Research Article

International Journal of Advanced Technology and Engineering Exploration, Vol 10(98)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2021.876382

An extrapolative model for price prediction of crops using hybrid ensemble learning techniques

G. Murugesan[1*] and B. Radha[2]
Assistant Professor, Department of Artificial Intelligence and Machine Learning, Sree Saraswathi Thyagaraja College, Pollachi, Tamil Nadu, India-642107[1]
Associate Professor, Department of Information Technology and Cognitive Systems, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu, India-641008[2]

Received: 13-July-2022; Revised: 10-January-2023; Accepted: 12-January-2023

Abstract
Agriculture is the basis for food and the backbone of a country's economy. In India, around 70% of the population is actively involved in growing crops for food or providing direct raw materials to a variety of industries, including textile, food processing, and non-agricultural sectors. The development of technology aids the agricultural sector in forecasting a variety of factors, including crop quality, disease detection, and soil quality, to increase crop yield. However, increased agricultural yield may not always result in a profit due to price reductions. Thus, price forecasting is crucial before choosing the crop to plant since it aids in making informed choices that reduce the risk and loss associated with market price instabilities. This study provides a hybrid model for price prediction that combines an autoregressive integrated moving average (ARIMA) model, a linear statistical analysis for time series data, and an ensemble machine learning approach using support vector regression (SVR). The work has three models: 1) a statistical model is applied over the input features related to crop price, and the residuals are evaluated using SVR, 2) SVR is applied over the predicted price from the statistical model along with the other input features, 3) SVR is applied to the results obtained from the statistical model and its residuals, in addition to the input features. After analysing the results, the model for price forecasting that produced better outcomes was finally chosen. The experimental and result analysis reveals that model 3 has improved results with a 13.37% deviation from actual observation compared to models 1 and 2, which have resulted with deviations of 14.68% and 16.48%, respectively. Additionally, compared to other models, the suggested model has the lowest average prediction errors and average divergence from actual values. Thus, the proposed model is suitable for reliable price forecasting and optimal performance.

Keywords
Agriculture, Crop price prediction, ARIMA, Support vector regression, Ensemble model, Machine learning technique.

1.Introduction
As agriculture is the backbone of the country, agricultural marketing is the basis for most of the nation's economic activities [1]. In addition to increasing production and consumption through farming, it also speeds up the development of an economy [2, 3]. In general, the goal of agricultural marketing is to get goods from farms to customers [4]. The forecasting of prices for agricultural commodities is always a major issue faced by farmers. The price for the commodity decreases after the cultivation of crops, and so they are not able to get the anticipated amount. The severe impact of this problem increases the suicide rate every year [5].

According to the report given by the National Crime Record Bureau (NCRB), in India, the number of suicides in the agricultural sector increased by 18% in the year 2020 compared to the year 2019 [6]. For many agricultural commodities, such as tomatoes and onions, the fluctuations in price are more irregular.

Moreover, these commodities cannot be preserved for a long time like turmeric or paddy, and so the farmers cannot store them until the optimal price is perceived in the market [7]. Thus, the arrival of commodities on the market directly affects the price since a decrease in supply increases the price and an increase in supply decreases the price [8]. Apart from the arrival of agricultural commodities, various other factors influence the price of the commodities, such as

*Author for correspondence

demand, climatic conditions, temperature, supply of production, transportation issues, and so on [9]. The Indian government has taken a number of steps to offer information about markets and commodity pricing to farmers and other users in order to give them a broad understanding of market prices. Utilizing the information effectively may provide greater insight into the risks involved and have a positive impact on making decisions. However, gathering, analysing, and turning the data into valuable knowledge is more difficult. Thus, data mining, machine learning and statistics play a huge role in extracting knowledge and predicting the future price of the commodities precisely [10, 11]. Several statistical models are available in the literature that can be used in forecasting time series data, among which the most popular and widely used techniques are autoregressive moving average (ARMA) [12], autoregressive integrated moving average (ARIMA) [13], simple exponential smoothing (SES) [14], naïve forecasting (NF) [15], Holt's method [16] and more. Many machine learning algorithms, such as linear regression (LR) [17], Gaussian process (GP) [18], neural networks [19], support vector machines (SVM) [20], multilayer perceptron (MP), and fuzzy logic [21], are also widely used in forecasting prices. However, most of the existing models face several challenges while forecasting prices. Several models are available in the literature for predicting the price of agricultural commodities, nevertheless many of them are only appropriate for linear data. These models provide better results with dataset having less instances. Thus, the significant problem with these models is the forecast accuracy, which is negatively associated with the volume of past data [11]. Another significant flaw in the models is that they were developed for a single commodity, hence their applicability to a variety of commodities has not been examined.

Researchers that use statistical models in conjunction with machine learning algorithms have recently given hybrid models a great deal of attention because of their enhanced performance [22]. These models are intended to interact with the data in various ways in order to complement and enhance one another. As a result, methods used in hybrid models can deal with problems collectively that they were not designed to manage individually. This gives rise to the idea of developing a hybrid price prediction model that performs better than other techniques now in use. Consequently, the primary goal of the study is to propose a reliable price prediction model with better performance and minimum prediction errors or price deviations that is suitable for a variety of commodities. This paper proposes a hybrid model that makes use of ARIMA, which analyses the statistical data linearly for predicting price from the time series data and then applies a machine learning based ensemble model by utilising support vector regression (SVR). The entire framework of the proposed model has three forms. The first model applies a machine learning model to the residuals obtained from the ARIMA model, and the results of both models are summed for a final prediction. The second model applies SVR over the forecasted price from ARIMA along with the other input features, and the final predicted price is the result obtained from the SVR model, whereas the third model applies SVR to the input features along with the forecasted price and residuals from ARIMA, and the final result is the sum of the predicted values of the forecasted price and forecasted residuals from the SVR. Finally, the results obtained are compared, and the improved results are chosen for price forecasting. An extensive result investigation conducted to analyse the performance of the models reveals that the results significantly outperform other existing models.

The organisation of the paper is as follows: The detailed review of literature related to the study of forecasting prices for agricultural commodities is summarised in Section 2. Section 3 discusses the background study and the proposed hybrid model in detail. Section 4 presents the experimental study carried out, and the results obtained from the experiments are analysed and compared with the other models. Section 5 discusses the findings of the results analysis. Finally, Section 6 concludes the work and discusses the scope for future research.

2.Literature review

Due to the tremendous progress made in previous decades, enormous research has been conducted on predicting agricultural commodity prices. Market analysis and subjective probability were used in the preliminary research to forecast prices. Later quantitative models, including statistical models, time series analysis, and intelligent analysis, come into existence for predicting future trends [23]. Statistical models for time series analysis are simple methods that look at history or past data to predict future trends. Though this model seems to be simple, it may not be used widely for all applications since it does not consider other factors that influence the price other than time. Variations of seasonal exponential smoothing, like quadratic smoothing, the Holter-Winters method, variations of ARIMA models, and moving average (MA) models belong to this category [24, 25].

For forecasting prices through intelligent models, researchers often use various artificial neural networks [26]. For price prediction, a back propagation neural network (BPNN) with a genetic algorithm was proposed [27]. For forecasting the price of soybeans in China, a radial basis function neural network (RBFNN) was used with quantile regression and parameter optimization [28]. A hybrid model that utilises seasonal-trend decomposition and extreme learning machines (ELMs) for seasonal forecasting of the vegetable price in China was anticipated. Though the model has a complex procedure, it offers better prediction results [29]. A feed-forward time-delay neural network (TDNN) for time series forecasting of oilseeds in India was made, and the results were compared with the ARIMA model, which shows that the model offers improved results for linear data with a clear insight on price change [30]. A long short-term memory (LSTM) neural networks were applied for seasonal forecasting of Arecanuts prices, and the results were compared with the seasonal ARIMA and holt-winters (HW) seasonal model in which the LSTM offers improved prediction [31].

The models that are highly influential in predicting the numerical values, including price, are SVM, SVR, and their variations [23]. In general, the hybrid models were proposed by using SVR, in which the residuals of linear models were applied over the nonlinear SVR model. For predicting the future price fluctuations for cotton, fuzzy information granulation was applied to transform the data, and SVR was applied after optimising the parameters using particle swarm optimization (PSO) [32]. A similar paradigm was put forth in which genetic algorithms were used to tune the SVR hyperparameters [33]. A combination of fuzzy information granulation and SVM was utilised in predicting the price of the commodity, and the parameter optimization was made using the mind evolutionary algorithm (MEA). However, the model possesses the deficiency of underestimation [34].

Recently, the hybrid model has become quite popular in the machine learning domain due to its performance effectiveness. A hybrid model that utilises seasonal-trend decomposition-based preprocessing and LSTM for forecasting vegetable prices in Korea was suggested [35]. LSTM is considered to be the most effective model for predicting prices [36, 37]. A comparison of various machine learning models such as LR, decision trees, and extreme gradient boosting (XGBoost)-based regression was analyzed, and the results proved that XGBoost offers good prediction [38]. Apart from these models, the other models used in the various applications of forecasting were also studied, including the combination of linear and nonlinear models or ensemble machine learning models such as ARIMA with ELM and SVM for wind speed forecasting [39], ensemble ARIMA with an adaptive neuro-fuzzy inference system for predicting energy consumption [40], power load prediction using LSTM and XGBoost [41], a probability-based fuzzy ARIMA model for consumption prediction [42], and so on.

Even though there are a lot of models suggested in the literature for predicting the price of agricultural goods, many of them are only good for linear data. Another main drawback of these models is their accuracy, which is directly proportional to the amount of historical data used. Also, they provide accurate predictions only over a shorter period. Another notable drawback of the models is that they are built for a specific commodity, so their suitability for a wide range of commodities has not been analyzed. With this knowledge and motivation to overcome the drawbacks of the existing models, the proposed work offers a hybrid model that makes use of the combination of linear and nonlinear models such as ARIMA and SVR for improving prediction accuracy and minimising forecasting errors in predicting prices specifically for the agricultural sector.

3.Methods

The proposed work employs ARIMA and SVR models that offer improved performance with LR and non-linear regression (NLR) respectively. However, many researchers recorded that the use of disparate models often provided improved results, especially in regression problems [43]. Thus, a hybrid strategy is proposed that combines ARIMA and SVR models for better price forecasting than the individual models.

3.1Study background

ARIMA is a statistical model used in regression analysis to comprehend time series data or forecast future values based on lags. The term “lag” refers to a constant amount of passing time. It is frequently used in forecasting price or financial markets by assuming that the future will be similar to the past. The model is outlined with three components (AR, I, MA) to make them fit in the data, indicating 1) the variable is regressed on its own past lag, 2) the data is stationary and the difference in the non-seasonal observations, and 3) the forecast errors as linear combinations of errors from the past respectively. Thus, the model can be represented in a common notation as ARIMA (p, d, q), where p represents the number of AR terms or number of lag observations, q represents the number of regression errors and refers to the size of the moving window, and d is the degree of differencing. The model can be built with four basic steps, including model identification, parameter estimation, model evaluation, and forecasting [40, 44].

