Predicting Agricultural Product and Supplies Prices Using Artificial Intelligence

Ioannis Dionissopoulos[1][a], Fotis Assimakopoulos[2][b], Dimitris Spiliotopoulos[2][c], Dionisis Margaris[3][d] and Costas Vassilakis[1][e]

[1]*Department of Informatics and Telecommunications, University of the Peloponnese, Tripoli, Greece*
[2]*Department of Management Science and Technology, University of the Peloponnese, Tripoli, Greece*
[3]*Department of Digital Systems, University of the Peloponnese, Sparta, Greece*
{tst06057, fasimakopoulos}@go.uop.gr, {dspiliot, margaris, costas}@uop.gr

Keywords: Agricultural Products, Price Forecasting, Machine Learning, Deep Learning, Data Integration, Forecasting Models.

Abstract: This work focuses on the prediction of agricultural product and supply prices using historical data and artificial intelligence methods. Agricultural product and supply prices are important for the economy and growth of agriculture. Using modern data analysis and deep learning methods, a forecasting model was developed to help us predict future price trends. The data used include the sales prices of crop products and the purchase prices of agricultural inputs. The developed forecasting methods exhibit high accuracy for predicting the actual prices of products and supplies, with error margins ranging from 0.29% to 9.8%, while they can also predict price rises and falls, with respective success rates ranging from 73.29% to 84.96%.

1 INTRODUCTION

In recent years there has been an explosion in data collection. Developments in internet technology have led more and more organisations, both private and public, to organise the collection and dissemination of their data. Some of this data is posted on open-data portals for public use.

Machine learning (ML) frameworks offer a clear knowledge of the process by analysing the massive amounts of data and interpreting the information extracted. These technologies are employed in the construction of models that delineate the connections between elements and actions. Furthermore, ML models can be utilised to predict future actions in a specific scenario (Rashid et al., 2021).

Precision farming uses algorithmic approaches and data to improve productivity, by predicting weather conditions, soil analysis, crop recommendations, and fertilizer and pesticide usage. It uses advanced technologies like IoT, Data Mining, and Machine Learning (ML) to collect data and train the respective systems. This approach reduces manual labour and increases productivity. Farmers face challenges like crop failure and soil infertility (Durai & Shamili, 2022).

Artificial Intelligence (AI) is being used in agriculture to improve crop production, disease prediction, supply chain management, operational efficiency, and water waste reduction (Pallathadka et al., 2023). Machine learning (ML) and deep learning (DL) are commonly used for data prediction, disease prediction, water irrigation optimisation, sales growth, profit maximisation, inventory management, security, fraud detection, and portfolio management.

Various ML approaches can be utilised for crop price prediction, including regression-based methods, time series forecasting techniques, ensemble methods, DL strategies, and hybrid models (Singh & Sindhu, 2024). ML approaches have strengths, limitations, and practical applications. However, there are challenges like data accessibility, feature

[a] https://orcid.org/0009-0003-2482-5322
[b] https://orcid.org/0009-0006-1888-6411
[c] https://orcid.org/0000-0003-3646-1362
[d] https://orcid.org/0000-0002-7487-374X
[e] https://orcid.org/0000-0001-9940-1821

Dionissopoulos, I., Assimakopoulos, F., Spiliotopoulos, D., Margaris, D. and Vassilakis, C.
Predicting Agricultural Product and Supplies Prices Using Artificial Intelligence.

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 3: KMIS*, pages 371-379
ISBN: 978-989-758-716-0; ISSN: 2184-3228

selection, model interpretability, scalability, and generalisation (Cravero et al., 2022).

Many works provide insights for researchers, practitioners, and policymakers, facilitating informed decision-making in agricultural contexts (Assimakopoulos et al., 2024). ML and IoT-enabled farm machinery are key components of the next agriculture revolution. ML applications in agriculture focus on soil parameters, crop yield prediction, disease detection, and species detection. ML with computer vision can monitor crop quality and yield assessment. This approach can enhance livestock production, predict fertility patterns, diagnose eating disorders, and reduce human labour. Knowledge-based agriculture improves sustainable productivity and product quality (Sharma et al., 2021).

Smart farming, utilising AI, addresses agricultural sustainability challenges (Akkem et al., 2023). ML, DL, and time series analysis are crucial for crop selection, yield prediction, soil compatibility classification, and water management. These algorithms classify soil fertility, crop selection, and forecast production. Time series analysis helps predict demand, commodity price, and crop yield. As population growth increases, crop production forecasting is crucial to overcome food insufficiency. Using ML and DL techniques, crop recommendations can be made based on time series analysis to reduce future food insufficiency (Benos et al., 2021).

