Indian Journal of Agricultural Economics 79: 2 (2024):214-230
DOI:10.63040/25827510.2024.02.003

Enhancing Forecasting Accuracy of Palm Oil Import to India Using Machine Learning Techniques

K. Naga Latha1[1], V. Srinivasa Rao[1], C. Sarada[2], A. Amarender Reddy[3] and K. N. Sreenivasulu[1]

ABSTRACT

Forecasting of palm oil imports to India has gained significant prominence in contemporary times due to huge exchequer is spending on vegetable oil imports. The government is keen on reducing imports. The quantity to be imported in future years is utmost important to make any policies or programs to enhance the oilseeds production. For forecasting ARIMA models have been the most widely used technique during the last few decades. When the assumption of homoscedastic error variance is violated then ARCH/GARCH models are applied to capture the changes in the conditional variance of the time-series data. The machine learning techniques, i.e., ANN and SVR, can also be applied in the field of forecasting of real time-series data successfully as an alternative to the traditional forecasting models as these are data-driven models and could capture nonlinearities existing in the data. The present study analyzed monthly time series data of palm oil import volume (thousand tonnes) from the world to India from April 2007 to March 2023. It is clear from the results that the machine learning models, viz, SVR and ANN, outperformed the traditional time series models (GARCH and ARIMA) with the least RMSE, MAPE, Theil's U statistic and the highest CDC values for both training and testing datasets. Empirical results revealed that SVR is the best model for forecasting palm oil import volume compared to all other models.

Keywords: ANN, machine learning, palm oil import, SVR.

JEL Classification: C45, C53, C63, Q17, Q18, Q56

I

INTRODUCTION

Vegetable oils are an integral component of the Indian diet, playing a vital role in many traditional and modern culinary dishes. India is one of the largest producers of oilseeds in the world, and this sector occupies an important position in the agricultural economy, accounting for the estimated production of 215.33 lakh tonnes of nine cultivated oilseeds during the year 2022-23 (Fourth Advance Estimates released by the Ministry of Agriculture).

Domestic production of edible oils is unable to meet domestic demand. As per the 4th Advance Estimates released by the Department of Agriculture, Cooperation and Farmers Welfare the estimated production of oilseeds for 2021-22 is 376.97 lakh tonnes as against 359.45 lakh tonnes in 2020-21. The total availability of edible oils from all sources (primary and secondary) for 2021-22 is estimated at 115.71 lakh

[1] Department of Statistics and Computer Applications, Agricultural College, ANGRAU, Bapatla, Andhra Pradesh, India-522101 [2]ICAR-Indian Institute of Oilseed Research, Hyderabad, Telangana, India-500030 and [3]ICAR-National Institute of Biotic Stress Management, Raipur, India-493225.

The authors thank the Department of Statistics and Computer Applications, Agricultural college, Bapatla, ANGRAU and ICAR-Indian Institute of Oilseed Research, Hyderabad for extending their support, guidance and technical assistance in conducting this study.

tonnes against 111.51 lakh tonnes in 2020-21. The gap between demand and supply is about 55 per cent and is met through imports. The per capita consumption, which was 19.5 kg per person per annum in 2017-18, has increased to around 21.0 kg at present (Ministry of Finance Economic Survey 2022-23, Statistical Appendix).

The constant increase in consumption, low productivity of oilseeds, the high price of traditional oils in India and low price in the international market, and liberalisation of trade policies resulted in the shift from self-sufficiency to highly import dependent in edible oils (Indhushree and Shivakumar, 2020). Out of all the imported edible oils, the share of palm oil is about 56 per cent, followed by soybean oil at 27 per cent, and sunflower at 16 per cent (National Mission on Edible Oils, 2022). Among the diverse range of vegetable oils available, palm oil, derived from the oil palm tree, has gained significant popularity in India. Almost all food products in supermarkets use palm oil (Septyari, 2021).

The import of palm oil from other countries has become necessary for India for several reasons. India's demand for edible oils, including palm oil, far exceeds its domestic production capacity. The oil palm, from which palm oil is derived, is primarily cultivated in countries with tropical climates, such as Malaysia and Indonesia, and these are the largest producers of palm oil, with 85 per cent of global production (Sagar *et al*., 2019). These nations have well-established palm oil industries, which produce palm oil in large quantities, ensuring a steady supply to meet the growing demands of the Indian market. In 2022, the imported quantity of palm oil and its fractions accounted for 25.24 lakh tonnes (ITC Trade Map).

As Indians continue to incorporate palm oil into their cuisine for convenience, there is a growing need to promote oil palm plantation and awareness about the need for domestic production to decrease the dependency on other countries for domestic consumption.

Recently, a new Centrally Sponsored Scheme, namely, National Mission on Edible Oil -Oil Palm (NMEO-OP), has been launched by the Government to promote oil palm cultivation for making the country Aatmanirbhar in edible oils with special focus on North Eastern States and A&N Islands. The Mission will bring an additional area of 6.5 lakh ha under oil palm plantation, with 3.28 lakh ha in north- eastern states and 3.22 in the Rest of India in the next five years from 2021-22 to 2025-26 (National Mission on Edible Oils, 2022).

The country's economic stability, the reduction of import dependency, and regional development depend heavily on the success of NMEO-OP. It becomes essential to precisely forecast imports of palm oil to meet these goals. These forecasts are essential for informing decision-makers about the country's import-export balance, which helps them create trade and agricultural policies that are effective in practice. Moreover, forecasting changes in the import of palm oil is essential for averting any economic setbacks, ensuring stable commodity prices, and safeguarding the food security of the nation.

Several studies have been conducted on forecasting the export or import of different products. Most of the forecasting techniques used are conventional forecasting techniques such as autoregressive integrated moving average (ARIMA) and generalized autoregressive conditional heteroskedasticity (GARCH) model (Khalid *et al*., 2018; Upadhyay, 2013; Rakhmawan *et al.*,2015). Besides that, there are also other forecasting techniques used, such as artificial neural network (ANN) (Pannakkong *et al.*, 2019) and Support Vector Regression (SVR) (Rathod *et al.*, 2018).

