Neural Computing and Applications (2023) 35:15109–15128
https://doi.org/10.1007/s00521-023-08528-7

ORIGINAL ARTICLE

Agricultural commodity price prediction model: a machine learning framework

Manas Kumar Mohanty[1] · Parag Kumar Guha Thakurta[1] · Samarjit Kar[2]

Received: 17 August 2022 / Accepted: 21 March 2023 / Published online: 6 April 2023

Abstract

An efficient machine learning-based framework for crop price prediction is proposed in this paper to assist the farmers in estimating their profit-loss beforehand. The proposed work is composed of four functional blocks, such as crop yield prediction, determination of supply, demand prediction and crop price prediction. The input datasets consist of the various field values, such as yield, remaining crop at the end of the year, import, demand and price of a crop. Various time series-based algorithms, such as autoregression, moving average, autoregressive moving average, autoregressive integrated moving average and exponential smoothing, are used to forecast the crop yield. The supply of the crop is determined as a sum of three variables, i.e., the predicted crop yield, residue and import values. The demand for the crop is predicted from a year alone as the demand has more correlation with year over other factors. The crop price from demand, supply and year is predicted using different approaches, which include the time series method, statistical approaches and machine learning techniques. Finally, these three techniques for price prediction are compared to determine the best model having minimum root-mean-square error value. In the proposed work, the decision tree regressor is found to be the best model, for predicting crop price, over others. The superiority of the proposed work over existing approaches, in terms of various aspects, is shown by simulation results.

Keywords Crop price prediction · Crop yield · Demand and supply · Machine learning · Agriculture

1 Introduction

Agricultural development is one of the most powerful tools to end extreme poverty, boost prosperity, and feed a projected 9.7 billion people by 2050 [1]. Growth in the agriculture sector is two to four times more effective in raising incomes among the poorest compared to other sectors. In 2016, it was found by an analysis [2] that 65% of poor working adults made a living through agriculture. However, more than 10% of the world's population [3] suffered from food scarcity in 2020. Under such a scenario, expanding food production is a compelling process to cope with a shortfall of food with respect to the growing population. In agriculture, crop yield is a measurement of the amount of a crop produced during a particular time period. Hence, crop yield prediction [4] is most significant for global food production. Various factors affecting crop yield are soil type, soil nutrients, temperature, rainfall, etc. Farmers are benefited from yield forecasts to get assistance in financial and managerial decisions. Hence, an increase in crop yield also improves farmer's profit, enhancing their socioeconomic status.

Meanwhile, accurate crop price forecasting can be helpful for farmers to obtain a reasonable price for their yield [5]. At the same time, it is very useful for the farmer to make better decisions like when to sell their products or harvest the crop. Such crop price measures the unit value

Parag Kumar Guha Thakurta and Samarjit Kar have contributed equally to this work.

✉ Manas Kumar Mohanty
mkm.19cs1102@phd.nitdgp.ac.in

Parag Kumar Guha Thakurta
paragkumar.guhathakurta@cse.nitdgp.ac.in

Samarjit Kar
samarjit.kar@maths.nitdgp.ac.in

[1] Department of Computer Science and Engineering, National Institute of Technology Durgapur, A-Zone, Durgapur, West Bengal 713209, India

[2] Department of Mathematics, National Institute of Technology Durgapur, A-Zone, Durgapur, West Bengal 713209, India

received by farmers from the domestic market for a specific agricultural commodity produced within a year [6]. The important determinants influencing the crop price are yield, crop type, rainfall, historical prices etc. Another promising factor in predicting the crop price is the demand [7] which measures the quantity of the crop that consumers can purchase during a given period. Furthermore, the amount of a crop available for consumers as a supply [8] is obtained from yield, residue and import, which can affect the variation of crop price. Here, the residue is the quantity of the crop which remains unsold at the end of the year and is available for consumption in the next year, while import denotes the amount of the crop, not produced within the country, which is brought from another country [9].

In the early days, crop price prediction was performed by the farmers using their past experience in terms of historical data [10]. Gradually, much research is being done to predict the crop price with growing interest. Prediction of the crop price of agricultural products is also needed to anticipate the negative impact of future price changes ahead of time. In a real scenario, every farmer desires to know about his profit if he cultivates a particular crop in the next season. This can improve farmer's living standards. Most existing research works on predicting crop price using time series methods. However, supply and demand are the foremost factors in deciding the price of a commodity [11]. Hence, the question remains open to predicting the crop price more accurately considering those factors. Usually, a considerable part of forecasting crop yield and crop price in the agricultural framework cannot be delineated in a fundamental stepwise procedure, especially with complex, incomplete, ambiguous and strident datasets [12]. In such a scenario, numerous studies indicate that machine learning (ML) algorithms [13] have comparatively an improved potential over conventional statistics in order to develop an agricultural framework with remarkable forecasting ability. Henceforth, it motivates us to develop an ML-based efficient crop price prediction model using significant parameters related to it, which can assist the farmers' anticipations of substantial price changes in the future and their consequences.

An efficient crop price prediction using ML is proposed here to assist the farmers in estimating their profit-loss beforehand. The input datasets for the proposed work consist of the various field values of the crop (corn) for 47 number of years. The fields of the input datasets are yield, remaining crop at the end of the year, import, demand and price of the crop for each year. Initially, the various time series-based ML algorithms, such as exponential smoothing (ES), autoregression (AR), theta model (Theta), autoregressive integrated moving average (ARIMA), moving average (MA), autoregressive moving average (ARMA) and seasonal-trend decomposition (STL) using LOESS are used to forecast the crop yield. The ARIMA method is found to be the best, in terms of minimum absolute percentage error (MAPE) value, among others used in predicting the crop yield. After that, the supply of the crop is determined as a sum of three variables, such as the predicted crop yield, residue and import values. Next, the demand for the crop is predicted from the year alone as the demand has more correlation with year over other factors. Furthermore, the crop price from demand, supply and year is predicted using different approaches, which include the time series method (ARIMA), statistical regression (SR) and various ML techniques such as decision tree regressor (DTR), random forest (RF), K-nearest neighbor regressor (KNN) and kernel ridge(KR). Finally, these three techniques for price prediction are compared to determine the best model having a minimum root-mean-square error (RMSE) value. In the proposed work, the best model DTR having the least RMSE is found over other models. The superiority of the proposed work over existing approaches is shown in various aspects. Hence, the major contributions of our work are summarized as follows:

- Crop yield is forecasted using different time series-based ML algorithms, and subsequently, the best one among alternatives is determined in terms of having minimum MAPE.
- The amount of supply is determined from predicted yield, residue data and import data.
- Forecasting of demand is obtained by the most accurate time series prediction method from historical data.
- Predicting the crop price for the next year is obtained by using different methods such as time series, statistical and ML techniques and the best one amidst others having a minimum RMSE value is subsequently determined.
- The superiority of the proposed work over existing approaches is shown by various performance comparison through simulation results.

The rest of this paper is organized as follows: Section 2 presents the literature survey for the completeness of the proposed work. Section 3 introduces useful preliminaries with respect to the proposed work. Section 4 describes the proposed methodology. Section 5 covers results from different models and forecasting the real price data. Finally, Sect. 6 concludes the paper.

2 Literature survey

The price of different crops is predicted by analyzing the previous rainfall data [14, 15]. In order to predict the crop price, several contemporary ML techniques are used [5]. In another work [16], an integrated feature selection of

ARIMA with computational intelligence approaches has been applied for crop price prediction. Other than the ARIMA, the components of the proposed integrated forecasting models include artificial neural networks (ANNs) and support vector regression (SVR). The paper by [12] predicts the crop price for the next rotation. This work is based on finding suitable data models that help in achieving generality for price prediction. Again, different data mining techniques were evaluated on different datasets to solve crop price prediction issues of seasonal crops [17]. Here, the equipped marketing information is analyzed for integration of domestic markets. Subsequently, the price transmission from markets to farm gates is perceived, which can lead to a sustainable profit. Another paper [18] highlights yield and price forecasting for crop decision planning. For yearly yield prediction, they introduce a weather-based time-dependent dataset. Here, the price is predicted as a sum of the future prices on a commodity basis.

In order to make decisions for agricultural marketing, an intelligent crop price prediction model is described in [19]. It is addressed by different ML models which can predict the prices of crops in advance. An adaptive crop price prediction for agriculture applications is proposed by [20]. Another time series-based approach [21] is proposed to explore the crop price and crop yield prediction of selected crops to identify the relevant information with respect to the market prices and crop yields. Some of the uncertain conditions, such as climate changes, fluctuations in the market and flooding, to the agricultural process have been addressed here. The research work by [22] discusses the prediction of crop yield using regression analysis. In this work, several influencing factors are mentioned which are related to crop yield prediction. In [23], a framework for crop price forecasting is designed by analyzing the time series data. Here, the major features related to developing crop price prediction models are the historical weather data that influence crop production and transportation and data quality-related features obtained by performing statistical analysis. Using ML approaches, another location-aware crop price prediction is discussed [24]. In this work, the future price of vegetables is predicted depending on the suitable location for the profitable production of agricultural products. Forecasting of both crop yield and price using ML techniques have presented in [25]. The prediction of crop price is mainly dependent on factors, such as rainfall, temperature, market prices, land area and past crop yield. The authors of this paper also predict the price and the gain for the next twelve months over the past twelve months. Another research work employs intelligent techniques on historical prices to predict the crop price for different agricultural products [26].

Another prediction model for crop price is discussed using time series methods [27]. However, this work has yet to consider other input parameters except historical prices in order to develop this price prediction model. In [28], a comparison based on different algorithms, such as ARIMA, SVR and XGBoost, is introduced concerning price prediction. In this context of price forecasting, the authors of [29] have used optimal lag selection to improve the prediction performance in terms of time and reduced error. However, the significance of supply and demand factors [30] in forecasting agricultural commodity prices was beyond the scope of their work. Similarly, another paper [31] predicts the crop price using statistical regression on time series data of monthly price for 14 years. The price forecast of a crop using ARIMA is presented by [32]. However, it is restricted only to the prediction of the monthly price, which would not be beneficial for the farmers in their season-based agriculture. The work of [33] compares various models and finds ARIMA as suitable for crop price prediction. An ANN-based approach is used to predict the future crop price [34].

In a nutshell, several procedures to predict crop price accurately are discussed in various ways. It is observed that most of these techniques have used the time series method or weather data to predict the price of the crop. However, the utility of supply and demand to predict the crop price is beyond the scope of their work. As per our best knowledge, no prior research was considered earlier to predict price using the supply and demand of the crop, even after these being the prime factors of pricing a crop. Another scope of research is still open for further reduction of training and testing errors in developing a crop price prediction model, which in turn can be beneficial to assist the farmers in decision-making on their profit. Hence, the work proposed in this paper addresses an efficient ML-based technique to obtain the reduced error in price prediction using the supply and demand of that particular crop. Furthermore, the best alternative training algorithms are determined from the perspective of error reduction in crop price prediction. Hence, this ML-based crop price prediction framework, using crop yield, supply and demand in agriculture, is comprehensively presented next.

3 Preliminaries

For a comprehensive presentation of the proposed work, the descriptive statistics of the input data is shown next. In addition, various performance measures used to highlight efficiency of the proposed work are introduced.