The purpose of this paper is to develop a model for forecasting agricultural commodity prices using historical data. The development includes the entire process flow, from data collection to the evaluation of the results using performance metrics of our forecasting model. Data analysis and the use of advanced ML techniques will enable the prediction of future prices of these products and the future agricultural production.

For the purposes of this work, data from Eurostat were used, as well as ancillary data of other parameters from other internet sources. Eurostat's open data portal offers us in a user-friendly and structured way information relating to the European Union and figures concerning a wide range of sectors and activities in its area of competence, including data relating to agricultural production.

In the remainder of the paper, Section 2 discusses the related work in ML and other related technologies for precision agriculture. Section 3 presents the data used and the preprocessing and integration methods that were applied to construct the training dataset. Section 4 presents the price forecasting methods that were implemented, while section 5 presents and discusses the results obtained. Finally, section 6 concludes the paper and outlines future work.

2 RELATED WORK

Agricultural data, economy data (market, local economy, wholesale), and world data are but a few domains that are useful for price prediction of agricultural goods. The data provide a strong foundation for innovative agricultural economic management and contributes to scientifically sound price prediction, as well as decision making in precision agriculture (Su & Wang, 2021).

Kumar et al. researched crop yield prediction using historical data to forecast crop yields, considering factors like temperature, humidity, and rainfall. The approach found that the Random Forest (RF) algorithm provides the best predictions, considering the least number of models, making it useful in the agriculture sector (Kumar et al., 2020).

Zhao used a wavelet method to smooth multiple sources of data and build a model to process the hierarchical information after signal decomposition (Zhao, 2021). Another study compared predictive accuracies of various ML techniques, focusing on GRNN, with the Autoregressive integrated moving average (ARIMA) model (Paul et al., 2022). Results showed GRNN outperforms other techniques in all seventeen markets, while RF is comparable in four. The Diebold-Mariano test confirmed these superior performances. Other techniques like SVR, GBM, and ARIMA are not as effective.

Xu & Zhang investigated corn cash price forecasting using univariate neural network (NN) modelling and bivariate NN modelling with futures prices. Results show high accuracy for one-day ahead horizons, with futures prices benefiting cash price forecasting. The framework was deemed easy to deploy and can be generalised to other commodities (Xu & Zhang, 2021).

Oktoviany et al. proposed a two-step hybrid model using ML methods to incorporate external factors in price changes (Oktoviany et al., 2021). The model assigns price states to historical prices and predicts future price states using short-term predictions. The model is applied to real corn futures data and generates price scenarios through Monte Carlo simulations. The simulations can be used to assess price risks in risk management systems or support trading strategies under different price states.

Another research used supervised ML for intelligent information prediction analysis to improve farming efficiency and profitability (Shakoor et al., 2017). The approach suggests area-based beneficial crop rankings, based on static data from previous years. This happens before the cultivation process. It indicates the crops that are cost effective for cultivation for a particular area of land. The study used Decision Tree Learning-ID3 and K-Nearest Neighbours Regression algorithms.

Time-series and ML models have also been deployed to predict monthly areca nut prices using SARIMA, Holt-Winter's Seasonal method, and LSTM neural networks (NNs). The LSTM NN model was found to be the best fit for the data (Sabu & Kumar, 2020). ANNs have also been used to predict soybean harvest area, yield, and production, comparing it with classical methods of Time Series Analysis (Abraham et al., 2020).

The work in (Purohit et al., 2021) proposed two additive hybrid methods and five multiplicative hybrid methods to predict the monthly retail and wholesale prices of three commonly used vegetable crops in India: tomato, onion, and potato (TOP). Extensive statistical analyses confirmed the superiority of the hybrid methods against existing statistical models, ML models, and existing hybrid methods in predicting TOP prices.

An alternative method that addresses the nonlinearity problem if time series approaches is wavelet transformation in generating hybrid models for predicting monthly prices markets. This hybrid model approach significantly improved over conventional techniques, utilising a combination of ANN and ML techniques (Paul & Garai, 2021).

Xu & Zhang investigated the use of nonlinear autoregressive neural networks (NARNN) and NARNN with exogenous inputs (NARNN–X) for price forecasting soybeans and soybean oil for periods that spanned over fifty years. The models exhibited accurate and stable performance, with relative root mean square errors of 1.701% and 1.777% for soybeans and 1.757% for soybean oil, respectively. Also, the approach can be generalised for other similar commodities (Xu & Zhang, 2022).

