Indian Journal of Agricultural Economics 79: 4 (2024):978-992
DOI:10.63040/25827510.2024.04.006

A Novel Ensemble Machine Learning Approach for Forecasting Oilseeds Prices in India

Ankit Kumar Singh[1], Ranjit Kumar Paul[2], Anita Sarkar[1], Md. Yeasin[2], Kanchan Sinha[2], Soumen Pal[2] and A.K. Paul[2]

ABSTRACT

The oilseed crops hold significant agricultural importance globally, particularly in India, where they contribute significantly to economic growth and food security. However, predicting its prices accurately remains challenging due to complex factors such as supply-demand dynamics, weather fluctuations, and global economic conditions. This study investigates price forecasting for major oilseed crops using traditional statistical models like autoregressive integrated moving average (ARIMA) and simple exponential smoothing (SES), alongside machine learning (ML) algorithms including artificial neural networks (ANN), random forest (RF), support vector regression (SVR), and K-nearest neighbors (KNN). The research employs a weighted ensemble (WE) approach based on the cuckoo search (CS) algorithm for weight optimization and model confidence set (MCS) for superior model selection. The ensemble model integrates predictions from multiple models, aiming to enhance forecast accuracy and reliability. Monthly wholesale price data of major oilseed crops from January 2010 to June 2024 across various markets from different states of India, sourced from AGMARKNET, have been analyzed. Results show that the proposed CS-WE model outperformed all the candidate models and suggest using an ensemble approach rather than a single model in managing risks and improving decision-making in the oilseed sector.

Keywords: Ensemble, machine learning, MCS, price prediction, CS, oilseed crops

JEL codes: C53, Q11, Q13

I

INTRODUCTION

The oilseeds sector has shown remarkable growth in global agriculture over the past three decades, expanding annually by a rate of 4.1 per cent, which exceeds the growth rates of agriculture and livestock products. Oilseed crops are crucial in achieving various Sustainable Development Goals (SDGs) by improving economic growth, ensuring food security, and promoting environmental sustainability. They provide essential nutrients and income to millions of smallholder farmers, alleviating poverty (SDG 1) and addressing hunger issues (SDG 2). Oilseed crops are cultivated for their nutritious seeds, which are rich in oil, proteins and vitamins A, B, E & K. The oil extracted from oilseeds is used in cooking and as a healthy alternative to other oils due to its high content of unsaturated fats. The oilseed cake is nutritious and can feed the milk cattle, poultry and pigs. The oilseed crops like Sunflowers also play a crucial role in the ecosystem. They attract pollinators such as bees and butterflies, supporting biodiversity and positively affecting the yields of other crops. Despite challenges such as adverse weather and fluctuating global prices, oilseeds have performed well in meeting increasing domestic demand. India ranks fourth globally in the vegetable oil economy, trailing only the USA, China, and Brazil. Oilseeds cover 13 per cent of

[1]ICAR-Indian Agricultural Research Institute, New Delhi 110 012. [2]ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110 012.

India's total cropped area, contribute 3 per cent to the Gross National Product (GNP), and represent 10 per cent of the total value of agricultural commodities. The oilseed crops are divided into two groups, the edible and the non-edible. The edible groups include Groundnut, Rapeseed, Soybean, Sunflower, Sesame, Safflower, and Niger, and the non-edible group, which consists of Castor and Linseed. In India, soybean dominates with 38 per cent of total production and 44 per cent of cultivation area, followed by rapeseed-mustard and groundnut, each contributing 27 per cent to production while occupying 24 per cent and 20 per cent of the total area, respectively (https://oilseeds.dac.gov.in/). These scenarios signify the importance of oilseed crops, mainly soybean, mustard, and groundnut, in policy formation, economic growth, food security, and eco-friendly farming practices. Consequently, this study aims to predict oilseed prices, allowing stakeholders to make well-informed decisions, such as developing government policies and planning for farmers. This approach seeks to stabilize markets and improve production efficiency, as accurate forecasts are vital for effective risk management and optimal resource allocation.

Numerous stochastic processes have been applied in the literature to model and forecast a time series. Autoregressive integrated moving average (ARIMA) model has been a dominant approach in time series analysis. In agricultural data, numerous applications of the ARIMA model are well-documented in the literature (Bhardwaj *et al*., 2014; Box, 2015; Paul, 2015; Jadhav and Kamble, 2017; Noureen *et al*., 2019; Rakshit *et al*., 2021; Srivastava *et al*., 2022; Mapuwei *et al*., 2022). Alternative time series forecasting models like exponential and trend-seasonal component models are also considered due to the limitations of traditional statistical models in capturing complex patterns, nonlinear relationships, and high-dimensional data in real-world applications (Goodwin, 2010; Livera *et al*., 2011). Simple exponential smoothing (SES) is a forecasting method that works best when data shows neither a trend nor seasonality in its patterns. Forecasts are calculated using weighted averages, where the weights decrease exponentially, and the smallest weights are associated with the oldest observations (Hyndman *et al*., 2008). However, due to frequent fluctuations, predicting non-linear time series data and series with complex patterns may not be easy to forecast using stochastic models. Machine learning (ML) algorithms were developed to tackle the non-linear and complex patterns in time series. These advantages of ML models have led to a growing trend towards utilising ML algorithms for predicting nonlinear time series data in recent times (Ahmed *et al*., 2010; Vui *et al*., 2013; Liakos *et al*., 2018; Chaitra and Meena, 2023). These methods include artificial neural networks (ANN), support vector regression (SVR), random forest (RF) and K-Nearest Neighbors (KNN) etc. Shengwei *et al*. (2017) explored factors influencing agricultural price fluctuations and developed a prediction model using least squares SVR for wholesale agricultural prices. Paul *et al*. (2022) explored the efficacy of various ML algorithms like ANN, SVR and RF in predicting wholesale prices of brinjal, demonstrating the superior performance of ML techniques compared to other models. Jena *et al.* (2023) focused on constructing a low-complexity, adaptive ANN-based model for crop yield prediction. Several authors (Paul *et al*., 2023; Mohanty *et al*., 2023; Chelliah *et al*., 2024) in the literature have explored the performance of machine learning models for price forecasting in agriculture. Lathal *et al.* (2024) explored the ANN and SVR models for improving palm oil import forecast accuracy and reported that these models outperform the traditional time series models.

