Journal of Animal & Plant Sciences, 34(5): 2024, Page: 1290-1302
ISSN (print): 1018-7081; ISSN (online): 2309-8694
https://doi.org/10.36899/JAPS.2024.5.0811

PRICE ESTIMATION OF SELECTED GRAINS PRODUCTS BASED ON MACHINE LEARNING FOR AGRICULTURAL ECONOMIC DEVELOPMENT IN TÜRKİYE

Abdulkadir Keskin[1,*], İrfan Ersin[2] and Abdulkadir Atalan[3]

[1]Department of Statistics, Istanbul Medeniyet University, Istanbul, Türkiye, https://orcid.org/0000-0002-4795-1028
[2]Vocational School of Social Sciences, Istanbul Medipol University, Istanbul, Türkiye, https://orcid.org/0000-0002-7407-3654
[3]Industrial Engineering Department, Çanakkale Onsekiz Mart University, Çanakkale, Türkiye, https://orcid.org/0000-0003-0924-3685
*** Corresponding author's Email:** abdulkadir.keskin@medeniyet.edu.tr

ABSTRACT

This study aims to estimate the price fluctuations of essential grain products, namely bread wheat (Triticum aestivum), durum wheat (Triticum durum), barley (Hordeum vulgare), and corn (Zea mays), in Türkiye using machine learning (ML) algorithms. Using data from January 2, 2020, to January 10, 2023, the study employs algorithms such as random forest (RF), neural network (NN), support vector machine (SVM), and linear regression (LR). Independent variables include oil prices, currency exchange rates, and grain production volumes. The random forest (RF) algorithm provided the best results with the highest R^2 values, while NN and LR showed relatively lower performance. The study highlights the significant impact of production and consumption volumes on grain prices and underscores the importance of ML algorithms in predicting these prices amidst changing conditions. Investments in agricultural technologies should be increased to improve data collection and analysis processes, as this is crucial for preventing price fluctuations in the agricultural sector.

Keywords: Agricultural products; grains; durum wheat; bread wheat; corn; barley; machine learning algorithms; price estimation

Published first online September 23, 2024 Published Final October 22, 2024

INTRODUCTION

In agricultural economics, prices significantly influence sector production. Uncertainties in commodity prices can lead to stagnation in the agricultural economy and reduce long-term investments (Elder, 2018; Elder and Serletis, 2010; JO, 2014; Ravazzolo and Rothman, 2013). Moreover, inflation significantly affects commodity prices (Abaidoo and Agyapong, 2022). The economic restrictions imposed by COVID-19, the contraction of trade, and the rise in production costs have led to inflation in food prices (Adewopo *et al.,* 2021; Armantier *et al.,* 2021; Bonam and Smădu, 2021). Nicola *et al.* (2020) found that, during the COVID-19 pandemic, the decrease in demand from the hospitality sector led to a 20% drop in agricultural commodity prices globally. Swinnen and McDermott (2020) noted that the pandemic and quarantine measures caused unexpected fluctuations in commodity prices and significant disruptions in food systems.

Since commodity prices affect global economic activities, accurate price forecasts are critical for both exporting and importing countries, producers, and intermediaries. Agricultural commodities are particularly significant in this regard. These products have historically been central to global trade, and events such as COVID-19 and the Russia-Ukraine war have underscored their importance in global markets (Sadiq *et al.,* 2022; Sokhanvar and Bouri, 2023). Agricultural prices, like other commodities such as gold, oil, aluminum, and coffee, are determined by the market, but market disruptions can impact prices (Sasmal, 2015). Thus, it is crucial to reduce agricultural price uncertainties and develop reliable forecasting models. The pricing of agricultural products involves big data, which can be utilized through data mining techniques. The use of artificial intelligence (AI), machine learning (ML), and deep learning algorithms in the agricultural sector is becoming increasingly crucial for predicting price fluctuations, analyzing market trends, and improving decision-making processes (Şengül and Gerşil, 2018; Bermpei *et al.,* 2023).

The literature offers various suggestions on ML algorithms used for forecasting agricultural product prices. Özdemir and Çılgın (2022) compared Support Vector Regression (SVR), Artificial Neural Networks (ANN), Random Forest (RF), and Autoregressive Integrated Moving Average (ARIMA) models for wheat prices. They found that the ARIMA model provided more effective results. Zou *et al*. (2007) significantly improved forecasting performance by combining ARIMA and ANN models. Zong and Zhu (2012) predicted the prices of ten agricultural products using ANN models and found that the radio-based ANN model performed best. Jha and Sinha (2013) revealed that a time-delay ANN model outperformed the ARIMA model in forecasting soybean and canola prices. Other studies have shown that RF and ANN models effectively forecast agricultural prices (Garai *et al.,* 2023; Kuradusenge *et al.,* 2023).

Machine learning algorithms are widely used for classifying complex relationships and price forecasting (Adetunji *et al.,* 2022; Çepni *et al.,* 2022; Sadorsky, 2022). This study employs four different machine learning algorithms: Random Forest (RF), Neural Networks (NN), Support Vector Machines (SVM), and Linear Regression (LR) to forecast the daily unit prices of major agricultural products (wheat, corn, barley) in Türkiye. The performance of these models is evaluated using metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Squared Error (MSE), and the Coefficient of Determination (R^2). The research aims to contribute to the development of agricultural economics by highlighting the importance of agricultural price forecasting.

The existing literature has demonstrated the effectiveness of machine learning algorithms in forecasting agricultural product prices. However, these algorithms have limited comparative performance analysis in predicting agricultural prices in Türkiye. This study contributes significantly to the literature by comparing the performance of four different machine learning algorithms (RF, NN, SVM, LR) in forecasting the prices of major agricultural products in Türkiye. The innovative aspect of this study is its comprehensive evaluation of these algorithms' performance to determine the most effective model for agricultural price forecasting.

MATERIALS AND METHODS

Data set and descriptive statistics: The research methodology flowchart is visualized in **Figure 1**. Briefly, after determining the prices of the grain products to be estimated in the research, daily data to be used in price estimation were obtained. Then, the data were arranged and made suitable for analysis. Firstly, the missing data were determined, and the missing data were completed using the nearby points method. Then, the data were subjected to logarithmic transformation. Dependent and independent variables were determined for four different agricultural products, and the analysis phase was started. Four suitable ML algorithms have been identified to determine price estimations of agricultural RF, NN, SVM, and LR. The algorithms were trained at first and then tested with the allocated 30% of the data with the determined ML algorithms. Finally, price estimations of bread wheat, durum wheat, barley, and corn were made using the algorithms, and the performances of the algorithms were compared.