SVR is a regression technique based on the SVM principles used in classification problems. It supports both linear and nonlinear data. The main idea of SVR is to identify the best fit line, thereby approximating the best value inside the given boundaries. The overall aim is to reduce error by personalising the hyperplane, which increases the margin [45].

Generally, the hybrid model has the ability to capture the data behaviour in the linear and nonlinear domains by enlightening the overall performance of the prediction model. This can be achieved by applying linear and nonlinear models sequentially.

More specifically, the linear model is applied to the given input data that ends with an outcome and residuals, and then the nonlinear model is applied to the residuals obtained from the linear model that ends with an outcome, which is then added to the outcome of the linear model [46]. The prediction using a hybrid model can be represented as in Equation 1.

$$y_t = L_t + N_t \quad (1)$$

Where Y_t is the final forecasted value at time t that is obtained by summing the outcomes of the linear model L_t and the nonlinear model N_t. The residuals of the linear model are represented as ε which can be obtained by subtracting the forecasted value from the actual value as given in Equation 2.

$$\varepsilon_t = x_t - l_t \quad (2)$$

Here, ε_t is the residual at time t which is obtained by subtracting the actual value x_t from the linearly forecasted value l_t using a linear model $L_t$at time t.

With this basic knowledge of the hybrid model acquired from the previous research works, the proposed model utilizes an ensemble model that makes use of three hybrid strategies for predicting the price of the crops. The first model applies the SVR to the ARIMA residuals (ε_t) obtained on implementing the ARIMA and finally, the linear prediction of ARIMA (l_t) and outcome of SVR are summed to predict the values. The second version applies SVR to the outcome of the ARIMA model (l_t) with other input variables to identify the final prediction and finally, the third model applies SVR to the residuals (ε_t) and outcome (l_t) of the ARIMA model with other input variables for the final prediction in which the final forecasted values are the sum of the SVR forecasted value and residual. The overall idea of the proposed ensemble forecasting model is presented in *Figure 1*.

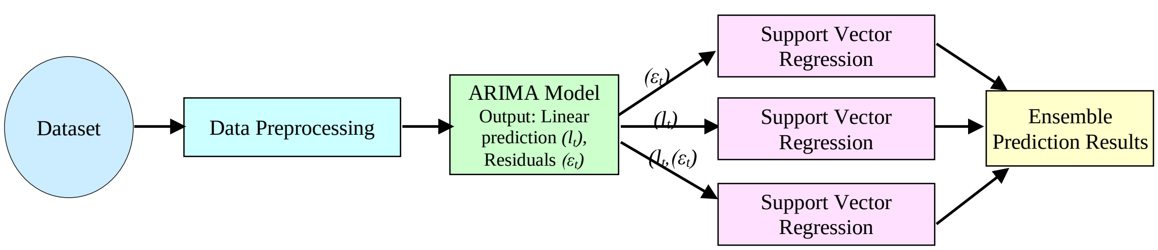


Figure 1 Overall idea of the proposed forecasting model

The initial phase of the model is to collect data about crop prices from previous years, which will be employed to predict the price of crops in the future. The data is pre-processed in such a way by removing the missing values, and then the feature values are scaled to avoid the dominance of larger ranges on lower ranges. The data normalisation used in the proposed model is given in Equation 3.

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3)$$

The next step is to apply the ARIMA model, which is common for all three models used in the study. The ARIMA model can be applied to the given input data by determining the values for the parameters p, q and d that end with the forecasted value (l) specifying the linear part and the error residuals (ε) identified by subtracting the forecasted value from the actual value as in Equation 2.

3.2Model 1
The first model is the traditional hybrid model, which applies the ARIMA model to the input dataset to predict the price linearly. The SVR, which works with NLR, is used on the ARIMA model's linear residuals. The residuals are partitioned into training and test sets, and the SVR model is trained using the residuals in the training set with a radial basis function as a kernel. By adding up the two predictions from ARIMA and SVR, the test set is used to make the final prediction. This model with residuals as input is given in Equation 4.

$$y_t = l_t + f(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-n}) + \Delta_t \quad (4)$$

Here y_t is the final forecasted value at time *t*, l_t is the linear part of the forecasted value using ARIMA at time *t*, *f* denotes the nonlinear SVR function, ε is the residual from the past generated by the ARIMA model, Δ_t is the random error. However, since the input values are scaled, the forecasted value obtained from this model is also scaled which must be rescaled to its original value as $x_{scaled} \times (x_{max} - x_{min}) + x_{min}$.

3.3Model 2
The second model is a popular hybrid strategy that is often used to add NLR to the ARIMA model. Here, the ARIMA linearly predicted values are used with the SVR model instead of the ARIMA residuals. Thus, the predicted outcomes from the ARIMA model are passed as one of the features along with the other input features, which are then partitioned into training and test sets, and the SVR model is trained using training samples and tested using the test set. Here, the outcome obtained from the SVR is the final forecasted price. This model of using the forecasted values from the linear model as input for the nonlinear model is given in Equation 5.

$$y = f(\gamma_{t-1}, \varepsilon\gamma_{t-2}, \dots, \gamma_{t-n}) + \Delta_t \quad (5)$$

Here, $y_t$denotes the final forecasted value at time *t*, γ represents all the input features from the input dataset in addition to the predictions *l* obtained from the linear model ARIMA in the past before time *t*, *f* denotes the nonlinear SVR function and Δ_t is the random error.

3.4Model 3
The third model in the ensemble technique employs the combination of both residuals and forecasted outcomes from the ARIMA model as the input for the SVR model.

Instead of applying either linear residuals or a linearly forecasted value, both of them are included as separate features along with the other input models and fed as an input for the SVR model.

This data is partitioned into the training set and test set in which the training set is used to train the SVR model and the test data is applied to predict the final values for the test samples. This model operates both the outcome and residuals of the linear model as input using a nonlinear model as given in Equation 6.

$$y_t = f(\gamma_{t-1}, \varepsilon\gamma_{t-2}, \dots, \gamma_{t-n}) + f(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-n}) + \Delta_t \quad (6)$$

Here, y_t is the final forecasted value at time *t*, γ represents all the input features from the input dataset in addition to the predictions *l* obtained from the ARIMA from past before time *t*, ε is the residual from past generated by the ARIMA model and Δ_t is the random error. Thus, the results obtained from the SVR model have two parts of which one is the forecasted values and the other is the forecasted residuals. Eventually, the final forecasted value is the sum of the forecasted value and the forecasted residual which is then rescaled to its original value.

The results obtained from the models are evaluated using various performance metrics, and the model with the improved prediction can be used for the prediction of prices. The detailed representation of the proposed ensemble forecasting model is presented in *Figure 2*.

To summarize, the ARIMA model with nonlinear values as input provides linear prediction outcome along with the non-linear part known as residuals. Thus, the SVR, a machine learning algorithm is applied to the output of ARIMA model in different ways.

In the proposed model 1, the SVR model is applied to the ARIMA residuals and the linear ARIMA prediction and SVR prediction is summed to provide final prediction. In the model 2, the final predicted value is the outcome of SVR model that is applied on the dataset with linearly predicted outcome of ARIMA. The final predicted value in the model 3 is obtained by applying SVR model on the dataset along with the ARIMA results i.e., linear predicted values and residuals. The working procedure of the models is depicted in *Figure 3*.

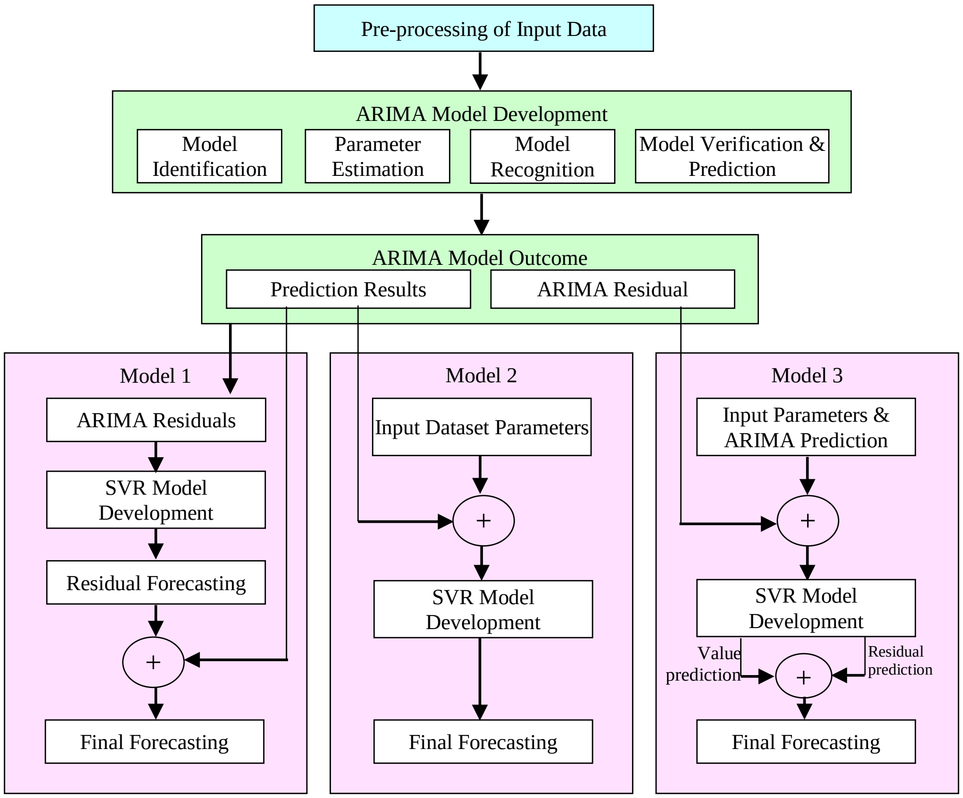


Figure 2 Detailed framework of the proposed forecasting model

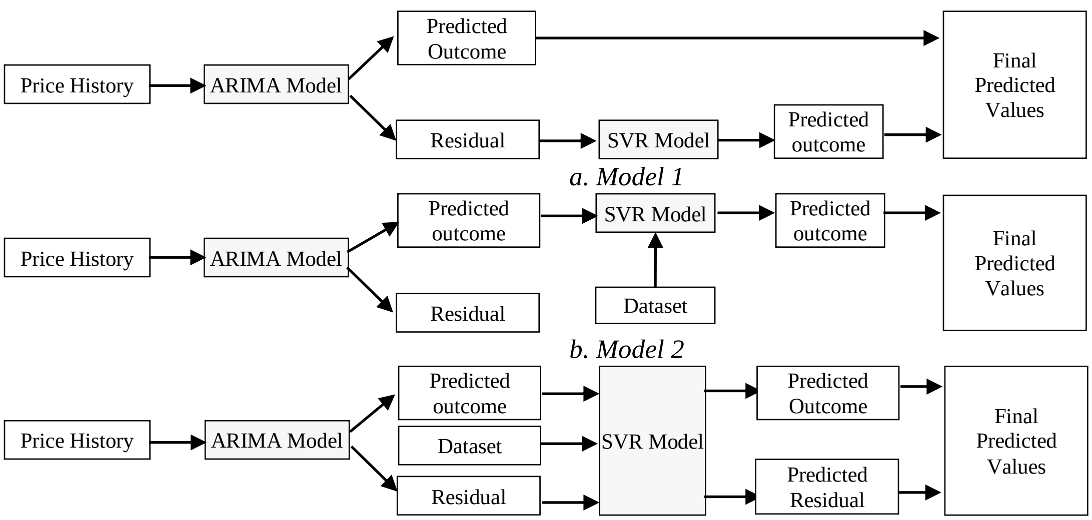


Figure 3 Working procedure of proposed hybrid models

4.Results

This section presents the experiments carried out for the proposed model and the analysis of the results obtained for the model, along with a comparison of the results with those of other existing models using various performance techniques. The work was carried out on a system with the following hardware configurations: an Intel(R) Pentium (R) Processor @ 2.40 GHz, 8 GB of RAM, and a 64-bit Windows operating system. The software used in the analysis is R programming for implementing the ARIMA model and the WEKA tool to analyse the machine learning part of the proposed model.