While ML models generally exhibit higher accuracy than traditional statistical models, predicting agricultural commodity prices remains challenging due to numerous influencing factors. These include supply and demand dynamics, weather fluctuations, global economic conditions, government policies, exchange rates, technological advancements, geopolitical events, etc. Both traditional stochastic time series models and ML models often struggle to capture the complex behaviours of these data, making it difficult to rely on a single model for accurate price forecasting. In response to this challenge, ensemble machine-learning approaches have become essential. These methods combine predictions from multiple models, offering a more robust and reliable forecasting strategy in such complex and dynamic environments. Sinta *et al*. (2014) utilised the ensemble method to forecast rice crop prices. Mitra and Paul (2017) integrated traditional time-series models with machine-learning algorithms to enhance the precision of price predictions. Galicia *et al*. (2019) introduced ensemble learning approaches for forecasting time series data, exploring dynamic and static strategies for updating weights. Ribeiro and Coelho (2020) recommended using ensembles to predict agricultural commodity prices, highlighting its ability to improve model accuracy and reduce decision-making risks. Anshori *et al*. (2021) used the cuckoo search (CS) method to optimise world crude oil price estimation. Swathi *et al.* (2022) introduced a unique CS optimisation-based model for stock prediction. Sun *et al*. (2021) proposed a model for carbon price forecasting that combines decomposition techniques with ensemble models optimised through an optimization algorithm. Abdelhamid and Alotaibi (2022) developed a two-level ensemble model, where the first level integrates RF, SVR, and light gradient boosting machine models, and the second level employs elastic net regression. Yeasin *et al*. (2024) developed an ensemble model for forecasting tropical cyclones in India.

This study mainly focuses on the price prediction of important oilseed crops, viz., soybean, groundnut, and mustard. The cuckoo search optimization-based weighted ensemble (CS-WE) approach is proposed to predict prices of oilseed crops using machine learning models (ANN, RF, KNN, SVR and XGBoost), ARIMA and SES as candidate models. The study utilises monthly wholesale price data for oilseed crops from January 2010 to June 2024 taken from AGMARKNET (https://agmarknet.gov.in/) for different markets in India. The detailed framework of the paper is mentioned as follows. Section II gives a detailed introduction to the models used in this study. Section III provides the empirical studies. Section IV discussed the results and findings of the studies. In Section V, conclusions are mentioned.

II

METHODOLOGY

2.1 *Candidate Model*

This section outlines various well-known machine learning models considered candidate models within an ensemble approach.

2.1.1 *Auto Regressive Integrated Moving Average (ARIMA) Model*

A popular statistical method for analyzing and predicting time series data is the ARIMA (Auto Regressive Integrated Moving Average) model. It combines three essential elements—the moving average (MA), autoregressive (AR), and differencing (I)—to forecast future values based on historical data. It is represented as ARIMA (p, d, q), where p stands for the AR order, d for the degree of differencing, and q for the MA order. This paradigm works well for handling time series data that exhibit seasonality and trends. Autocorrelation function (ACF) and partial autocorrelation function (PACF) studies are used to determine values of p and q. The form of ARIMA (p, d, q) model is given by

$$\Phi(B)(1-B)^d X_t = \Theta(B)\varepsilon_t \qquad \dots (1)$$

where, $\Phi(B)$ is AR and $\Theta(B)$ is MA polynomial of order p and q respectively; B is the backshift operator and ε_t is a white noise process.

2.1.2 *Exponential Smoothing*

Simple exponential smoothing (SES) is the most straightforward of the exponential smoothing techniques. Time series without a discernible trend or seasonal pattern can be forecasted using this method. Weighted averages are used in forecast calculations; the earliest observations are assigned the smallest weights, and the weights decline exponentially. The one-step-ahead forecast for time t+1 based on the observations in the series $y_1, \ldots, y_t$ is given by equation (2).

$$y_{t+1|t} = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \cdots \qquad \dots\dots(2)$$

Where $0 \le \alpha \le 1$ is the smoothing parameter. The parameter α controls the rate at which the weights decrease.

2.1.3 *Artificial Neural Network (ANN)*

The ANN is an ML model of neurons grouped into the input, hidden, and output layers. The input layer takes input data, the output layer gives us the predicted results, and the hidden layer captures complicated connections between the inputs and outputs. In the time series domain, ANN incorporates a sequence ranging from y_{t-1} to y_{t-n}, where n represents the number of lagged observations considered. To model time series data, a nonlinear function (f) works on time series ranges (y_t) from y_{t-1} to y_{t-n} where n is the number of lag using the equation (3).

$$y_t = w_0 + \sum_{j=1}^{h} w_j f(w_{0j} + \sum_{i=1}^{n} w_{ij} y_{t-i}) + e_t \qquad \dots (3)$$

In equation (3), w_{ij}, and w_j represent the weights to model, h represents the number of hidden nodes, n represents the number of input nodes, and e_t represents an error term. ANN is widely recognized and applied to forecast the prices of agricultural commodities.