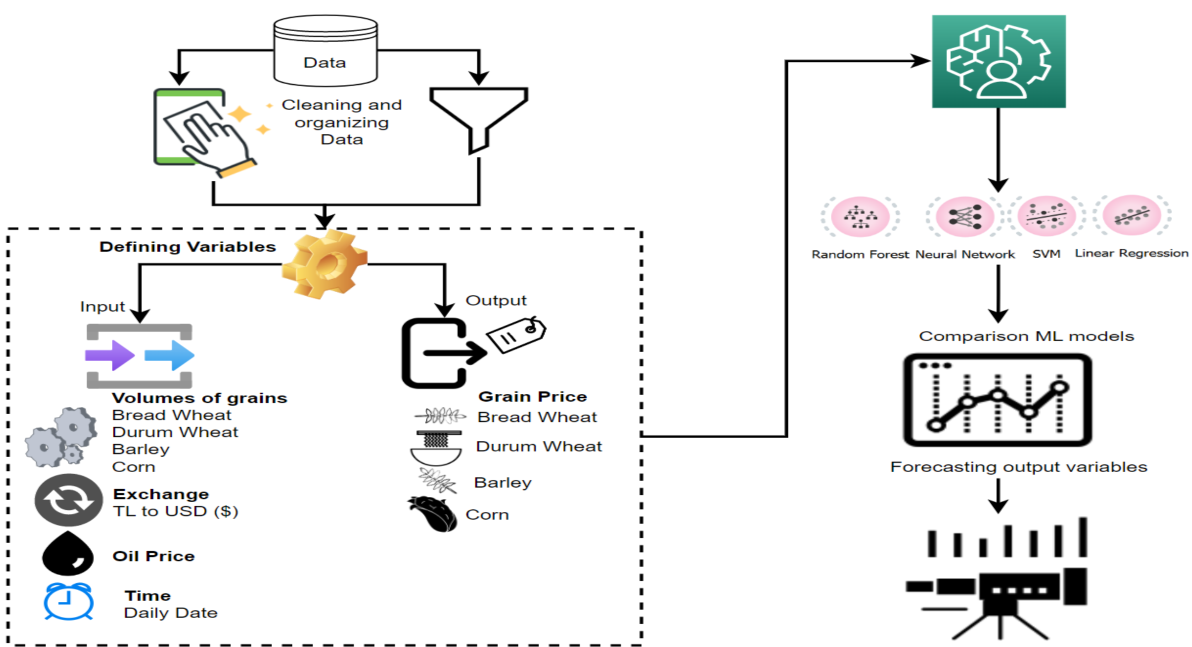


Figure 1. The flowchart of the research methodology

The data for this study were collected daily and included information on grain prices and trading volumes for the grains selected. The data was obtained from the Turkish Mercantile Exchange (TME) and covered January 2, 2020, and January 10, 2023 (https://www.turib.com.tr/en/). Table 1 summarizes the acronyms and data sources of the independent and dependent variables used in ML algorithms.

Table 1. Characteristics of the dependent and independent variables

Variable	Abbreviations	Unit*	Status	Source
Oil price	OP	$/lt	Input/ independent	INV
Exchange	USD	$/₺	Input/ independent	CBRT
Durum wheat volume	DWV	Kg	Input/ independent	TME
Durum wheat price	DWP	₺/kg	Output/dependent	TME
Bread wheat volume	BWV	Kg	Input/ independent	TME
Bread wheat price	BWP	₺/kg	Output/dependent	TME
Corn volume	CV	Kg	Input/ independent	TME
Corn price	CP	₺/kg	Output/dependent	TME
Barley volume	BV	Kg	Input/ independent	TME
Barley price	BP	₺/kg	Output/dependent	TME
Time (Daily)	DT	Day	Metadata/independent	**

Abbreviations=*$, US Dollar; TL₺, Turkish Lira; CBRT, Central Bank of the Republic of Türkiye; INV, Investing.com; lt, liter; kg, kilogram ** undefined

The daily prices of four agricultural products on the stock exchange were used as the dependent variable. The transaction volume refers to the total monetary value that security receives from purchase and sale. The volume of stock exchange transactions plays a vital role in determining the prices of products. ML algorithms included this study's daily transaction volume of four agricultural products.

Petrol price does have a significant impact on agricultural products both during the production process and transportation stage. The cost of fuel is a major factor in the cost of producing and transporting agricultural goods. High fuel prices can increase the cost of farming operations, including planting, harvesting, and transporting crops, which can, in turn, lead to higher prices for consumers. In this study, we used the exchange rate of the US dollar ($) to Turkish lira (₺) to predict the prices of four-grain products. The data of oil prices and US dollar prices are obtained from the website investing.com. In this study, we have used seven independent variables for each agricultural product, and the time variable is included as metadata in all models. Before analyzing the data, it was found that the observations were missing by 2%. The missing observations were completed using the mean of nearby points method (George and Mallery, 2021). Then, the logarithm of the dataset was taken to make it suitable for analysis. Descriptive statistics of the data set used in the research are presented in **Table 2**.

Table 2. Descriptive statistics of dependent and independent variables

Variable	Mean	Mode	Median	Variance	Minimum	Maximum	Skewness	Kurtosis	N
OP	4.1935	3.755140	4.2690	0.1512	2.9557	4.8605	-0.63	-0.22	789
USD	2.3144	2.166040	2.1408	0.1509	1.7688	2.9314	0.43	-1.41	789
DWV	13.989	12.81700	14.164	4.2790	6.1090	20.650	-0.10	0.68	789
DWP	1.1873	0.574504	0.9839	0.3032	0.4187	2.1041	0.38	-1.39	789
BWV	15.642	15.78750	15.728	2.7450	10.939	20.896	0.14	0.47	789
BWP	1.0677	0.477786	0.8775	0.2759	0.2945	2.0412	0.44	-1.31	789
CV	15.785	15.15220	15.958	2.3070	10.812	20.047	-0.17	0.00	789
CP	0.9393	1.677100	0.8618	0.3209	0.1689	1.8718	0.24	-1.48	789
BV	14.561	13.81640	14.562	2.4360	2.6250	19.394	-0.52	5.61	789

Abbreviations = Oil price; OP, Exchange; USD, Durum wheat volume; DWV, Durum wheat price; DWP, Bread wheat volume; BWV, Bread wheat price; BWP, Corn volume; CV, Corn price; CP, Barley volume; BV, Barley price; BP

Grain prices are determined by the interaction of several complex factors. These factors include weather conditions, harvest amount, demand and supply balance, prices in world markets, consumer preferences, agricultural policies, and exchange rates. For example, adverse weather conditions or pest outbreaks can affect the yield of grain crops, causing prices to rise. Increased demand or increased use of grain for biofuel production could also increase prices. Additionally, governments' subsidy policies and trade restrictions can also affect prices. Grain prices are shaped at a balance point formed by the combination of these complex factors and are often very volatile. The price changes of four-grain products four different ML algorithms predict are shown in **Figure 2.** Generally, until January 2022, there was little upward trend or price fluctuation. Prices have typically followed a stable course. However, significant fluctuations and increases in the prices of the four agriculture products occurred after January 2022. This period coincides with the global epidemic crisis, the rise in oil prices due to the Russia-Ukraine war, and a massive shipping problem for grain products.