3.1 Materials

The dataset ($D2$-corn) (https://www.ers.usda.gov/data-products/feed-grains-database/feed-grains-yearbook-tables/) [35] is obtained from the department of agriculture, United States (U.S.). The $D2$-corn was created on Friday, December 10, 2021. It consists of the fields such as yield, residue and import of the crop, corn, year wise for 47 years. In Table 1, the statistical summary of $D2$-corn is shown. Another dataset, $D3$-corn, obtained from the same [35] was also created on Friday, December 10, 2021. The $D3$-corn includes the fields year, demand and price for the same 47 years. It has similar statistics on the crop corn, like $D2$-corn, for 47 years. The statistical summary of $D3$-corn is shown in Table 2. It is significant to mention here that the field, year, is common to both $D2$-corn and $D3$-corn, so that an efficient integration can be performed for successive operations. In Tables 1 and 2, 'N' denotes the number of samples in that corresponding dataset. In addition, the description of all parameters of $D2$-corn and $D3$-corn are outlined in Table 3 for a better understanding of the proposed work.

3.2 Performance measures

The following measures [36], shown in Table 4, are used to determine the performance of the work proposed in this paper.

4 Proposed methodology

The proposed work aims to predict the crop price for the next farming season. The proposed methodology is comprised of four individual functional blocks as shown in Fig. 1. The respective functions of these are-(i) Block 1: Crop yield prediction, (ii) Block 2: Determination of supply, (iii) Block 3: Demand prediction and (iv) Block 4: Crop price prediction. These functional blocks are numbered in a sequence according to the order of their occurrence in the proposed methodology. The entire methodology is described next.

4.1 Crop yield prediction

The Block 1 considers an input $D2-corn["\text{yield}"]$. Different time series-based ML algorithms, as shown in Fig. 2, are used to train the model for crop yield prediction in Block 1 which is represented as follows:

$$\begin{aligned}\hat{Y_X^i} &= f_X(D2-\text{corn}["\text{yield}"].\text{head}(i-1)) \\ &\quad + b1_X + \text{err}_X^i,\ \text{i=1,2, ..., n}\end{aligned} \tag{1}$$

In (1), $\hat{Y_X^i}$ is the amount of crop yield predicted by an algorithm "X" for ith instance, $head(i-1)$ returns the first (i-1) number of instances, $b1_X$ is an additive term for X and err_X^i is the error in prediction for ith instance of $D2-corn["yield"]$ by the algorithm X. In Block 1, time series-based ML algorithms, such as ES, AR, Theta, ARIMA, MA, ARMA and STL as shown in Fig. 2, are used to train the prediction model using Eq. (1). Finally, all of these seven predicted values, such as $\hat{Y}_{\text{ES}}$, $\hat{Y}_{\text{AR}}$, $\hat{Y}_{Theta}$, $\hat{Y}_{\text{ARIMA}}$, $\hat{Y}_{\text{MA}}$, $\hat{Y}_{\text{ARMA}}$ and $\hat{Y}_{\text{STL}}$, are compared. Subsequently, the best one ($\hat{Y}_{\text{best}}$) for the crop yield prediction is determined by the following:

$$\hat{Y}_{\text{best}} = \min_{\text{method}\in(\text{ES,AR,Theta,ARIMA,MA,ARMA,STL})} \hat{Y}_{\text{method}}.\text{mape}() \tag{2}$$

In (2), the min operator returns the algorithm for which MAPE is minimum and mape() returns the MAPE of the corresponding algorithm. Thus, $\hat{Y}_{\text{best}}$ corresponds to one of the algorithms, addressed earlier, has the minimum MAPE in prediction over others.

4.2 Determination of supply

In Block 2, the value of supply is determined from values of other parameters such as the amount of predicted crop yield from Block 1, i.e., $\hat{Y}_{\text{best}}$, the import value of that crop in the current year, i.e., $D2-\text{corn}["\text{import}"]$ and residue left-over at the end of last year, i.e., $D2-\text{corn}["\text{residue}"]$. Henceforth, the residue can be expressed as follows:

$$R^n = S^{(n-1)} - C^{(n-1)} \tag{3}$$

where $S^{(n-1)}$ is the supply amount for the crop in $(n-1)th$ year and $C^{(n-1)}$ denotes the total amount of the crop consumed in the $(n-1)th$ year. Therefore, the supply of the

Table 1 Statistical summary of $D2$-corn

Parameter	N	Mean	SD	Sum	Minimum	Maximum
Year	47	1998.5	14	95,928	1975	2021
Yield	47	10,218.2	3102.6	480,255.4	6206	15,454
Residue	47	1351.35	272.4	63,513.45	821	1879
Import	47	82.35	19.48	3870.45	8.2	205.5

Table 2 Statistical summary of $D3$-corn

Parameter	N	Mean	SD	Sum	Minimum	Maximum
Year	47	1998.5	14	95,928	1975	2021
Demand	47	9908.15	2847.35	465,676	5767	14,830
Price	47	3.00149	1.17907	141.07	1.5	6.89

Table 3 Description of parameters used in $D2$-corn and $D3$-corn

Parameter	Description
Year	Year
Yield	Yearly yield of the crop
Residue	Remaining crop at the end of the year
Import	Yearly import of the crop
Demand	Yearly demand of the crop
Price	Yearly price of the crop

crop is determined as a sum of the values of three parameters by the following:

$$S^n = \hat{Y}_{\text{best}} + I^n + R^n \tag{4}$$

where S^n is the supply amount for *nth* year and $\hat{Y}_{\text{best}}$ is the predicted crop yield obtained in Block 1. Here, for *nth* year, the parameter I^n denotes the import value. It is to be noted here that the parameter S^n has a significant impact on the crop price prediction.

4.3 Demand prediction

It is already discussed that the demand has a positive correlation with the parameter “Year”. So, the demand can be predicted using various time series methods. In Block 3, four methods, such as exponential smoothing (ES), autoregression (AR), moving average (MA) and autoregressive integrated moving average (ARIMA), are used to obtain a predicted value for the demand of the corresponding crop. Each of these methods is briefly introduced here.

- ES: This method predicts the demand value by using a weighted sum of past observations with exponentially decreasing weight for older observations. It can be expressed by the following:

$$\begin{aligned}\hat{D}^n_{\text{ES}} =& \alpha \times D3 - \text{corn}[''\text{demand}''].\text{head}(n-1)\\ &+ (1-\alpha) \times D^{(n-1)}_{\text{ES}}\end{aligned} \tag{5}$$

where $\hat{D}^n_{\text{ES}}$ is the predicted demand for *nth* year, $D3 - \text{corn}[''\text{demand}''].\text{head}(n-1)$ is the demand data up to $(n-1)th$ year, $D^{(n-1)}_{\text{ES}}$ is the actual demand of $(n-1)th$ year and α is the smoothing factor having a value from

Table 4 Performance measures

Name of the measure	Definition	Formula	Symbols
Mean absolute error (MAE)	The average of difference between predicted values and actual values	$\text{MAE} = \frac{\sum_{i=1}^{n}(\hat{Y^i} - Y^i)}{n}$	Y^i = Actual value $\hat{Y^i}$ = Predicted value n = Number of instances
Mean square error (MSE)	The average of the square of difference between predicted values and actual values	$\text{MSE} = \frac{\sum_{i=1}^{n}(\hat{Y^i} - Y^i)^2}{n}$	
Root-mean-square error (RMSE)	The root of the average of the square of difference between predicted values and actual values	$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{Y^i} - Y^i)^2}{n}}$	
Mean absolute percentage error (MAPE)	The mean of the percentage of difference between predicted values and actual values	$\text{MAPE} = \frac{\sum_{i=1}^{n}(\hat{Y^i} - Y^i)}{Y^i}$	
R-squared (R^2)	The proportion of the variance of the output variable that is explained by input variable(s)	$R^2 = 1 - \frac{\text{SSE}}{\text{SST}}$	SSE = Sum of squares error SST = Sum of squares total
Adjusted R-squared (R^2_{Adjusted})	Modified version of R^2 which also takes into account the number of independent variable(s)	$R^2_{\text{Adjusted}} = 1 - \frac{(1-R^2)\times(n-1)}{n-\ell-1}$	ℓ = Number of independent variable(s)
Theil's U (U)	The accuracy of predictive model over the last observation as forecast	$U = \frac{\left(\frac{\sum_{i=1}^{n}(\hat{Y^i} - Y^i)^2}{n}\right)^{\frac{1}{2}}}{\left(\frac{\sum_{i=1}^{n}Y^2}{n} + \frac{\sum_{i=1}^{n}\hat{Y}^2}{n}\right)^{\frac{1}{2}}}$	