Menculini et al. examined various techniques for forecasting sale prices in an Italian food wholesaler, comparing ARIMA models, Prophet, which is a scalable forecasting tool from Facebook, and deep learning models, such as LSTM and CNNs. Results showed that ARIMA models and LSTM neural networks perform similarly, while the combination of CNNs and LSTMs achieved the best accuracy but requires more tuning time. Prophet was quick and easy to use but less accurate (Menculini et al., 2021).

3 DATA AND PREPROCESSING

Agricultural price prediction is a highly complex task, due to the fact that prices depend on numerous factors, both within the agricultural value chain and in the macroeconomic environment. Besides building a comprehensive dataset, encompassing the widest possible range of factors affecting prices, the quality and trustworthiness of data are of critical importance, in order to achieve high prediction accuracy. In the following paragraphs we describe the data sources used, as well as the integration and preprocessing methods used to formulate the training datasets.

3.1 Data Sources

Two key datasets for this research were obtained from Eurostat. These datasets are as follows:

1. Selling prices of crop products. These data cover the historical dimension of agricultural sales, supporting the dimension of analysis, which assumes that future patterns of agricultural prices will follow similar patterns, already been observed in the past.
2. Purchase prices of agricultural production means. The analysis considers this data, since the selling prices of agricultural products obviously depend on the prices of the means used for their production.

These datasets contain information spanning from 1969 to 2023; each dataset is provided in two parts, with the first covering the period 1969-2000 and the second spanning from 2001 to 2023. Since our price data is sourced from Eurostat, they contain only data for EU countries, hence price predictions in our experiments are limited to member states of the EU.

Energy cost is an important factor in the cost of agricultural production since oil is extensively used to operate motorised equipment, such as tractors and tillers, and is thus involved in the production cost. Consequently, we take Brent oil prices into account, in our predictions. It is considered as the most important indicator of energy spent in agricultural production, since its two main fuels, diesel and gasoline, are used to drive motorised equipment with internal combustion engines. Data concerning Brent oil prices were obtained from statista.com. At this stage of our research, Brent oil price is used as an overall indicator for energy cost. The inclusion of more detailed energy costs, notably electricity costs, is considered as part of our future research.

Land use data, from the World In data website, were also considered in our work. This dataset provides information on overall land use, cropland

land use, grazing land use and built-up area, per country and year.

The availability of human resources in agricultural production is also a factor impacting the prices of agricultural production. These data were obtained from Eurostat and cover the period from 1973 to 2023. The dataset provides a detailed breakdown of the total labour force to salaried and non-salaried workers. In our work, we consider all types of employees and hence we maintained only the sum of these two categories.

In our work, we also consider indicators of economic nature, concerning agriculture. From the Economic Accounts for Agriculture dataset, sourced from Eurostat, we extract and use the following data: (a) Production Value at Basic Price, (b) Subsidies on Products, (c) Tax on Products, and (d) Production Value at Producer Price.

3.2 Preprocessing and Integration

The data obtained from the above listed data sources were not directly utilisable for model training, necessitating preprocessing and integration activities.

Preprocessing activities concern the handling of missing data, noisy data, inconsistent data, encoding and value range discrepancies, and handling of textual data. In the following paragraphs, we outline the specific activities taken to address these issues.

Missing Data. Some attribute values may be missing due to an error, either in the registration process or because they were not provided by the relevant agency. For these cases, we considered firstly to find supplemental datasets that provided the missing values and integrate them into our dataset.

Values that were still missing, we applied imputers to fill in the missing values. For each data element, different imputers were considered and the effectiveness of the use of each imputer to predict a data element on the accuracy of the predictions was assessed. Experimental results demonstrated that the most accurate results were obtained by using the following imputers: (a) for Brent oil prices, backward fill (i.e. if a price is missing, use the price for the next known data point); (b) for the labour workforce, linear regression. For the agricultural economic accounts, a KNN-based imputer was applied, using N=20 (N denotes the number of nearest neighbours considered for computing a missing value).

For the cases which after the use of imputers data were still missing (because the imputers could not calculate the missing data due to the sparsity of the original dataset), the relevant records were dropped.

Noisy data: Data containing errors or outliers, which are highly deviant from the normal pattern, were discarded, since their use affected negatively the accuracy of predictions. The interquartile distance method (Vinutha et al., 2018) was used for identifying potential outliers and subsequently visual verification was conducted using graphs.

