Revista ARACÊ
ISSN: 2358-2472

APPLICATION OF RECURRENT NEURAL NETWORKS IN FRUIT PRICE PREDICTION IN THE AGRICULTURAL SECTOR OF BAHIA

https://doi.org/10.56238/arev7n1-073

Submission date: 12/06/2024 **Publication date:** 01/06/2025

Iara Fábia Castro Viana Silva[1], Caio Henrique Rodrigues Carvalho[2], Tiago do Carmo Nogueira[3], Danyele de Oliveira Santana[4], Gelson da Cruz Junior[5], Cássio Dener Noronha Vinhal[6] and Matheus Rudolfo Diedrich Ullmann[7]

ABSTRACT
This study investigates Recurrent Neural Networks (RNN), specifically LSTM and GRU, in the price forecasting of fruits grown by family farmers in Bahia, based on 1,883 records of banana nana, banana prata, and papaya formosa. The LSTM model presented the best performance, with RMSE results ranging from 0.186 to 0.606, MAE from 0.142 to 0.483, MAPE from 7.286 to 16.624, and MSE from 0.035 to 0.367 for the fruits analyzed. The potential of RNNs in supporting decision-making in the agricultural sector is highlighted, with proposals for future work that include incorporating exogenous variables and the development of a free platform for small producers.

Keywords: Agriculture. Time Series. Price Forecasting. Recurrent Neural Networks.

[1]Graduated in Information Technology Management (IFBAIANO)
Instituto Federal Baiano (IFBAIANO)
E-mail: zziarafcastro@gmail.com
[2]Master in Electrical Engineering (UFPI)
Instituto Federal Baiano (IFBAIANO)
E-mail: caio.carvalho@ifbaiano.edu.br
[3]Dr. in Electrical and Computer Engineering (UFG)
Instituto Federal Baiano (IFBAIANO)
E-mail: tiago.nogueira@ifbaiano.edu.br
[4]Master in Computer Science (UEFS)
Instituto Federal Baiano (IFBAIANO)
E-mail: danyele.santana@ifbaiano.edu.br
[5]Dr. in Electrical Engineering (Unicamp)
Federal University of Goiás (UFG)
E-mail: gcruzjr@ufg.br
[6]Dr. in Electrical Engineering (Unicamp)
Federal University of Goiás (UFG)
E-mail: vinagrel@ufg.br
[7]Doctorate student in Electrical and Computer Engineering (UFG)
Federal Institute of Bahia (IFBA)
E-mail: matheusullmann@ifba.edu.br

INTRODUCTION

Bahia is the second-largest producer of fresh fruit in Brazil, with 664,000 hectares dedicated to cultivation, according to data from the Bahia Federation of Agriculture and Livestock - FAEB (2023). In this context, the state stands out as the second-largest banana producer in the country, with a production of 866,000 tons (SEAGRI, 2023). Bananas are grown in all regions of Bahia, with the city of Bom Jesus da Lapa being one of the main production hubs. Another highlight is papaya, which, although cultivated throughout the national territory, has had more than 50% of its production concentrated in Bahia in recent decades, with the city of São Félix do Coribe as the main production center (LANDAU; SILVA, 2020).

Family farming is an essential part of Bahia's agricultural production, accounting for about 298,000 tons of products (CODEVASF, 2024). Characterized by cultivation on small rural properties, this mode of agriculture not only provides food for the farmers themselves but also significantly contributes to the local economy. However, family farmers face challenges such as socioeconomic vulnerabilities and low productive technology, which can negatively impact the production of fruits such as bananas and papayas (AQUINO; ALVES; VIDAL, 2020).

For agricultural production to be marketed viably, it is crucial that selling prices cover costs and prevent losses. In this scenario, accounting plays a fundamental role, helping rural producers make informed decisions that increase economic results (ASSIS et al., 2021). The lack of proper accounting treatment can lead to misinformation about costs, negatively impacting profit projections (DUTRA, 2020).

The instability of rural product prices, often caused by seasonal supply throughout the year, presents significant challenges in the financial management of family farmers. Seasonality, divided into long- and short-term periods, is an essential factor in agricultural forecasting and management (YOO; OH, 2020). While long-term seasonality relates to permanent changes in supply and demand, short-term seasonality involves temporary changes, such as variations in temperature and precipitation conditions (CHU et al., 2020).

Given these challenges, accurate and timely information on seasonal dynamics is increasingly necessary to optimize the management of cultivation systems and detect seasonal anomalies (BOSCHETTI et al., 2015). Seasonality results in price variations that often cause economic losses (RODRIGUES et al., 2021). Therefore, producers need to have effective financial controls to evaluate and decide on selling prices, even if they are influenced by the market (PEREIRA; SANTOS, 2022).

Additionally, traders also face difficulties in formulating agricultural prices due to variations in supply and demand, climate changes, and the cultivation process (OKTOVIANY; KNOBLOCH; KORN, 2021). In this context, the demand for methods to predict the seasonality of agricultural products has increased (LIU et al., 2021; SUN et al., 2023b). Under these circumstances, Digital Agriculture, which uses data and technological approaches to improve productivity, has proven to be a promising tool (KURUMATANI, 2018). It processes data generated in the agricultural field (KULBA; MEDENNIKOV, 2020), which is essential for the operation of forecasting models (YUAN; LING, 2020), allowing significant advances in understanding and solving problems in the sector.

Revista ARACÊ
ISSN: 2358-2472

Time series forecasting, which involves analyzing sequences of observations over time, such as days, months, quarters, or years (RAHMAN et al., 2023), to make predictions using statistical methods and modeling (XU; HSU, 2022), has emerged as a powerful technique in detecting seasonal anomalies. In this way, Artificial Neural Networks (ANNs) have attracted great attention from researchers in this area of time series forecasting (LIU et al., 2021), being applied to predict product prices in the agricultural market, which contributes to stabilizing supply and demand (YOO; OH, 2020). ANNs offer a wide range of possibilities, surpassing traditional static methods (YUAN; LING, 2020), and are particularly effective in capturing nonlinear patterns that may exist without an apparent reason (OKTOVIANY; KNOBLOCH; KORN, 2021; KURUMATANI, 2018).

However, when it comes to forecasting with seasonal factors, Recurrent Neural Networks (RNNs) are generally preferred. Their "memory" capability in data transmission allows them to capture sequential patterns, such as seasonality, more effectively, overcoming the limitations of other models (LIU et al., 2021). Thus, RNNs stand out as an essential tool for efficient agricultural price forecasting, facilitating both data training and decision-making based on complex data sets.