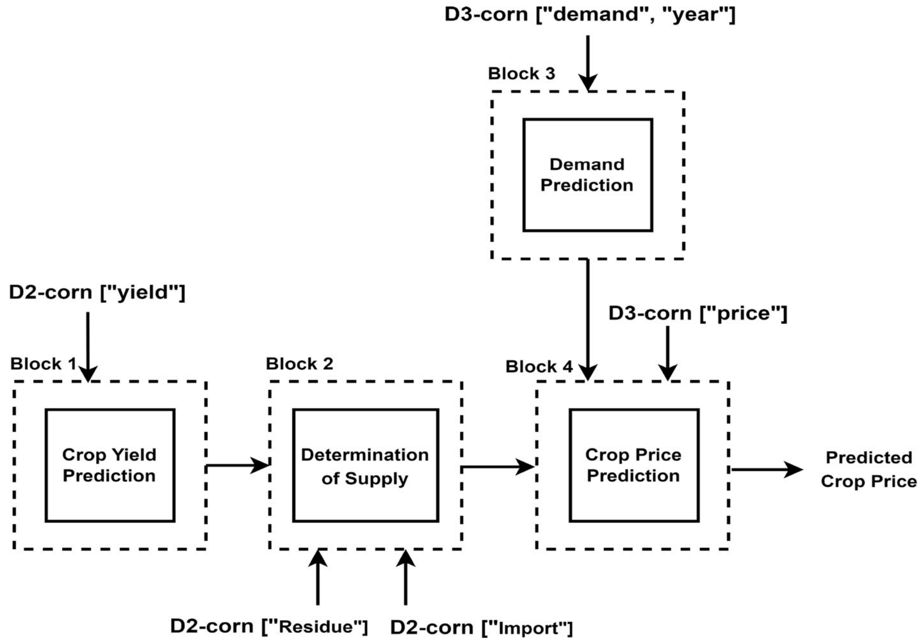


Fig. 1 Evaluation framework of proposed methodology

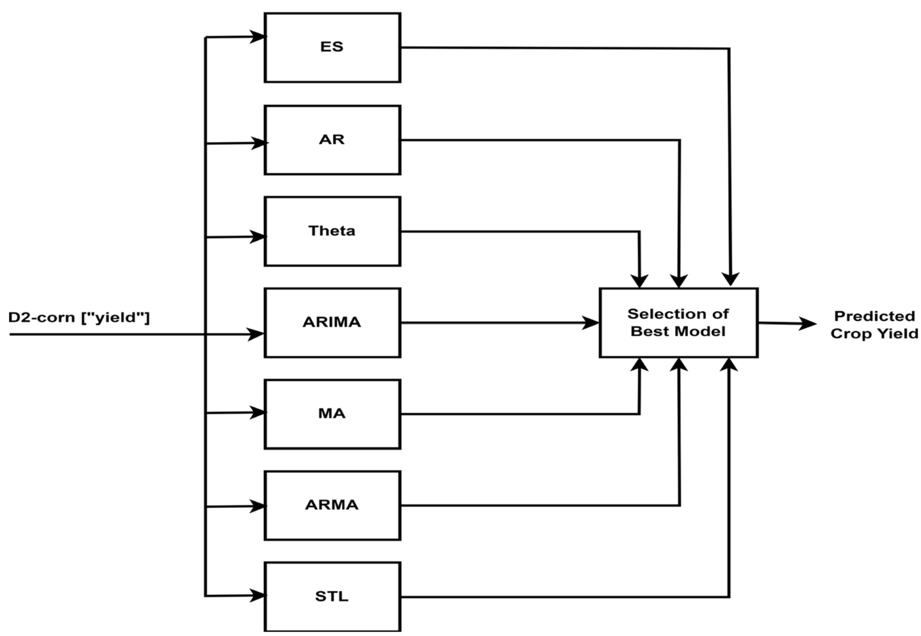


Fig. 2 Crop yield prediction by different ML algorithms

0 to 1. If the value of α is larger, then the level of smoothing is reduced. Thus, the value of α close to 1 has less smoothing effect and provides greater weight to past observations, while the value of α closer to zero has a greater smoothing effect and emphasizes on $D_{\mathrm{ES}}^{(n-1)}$.

- AR: It uses a series of observations from previous time steps as an input to predict the value of the next time step. The order of AR is the number of immediately preceding values in the series which are used to predict the value at the present time. The demand value by AR is obtained by the following:

$$\hat{D}_{\mathrm{AR}}^{n} = \sum_{i=1}^{p} \Phi_{n-i} \times D3 - \mathrm{corn}[''\mathrm{demand}''].\mathrm{loc}(n-i) + b2_{\mathrm{AR}} + \in^{n} \tag{6}$$

where Φ_i is the coefficient of *ith* lagged demand, 'p' is the order of the autoregression, $b2_{\mathrm{AR}}$ is an additive term and $\hat{D}_{\mathrm{AR}}^{n}$ is predicted demand value of *nth* year and $\in^{n}$ is the residual.

- MA: It predicts the demand by adding up all the data values up to a specific time which is further divided by the sum of the number of time periods. It can find a constantly updated average demand by considering equal weightage to all the past values. This MA method uses a way to predict the demand from $D3-corn$ as follows:

$$\hat{D}^{n}_{\mathrm{MA}} = \frac{1}{t}\sum_{i=n-1-t}^{n-1} D3 - \mathrm{corn}[\text{"demand"}].\mathrm{loc}(i) \tag{7}$$

where $\hat{D}^{n}_{\mathrm{MA}}$ is the value of demand for the nth year and 't' is the window size of moving average.

- ARIMA: It is based on both AR and MA processes. It has an ability to generalize for non-stationary series. The predicted demand by ARIMA is obtained by the following:

$$\begin{aligned}\hat{D}^{n}_{\mathrm{ARIMA}} =& \phi_1 D3 - \mathrm{corn}[\text{"demand"}].\mathrm{loc}(n-1) + \ldots \\ &+ \phi_p D3 - \mathrm{corn}[\text{"demand"}].\mathrm{loc}(n-p) + \theta_1\lambda_{n-1} + \ldots \\ &+ \theta_q\lambda_{n-q} + b2_{\mathrm{ARIMA}} + \mu_1\end{aligned} \tag{8}$$

where $\hat{D}^{n}_{\mathrm{ARIMA}}$ is the predicted demand for nth instance of D3-corn["demand"], 'p' is the order of the autoregression, 'q' denotes the order of the moving average, θ is MA parameter, ϕ is AR parameter, $D3-\mathrm{corn}[\text{"demand"}].\mathrm{loc}(n-p)$ are lagged values, λ_{n-q} are lagged errors of moving average, $b2_{\mathrm{ARIMA}}$ is an additive term and μ_1 denotes a constant.

Thus, these four time series methods have different processing techniques, which can lead to distinct values of prediction for each. Then, the outputs obtained from the four methods are compared, and subsequently, the best one (D^{n}_{best}) is selected, which has minimum error for demand prediction over others.

4.4 Crop price prediction

The function of crop price prediction for next year (Y_n) in Block 4 is detailed in Fig. 3. Initially, the dataset D3-corn and the supply (S_n) are considered as the inputs of Block 4. The feature "price" from D3-corn is selected. It is noted that $D_3-\mathrm{corn}[\text{"price"}]$ is partitioned to keep the first 70% samples as $D_3-\mathrm{corn}[\text{"price"}].Train$ and the rest 30% samples as $D_3-\mathrm{corn}[\text{"price"}].Test$ for training and testing, respectively. Then, $D_3-\mathrm{corn}[\text{"price"}].Train$ is used to train the model by the ARIMA method as a time series data. Meanwhile, the inputs D3-corn and S_n are integrated to obtain the dataset, $D_{23}-corn$ which is partitioned into two subsets, such as $D_{23}-corn.Train$ and $D_{23}-corn.Test$, in a 70:30 ratio using random sampling. Thus, the $D_{23}-corn.Train$ is used to train the model by using statistical regression (SR) and different ML approaches. The predicted crop price by these algorithms are obtained by the following equations:

- For SR:

$$\begin{aligned}\hat{P}^{i}_{\mathrm{REG}} =& F_{\mathrm{REG}}(D23 - \mathrm{corn}, a3_{\mathrm{reg}}) \\ &+ b3_{\mathrm{reg}} + e^{i}_{\mathrm{reg}},\ i = 1, 2, \ldots, n\end{aligned} \tag{9}$$

where $\hat{P}^{i}_{\mathrm{REG}}$ denotes the predicted price for ith instance of $D_{23}-\mathrm{corn}$, $a3$reg is the coefficient matrix. Another term $b3_{\mathrm{reg}}$ acts as an additive term in (9) and e^{i}_{reg} denotes the error in prediction of ith instance of $D_{23}-\mathrm{corn}$.

- For ARIMA:

$$\begin{aligned}\hat{P}^{n}_{\mathrm{ARIMA}} =& \Phi_1 D3 - \mathrm{corn}[\text{"price"}].\mathrm{loc}(n-1) + \ldots \\ &+ \Phi_p D3 - \mathrm{corn}[\text{"price"}].\mathrm{loc}(n-p) \\ &+ \Theta_1\Lambda_{n-1} + \ldots + \Theta_q\Lambda_{n-q} + b3_{\mathrm{ARIMA}} + \mu_2\end{aligned} \tag{10}$$

In (10), $\hat{P}^{n}_{\mathrm{ARIMA}}$ is the predicted price for nth instance, $D3-\mathrm{corn}[\text{"price"}].\mathrm{loc}(i)$ is the ith instance, 'p' is the order of the autoregression, 'q' is the order of the moving average, Λ is MA parameter, Φ is AR parameter, P_{n-p} are lagged values, Λ_{n-q} are lagged errors of moving average, $b3_{\mathrm{ARIMA}}$ is an additive term and μ_2 is the constant.

- For ML:

$$\hat{P}^{i}_{X} = f_X(D23 - \mathrm{corn}, a3_{ml}) + b3_{ml} + e^{i}_{ml}, \quad i = 1, 2, \ldots, n \tag{11}$$

where $\hat{P}_X$ denotes the price prediction using algorithm 'X', $a3_{ml}$ is the coefficient matrix, $b3_{ml}$ is the additive term and e^{i}_{ml} is the error for the ith instance. It is to be mentioned here that four different ML algorithms, such as DTR, KNN, RF and KR, are used to train over the training dataset to develop the crop price prediction model.

After the completion of the training phase, the performance of those training models are evaluated using test datasets, such as $D_3-\mathrm{corn}[\text{"price"}].\mathrm{Test}$ and $D_{23}-corn.Test$. Finally, the best one amidst others having the least error is selected as the proposed crop price prediction model, $\mathrm{Model}_{\mathrm{proposed}}$, as shown as follows:

$$\mathrm{Model}_{\mathrm{proposed}} = \min_{\mathrm{method}\in(\mathrm{REG,ARIMA,ML})} \hat{P}_{\mathrm{method}}.\mathrm{rmse}() \tag{12}$$

In (12), the min operator returns the algorithm for which RMSE is minimum and rmse() returns the RMSE of the corresponding algorithm. Thus, for any unseen sample comprising of (S^n, D^{n}_{best}, Y_n), the proposed model can predict the crop price for the next year more accurately.

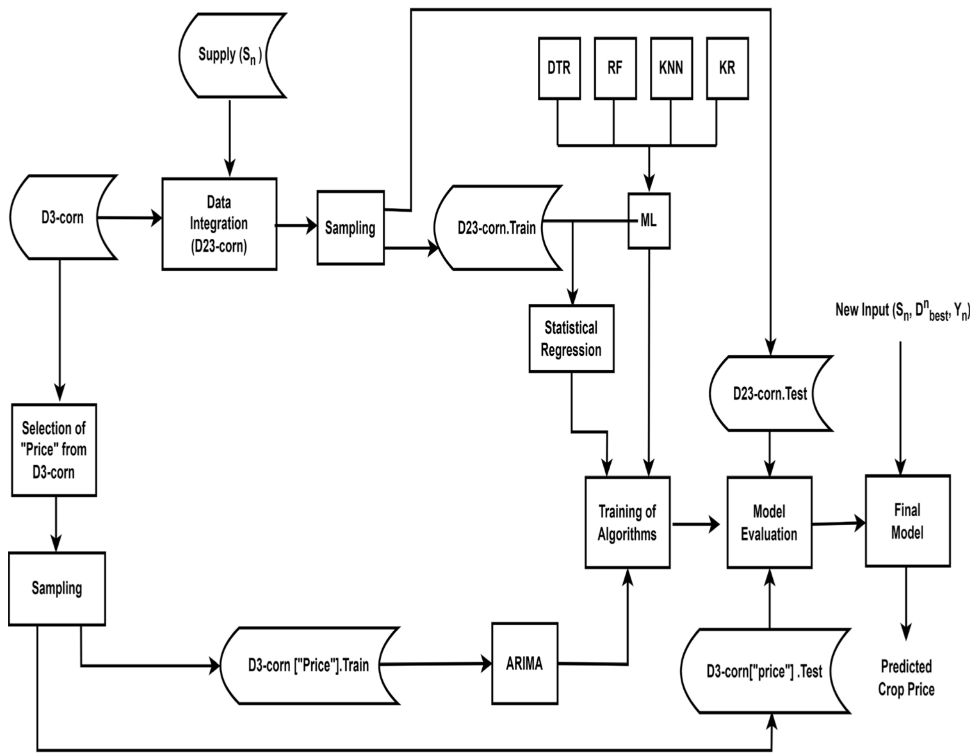


Fig. 3 Workflow of the model for crop price prediction

5 Experimental results

Various simulation results are shown to highlight the efficacy of the proposed work. In this context, the simulation setup and the dataset description are shown next.