Inconsistent Data. Either duplicate values or data providing different values for a specific data element, for the same country and period. Data that were verified to be duplicates were discarded.

Differences in Units. Due to the currency change in many European countries, our data contained prices in both Euros and the previous local currency. For the algorithm to have comparable data at its disposal, price conversions to Euro were performed for countries that underwent currency changes.

Differences in Encoding. The price datasets obtained from Eurostat used different codes for agricultural products and supplies for the period 1969-2000 than for the period 2001-2023. To produce the integrated dataset, the product/supplies codes for the data concerning the period 1969-2000 were replaced by the respective codes used for the period 2001-2023. A fuzzy match on the names of the products was used to perform the mapping.

Handling of Textual Data. AI-based regression methods that were used for price prediction mainly work with numeric data and not textual data. Our datasets contain multiple cases where textual data are present, e.g., country names/codes and agricultural products/supplies names and codes For these cases, *label encoding* was employed, i.e., each distinct value of the respective data element was mapped to a unique integer, and only the mapped value was considered in the prediction process.

Different Scales. Different data elements had highly divergent scales (e.g. land availability and Brent oil prices), and this aspect negatively affected the accuracy of the predictions, due to overfitting. To mitigate this issue, each data column (except encoded labels and prices) was normalized to the range [0,1] using the *Min Max Scaler*; the normalized value NV produced by the Min Max Scaler for a value V is computed as $NV = \frac{value - MinVal}{MaxVal - MinVal}$, where $MinVal$ and $MaxVal$ are the minimum and maximum values for the specific column, respectively.

4 PRICE FORECASTING METHODS

In the previous section we presented the data collection, preprocessing and integration process. Following the above, all input data have been formulated in two comprehensive datasets:

- The crop products selling prices dataset,
- The agricultural production means dataset.

Each of these two datasets contains records with the following data elements: (i) country, (ii) agricultural product or means of production, (iii) year, (iv) price, (v) availability of labor in agricultural production, (vi) purchase and rental prices of the land, (vii) Brent oil prices and (viii) economic indicators of agricultural production (production value at basic price, subsidies on products, tax on products, and production value at producer price). These datasets can be used to train ML algorithms to perform predictions.

Since multiple AI-based methods and configurations are available for performing predictions, and each of these can be tuned through a number of hyperparameters, we resorted to the use of automatic machine learning (autoML) toolkits which underpin the tasks of method selection and hyperparameter tuning. To this end, the Autokeras and the TPOT autoML toolkits were used.

AutoKeras (https://autokeras.com/) is an open-source ML library, based on Keras and Tensorflow, which aims to build and optimise NNs automatically. In its basic function, the user only specifies whether a classification or a regression model is required, and the columns that are used for training, designating the target column for prediction.

TPOT (https://epistasislab.github.io/tpot/) is an open-source library that explores the performance of ML models in an automatic way, as well. It allows to search for the most efficient ML algorithm for the dataset used each time.

The hyperparameters used for the autoconfiguration process performed by the AutoKeras toolkit are as follows:

- *Tries.* The number of attempts AutoKeras will perform to arrive at the most efficient model. In this work we will experiment with 25 attempts for each dataset.
- *Test Size.* The percentage of training data that we will use for testing, in order to avoid Overfitting. In this work we will experiment with 30% of the data.
- *Number of Training Epochs*: i.e. the number of iterations in which each of our models is trained to approach the best result. In this work we will experiment with 30 seasons.

Table 1 illustrates the topology of the NN. This topology is designated as optimal for both price prediction tasks (agricultural products and supplies).

Table 1: The topology of the neural network.

Layer (type)	**Output Shape**	**Parameter value**
input_1 (InputLayer)	(None, 20)	0
multi_category_encoding (MultiCategoryEncoding)	(None, 20)	0
normalization (Normalization)	(None, 20)	41
dense (Dense)	(None, 32)	672
re_lu (ReLU)	(None, 32)	0
dense_1 (Dense)	(None, 128)	4224
re_lu_1 (ReLU)	(None, 128)	0
regression_head_1 (Dense)	(None, 1)	129

Table 2: Parameters for the Random Forest regressor.

Parameter	**Value**
n_estimators (The number of trees in the forest)	100
max_features (number of features to consider when looking for the best split)	75% of the number of input features
min_samples_leaf (the minimum number of samples a leaf node must contain)	7
min_samples_split (minimum number of samples required to split an internal node)	19

Table 3: Parameters for the Gradient Boosting regressor.

