVM model was found unsuitable for capturing the nonlinear relationships in the data. The ANN predictions fail to capture the full range of the actual prices.

The model consistently under predicts, especially for the higher values. The KNN model has limitations such as the number of nearest neighbours' must first be determined, can be computationally expensive, Memory limitation and Sensitive to the local structure of the data (Farahani & Hajiagha, 2021) Like ANN model, KNN predictions also consistently under predict the higher actual price values and also show rapid oscillations.

The Random Forest model struggles with outliers or extreme values in the dataset. This model often results in conservative estimates for average predictions, high or low extremes. The decision tree model has the issue of over fitting and the decision tree loses some of the valuable information while categorizing variables in various categories (Bansal et al., 2021; Bansal et al., 2022). The decision tree model model found also inadequate on accurate cardamom price forecasting. All these individually employed models face limitations, including risks of over fitting, poor scalability and interpretability, and dependency on the quality of input features. Therefore, the present study goes for ensemble models by combining all those individually treated models for price predictions.

The ensemble model which constitutes SVM, ANN, KNN, Random Forest and Decision Tree. The price prediction was accurate to an extent but still have some limitations such as less accurate aggregation of individual model outputs, leading to poorer alignment with actual values and weak predictive power.

Therefore, the study includes XG Boost model along with ensemble model 1 to form comprehensive hybrid model for cardamom price prediction. It was found that ensemble Model 2 with XG Boost outperforms Ensemble Model 1 and the XG Boost model significantly across various evaluation metrics, including Mean Squared Error (MSE), R^2, and Mean Absolute Error (MAE). Ensemble Model 2 demonstrates a highly optimised performance with an MSE of 13,202, an R^2 value of 97.25%, and an MAE of 54.19, compared to Ensemble Model 1's MSE of 87,651, R^2 of 81%, and MAE of 178.19.

This stark difference highlights Ensemble Model 2's superior ability to capture the variance in the data and minimise prediction errors. The study also similar to findings of the past literature on ensemble models are more comprehensive and adequate in price forecasting for electricity (Hussain et al., 2024), crude oil (Yu et al., 2015). The present research emphasises that the ensemble model with XG Boost is more accurate for cardamom price forecasting.

6. CONCLUSION

The study concludes by revealing the efficacy and efficiency of ensemble models in price forecasting. Given the substantial fluctuations in the cardamom prices, it is imperative to adopt advanced price forecasting methodologies to enhance the accuracy and reliability in price predictions. The ensemble model which incorporating SVM, ANN, KNN, RF, DT and XG Boost emerges as the best ensemble model and an insightful measure as evidenced by the results reinforcing the effectiveness in forecasting of future cardamom prices fluctuations.

7. THEORETICAL AND PRACTICAL IMPLICATIONS

The study significantly contributes to the literature of machine learning techniques for price forecasting. The demonstrated efficacy of XG Boost model highlight is potential as a robust machine learning algorithm for price forecasting. The study also put forward practical implications such as policy makers and cardamom farmers. By utilizing the prediction capacity of the proposed ensemble model it can be able to formulate and apply for prediction of cardamom prices based on the daily average price. Successful forecasting models empower both the policy makers and farmers to make decisions and plans for future to enhance favourable circumstances.

8. LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

Only the daily average prices of cardamom in considered for price forecasting in this study. Further research considers all the other factors such as seasonality, government regulations, production and processing cost, demand and supply etc. The study does not assess the individual performance of XG Boost algorithm in cardamom price forecasting. Future research can be done on a new composition of machine learning algorithms in the price prediction process.

REFERENCES:

[1]. Almaafi, A., Bajaba, S., Alnori, F. (2023). Stock price prediction using ARIMA versus XGBoost models: the case of the largest telecommunication company in the Middle East. *International Journal of Information Technology*, *15*(4), 1813–1818. https://doi.org/10.1007/s41870-023-01260-4

[2]. Bansal, M., Malik, S., Kumar, M., Meena, N. (2020). Arduino based Smart Walking Cane for Visually Impaired People. *2020 Fourth International Conference on Inventive Systems and Control (ICISC)*, 462–465. https://doi.org/10.1109/icisc47916.2020.9171209

[3]. Bansal, M., Goyal, A., Choudhary, A. (2021). Industrial Internet of Things (IIOT): A vivid perspective. In *Lecture notes in networks and systems* (pp. 939–949). https://doi.org/10.1007/978-981-16-1395-1_68