To develop a comprehensive model for palm oil import forecasting, this study explores a variety of forecasting techniques, including traditional methods like the generalized autoregressive conditional heteroskedasticity (GARCH) model and autoregressive integrated moving average (ARIMA) model, as well as machine learning techniques viz., artificial neural networks (ANN) and support vector regression (SVR).

II

MATERIALS AND METHODS

*Data***:** monthly time series data of palm oil import quantity in India from the world based on the 6-digit HS (Harmonized System) code was gathered from multiple sources such as ITC Trade map (https://www.trademap.org), World Bank portal (https://data.worldbank.org), etc. The monthly data collected ranges from April 2007 to March 2023. The data was divided into training and testing sets. Out of 192 data points, 180 were allocated to the model training, and the remaining 12 were reserved for testing the model.

The methods that were used to analyse the data in this study- namely, the SVR, ANN, ARIMA, GARCH models and stationarity and linearity tests are briefly described in the following section:

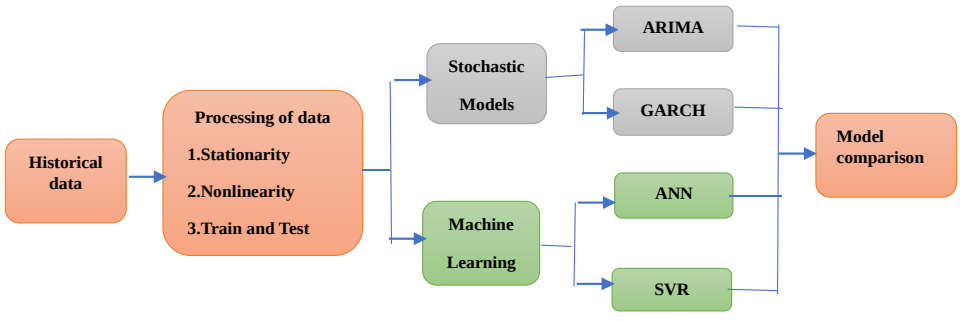


Figure 1. Process Flow Diagram of Methodology

*ADF Test for Stationary***:** The most commonly utilized test for detecting the unit root (non-stationary) of time series is Augmented Dickey Fuller (ADF) test. In conducting Dickey Fuller (DF) test, it was assumed that the error term is uncorrelated. But in most cases, it is correlated. Hence Dickey-Fuller (1979) has developed another test, popularly known as the ADF test. This test is conducted by augmenting the regression [e.g. $(\Delta Y_t = \beta_1 + \delta Y_{t-1} + e_t)$] equation by adding the dependent variable lag values (ΔY_t). The ADF test consists of estimating the following regression,

$$\Delta Y_t = \beta_1 + \delta Y_{t-1} + \sum_{i=1}^{m} \alpha_i \Delta Y_{t-i} + e_t \quad \ldots (1)$$

Where m is the number of lagged difference terms required so that the error term et is serially independent. The test can be carried out by performing a t (= τ) statistic of $'\delta'$ with dickey fuller table values at 5 and 10 percent level of significance. The series is said to be stationary if the absolute t (= τ) statistic of $'\delta'$ is higher than its table value.

*BDS (Brock-Dechert-Scheinman) Test for Non-Linearity***:**

The BDS test (Brock *et al.,* 1996) is a non-parametric test and its null hypothesis is the data is independently and identically distributed (*iid*) against an unspecified alternative. This test enables one to test for nonlinear dependence because it is not affected by linear dependencies in the time series data. BDS test is a two-tailed test, one should reject the null hypothesis if the BDS test statistic is higher than or equal to the critical values (e.g., if a=0.05, the critical value = ±1.96).

QS Statistic for Seasonality:

QS is a statistic that tests the hypothesis of no seasonality. It is applied to appropriate series associated with the modeling and the seasonal adjustment of a given series (Bhattacharya *et al.,* 2016). The statistic QS is assumed to be adequately approximated by a chi-squared. The p-value of the QS-test below 0.01 will classify the corresponding time series as seasonal (Weerasinghe *et al.,* 2021). This test is used in this study to identify seasonality theoretically.

Auto-Regressive Integrated Moving Average (ARIMA):

The ARIMA model, developed by Box and Jenkins, describes time series changes using a mathematical approach known as the Box-Jenkins model, involving procedures to identify, fit, and check ARIMA models with time series data (Box and Jenkins, 1970). A combination of autoregressive and moving average processes is often beneficial to enhance adaptability in fitting real-time series data. The amalgamation gives rise to the combined autoregressive moving average model, denoted as ARMA (p, q). The fundamental assumption in this class of models is that

the data should be time-invariant or stationary. The ARMA (p, q) model is expressed as follows:

$$y_t = \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots \theta_q \varepsilon_{t-q} + \varepsilon_t \quad \ldots (2)$$

Here, ε_t represents independent and identically distributed errors with N (0, σ^2). The integration parameter d is a nonnegative integer. To accommodate a broader class of nonstationary time series, a generalization of ARMA models is achieved by introducing "differencing" in the model, resulting in ARIMA (p, d, q). When d=0, the ARIMA (p, d, q) model simplifies to the ARMA (p, q) model.

Generalized Autoregressive Conditional Heteroskedasticity (GARCH):

In 1986, Bollerslev introduced the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model, wherein the conditional variance is a linear function of its own lagged values. Notably, the GARCH model's conditional variance possesses a characteristic allowing for a slow decay in the autocorrelation function of ε_t^2.

The GARCH model is stated as follows:

$$h_t = a_0 \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} b_j \, h_{t-j} \quad \ldots . (3)$$

For the ARCH family, the decay rate is rapid. The GARCH (1,1) model is prevalent in practical applications. The GARCH (p, q) procedure is weakly stationary if and only if:

$$\sum_{i=1}^{q} a_i + \sum_{j=1}^{p} b_j < 1 \quad \ldots . (4)$$

It is worth noting that the GARCH model can be viewed as an application of the Autoregressive Moving Average (ARMA) model to the squared series $\mathcal{E}_t^2$, where *a* and *b* are constants.