2.1.4 *k-Nearest Neighbour (k-NN)*

The k-nearest neighbours (k-NN) algorithm is a non-parametric technique that predicts numerical outcomes by comparing the similarity of data points, often using distance measures. In kNN regression, an approach is to compute the average numerical outcomes from the k nearest neighbours. The formulation of the k-NN model can be represented mathematically, as shown in the equation (4).

$$Output_{kNN} = f(u) \qquad \dots (4)$$

Where u contains a sequence of time series from y_{t-1} to y_{t-n} where n is the number of lag.

2.1.5 *Support Vector Regression (SVR)*

Support Vector Regression (SVR) effectively addresses non-linear relationships between input variables and the target variable by using kernel functions to transform data into higher-dimensional spaces. This characteristic enables SVR to be useful in regression tasks where complex associations exist between input and target variables. The SVR model can be formulated as in equation (5).

$$f(y) = \boldsymbol{z}.K(y) + b \qquad \dots (5)$$

Where $\boldsymbol{y}$ contains a sequence of time series from y_{t-1} to y_{t-n} where n is the number of lag, kernel function $K(.)$ relocates non-linear data to higher-dimensional feature space, weight vector represented as 'z' and bias term as 'b'. The estimated function of the dataset, i.e., $f(y)$ in equation (5) is the output from the model.

2.1.6 *Random Forest (RF)*

Random Forest (RF) is an ensemble learning method based on decision trees. It demonstrates strong performance across various applications. Decision trees are useful in handling both classification and regression problems. Decision trees are particularly effective for regression tasks where the target variable is continuous. RF employs a technique known as bootstrap aggregation or bagging, where each decision tree is trained using a randomly selected subset from the entire training dataset. Let $r(\cdot)$ be the function derived from the training of RF, which is utilized for predicting y_t based on the historical time series values employing n lagged variables, the forecasted value $(\hat{y}_t)$ is determined by the equation (6):

$$\hat{y}_t = r(y_{t-1}, y_{t-2}, \dots, y_{t-n}) \ , t = n+1, \dots, k. \qquad \dots (6)$$

Some applications of the RF model for prediction in time series data can be found in Lahouar *et al*. (2015) and Moon *et al.* (2018).

2.1.7 *Extreme Gradient Boosting (XGBoost)*

XGBoost is a powerful ML algorithm commonly utilized for regression and classification tasks. Its main goal is to optimize the weights and predictions of individual trees to minimize the loss function, which quantifies the difference between the predicted and actual values. During the training process, XGBoost progressively adds new trees to the ensemble, with each tree aimed at correcting the errors made by the previous ones. It continuously adjusts weights and predictions by applying gradient boosting to reduce the loss. In the end, the final prediction is obtained from a weighted average of the predictions from each individual tree. The prediction using XGBoost is given by equation (7):

$$\mathrm{P(X)} = \sum_i \mathrm{w_i} * \mathrm{g_i(X)} \qquad \dots (7)$$

Where P(X) is the predicted value, w_i is the weight assigned to the ith tree, and g_i(X) is the prediction of the ith tree.

2.2 Ensemble model

Ensemble methods capture different aspects of the underlying data patterns by combining different models and providing more accurate and reliable forecasts. Prediction of multiple individual models is combined using weights. As the unweighted ensemble (UWE) approach provides equal weights to each model, thus it ignores the performance of the model. The fixed-weighted ensemble methods may overcome this limitation as the weights of candidate models in the weighted ensemble approach are different and determined using an optimization algorithm based on the performance of candidate models. So, the forecast from the weighted ensemble method $(\widehat{y_{fw}})$ is computed by equation (8):

$$\widehat{y_{fw}} = \left(\sum_{i=1}^{N} w_i \ \widehat{y_i} \right) \qquad \dots (8)$$

In equation (8), the weight associated with ith candidate model is w_i such that $\sum_{i=1}^{N} w_i = 1$.

Cuckoo search (CS) is a population-based optimization algorithm that draws inspiration from the breeding behaviour of cuckoo species. The population-based optimization technique optimises the weights, which enhances the chances of discovering the global optimum while preventing entrapment in local optima (Yeasin and Paul, 2024). Cuckoo search is preferable to other optimization algorithms, such as Artificial Bee Colony (ABC) and Genetic Algorithm (GA). CS is an algorithm par excellence because it entails fewer parameters than other state-of-the-art algorithms (Kirti and Singla, 2020).

The output of the ensemble model depends on the performance of candidate models. The better performance of the model leads to improved forecasts from the ensemble model. The model confidence set (MCS) algorithm selects the better-performing candidate model. Suppose $C_0=\{C_1,\ldots C_m\}$ be a group of m candidate models based on time series $\{y_t\}$ ($t = 1, 2 \ldots n$) and being evaluated by a loss function, say $l_i = \{l_{i,t}\}$. Then, the t-test statistic of the MCS procedure is represented in equation (9):

$$t_{ij,\tau} = \frac{\bar{\Delta}_{ij}}{\sqrt{\widehat{var}(\Delta_{ij})}} \text{ and } t_{i.} = \frac{\bar{\Delta}_{i.}}{\sqrt{\widehat{var}(\bar{\Delta}_{i.})}} \quad \ldots.(9)$$

Where $\bar{\Delta}_{ij} \equiv \frac{1}{n}\sum_{t=1}^{n} \Delta_{ij,t}$, $\bar{\Delta}_i \equiv \frac{1}{m}\sum_{j=1}^{m} \bar{\Delta}_{ij}$, $\Delta_{ij,t} = l_{i,t} - l_{j,t}$ for all i,j belongs to C_0

Test statistics for the MCS algorithm can be written as follows:

$$T_D \equiv \sum_{j=1}^{m} t_{i.}^2$$

The Test statistics T_D has χ^2_{m-1}.