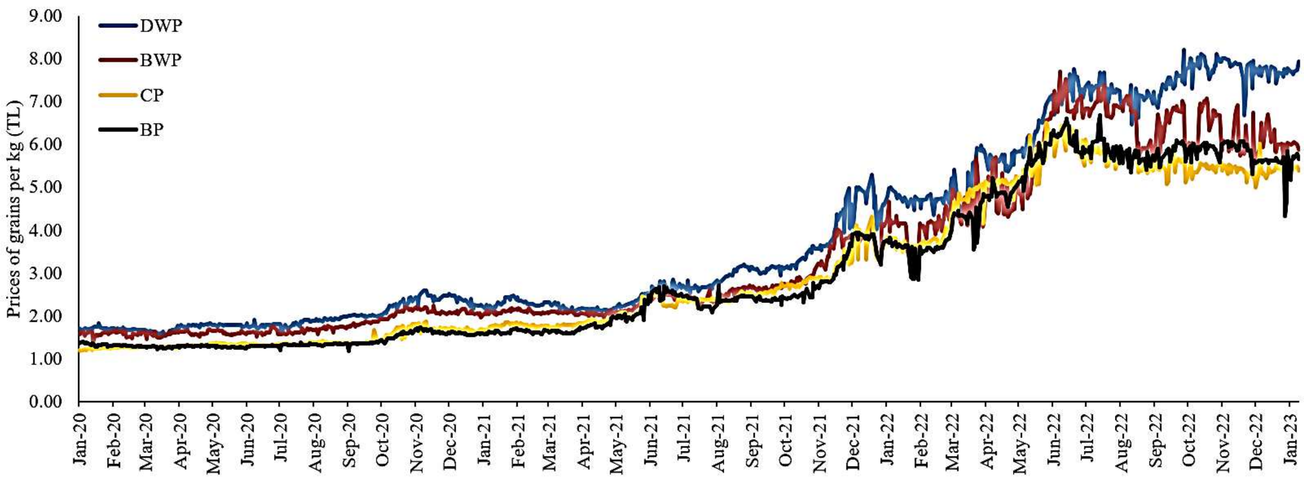


Figure 2: Durum wheat, bread wheat, corn, and barley prices based on the date considered

Table 3 shows the correlation matrix containing the dependent and independent variables we will use to estimate agricultural product prices. The correlation coefficient takes values between -1 and +1 and shows the strength of the linear relationship between two variables. A negative coefficient indicates an inverse relationship, while a positive coefficient indicates a positive relationship. As the absolute value of the obtained correlation coefficient approaches zero, the strength of the relationship decreases; as it approaches 1, the power of the relationship increases.

Table 3. Correlation matrix of variables considered

	OP	USD	DWV	DWP	BWV	BWP	CV	CP	BV	BP
OP	1.000									
USD	0.781	1.000								
DWV	0.579	0.696	1.000							
DWP	0.819	0.985	0.705	1.000						
BWV	0.592	0.664	0.653	0.659	1.000					
BWP	0.827	0.982	0.698	0.991	0.650	1.000				
CV	0.499	0.678	0.457	0.661	0.485	0.652	1.000			
CP	0.864	0.974	0.707	0.985	0.673	0.985	0.661	1.000		
BV	0.466	0.579	0.570	0.562	0.693	0.571	0.497	0.578	1.000	
BP	0.855	0.977	0.697	0.987	0.667	0.987	0.659	0.993	0.555	1.000

Abbreviations = Oil price; OP, Exchange; USD, Durum wheat volume; DWV, Durum wheat price; DWP, Bread wheat volume; BWV, Bread wheat price; BWP, Corn volume; CV, Corn price; CP, Barley volume; BV, Barley price; BP

We have observed that there is a strong relationship between the variables in terms of the correlation coefficient values of the dependent and independent variables. The correlation coefficient showing the strong relationship between BP and CV was calculated as 0.993. Barley price and corn price generally show a high correlation with each other depending on agricultural factors such as climatic conditions supply and demand factors, as both are subject to similar growing conditions and markets. Similar features are also present in variables with high correlation values. However, the correlation value between DWV and CV was calculated as 0.457, leading to the weakest relationship between the variables in the data set. This value shows a moderate relationship according to the correlation coefficient evaluation situations (Taylor, 1990). In general, we have observed moderate and high correlations between the variables in the ML model in which we estimated the selected grain prices. There is a strong positive relationship between agricultural product prices and oil prices. A similar situation can be observed between product prices and exchange rates. In addition, there is a moderate relationship between product prices and transaction volumes in the market.

Machine Learning (ML) Algorithms for Forecasting: Machine learning (ML) algorithms have recently been widely used to classify or predict complex relationships in extensive complex data. This study also applied this method to estimate the dependent variables using data sets belonging to the different independent variables. Four different ML algorithms were used in this study, which aims to predict the prices of four essential grain products in Türkiye. Durum wheat, bread wheat, corn, and barley prices were calculated using the SVM, RF, NN, and LR algorithms through the Orange 3.14 software preferred for this study. Orange is a free and open-source software that uses the infrastructure of Python and Anaconda software programs in data mining. The schema of the developed ML models is shown in Figure 3.

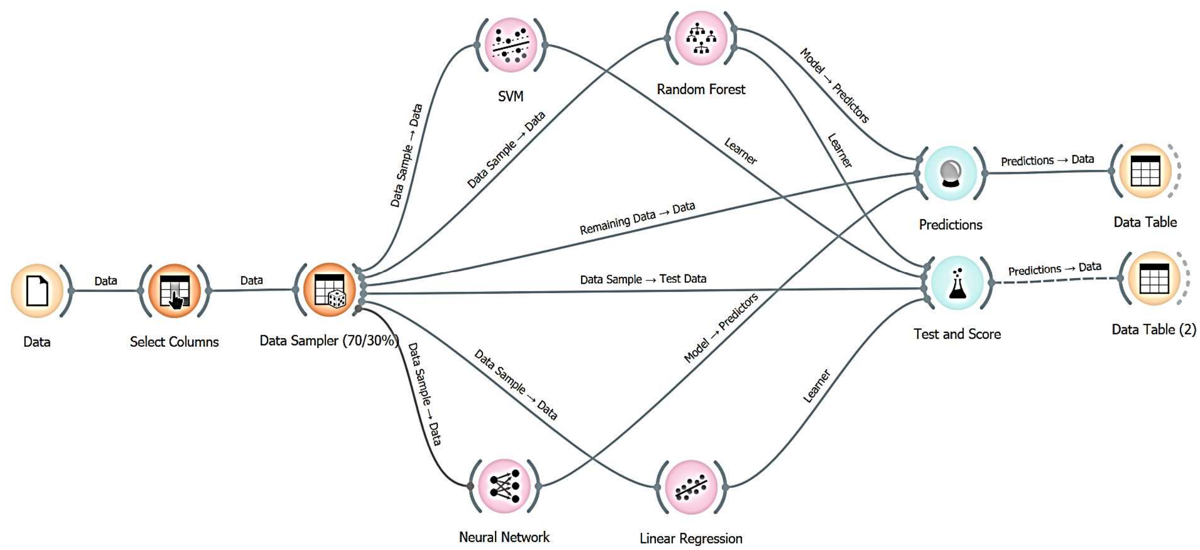


Figure 3. The algorithm of ML models

Support Vector Machine (SVM): Support Vector Machine (SVM) is a data-based ML technique that is primarily used in binary classification. The basic goal of the SVM function is to completely separate multidimensional data into two distinct classes with the help of a hyperplane (Al-Anazi and Gates, 2010). Furthermore, SVM is not only used in classification problems but also in solving regression problems. The primary goal of SVM is to achieve the correct solution in an unknown data set that minimizes the classification error (Atalan, 2022).