Artificial Neural Network (ANN):

The ANNs are typically structured with layers of units, i.e., artificial neurons or nodes, hence commonly referred to as multilayer ANNs. Every unit within a layer of this architecture is intended to perform a certain task. The first layer consists of input units, referred to as independent variables in statistical terms. Similarly, the last layer contains the output units, statistically denoted as response or dependent variables. Hidden units are positioned between these layers to make up the hidden layers in the model. This layered arrangement makes complex information processing possible within the network.

In time series analysis, the inputs are typically the past observations series, and the output is the future value. The ANN performs the following nonlinear function mapping between the input and output variables of interest.

$$y_t = f(y_{t-1}, y_{t-2}, \ldots, y_{t-p}, w) + \varepsilon_t \quad \ldots (5)$$

Where, w is a vector of all parameters, ε_t is the bias term, and f is a function of network structure and connection weights. Therefore, the neural network resembles a nonlinear autoregressive model.

This study used a single hidden layer within a multilayer feedforward network to construct our Artificial Neural Network (ANN) model. This model is characterized by a network composed of three layers of processing units. The first layer serves as the input layer, followed by the hidden layer, and the last layer is the output layer.

The relation between the output y_t and the inputs (y_{t-1}, y_{t-2}, …, y_{t-p}) can be mathematically represented as follows:

$$y_t = f\left(\sum_{j=0}^{q} w_j g\left(\sum_{j=0}^{p} w_{ij} y_{t-1}\right)\right) + \varepsilon_t \quad \ldots (6)$$

Where, $w_j\,(j = 0,1,2...q)$ and $w_{ij}\,(i = 0,1,2...p,\ j = 0,1,2...q)$ are the model parameters often called the connection weights, p is the count of input nodes, and q is the count of hidden nodes, ε_t is the bias term, g, and f denote the activation function at the hidden and output layers, respectively. The activation function defines the relation between inputs and outputs of a network in terms of the degree of non-linearity. This study employed the sigmoid activation function in the hidden layer, and the identity activation function was utilized for the output layer (Rathod *et al.*, 2017).

Support Vector Regression (SVR):

Vapnik (1998) proposed using Support Vector Machine in supervised learning frameworks for data analysis and pattern recognition. It is a non-linear algorithm. Vapnik (1998) introduced support vector regression model by incorporating Ɛ-loss function. SVR maps input vectors into a high dimensional space and then fits linear regression in outer space. The model has been built in two steps, i.e., the training and the testing.

For the given dataset $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ ($x_i \in R^k$, $y_i \in R^1$), SVR maps the original data into a higher or infinite dimensional space by nonlinear function φ, then seeks mapping function φ: $R^k \rightarrow R^1$. The general formula for linear support vector regression is given as:

$$y = \varphi(x) = w^T \varphi(x) + b \quad \ldots (7)$$

Where w defines the weight vector, φ denotes the mapping function, and b is the bias term. LS-SVR is the least square of SVR, where a set of linear equations is used to find

the solution. The solution of W and b in the above equation can be obtained by solving the following minimization problem

$$\min_{w,b,\zeta,\zeta^*} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\zeta + \zeta_i^*) \quad \ldots (8)$$

Such that $w^T\varphi(x_i) + b - y_i \leq \varepsilon + \zeta_i\,;\; y_i - w^T\phi(x_i) - b \leq \varepsilon + \zeta_i^*$
$\zeta, \zeta^* \geq 0, i = 1, \ldots, N.$

It is a primal function, and the solution of the function is quite complex. So, the dual of the function can be used. Its dual will be

$$\min_{\alpha,\alpha^*} \frac{1}{2}\sum_{i,j}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)Ker(x_i, x_j) + \varepsilon\sum_{i}^{N}(\alpha_i - \alpha_i^*) + \sum_{i}^{N}(\alpha_i - \alpha_i^*) \ldots (9)$$

s.t $\sum_i^N(\alpha_i - \alpha_i^*) = 0$ and $0 \leq \alpha, \alpha^* \leq C$, i=1, …, N
where $Ker(x_i, x_j) = \varphi(x_i)^T\varphi(x_j)$ is a kernel function. For getting the estimated value α, α^* the dual function will be used. Thus, the coefficient b will be calculated as

$$\tilde{b} = y_j - \sum_{i}^{N}(\alpha_i - \alpha_i^*)Ker(x_i, x_j) - \varepsilon.\,;\, \tilde{\alpha_i} \,\epsilon(0, C) \quad \ldots (10)$$

$$\tilde{b} = y_j - \sum_{i}^{N}(\alpha_i - \alpha_i^*)Ker(x_i, x_j) - \varepsilon.\,;\, \widetilde{\alpha_i^*} \,\epsilon(0, C) \quad \ldots (11)$$

The bias term, b in Eq. 5, can be accommodated within the kernel function $Ker(x_i, x_j)$
The regression function is given by:

$$f(x) = \sum_i^N(\bar{\alpha}_i - \alpha_i^*)K(x_i, x_j) \quad \ldots (12)$$

The SVR model (Eq. 10) contains three tuning parameters in the $K(x_i, x_j)$:

1. € epsilon of the loss function 2. C, the constraints, and 3. Sigma of the kernel.

Model Performance Measures:

Forecasting accuracy is key for finding out the practicability of developed models. To compare the model accuracy of the machine learning models with the statistical models, the following evaluation criteria have been used:

Root Mean Squared Error (RMSE):

RMSE is defined as:

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

Where y_i and $\hat{y}_i$ are the actual and predicted values of the dependent variable, respectively. If any model has a higher RMSE value, then the forecasting accuracy is lower for that particular model.

Mean Absolute Percentage Error (MAPE):

MAPE defines as

$$MAPE = \frac{\sum_{i=1}^{N} |y_i - \hat{y}_i| / y_i}{N} \quad \ldots (14)$$

Where y_i and $\hat{y}_i$ are the actual and predicted values of the response variable, respectively. The more it is, the less accurate is the forecast.