5.1 Simulation setup

In order to simulate the proposed model, two datasets such as D2-corn and D3-corn having values for a specific crop, i.e., corn, are used. Various useful information regarding the parameters of the datasets is shown in Table 5. Firstly, we have simulated the function of Block 1 to find the yield predicted for the next season. Next, demand is predicted by Block 3. At last, the crop price is predicted in Block 4. The proposed work is implemented using Python and Origin software in the machine, having an Intel I7 processor and 16 GB RAM.

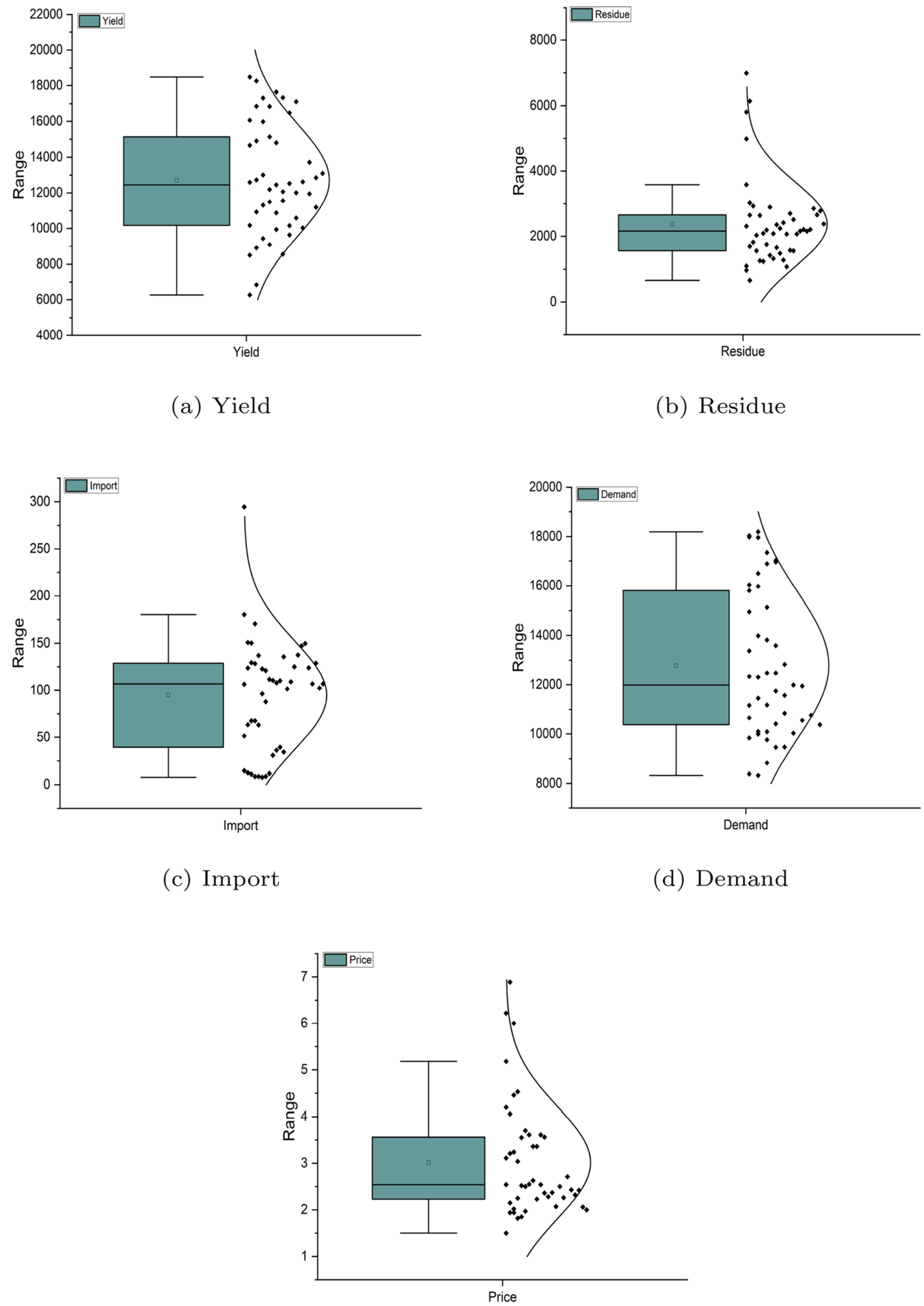
Table 5 Parameter description in *D*2-corn and *D*3-corn

Name	Unit	Data type	Dimension	Attribute type
Yield	Million bushel	Float	1	Continuous
Residue	Million bushel	Float	1	Continuous
Import	Million bushel	Float	1	Continuous
Demand	Million bushel	Float	1	Continuous
Price	USD	Currency	1	Continuous

5.2 Simulation results

Figure 4 shows the probability distributions of all input parameters of D2-corn and D3-corn. The probability distributions of the parameters can span over the entire range of the sample values for that parameter. These are shown in Fig. 4a–e, which correspond to the data of both D2-corn and D3-corn. More number of these parameters make the distribution close to normal. In contrast, fewer parameters appeared to be less normally distributed with slight skewness. From this analysis, it is observed that: (i) most of the parameters are well distributed around the mean and very few parameters have outliers, (ii) the mean, median, standard deviation, variance and outliers of input parameters are understood, and (iii) Both D2-corn and D3-corn are suitable for further analysis as any statistical tests can perform better when the input data follows a normal distribution. The scatter matrix with histogram for all the parameters of D2-corn and D3-corn are shown in Fig. 5a and b, respectively. These two figures show the scatter plots of each parameter with respect to all of the other parameters to highlight the relationship between those. Here, it is observed that all the parameters are linearly related. The histograms of each parameter are also shown on the diagonal line of scatter matrix, which represents the

Fig. 4 Box normal plots of parameters of D2-corn and D3-corn



(a) Yield

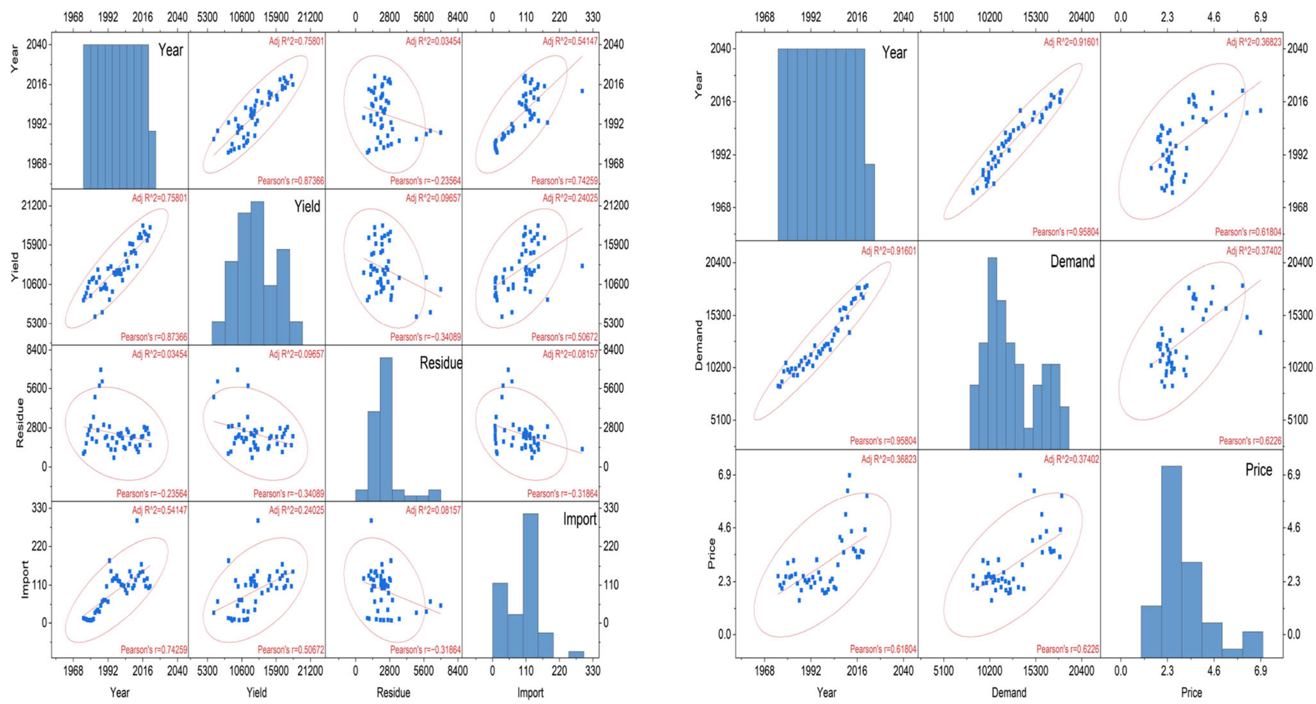
(b) Residue

(c) Import

(d) Demand

(e) Price

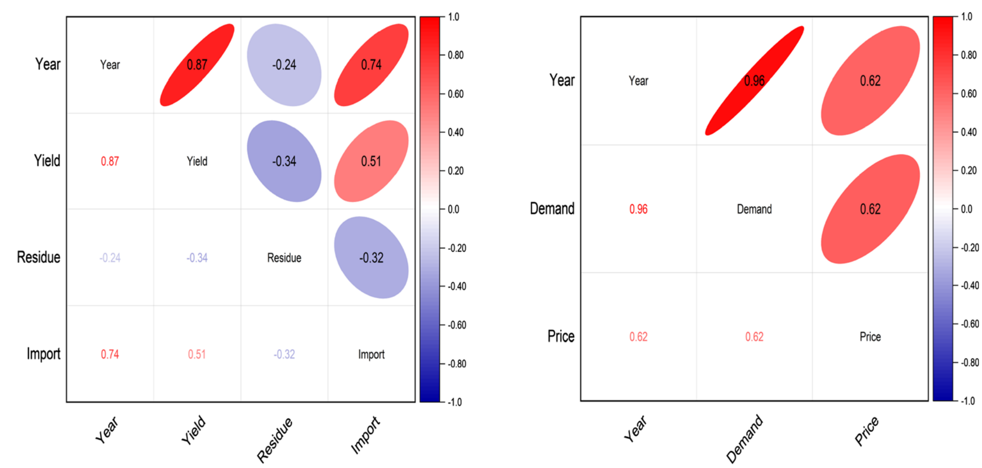
frequency distributions of the parameters. The frequency distributions of more number of parameters are nearly normal, except a few parameters have positive skewness. As these variables are normally distributed in both D2-corn and D3-corn, Pearson's correlation is preferred over Spearman's correlation to highlight the linear relationships between the variables. The correlation analysis is significant here as it would be very difficult to understand the relationship between each variable by simply staring at the parameters of datasets. Figure 6a can indicate the correlations between each pair of variables for $D2$-corn. Here, it is shown that most of the variables are highly correlated with each other. In Fig. 6b, the correlations between each pair of variables for $D3$-corn are shown. Here, a few of the variables are highly correlated with each other.