[4]. Bansal, M., Goyal, A., Choudhary, A. (2022). A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning. Decision Analytics Journal, 3, 100071. https://doi.org/10.1016/j.dajour.2022.100071

[5]. Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

[6]. Chen, T., Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, 785–794. Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

[7]. Dezhkam, A., Manzuri, M.T. (2022). Forecasting stock market for an efficient portfolio by combining XGBoost and Hilbert–Huang transform. *Engineering Applications of Artificial Intelligence*, *118*, 105626. https://doi.org/10.1016/j.engappai.2022.105626

[8]. Farahani, M.S., Hajiagha, S.H.R. (2021). Forecasting stock price using integrated artificial neural network and metaheuristic algorithms compared to time series models. *Soft Computing*, *25*(13), 8483–8513. https://doi.org/10.1007/s00500-021-05775-5

[9]. González, C., Mira-McWilliams, J., Juárez, I. (2015). Important variable assessment and electricity price forecasting based on regression tree models: classification and regression trees, Bagging and Random Forests. IET Generation Transmission & Distribution, 9(11), 1120–1128. https://doi.org/10.1049/iet-gtd.2014.0655

[10]. Gordon, A.D., Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. (1984). Classification and Regression Trees. *Biometrics*, *40*(3), 874. https://doi.org/10.2307/2530946

[11]. Government of Kerala, Department of Economics and Statistics (2023). A compendium of agricultural statistics: Kerala 2023. Department of Economics and Statistics, Kerala.

[12]. Gumus, M., & Kiran, M. S. (2017). Crude oil price forecasting using XGBoost. *International Conference on Computer Science and Engineering (UBMK)*, October, 1100–1103. IEEE. https://doi.org/10.1109/UBMK.2017.8093500

[13]. Hussain, S., Teni, A.P., Hussain, I., Hussain, Z., Pallonetto, F., Eichman, J., Irshad, R.R., Alwayle, I.M., Alharby, M., Hussain, M.A., Zia, M.F., Kim, Y. (2024). Enhancing electric vehicle charging efficiency at the aggregator level: A deep-weighted ensemble model for wholesale electricity price forecasting. *Energy*, *308*, 132823. https://doi.org/10.1016/j.energy.2024.132823

[14]. Jabeur, S.B., Mefteh-Wali, S., Viviani, J. (2021). Forecasting gold price with the XGBoost algorithm and SHAP interaction values. *Annals of Operations Research*, *334*(1–3), 679–699. https://doi.org/10.1007/s10479-021-04187-w

[15]. Kontopoulou, V.I., Panagopoulos, A.D., Kakkos, I., Matsopoulos, G.K. (2023). A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks. Future Internet, 15(8), 255. https://doi.org/10.3390/fi15080255

[16]. Lin, G., Lin, A., Cao, J. (2020). Multidimensional KNN algorithm based on EEMD and complexity measures in financial time series forecasting. *Expert Systems With Applications*, *168*, 114443. https://doi.org/10.1016/j.eswa.2020.114443

[17]. Menculini, L., Marini, A., Proietti, M., Garinei, A., Bozza, A., Moretti, C., Marconi, M. (2021). Comparing Prophet and Deep Learning to ARIMA in Forecasting Wholesale Food Prices. Forecasting, 3(3), 644–662. https://doi.org/10.3390/forecast3030040

[18]. Myneedi, J. B., Lautre, N. K., Dumpala, R. (2023). Forecasting Price of Small Cardamom in Southern India Using ARIMA Model. In *Lecture notes in mechanical engineering* (pp. 33–44). https://doi.org/10.1007/978-981-99-1328-2_4

[19]. Narayana, D., Sivanandan, P., Nair, K.N. (1985). Trends and fluctuations in prices and output of cardamom in India. Indian Journal of Agricultural Economics, 40(2), 160–169. https://doi.org/10.22004/ag.econ.271320

[20]. Naveenkumar, R., Bhadra, S., Haldurai, L., Visalakshi, S., Kumar, N. (2024). Environmental Artificial Intelligence Paving the Way for a Greener and More Resilient

Planet using Machine learning. *Journal of Computational Analysis and Applications*, *33*(7).