The basic goal of this supervised ML algorithm is to divide the data into two groups, positive and negative, by maximizing the distance or margin as explained in equations (1) and (2) (DeLong *et al.,* 2010).

$$X_i W + b \geq +1, \forall y_i = +1 \quad (1)$$

$$X_i W + b \leq -1, \forall y_i = -1 \quad (2)$$

The training data is represented by x and y, the weight of vectors is symbolized by w and the bias value is denoted by b in this equation. The SVM approach is to identify a hyperplane with the most significant possible distance between the margins which is represented by the equation $w^T x = 0$. The objective function for the SVM model is defined as below:

$$min_{w,b} = \frac{1}{2} ||w||^2 \quad (3)$$

The features of SVM algorithm developed for this research is discussed in Table 4.

Table 4. The features of the SVM algorithm.

The type of ML	SVM
Cost (C)	0.10
Regression loss epsilon (ε)	0.10
Kernel: RBFa, g	auto
Numerical tolerance	0.0010
Iteration limit	100

The cost parameter (C) controls the trade-off between achieving a low training error and a low testing error. The regression loss epsilon (ε) is the tolerance level for the distance between the predicted and actual values. If the difference is larger than epsilon, it is counted as an error. The kernel function, in this case, is the RBF (Radial Basis Function) kernel, which is commonly used in SVM because of its ability to model non-linear decision boundaries. The gamma parameter is used to adjust the shape of the decision boundary. The numerical tolerance specifies the precision level for the algorithm to terminate. The iteration limit specifies the maximum number of iterations allowed for the algorithm to converge. If the algorithm has not converged within this limit, it stops and returns the current solution.

Random Forest (RF): Random Forest (RF) algorithm is a classification and regression algorithm that is commonly used in data with a large number of variables. The RF regression model is developed as an extension of the classification regression tree (CART). It generally provides better results than CART model. (Breiman *et al.,* 2017). The RF method produces hundreds of decision trees independently on its own. Then, the final result is obtained by taking the average of the prediction values of the obtained decision trees. (Li *et al.,* 2018).

If x, the input vector is considered for a specific training, k number of regression trees are obtained. The average of the k obtained regression trees is considered. The RF regression predictor is given by the equation. (Rodriguez-Galiano *et al.,* 2015).

$$f_{rf}^{K}(x) = \frac{1}{K}\sum_{k=1}^{K} T(x) \quad (4)$$

The features of RF algorithm developed based on ML method for this research is discussed in Table 5.

The number of trees option determines the quantity of decision trees that will be generated and combined in the RF model. The "number of trees considered at each split" option determines how many attributes will be randomly selected and considered at each node when building the decision trees in the RF model. The "limit depth of individual trees" option allows the user to set a maximum depth to which each individual decision tree in the RF model can grow. This is often used as a pre-pruning technique to improve the model's speed and avoid overfitting the training data.

Table 5. The features of RF algorithm.

The type of ML	RF
Number of trees	10
Number of attributes considered at each split	5
Limit depth of individual trees	3
Maximal tree depth	Unlimited

Neural Network (NN): The structure and function of the human brain served as the inspiration for the ML technique known as neural networks (NN) (Zupan and Gasteiger, 1991). These networks are composed of layers of interconnected "neurons" that process and transmit information. NNs are used for a wide range of objectives, such as robotics and control systems, Weather forecasting and climate modeling, gaming and decision-making, and prediction (Agatonovic-Kustrin and Beresford, 2000). They are particularly useful for tasks involving large amounts of complex or unstructured data. They are also widely used for non-linear regression or classification tasks. A NN algorithm typically consists of an input layer, one or more hidden layers, and an output layer (Oguz *et al.,* 2008) (Tu, 1996). The input layer receives the input data, and each subsequent layer processes the information and passes it on to the next layer. The output layer produces the final output or prediction (Agatonovic-Kustrin and Beresford, 2000). The structure of NNs can be visualized in **Figure 4**, which illustrates the general layout of NNs. The parameters of the NN model used in this research in **Figure 4**.

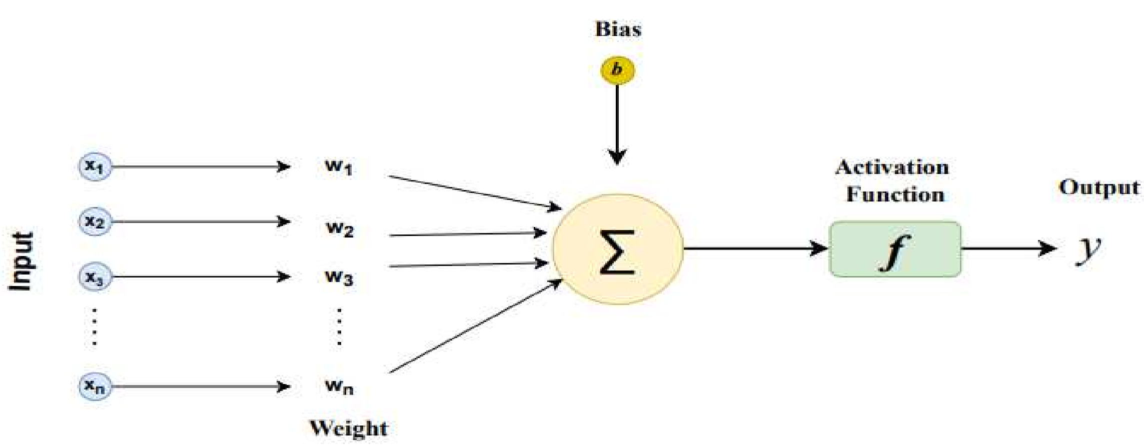


Figure 4. Neural network structure

The features of NN algorithm developed based on ML method for this research is discussed in Table 6.

Table 6. The features of NN algorithm.

The type of ML	NN
Neurons in hidden layers	100
Activation	Relub
Solver	Adam
Regularization	α = 0.0001
Maximal number of iterations	200
Replicable training	yes

The neurons per hidden layer option specifies the number of neurons that will be used in each of the hidden layers of a NN. The value for each element in the list corresponds to the number of neurons in the corresponding hidden layer. The "activation function for hidden layers" option refers to the mathematical function that is applied to the output of each neuron in the hidden layers of a NN. The "solver for weight optimization" option specifies the algorithm used to optimize the weights in a NN during training. "Maximal number of iterations" option has been set to a value of 200, which means that the NN will undergo a maximum of 200 iterations during the training process.

Linear Regression (LR): Linear regression (LR) model is one of the most widely used methods in academic research. Its simplicity and ease of understanding is the biggest reason for this. The LR algorithm allows us to obtain the estimated values of the dependent variable by showing the linear relationship between the dependent variable and independent variables. The basic notation of LR model is revealed in Eq [5].

$$y = \alpha_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots \beta_k x_k + \varepsilon \quad (5)$$

In the above simple LR model, y represents the dependent variable, α_0 represents the constant parameter in the regression model, and β represents the coefficients in the model. The number of parameters in the model is k, and ε represents the model's standard error.