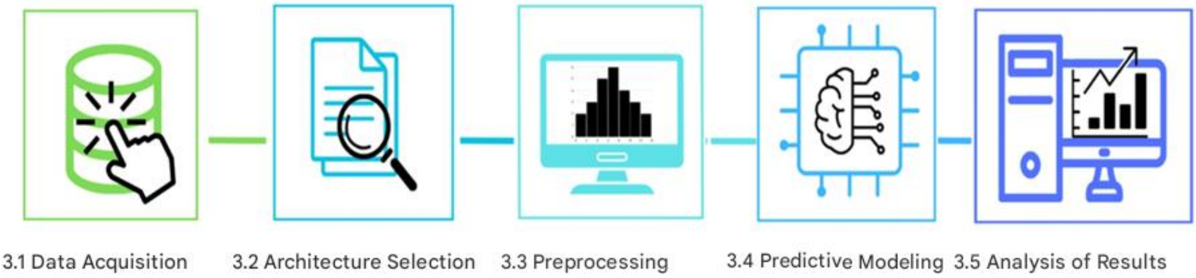
Accurate price forecasting can enable the agricultural production network to make strategic decisions to deal with price variations (REDDY et al., 2022). Based on this, this study proposes the use of models based on Recurrent Neural Networks to develop an effective methodology for price forecasting of some of the main fruits cultivated in the state of Bahia, such as bananas and papayas. The aim is to support small agricultural producers in economic decision-making and optimizing their operations, assisting in financial management and risk reduction.

To achieve this goal, the Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models were selected and evaluated using various metrics, such as Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), based on a dataset available in the online repository of the Center for Advanced Studies in Applied Economics (CEPEA) at the University of São Paulo (USP). This approach aims to identify the most suitable model for the agricultural scenario in Bahia.

This article is organized as follows: Section 2 presents related works on machine learning techniques in agricultural product price forecasting, Section 3 presents the methodology used in this study, Section 4 presents the results and discussions, and Section 5 discusses the conclusions of this study.

Revista ARACÊ
ISSN: 2358-2472

Figure 1. Steps of this Methodology.

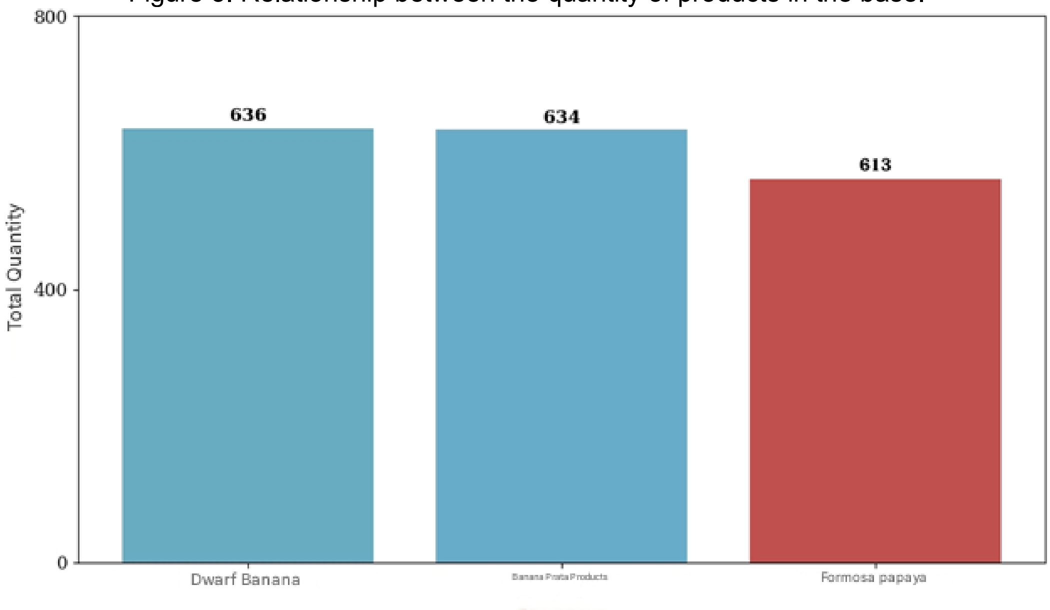


DATA ACQUISITION

To develop this work, a dataset containing daily information on the prices of three typical fruits from the Bahia region was used: banana nana, banana prata, and papaya formosa, available in the online repository of the Center for Advanced Studies in Applied Economics (CEPEA, 2024), of the University of São Paulo (USP).

The database contains 1,270 price records for bananas (Monica and Prata) and 613 records for papaya Formosa, from 2012 to 2023. Organized in a table format, it contains five columns: Product (type of fruit), Region, Date, Unit (kilogram, 22 kg box, and 20 kg box), and Price. Figure 2 illustrates the products and their respective quantities in the database.

Figure 3. Relationship between the quantity of products in the base.



Revista ARACÊ
ISSN: 2358-2472

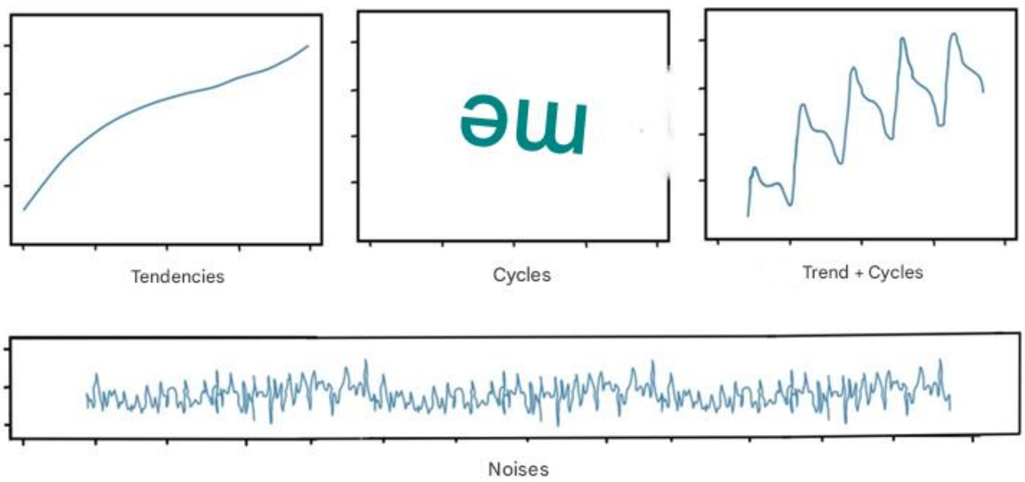
SELECTION OF ARCHITECTURES

A time series consists of a sequence of well-defined observations obtained through repeated measurements over time, such as days, months, quarters, or years (RAHMAN et al., 2023). Forecasting these series involves analyzing the data to make a prognosis, using statistical methods and modeling (XU; HSU, 2022). Time series analysis is a powerful technique for detecting seasonal anomalies. It is assumed that the observed values present repeating patterns over time. The general model to represent these patterns is given by the equation:

$$Y_t = T + S + \epsilon$$

Y_t represents the value of the time series at an instant t, T corresponds to the trend, S to the seasonality, and ϵ to the error or random noise that can interfere with the analysis SUN et al., 2023b). Figure 3 illustrates how the components can appear graphically.