Theil's U Statistic:

It was developed by Theil (Theil *et al.*, 1966). This measurement of accuracy allows relative comparison of formal forecasting methods with benchmark method (Naïve forecast model). Naïve model is the most frequently used effective threshold to differentiate different forecast models, and it simply takes the most recent values of the variable in question and extrapolates the future value (Gohain, 2021).
Theil's U-statistic is defined as:

$$Theil's\ U = \frac{RMSE\ forecast\ model}{RMSE\ naive\ model} \quad \ldots. (14)$$

A lower value of Theil's U statistic indicates better forecasting accuracy, and values closer to 0 are considered more desirable.
The range of the U-statistic can be summarised as follows.
U = 1: the forecasting model provides no improvement over using benchmark.
U < 1: the forecasting model being used is better than the benchmark.
U > 1: benchmark model produces better results than the forecasting model.

Correct Directional Change (CDC):

It provides the direction of change that is given by

$$CDC = (100/n) \sum_{t=1}^{n} D_t \quad \ldots.(15)$$

$$where\ D_t = \begin{cases} 1, & if\ (y_t - y_{t-1}).(\hat{y}_t - \hat{y}_{t-1}) > 0 \\ 0, & elsewhere \end{cases}$$

It is employed to determine whether the volatility forecast's direction matches the actual change that has taken place. The higher the value of the CDC, the better the model forecasting accuracy. In their study, Lama *et al.*, 2016 also applied CDC as a model performance metric.

III

RESULTS AND DISCUSSION

This section deals with presenting and interpreting the results and relevant discussions. The analysis was done with the help of R software. There are different packages available for analysis. The packages which are used in this study include "tseries" (ADF test, BDS test), "FinTS" (ARIMA, ARCH-LM test), "seastests" (QS

test), "fGarch" (GARCH), "forecast" (ANN), "e1071" (SVR) "MLmetrics" (Accuracy measures), "ggplot2" (graphics).

4.1 Dataset and Basic Features

For the present study, we have considered monthly times series data of palm oil import quantity from the World to India. The data are collected from the ITC Trade map (https://www.trademap.org). The data points collected are ranging from Apr 2007 to Mar 2023. The time plot of data is demonstrated in Figure 2. A perusal of the plot indicates that the quantity of palm oil imports gradually increased over the years, and the fluctuations over months and years indicate the presence of heteroskedasticity. The summary statistics of the dataset are presented in Table 1. A perusal of Table 1 indicates the mean monthly palm oil import quantity is 144.116 thousand tonnes. A higher CV value indicates the presence of a high variability nature in the data. The data is positively skewed (0.56) and platykurtic (-0.24).

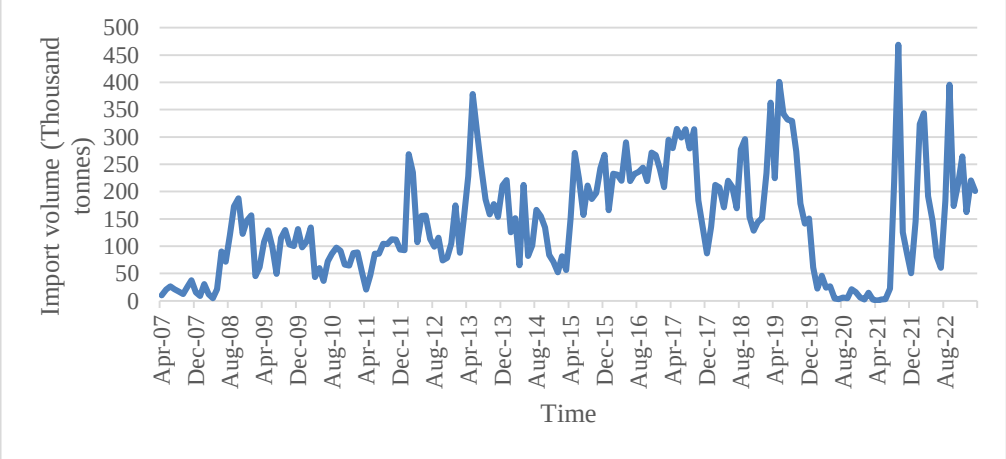


Figure 2. Palm Oil Imports to India In Terms of Volume (April 2007 To March 2023)

TABLE 1. SUMMARY STATISTICS OF PALM OIL MONTHLY IMPORT VOLUME TO INDIA

(*'000 tonnes*)

Statistic	Import volume
(1)	(2)
Observations	192
Minimum	0.04
Maximum	468.65
Mean	144.12
Standard Deviation	99.29
Skewness	0.56
Kurtosis	-0.24
CV (per cent)	68.89

4.2 Preliminary tests for the Data

Augmented Dickey-Fuler (ADF) test has been carried out to test the stationarity of the dataset. The test statistic was calculated as -2.82, and the corresponding P value was 0.24. Since the P value exceeds the common significance level of 0.05, the null hypothesis that the data contains a unit root and is non-stationary is accepted. Therefore, based on the test results, it is clear that data is not stationary. First order differencing was done to bring the time series stationary.

To check for the presence of a nonlinearity pattern in the import volume of palm oil, the Brock-Dechert-Scheinkman (BDS) test is performed, and test results are presented in Table 2. It can be seen that at a 5 per cent level of significance, palm oil import volume has a clear nonlinear pattern.

TABLE 2. BDS TEST RESULTS FOR LINEARITY

Embedding dimension				
2		3		
Statistics	Probability	Statistics	Probability	Conclusion
60.63	<0.001	88.66	<0.001	Nonlinear
31.63	<0.001	36.09	<0.001	
21.41	<0.001	22.21	<0.001	
16.89	<0.001	17.37	<0.001	

This study used the QS test to test the data's seasonality theoretically. The results in Table 3 show that the p value of the QS test (0.10) is greater than the commonly used significance level of 0.05, we fail to reject the null hypothesis, indicating that there is no strong evidence to suggest the presence of seasonality in the data.