The LR algorithm often overfits the data due to relatively few tracking stages and many potential predictive variables proposed to train a model. Generally, this algorithm ignores the interaction between the variables and provides a linear relationship directly to the prediction data (Chen *et al.,* 2019). For these reasons, we chose to use more than one ML algorithm for this algorithm to avoid these concerns. The feature data of the LR algorithm used in this study are shared in Table 7.

Comparison Criteria of ML Algorithms: In ML algorithms, the performance of models is evaluated by considering various criteria for evaluating model performance. In this study, where we estimated four different grain prices on a daily basis, we used the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Square Error (MSE) and Coefficient of Determination (R^2) to compare the performance of different ML models. The mathematical formulas of the criteria used to compare the ML models in this study are presented below:

Table 7. The features of LR algorithm.

The type of ML	**LR**
Parameters	Fit intercept (unchecking fixes it to zero)
No regulations	

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \tilde{y}_i|^2 \quad (6)$$

$$MSE = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 \quad (7)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2}{n}} \quad (8)$$

$$R^2 = \sum_{i=1}^{n} \left[\frac{y_i - \tilde{y}_i}{y_i - \bar{y}_i}\right]^2 \quad (9)$$

In the equations above, the number of observations is indicated by n, while the predicted values are represented by ỹ, and the actual values are represented by y_i. The R^2 (R-squared) value ranges between 1(or %100) to 0 (or %0.00), and a value closer to 1 indicates a better model. MSE, MSE, and MAE values must reach zero to ensure the perfect fit of ML algorithms (Chen *et al.,* 2019)

RESULTS

In this study, we have estimated the daily prices of four agricultural products in Türkiye using four ML models. Each model created the general features of the ML algorithms, and prediction data were obtained. Feature selections of ML algorithms are discussed in detail in the method part of the study. For all ML models, 70% of the data set was randomly selected, and the model was trained. The models obtained from the training data were tested using 30% of the data set. Then, predictions were made for all models on the remaining data. The validity of the prediction results of the ML algorithms was verified by comparing the performance criteria of the models for all models. The numerical results of ML models are discussed in the following subsections.

Results of Performance Criteria: The RF algorithm performs best according to the measurement metrics for ML algorithms' durum wheat price train phase. This algorithm's MSE, RMSE, MAE, and R^2 values were calculated as 0.001, 0.024, 0.015, and 0.998, respectively.

Likewise, the RF algorithm gave better results in the durum wheat prices testing phase than the other three ML algorithms. MSE, RMSE, MAE, and R^2 values of the RF algorithm performance results of the durum wheat price test phase were calculated as 0.005, 0.071, 0.042, and 0.983, respectively. The performance criteria result of ML models are shared in Table 8. Within the scope of this study, the RF algorithm was found to be the best algorithm in the training stage for the second agricultural product, bread wheat, which we made with the price estimation ML algorithm. MSE, RMSE, MAE, and R^2 values were calculated as 0.001, 0.029, 0.018, and 0.997, respectively.

Table 8. Numerical results of Performance Criteria for ML Algorithms

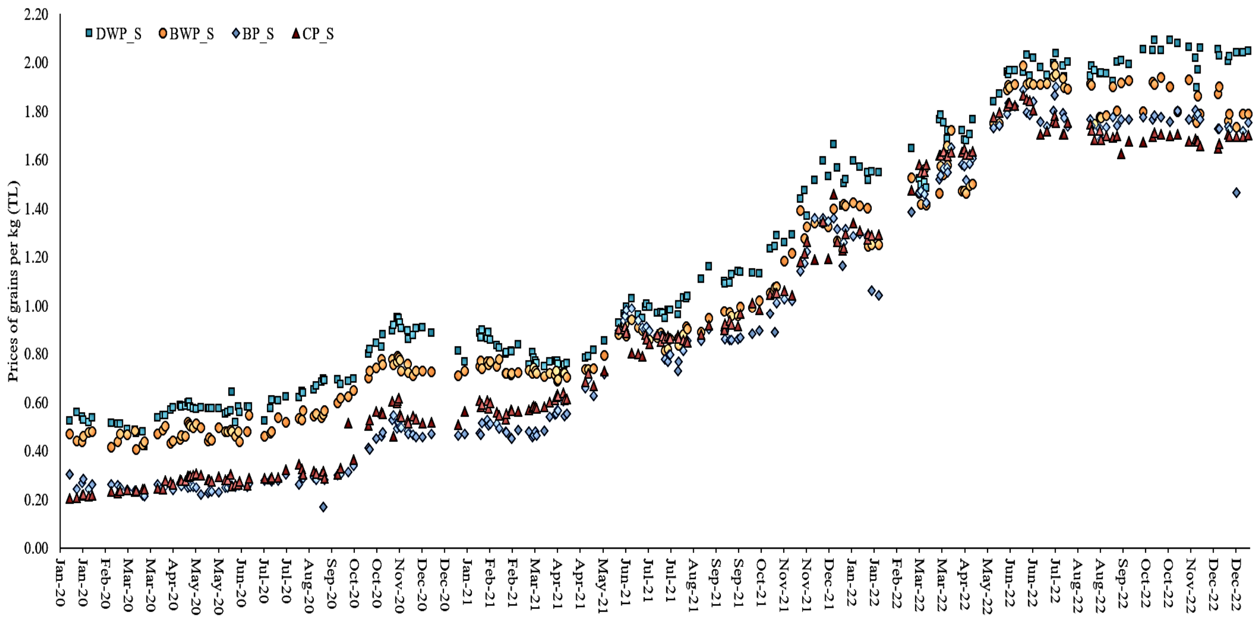
Variables	Model	Test Data				Train Data			
		MSE	RMSE	MAE	R^2	MSE	RMSE	MAE	R^2
DWP	SVM	0.008	0.087	0.067	0.974	0.005	0.069	0.056	0.985
	RF	0.005	0.071	0.042	0.983	0.001	0.024	0.015	0.998
	NN	0.011	0.105	0.077	0.962	0.010	0.098	0.076	0.969
	LR	0.008	0.087	0.065	0.974	0.006	0.078	0.060	0.980
BWP	SVM	0.006	0.080	0.065	0.976	0.006	0.077	0.064	0.979
	RF	0.003	0.059	0.035	0.987	0.001	0.029	0.018	0.997
	NN	0.008	0.089	0.068	0.970	0.008	0.087	0.068	0.973
	LR	0.007	0.087	0.069	0.972	0.007	0.082	0.066	0.976
BP	SVM	0.008	0.091	0.066	0.975	0.005	0.068	0.053	0.987
	RF	0.004	0.064	0.039	0.988	0.001	0.026	0.015	0.998
	NN	0.010	0.100	0.069	0.970	0.007	0.087	0.063	0.978
	LR	0.010	0.098	0.075	0.971	0.007	0.085	0.065	0.979
CP	SVM	0.006	0.075	0.060	0.982	0.005	0.068	0.056	0.986
	RF	0.004	0.064	0.034	0.987	0.001	0.025	0.015	0.998
	NN	0.006	0.078	0.059	0.980	0.007	0.081	0.061	0.980
	LR	0.008	0.090	0.074	0.974	0.007	0.084	0.070	0.978

Abbreviations = Durum wheat price; DWP, Bread wheat price; BWP, Barley price; BP, Corn price; CP

The RF model gave the best results compared to the other three ML algorithms in the testing phase. MSE, RMSE, MAE, and R^2 values were calculated as 0.003, 0.059, 0.035, and 0.987, respectively. The RF model was found to be the best model in both the training and testing phases among the barley and corn prices we are estimating. In this study, in which we estimated the daily prices of four agricultural products with ML algorithms, the RF model, one of the ML algorithms, gave better results in both the training and test stages in the price estimation of four-grain products.