Figure 4. Trend, seasonality (cycles), and noise in time series.



Source: Adapted from Seber and Wild (2004).

The state-of-the-art review aimed to identify the most commonly used architectures for predicting agricultural product prices. Table 2 summarizes the key information from the related studies. A clear predominance of RNNs in time series forecasting tasks was observed, highlighting their effectiveness and popularity in this field.

Table 2. Architectures used in related studies.

Studies	Architectures
Kurumatani (2020)	SRNN, LSTM, and GRU

Revista ARACÊ
ISSN: 2358-2472

Yuan and Ling (2020)	ARIMA, LSTM, SVR, Prophet, and XGBoost
Sabu and Kumar (2020)	SARIMA, Holt-Winter, and LSTM
Chu et al. (2020)	EEMD, LSSVM, EEMD-R-ADD, SVM, BPNN, and ELM
Zheng et al. (2020)	MLP
Oktoviany, Knobloch, and Korn (2021)	K-NN and Random Forest
Ozden (2023)	CNN, LSTM, and Random Forest
Sun et al. (2023a)	Random Forest, ELM, LSTM, EEMD–LSTM, VMD–LSTM
Ray et al. (2023)	ARIMA, GARCH, LSTM, and ARIMA-LSTM

Among the analyzed models, Long Short-Term Memory (LSTM) stood out as the most frequently used in related studies. Its selection for this study is due to its effectiveness in processing long-time series and its ability to address the vanishing gradient problem during training (ZHANG et al., 2019; LIU; CHEN; WU, 2020). LSTM is distinguished by its gating mechanism, which allows the network to retain relevant information for extended periods, making it ideal for agricultural price forecasting tasks.

Additionally, the Gated Recurrent Unit (GRU) model was also chosen due to its similarity to LSTM but with a more simplified structure and fewer parameters. While it shares the ability to process sequences through input, output, and forget gates, GRU transmits only a single hidden state between cells, which can lead to a reduction in training time without significantly compromising performance (KURUMATANI, 2020; FOROUTAN; LAHMIRI, 2024). Therefore, the LSTM and GRU models were selected for price forecasting in the agricultural scenario of Bahia.

Preprocessing

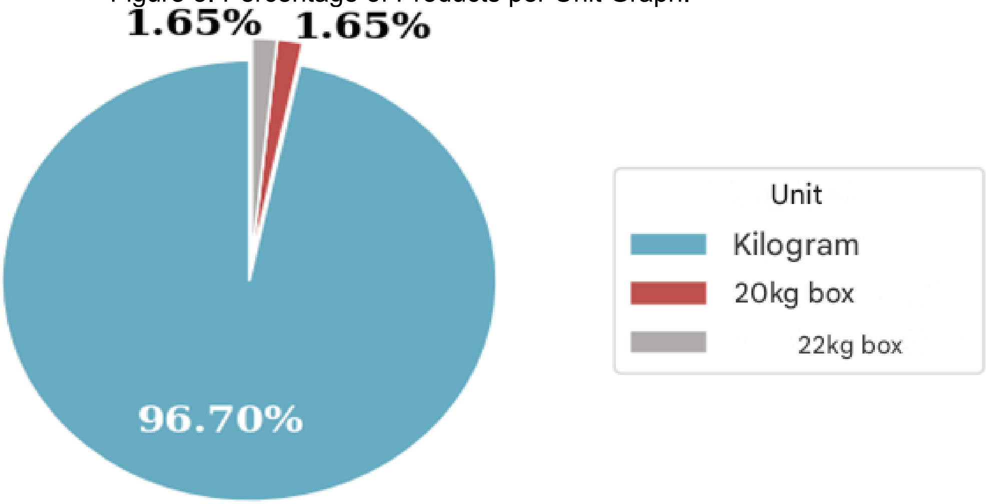
The preprocessing steps involved data cleaning, removal of unwanted values, selection of relevant features, and data scaling (QULMATOVA; KARIMOV; AZIMOV, 2022).

First, the presence of null/invalid values in the dataset was checked, and the identified records were removed. Additionally, it was observed that the products were classified into three categories based on their unit of mass, as illustrated in Figure 4. The “box” categories represent products sold in larger quantities, resulting in a higher average price compared to the price per kilogram. Since entries associated with the "box" category account for only 3.3% of the total dataset, they were discarded to prevent distortions in the analysis.

.

Revista ARACÊ
ISSN: 2358-2472

Figure 5. Percentage of Products per Unit Graph.



The data was then normalized to fit a range between 0 and 1. This normalization process is essential because it helps standardize the scale of the data, facilitating the convergence of machine learning models during training. By keeping all values within the same range, models can learn more efficiently and prevent variables with larger scales from dominating the optimization process (SEVERO, 2023). With the preprocessing completed, it was possible to move on to the next stage of the methodology.

PREDICTIVE MODELING

To perform the prediction, following state-of-the-art practices, the processed database was divided into two parts: data from January 2012 to December 2022 were used for training, and the year 2023 was reserved for testing. The training data is used for the model to learn to identify patterns in the information, while the test data allows the model's efficiency to be analyzed. This division allows the model's ability to make predictions with new data to be assessed. The data were transformed into a time series with daily frequency.