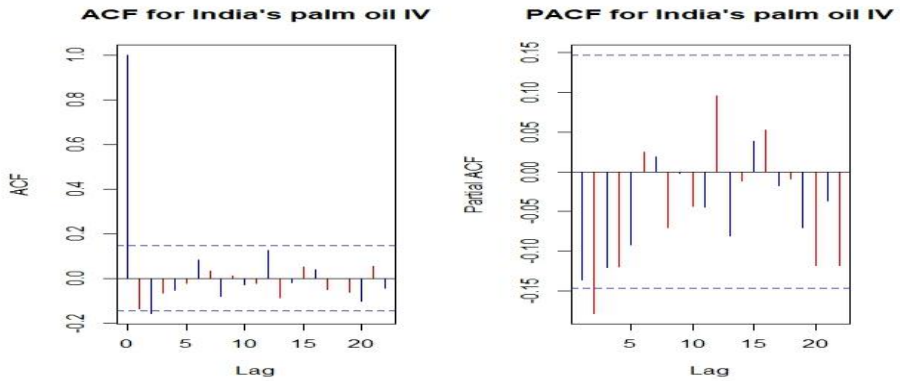
TABLE 3. QS AND ARCH-LM TEST RESULTS

QS test		
Test statistic	P value	Conclusion
4.59	0.10	No seasonality
ARCH-LM test		
Chi Squared value	P value	Conclusion
61.64	0.00	ARCH-effect present in residuals.

4.3 Fitting of Different Time Series Models

In the present study, we have considered stochastic models, e.g., ARIMA and GARCH model, machine learning (ML) techniques, i.e., ANN and SVR, to examine India's palm oil import volume.

The Autocorrelation (ACF) and Partial Autocorrelation Function (PACF) plots of palm oil imports are depicted in Figure 3 below. The figures indicate that the monthly time series was autocorrelated, which was supported by Box-Jung test statistic as the probability value was less than 0.05, which shows that data under



Note: IV = Import Volume

Figure 3. ACF and PACF for India's Palm Oil Import Volume

consideration was autocorrelated in nature. Once the time series data of palm oil imports were autocorrelated, the ARIMA model was built for the series. Further, the QS test was applied to the data to test the presence of seasonality. The results suggest there's no seasonality in the data, so it may not be necessary to consider a seasonal ARIMA model for modeling the data.

4.3.1 ARIMA Model:

Auto Regressive Integrated Moving Average model is a linear time series model. The ARIMA model has been built for India's palm oil import volume. The original data series was found to be non-stationary, so first differencing was done to make the time series stationary. The adequate model, i.e., ARIMA (1,1,1), has been identified based on Autocorrelation and Partial Autocorrelation Function (ACF and PACF) plots (Figure 3). The parameters of the ARIMA model are estimated using the maximum likelihood method. Further, the model performance criteria of RMSE, MAPE, Theil's U statistics, and CDC for both training and testing sets were given in *Table 5*. Petrevska (2017) also found a similar result, where the ARIMA (1,1,1) model was appropriate for predicting tourism demand.

4.3.2 GARCH Model

After fitting the ARIMA model, the ARCH Lagrange Multiplier (LM) test, a heteroscedastic test developed by Engle (1982), was used to determine the presence of the ARCH effect in the residuals. The null hypothesis of this test is that there is no ARCH effect in the residual series (Rakshit *et al.*,2021).

Table 3 reveals the results of the ARCH-LM tests. The results revealed the ARCH effect on India's palm oil import volume; hence, we need to apply the GARCH model. AR (1,1)-GARCH (1,1) model is the best model for forecasting palm oil import and was selected based on the least RMSE, MAPE, Theil's U statistic, and the highest

CDC values. The findings are presented in Table 5. The GARCH model outperformed the linear model, i.e., ARIMA, in the case of both in-sample and out-sample data sets. Paul *et al.* (2009) also applied ARIMA and GARCH for modeling and forecasting India's volatile spices export dataset.

4.3.3 Artificial Neural Network (ANN) Model

A multilayer network was trained using the feedforward backpropagation algorithm to analyse India's palm oil import volume. The results presented in Table 4 indicated that the optimal input lag for import volume was determined to be 3. Different network topologies were trained by increasing the number of hidden nodes from 1 to 25, with the sigmoid function employed as the activation function in the hidden layer. Among the various models tested, the top-performing model was considered. With the configuration of 3-6-1, the neural network model emerged as the best performer for import volume. The model was selected based on the lowest values of RMSE and Theil's U statistic. Muharni and Denisa (2021) also found that the ANN (3-6-1) model is best for quality prediction of industrial standard water. The model performance in the training and testing set is presented in Table 5. Model Building and testing of ANN was done with the help of R software using a package called "forecast" (Hyndman *et al.*, 2022).

TABLE 4. ANN MODEL DESCRIPTION

Particulars	ANN parameter
(1)	(2)
Optimum lag	3
Optimum hidden node	6
Network type	(3, 6, 1): Feed forward
Activation function	Sigmoidal

TABLE 5. OVERALL MODEL PERFORMANCE FOR TRAINING AND TESTING SET OF INDIA'S PALM OIL IMPORT VOLUME (THOUSAND TONNES)

	ARIMA	GARCH	ANN	SVR
(1)	(2)	(3)	(4)	(5)
RMSE (train)	61.74	57.71	36.78	10.65
RMSE (test)	102.54	96.86	67.95	29.01
MAPE (train)	484.44	294.75	277.37	39.61
MAPE (test)	102.54	68.19	46.46	11.09
CDC (train) (per cent)	42.45	44.13	57.39	93.85
CDC (test) (per cent	54.54	54.54	63.64	100
Theil's U statistic (train)	0.18	0.17	0.11	0.03
Theil's U statistic (test)	0.25	0.22	0.15	0.07

4.3.4 Support Vector Regression (SVR)

The support vector regression model for India's palm oil import volume time series was analysed using R software with the help of package 'e1071' (David, 2017). The SVR model was specifically configured with a radial basis function (RBF) kernel, which is well-suited for capturing complex relationships in the data. The hyperparameters of the SVR model were fine-tuned to enhance its performance, with

the regularization parameter (C) set to 1, the gamma parameter of the RBF kernel set to 0.33, and the epsilon parameter set to 0.01. These parameters were the most effective in improving the model's predictive performance. Additionally, the SVR model comprised 31 support vectors, indicating the complexity of the relationship between the input and target variables.