The detailed architecture of the Ensemble model is mention in Figure 1.

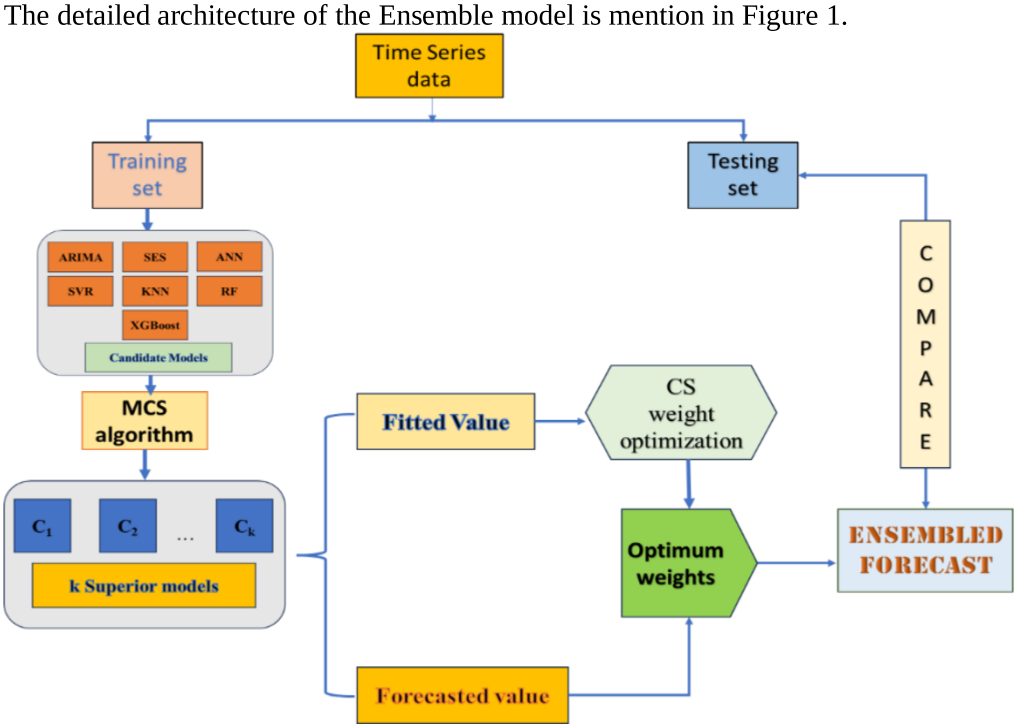


Figure 1. The Flowchart of Ensemble Model.

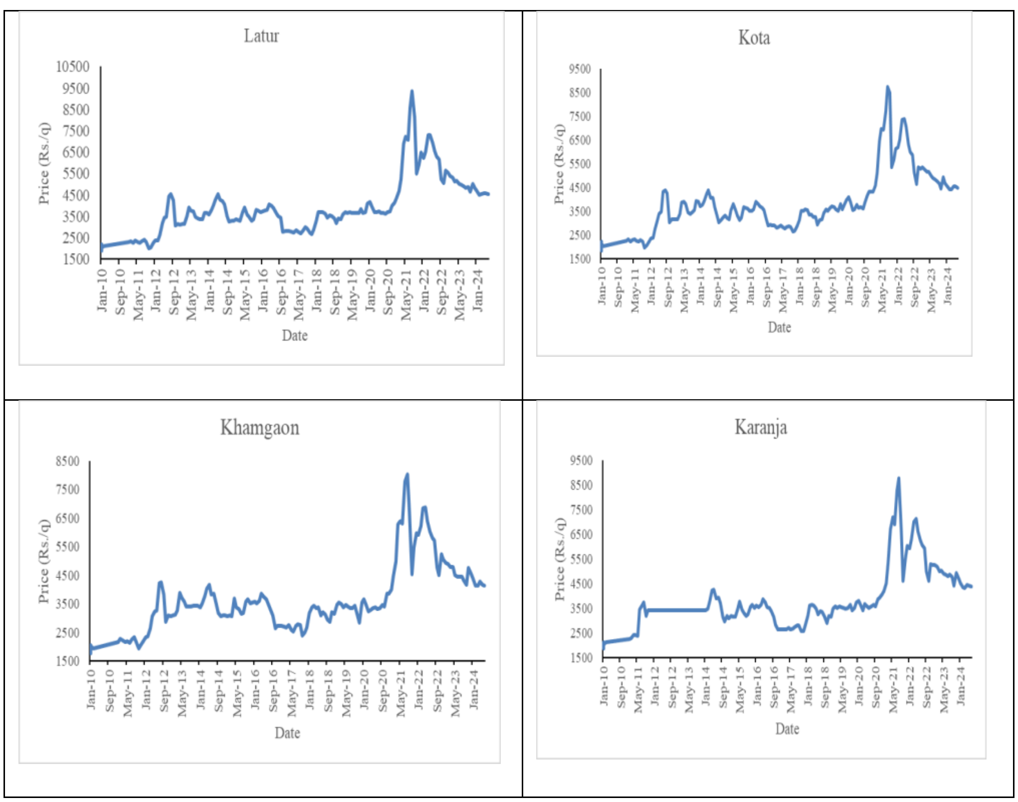
III
EMPIRICAL STUDIES

3.1 *Dataset*

The price data for the oilseed crops, viz., soybean, mustard, and groundnut, are considered from different markets in India. A detailed list of crop-wise markets are mentioned in Table 1. The study utilizes monthly wholesale price data for oilseed crops from January 2010 to June 2024, taken from AGMARKNET (https://agmarknet.gov.in/). The missing observations in the dataset were imputed using appropriate statistical techniques. This price dataset has 174 observations divided into training and testing sets in an 80:20 ratio. The training set with 142 observations has been used for model development, and the remaining 32 observations have been used for checking the accuracy and performance of models. Figure 2 shows the time series plot for the price of oilseed data corresponding to all 12 markets.

