Application of an LSTM Network to CEPEA Indicators to Estimate the Price of Arabica Coffee

1st Deivison Oliveira Costa
Department of Engineering and Computing (DEC)
Instituto Federal de Minas Gerais (IFMG)
Bambuí, Brazil
0034436@academico.ifmg.edu.br

2nd Artur Francisco Pereira Carvalho
Department of Engineering and Computing (DEC)
Instituto Federal de Minas Gerais (IFMG)
Bambuí, Brazil
0034430@academico.ifmg.edu.br

***Abstract*—This paper presents a method for estimating the price of Arabica coffee using Long Short-Term Memory (LSTM) networks, a type of Artificial Neural Networks (ANNs). The approach leverages Center for Advanced Studies on Applied Economics (CEPEA) indicators as input features for training and validating the LSTM model, which exploits temporal dependencies in coffee price data to predict future trends. Experimental results demonstrate the effectiveness of the proposed approach, with the trained LSTM network achieving a coefficient of determination of 0.9679 and a Mean Absolute Percentage Error (MAPE) of 4.7576 when applied to the separated test set. These findings support informed decision-making in the coffee market.**

***Index Terms*—Long Short-Term Memory, Artificial Neural Networks, Price estimation, Arabica coffee, CEPEA indicators.**

I. INTRODUCTION

Fluctuations in agricultural commodity prices pose significant challenges to the stability of global markets and impact stakeholders across the supply chain, from producers to consumers [1]–[4], [9]. These fluctuations necessitate accurate prediction models to mitigate risks and facilitate informed decision-making. Traditional methods like Auto-Regressive Integrated Moving Average (ARIMA) and Vector Auto-Regression (VAR) often fall short in capturing the complex dynamics inherent in agricultural commodity markets [2]. However, recent advancements in machine learning, particularly the application of LSTM networks, have shown promise in enhancing predictive capabilities [1], [3].

In this context, this research proposes the application of an LSTM network to CEPEA indicators, which comprise the daily prices in dollars of Arabica coffee since 09/02/1996, to estimate the same. Leveraging LSTM networks enables the identification of feature correlations and temporal relationships within multivariate time series data, thereby enhancing predictive accuracy [1], [3].

Existing research has demonstrated the efficacy of LSTM networks in forecasting agricultural commodity prices by leveraging dynamic information, such as the dynamic main production area [1]. This departure from conventional static models has yielded promising results, showcasing lower MAPE compared to benchmark models [1]. Additionally, the fusion of LSTM forecasts with traditional methods has shown further improvements in predictive accuracy, highlighting the potential of hybrid approaches [4].

The utilization of LSTM networks for commodity price prediction offers a novel avenue for research, with implications for stakeholders seeking to navigate volatile agricultural markets. By harnessing freely accessible datasets and advanced machine learning algorithms, this research seeks to contribute to the development of robust predictive models, ultimately fostering market stability and informed decision-making.

In this paper, we present an LSTM-based model trained on CEPEA indicators to forecast Arabica coffee prices, building upon the foundation laid by previous research [1]–[4], [9]. The following sections will delve into the methodology, experimental setup, results, and implications of our approach, culminating in a comprehensive analysis of the application of LSTM networks to agricultural commodity price forecasting.

II. RELATED WORKS

A. Machine Learning Models for Agricultural Commodity Price Prediction

GU et al. [1] proposed a dual input attention LSTM (DIA-LSTM) model for forecasting agricultural commodity prices, achieving 2.8% to 5.5% lower MAPE compared to conventional methods. Furthermore, their approach resulted in 1.41% to 4.26% lower MAPE than benchmark models.

OUYANG et al. [2] introduced the Long- and Short-Term Time-series Network (LSTNet), which outperformed the neural baseline Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), ARIMA, VAR by 6.52%, 21.68%, 91.70%, and 91.69% in the Root Squared Error (RSE) metric, and by 10.10%, 33.75%, 93.67%, and 93.66% in the Relative Absolute Error (RAE) metric respectively when the horizon is 24.

FOFANAH [5] explored linear regression, Extreme Gradient Boosting (XGB), and LSTM techniques for price prediction in the coffee market. Their results showed that the Root Mean Squared Error (RMSE) of the average forecast was 0.21 and 21.49 percent lower, respectively, for cotton. For oil, the forecast averaging did not provide improvements in terms of RMSE.

B. Forecasting Techniques for Commodity Markets

SUMAN et al. [4] emphasized the significance of accurate commodity price prediction for informed trading decisions, with their LSTM algorithm achieving a RMSE that was 21.49% lower for cotton. They suggested using a forecast averaging method for improved results.