4.1Dataset used

For forecasting the price trend, the dataset is collected from the website of the Directorate of Marketing and Inspection, Department of Agriculture and Cooperation, Ministry of Agriculture and Farmers Welfare, Government of India, developed by the national informatics centre (NIC) (https://agmarknet.gov.in/) [47]. This site offers various information for a wide range of users, including farmers, customers, dealers, and researchers. The reliable and complete details about the arrivals and the prices of various commodities, including cereals, pulses, vegetables, and fruits, from all over the markets covering all the districts in all the states of India. From January 2012 to March 2022, 10 years of price data for the commodity coconut were collected in various markets in Coimbatore, Tamil Nadu, including Pollachi, Annur, Anaimalai, Karamadai, Sulur, Palladam, Kinathukadavu, Thondamuthur, and Udumalpet. This dataset contains 18489 instances with 5 attributes, including district name, market name, commodity, prices per quintal (Q), minimum price, maximum price, modal price, and price date. The details of the attributes are given in *Table 1*. However, the numeric attributes such as minimum price, maximum price, and modal price, as well as price date, are used in analysis. *Table 2* shows a sample of ten records retrieved for the commodity Coconut from the Pollachi market in Coimbatore.

Table 1 Attribute description for coconut dataset

Attributes	Data type	Parameters/Range
Market Name	Categorical	Pollachi, Annur, Anaimalai, Karamadai, Sulur, Palladam, Kinathukadavu, Thondamuthur, Udumalpet
Commodity	Categorical	Coconut
Minimum Price	Numeric	400-10100
Maximum Price	Numeric	450-20000
Modal Price	Numeric	425-10400
Price Date	Temporal	2-Jan-12 to 31-Mar-22

Table 2 Samples of the dataset used for the analysis

S. No.	Market name	Min (Rs./Q)	Max (Rs./Q)	Modal (Rs./Q)	Price Date
1	Anaimalai	400	600	500	11-May-12
2	Sulur	400	600	500	15-Sep-12
3	Kianthukadavu	3400	3600	3500	30-Aug-18
4	Karamadai	3000	3500	3250	4-Jan-18
5	Anaimalai	10100	10700	10400	27-Feb-19
6	Thondamuthur	4000	4200	4100	27-Nov-17
7	Palladam	5000	5500	5225	27-Jun-16
8	Annur	8000	12000	10000	4-Oct-21
9	Pollachi	3400	3600	3500	1-Jun-18
10	Udumalpet	2800	2900	2850	9-Nov-20

4.2Evaluation metrics

To perform the analysis of the results obtained for the proposed model and to compare the results with the results of other models, various performance metrics specifically used in forecasting applications are utilized. It includes 1) the mean absolute error (MAE) that measures the average errors between the forecasted and actual values, 2) root relative squared error (RRSE), calculated as the square root of the sum of squared errors to the squared errors of a simple model, 3) direction accuracy (DA) representing forecasting accuracy 4) relative absolute error (RAE), defined as the total absolute errors divided by the total absolute errors of a simple model, 5) mean absolute percentage error (MAPE), defined as the average absolute percentage error for each period divided by the actual values, 6) root mean squared error (RMSE) that measures the standard deviation of forecasted errors, 7) mean squared error (MSE) that measures the mean of squares of errors and 8) execution time (ET) that denotes the time in seconds taken by the model for training and testing. The formulas to compute these metrics are shown in Equations 7–13.

$$MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} \tag{7}$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - x_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \tag{8}$$

$$DA = \frac{1}{N}\sum_t 1_{sign(x_t - x_{t-1}) = sign(y_t - x_{t-1})} \tag{9}$$

$$RAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{\sum_{i=1}^{n}|x_i - \bar{x}|} \tag{10}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_i - y_i}{x_i}\right| \quad (11)$$

$$MASE = mean\left(\frac{1}{\frac{1}{T-m}\sum_{t=m+1}^{T}|y_t - y_{t-m}|}\right) \quad (12)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2 \quad (13)$$

where x_i is the actual values, y_i is the predicted values, n is the number of observations and $\bar{x}$ is the mean of actual values.

4.3ARIMA

The ARIMA model that performs the forecasting on the linear data has been implemented using R programming for forecasting the prices [48]. The analysis has been performed by implementing the proposed model on 2688 entries of coconut prices in the Pollachi market. Initially, the given monthly average price dataset is converted to time series data, and the values are presented in *Figure 4.* Here, the average price for each month from the years 2012 to 2020 is considered the training dataset. The summary of the training set is given in *Table 3,* in which the minimum price is reported for December 2012 and the maximum price for December 2020. The mean of the observations is 1871.2, and the median is 1775.

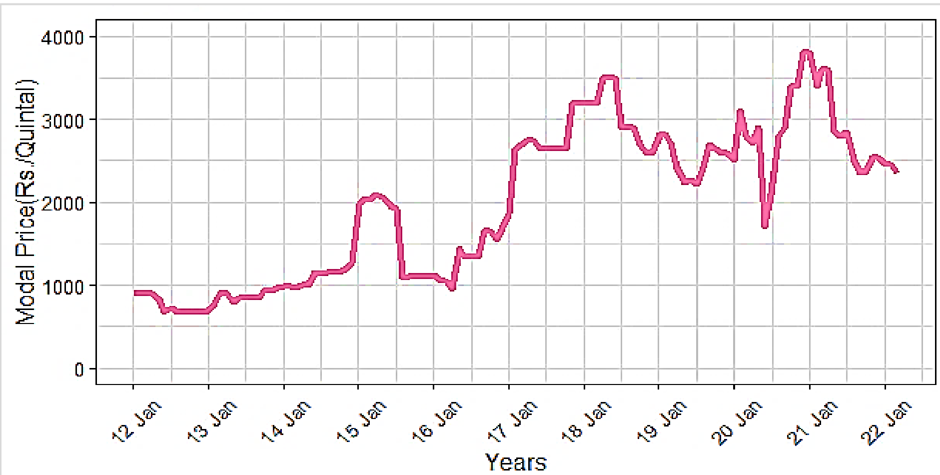


Figure 4 Monthly average price of coconut for the past 10 years

Table 3 Summary of converted time series dataset

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2012	900	900	915	900	825	675	720	680	680	680	680	670
2013	690	780	915	900	795	840	855	855	850	930	935	975
2014	980	990	970	1000	1000	1150	1140	1150	1155	1150	1190	1275
2015	1975	2050	2025	2090	2040	1960	1920	1090	1100	1110	1100	1110
2016	1115	1050	1050	950	1450	1350	1350	1350	1650	1660	1550	1700
2017	1850	2650	2700	2750	2750	2650	2650	2650	2650	2650	3200	3200
2018	3200	3200	3200	3500	3500	3500	2900	2900	2900	2700	2600	2600
2019	2810	2810	2720	2430	2230	2260	2210	2420	2700	2650	2600	2600
2020	2500	3100	2800	2710	2900	1700	2100	2800	2900	3400	3400	3800

Minimum: 670, Median: 1775.0, Mean: 1871.2, Maximum: 3800

The exploratory analysis has been carried out on the components of the time series dataset as trends, indicating the long-term increase or decrease of the price, and seasonal, indicating the monthly or yearly pattern in price, along with the random and observed price. This is plotted as a graph in *Figure 5*. The graph shows that the dataset is suitable for applying the ARIMA model since it satisfies the stationarity requirement, indicating that the data is not dependent on time.

The analysis of the time series data has been carried out, in which the trend line is evaluated for the price of the commodity and the evaluation of data based on seasonality has been made. The obtained results are shown in *Figure 6.* The trend line depicted on the left side graph represents the price's high in 2015, decrease in 2016, and increase in 2018. Similarly, the right-side box plot in the figure shows that the price typically decreases in June and July and increases in February each year. To attain stationarity, the unit root test using the kwiatkowski-phillips-schmidt-shin (KPSS) test is applied, and the results show that the data is stationary against the unit root. The results obtained from the KPSS test are presented in *Figure 7*.