TABLE 1. LIST OF MARKETS FOR DIFFERENT OILSEED CROPS

Crop (1)	Market (2)
Soybean	Latur, Khamgaon, Karanja, Kota, Bhawani mandi, Ujjain, Dewas
Mustard	Baran, Kota, Sriganganagar
Groundnut	Bikaner, Chomu



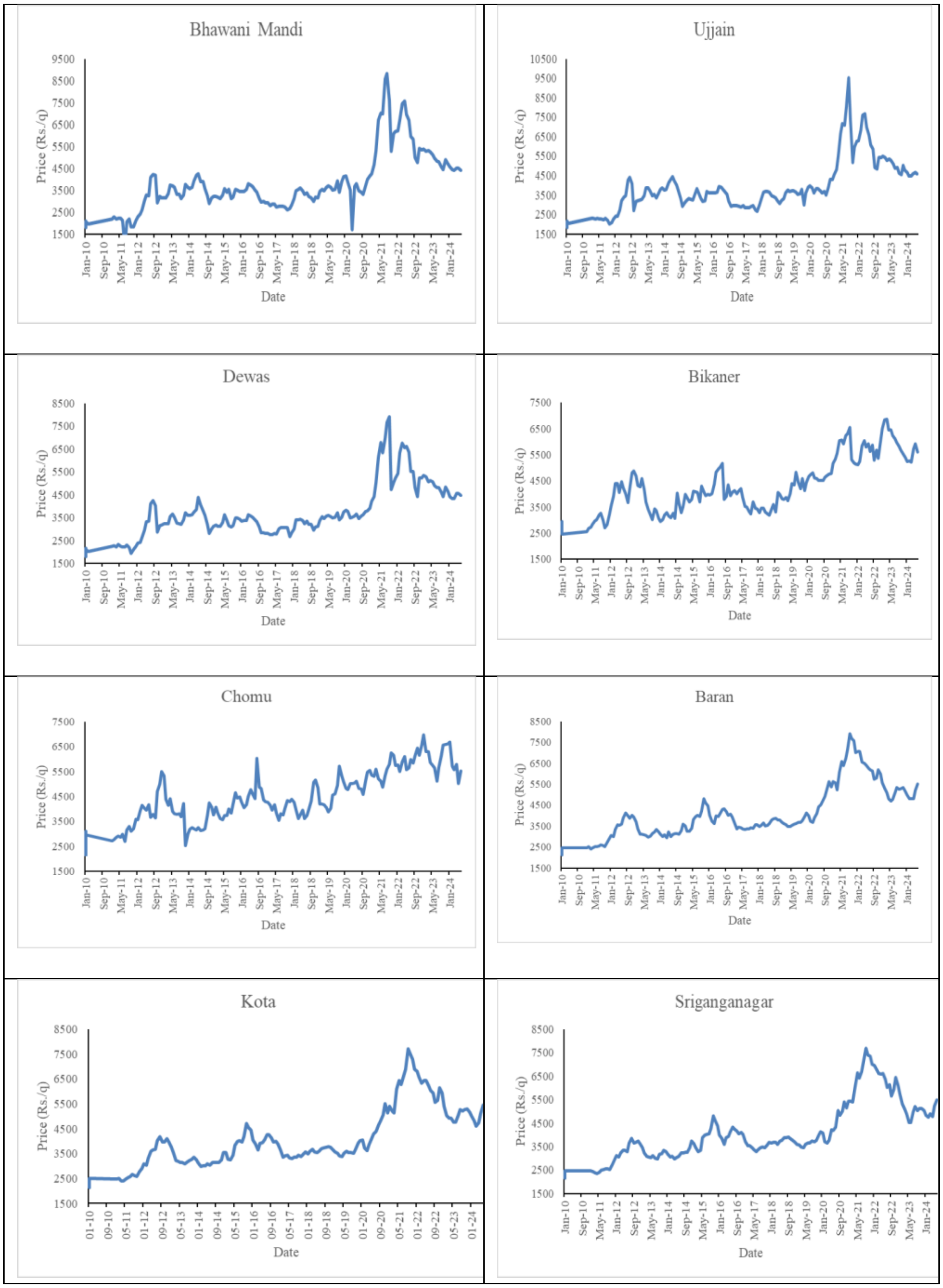


Figure 2. The Time Series Plot of All the Markets

The summary statistics of the wholesale price of different oilseeds from respective markets are reported in Table 2. A perusal of Table 2 indicates that the average price of soybean is highest in the Latur market, whereas the lowest average price is observed in the Khamgaon market. The highest and lowest average prices of groundnut have been observed in Chomu and Bekaner market. The average price of mustard remains almost the same in all the studied markets. All the price series are positively skewed, and most are leptokurtic. The variability, as measured in terms of coefficient of variation (CV), is the highest in the Bhawani market for soybean and lowest in the Chomu market for groundnut.