Parameter	**Value**
loss (Loss function used in optimization; the value huber combines squared error and absolute error)	huber
alpha (The alpha-quantile of the huber loss function and the quantile loss function)	0.8
learning_rate (moderates the contribution of each tree)	0.1
max_depth (moderates the maximum number of nodes in a tree, setting the maximum depth of the individual regression estimators)	7
max_features (number of features are considered in each split; value 1 indicates that all features are taken into account)	1.0
min_samples_leaf (the minimum number of samples a leaf node must contain)	1
min_samples_split (minimum number of samples required to split an internal node)	11
n_estimators (number of boosting stages that will be performed)	100
Subsample (percentage of samples used for fitting the individual base learners)	0.65

For the TPOT toolkit, the number of generations was set to 15, while the population size was set to 15. The population size refers to the number of individuals in each generation that retain their characteristics, as compared to the previous generation. The output of the TPOT toolkit determined that the optimal prediction method for agricultural product price prediction would be the random forest regression method, under the parameters illustrated in Table 2. Agricultural supplies prices, on the other hand, are more accurately predicted using Gradient Boosting, under the parameters listed in Table 3.

In the following section, the results and evaluation of this work will be presented and analysed.

5 RESULTS AND EVALUATION

In this section, the results and evaluation of this work are presented and analysed.

The prediction accuracy of our model can be assessed using performance metrics, which evaluate the closeness between the prediction result and the actual result. The metrics used in this work, are widely used in related research works that measure prediction The metrics are illustrated in Table 4, along with their respective formulas.

Table 4: The performance metrics used in our work.

Metric Name	Formula
Root Mean Square Error (RMSE)	$\sqrt{\sum_{i=i}^{n}(y_i - \hat{y}_i)^2 / n}$
Mean Average Error (MAE)	$\frac{1}{n} \times \sum_{i=i}^{n} \lvert y_i - \hat{y}_i \rvert$
Normalized MAE (NMAE)	$\left(\frac{1}{n} \times \sum_{i=i}^{n} \lvert y_i - \hat{y}_i \rvert\right) \Big/ \left(\frac{1}{n} \times \sum_{i=i}^{n} \lvert y_i \rvert\right)$

The RMSE metric boosts the significance of large deviations between the prediction result and the actual result, while the MAE handles all errors uniformly. The NMAE has the property of amortizing differences in the scale of the predicted variables, however, when the actual values are very small, errors are over-emphasised. In all the aforementioned metrics, lower values indicate smaller divergence and hence more accurate predictions.

In addition to the above, in this work, we include an additional performance metric, namely the *Percentage of Successful Predictions* (PSP); this metric computes the percentage of predictions that are deemed to be 'successful', and a price prediction $\hat{y}_i$ for time point *i* is considered successful iff

$$(y_i - y_{i-1}) * (\hat{y}_i - y_{i-1}) > 0$$

where y_i and y_{i-1} are the actual prices at time points *i* and *i-1*, respectively. Effectively, a prediction is considered to be successful iff either (a) a rise in the price is predicted and a rise actually occurred or (b) a drop in the price is predicted and a drop actually took place, otherwise the prediction is deemed unsuccessful. The *percentage of successful predictions* metric can be useful for assessing the utility of the approach for investment decisions, e.g., to invest on a particular product.

Tables 5 and 6 depict the accuracy metrics obtained from our experiments regarding the prediction of agricultural product sale prices and agricultural supplies, respectively.

In Table 5 we can observe that the NN optimised and proposed by AutoKeras achieves predictions that deviate from the actual prices by 6.6% on average (c.f. the NMAE metric), surpassing the accuracy of the Random Forest predictor proposed by TPot (average deviation 9.8%). The AutoKeras NN also achieves superior performance in predicting price rises or drops (80.96% vs. 73.29%).

In Table 6 we notice that both the AutoKeras NN and the gradient boosting predictor, proposed by TPot, formulate predictions with very small deviations from the actual prices (2.7% and 0.29%, respectively). While the gradient boosting predictor estimates actual prices better than the AutoKeras NN, it lags behind concerning the prediction of price rises or drops.

Table 5: Prediction accuracy for agricultural product sale prices.

Metric	Neural network (AutoKeras)	Random forest (Tpot)
RMSE	28.66	29.69
MAE	11.76	11.49
NMAE	0.0659	0.098
PSP	80.96%	73.29%

Table 6: Prediction deviation agricultural supplies prices.