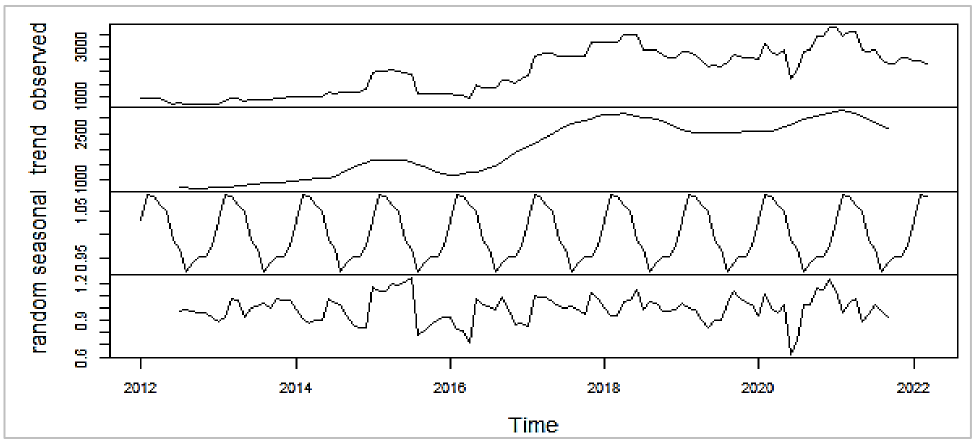


Figure 5 Components of time series data of commodity price

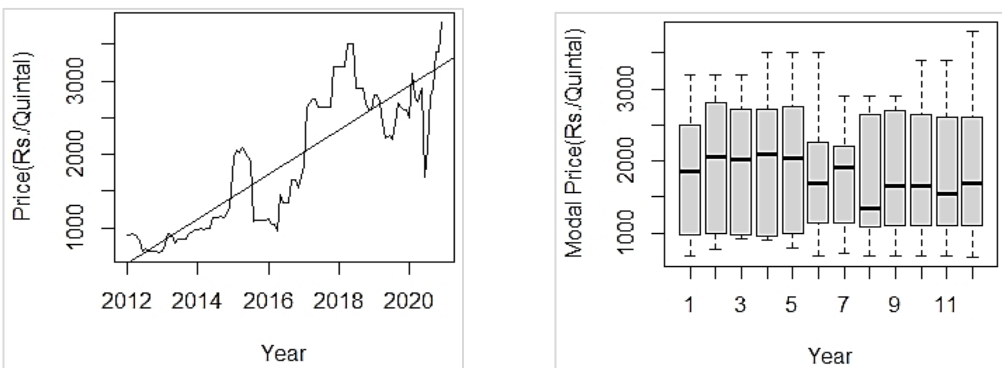


Figure 6 Statistical analysis of the time-series data

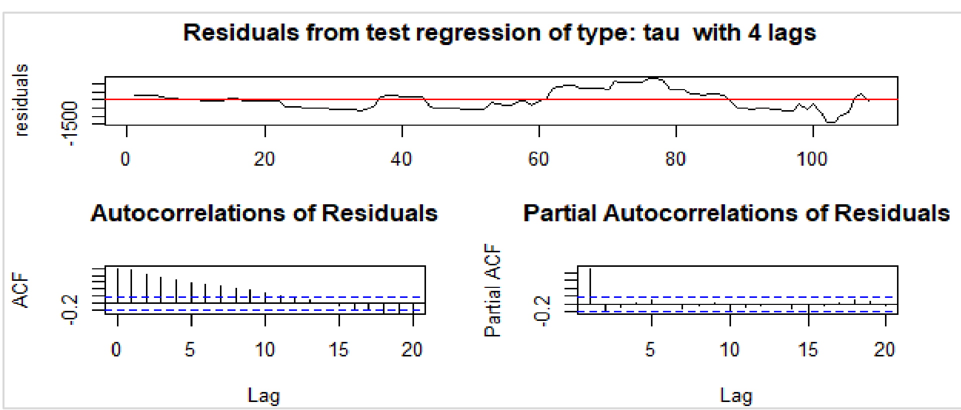


Figure 7 Results of unit root KPSS test

Consecutively, the non-stationarity is removed by assessing the differences between consecutive observations, and the seasonality is removed by deducting the seasonal component from the actual time data. This makes the data stationary. Once the data is prepared for analysis, the next step is to fit the

ARIMA model by determining the values for p, d, and q which represent the AR, I, and MA parameters of the ARIMA model, respectively.

The right degree of differencing (d) specifies the least amount of differencing needed to obtain a stationary series over a mean. The order can be chosen if it produces the smallest standard deviation in the differenced series. To determine the value of p and q, it is necessary to evaluate the auto correlation function (ACF) and partial autocorrelation graph (PACF) and evaluate the ACF and PCAF graph data [49]. This can be done by applying the ACF and PACF functions at each k-lag to perform autocorrelations between data points in the time series dataset. From the plotted graph, the parameters for p and q, i.e., the AR and MA parameters, can be determined. The plotted graph for ACF and PACF is shown in *Figures 8* and *9*, respectively.

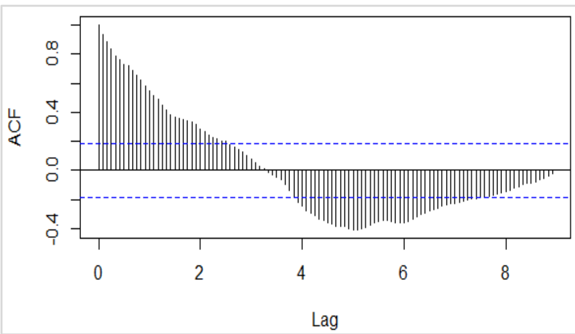


Figure 8 Interpretation of ACF plots

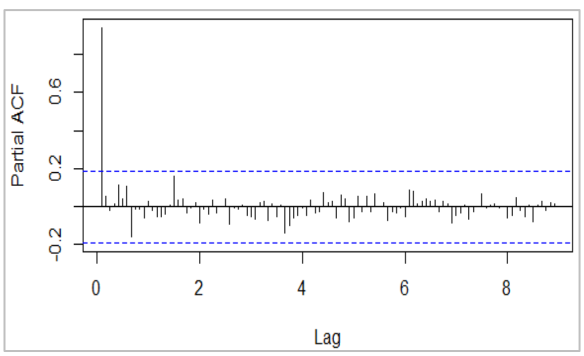


Figure 9 Interpretation of PACF plots

Upon identifying the model by determining the d, p, and q, the other parameters such as coefficients of lag and forecast error lag can be determined using probabilistic models such as maximum likelihood estimation (MLE).

In general, a good ARIMA model can be built by estimating and selecting the optimised parameters that minimise the various criteria such as akaike's information criterion (AIC) [50], corrected akaike's information criterion (AICc) [51], and bayesian information criterion (BIC) [52]. The formula for each criterion is given in Equations 14, 15, and 16.

$$AIC = -2\log(L) + 2(p+q+k+1) \quad (14)$$

$$AICc = AIC + \frac{2(p+q+k+1)(p+q+k+2)}{T-p-q-k-2} \quad (15)$$

$$BIC = AIC + (log(T) - 2)\,(p+q+k+1) \quad (16)$$

Thus, the different ARIMA models assessed for the dataset used in the analysis are shown in *Table 4.* Thus, the model ARIMA (0, 1, 0) having the minimum likelihood is chosen as the best fit model for the time series dataset used in the study. The various analyses of the identified ARIMA model are given in *Table 5*.

Table 4 Estimation of different ARIMA models

ARIMA Models	AICc Values
ARIMA(2,1,2)(1,0,1)[12] with drift	Infinity
ARIMA(0,1,0) with drift	1465.496
ARIMA(1,1,0)(1,0,0)[12] with drift	1467.837
ARIMA(0,1,1)(0,0,1)[12] with drift	1467.949
ARIMA(0,1,0)	**1464.127**
ARIMA(0,1,0)(1,0,0)[12] with drift	1467.389
ARIMA(0,1,0)(0,0,1)[12] with drift	1467.388
ARIMA(0,1,0)(1,0,1)[12] with drift	1469.547
ARIMA(1,1,0) with drift	1465.924
ARIMA(0,1,1) with drift	1466.017
ARIMA(1,1,1) with drift	1468.078

Table 5 Analysis of efficient ARIMA (0,1,0) identified

Evaluators	Values	Evaluators	Values
$Sigma^2$	50337	RMSE	223.3188
Log-likelihood	-731.04	MAE	117.1366
AIC	1464.09	MPE	0.1322339
AICc	1464.13	MAPE	7.572456
BIC	1466.76	MASE	0.1925203
Mind evolutionary (ME)	18.0625	-	-

Once the model is identified, the next step is to verify whether there is a pattern in the ARIMA residuals. The forecast can only be made if the residuals of the model are white noise. This can be achieved by plotting an autocorrelation graph for the residuals, through which it can be shown that the residuals are not correlated. It is given in *Figure 10*.

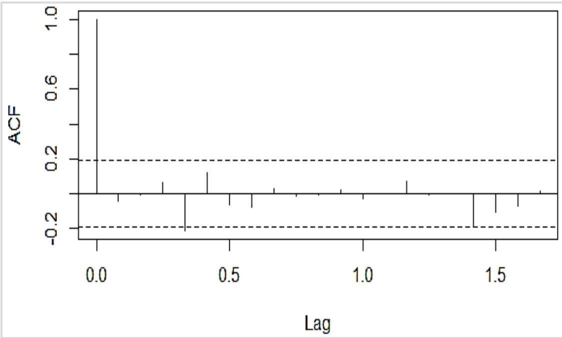


Figure 10 ACF plot for residuals of ARIMA (0,1,0)

The Box–Jenkins Analysis, a modelling approach, offers a systematic way to identify, fit, and verify the ARIMA model and use it to forecast time series data up to 18 lags [53]. It is also used for the analysis, in which it verifies the independence at all lags and computes the overall randomness based on the lags used. Upon applying the test to the residuals of the fitted model, the results show that most of the values are above p = 0.05, which indicates that there is no significance between the residuals. This is plotted in a graph in *Figure 11*. A normal quantile-quantile (Q-Q) plot can also be used to analyse the distribution of residuals, in which the values rest on the line, indicating that they are normally distributed. This is given in *Figure 12*. Thus, all three tests indicate that there is no pattern in the ARIMA residuals, so forecasting of values can be carried out.

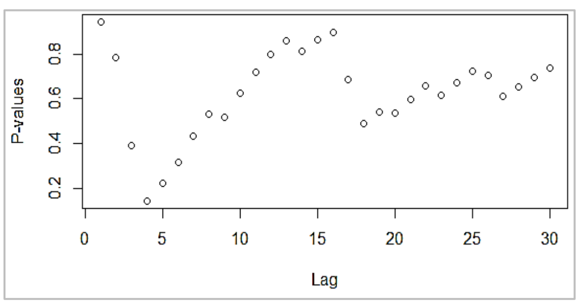


Figure 11 Box-ljung test on ARIMA residuals

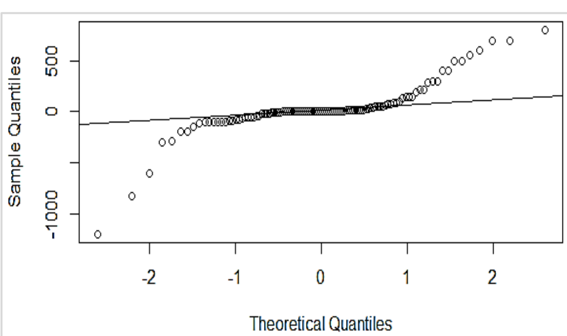


Figure 12 Normal Q-Qplot for ARIMA residuals

4.4Support vector regression

SVR is used to forecast values for nonlinear data. Generally, in the process of fitting the SVR model, the three significant hypermeters, such as *C, Gamma, and Epsilon*, must be tuned to provide an optimised result. The validation tests are performed for different parameter values, and the set of values having the minimum error and maximum accuracy can be selected to be utilised in the best-fit SVR model. If the parameters are not properly tuned based on the input data, then it often leads to under or over-fitting problems. However, the selected parameters may not be effective for all the datasets. In the proposed model, three ensemble models are utilised in combination with ARIMA and SVR, with different input sets for the SVR model. Thus, the parameter for the SVR algorithm used in the three models is to be tuned separately. For model 1, which utilises the residuals of ARIMA as its input, the hyperparameters in SVR are adjusted such that *C=11, gamma=3, and epsilon = 0.001* are chosen as they give the minimum error values. Similarly, for model 2, which utilises the outcome of ARIMA as one of the inputs of the SVR, the tuned hyperparameters are *C=11, gamma= 6, epsilon=0.001*, and for the model that takes ARIMA outcome and residuals as additional input, the tuned parameters of the SVM are *C=8, gamma=8, epsilon=1.0E-5*, as these sets of parameters provide fewer errors than other combinations.