Based on the lowest root mean square error (RMSE), Mean Absolute Percentage Error (MAPE), Theil's U statistic, and highest correct directional change (CDC) values of all the fitted models obtained for both the training and testing sets (Table 5) considered, one can infer that machine learning techniques, NLSVR and ANN outperformed over the stochastic models. Rathod *et al.* (2018) aimed to forecast India's oilseed production using traditional and AI models, and they also found that AI models, including SVR, performed best.

4.4 Model Comparison:

The comparative analysis of palm oil import volume forecasting models reveals a distinct performance hierarchy, with the Support Vector Regression (SVR) model demonstrating the most accurate predictions. SVR's lower Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) values in both training (10.65 and 39.61) and testing datasets (29.01 and 11.09) and higher Correct Directional Change (CDC) percentages (93.85 per cent in training and 100 per cent in testing) underscore its robustness. This superior performance can be attributed to SVR's ability to manage complex, non-linear data patterns effectively, which is evident in fluctuating palm oil import volumes. Such efficacy accentuates the potential of machine learning models in handling forecasting challenges, especially in sectors characterized by high volatility and unpredictability. Similar results were found by Paul *et al.* (2019) in their study of Prediction of early blight severity in tomato by machine learning technique.

The SVR model's impressive performance is further highlighted by its Theil's U statistic of 0.03 in the training phase and 0.07 in the testing phase, indicating high forecasting reliability. This is contrasted with the Artificial Neural Network (ANN), which, despite its advanced capabilities, showed Theil's U values of 0.11 in training and 0.15 in testing, pointing to a lesser degree of precision compared to SVR.

In contrast, while useful, traditional time series models like ARIMA and GARCH fell short in accuracy. The GARCH model generally outperforms ARIMA across several metrics. In terms of accuracy, GARCH exhibited lower Root Mean Squared Error (RMSE) values for both training (57.71) and testing (96.86) datasets compared to ARIMA (61.74 and 102.54, respectively). Similarly, GARCH shows lower Mean Absolute Percentage Error (MAPE) values for training (294.75 per cent) and testing (68.19 per cent) compared to ARIMA (484.44 per cent and 102.54 per cent, respectively), indicating better overall accuracy in percentage terms. However, SVR performed well compared to all the models presented in this study. For instance, ARIMA, GARCH, and ANN model performance metrics, including RMSE and MAPE values, were significantly higher than SVR, reflecting their limitations in capturing the

complex dynamics of palm oil imports. These results suggest a paradigm shift in forecasting methodologies, urging a transition towards more sophisticated, data-driven approaches like machine learning for enhanced predictive accuracy. Consequently, this shift holds profound implications for strategic decision-making in agricultural imports, potentially influencing policy-making and economic planning at national levels.

IV

CONCLUSIONS

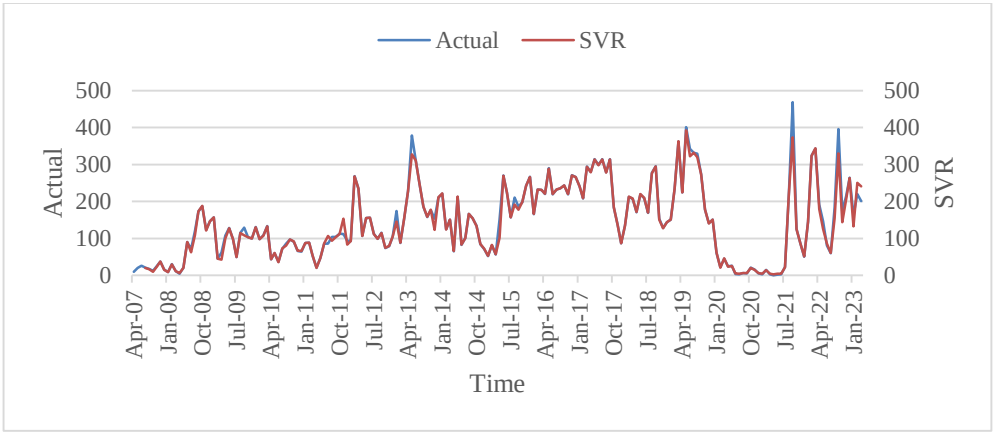
Time series models play a crucial role in forecasting, yet they may fall short when dealing with highly volatile data. In that case, machine learning models like artificial neural network and support vector regression can effectively improve forecasting performance. The distribution of prediction accuracy for the models chosen (ARIMA, GARCH, ANN, and SVR) for predicting India's imports of palm oil is given in Table 6. SVR is particularly good at making accurate predictions with a high prediction accuracy rate of 84.66 per cent over 95 per cent, compared to 62.43 per cent for ANN in the same range. ANN performs better than other models with 9.52 per cent accuracy within the 90-95 per cent range. Particularly below 75 per cent, where GARCH predicted 14.58 per cent and 21.88 per cent of data points, respectively, GARCH and ARIMA showed strengths in lower precision predictions. A thorough analysis of the model's performance across a range of accuracy thresholds is given in this table, which also offers insightful information about how well the models forecast trends in India's import of palm oil. The actual versus predicted values of all the models are displayed in Figure 4.