TABLE 2. SUMMARY STATISTICS OF PRICE SERIES

Crop (1)	Market (2)	Mean (Rs/Q) (3)	Median (4)	Standard Deviation (5)	CV (per cent) (6)	Kurtosis (7)	Skewness (8)	Minimum (9)	Maximum (10)
Soybean	Latur	3867.35	3672.33	1358.07	35.12	2.24	1.28	1901.27	9342.86
	Khamgaon	3581.50	3387.65	1197.02	33.42	1.80	1.20	1769.44	8023.05
	Karanja	3771.21	3477.08	1220.82	32.37	2.67	1.41	1865.39	8763.00
	Kota	3800.07	3574.28	1307.23	34.40	2.06	1.25	1853.85	8738.10
	Bhawani	3720.62	3478.73	1374.28	36.94	1.93	1.21	903.14	8832.06
	Ujjain	3831.68	3623.93	1347.79	35.17	2.23	1.28	1828.72	9509.56
	Dewas	3654.40	3429.41	1188.24	32.52	1.52	1.12	1799.48	7930.00
Groundnut	Bekaner	4250.12	4079.31	1081.23	25.44	-0.66	0.44	2456.59	6863.33
	Chomu	4431.53	4261.68	1081.72	24.41	-0.75	0.22	2185.42	6968.75
Mustard	Baran	4049.78	3689.20	1269.61	31.35	0.39	0.93	2132.61	7912.80
	Kota	4018.47	3669.91	1210.02	30.11	0.36	0.90	2134.38	7706.52
	Srigangnagar	4037.28	3708.85	1251.07	30.99	0.32	0.92	2153.15	7679.16

3.2 *Ensemble Methods*

Seven different traditional time series and machine learning models, viz., ARIMA, SES, ANN, SVR, kNN, RF and XGBoost models, are used to fit the monthly wholesale price data of oilseed crops from different markets. The results are combined using a fixed weighted ensembled approach (CS-WE). The weights corresponding to different models are optimised using population-based optimization algorithms (CS). The ensemble model uses information from all the candidate models. To eliminate poor-performing models, the MCS algorithm reduces the noise and makes the forecast more robust. Table 3 shows the list of superior candidate models in different markets. The data indicate that no single model performs better in all the cases. Each model is eliminated by the MCS algorithm at least once, which suggests the need for an ensemble-based approach for prediction. The result obtained from CS-WE is compared with all the candidate models used in this study, viz., ARIMA, SES, ANN, SVR, kNN, RF and XGBoost model, using various error metrics mentioned in section 3.4. All the analyses in the present study are carried out using the R software package.

TABLE 3. LIST OF SUPERIOR CANDIDATE MODELS IN DIFFERENT MARKETS.

Crops	Market	Candidate models						
		ARIMA	SES	ANN	SVR	RF	kNN	XGBoost
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Soybean	Latur			✓	✓	✓	✓	
	Khamgaon			✓	✓	✓	✓	
	Karanja				✓	✓	✓	
	Kota			✓	✓	✓	✓	
	Bhawani Mandi			✓		✓	✓	
	Ujjain			✓	✓	✓	✓	
	Dewas			✓	✓	✓	✓	
Groundnut	Bikaner					✓	✓	
	Chomu	✓	✓	✓		✓		✓
Mustard	Baran		✓	✓	✓	✓	✓	✓
	Kota			✓		✓	✓	✓
	Sriganganagar			✓			✓	

3.3 Performance Measure

The ensemble models are empirically evaluated using accuracy metrics, specifically Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE) and symmetric mean absolute percentage error (SMAPE), which are estimated as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad ....(10)$$

$$\text{MAPE} = \frac{1}{N}\sum_{\text{i}=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \quad ....(11)$$

$$\text{MAE} = \frac{1}{N}\sum_{\text{i}=1}^{N}|y_i - \hat{y}_i| \quad ....(12)$$

$$SMAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{\left(\frac{|y_i|+|\hat{y}_i|}{2}\right)} \quad ....(13)$$

Where, y_i is the actual value, $\hat{y}_i$ is the forecasted value, and N is the total number of observations.

IV

RESULTS AND DISCUSSION

The Accuracy measures of different selected superior candidate models for predicting monthly wholesale prices of oilseed crops of different markets are given in Table 4. The CS-WE significantly outperforms all the candidate models, including stochastic and machine learning models, across all three oilseed crops and twelve markets. A detailed analysis based on the four accuracy metrics (i.e., RMSE, MAPE, MAE and SMAPE) values shows that among candidate models, The RF, ANN and kNN model consistently performs well in most markets compared to other models. Still, no model performs best in all the cases. These differences in the performance of models from market to market and crop to crop provide strong evidence for the usefulness of the ensemble model. The UWE approach performs better than the candidate models in most cases. In the case of the soybean crop, the UWE model outperforms all the candidate models in terms of the lower value of all four accuracy measures. In the case of the mustard crop, the UWE model outperforms all the candidate models except the kNN model in the Baran market and the Kota market except the ANN model in the Sriganganagar market. In the case of the groundnut crop, the UWE model outperforms all the candidate models except the RF model in the Bikaner market and the ARIMA and XGBoost model in the Chomu market. The proposed model CS-WE has the lowest value for all four accuracy measures corresponding to all 12 markets of all three different oilseed crops.

TABLE 4. ACCURACY MEASURES OF DIFFERENT SELECTED SUPERIOR CANDIDATE MODELS FOR PREDICTING MONTHLY WHOLESALE PRICES OF OILSEED CROPS OF DIFFERENT MARKET.