4.5Comparative analysis

This section presents the results obtained for the three models in the proposed work and the comparison of the results with the existing models. To perform the analysis, Coconut dataset has been used. Notably, the dataset contains a minimum number of attributes, and hence no feature selection technique is applied. Furthermore, for all machine learning algorithms involving machine learning models, 80% of the data is used as the training set and 20% as the test set. The analysis of forecasted values obtained from the three models is examined in detail to evaluate the performance of the models. The deviation of the predicted values from the three models is compared with the actual values. Thus, the absolute deviation between the predicted and actual values is computed. The formula to compute the average absolute deviation percentage (DP) is the percentage of the average difference between the actual and predicted values, as in Equation 17.

$$Deviation\ Percentage = \frac{\sum_{i=1}^{n}|x_i-y_i|}{\mu_x+\mu_y} \qquad (17)$$

where n is the number of forecasted values, x_iandy_iare the ith actual and forecasted values, and μ_x and μ_y are the averages of actual and forecasted values. *Table 6* shows the predicted values as well as the absolute deviation. The absolute deviations in the predicted values for the three models are presented as a graph in *Figure 13.*

Table 6 Predicted values and absolute deviation

Time/Year	Actual Values	Model 1		Model 2		Model 3	
		Predicted	**Absolute Deviation**	**Predicted**	**Absolute Deviation**	**Predicted**	**Absolute Deviation**
January 2021	3800	3814.634	14.6336	2048.058	1751.942	2162.481	1637.519
February 2021	3400	3814.223	414.2228	2026.958	1373.042	2162.481	1237.519
March 2021	3600	3814.366	214.3661	2029.217	1570.783	2162.481	1437.519
April 2021	3600	3814.281	214.2805	2029.203	1570.797	2162.481	1437.519
May 2021	2850	3814.264	964.2635	2028.54	821.4598	2162.481	687.5193
June 2021	2800	3814.381	1014.381	2027.633	772.3667	2162.481	637.5193
July 2021	2850	3809.946	959.9463	2027.544	822.456	2162.481	687.5193
August 2021	2500	3806.943	1306.943	2027.113	472.8875	2162.481	337.5193
September 2021	2350	3785.126	1435.126	2026.902	323.0982	2162.481	187.5193
October 2021	2350	3771.418	1421.418	2026.855	323.1451	2162.481	187.5193
November 2021	2550	3815.267	1265.267	2026.938	523.0623	2162.481	387.5192
December 2021	2550	3790.729	1240.729	2027.141	522.8586	2162.482	387.5184
January 2022	2450	3795.145	1345.145	2030.334	419.6659	2162.571	287.4286
February 2022	2450	3802.592	1352.592	2029.287	420.7132	2162.527	287.4726
March 2022	2350	3793.243	1443.243	2028.553	321.4468	2162.503	187.4966

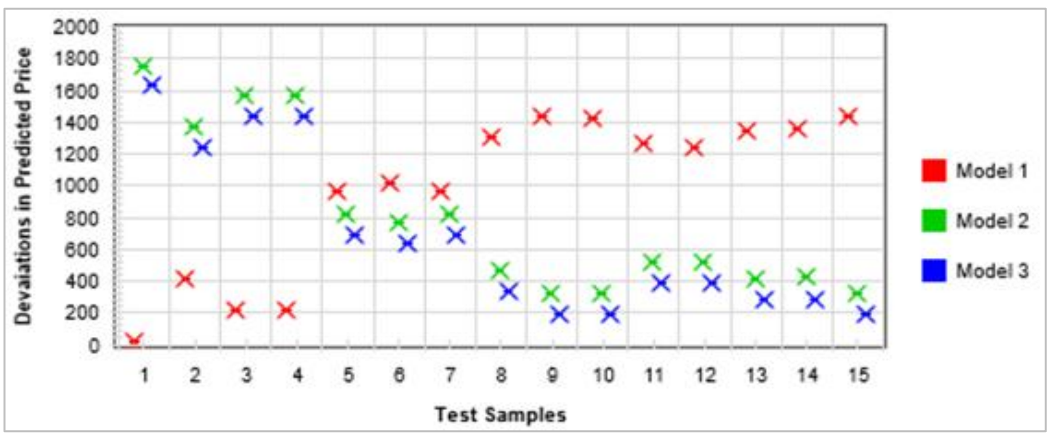


Figure 13 Deviations in forecasted values for the proposed model

Another analysis has been made with the experiments performed for the proposed models with various metrics at 1-step-ahead, 2-steps-ahead, and 3-steps-ahead with the coconut dataset. The results are presented in *Table 7.* The bold values represent the improved performance in each column with respect to the specific performance metrics.

Table 7 Performance analysis of the proposed model

	Metrics	MAE	RRSE	DA	RAE	MAPE	RMSE	MSE
	1-step-ahead	8.989	16.23	**86.316**	4.0489	12.722	60.128	3615.397
Model 1	2-steps-ahead	9.77	16.287	**85.106**	4.362	16.001	60.657	3679.214
	3-steps-ahead	11.238	16.869	**84.946**	5.096	24.836	62.201	3869.009
	1-step-ahead	3.533	1.608	72.632	2.585	**0.238**	4.18	17.472
Model 2	2-steps-ahead	3.869	1.92	71.277	2.808	**0.279**	5.016	25.158
	3-steps-ahead	3.788	1.404	70.968	1.615	**0.271**	5.104	26.048
	1-step-ahead	**1.868**	**0.775**	81.579	**1.139**	3.104	**2.309**	**2.731**
Model 3	2-steps-ahead	**1.936**	**0.794**	81.383	**1.172**	4.003	**2.377**	**2.886**
	3-steps-ahead	**1.938**	**0.655**	81.183	**0.849**	4.176	**2.395**	**2.909**

The obtained forecasted values from the proposed three models and the other existing statistical and learning models are compared with the actual observations for the Coconut dataset. The machine learning models used in the comparison are the LR, GP, MP, HW, NF, SES, and ARIMA models. The obtained models are plotted as a graph in *Figure 14*. From the figure, it is clear that most of the forecasted values of models 3 and 2 are closer to the actual observations than those of model 1 and other existing models.

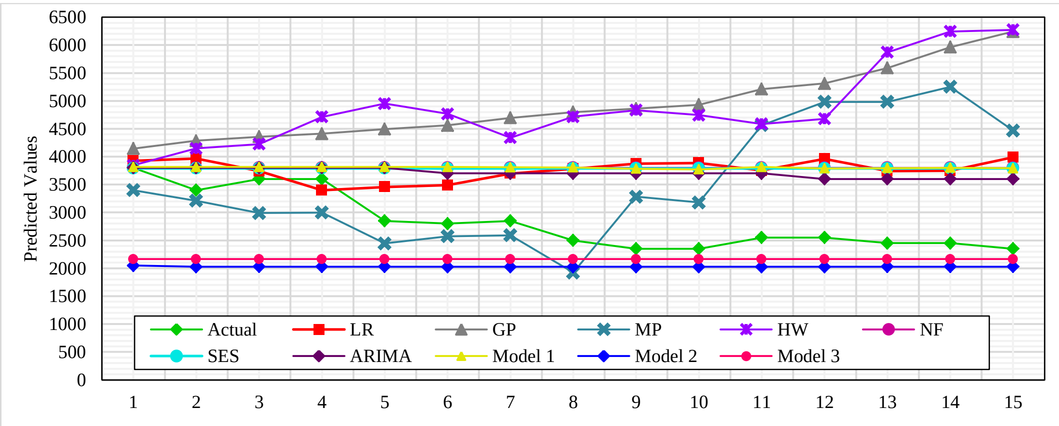


Figure 14 Comparison of forecasted values of the proposed model

Apart from forecasted values, the performance of the proposed models using the 1-step-ahead metric is also compared with other existing models using various evaluation metrics. The results obtained for the models are presented in *Table 8*. The percentage of deviation of the forecasted value from the actual value has been calculated using the formula given in Equation 17. The result obtained for the average absolute deviation is shown in *Figure 15*.

Table 8 Performance comparison of the proposed model

Forecasting Models	MAE	RRSE	DA	RAE	MAPE	RMSE	MSE	ET
LR	152.56	91.28	43.16	111.78	8.20	237.17	56251.07	0.132
GP	384.95	182.41	35.79	274.99	29.38	483.94	23420.02	0.342
MP	113.52	64.79	50.53	82.55	6.48	168.59	28424.78	0.372
HW	810.74	398.12	43.66	505.26	32.74	1141.39	13027.65	0.314
NV	125.79	52.39	77.11	61.73	6.73	246.74	220.39	0.398
SES	125.62	53.33	78.72	63.46	6.69	245.49	220.09	0.107
ARIMA	124.64	51.99	79.99	62.48	6.67	245.59	218.36	0.189
Model1	8.989	16.23	**86.32**	4.049	12.72	60.13	3615.39	0.236
Model2	3.533	1.608	72.63	2.59	**0.24**	4.18	17.47	0.302
Model3	**1.938**	**0.655**	81.18	**0.85**	4.18	**2.39**	**2.91**	0.311