Revista ARACÊ
ISSN: 2358-2472

Figura 6. Time series

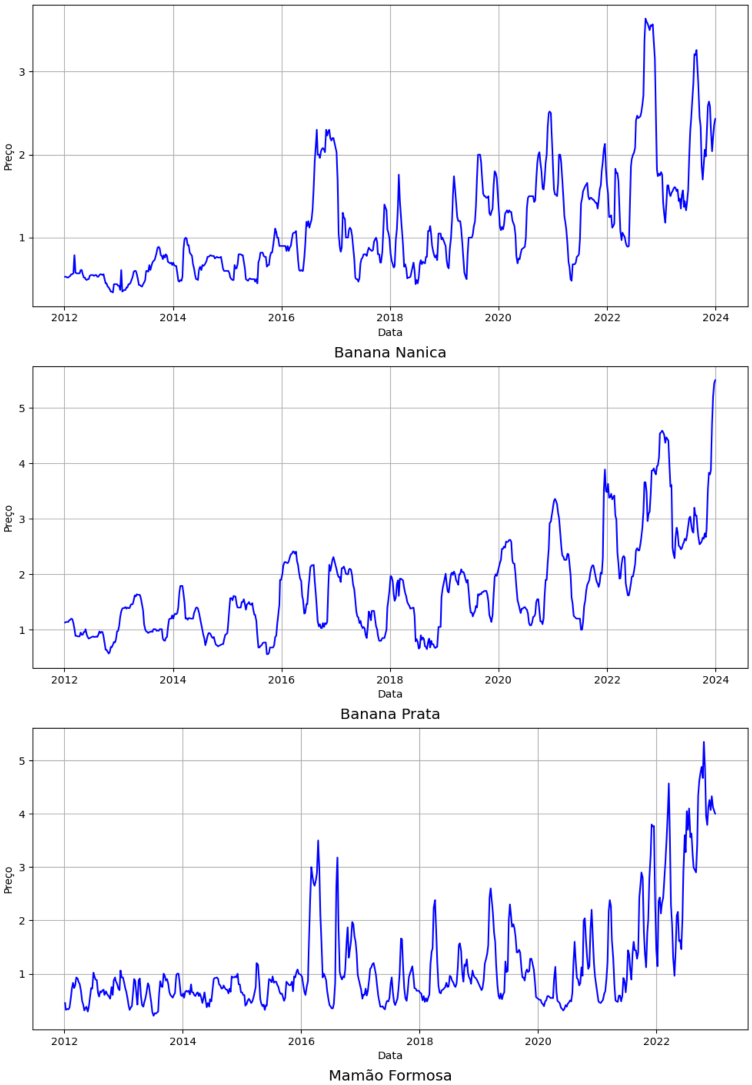


Figure 5 shows that the prices of the base products exhibit trends, seasonality, and noise, confirming that the data are time series. As indicated in the work of Yuan, San, and Leong (2020), it is clear that prices are unstable, with clear trends and seasonal patterns. To deal with these data, choosing a model that can capture both the characteristics of the time series and seasonality is the most appropriate.

The models applied in this study use RNNs, which are effective for analyzing time series due to their short-term memory capacity (LIU et al., 2021). RNNs are composed of artificial neurons that simulate the structure of a biological neural network (LIU; CHEN; WU, 2020), having an architecture composed of input, hidden, and output layers (KURUMATANI, 2020). The input layer receives and processes the initial information through activation functions. The hidden layer then processes and analyzes the data, generating predictions. Finally, the output layer provides the result.

Revista ARACÊ
ISSN: 2358-2472

Gradient descent is used during the training process and tries to find coefficients that best fit the data until the best weights for the model are found (CHENG; WEI; CHENG, 2020). The weights are used to weigh the influence of each neuron and are adjusted during training to improve prediction performance. Its mathematical formula is as follows::

$$yt = g(VA_t)$$
$$A_t = f(U_{x_t} + WA_{t-1}$$

From the RNN formula, it can be seen that at time t: the input value, hidden value, and output value of the network are: X_t, A_t, Y_t (LIU et al., 2021). However, RNNs suffer from gradient loss when dealing with long data sequences (YUAN; LING, 2020), the error value of the gradients can become very small, making it difficult for the network to learn (MÁRQUEZ et al., 2021). This problem is overcome using the Long Short-Term Memory model - LSTM (SABU; KUMAR, 2020).

LSTM is a special type of RNN that solves the vanishing gradient problem (YUAN; LING, 2020). It has an Input Gate, Output Gate, and Forget Gate. Activations are calculated in each layer and compared with the desired output. The forget gate and input gate are used to control the position of the unit, and the output controls the outputs generated by the LSTM (RAY et al., 2023). In this way, the LSTM can analyze time series with a long prediction range, allowing the exploration of dynamic changes in the input sequences (YUAN; LING, 2020). The mathematical formula of the LSTM is as follows::

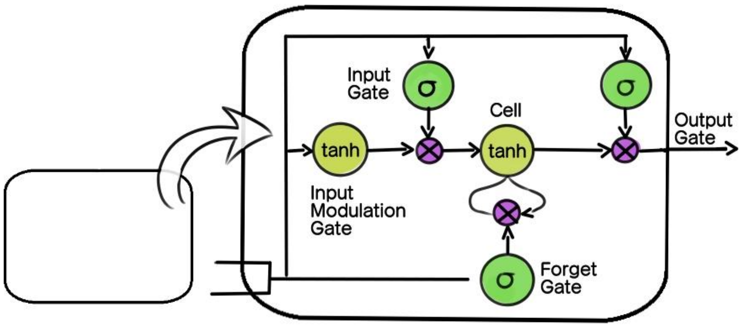
$$f_t = \sigma(W_{f\,A}A_{t-1} + W_{f\,X}X_t + b_f$$
$$i_t = \sigma(W_{i\,A}A_{t-1} + W_{i\,X}X_t + b_i$$
$$\tilde{c}_t = tanh(W_{c\,A}A_{t-1} + W_{c\,X}X_t + b_c$$
$$c_t = f_t{\circ}c_{t-1} + i_t{\circ}\tilde{c_t}$$
$$Y_t = \sigma(W_{Y\,A}A_{t-1} + W_{Y\,X}X_t + b_Y$$
$$A_t = Y_t \circ \tanh(c_t)$$

In the LSTM cell, it represents the forget gate, which is what to discard from memory. The input gate, i_t, determines what to add to memory, while ~ct generates the new data to potentially store. The memory state, ct, updates the memory state based on the forget gate and the input gate. The output gate, Y_t, decides the value to be used as

Revista ARACÊ
ISSN: 2358-2472

the output of the cell, and the final hidden state, At, is generated based on the memory state and the output gate. Figure 6 shows a graphical representation of the LSTM cell.

Figure 6. Graphical representation of LSTM



Source: Adapted from Abdelwahed, Letaifa, and Ksouri (2023).

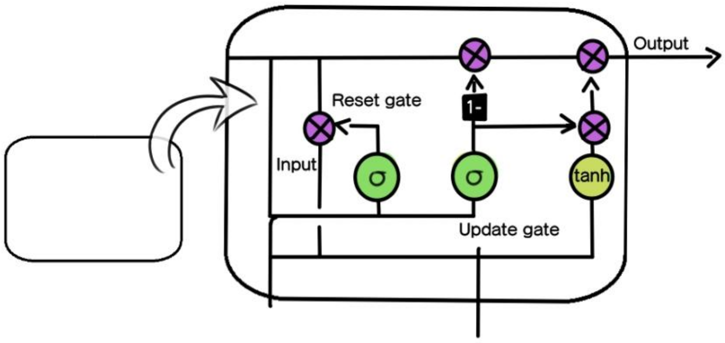
The GRU, like the LSTM, is a special type of RNN, but with a simpler structure (MIGLIATO, 2021). It has only two gates, a reset gate and an update gate, which decide which information should be passed to the output (KURUMATANI, 2018). The GRU cell equations can be defined as:

$$
\begin{aligned}
z_t &= \sigma(W_x X_t + U_z h_{t-1} + b_z) \\
r_t &= \sigma(W_r X_t + U_r h_{t-1} + b_r) \\
\hat{h}_t &= \tanh\,(W_h X_t + U_h(r_t \odot h_{t-1}) + b_h) \\
h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \hat{h}_t
\end{aligned}
$$

In the GRU cell, X_t represents the input value that contains the information to be processed. The update gate vector z_t controls which old information should be kept and which new information should be incorporated. The reset gate vector, rt decides how much of the old information should be forgotten. The output vector, h_t shows the hidden state of the cell at the current time, and hˆ_t suggests the new state before the control gate decisions are applied. Figure 7 shows a representation of the GRU cell.