Oilseed Crop	Market	Model	Accuracy Measures			
			RMSE (4)	MAPE (per cent) (5)	MAE (6)	SMAPE (7)
(1)	(2)	(3)				
Soybean	Latur	ANN	1555.882	25.9	1396.030	0.242
		SVR	2236.933	30.6	1863.926	0.377
		RF	2001.764	35.5	1827.382	0.292
		kNN	1463.816	25.0	1300.774	0.218
		UWE	1210.748	19.9	1100.712	0.189
		CS-WE	1108.622	15.5	897.129	0.154
	Khamgaon	ANN	1572.689	29.3	1370.092	0.246
		SVR	1915.264	30.5	1672.343	0.370
		RF	1888.558	36.6	1704.393	0.297
		kNN	1266.268	23.9	1121.510	0.209
		UWE	1069.187	19.9	967.350	0.183
		CS-WE	930.735	15.2	797.745	0.151
	Karanja	SVR	2012.971	28.9	1670.798	0.353
		RF	1932.039	35.9	1764.229	0.294
		kNN	1808.412	33.4	1642.478	0.277
		UWE	1133.382	19.5	1006.103	0.180
		CS-WE	977.402	13.6	762.284	0.138
	Kota	ANN	1909.665	34.8	1755.104	0.288
		SVR	2142.144	29.1	1757.200	0.357
		RF	1757.432	31.9	1619.203	0.269

		kNN	1565.388	28.1	1433.891	0.242
		UWE	1270.127	21.6	1147.865	0.199
		CS-WE	1115.709	14.9	864.907	0.151
	Bhawani Mandi	ANN	2167.540	34.5	1882.284	0.373
		RF	1973.712	35.9	1795.099	0.293
		kNN	2014.644	30.5	1754.340	0.373
		UWE	1208.857	17.4	983.572	0.176
		CS-WE	974.200	12.9	737.429	0.130
	Ujjain	ANN	1676.669	24.7	1395.310	0.270
		SVR	2120.065	27.8	1702.795	0.340
		RF	1565.109	27.2	1409.799	0.235
		kNN	1692.486	29.5	1518.025	0.251
		UWE	1113.423	15.6	904.680	0.156
		CS-WE	1096.491	15.1	873.703	0.150
	Dewas	ANN	1346.795	22.6	1182.129	0.231
		SVR	1909.028	29.6	1658.114	0.358
		RF	1395.891	26.0	1278.943	0.226
		kNN	1307.080	24.3	1207.380	0.215
		UWE	927.971	13.3	721.727	0.132
		CS-WE	904.313	12.3	681.047	0.124
GroundNut	Bikaner	RF	494.739	7.2	411.020	0.071
		kNN	511.750	7.6	430.163	0.074
		UWE	502.311	7.4	419.610	0.072
		CS-WE	485.605	6.9	399.860	0.069
	Chomu	ARIMA	769.592	10.7	661.036	0.115
		SES	822.425	11.6	717.319	0.125
		ANN	838.207	11.8	725.469	0.127
		RF	809.823	11.4	707.306	0.123
		XGBoost	803.333	11.1	687.374	0.120
		UWE	804.390	11.3	699.221	0.122
		CS-WE	438.601	6.2	366.250	0.061
Mustard	Baran	SES	1264.280	20.8	1116.177	0.184
		ANN	1058.451	16.4	891.845	0.149
		SVR	1768.790	23.7	1481.233	0.271
		RF	1068.127	17.2	937.890	0.157
		kNN	980.903	15.3	854.745	0.144
		XGBoost	1138.780	17.4	984.826	0.165
		UWE	1026.943	16.1	908.893	0.153
		CS-WE	924.189	14.1	807.497	0.136
	Kota	ANN	1126.520	17.1	900.224	0.152
		RF	1017.551	16.5	886.417	0.151
		kNN	928.170	14.9	814.229	0.140
		XGBoost	1010.596	16.4	875.329	0.150
		UWE	997.986	15.9	851.409	0.146
		CS-WE	847.706	12.7	732.656	0.126
	Sriganganagar	ANN	989.527	13.9	738.777	0.125
		kNN	1014.646	16.6	894.367	0.153
		UWE	947.715	14.5	773.134	0.132
		CS-WE	860.813	13.3	757.953	0.129

V
CONCLUSION

This study highlights the importance of employing ensemble techniques for agricultural price forecasting, particularly in dynamic and complex markets like oilseed. Among the candidate models, the RF, ANN, and kNN models demonstrate robust performance across most markets, underscoring their reliability in capturing price trends. The ensemble model emerges as the most effective strategy for predicting monthly wholesale prices of oilseed crops across diverse markets, viz., soybean, mustard and groundnut. It consistently outperforms individual candidate models, including stochastic and machine learning approaches, regarding RMSE, MAPE, MAE, and SMAPE metrics across five markets. This study underscores the value of ensemble methods in enhancing predictive accuracy and informing strategic decisions in agricultural commodity markets. The performance of the ensemble approach depends on the candidate models. So, the candidate model should be properly selected for ensemble purposes. The study may be extended by incorporating deep learning models in the ensemble framework.

Received August 2024. *Revision accepted August 2024.*

REFERENCES

Abdelhamid, A. A., and Alotaibi, S. R. (2022). Optimized two-level ensemble model for predicting the parameters of metamaterial antenna, *Computers, Materials and Continua*, *73*(1).

Ahmed, N. K., Atiya, A. F., Gayar, N. E., and El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric reviews*, *29*(5-6), 594-621.

Anshori, M. Y., Rahmalia, D., Herlambang, T., and Karya, D. F. (2021). Optimizing adaptive neuro fuzzy inference system (ANFIS) parameters using cuckoo search (Case study of world crude oil price estimation), *Journal of Physics: Conference Series*, 1836, 012041.

Bhardwaj, S. P., Paul, R. K., Singh, D. R., and Singh, K. N. (2014). An empirical investigation of ARIMA and GARCH models in agricultural price forecasting, *Economic Affairs*, 59(3), 415-428.

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: forecasting and control*, John Wiley and Sons.

Chaitra, B., & Meena, K. (2023). Forecasting crop price using various approaches of machine learning, *2023 International Conference on Innovations in Engineering and Technology (ICIET)*,1-5. IEEE, July.

Chelliah, B. J., Latchoumi, T. P., & Senthilselvi, A. (2024). Analysis of demand forecasting of agriculture using machine learning algorithm. *Environment, Development and Sustainability*, *26*(1), 1731-1747.