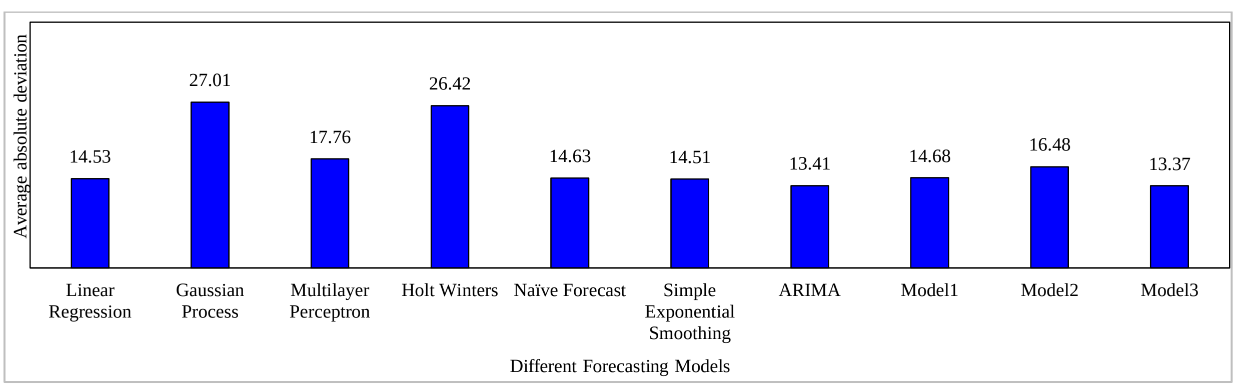
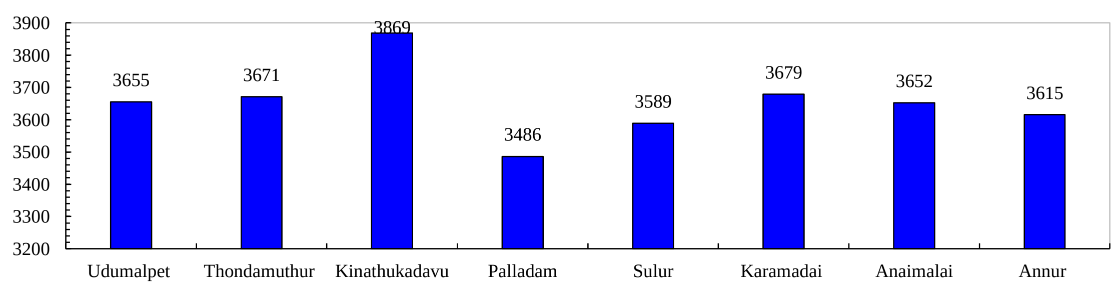


Figure 15 Comparison of average absolute deviation

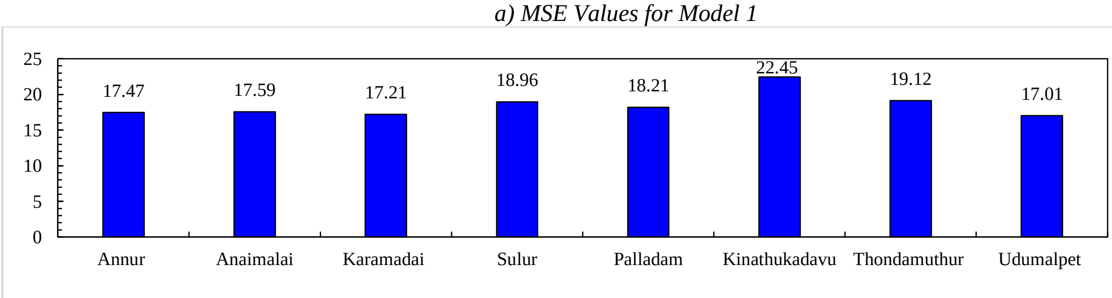
To justify the use of SVR with the statistical ARIMA, an analysis has been performed by utilizing various machine learning prediction algorithms in the proposed model in the place of SVR with 1-step-ahead metric for the Coconut dataset. Some of the prediction algorithms used are LR, GP, MP, NF and SVR. The obtained results are presented in *Table 9*. The results indicate that the utilizing machine learning techniques with ARIMA model improves the predication accuracy than using the machine learning algorithms individually. The proposed models have been analysed using the price of coconut extracted from other markets in the Coimbatore district, including Annur, Anaimalai, Karamadai, Sulur, Palladam, Kinathukadavu, Thondamuthur, and Udumalpet markets, where coconuts are the primary commodity. The past 10 years of prices are used as a training set for the proposed models. The MSE for the models used in the proposed work for the different datasets is computed, and the results obtained are presented as a graph in *Figure 16*, in which a) MSE for Model 1 b) MSE for Model 2 and c) MSE for Model 3.

Table 9 Result comparison of proposed model with other machine learning algorithms

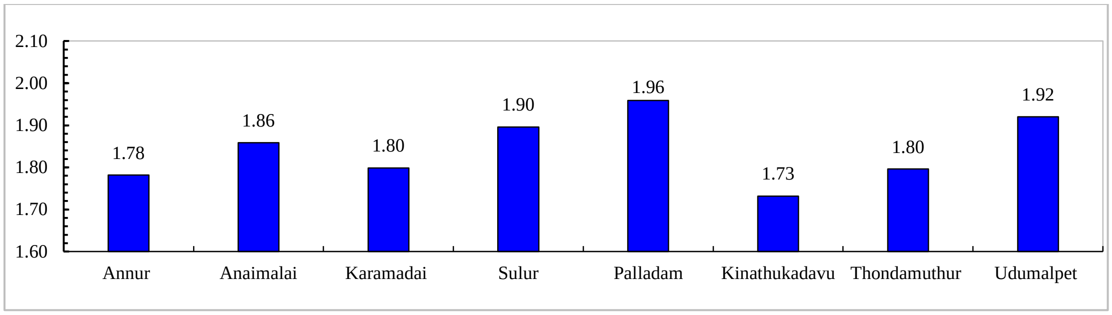
ML Technique	Metrics	MAE	RRSE	DA	RAE	MAPE	RMSE	MSE	ET
LR	Model 1	148.63	89.75	47.15	102.35	7.9	215.69	48569.25	0.189
	Model 2	132.53	85.63	48.25	98.63	6.8	197.85	3485.12	0.238
	Model 3	112.3	79.36	55.63	72.36	5.2	98.56	2478.2	0.289
	Average	*131.15*	*84.91*	*50.34*	*91.11*	*6.63*	*170.70*	*18177.52*	*0.24*
GP	Model 1	375.62	168.99	39.63	212.56	20.56	412.36	12489.61	0.399
	Model 2	312.32	128.41	42.36	165.22	15.20	285.63	5698.3	0.412
	Model 3	118.25	98.65	49.97	115.36	9.87	125.11	124.87	0.428
	Average	*268.73*	*132.02*	*43.99*	*164.38*	*15.21*	*274.37*	*6104.26*	*0.41*
MP	Model 1	108.7	56.93	57.23	80.23	7.89	98.75	22145.36	0.389
	Model 2	95.23	23.69	59.56	56.36	5.63	56.33	5896.36	0.401
	Model 3	56.23	12.01	55.36	19.63	6.61	18.12	1248.32	0.418
	Average	*86.72*	*30.88*	*57.38*	*52.07*	*6.71*	*57.73*	*9763.35*	*0.40*
NF	Model 1	112.3	42.22	80.98	58.71	5.32	214.56	202.37	0.425
	Model 2	92.36	30.11	69.36	12.36	1.55	60.86	18.36	0.451
	Model 3	58.63	11.63	78.96	6.32	2.36	10.99	5.36	0.475
	Average	*87.76*	*27.99*	*76.43*	*25.80*	***3.08***	*95.47*	***75.36***	*0.45*
SVR	Model 1	8.989	16.23	**86.316**	4.0489	12.722	60.128	3615.397	0.236
	Model 2	3.533	1.608	72.632	2.585	**0.238**	4.18	17.472	0.302
	Model 3	**1.868**	**0.775**	81.579	**1.139**	3.104	**2.309**	**2.731**	0.311
	Average	***4.80***	***6.20***	***80.18***	***2.59***	*5.35*	***22.21***	*1211.87*	*0.28*



a) MSE Values for Model 1



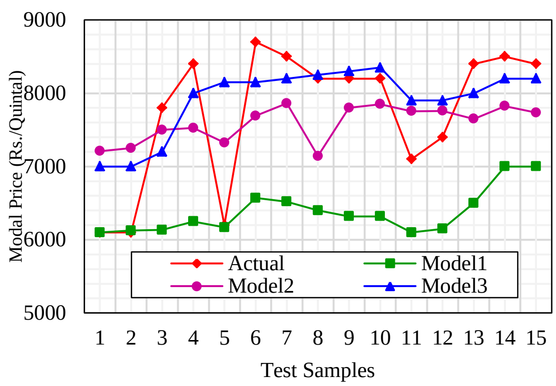
b) MSE Values for Model 2



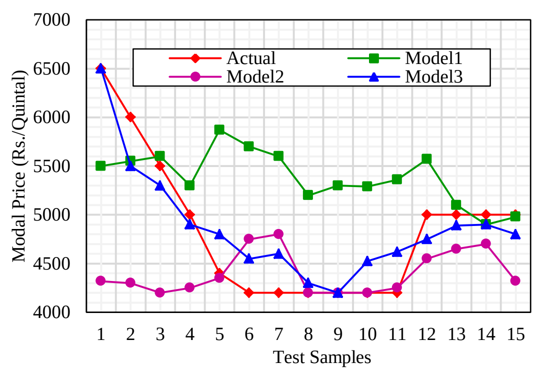
c) MSE Values for Model 3

Figure 16 MSE for coconut price prediction at various markets in Coimbatore

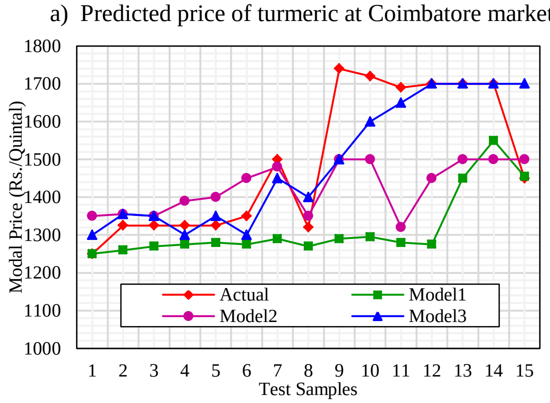
Similarly, to the price of the coconut, the price of various other commodities is extracted from various markets in and around the Coimbatore district via the website agmarknet.gov.in. The models used in the proposed work are executed for the various datasets that contain the prices of the commodities from the various markets in Coimbatore, such as turmeric at the Coimbatore market, cotton at the Tirupur market, maize at the Sulur market, tobacco at the Palladam market, groundnut at the Sevur market, and paddy at the Udumalaipet market. Moreover, turmeric had 108 samples, cotton had 45 samples, maize had 48 samples, tobacco had 72 samples, groundnut had 108 samples, and paddy had 36 samples, based on which the prices were predicted for the commodities. The values predicted for the three models used in the proposed study and the actual observations for various commodities are plotted as a graph in *Figure 17*, in which a) the forecasted price for turmeric; b) the forecasted price comparison for cotton; c) the price predicted for maize; d) the predicted price of tobacco; e) the forecasted price for groundnut; f) the predicted price for paddy.