Revista
ARACÊ
ISSN: 2358-2472

Figure 7. Graphical representation of the GRU.



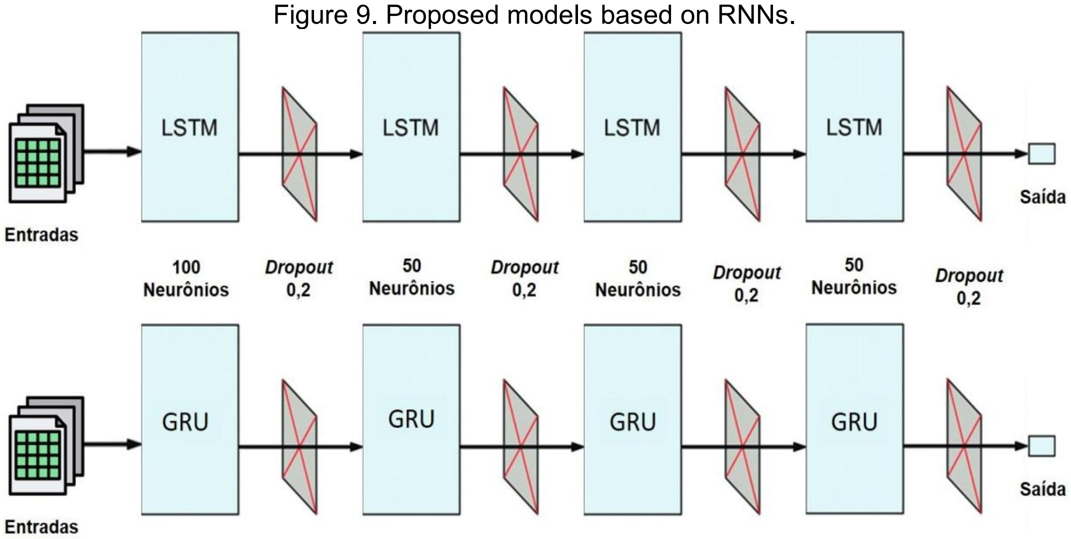
Source: Adapted from Wang et al. (2024).

While LSTM carries long-term and short-term memory through the cell state and the hidden state, GRUs have only one hidden state, which carries short-term memory, although both models are superior to a simpler RNN model (SILVA, 2020).

To build the LSTM and GRU models used in this study, the input data were prepared in 30-day lag windows. After empirical analysis, the models were configured with four layers, the first with 100 neurons and the three subsequent layers with 50 neurons each. Each layer is followed by a Dropout layer with a 20% rate to reduce the risk of overfitting, a situation where the model fails to generalize and instead fits too closely to the training data set. In addition, a final dense layer is added, containing an output unit with linear activation, responsible for price prediction. Figure 8 illustrates the architecture of the applied models..

Revista ARACÊ
ISSN: 2358-2472

Figure 9. Proposed models based on RNNs.



ANALYSIS OF RESULTS

To assess the ability of price forecasting models, several authors use error metrics to validate simulation results (ZHANG et al., 2019; LIU; CHEN; WU, 2020; QULMATOVA et al., 2022; DWIVEDI et al., 2021). These metrics compare predicted values with actual values, focusing on the difference between them. The main error metrics include RMSE, MAE, MAPE, and MSE, where Y_t represents the actual value at time i, Yˆis the predicted value for the same time, and n is the total number of data points (SUN et al., 2023a).

MAE calculates the average of the absolute errors between predicted and actual values, directly measuring the difference between the actual and predicted values. Absolute errors are expressed in the real units of the data. A MAE closer to 0 indicates a more accurate model, as illustrated by the formula below:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |Y_i - \widehat{Y_i}|$$

MAPE, in turn, calculates the absolute error in percentage terms, offering a relative view of the error. However, there is no universal standard for a MAPE considered excellent, as variations in prices can make accurate forecasting difficult (QULMATOVA et al., 2022):

$$MAPE = \frac{100}{n}\sum_{i=1}^{n} |\frac{Y_i - \widehat{Y_i}}{Y_i}|$$

Revista ARACÊ
ISSN: 2358-2472

MSE is a common metric in assessing the accuracy of time series (DWIVEDI et al., 2021). MSE calculates the mean of the squares of the differences between predicted and actual values. High MSE values indicate that the model performed poorly in forecasting.:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2$$

The RMSE, in turn, provides the square root of the mean of the squares of the differences between the predicted and actual values (AGARWAL; RAY; TRIPATHI, 2023). Using the square root helps to keep the units consistent and makes it easier to interpret the results. A lower RMSE indicates a more effective model.:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2}$$

RESULTS AND DISCUSSION

The hyperparameters used in training the models were defined based on an empirical approach, where different combinations were tested and adjusted to achieve the best possible performance in terms of accuracy and generalization ability. Table 3 summarizes the chosen hyperparameters, which include both model architecture aspects and training parameters.

Table 3. Hyperparameters Used in Model Training.

Type	Hyperparameter
Architecture	LSTM and GRU
Dropout	0.2
Activation Function in Output Layer	Linear
Optimizer	RMSprop
Learning Rate	0.001
Training Loss Function	MSE
Training Evaluation Metric	MAE
Batch Size	8, 16, and 32
Training Epochs	50, 100, 150, and 200

The models were implemented using LSTM and GRU layers, chosen for their ability to capture temporal dependencies in sequential data. Both architectures were tested due to their proven effectiveness in time series tasks, such as agricultural price forecasting, which

Revista ARACÊ
ISSN: 2358-2472

is the focus of this study. The Dropout technique was applied with a rate of 0.2 to reduce the risk of overfitting, ensuring better generalization during training.