Metric	Neural network (AutoKeras)	Gradient boosting (TPot)
RMSE	10.60	10.08
MAE	3.66	2.84
NMAE	0.0269	0.0029
PSP	84.96%	79.34%

The performance recorded for price predictions in our experiment surpasses the price prediction accuracy recorded for the works surveyed in section 2, which exhibit deviations from actual prices ranging from 12% to 26%. Since our experiment is limited to EU countries only, involving only countries for which historical data of high accuracy and ample time depth are available, more experimentation is required to fully compare the proposed algorithm against approaches proposed in the literature. This is considered a part of our future work.

Finally, in our experiments we can observe that prices of the means of agricultural production are predicted with higher accuracy than prices of agricultural products. This may be attributed to a dependence of agricultural product prices to additional factors than the ones considered in our work, while these factors suffice for the prediction of prices of means of agricultural production; this aspect will also be examined in our future work.

6 CONCLUSION

In this paper, we have presented a model for forecasting agricultural product and supply prices using historical data. We analysed the entire process flow, including data selection, preprocessing and integration, model training and algorithm tuning, as well as performance metrics and model evaluation.

The proposed model exhibits high accuracy for price predictions, especially for agricultural supplies, while it is also able to predict price rises or drops. Thus, the proposed algorithm can be used for budgeting production, estimating earnings and investment planning.

As richer datasets become available, especially with the advent of IoT, additional data can be taken into account for performing price predictions. Yet, developing countries are still challenged regarding the availability and accuracy of data. These aspects will be surveyed in our future work, elaborating on methods and techniques that are able to achieve high prediction accuracy over more sparse datasets.

ACKNOWLEDGEMENTS

This research was funded by project SODASENSE (https://sodasense.uop.gr) under grant agreement No. MIS 6001407 (co-financed by Greece and the EU through the European Regional Development Fund).

REFERENCES

Abraham, E. R., Mendes Dos Reis, J. G., Vendrametto, O., Oliveira Costa Neto, P. L. D., Carlo Toloi, R., Souza, A. E. D., & Oliveira Morais, M. D. (2020). Time Series Prediction with Artificial Neural Networks: An Analysis Using Brazilian Soybean Production. *Agriculture*, *10*(10), 475. https://doi.org/10.3390/agriculture10100475

Akkem, Y., Biswas, S. K., & Varanasi, A. (2023). Smart farming using artificial intelligence: A review. *Eng. Appl. of Artificial Intelligence*, *120*, 105899. DOI: 10.1016/j.engappai.2023.105899

Assimakopoulos, F., Vassilakis, C., Margaris, D., Kotis, K., & Spiliotopoulos, D. (2024). The Implementation of "Smart" Technologies in the Agricultural Sector: A Review. *Information*, *15*(8), 466. https://doi.org/10.3390/info15080466

Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine Learning in Agriculture: A Comprehensive Updated Review. *Sensors*, *21*(11), 3758. DOI: 10.3390/s21113758

Cravero, A., Pardo, S., Sepúlveda, S., & Muñoz, L. (2022). Challenges to Use Machine Learning in Agricultural Big Data: A Systematic Literature Review. *Agronomy*, *12*(3), 748. https://doi.org/10.3390/agronomy12030748

Deepa, S., Alli, A., Sheetac, & Gokila, S. (2023). Machine learning regression model for material synthesis prices prediction in agriculture. *Materials Today: Procs.*, *81*, 989–993. https://doi.org/10.1016/j.matpr.2021.04.327

Durai, S. K. S., & Shamili, M. D. (2022). Smart farming using Machine Learning and Deep Learning techniques. *Decision Analytics Journal*, *3*, 100041. https://doi.org/10.1016/j.dajour.2022.100041

Falah, R. A., & Rachmaniah, M. (2022). Price Prediction Model for Red and Curly Red Chilies using Long Short Term Memory Method. *Indonesian Journal of Statistics and Its Applications*, *6*(1), 143–160. https://doi.org/10.29244/ijsa.v6i1p143-160

Ge, Y., & Wu, H. (2020). Prediction of corn price fluctuation based on multiple linear regression analysis model under big data. *Neural Computing and Applications*, *32*(22), 16843–16855. https://doi.org/10.1007/s00521-018-03970-4

Ghutake, I., Verma, R., Chaudhari, R., & Amarsinh, V. (2021). An intelligent Crop Price Prediction using suitable Machine Learning Algorithm. *ITM Web of Conf.*, *40*, 03040. DOI: /10.1051/itmconf/20214003040