(a) D2-corn (b) D3-corn

Fig. 5 Scatter matrix with histogram

Fig. 6 Correlation matrix



(a) D2-corn (b) D3-corn

Figure 7 shows the values of MAPE for all time series-based ML algorithms used to predict the crop yield. These ML algorithms are trained over the training subset of $D2-corn["yield"]$, and subsequently, these are tested over the testing subset of the same. The train-test split for $D2-corn["yield"]$ is 85% and 15%, respectively. The prediction performance of all these training algorithms is measured using the test dataset. It is observed that the MAPE is highest for ES and lowest for ARIMA(4,1,1). The MAPE emphasizes the deviations of predicted yield from actual yields by what percentage of actual yields, so it focuses on the magnitude of the error. Thus, the ARIMA(4,1,1) is selected as the best in terms of reduced MAPE, for crop yield prediction.

Figure 8 shows the normal distributions of all crop yield algorithms used in the proposed work and that of test data. It shows the spread of the yield and subsequent weightage of the various values of the yield. No test dataset can show

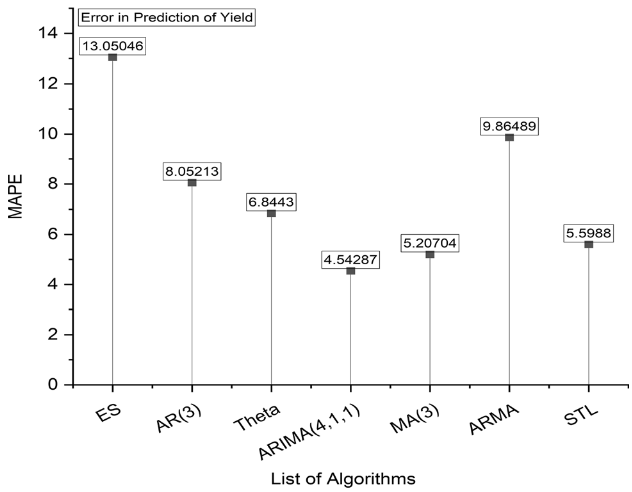


Fig. 7 Performance measures in crop yield prediction using different time series-based ML algorithms

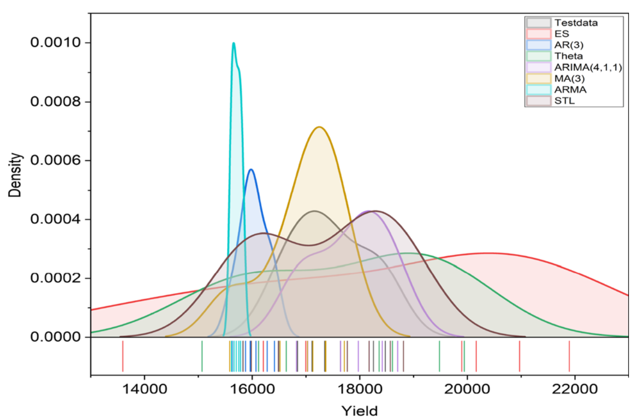
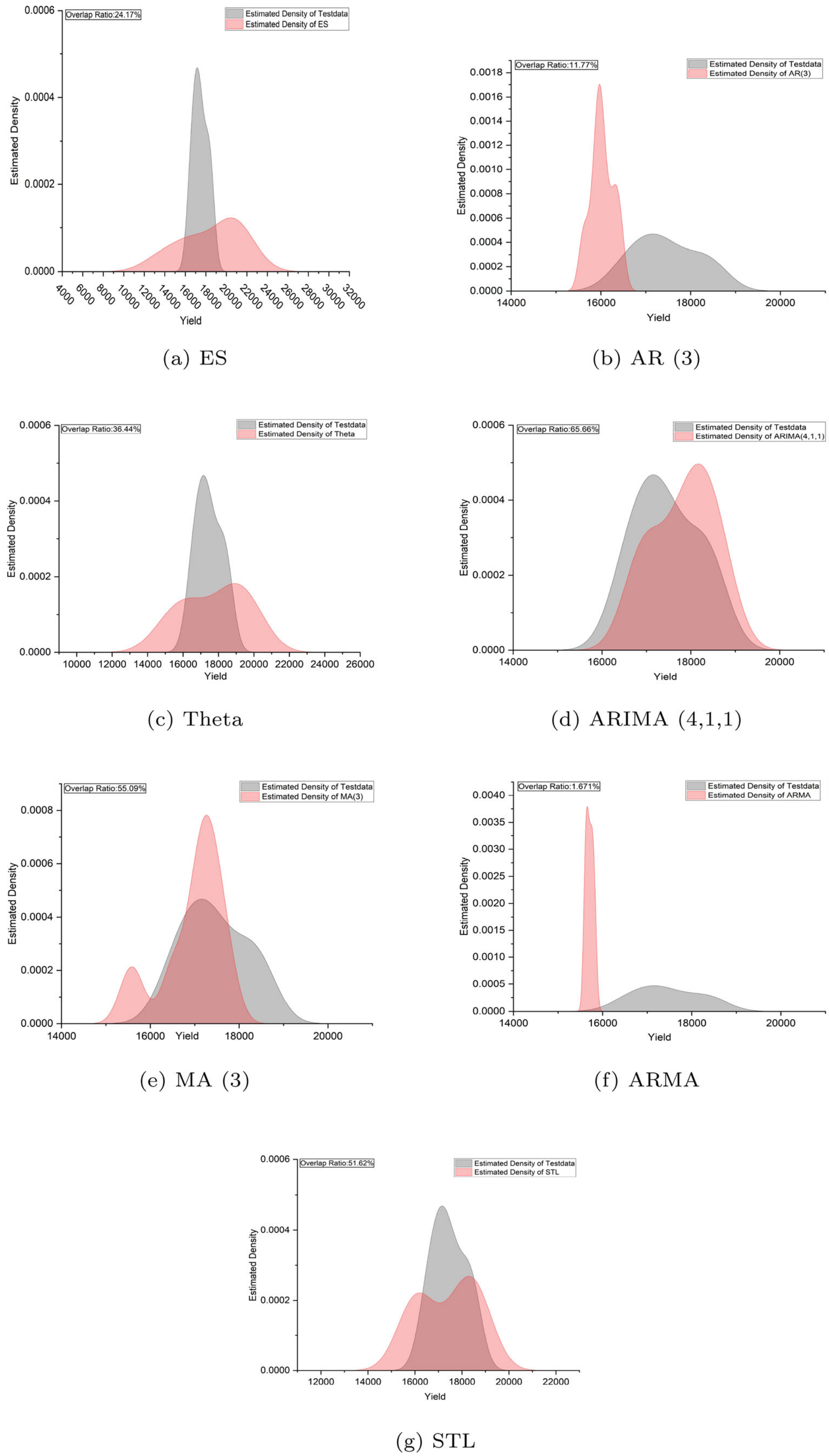


Fig. 8 Normal distribution of crop yield prediction using different time series-based ML algorithms and test data

perfect symmetry around the mean in normal distributions. Each of these performances has modest positive skewness, which is obtained due to the nature of the test dataset. All of these algorithms have different standard deviations. It is to be noted here that all the normal curves are run on 15% test data of $D2 - corn[''yield'']$. The normal curves show that the performance of ARMA and Theta have the larger peaks than others. Therefore, it shows a higher frequency of predictions near the mean yield. Finally, ARIMA(4,1,1) is performing best as it shows a similar curve to actual test data. In order to obtain a detailed analysis of individual algorithms mentioned in Fig. 8, the normal distributions to predict the corn yield against test data are shown in Fig. 9. The overlapping ratio of ARIMA(4,1,1) with test data is 65.66% which is the highest among all, while the overlapping ratio of ARMA is the lowest. In addition, the detailed parameters of the various time series-based ML algorithms used in the crop yield prediction, along with the values of the coefficients of the best algorithm among all the algorithms, are shown in Tables 10 and 11 (in Appendix), respectively.

Figure 10 shows the normal distribution curves of demand and supply data obtained from $D2$-corn and $D3$-corn. It is to be noted that both the normal curves are run on the same historical data from 1975 to 2021. Due to different variations in both demand and supply data, the curves differ from each other. The normal curve of demand data has a higher peak than that of supply data, as the weightage of the mean demand is more than the weightage of the mean supply. Both the normal curves start at the same point; however, they end at different points. The supply is stretched more, as shown here. It means that the normal curve of supply data has a higher standard deviation than that of demand data as the supply values at different years are more spread out and demand values of each year are more closer to the mean. Although the areas under both curves are nearly equal, however, the overlapping ratio is only 60.81% due to the difference in mean values of both. The difference in mean values is mentioned in Table 6. The yield value under the peak of the curve is called the mean yield. It is observed from Fig. 10 that the mean value of supply is slightly higher than that of demand. So, due to this difference, there is the residue of the crop at the end of each year. Here, both plots are not perfectly symmetrical around the mean. The presence of small positive skewness in both the normal curves indicates that the right tail is longer than the left tail for both demand and supply. It occurs due to a rise in both demand and supply for a few years. During the training phase of the price prediction model, the supply is calculated by adding all the three parameters, such as yield, residue and import, of D2-corn. This supply acts as one of the three inputs for the training of Block 4. During the prediction of the crop price for next year by Block 4, the yield ($\hat{Y}_{\text{best}}$) obtained from Block 1 is added with residue and import values of next year to obtain S^n. This S^n acts as one of the three inputs for price prediction.

The linear fits of demand and supply are shown in Fig. 11, The fit of demand has higher R^2 and $R^2_{Adjusted}$ values in comparison with that of supply. It indicates that the demand prediction can show the variance of demand effectively by using year as the only input variable. Here, the demand has a higher 'R'-value with the year parameter compared to that of supply. The demand has a correlation value of 0.97 with year, as shown in Table 6. So, it is preferred to predict demand using time alone. Both plots are fitted over some of the fields of $D2$-corn and $D3$-corn. Due to different trends of supply and demand, the fits of both can vary from each other. Figure 12 shows the performance of the autocorrelation function (ACF) and partial autocorrelation function (PACF) for demand at various



(a) ES

(b) AR (3)

(c) Theta

(d) ARIMA (4,1,1)

(e) MA (3)

(f) ARMA

(g) STL

Fig. 9 Detailed graphical analysis of individual crop yield prediction algorithms

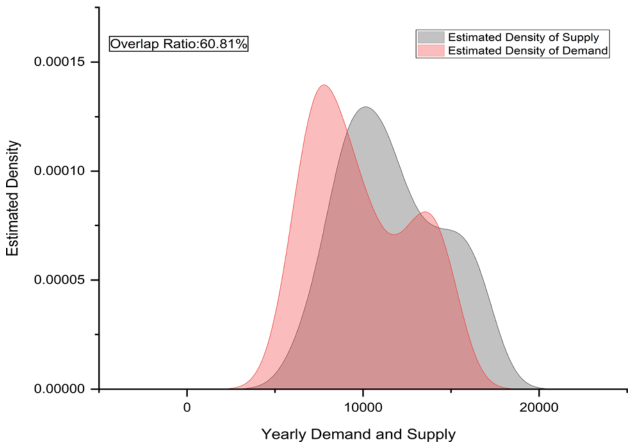


Fig. 10 Normal distribution of demand vs supply (1975 to 2021)