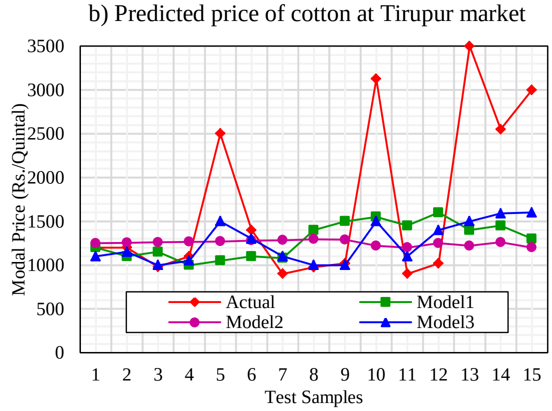
a) Predicted price of turmeric at Coimbatore market



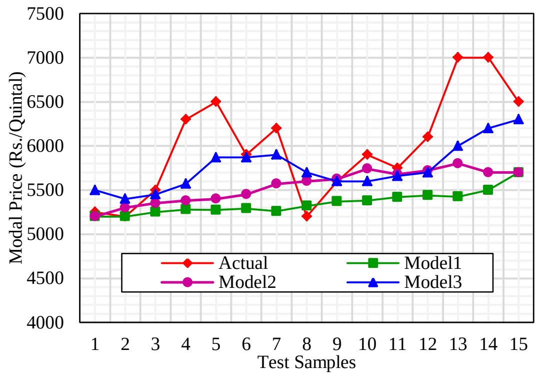
b) Predicted price of cotton at Tirupur market



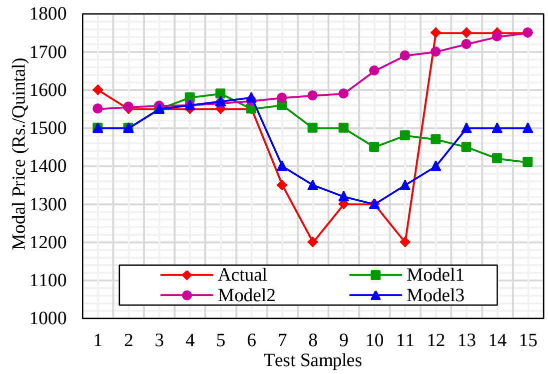
c) Predicted price of maize at Sulur market



d) Predicted price of tobacco at Palladam market



e) Predicted price of groundnut at Sevur market



f) Predicted price of paddy at Udumalaipet market

Figure 17 Predicted price of various commodities in different markets

5.Discussion

A detailed discussion of the results obtained in the above section has been made in this section. The experimental design and result analysis performed in the study is multi-fold. First, it evaluates the prediction performance of the hybrid model with a statistical ARIMA model using machine learning techniques. Second, comparing the predicted results of the proposed models with those of the other existing models. Third, justifying the preference of the SVR technique over other machine learning techniques. Fourth, identifying the better model among the proposed models for reliable prediction.

Thus, for model 1, which uses residuals from the linear part as the input for the nonlinear part, the ARIMA residuals are given as an input to the SVR model. For the final prediction, the SVR model results are combined with the ARIMA forecasted value. Thus, model 1 results in MAE = 8.989, DA = 86.31, RAE = 4.049, RMSE = 60.12, and MAPE = 12.721 (see *Table 7*). Similarly, for model 2, which uses the forecasted value of ARIMA as one of the inputs to the SVR model, the results obtained from the SVR model are considered the final forecasted value. The model has the performance with MAE = 3.533, DA = 72.632, RAE = 4.18, RMSE = 4.18, and MAPE = 0.2383. Finally, for model 3, which uses the forecasted value and residuals of ARIMA as additional inputs to the SVR model, the forecasted value and the forecasted residuals from the SVR model are summed to predict the final results. The model has the metric values MAE = 3.533, DA = 72.632, RAE = 4.18, RMSE = 4.18, and MAPE = 0.2383.

Moreover, the forecasting models such as HW and the GP has poor performance concerning most of the performance metrics used in the study. The models such as LR, GP and MP offer poor performance on MSE. The models such as NF, SES and ARIMA models offer similar yet better results than LR, GP, MP and HW. On the other hand, the models in the proposed works produce good results concerning various performance metrics (See *Table 8*). For the experiments performed to analyze the DP of forecasted value to the actual values, the forecasted results of the GP and HW have deviated more from the actual values. The results of MP and model 2 deviate about 16% from the actual values. With other models such as LR, NF, SES, ARIMA and Model1 have deviated about 14%. However, the results of ARIMA and Model 3 have deviated about 13% from actual observation which is minimum than other models. Though the ET of the proposed models is greater than LR, SES and ARIMA, nonetheless the time complexity of the proposed models is even less than the other four existing models. Moreover, the models having less ET have more prediction errors. To justify the use of the SVR model, other models such as LR, GP, MP, and NF are assessed in the proposed models instead of SVR. From the results obtained, it is found that using machine learning algorithms with the ARIMA model significantly improves the prediction performance compared to using them individually. ARIMA with LR appears to have a short average running time, but its prediction errors are higher than those of the SVR model. Also, while the average MAPE and MSE for a NF are lower than those of the SVR model (3.08 and 75.36, respectively), the SVR model outperforms it with better results for MAE, RRSE, DA, RAE, RMSE, and ET. Moreover, it is evident that the SVR model outperforms the GP and MP with all the evaluation metrics used (see *Table 9*). Thus, it is clear that the use of the SVR model is more suitable than other prediction algorithms in terms of minimum prediction error and time complexity.

According to the collective analysis, the sum of the absolute deviations for the three models is 14606.56, 12009.72, and 10012.63, respectively, resulting in an average absolute deviation of 973.77, 800.64, and 667.51 on which the percentage is calculated. Also, the percentage of the difference between the actual and predicted values for the three models is 14.68%, 16.48%, and 13.37%, respectively. This shows that the third model offers better prediction than the other two. Correspondingly, the time complexity of model 1 is 0.236 seconds, that of model 2 is 0.318 seconds, and that of model 3 is 0.331 seconds, respectively. Though the time complexity of model 3 seems to be a little higher than that of models 1 and 2, the results using various evaluation metrics show that, in comparison with accurate prediction, the increase in time is negligible for model 3 (see *Table 8*). When the DA is considered, model 1 outperforms models 2 and 3. On the other hand, the value of MAPE is lower with improved performance for model 2 than for model 1 and model 3. However, for the other metrics such as MAE, RRSE, RAE, RMSE, and MSE, model 3 offers improved performance over models 1 and 2.

So, to conclude, model 3 is better than the other proposed and existing models, as it makes predictions that are more accurate and closer to the actual value. Thus, the performance of model 3 is highly appreciable compared to models 1 and 2 for accurate prediction or forecasting of future trends in an effective way.

5.1Limitations

Like any other study, the proposed research on predicting prices has some limitations. The proposed research work has been implemented and verified with the market datasets that belong to and around the Coimbatore region. The model predicts prices using commodity price history, which it only uses the last 10 years of prices allotted for various commodities that are popular in a wide range of markets. Moreover, only a limited set of commodities, such as turmeric, cotton, maize, tobacco, groundnuts, and paddy, have been used for the analysis. Despite the price datasets on limited commodities, these datasets are also incomplete. The entries for the few commodity datasets extracted from the website have missing entries, so the few months or years of price allocation for the commodities could not be used for the analysis. This also reduces the performance of the model in forecasting prices for such commodities. Finally, the compassion of results obtained from the proposed models is only compared to traditional models. This is due to the fact that the studies in the literature utilise different datasets that are not publicly available. Thus, the lack of common datasets makes it difficult to perform comparative analysis with the other existing models specified in the literature review.

6.Conclusion and future work

Forecasting is a significant process that helps people by providing significant information and assisting in making decisions. Price forecasting is important for farmers as it reduces their potential losses and acts as a base for crop production and marketing decisions. This paper utilises a hybrid ensemble model with the ARIMA model as the base and utilises the SVR, a machine learning algorithm, in three different forms. The first model utilises the residuals from the linear model as input for the SVR algorithm, the second model utilises the forecasted value of the ARIMA model as one of the inputs and the third model employs residuals and forecasted ARIMA values as additional input to predict the desired price for the commodity. Extensive analysis has been done for the commodity price dataset available publicly for coconut, and the results show that model 3 has improved results with a 13.37% deviation from actual observation compared to the model 1, and model 2 has a deviation of about 14.68% and 16.48%, respectively. Though the DA of model 1 has improved to 86.32% and the MAPE of model 2 has a minimum value of about 0.24, model 3 can be considered as effective as the RRSE, RAE, RMSE, and MSE values are very low at 0.655, 0.85, 2.39, and 2.91, respectively, compared to other existing models. Also, an analysis has been made of other commodities that are sold in the markets in and around Coimbatore, including turmeric, cotton, maize, tobacco, groundnuts, and paddy. The future research will focus on improving the model's performance and combining demand and price forecasting analysis to recommend crops to farmers. The proposed model only utilises past prices as a factor for predicting prices, so future work can incorporate other factors in predicting commodity prices. Also, the comparison has been made with traditional algorithms, and so the comparison of results has to be made with other existing models in the future.

A complete list of abbreviations is shown in *Appendix I.*

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

Author's contribution statement

G. Murugesan: Conceptualization and design of the work, data acquisition, implementation, interpretation and analysis, writing and editing original draft. **B. Radha:** Conceptualization and design of the work, supervision, analysis and interpretation of results and review of results. Both authors discussed the results, contributed to the final manuscript and provided critical feedback at every stage.

References

[1] Selvaraj M, Ibrahim MS. Indian agricultural marketing-a review. Asian Journal of Agriculture and Rural Development. 2012; 2(1):69-75.

[2] Johnston BF, Mellor JW. The role of agriculture in economic development. The American Economic Review. 1961; 51(4):566-93.

[3] Matsuyama K. Agricultural productivity, comparative advantage, and economic growth. Journal of Economic Theory. 1992; 58(2):317-34.

[4] Acharya SS. Agricultural marketing in India, 4/E. Oxford and IBH publishing; 2004.

[5] Vohra A, Pandey N, Khatri SK. Decision making support system for prediction of prices in agricultural commodity. In Amity international conference on artificial intelligence 2019 (pp. 345-8). IEEE.

[6] https://timesofindia.indiatimes.com/india/suicides-of-agricultural-labourers-rise-by-18-ncrb-report/articleshow/87344897.cm. Accessed 8 March 2022.

[7] Noonari S, Irfana NM, Raiz AB, Muhammad IK, Ali S. Price flexibility and seasonal variations of major vegetables in Sindh Pakistan. Journal of Food Processing & Technology. 2015; 6(524):1-8.

[8] Anggraeni W, Mahananto F, Rofiq MA, Andri KB, Zaini Z, Subriadi AP. Agricultural strategic commodity price forecasting using artificial neural network. In international seminar on research of information technology and intelligent systems 2018 (pp. 347-52). IEEE.