Hegde, G., Hulipalled, V. R., & Simha, J. B. (2021). Price Prediction of Agriculture Commodities Using Machine Learning and NLP. *2021 Second International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*, 1–6. https://doi.org/10.1109/ICSTCEE54422.2021.9708582

Im, J.-J., Kim, T.-W., Lim, J.-S., Kim, J.-H., Yoo, T.-Y., & Lee, W. J. (2022). A Design and Implement of Efficient Agricultural Product Price Prediction Model. *Journal of the Korea Society of Computer and Information*, *27*(5), 29–36. https://doi.org/10.9708/JKSCI.2022.27.05.029

Kumar, Y. J. N., Spandana, V., Vaishnavi, V. S., Neha, K., & Devi, V. G. R. R. (2020). Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector. *2020 5th Intl. Conf. on Comm. and Electronics Systems (ICCES)*, 736–741. https://doi.org/10.1109/ICCES48766.2020.9137868

Kurumatani, K. (2020). Time series forecasting of agricultural product prices based on recurrent neural networks and its evaluation method. *SN Applied Sciences*, *2*(8), 1434. https://doi.org/10.1007/s42452-020-03225-9

Lanjewar, M. G., Parate, R. K., & Parab, J. S. (2022). Machine Learning Approach with Data Normalization Technique for Early Stage Detection of Hypothyroidism. In M. K. Ahirwal, N. D. Londhe, & A. Kumar, *Artificial Intelligence Applications for Health Care* (1st ed., pp. 91–108). CRC Press. https://doi.org/10.1201/9781003241409-5

Li, J., Li, G., Liu, M., Zhu, X., & Wei, L. (2022). A novel text-based framework for forecasting agricultural futures using massive online news headlines. *International Journal of Forecasting*, *38*(1), 35–50. https://doi.org/10.1016/j.ijforecast.2020.02.002

Luo, J., Zhao, C., Chen, Q., & Li, G. (2022). Using deep belief network to construct the agricultural information system based on Internet of Things. *The Journal of Supercomputing*, *78*(1), 379–405. https://doi.org/10.1007/s11227-021-03898-y

Mahto, A. K., Alam, M. A., Biswas, R., Ahmed, J., & Alam, S. I. (2021). Short-Term Forecasting of Agriculture Commodities in Context of Indian Market for Sustainable Agriculture by Using the Artificial Neural Network. *Journal of Food Quality*, *2021*, 1–13. https://doi.org/10.1155/2021/9939906

Mamoudan, M. M., Mohammadnazari, Z., Ostadi, A., & Esfahbodi, A. (2024). Food products pricing theory with application of machine learning and game theory approach. *International Journal of Production Research*, *62*(15), 5489–5509. https://doi.org/10.1080/00207543.2022.2128921

Menculini, L., Marini, A., Proietti, M., Garinei, A., Bozza, A., Moretti, C., & Marconi, M. (2021). Comparing Prophet and Deep Learning to ARIMA in Forecasting Wholesale Food Prices. *Forecasting*, *3*(3), 644–662. https://doi.org/10.3390/forecast3030040

Mohanty, M. K., Thakurta, P. K. G., & Kar, S. (2023). Agricultural commodity price prediction model: A machine learning framework. *Neural Computing and Applications*, *35*(20), 15109–15128. https://doi.org/10.1007/s00521-023-08528-7

Oktoviany, P., Knobloch, R., & Korn, R. (2021). A machine learning-based price state prediction model for agricultural commodities using external factors. *Decisions in Economics and Finance*, *44*(2), 1063–1085. https://doi.org/10.1007/s10203-021-00354-7

Pallathadka, H., Mustafa, M., Sanchez, D. T., Sekhar Sajja, G., Gour, S., & Naved, M. (2023). Impact of Machine Learning on Management, Healthcare and Agriculture. *Materials Today: Proceedings*, *80*, 2803–2806. https://doi.org/10.1016/j.matpr.2021.07.042

Paul, R. K., & Garai, S. (2021). Performance comparison of wavelets-based machine learning technique for forecasting agricultural commodity prices. *Soft Computing*, *25*(20), 12857–12873. https://doi.org/10.1007/s00500-021-06087-4

Paul, R. K., Yeasin, Md., Kumar, P., Kumar, P., Balasubramanian, M., Roy, H. S., Paul, A. K., & Gupta, A. (2022). Machine learning techniques for forecasting agricultural prices: A case of brinjal in Odisha, India. *PLOS ONE*, *17*(7), e0270553. https://doi.org/10.1371/journal.pone.0270553