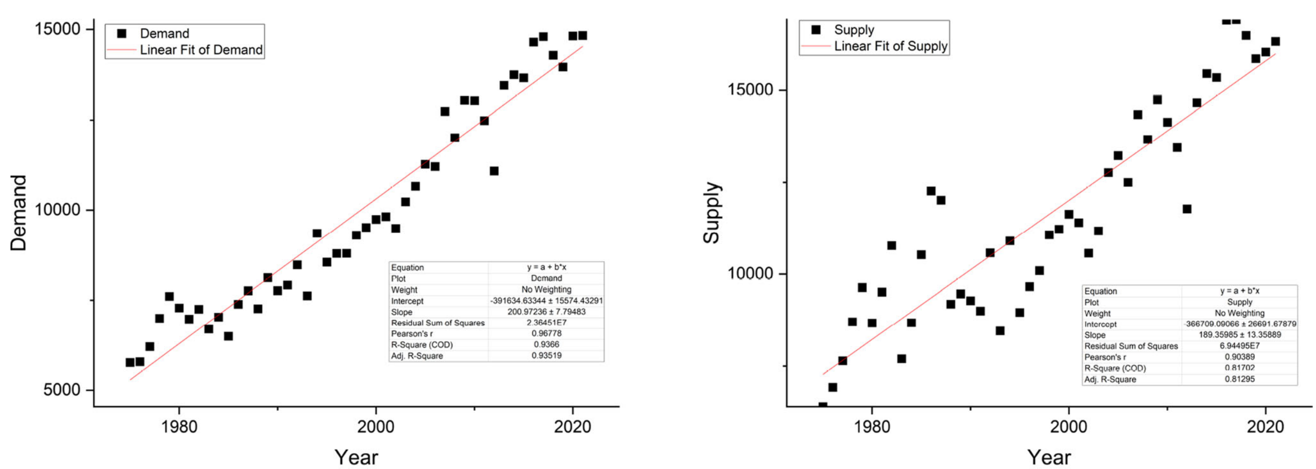
Table 6 Summary of supply demand analysis

Parameter	Yearly demand	Yearly supply
Mean	9908.15	11631.19
Standard deviation	2847.35	2872.44
Pearson correlation	0.968	0.904
R^2	0.937	0.817

lags. As the demand has very high ACF values for small lags, it shows the demand as nearly correlated with recent past values. The value of ACF decreases as the lag increases. It shows no seasonal pattern in demand over time, while the PACF has a significant value at lag 1 only. So, it shows that it is better to apply AR while predicting demand using time. Figure 13 shows the mean error in predicting demand for the time series methods, such as MA(3), MA(4), ES, AR(1) and ARIMA(1.0.1). All of these methods are fitted over the entire D3-corn. Then, the predictions are compared with the actual prices. It is observed that the MAE is lowest for ARIMA, while the same is worst for AR. The MA(3) and MA(4) have less error, hence it shows that the price trend depends upon the recent history to some extent. The large error of AR(1) shows the absence of any seasonal pattern in demand. The ARIMA(1,0,1), which includes both the AR and MA factors, has the least error in our proposed model. In addition, the detail of parameters of the various time series-based ML algorithms used in the demand prediction, along with the values of the coefficients of the best algorithm among all are shown in Tables 12 and 13 (in Appendix), respectively.

Figure 14 shows the ACF and PACF for price prediction at various lags. The price has very high ACF values for small lags, which shows the price as closely correlated with recent past values. The values of ACF decrease as the lag increases. Again this ACF value increases after some lags, the PACF has a significant value at lag 1 only. There is no understandable pattern in ACF and PACF values, so the time series method is ineffective for predicting price.

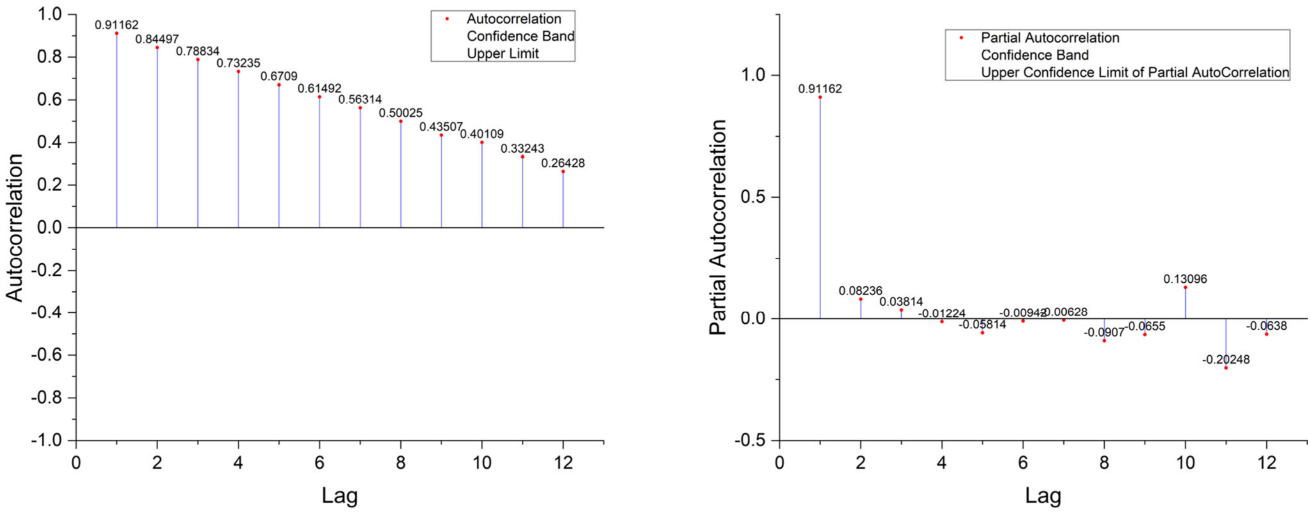
Fig. 15 shows the trend of change in crop price from 1975 to 2021. Initially, during 1994 and 1995, it is observed from Fig. 15b that MA(3) of supply is constant. However, that demand is increasing. This leads to an increase in price, as seen in Fig. 15a. Again, it is seen that, during the year 2006, there was a sharp transition in the price raising trend. It occurs due to a sudden decrease in supply with no variation in demand. During 2013, there was a solid down-trend which happened due to an increase



(a) Yearly Demand

(b) Yearly Supply

Fig. 11 Linear fit of supply and demand (1975–2021)



(a) ACF of Demand

(b) PACF of Demand

Fig. 12 ACF vs. PACF of demand (1975–2021)

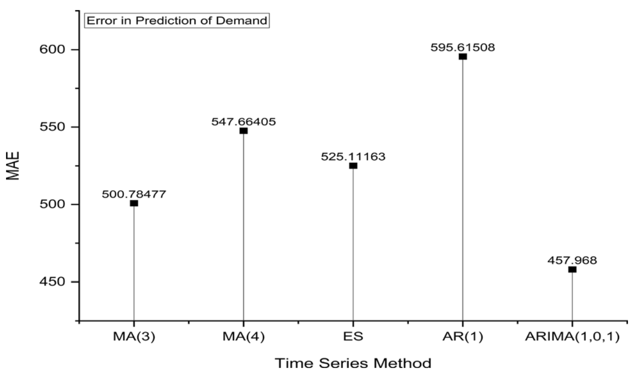


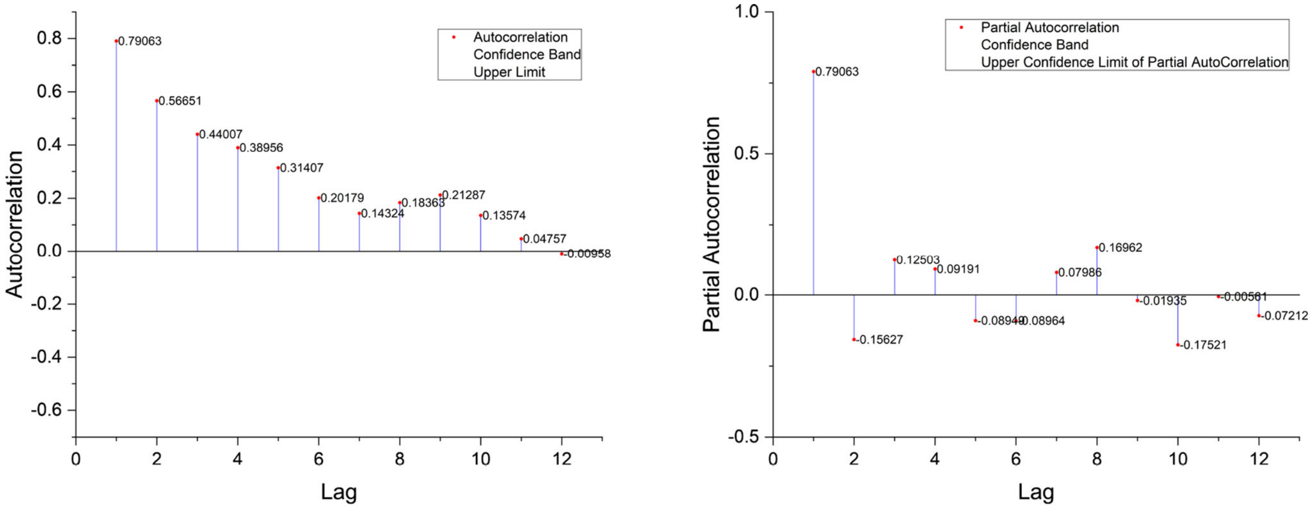
Fig. 13 Error in time series prediction of demand

in supply. This down-trend is evident from MA(3) value of supply, sma(3), and samples in Fig. 15b. Lastly, Fig. 15b shows that supply is decreasing at a higher rate than demand after 2017. This leads to an increase in price, which is noticeable in Fig. 15a. Hence, the changes in supply and demand affect the crop price, which indicates the significant importance of the proposed work over existing works.

Table 7 shows the summary of the price field of D3-corn. There are a total of 47 samples from the year 1975 to 2021. The mean price is 3.00149 USD per bushel. The median and standard deviation (SD) are also shown in this Table. The different measures of the statistical regression for predicting price are shown in Table 8. We have run the statistical regression up to degree 6. The value of R is found as 0.6902, which needs to be more significant to fit well with the input data.