De Livera, A. M., Hyndman, R. J., and Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing, *Journal of the American statistical association*, *106*(496), 1513-1527.

Galicia, A., Talavera-Llames, R., Troncoso, A., Koprinska, I., and Martínez-Álvarez, F. (2019). Multi-step forecasting for big data time series based on ensemble learning. *Knowledge-Based Systems*, *163*, 830-841.

Goodwin, P. (2010). The Holt-Winters Approach to Exponential Smoothing: 50 Years Old and Going Strong, *Foresight*, *19*(19), 30-33.

Hyndman, R., Koehler, A. B., Ord, J. K., and Snyder, R. D. (2008). *Forecasting with exponential smoothing: the state space approach*. Springer Science and Business Media.

Jadhav, A. V., & Kamble, K. B. (2017). Prediction of stock prices in oil sectors using ARIMA model, *International Journal of Mathematics Trends and Technology-IJMTT*, 51.

Jena, P. R., Majhi, B., Kalli, R., and Majhi, R. (2023). Prediction of crop yield using climate variables in the south-western province of India: a functional artificial neural network modeling (FLANN) approach. *Environment, Development and Sustainability*, 25(10), 11033-11056.

Kirti & Anshu Singla (2020). Csbiist: cuckoo search-based intelligent image segmentation, *Nature-Inspired Computation and Swarm Intelligence: Algorithms, Theory and Applications*, *323*.

Lahouar, A., and Slama, J. B. H. (2015). Random Forests Model for One Day Ahead Load Forecasting. *Irec2015 the sixth international renewable energy congress* 1-6. IEEE, March.

Lathal, K. N., Rao, V.S., Rao, S., Reddy, A., and Sreenivasulu, K. N. (2024). Enhancing forecasting accuracy of palm oil import to India using machine learning techniques, *Indian Journal of Agricultural Economics*, 79(2), 214-230.

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., and Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, *18*(8), 2674.

Mapuwei, T. W., Ndava, J., Kachaka, M., and Kusotera, B. (2022). An application of time series ARIMA forecasting model for predicting tobacco production in Zimbabwe, *American Journal of Modeling and Optimization*, *9*(1), 15-22.

Mitra, D., & Paul, R. K. (2017). Hybrid time-series models for forecasting agricultural commodity prices, *Model Assisted Statistics and Applications*, *12*(3), 255-264.

Mohanty, M. K., Thakurta, P. K. G., and Kar, S. (2023). Agricultural commodity price prediction model: A machine learning framework. *Neural Computing and Applications*, *35*(20), 15109-15128.

Moon, J., Kim, Y., Son, M., & Hwang, E. (2018). Hybrid short-term load forecasting scheme using random forest and multilayer perceptron, *Energies*, *11*(12), 3283.

Noureen, S., Atique, S., Roy, V., and Bayne, S. (2019). Analysis and application of seasonal ARIMA model in energy demand forecasting: A case study of small scale agricultural load, *2019 IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS)*, 521-524. IEEE, August.

Paul, R. K. (2015). ARIMAX-GARCH-WAVELET model for forecasting volatile data, *Model Assisted Statistics and Applications*, *10*(3), 243-252.

Paul, R. K., Yeasin, M., Kumar, P., Kumar, P., Balasubramanian, M., Roy, H.S., Paul, A.K. and Gupta, A. (2022). Machine learning techniques for forecasting agricultural prices: A case of brinjal in Odisha, India, *Plos one*, *17*(7), e0270553.

Paul, R. K., Das, T., and Yeasin, M. (2023). Ensemble of time series and machine learning model for forecasting volatility in agricultural prices, *National Academy Science Letters*, *46*(3), 185-188.

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

Ribeiro, M. H. D. M., and dos Santos Coelho, L. (2020). Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series, *Applied soft computing*, *86*, 105837.

Shengwei, W., Yanni, L., Jiayu, Z., & Jiajia, L. (2017). Agricultural price fluctuation model based on SVR, *2017 9th International Conference on Modelling, Identification and Control (ICMIC)*, 545-550. IEEE, July.

Sinta, D., Wijayanto, H., & Sartono, B. J. A. M. S. (2014). Ensemble K-nearest neighbors method to predict rice price in Indonesia. *Applied Mathematical Sciences*, *8*(160), 7993-8005.

Srivastava, A. B., Mishra, P., Singh, K. K., and Choudhri, H. P. S. (2022). Instability and production scenario of wheat production in Uttar Pradesh using ARIMA model and its role in food security. *Indian Journal of Economics and Development*, *18*(1), 181-188.

Sun, S., Jin, F., Li, H., and Li, Y. (2021). A new hybrid optimization ensemble learning approach for carbon price forecasting, *Applied Mathematical Modelling*, *97*, 182-205.

Swathi, T., Kasiviswanath, N., and Rao, A. A. (2022). Cuckoo search optimization and long-term memory-based stock price prediction model with improved classification performance, In *2022 IEEE 4th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA)*, 397-402. IEEE, October.

Vui, C. S., Soon, G. K., On, C. K., Alfred, R., and Anthony, P. (2013). A review of stock market prediction with Artificial neural network (ANN). In *2013 IEEE international conference on control system, computing and engineering*, 477-482. IEEE, November.

Yeasin, M., and Paul, R. K. (2024). Optisembleforecasting: optimization-based ensemble forecasting using MCS algorithm and PCA-based error index, *The Journal of Supercomputing*, *80*(2), 1568-1597.

Yeasin, M., Paul, R. K., and Shankar, S. V. (2024). Ensemble machine learning models for forecasting tropical cyclones in North Indian region, *Earth Science Informatics*, 1-10.