Purohit, S. K., Panigrahi, S., Sethy, P. K., & Behera, S. K. (2021). Time Series Forecasting of Price of Agricultural Products Using Hybrid Methods. *Applied Artificial Intelligence*, *35*(15), 1388–1406. https://doi.org/10.1080/08839514.2021.1981659

Rashid, M., Bari, B. S., Yusup, Y., Kamaruddin, M. A., & Khan, N. (2021). A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches With Special Emphasis on Palm Oil Yield Prediction. *IEEE Access*, *9*, 63406–63439. https://doi.org/10.1109/ACCESS.2021.3075159

Ribeiro, M. H. D. M., & 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. https://doi.org/10.1016/j.asoc.2019.105837

Sabu, K. M., & Kumar, T. K. M. (2020). Predictive analytics in Agriculture: Forecasting prices of Arecanuts in Kerala. *Procedia Computer Science*, *171*, 699–708. https://doi.org/10.1016/j.procs.2020.04.076

Shakoor, Md. T., Rahman, K., Rayta, S. N., & Chakrabarty, A. (2017). Agricultural production output prediction using Supervised Machine Learning techniques. *2017 1st International Conference on Next Generation Computing Applications (NextComp)*, 182–187. https://doi.org/10.1109/NEXTCOMP.2017.8016196

Sharma, A., Jain, A., Gupta, P., & Chowdary, V. (2021). Machine Learning Applications for Precision Agriculture: A Comprehensive Review. *IEEE Access*, *9*, 4843–4873. DOI: 10.1109/ACCESS.2020.3048415

Singh, N., & Sindhu, R. (2024). Crop Price Prediction Using Machine Learning. *Journal of Electrical Systems*, *20*(7s), 2258–2269. DOI: 10.52783/jes.3961

Su, Y., & Wang, X. (2021). Innovation of agricultural economic management in the process of constructing smart agriculture by big data. *Sustainable Computing: Informatics and Systems*, *31*, 100579. https://doi.org/10.1016/j.suscom.2021.100579

Vinutha, H. P., Poornima, B., & Sagar, B. M. (2018). Detection of Outliers Using Interquartile Range Technique from Intrusion Dataset. In S. C. Satapathy, J. M. R. S. Tavares, V. Bhateja, & J. R. Mohanty (Eds.), *Information and Decision Sciences* (Vol. 701, pp. 511–518). Springer Singapore. https://doi.org/10.1007/978-981-10-7563-6_53

Wang, J., Wang, Z., Li, X., & Zhou, H. (2022). Artificial bee colony-based combination approach to forecasting agricultural commodity prices. *International Journal of Forecasting*, *38*(1), 21–34. https://doi.org/10.1016/j.ijforecast.2019.08.006

Wang, L., Feng, J., Sui, X., Chu, X., & Mu, W. (2020). Agricultural product price forecasting methods: Research advances and trend. *British Food Journal*, *122*(7), 2121–2138. DOI: 10.1108/BFJ-09-2019-0683

Xu, X. (2020). Corn Cash Price Forecasting. *American Journal of Agricultural Economics*, *102*(4), 1297–1320. https://doi.org/10.1002/ajae.12041

Xu, X., & Zhang, Y. (2021). Corn cash price forecasting with neural networks. *Computers and Electronics in*

Agriculture, *184*, 106120. https://doi.org/10.1016/j.compag.2021.106120

Xu, X., & Zhang, Y. (2022). Soybean and Soybean Oil Price Forecasting through the Nonlinear Autoregressive Neural Network (NARNN) and NARNN with Exogenous Inputs (NARNN–X). *Intelligent Systems with Applications*, *13*, 200061. https://doi.org/10.1016/j.iswa.2022.200061

Yin, H., Jin, D., Gu, Y. H., Park, C. J., Han, S. K., & Yoo, S. J. (2020). STL-ATTLSTM: Vegetable Price Forecasting Using STL and Attention Mechanism-Based LSTM. *Agriculture*, *10*(12), 612. https://doi.org/10.3390/agriculture10120612

Zhang, D., Chen, S., Liwen, L., & Xia, Q. (2020). Forecasting Agricultural Commodity Prices Using Model Selection Framework With Time Series Features and Forecast Horizons. *IEEE Access*, *8*, 28197–28209. https://doi.org/10.1109/ACCESS.2020.2971591

Zhao, H. (2021). Futures price prediction of agricultural products based on machine learning. *Neural Computing and Applications*, *33*(3), 837–850. https://doi.org/10.1007/s00521-020-05250-6