Comparison of forecasting data based on the ML algorithms: Price estimation and forecasting of agricultural products are essential for the economy, as the agricultural sector plays a crucial role in the economies of many countries. Especially in developing countries, the agricultural industry generally constitutes a large part of the country's national income, and exporting agricultural products is an essential source of economic growth. Price estimations of agricultural products can also be used to plan economic development in these countries. The prices of agricultural products depend on many variables, such as production costs, consumer demand, climatic conditions, and other factors. Therefore, the price prediction of agricultural products helps economists understand the market's balance of supply and demand and is vital in developing economic policies. Price forecasts of agricultural products are also closely related to consumer prices and food security. If the prices of agricultural products are higher than expected, consumers may struggle to access these products, and food security issues may arise. The most recent crisis in grain products has emerged due to the Russian-Ukrainian war. Conversely, if prices are lower than expected, producers may face difficulties, and the value of their products may decrease. The fluctuations in the price of agricultural products directly affect the income of the producers in the sector, which is crucial for the survival of the producers and their businesses. In addition, the price forecasting of agricultural products helps investors and speculators follow the market, which can mitigate fluctuations and price changes in the agricultural products market. Price stability in the agricultural products mark*et al*lows producers and consumers to plan and make informed decisions, thereby supporting economic stability.

Ensuring price stability in agricultural products is an important part of economic development. For this purpose, we tried to determine which algorithms give more accurate results in the price estimation of agricultural products. In this study, we attempted to determine the best algorithm by estimating the prices of four agricultural products obtained daily with ML algorithms. Detailed results of estimation ML algorithms-based performance criteria are shown in Table 9.



*DWP_S, selected data for DWP; BWP_S, selected data for BWP; BP_S, selected data for BP; CP_S, selected data for CP

Figure 5. Selected data for testing phase of ML algorithms

Figure 5. Shows the test phase results of four agricultural products predicted by the ML algorithm. The algorithm randomly determined the observations used in the test phase. The data used in the testing phase of ML estimation methods should be chosen randomly. If the variables in the test phase of the model are not determined randomly and evaluated, it gives biased results. Generally, the models performed in the testing phase have a similar trend.

Table 9. Estimation Data of ML Algorithms-based Performance Criteria

Variables	Model	Prediction			
		MSE	RMSE	MAE	R^2
Durum wheat	SVM	0.008	0.087	0.067	0.974
	RF	0.004	0.055	0.039	**0.985**
	NN	0.001	0.105	0.077	0.962
	LR	0.008	0.087	0.065	0.974
Bread wheat	SVM	0.006	0.080	0.065	0.976
	RF	0.003	0.058	0.035	**0.987**
	NN	0.008	0.089	0.068	0.970
	LR	0.007	0.087	0.069	0.972
Barley	SVM	0.008	0.091	0.066	0.975
	RF	0.004	0.063	0.038	**0.988**
	NN	0.010	0.100	0.069	0.970
	LR	0.010	0.098	0.075	0.971
Corn	SVM	0.006	0.075	0.060	0.982
	RF	0.004	0.063	0.034	**0.987**
	NN	0.006	0.078	0.059	0.980
	LR	0.008	0.090	0.074	0.974

The algorithms that best predicted durum wheat prices were determined as RF, (R^2; 0.985) SVM, (R^2; 0.974) LR, (R^2; 0.974) NN, (R^2; 0.962), respectively. The best algorithm for estimating the bread wheat prices was the RF algorithm, as in the case of wheat. MSE, RMSE, MAE, and R^2 values were calculated as 0.003, 0.059, 0.035, and 0.987, respectively. According to the results, the SVM model is the second-best in bread wheat price prediction. Similar results were obtained from the corn price and barley price estimates. The best model for barley and corn prices was determined as the RF model. R^2 values of barley and corn in the RF model were calculated at 0.988 and 0.987, respectively.

Our study has certain limitations due to the unavailability of some data on a daily basis. Since the research relies on daily data, variables that could impact agricultural product prices, such as daily demand or population fluctuations, could not be included in the analysis as they were not available on a daily basis. Similarly, the research did not include other macroeconomic indicators such as inflation, economic growth, and current account deficit due to their unavailability in daily data. A significant limitation of our study is the exclusion of the impact of macroeconomic variables on grain product prices. Another limitation is the inflation problem that Türkiye has struggled with since 2020. The fact that food inflation is very high among the inflation experienced in Türkiye has led us to the last periods in choosing the time. Another study limitation is that unexpected shocks were not included in the analysis. For example, distinguishing the impact of the grain crisis due to the Russia-Ukraine war on prices is a significant problem. The last limitation is that since some factors, such as gold, silver, and aluminum prices, have lower frequencies, such factors were not included in the analysis in this study.

Another example is the need to measure climate developments' impact on grain product prices. Since this study's preferred pilot region has a large area, data sets of climatic and environmental factors were not considered. Speculation by personal or local authorities on grain prices causes fluctuations in grain unit prices, albeit for a short time. Data from such speculation periods were not used in this study. Considering all these reasons, the method we propose in this study plays a vital role in avoiding uncertainties regarding the future of grain unit prices (Turkish Lira per kg).

DISCUSSION

Maintaining normal living conditions is crucial for the sustainability of the agricultural economy. However, various factors, such as global epidemics, wars, and government policies, can profoundly impact both agriculture and the industrial economy. For instance, the COVID-19 pandemic, which emerged in late 2019, and the Ukraine-Russia war, starting in 2021, severely disrupted access to grain products, particularly in Europe, Africa, and parts of Asia. Russia, being a major player in the agricultural economy, significantly influences grain product prices (Svanidze and Götz, 2019).

The Covid-19 pandemic caused a global economic slowdown, impacting agriculture, industry, and services. Although lifting pandemic restrictions increased global demand, it also led to inflationary pressures worldwide, affecting agricultural product prices. In Türkiye, food inflation has been high for the past three years, significantly impacting the rural economy where agriculture is the primary income source. Ensuring the price stability of agricultural products is crucial for providing affordable and nutritious food for all income groups in Türkiye. Accurate price estimation can help prevent poverty and ensure food availability.

Forecasting and estimating agricultural prices are vital for economic development, as the agricultural sector is a primary industry and a major livelihood source for many countries. Changes in agricultural prices directly impact producers' income and economic growth. High agricultural prices can generate more producer income and company profits, enhancing the country's foreign trade and economic growth. However, high prices can also negatively affect consumer spending, leading to reduced consumption and economic contraction. Thus, accurate agricultural price forecasts are essential for financial planning.