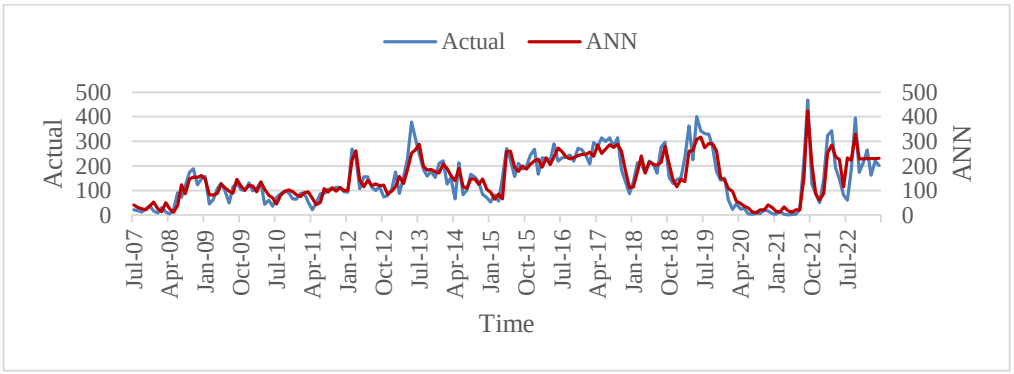
TABLE 6. DISTRIBUTION OF PREDICTION ACCURACY OF THE MODELS SELECTED FOR PALM OIL IMPORT TO INDIA

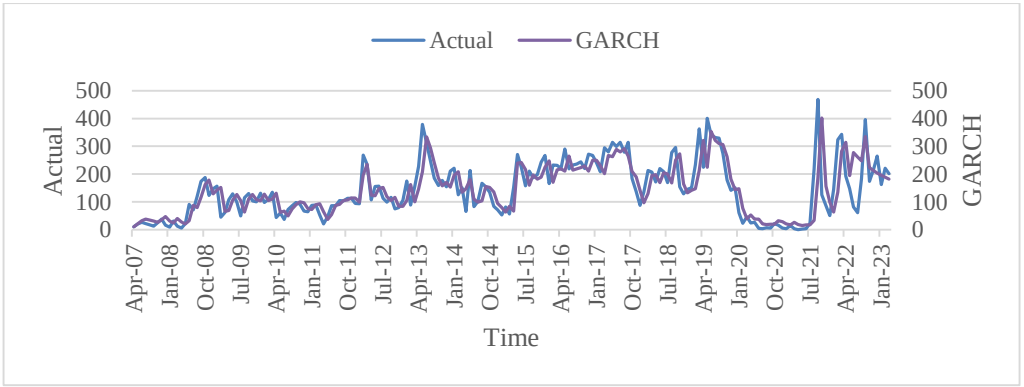
(per cent)

Percent prediction accuracy	Percent total of data points predicted			
	SVR	ANN	GARCH	ARIMA
(1)	(2)	(3)	(4)	(5)
Above 95	84.66	62.43	58.33	55.73
90-95	6.88	9.52	8.85	4.69
85-90	2.64	7.94	6.77	7.81
80-85	4.23	8.99	5.73	5.21
75-80	0.53	3.70	5.73	4.69
< 75	1.06	7.41	14.58	21.88

It is evident from Table 5 that Support Vector Regression (SVR) stands out as the most accurate predictor, boasting the lowest Root Mean Squared Error (RMSE) values, lowest Theil's U statistic, and highest CDC values for both training (10.65, 0.03, 93.85) and testing (29.01, 0.07, 100) data respectively and This signifies that SVR consistently provides more precise forecasts with substantially smaller errors in comparison to the other models. Therefore, for accurate predictions of India's palm oil


Actual
SVR
500
400
300
200
100
0
Actual
SVR
Apr-07
Jan-08
Oct-08
Jul-09
Apr-10
Jan-11
Oct-11
Jul-12
Apr-13
Jan-14
Oct-14
Jul-15
Apr-16
Jan-17
Oct-17
Jul-18
Apr-19
Jan-20
Oct-20
Jul-21
Apr-22
Jan-23
Time


Actual
ANN
500
400
300
200
100
0
Actual
ANN
Jul-07
Apr-08
Jan-09
Oct-09
Jul-10
Apr-11
Jan-12
Oct-12
Jul-13
Apr-14
Jan-15
Oct-15
Jul-16
Apr-17
Jan-18
Oct-18
Jul-19
Apr-20
Jan-21
Oct-21
Jul-22
Time


Actual
GARCH
500
400
300
200
100
0
Actual
GARCH
Apr-07
Jan-08
Oct-08
Jul-09
Apr-10
Jan-11
Oct-11
Jul-12
Apr-13
Jan-14
Oct-14
Jul-15
Apr-16
Jan-17
Oct-17
Jul-18
Apr-19
Jan-20
Oct-20
Jul-21
Apr-22
Jan-23
Time

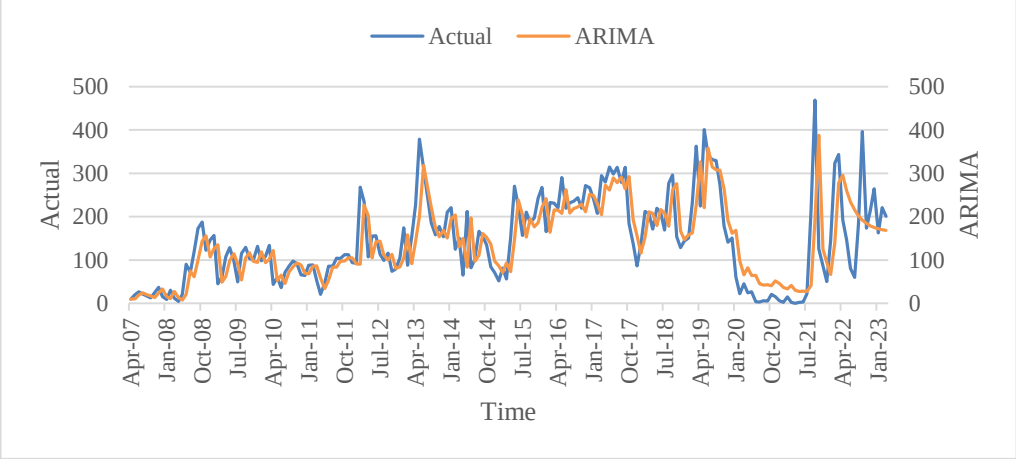


Figure 4. Performance Comparison of Different Models

import volume, the SVR model is the preferred choice. Researchers and policymakers may consider using SVR as a reliable tool for making informed decisions in the context of palm oil import forecasting in India.