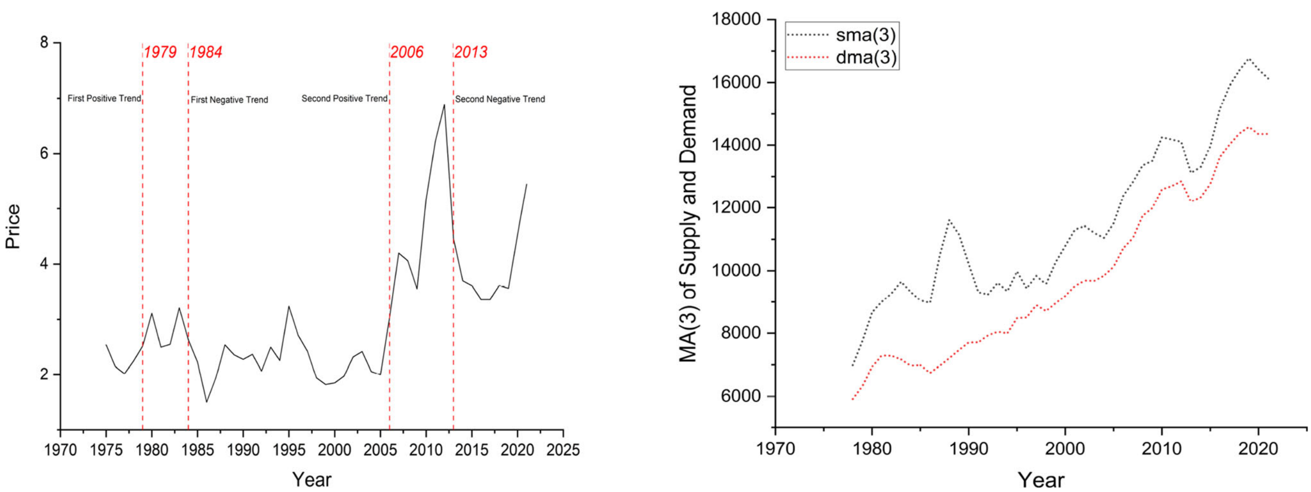
Figure 16 shows the predicted prices obtained by the methods such as SR, ARIMA and ML. It is to be noted here that DTR is used as an ML algorithm. These methods are trained over 70% of D23-corn, i.e., $D23 - corn.Train$, and subsequently tested over the remaining 30% test dataset, i.e., $D23 - corn.Test$. Due to the different predicting capabilities of these methods, the prediction results are notably varied from each other. Overall, it is seen that the SR algorithm deviates more from the actual price and the DTR algorithm runs most closely with the actual price, among others. There is a minimum value, pointed out in Fig. 16, at the year 1986 in actual data, which is best predicted by DTR. The ARIMA predicts it closer to the actual price. The detail of values of the coefficients of the ARIMA is shown in Table 14 (in Appendix). The minimum value is worst predicted by SR. There is a maximum value of the price in 2012, which is predicted to be identical to the proposed DTR. While SR could not attain this maximum value due to its polynomial terms. Similarly, the ARIMA can provide the predicted price lower than the value obtained by DTR. Finally, it is observed that the DTR method provides the predicted crop price value similar to the actual price. Thus, DTR is determined as the best price prediction model over others.

Figure 17 shows the results of various performance measures in testing the ML algorithms to predict crop price. The performance of test data varies from each other for different ML algorithms used in the proposed work. The crop price prediction in Block 4 is trained by DTR, RF, KNN and KR. The differences in MAE values of all the algorithms are noticed in Fig. 17. It is seen that DTR has the least value of MAE as it captures the pattern in the



(a) ACF of Price (b) PACF of Price

Fig. 14 ACF vs. PACF of Price



(a) Trends in price (b) MA(3) of Supply and Demand

Fig. 15 Price trend from 1975 to 2021

Table 7 Statistics of price (USD per bushel)

N	Mean	Median	SD	Sum	Minimum	Maximum
47	3.00149	2.54	1.17907	141.07	1.5	6.89

Table 8 Summary output of regression

Parameter name	Value
R	0.6902
R^2	0.4764
Standard error	0.8824
Average error	0.6209

training dataset very well among the four algorithms. The MAE by RF is the maximum among all, making RF the worst out of other ML algorithms used here for crop price prediction. The MSE is highest for RF, whereas the lowest is for DTR. The MSE of all the algorithms is highly varied from each other as it calculates upon squared errors. The DTR is the best as it fits best to the training dataset among these four algorithms. The errors by RF are found to be with higher magnitudes. The RMSE is also the highest for RF while the lowest for DTR. The RMSE also focuses on

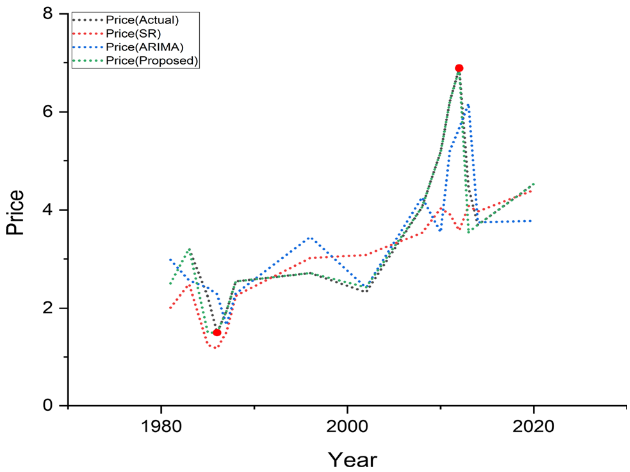


Fig. 16 Fit of price prediction using time series, regression, ML algorithm (proposed), ARIMA and actual

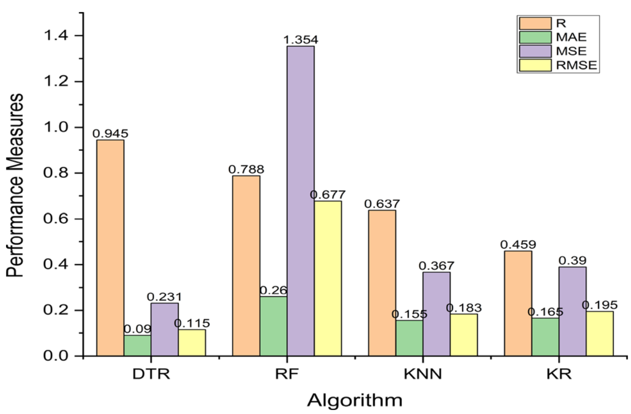


Fig. 17 Performance measures in price prediction using different ML algorithms

the magnitude of the error. However, it is lesser than MSE due to the square root terms after squaring. So the RMSE values of all ML algorithms used here are also far from each other. So, DTR is the best one in the proposed work as it can provide the highest accuracy during prediction. The errors by RF are found to be with high magnitudes also. The RMSE values show similar trends to MSE values. The second best training algorithm is found to be KNN and the results of KR are very near to KNN. In accordance with RMSE, the performance of the DTR algorithm used in the proposed work over existing works for price prediction is shown in Table 9. The DTR algorithm using rainfall data [15] is involved in the prediction of crop price. The method applies six different ML algorithms and finds DTR as best in terms of RMSE. The RMSE in this work is 3.8. Another work of crop price prediction using DTR in [12] obtains an RMSE of 63.8. Comparing the results obtained from such

Table 9 Result comparison with existing works

Method	RMSE	MAPE
Ranjani Dhanapal et al. [15]	3.8	–
Pandit Samuel et al. [12]	63.8	–
Wiwik Anggraeni et al. [34]	–	16.19%
Proposed	0.115	14.02%

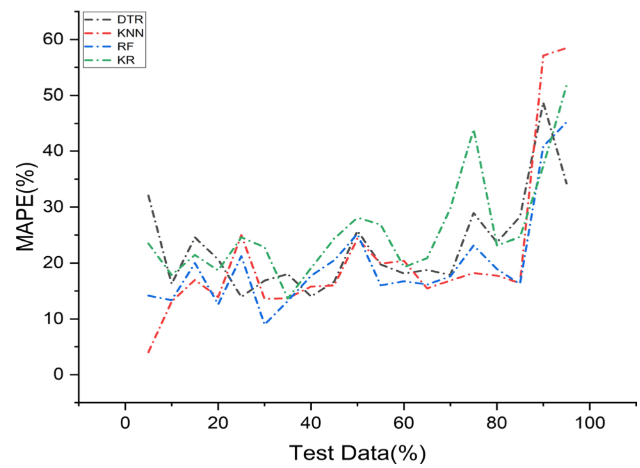


Fig. 18 MAPE in price prediction using different ML algorithms vs. test dataset size

existing works indicates that our proposed work predicts the crop price with the least RMSE.

Figure 18 shows the MAPE for all four ML algorithms used to predict the crop price. Here, 05% of the dataset is initially used for testing and 95% for training. This process is repeated in such a way that the test dataset increases by 5% more than the previous test data set size every time. The process continues up to 95% size of the test dataset. The MAPE results are varied for different ML algorithms at any specific percentage of the test dataset. The process is performed on DTR, RF, KNN and KR. It is observed that different algorithms perform better at different percentages of the test dataset size. The KNN performs best and DTR performs worst when test dataset is less than 20%. It means for smaller test dataset sizes, KNN fits the input data best. When the test dataset size is about 20%-70%, all algorithms have similar performance. Here, DTR shows a more steady and least volatile increase in the test dataset. When the test dataset increases beyond 70%, the spikes can be observed suddenly in the case of all four algorithms. It happens because of a decrease in training dataset sizes, making it difficult to train on the input datasets for all these algorithms. DTR performs best for a higher percentage of the test datasets, and KNN performs worst. It means that KNN finds it difficult to fit when the training dataset is less. Finally, DTR is shown as best for crop price prediction in terms of MAPE. The proposed price prediction using DTR is found to produce a better result of 14.02% than 16.19% [34] in terms of MAPE, as highlighted in Table 9. The Theil's U for predicting the crop price using DTR is found to be 0.456864. So, according to the U statistics measure mentioned earlier, it is revealed that the proposed model can be considered a better price prediction method.

6 Conclusion

An efficient crop price prediction model is proposed to predict the crop price using supply, demand and year data. As per our best knowledge, no prior research work predicting the crop price was done using such supply and demand of that particular crop, even after these being the prime factors of pricing a crop. So, from the proposed framework, a new research direction has emerged where the supply and demand of a particular crop are utilized under a single domain toward obtaining crop price prediction. The prediction model for crop pricing is well presented as a significant measure of crop yield. However, the effect of crop yield in predicting crop price needed to be addressed suitably in existing research works in this price prediction domain. An exhaustive analysis of the training datasets can provide ample insights to accurately forecast the required market price of the crop. Through the dataset prediction results, it is evident that the price prediction agent administers the process, suggesting that the proposed method can precisely define the characteristics of the crop price. Simulation results show that the ML approach is more suitable for crop price prediction than other time series methods and statistical regression techniques. A comprehensive result analysis is contributed after postprocessing of data obtained in simulations. Consequently, the convincing results obtained by data analysis can assist the farmers in deciding crop prices, which can improve their economy. Thus, the results observed from the precision and efficiency tests illustrate the effectiveness and versatility of the proposed ML algorithms for crop price prediction. By building an ML-based price prediction environment, the proposed method makes it feasible for the farmer to bring about the crop price prediction with less expert dependency and minimum prior knowledge. The proposed work can assist the farmers and organizations to take beneficial decisions ahead of time which can help them to improve their living standards. Experimenting with data prediction through a wide range of ML predictive algorithms can be observed as a basis for decision-making. Exploration of more crop price prediction parameters with respect to capita income, the health impact of crop, the effect of any disastrous event, the price of neighboring countries, future inflation, and availability of alternative crops can be included in the current framework to construct a more robust working model in the future. Further improvement in the computing efficiency of the training process is an intriguing option to be concentrated.