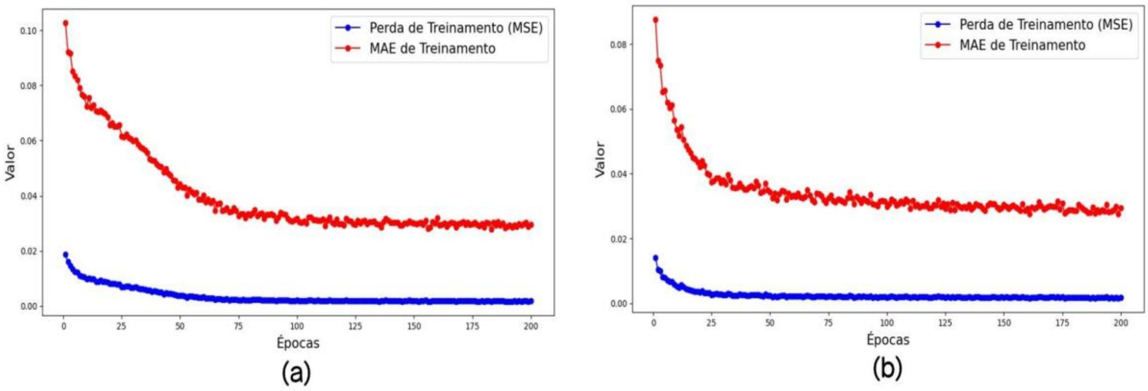
A linear activation function was adopted in the output layer due to the nature of the regression problem, where the goal is to predict continuous values, such as prices. The linear activation is the most natural choice for this type of task, as it allows the network to produce outputs unrestricted by fixed intervals. The RMSprop optimizer was selected for its effectiveness in time series problems, where gradient variances can vary significantly across iterations. RMSprop dynamically adjusts the learning rate for each parameter, contributing to a more stable and efficient training process (DIJKINGA, 2024). A learning rate of 0.001 was chosen based on preliminary tests. This value is commonly used as a starting point for the RMSprop optimizer and has proven suitable for ensuring stable convergence while preventing slow learning.

The Mean Squared Error (MSE) loss function was employed as it is widely used in regression problems. MSE penalizes larger deviations more severely, helping the model focus on minimizing high-magnitude errors. Mean Absolute Error (MAE) was chosen as the evaluation metric because it provides a direct interpretation of the average error in terms of output units. By complementing the MSE loss function, MAE facilitates model performance analysis throughout training, particularly in terms of predictability (JUNIOR, 2021).

Batch sizes of 8, 16, and 32 were tested to determine the best balance between computational efficiency and gradient stability. These values are commonly used in deep learning tasks and offer different trade-offs between accuracy and training time. Finally, the number of epochs was adjusted to 50, 100, 150, and 200, allowing observation of model behavior at different training stages. This range of values enabled an assessment of whether the model was improving with increased training epochs.

As observed in Figure 10, in the experiment for "banana nanica" using hyperparameters of 8 Batch Size and 200 Epochs, it is evident that the MSE of the LSTM model is significantly lower than that of the GRU model during training. This indicates that LSTM better adapted to the time series of the product.

. Figure 10. (a) Example of LSTM training and (b) Example of GRU training.



Revista ARACÊ
ISSN: 2358-2472

The evaluation of the models was conducted using the metrics presented in subsection 3.5. Table 4 presents the performance of the LSTM model on the test set under different hyperparameter configurations for each fruit type, comparing the RMSE, MAE, MAPE, and MSE values. The best results were obtained with: 200 epochs and a batch size of 8 for "banana nanica," 100 epochs and a batch size of 8 for "banana prata," and 100 epochs with a batch size of 16 for "mamão formosa."

Table 4. Best Results Using LSTM.

Product	Epochs	Batch Size	RMSE	MAE	MAPE	MSE
Banana Nanica	200	8	0.186	0.142	7.286	0.035
Banana Prata	100	8	0.278	0.196	5.853	0.077
Mamão Formosa	100	16	0.606	0.483	16.624	0.367

In Table 5, a comparison was made using the same hyperparameters, but applied to the GRU model. The best result for "banana nanica" was 200 epochs and a batch size of 8, for "banana prata" 50 epochs and a batch size of 8, and for "mamão formosa" 100 epochs and a batch size of 32.

Table 5. Best Results Using GRU.

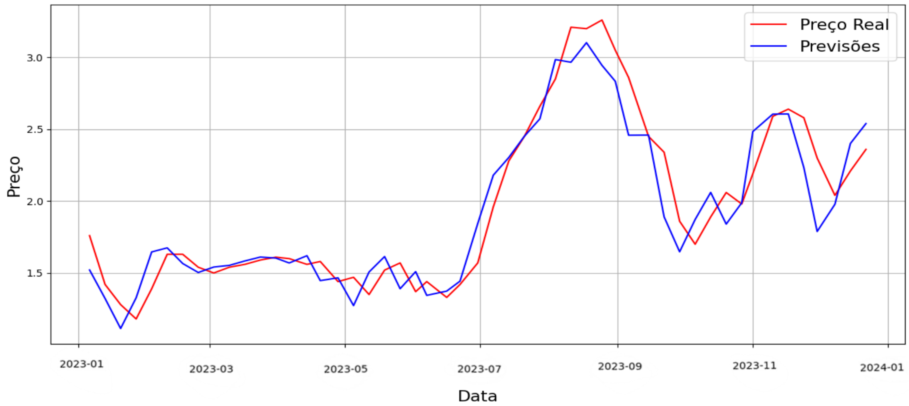
Product	Epochs	Batch Size	RMSE	MAE	MAPE	MSE
Banana Nanica	200	8	0.192	0.145	7.391	0.037
Banana Prata	50	8	0.283	0.193	5.938	0.080
Mamão Formosa	100	32	0.639	0.503	17.830	0.409

Figure 11 shows the comparison between the prices generated by the selected models in this research (LSTM) and the actual price of each product. The solid red line represents actual values, while the blue line represents predicted values.

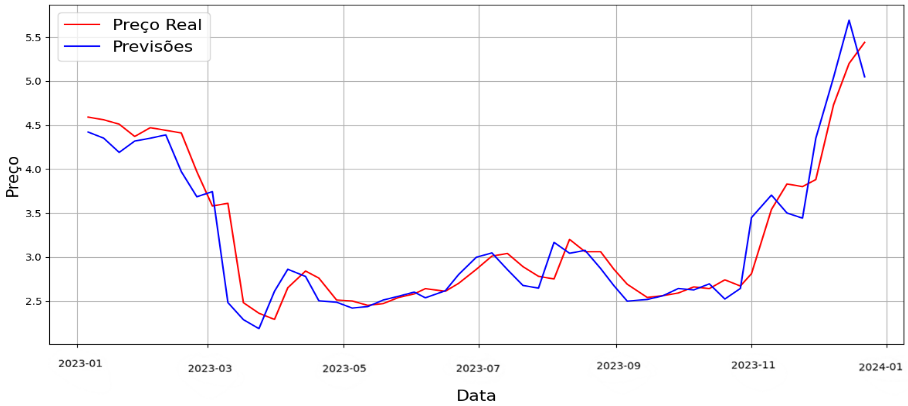
.