[21]. Pathak, N.A. (2020). Study of Machine learning Algorithms for Stock Market Prediction. *International Journal of Engineering Research And*, *V9*(06). https://doi.org/10.17577/ijertv9is060064

[22]. Raman, G.K. (2022). Edamalakkudy farmers stick to indigenous cardamom. The Hindu. April 21. https://www.thehindu.com/news/national/kerala/edamalakkudy-farmers-stick-to-indigenous-cardamom/article65341588.ece

[23]. Simsek, A.I., Bulut, E., Gur, Y.E., Tarla, E.G. (2024). A Novel Approach To Predict WTI Crude Spot Oil Price: LSTM-Based Feature Extraction With Xgboost Regressor. *Energy*, *309*, 133102. https://doi.org/10.1016/j.energy.2024.133102

[24]. Somkunwar, R.K., Pimpalkar, A., Srivastava, V. (2024). A Novel Approach for Accurate Stock Market Forecasting by Integrating ARIMA and XGBoost. *2020 IEEE International Students' Conference on Electrical,Electronics and Computer Science (SCEECS)*, *3*, 1–6. https://doi.org/10.1109/sceecs61402.2024.10481891

[25]. Sonkavde, G., Dharrao, D.S., Bongale, A.M., Deokate, S.T., Doreswamy, D., Bhat, S.K. (2023). Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications. *International Journal of Financial Studies*, *11*(3), 94. https://doi.org/10.3390/ijfs11030094

[26]. Spyrou, E. D., Tsoulos, I., Stylios, C. (2022). Applying and comparing LSTM and ARIMA to predict CO levels for a Time-Series measurements in a port area. Signals, 3(2), 235–248. https://doi.org/10.3390/signals3020015

[27]. Sun, F., Meng, X., Zhang, Y., Wang, Y., Jiang, H., Liu, P. (2023). Agricultural Product Price Forecasting Methods: A Review. *Agriculture*, *13*(9), 1671.https://doi.org/10.3390/agriculture13091671

[28]. Tissaoui, K., Zaghdoudi, T., Hakimi, A., Nsaibi, M. (2022). Do Gas Price and Uncertainty Indices Forecast Crude Oil Prices? Fresh Evidence Through XGBoost Modeling. *Computational Economics*, *62*(2), 663–687. https://doi.org/10.1007/s10614-022-10305-y

[29]. Thomas, M., Menon, P.P. (2022). Forecasting of small cardamom price using SARIMA model. The Journal of Research ANGRAU, 50(4), 116-123.

[30]. Varghese, P.K. (2004). Trend analysis in area, production, productivity and price behaviour of cardamom in Kerala. Indian Journal of Agricultural Economics, 59(4), 798–807. https://eurekamag.com/research/004/372/004372735.php

[31]. Vellaram, S. (2024). Cardamom prices show a declining trend despite fears of crop loss. The Hindu. August 21. https://www.thehindu.com/news/national/kerala/cardamom-prices-show-a-declining-trend-despite-fears-of-crop-loss/article68550690.ece

[32]. Vuong, P.H., Dat, T.T., Mai, T.K., Uyen, P.H., Bao, P. (2021). Stock-Price Forecasting Based on XGBoost and LSTM. *Computer Systems Science and Engineering*, *40*(1), 237–246. https://doi.org/10.32604/csse.2022.017685

[33]. Wang, L., Feng, J., Sui, X., Chu, X., Mu, W. (2020). Agricultural product price forecasting methods: research advances and trend. *British Food Journal*, *122*(7), 2121–2138. https://doi.org/10.1108/bfj-09-2019-0683

[34]. Wang, J., Cheng, Q., Dong, Y. (2022). An XGBoost-based multivariate deep learning framework for stock index futures price forecasting. *Kybernetes*, *52*(10), 4158–4177. https://doi.org/10.1108/k-12-2021-1289

[35]. Wu, K., Chai, Y., Zhang, X., Zhao, X. (2022). Research on Power Price Forecasting Based on PSO-XGBoost. *Electronics*, *11*(22), 3763. https://doi.org/10.3390/electronics11223763

[36]. Yu, L., Wang, Z., Tang, L. (2015). A decomposition–ensemble model with data-characteristic-driven reconstruction for crude oil price forecasting. *Applied Energy*, *156*, 251–267. https://doi.org/10.1016/j.apenergy.2015.07.025

[37]. Zhao, H. (2020). Futures price prediction of agricultural products based on machine learning. *Neural Computing and Applications*, *33*(3), 837–850. https://doi.org/10.1007/s00521-020-05250-6

[38]. Zou, H., Xia, G., Yang, F., Wang, H. (2007). An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting. *Neurocomputing*, *70*(16–18), 2913–2923. https://doi.org/10.1016/j.neucom.2007.01.009