[9] Tong AH. Factors influencing price of agricultural products and stability countermeasures. Asian Agricultural Research. 2012; 4(4):17-43.

[10] Chun-yan Y, Jun M, Yu-yan Z. Online price extraction and decision support for agricultural products. In international conference on information management, innovation management and industrial engineering 2009 (pp. 337-40). IEEE.

[11] Hamzaçebi C, Akay D, Kutay F. Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting. Expert Systems with Applications. 2009; 36(2):3839-44.

[12] Erdem E, Shi J. ARMA based approaches for forecasting the tuple of wind speed and direction. Applied Energy. 2011; 88(4):1405-14.

[13] Jadhav V, Chinnappa RB, Gaddi GM. Application of ARIMA model for forecasting agricultural prices. Journal of Agricultural Science and Technology. 2017; 19(4):981-92.

[14] Mgale YJ, Yan Y, Timothy S. A comparative study of ARIMA and holt-winters exponential smoothing models for rice price forecasting in Tanzania. Open Access Library Journal. 2021; 8(5):1-9.

[15] Gjølberg O, Bengtsson BA. Forecasting quarterly hog prices: simple autoregressive models vs. naive predictions. Agribusiness: An International Journal. 1997; 13(6):673-9.

[16] Maulana H, Mulyantika U. The prediction of export product prices with holt's double exponential smoothing method. In 3rd international conference on computer and informatics engineering 2020 (pp. 372-5). IEEE.

[17] Ge Y, Wu H. Prediction of corn price fluctuation based on multiple linear regression analysis model under big data. Neural Computing and Applications. 2020; 32(22):16843-55.

[18] Gastli MS, Nassar L, Karray F. Satellite images and deep learning tools for crop yield prediction and price forecasting. In international joint conference on neural networks 2021 (pp. 1-8). IEEE.

[19] Jha GK, Sinha K. Agricultural price forecasting using neural network model: an innovative information delivery system. Agricultural Economics Research Review. 2013; 26(2):229-39.

[20] Bao Y, Lu Y, Zhang J. Forecasting stock price by SVMs regression. In international conference on artificial intelligence: methodology, systems, and applications 2004 (pp. 295-303). Springer, Berlin, Heidelberg.

[21] Kozlovskyi S, Mazur H, Vdovenko N, Shepel T, Kozlovskyi V. Modeling and forecasting the level of state stimulation of agricultural production in Ukraine based on the theory of fuzzy logic. Montenegrin Journal of Economics. 14(3):37-53.

[22] Wang B, Liu P, Chao Z, Junmei W, Chen W, Cao N, et al. Research on hybrid model of garlic short-term price forecasting based on big data. Computers, Materials and Continua. 2018; 57(2):283-96.

[23] Wang L, Feng J, Sui X, Chu X, Mu W. Agricultural product price forecasting methods: research advances and trend. British Food Journal. 2020; 122(7):2121-38.

[24] Darekar A, Reddy AA. Cotton price forecasting in major producing states. Economic Affairs. 2017; 62(3):373-8.

[25] BV BP, Dakshayini M. Performance analysis of the regression and time series predictive models using parallel implementation for agricultural data. Procedia Computer Science. 2018; 132:198-207.

[26] Xu X, Zhang Y. Corn cash price forecasting with neural networks. Computers and Electronics in Agriculture. 2021; 184:1-13.

[27] Yu Y, Zhou H, Fu J. Research on agricultural product price forecasting model based on improved BP neural network. Journal of Ambient Intelligence and Humanized Computing. 2018:1-6.

[28] Zhang D, Zang G, Li J, Ma K, Liu H. Prediction of soybean price in China using QR-RBF neural network model. Computers and Electronics in Agriculture. 2018; 154:10-7.
[29] Xiong T, Li C, Bao Y. Seasonal forecasting of agricultural commodity price using a hybrid STL and ELM method: evidence from the vegetable market in China. Neurocomputing. 2018; 275:2831-44.
[30] Jha GK, Sinha K. Time-delay neural networks for time series prediction: an application to the monthly wholesale price of oilseeds in India. Neural Computing and Applications. 2014; 24(3):563-71.
[31] Sabu KM, Kumar TM. Predictive analytics in agriculture: forecasting prices of arecanuts in Kerala. Procedia Computer Science. 2020; 171:699-708.
[32] Zhang YL, Zhao L, Dong ZL. Prediction of fluctuation range and change trend of cotton price based on fuzzy information granulation and PSO-SVR model. Guangdong Agricultural Sciences. 2015; 42(11):180-5.
[33] Duan Q, Zhang L, Wei F, Xiao X, Wang L. Forecasting model and validation for aquatic product price based on time series GA-SVR. Transactions of the Chinese Society of Agricultural Engineering. 2017; 33(1):308-14.
[34] Zhang Y, Na S. A novel agricultural commodity price forecasting model based on fuzzy information granulation and MEA-SVM model. Mathematical Problems in Engineering. 2018; 2018:1-11.
[35] Yin H, Jin D, Gu YH, Park CJ, Han SK, Yoo SJ. STL-ATTLSTM: vegetable price forecasting using STL and attention mechanism-based LSTM. Agriculture. 2020; 10(12):1-17.
[36] Yuan CZ, Ling SK. Long short-term memory model based agriculture commodity price prediction application. In proceedings of the 2nd international conference on information technology and computer communications 2020 (pp. 43-9).
[37] Chen Z, Goh HS, Sin KL, Lim K, Chung NK, Liew XY. Automated agriculture commodity price prediction system with machine learning techniques. Automated Agriculture Commodity Price Prediction System with Machine Learning Techniques. 2021; 6(4): 376-84.
[38] Samuel P, Sahithi B, Saheli T, Ramanika D, Kumar NA. Crop price prediction system using machine learning algorithms. Quest Journals: Journal of Software Engineering and Simulation. 2020; 6(1):14-20.
[39] Wang J, Hu J. A robust combination approach for short-term wind speed forecasting and analysis–combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM) forecasts using a GPR (Gaussian Process Regression) model. Energy. 2015; 93:41-56.
[40] Barak S, Sadegh SS. Forecasting energy consumption using ensemble ARIMA–ANFIS hybrid algorithm. International Journal of Electrical Power & Energy Systems. 2016; 82:92-104.
[41] Li C, Chen Z, Liu J, Li D, Gao X, Di F, et al. Power load forecasting based on the combined model of LSTM and XGBoost. In proceedings of the international conference on pattern recognition and artificial intelligence 2019 (pp. 46-51). ACM.
[42] Torbat S, Khashei M, Bijari M. A hybrid probabilistic fuzzy ARIMA model for consumption forecasting in commodity markets. Economic Analysis and Policy. 2018; 58:22-31.
[43] Suganthi L, Iniyan S, Samuel AA. Applications of fuzzy logic in renewable energy systems–a review. Renewable and Sustainable Energy Reviews. 2015; 48:585-607.
[44] Li C, Hu JW. A new ARIMA-based neuro-fuzzy approach and swarm intelligence for time series forecasting. Engineering Applications of Artificial Intelligence. 2012; 25(2):295-308.
[45] Xu S, An X, Qiao X, Zhu L, Li L. Multi-output least-squares support vector regression machines. Pattern Recognition Letters. 2013; 34(9):1078-84.
[46] Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing. 2003; 50:159-75.
[47] https://agmarknet.gov.in/. Accessed 24 April 2022.
[48] https://datascienceplus.com/time-series-analysis-using-arima-model-in-r/. Accessed 24 April 2022.
[49] Khan FM, Gupta R. ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India. Journal of Safety Science and Resilience. 2020; 1(1):12-8.
[50] Bozdogan H. Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. Psychometrika. 1987; 52(3):345-70.
[51] Hurvich CM, Tsai CL. A corrected Akaike information criterion for vector autoregressive model selection. Journal of Time Series Analysis. 1993; 14(3):271-9.
[52] Neath AA, Cavanaugh JE. The Bayesian information criterion: background, derivation, and applications. Wiley Interdisciplinary Reviews: Computational Statistics. 2012; 4(2):199-203.

G. Murugesan is currently working as an Assistant Professor in Sree Saraswathi Thyagaraja College, Pollachi. He has completed Bachelor of Science in computer science and Master of Computer applications. His area of interest includes Web Technologies, IoT, Object Oriented Programming and Software Engineering. He has published articles in various international/national journals and presented papers in national/international conferences and workshops. He has been a resource person for various seminars and workshops and guided around 50 technical projects.
Email: g29.murugesan@gmail.com

Dr. B. Radha is presently working as Associate Professor in Sri Krishna college of Arts and Science, Coimbatore. She has completed her Ph.D from Anna University, Chennai. She has attended and published many papers in International/National conferences and also has written many books in different areas. She has acted as a chairperson in many international conferences. Her current research interests are Data Mining, Networking and Cloud Computing.
Email: radhakbr10@gmail.com

Appendix I

S. No.	Abbreviation	Description
1	x_t	Actual Alue at Time t
2	d	Degree of Differencing
3	y_t	Final forecasted Value at Time t
4	l_t	Linearly Forecasted Value at Time t
5	p	Number of Autoregressive Terms
6	q	Number of Regression Errors
7	ε_t	Residuals of the Linear Model
8	ACF	Auto Correlation Function
9	AIC	Akaike’s Information Criterion
10	AICc	corrected Akaike’s Information Criterion
11	ARIMA	Autoregressive Integrated Moving Average
12	ARMA	Autoregressive Moving Average
13	BIC	Bayesian Information Criterion
14	BPNN	Back Propagation Neural Network
15	DA	Direction Accuracy
16	ELM	Extreme Learning Machine
17	ET	Execution Time
18	GP	Gaussian Process
19	HW	Holt-Winters
20	KPSS	Kwiatkowski-Phillips-Schmidt-Shin
21	LR	Linear Regression
22	LSTM	Long Short-Term Memory
23	MA	Moving Average
24	MAE	Mean Absolute Error
25	MAPE	Mean Absolute Percentage Error
26	MEA	Mind Evolutionary Algorithm
27	MLE	Maximum Likelihood Estimation
28	MP	Multilayer Perceptron
29	MSE	Mean Squared Error
30	NCRB	National Crime Record Bureau
31	NF	Naïve Forecasting
32	NIC	National Informatics Centre
33	NLR	Non-Linear Regression
34	PACF	Partial Autocorrelation Graph
35	PSO	Particle Swarm Optimization
36	Q	Quintal
37	Q-Q	Quantile-Quantile
38	RAE	Relative Absolute Error
39	RBFNN	Radial Basis Function Neural Network
40	RMSE	Root Mean Squared Error
41	RRSE	Root Relative Squared Error
42	SES	Simple Exponential Smoothing
43	SVM	Support Vector Machines
44	SVR	Support Vector Regression
45	TDNN	Time-Delay Neural Network
46	XGBoost	Extreme Gradient Boosting