Revista ARACÊ
ISSN: 2358-2472

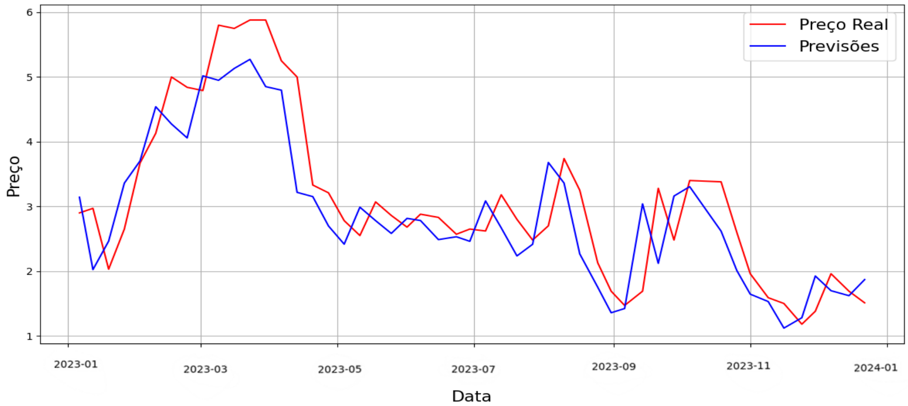
Figure 11. Comparison between Real Price and Predicted Price of Fruits in the model.



Banana Nanica



Banana Prata



Mamão Formosa

Revista ARACÊ
ISSN: 2358-2472

CONCLUSIONS AND FUTURE WORK

Price forecasting is a globally discussed topic due to the losses caused by the lack of financial control. To assist in the creation of decision-making strategies, or to avoid the damage caused by price fluctuations, researchers from different countries are looking for methods that can automatically predict these prices, bringing the lowest possible error rate. Predictive analysis based on RNNs has proven to be a promising technology. In this research, predictive analysis was used as a resource to predict the price of the 3 main fruits produced in the territory of Bahia by family farmers. The models were applied to a database containing, in total, 1,883 price records for the fruits, nanica banana, prata banana and formosa papaya. Through the results obtained, and after comparing the models, it is concluded that the LSTM model applied in this work presented a satisfactory performance in the context of academia, achieving the objective of the research. It is worth noting that predictive analysis can help with financial and risk management, identifying periods where prices fluctuate the most, creating a way for producers to gradually adjust their resale price, increasing profitability. In this way, it is expected that this model can help producers create measures that facilitate management control.

As a proposal for future work, we intend to improve the model presented with exogenous variables, since the prices of agricultural products are influenced by several factors, such as temperature fluctuations, holidays and consumption levels. Combining prices with these factors can be a valuable way to improve their accuracy. Additionally, we will build an online platform where small producers will be able to consult price projections free of charge and intuitively, in addition to analyzing other crops with a focus on more regions of Brazil, aiming to provide independent and reliable information to the agricultural sector..

Revista ARACÊ
ISSN: 2358-2472

REFERENCES

1. Abdelwahed, N., Letaifa, A. B., & Ksouri, A. (2023). Predicting sentiment analysis for web users with a deep learning approach. In 2023 IEEE Tenth International Conference on Communications and Networking (ComNet) (pp. 1–11). IEEE.

2. Agarwal, N., Ray, S., & Tripathi, K. (2023). Time series forecasting of agriculture yield of cotton with regression model implementation. In 2022 OPJU International Technology Conference on Emerging Technologies for Sustainable Development (OTCON) (pp. 1–6). IEEE.

3. Aquino, J. R. de, Alves, M. O., & Vidal, M. de F. (2020). Agricultura familiar no nordeste do Brasil: Um retrato atualizado a partir dos dados do censo agropecuário 2017. Revista Econômica do Nordeste, 51(Suplemento Especial), 31–54.

4. Assis, B. H., et al. (2021). A importância da contabilidade e do direito no agronegócio. Revista Projetos Extensionistas, 1(1), 195–208.

5. Boschetti, M., et al. (2015). Assimilating seasonality information derived from satellite data time series in crop modelling for rice yield estimation. In 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (pp. 157–160). IEEE.

6. CEPEA. (2024). Banco de dados Cepea. https://www.hfbrasil.org.br/br/banco-de-dados-precos-medios-dos-hortifruticolas.aspx?produto=4®iao%5B%5D=6&periodicidade=anual&anoinicial=2020&anofinal=2024&pagina=1

7. Cheng, W., Wei, S., & Cheng, F. (2020). Grey system correlation-based feature selection for time series forecasting. In 2020 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS) (pp. 837–841). IEEE.

8. Chu, X., et al. (2020). A novel machine learning-based strategy for agricultural time series analyzing and forecasting: A case study in China's table grape price. In 2020 IEEE 8th International Conference on Computer Science and Network Technology (ICCSNT) (pp. 75–80). IEEE.

9. Codevasf. (2024). Cerca de 298,3 mil toneladas de itens agrícolas foram produzidas em 2023 nos projetos de irrigação da Codevasf no oeste da Bahia. https://www.codevasf.gov.br/noticias/2024

10. Dijkinga, F. J. (2024). The RMSprop optimizer. https://medium.com/@fernando.dijkinga/the-rmsprop-optimizer-78f02efb63e9

11. Dutra, T. R. (2020). A aplicação da contabilidade na propriedade rural: As ferramentas contábeis como método de gestão [Trabalho de Conclusão de Curso]. Repositório de Trabalhos de Conclusão de Curso.

12. Dwivedi, S. A., et al. (2021). Analysis and forecasting of time-series data using SARIMA, CNN and LSTM. In 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS) (pp. 131–136). IEEE.

13. Econômica, B. (2023). Bahia é segundo estado no ranking de produção de banana do país. https://bahiaeconomica.com.br/wp/2023/09/22/bahia-e-segundo-estado-no-ranking-de-producao-de-banana-do-pais/

14. FAEB, F. da Agricultura e Pecuária do Estado da Bahia. (2023). Doce Bahia: Estado é potência nacional na fruticultura. https://sistemafaeb.org.br/doce-bahia-estado-e-potencia-nacional-na-fruticultura

15. Foroutan, P., & Lahmiri, S. (2024). Deep learning systems for forecasting the prices of crude oil and precious metals. Financial Innovation, 10(1), 111. https://doi.org/10.1186/s40854-023-00509-6

16. Inbox, O. (2024). Banana Prata 1kg. https://hmg.organicosinbox.com.br/produto/banana-prata-1kg/

17. Junior, C. de O. (2021). Métodos para regressão. https://medium.com/data-hackers/prevendo-números-entendendo-métricas-de-regressão-35545e011e70

18. Kulba, V., & Medennikov, V. (2020). A model of Russia's agriculture readiness for digital transformation. In 2020 13th International Conference "Management of Large-Scale System Development" (MLSD) (pp. 1–5). IEEE.