Appendix

See Tables 10, 11, 12, 13 and 14.

Table 10 Input parameters of different algorithms for yield prediction

Algorithm	Input parameters of algorithm										
	SP	*P*	*D*	*Q*	Trend	TS	SS	*S*	Deseasonalize	*S* degree	*S* jump
ES	12	–	–	–	Add	–	–	Auto	–	–	–
AR	–	3	–		–			–	–		
Theta	12	–	–					Auto	True		
ARIMA (4,1,1)	–	4	1	1				–	–		
MA		–	–	3							
ARMA		3		1							
STL		–		–		3	7			1	1

P = number of lagged forecast errors; *D* = number of differencing needed; *Q* = number of autoregressive terms; SP= number of observations in seasonal period; *S* = smoothing parameter; *SS* = length of seasonal smoother; *TS* = length of trend smoother; *S degree* = seasonal degree; *S jump* = seasonal jump

Table 11 Coefficients and performance summary of the model using ARIMA (4,1,1) for crop yield

Components	Coef	Std err	*z*	*P*> | z |	[0.025	0.975]
ARIMA (4,1,1)	Coefficients of the model for crop yield					
ar.L1	− 0.5124	0.235	− 2.184	0.029	− 0.972	− 0.053
ar.L2	− 0.3712	0.312	− 1.191	0.234	− 0.982	0.240
ar.L3	− 0.2561	0.261	− 0.983	0.326	− 0.767	0.255
ar.L4	− 0.1349	0.174	− 0.774	0.439	− 0.476	0.207
ma.L1	− 0.9467	0.229	− 4.134	0	− 1.396	−0.498
sigma2	4.737e+06	1.2e+06	3.956	0	2.39e+06	7.08e+06
ARIMA (4,1,1)	Performance summary of the model for crop yield					
Ljung-Box (L1) (Q):	0.41	Jarque-Bera (JB):	3.01	Log Likelihood	− 346.182	–
Prob(Q):	0.52	Prob(JB):	0.22	AIC	704.364	
Heteroscedasticity (H):	0.27	Skew:	− 0.62	BIC	714.189	
Prob(H) (two-sided):	0.02	Kurtosis:	3.61	HQIC	707.859	

coef Coefficients; *std err* Standard error; *AIC* Akaike information criterion; *BIC* Bayesian information criterion; *HQIC* Hannan-Quinn information criterion

Table 12 Input parameters of different algorithms for demand prediction

	Input parameters of algorithm							
	SP	*P*	*D*	*Q*	Trend	TS	SS	*S*
MA(3)	–	–	–	3	–	–	–	–
MA(4)				4				
ES	12			–	Add			Auto
AR (1)	–	1			–			–
ARIMA (1,0,1)		1	0	1				

Table 13 Coefficients and performance summary of the model using ARIMA(1,0,1) for demand prediction

Components	Coef	Std err	*z*	*P*>| z |	[0.025	0.975]
ARIMA(1,0,1)	Performance coefficients of the model for price prediction					
const	396.3324	273.511	1.449	0.147	− 139.739	932.404
ar.L1	0.9971	0.028	35.305	0	0.942	1.052
ma.L1	0.1525	0.19	0.803	0.422	− 0.22	0.525
sigma2	327.5957	56.788	5.769	0	216.293	438.898
ARIMA (1,0,1)	Performance summary of the model for demand prediction					
Ljung-box (L1) (Q)	4.9	Jarque-Bera (JB):	41.16	Log likelihood	− 231.42	–
Prob (Q)	0.03	Prob(JB):	0	AIC	470.839	
Heteroscedasticity (H)	4.03	Skew:	1.13	BIC	478.721	
Prob (H) (two-sided)	0	Kurtosis:	6.68	HQIC	473.87	

Table 14 Coefficients and performance summary of the model using ARIMA (1,0,1) for price prediction

Components	Coef	Std err	z	P>| z |	[0.025	0.975]
ARIMA(1,0,1)	Coefficients of the model for price prediction					
Const	3.0495	1.538	1.983	0.047	0.035	6.064
ar.L1	0.9486	0.091	10.466	0	0.771	1.126
ma.L1	0.2624	0.137	1.908	0.056	− 0.007	0.532
sigma2	0.2534	0.045	5.677	0	0.166	0.341
ARIMA (1,0,1)	Performance summary of the model for price prediction					
Ljung-box (L1) (Q)	0.01	Jarque-Bera (JB):	9	Log Likelihood	− 40.233	–
Prob(Q)	0.93	Prob(JB):	0.01	AIC	88.466	
Heteroscedasticity (H)	4.56	Skew:	0.85	BIC	96.347	
Prob(H) (two-sided)	0	Kurtosis:	4.1	HQIC	91.497	

Acknowledgements We thank National Institute of Technology Durgapur, India, for providing access to the computational resources and literatures.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest All authors of this research paper declare that they have no conflict of interest.

References

1. Rehman U, Bahar F, Saad A, Bhat M, Singh L, Mahdi S, Mir MS, Bashir M, Bhat TA, Palmo T (2022) Trends, challenges and prospects of underutilized food crops under temperate ecologies: a review
2. Rob V (2021) Cattaneo a poverty reduction through the development of inclusive food value chains. J Integ Agricult 20(4):964–978
3. von Braun J, Afsana K, Fresco LO, Hassan M (2021) Food systems: seven priorities to end hunger and protect the planet. Nature Publishing Group
4. Tian X, Engel BA, Qian H, Hua E, Sun S, Wang Y (2021) Will reaching the maximum achievable yield potential meet future global food demand? J Clean Prod 294:126285
5. Rakhra M, Soniya P, Tanwar D, Singh P, Bordoloi D, Agarwal P, Takkar S, Jairath K, Verma N (2021) Crop price prediction using random forest and decision tree regression: a review. Mater Today Proc. https://doi.org/10.1016/j.matpr.2021.03.261
6. Fafchamps M (1992) Cash crop production, food price volatility, and rural market integration in the third world. Am J Agricult Econ 74(1):90–99
7. Bennett AJ, Bending GD, Chandler D, Hilton S, Mills P (2012) Meeting the demand for crop production: the challenge of yield decline in crops grown in short rotations. Biol Rev 87(1):52–71
8. Fischer R, Byerlee D, Edmeades G (2014) Crop yields and global food security. Canberra ACT ACIAR, 8–11
9. Smith TR (1935) The wheat surplus. Geograph Rev 25(1):107–116
10. James HS, Sykuta ME (2006) Farmer trust in producer-and investor-owned firms: evidence from Missouri corn and soybean producers. Agribusiness Int J 22(1):135–153
11. Choi J-S, Helmberger PG (1993) How sensitive are crop yields to price changes and farm programs? J Agricult Appl Econ 25(1):237–244
12. Samuel P, Sahithi B, Saheli T, Ramanika D, Kumar NA (2020) Crop price prediction system using machine learning algorithms. Quest J J Softw Eng Simul
13. Medar R, Rajpurohit VS, Shweta S (2019) Crop yield prediction using machine learning techniques. In: 2019 IEEE 5th international conference for convergence in technology (I2CT), pp 1–5. IEEE
14. Akshay Prassanna S, Harshanand B, Srishti B, Chaitanya R, Kirubakaran Nithiya Soundari S, Kumar VM, Varshitha Chennamsetti VG, Maurya PK (2020) Crop value forecasting using decision tree regressor and models. Eur J Mol Clin Med, 7(2): 2020
15. Dhanapal R, AjanRaj A, Balavinayagapragathish S, Balaji J (2021) Crop price prediction using supervised machine learning algorithms. J Phys Conf Ser 1916:012042
16. Shao YE, Dai J-T (2018) Integrated feature selection of Arima with computational intelligence approaches for food crop price prediction. Complexity 2018:1–17
17. Lavanya K, Raguchander T, Iyengar N (2013) Svm regression and Sonn based approach for seasonal crop price prediction. Int J u-and e-Serv Sci Technol 6(4):155–168
18. Kantanantha N, Serban N, Griffin P (2010) Yield and price forecasting for stochastic crop decision planning. J Agricult Biol Environ Stat 15(3):362–380
19. Ghutake I, Verma R, Chaudhari R, Amarsinh V (2021) An intelligent crop price prediction using suitable machine learning algorithm. ITM Web Conf 40:03040
20. Bayona-Oré S, Cerna R, Tirado Hinojoza E (2021) Machine learning for price prediction for agricultural products
21. Dhanasekaran K, Ramprasath M, Sathiyamoorthi V, Poornima N, Jayaraj IA (2021) Meta-learning based adaptive crop price prediction for agriculture application. In: 2021 5th international conference on electronics, communication and aerospace technology (ICECA), pp 396–402. IEEE
22. Sellam V, Poovammal E (2016) Prediction of crop yield using regression analysis. Indian J Sci Technol 9(38):1–5
23. Jain A, Marvaniya S, Godbole S, Munigala V (2020) A framework for crop price forecasting in emerging economies by analyzing the quality of time-series data. arXiv preprint arXiv:2009.04171
24. Gamage A, Kasthurirathna D (2019) Agro-genius: crop prediction using machine learning. Int J Innov Sci Res Technol, 4(10)
25. Mulla SA, Quadri S (2020) Crop-yield and price forecasting using machine learning. Int J Anal Exp Modal Anal 12:1731–1737

26. Brunda SS, Nimish L, Khan A (2020) Crop price prediction using random forest and decision tree regression
27. Sabu KM, Kumar TM (2020) Predictive analytics in agriculture: forecasting prices of arecanuts in kerala. Procedia Comput Sci 171:699–708
28. Yuan CZ, Ling SK (2020) Long short-term memory model based agriculture commodity price prediction application. In: Proceedings of the 2020 2nd international conference on information technology and computer communications, pp 43–49
29. Yuan CZ, San WW, Leong TW (2020) Determining optimal lag time selection function with novel machine learning strategies for better agricultural commodity prices forecasting in malaysia. In: Proceedings of the 2020 2nd international conference on information technology and computer communications, pp 37–42
30. Gale D (1955) The law of supply and demand. Math Scandinavica, 155–169
31. Mishra MK, Sisodia B, Rai V (2019) Forecasting technique of price of potato of Uttar Pradesh. J Pharm Phytochem 8(3):60–62
32. Kumari RV, Venkatesh P, Ramakrishna G, Sreenivas A (2019) Chilli price forecasting using auto regressive integrated moving average (Arima). Int Res J Agricul Econ Stat 10(2):290–295
33. Ohyver M, Pudjihastuti H (2018) Arima model for forecasting the price of medium quality rice to anticipate price fluctuations. Procedia Comput Sci 135:707–711
34. Anggraeni W, Mahananto F, Rofiq MA, Andri KB, Zaini Z, Subriadi AP, et al. (2018) Agricultural strategic commodity price forecasting using artificial neural network. In: 2018 International seminar on research of information technology and intelligent systems (ISRITI), pp 347–352. IEEE
35. Feed Grains: Yearbook Tables. https://www.ers.usda.gov/data-products/feed-grains-database/feed-grains-yearbook-tables/. Accessed: 2010-09-30
36. Chicco D, Warrens MJ, Jurman G (2021) The coefficient of determination r-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. Peer J Comput Sci 7:623