Machine Learning (ML) algorithms play a critical role in predicting agricultural product prices. By identifying patterns in supply and demand, these algorithms can foresee potential price shocks, allowing policymakers to take preventive measures and stabilize prices. This contributes to achieving price stability, income stability for farmers, and poverty prevention, which are essential for Türkiye's agricultural economy and food security.

This study calculated the values of performance measurements of different ML algorithms for four different grain types. Performance metrics are used to evaluate how close predictions are to actual data. Each algorithm has four other measurements for each grain type: RMSE, MSE, MAE, and R^2. The RF algorithm generally produced predictions with the lowest MSE, RMSE, and MAE values. For this reason, it shows that the estimation of this algorithm are closer to the actual data. Additionally, the R^2 value of the RF model is higher than most other algorithms, indicating that the predictions better explain the data. The ML algorithms run in this study show that RF provides a better prediction performance for these grain types.

On the other hand, while SVM and NN have higher MSE, RMSE, and MAE values in some cases, they do not have very high R^2 values. It shows that the estimations of SVM and NN algorithms are farther from the actual data, and at the same time, these algorithms explain the data less well. These results show that RF and LR perform better for some grain types, but there are different results for each grain type. The performance of different ML algorithms may vary depending on the grain species' characteristics and the data set's structure. When the findings obtained in the study were compared with the literature, similar results were found (Garai *et al*. 2023; Kuradusenge *et al* 2023). Garai *et al*. (2023) stated that RF was the most robust prediction model, while another study by Kuradusenge *et al*. (2023) found that RF produced strong results in predicting potato and corn yields.

In this study, we tried to directly estimate the unit prices of grain products belonging to ML algorithms based on dependent and independent variables. In addition, this study provided numerical data and prediction data to be obtained. In one study, the effects of political practices were considered by considering the supply-demand balance in grain products (Wright, 2011). Some studies have tried to obtain estimation data on grain productivity instead of grain unit price. One study used digitally estimated canopy cover and plant density data, as well as crop indices, to estimate productivity in maize production using an NN algorithm (García-Martínez *et al.,* 2020). Another study estimated the yield response of corn in sub-Saharan Africa by analyzing environmental factors in corn production and using an empirical ML model and a mechanical model based on 12,081 trial observations (Bonilla-Cedrez *et al.,* 2021). Tian *et al.,* using the Rural Fixed Point Observation data set between 2004-2010, studied wheat, corn, and rice products from cereal crops (Tian *et al.,* 2020). Compared to other studies, the estimation data obtained in this study directly provides information about the unit prices of grain products.

In conclusion, accurate estimation of agricultural product prices is vital for Türkiye's agricultural sector. Given the importance of ML algorithms in predicting prices, policymakers and stakeholders should prioritize their development and implementation to ensure the effective functioning of Türkiye's agricultural economy.

Conclusion: Wheat, corn, and barley are essential grains for the agricultural economy, subject to daily price fluctuations influenced by various factors. This study uses machine learning (ML) algorithms, specifically RF, NN, LR, and SVM, to predict grain unit prices in Türkiye. Among these models, the Random Forest (RF) algorithm performed best, accurately estimating prices with the lowest error margins. SVM also performed well, especially in handling seasonal fluctuations. NN and LR models were less accurate.

This study provides a practical framework for predicting and estimating grain prices using ML algorithms that are adaptable to different regions. Investing in agricultural technologies and implementing early warning systems with effective algorithms like RF and SVM can significantly enhance data collection, decision-making, and overall agricultural efficiency.

The key takeaway is that integrating ML algorithms into agriculture can greatly improve efficiency and profitability.

Conflict of Interest: The authors declare that no conflicts of interest exist for this research

Funding Statement: The author received no funding from an external source.

Ethics statement: Not applicable

Author's Contributions Statement: The study was conceptualized by AK and İE, while AA and AK contributed to the methodology. The formal analysis involved AA, AK, and İE; AA and AK were responsible for the investigation and data curation. The original draft preparation was the joint effort of AA and İE. Writing and editing were conducted collaboratively by AA, AK, and İE. All authors have thoroughly reviewed and agreed upon the final version of the manuscript before its publication.

REFERENCES

Abaidoo, R. and E.K. Agyapong (2022). Commodity prices, inflation and inflation uncertainty among emerging economies. J. Economic and Administrative Sciences, doi: 10.1108/JEAS-10-2021-0203.

Adetunji, A.B., O.N. Akande, F.A. Ajala, O. Oyewo, Y.F. Akande and G. Oluwadara (2022). House price prediction using random forest machine learning technique. Procedia Comput Science 199:806–813. doi: 10.1016/j.procs.2022.01.100

Adewopo, J.B., G. Solano-Hermosilla, L. Colen and F. Micale (2021). Using crowd-sourced data for real-time monitoring of food prices during the COVİD-19 pandemic: Insights from a pilot project in northern Nigeria. Glob Food Security 29:100523. doi: 1016/j.gfs.2021.100523

Agatonovic-Kustrin, S., & Beresford, R. (2000). Basic concepts of Artificial Neural Network (ANN) modeling and its application in pharmaceutical research. J. Pharmaceutical and Biomedical Analysis, 22(5), 717–727. doi:10.1016/s0731-7085(99)00272-1

Al-Anazi, A. and I.D. Gates (2010). A support vector machine algorithm to classify lithofacies and model permeability in heterogeneous reservoirs. Engineering Geology 114:267–277. doi: 10.1016/j.enggeo.2010.05.005

Armantier, O., G. Koşar, R. Pomerantz, D. Skandalis, K. Smith, G. Topa and W. van der Klaauw (2021). How economic crises affect inflation beliefs: Evidence from the Covid-19 pandemic. J. Economic Behavior & Organization 189:443–469. doi: 10.1016/j.jebo.2021.04.036

Atalan, A. (2022). Forecasting drinking milk price based on economic, social, and environmental factors using machine learning algorithms. Agribusiness, 39(1), 214–241. doi:10.1002/agr.21773

Bermpei, T., Karadimitropoulou, A., Triantafyllou, A., & Alshalahi, J. (2023). Does commodity price uncertainty matter for the cost of credit? evidence from developing and Advanced Economies. J. Commodity Markets, 29, 100306. doi:10.1016/j.jcomm.2022.100306

Bonam, D., & Smădu, A. (2021). The long-run effects of pandemics on inflation: Will this time be different? Economics Letters, 208, 110065. doi:10.1016/j.econlet.2021.110065

Bonilla-Cedrez, C., Chamberlin, J., & Hijmans, R. J. (2021). Fertilizer and grain prices constrain food production in sub-saharan africa. Nature Food, 2(10), 766–772. doi:10.1038/s43016-021-00370-1

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). Regression trees. Classification And Regression Trees, 216–265. doi:10.1201/9781315139470-8

Çepni, O., Gupta, R., Pienaar, D., & Pierdzioch, C. (2022). Forecasting the realized variance of oil-price returns using machine learning: Is there a role for U.S. state-level uncertainty? Energy Economics, 114, 106229. doi:10.1016/j.eneco.2022.106229