19. Kurumatani, K. (2018). Time series prediction of agricultural products price based on time alignment of recurrent neural networks. In 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 81–88). IEEE.

20. 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/s42452-020-03229-7

21. Landau, E. C., & Silva, G. A. da. (2020). Evolução da produção de mamão (Carica papaya, Caricaceae). In E. C. Landau, G. A. da Silva, L. Moura, A. Hirsch, & D. P. Guimarães (Eds.), Produção de mamão (pp. 1–XX). [Publisher not specified].

22. Liu, D., Chen, A., & Wu, J. (2020). Research on stock price prediction method based on deep learning. In 2020 2nd International Conference on Information Technology and Computer Application (ITCA) (pp. 69–72). IEEE.

23. Liu, Z., et al. (2021). Forecast methods for time series data: A survey. IEEE Access, 9, 91896–91912. https://doi.org/10.1109/ACCESS.2021.3090755

24. Márquez, J. P., et al. (2021). Ethanol fuel demand forecasting in Brazil using an LSTM recurrent neural network approach. IEEE Latin America Transactions, 19(4), 551–558. https://doi.org/10.1109/TLA.2021.9448534

25. Migliato, A. L. T. (2021). Detecção de outliers em dados não vistos de séries temporais por meio de erros de predição com SARIMA e redes neurais recorrentes LSTM e GRU [Tese de doutorado]. Universidade de São Paulo.

26. 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/s10203-021-00354-6

27. Ozden, C. (2023). Comparative analysis of CNN, LSTM and random forest for multivariate agricultural price forecasting. Black Sea Journal of Agriculture, 6(4), 422–426. https://doi.org/10.47115/bsagriculture.1234567

28. Pereira, J. L. S. R., & Santos, N. S. dos. (2022). Controle financeiro na agricultura familiar: Uma investigação sobre sua utilização e relevância. Epitaya E-books, 1(9), 132–143.

29. Proativa. (2024). Mamão Formosa. https://www.proativaalimentos.com.br/mamao-formosa

30. Qulmatova, S., et al. (2022). Crop production under different climatic conditions by analyzing agricultural data using multiple linear regression, Winter Holt, and artificial intelligence. In Proceedings of the 6th International Conference on Future Networks & Distributed Systems (pp. 242–252). [Publisher not specified].

31. Qulmatova, S., Karimov, B., & Azimov, D. (2022). Data analysis and forecasting in agricultural enterprises. In Proceedings of the 6th International Conference on Future Networks & Distributed Systems (pp. 536–541). [Publisher not specified].

32. Rahman, M. T., et al. (2023). Time series forecasting of agricultural products sale using deep learning. In 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1–6). IEEE.

33. Ray, S., et al. (2023). An ARIMA-LSTM model for predicting volatile agricultural price series with random forest technique. Applied Soft Computing, 149, 110939. https://doi.org/10.1016/j.asoc.2023.110939

34. Reddy, P. C. S., et al. (2022). Data analytics in farming: Rice price prediction in Andhra Pradesh. In 2022 5th International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT) (pp. 1–5). IEEE.

35. Rodrigues, J. L., et al. (2021). Desafios para a sustentabilidade da cadeia produtiva do abacaxi em Itaberaba, Bahia. Revista Metropolitana de Sustentabilidade, 11(3).

36. Sabu, K. M., & Kumar, T. 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

37. Santos, V. F. dos, Maciel, L. dos S., & Ballini, R. (2020). Efeito das operações de hedge e especulação sobre a volatilidade dos preços de commodities agrícolas nos EUA. Economía Aplicada, 24(3), 343–366.

38. SEAGRI. (2023). Dia da Banana: Segundo maior produtor do Brasil, estado da Bahia investe em. http://www.seagri.ba.gov.br/noticias/2023/09/22/dia-da-banana-segundo-maior-produtor-do-brasil-estado-da-bahia-investe-em

39. Seber, G. A. F., & Wild, C. J. (2004). Encontros com o acaso: Primeiro curso de análise de dados e inferência (1st ed.). [Publisher not specified].

40. Severo, J. (2023). Normalização de dados contínuos. https://pt.linkedin.com/pulse/normalização-de-dados-johanes-severo

41. Silva, E. F. da. (2020). Alocação das reservas internacionais com base em previsão de fatores econômicos e financeiros utilizando redes neurais recorrentes e modelo Nelson-Siegel dinâmico [Dissertação]. [Institution not specified].

42. Sun, C., et al. (2023). A study on agricultural commodity price prediction model based on secondary decomposition and long short-term memory network. Agriculture, 14(1), 60. https://doi.org/10.3390/agriculture14010060

43. Sun, F., et al. (2023). Agricultural product price forecasting methods: A review. Agriculture, 13(9), 1671. https://doi.org/10.3390/agriculture13091671

44. Wang, T., et al. (2024). Application of a multi-model fusion forecasting approach in runoff prediction: A case study of the Yangtze River source region. Sustainability, 16(14), 5964. https://doi.org/10.3390/su16145964

45. Xu, J.-L., & Hsu, Y.-L. (2022). Analysis of agricultural exports based on deep learning and text mining. The Journal of Supercomputing, 78(8), 10876–10892. https://doi.org/10.1007/s11227-021-04245-0

46. Yoo, T.-W., & Oh, I.-S. (2020). Time series forecasting of agricultural products' sales volumes based on seasonal long short-term memory. Applied Sciences, 10(22), 8169. https://doi.org/10.3390/app10228169

47. Yuan, C. Z., & Ling, S. K. (2020). Long short-term memory model based agriculture commodity price prediction application. In Proceedings of the 2020 2nd International Conference on Information Technology and Computer Communications (pp. 43–49). [Publisher not specified].

48. Yuan, C. Z., San, W. W., & Leong, T. W. (2020). Determining optimal lag time selection function with novel machine learning strategies for better agricultural commodity prices forecasting in Malaysia. In Proceedings of the 2020 2nd International Conference on Information Technology and Computer Communications (pp. 37–42). [Publisher not specified].

49. Zhang, S., et al. (2019). Optimizing time-series prediction on China's green trade economy. In 2019 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 1579–1584). IEEE.

50. Zheng, G., et al. (2020). The research on agricultural product price forecasting service based on combination model. In 2020 IEEE 13th International Conference on Cloud Computing (CLOUD) (pp. 4–9). IEEE.