Received March 2024. *Revision accepted May 2024.*

REFERENCES

Bhattacharya, R., Pandey, R., Patnaik, I., & Shah, A. (2016). Seasonal adjustment of Indian macroeconomic time-series, 16/160.

Box, G., & Jenkins, G. (1970). Time-series analysis: Forecasting and control. San Francisco, CA: Holden-Day Press.

Brock, W. A., Scheinkman, J. A., Dechert, W. D., & LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econometric Reviews*, 15(3), 197-235.

David, M. (2017). E1071: Misc functions of the department of statistics, probability theory group. R package version 1: 6-8.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a), 427-431.

Gohain, K. (2021). Evaluation of Theils U: A naïve forecast application. *Quantum Journal of Engineering, Science and Technology*, 2(5), 26-31.

Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., Kuroptev, K., O'Hara-Wild, Petropoulos, F., M., Razbash, S., Wang, E. &Yasmeen, F. (2022). Forecasting Functions for Time Series and Linear Models (Version 8.18). Retrieved from https://cran.r-project.org/web/packages/forecast/index.html

Indhushree, A., & Shivakumar, K. M. (2020). Dynamics of palm oil import on prices, income and trade of Indian edible oil sector. *Indian Journal of Agricultural Economics*, 75(4), 578-586.

ITC Trade map (2023). Trade statistics for international business development. Retrieved from www.trademap.org

Khalid, N., Hamidi, H. N. A., Thinagar, S., & Marwan, N. F. (2018). Crude palm oil price forecasting in Malaysia: An econometric approach. *Jurnal Ekonomi Malaysia*, 52(3), 263-278.

Lama, A., Jha, G. K., Gurung, B., Paul, R. K., Bharadwaj, A., & Parsad, R. (2016). A comparative study on time-delay neural network and GARCH models for forecasting agricultural commodity price volatility. *Journal of the Indian Society of Agricultural Statistics*, 70(1), 7-18.

Ministry of Agriculture & Farmers Welfare (2022). Agricultural Statistics at a Glance. Retrieved from https://desagri.gov.in/wp-content/uploads/2023/05/Agricultural-Statistics-at-a-Glance-2022.pdf

Ministry of Finance (2023). Economic Survey 2022-23. Retrieved from https://www.indiabudget.gov.in/economicsurvey

Muharni, Y., & Denisa, A. (2021). The application of artificial neural network for quality prediction of industrial standard water. In IOP Conference Series: Earth and Environmental Science, 926(1), 012048. IOP Publishing, November.

National Mission on Edible Oils (2022). NMEO-OP guidelines. Retrieved from https://nmeo.dac.gov.in/nmeodoc/NMEO-OPGUIEDELINES.pdf

Pannakkong, W., Huynh, V. N., & Sriboonchitta, S. (2019). A novel hybrid autoregressive integrated moving average and artificial neural network model for cassava export forecasting. International Journal of Computational Intelligence Systems, 12(2), 1047-1061.

Paul, R. K., Vennila, S., Bhat, M. N., Yadav, S. K., Sharma, V. K., Nisar, S., & Panwar, S. (2019). Prediction of early blight severity in tomato (Solanum lycopersicum) by machine learning technique. Indian Journal of Agricultural Sciences, 89(11), 1921-1927.

Paul, R. K., Prajneshu, P., & Himadri Ghosh, H. G. (2009). GARCH nonlinear time series analysis for modelling and forecasting of India's volatile spices export data. Journal of the Indian Society of Agricultural Statistics, 63, 123-131.

Petrevska, B. (2017). Predicting tourism demand by ARIMA models. Economic Research-Ekonomska Istraživanja, 30(1), 939-950.

Rakhmawan, S., Notodiputro, K. A., & Sumertajaya, I. M. (2015). A Study of ARIMA and GARCH models to forecast crude palm oil (CPO) export in Indonesia. in Proceeding of international conference on research, implementation and education of mathematics and sciences 2015 (ICRIEMS 2015), Yogyakarta State University, 17-19 May.

Rakshit, D., Paul, R. K., & Panwar, S. (2021). Asymmetric price volatility of onion in India. Indian Journal of Agricultural Economics, 76(2), 245-260.

Rathod, S., Singh, K. N., Patil, S. G., Naik, R. H., Ray, M., & Meena, V. S. (2018). Modeling and forecasting of oilseed production of India through artificial intelligence techniques. *Indian Journal of Agricultural Sciences*, 88(1), 22-27.

Rathod, S., Singh, K. N., Paul, R. K., Meher, S. K., Mishra, G. C., Gurung, B., Ray, M. & Sinha, K. (2017). An improved ARFIMA Model using Maximum Overlap Discrete Wavelet Transform (MODWT) and ANN for forecasting agricultural commodity price. Journal of the Indian Society of Agricultural Statistics, 71(2), 103-111.

Sagar, H. S. C., Mabano, A., Roopa, R., Sharmin, M., Richard, F. J., & Clause, J. (2019). India in the oil palm era: Describing India's dependence on palm oil, recommendations for sustainable production, and opportunities to become an influential consumer. Tropical Conservation Science, 12, 1940082919838918.

Septyari, F. M. (2021). Grey forecasting of the exports of Indonesian palm oil to India. International Journal of Grey Systems, 1(2), 60-68.

The World Bank (2023). World bank indicators database. Retrieved from https://data.worldbank.org

Theil, H., Beerens, G. A. C., Tilanus, C. G., & De Leeuw, C. B. (1966). Applied economic forecasting, 4, Amsterdam: North-Holland.

Upadhyay, V. K. (2013). Modelling and forecasting export and import of Indian wood-based panel using ARIMA models. Elixir Stat, 63, 18145-18148.

Vapnik, N. V. (1998). Statistical learning theory. Wiley-Interscience.

Weerasinghe, W. P. M. C. N., & Jayasundara, D. D. M. (2021). Modelling pepper export income in Sri Lanka using deterministic decomposition and seasonal ARIMA models. Statistics and Applications, 19(2), 89-100.