Chen, J., de Hoogh, K., Gulliver, J., Hoffmann, B., Hertel, O., Ketzel, M., … Hoek, G. (2019). A comparison of linear regression, regularization, and machine learning algorithms to develop Europe-wide spatial models of fine particles and Nitrogen Dioxide. Environment International, 130, 104934. doi:10.1016/j.envint.2019.104934

DeLong, J. P., Burger, O., & Hamilton, M. J. (2010). Current demographics suggest future energy supplies will be inadequate to slow human population growth. PLoS ONE, 5(10). doi:10.1371/journal.pone.0013206

Elder, J. (2018). Oil price volatility: Industrial production and special aggregates. Macroeconomic Dynamics, 22(3), 640–653. doi:10.1017/s136510051600047x

Elder, J. and A. Serletis (2010). Oil price uncertainty. J Money Credit Bank 42:1137–1159. 42(6), 1137–1159. doi:10.1111/j.1538-4616.2010.00323.x

Garai, S., Paul, R. K., Rakshit, D., Yeasin, M., Emam, W., Tashkandy, Y., & Chesneau, C. (2023). Wavelets in combination with stochastic and machine learning models to predict agricultural prices. Mathematics, 11(13), 2896. doi:10.3390/math11132896

García-Martínez, H., Flores-Magdaleno, H., Ascencio-Hernández, R., Khalil-Gardezi, A., Tijerina-Chávez, L., Mancilla-Villa, O. R., & Vázquez-Peña, M. A. (2020). Corn grain yield estimation from vegetation indices, canopy cover, plant density, and a neural network using multispectral and RGB images acquired with unmanned aerial vehicles. Agriculture, 10(7), 277. doi:10.3390/agriculture10070277

George, D. and P. Mallery (2021). IBM SPSS Statistics 27 Step by Step. Routledge, New York.

Jha, G. K. and K.Sinha (2013). Agricultural price forecasting using neural network model: An innovative information delivery system. Agricultural Economics Research Review, 26(2), 229-240.

JO, S. (2014). The effects of oil price uncertainty on global real economic activity. J. Money, Credit and Banking, 46(6), 1113–1135. doi:10.1111/jmcb.12135

Kuradusenge, M., Hitimana, E., Hanyurwimfura, D., Rukundo, P., Mtonga, K., Mukasine, A., … Uwamahoro, A. (2023). Crop yield prediction using machine learning models: Case of Irish potato and maize. Agriculture, 13(1), 225. doi:10.3390/agriculture13010225

Li, Y., Zou, C., Berecibar, M., Nanini-Maury, E., Chan, J. C.-W., van den Bossche, P., … Omar, N. (2018). Random Forest regression for online capacity estimation of lithium-ion batteries. Applied Energy, 232, 197–210. doi:10.1016/j.apenergy.2018.09.182

Nicola, M., Z. Alsafi, C. Sohrabi, A. Kerwan, A. Al-Jabir, C. Iosifidis, M. Agha and R. Agha (2020). The socio-economic implications of the coronavirus and covid-19 pandemic: a review. International J. Surgery 78:185-193 doi: https://doi.org/10.1016/j.ijsu.2020.04.018.

Oguz, E., Tortum, A., & Keskinler, B. (2008). Determination of the apparent rate constants of the degradation of humic substances by ozonation and modeling of the removal of humic substances from the aqueous solutions with Neural Network. J. Hazardous Materials, 157(2–3), 455–463. doi:10.1016/j.jhazmat.2008.01.018

Özdemir, M.O. and C. Çilgin (2022). A comparison of the performance of machine learning and time series forecasting models in predicting wheat price. Understanding Economics in the 21st Century. Gazi Bookstore, Ankara. pp.203–218.

Ravazzolo, F., & Rothman, P. (2013). Oil and U.S. GDP: A real-time out-of-sample examination. J. Money, Credit and Banking, 45(2–3), 449–463. doi:10.1111/jmcb.12009

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geology Reviews, 71, 804–818. https://doi.org/10.1016/j.oregeorev.2015.01.001

Sadiq, M., Lin, C.-Y., Wang, K.-T., Trung, L. M., Duong, K. D., & Ngo, T. Q. (2022). Commodity dynamism in the COVID-19 crisis: Are gold, oil, and stock commodity prices, symmetrical? Resources Policy, 79, 103033. doi:10.1016/j.resourpol.2022.103033

Sadorsky, P. (2022). Using machine learning to predict clean energy stock prices: How important are market volatility and economic policy uncertainty? J. Climate Finance, 1, 100002. doi:10.1016/j.jclimf.2022.100002

Sasmal, J. (2015). Food price inflation in India: The growing economy with sluggish agriculture. J. Economics, Finance and Administrative Science, 20(38), 30–40. doi:10.1016/j.jefas.2015.01.005

Şengül, C. M.Gerşil (2018). Estimating of Manisa Cotton Prices Using Time Series and Artificial Neural Network Techniques and Comparison of Their Estimating Performance. J. Management and Economics 25:1017–1031. doi:10.18657/yonveek.457761

Sokhanvar, A., & Bouri, E. (2023). Commodity price shocks related to the war in Ukraine and exchange rates of commodity exporters and importers. Borsa Istanbul Review, 23(1), 44–54. doi:10.1016/j.bir.2022.09.001

Svanidze, M., & Götz, L. (2019). Determinants of spatial market efficiency of grain markets in Russia. Food Policy, 89, 101769. doi:10.1016/j.foodpol.2019.101769

Swinnen, J., & McDermott, J. (2020). Covid-19 and Global Food Security. EuroChoices, 19(3), 26–33. doi:10.1111/1746-692x.12288

Taylor, R. (1990). Interpretation of the correlation coefficient: A basic review. J. Diagnostic Medical Sonography, 6(1), 35–39. doi:10.1177/875647939000600106

Tian, X., Yi, F., & Yu, X. (2019). Rising cost of labor and transformations in grain production in China. China Agricultural Economic Review, 12(1), 158–172. doi:10.1108/caer-04-2018-0067

Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. J. Clinical Epidemiology, 49(11), 1225–1231. doi:10.1016/s0895-4356(96)00002-9

Wright, B. D. (2011). The economics of grain price volatility. Applied Economic Perspectives and Policy, 33(1), 32–58. doi:10.1093/aepp/ppq033

Jiajun Zong and Quanyin Zhu (2012). Price forecasting for agricultural products based on BP and RBF Neural Network. 2012 IEEE International Conference on Computer Science and Automation Engineering. IEEE. pp.607–610.

Zou, H. F., Xia, G. P., Yang, F. T., & Wang, H. Y. (2007). An investigation and comparison of artificial neural network and time series models for Chinese Food Grain Price forecasting. Neurocomputing, 70(16–18), 2913–2923. doi:10.1016/j.neucom.2007.01.009

Zupan, J., & Gasteiger, J. (1991). Neural networks: A new method for solving chemical problems or just a passing phase? Analytica Chimica Acta, 248(1), 1–30. doi:10.1016/s0003-2670(00)80865-x.