NOVANDA et al. [9] conducted a comparative analysis of forecasting techniques for coffee prices, where the ARIMA model demonstrated superior forecasting accuracy compared to other methods, with a MAPE of 3.76%, Mean Absolute Deviation (MAD) of 0.074, and Mean Squared Deviation (MSD) of 0.010 for world coffee prices, and a MAPE of 0.9%, MAD of 141.2, and MSD of 43455 for domestic coffee prices.

C. Coffee Price Prediction Using Deep Learning Models

MEKALA et al. [7] proposed a novel approach combining CNN and Bidirectional LSTM (BLSTM) models for coffee price prediction, achieving performance levels superior to traditional time-series models.

SETIYANI and UTOMO [6] utilized the LSTM algorithm to predict Arabica coffee prices, achieving accurate price predictions with a Mean Absolute Error (MAE) value of less than 0.6. The LSTM model demonstrated good performance, as evidenced by the Mean Squared Error (MSE) loss function graph and the MAE metric, which approached close to 0 during training and testing.

D. Advanced Techniques for Price Forecasting

DEINA et al. [8] proposed a methodology for coffee price forecasting based on Extreme Learning Machines (ELM). The ELM consistently outperformed other models, achieving lower MSE, MAE, and MAPE values. For instance, for Arabica coffee, the ELM achieved an MSE of 125.5083 compared to the next best model Multi-Layer Perceptron (MLP) with an MSE of 155.3316, indicating a 19% improvement. Similarly, for Robusta coffee, the ELM outperformed MLP by 30% in terms of MSE.

E. Observation

By examining the literature, it is evident that a variety of machine learning and traditional forecasting methods have been applied to predict agricultural commodity prices, particularly in the context of coffee markets. These studies offer valuable insights into the effectiveness of different techniques and provide a foundation for further research in this domain.

III. Theoretical Foundation

A. LSTM Algorithm

The LSTM algorithm, devised by Hochreiter and Schmidhuber in 1997 [10], addresses the vanishing gradient problem in traditional RNNs, facilitating effective learning of long-range dependencies in sequential data. LSTM architecture comprises interconnected LSTM units with essential components: input gate, forget gate, output gate, memory cell, and hidden state (Fig. 1). These components operate as follows:

1) Input Gate (i_t): controls new information flow into the memory cell by integrating current input x_t, previous hidden state h_{t-1}, weight matrices W_{xi} and W_{hi}, and bias term b_i through a sigmoid activation function:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \tag{1}$$

2) Forget Gate (f_t): decides what information to discard from the memory cell by considering x_t, h_{t-1}, W_{xf}, W_{hf}, and b_f, calculating the forget gate activation via a sigmoid function:

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \tag{2}$$

3) Output Gate (o_t): governs information flow from the memory cell to the hidden state using x_t, h_{t-1}, W_{xo}, W_{ho}, and b_o, computing the output gate activation with a sigmoid function:

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \tag{3}$$

4) Memory Cell (c_t): stores and updates information by combining input and forget gates, and x_t through element-wise multiplication, addition, and the hyperbolic tangent activation function:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{4}$$

5) Hidden State (h_t): derived by multiplying the updated memory cell c_t with the output gate activation o_t, after applying the hyperbolic tangent activation function. The hidden state h_t encapsulates important information relevant to subsequent tasks:

$$h_t = o_t \odot \tanh(c_t) \tag{5}$$

The LSTM algorithm effectively captures long-term dependencies in sequential data, making it suitable for tasks involving temporal dynamics and extended context.

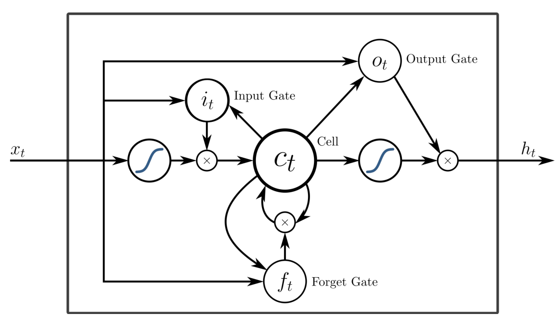

Fig. 1. LSTM Model

B. Attention Mechanisms

In the context of LSTM networks applied to time series forecasting, attention mechanisms play a important role in capturing relevant information from historical data. Bahdanau et al. (2016) proposed an attention mechanism for sequence-to-sequence tasks, which has since been widely adopted in various applications, including natural language processing and time series analysis [11].

The attention mechanism allows the LSTM model to focus on specific parts of the input sequence when making predictions, mimicking the human cognitive process of selectively attending to relevant information. This is particularly useful in scenarios where long sequences of historical data are involveds.

In the Bahdanau attention mechanism [11], the decoder network generates an output sequence while aligning each generated element to the encoded input sequence. This alignment is achieved by assigning attention weights to different parts of the input sequence, indicating their relevance to the current prediction. The attention weights are computed dynamically at each time step based on the hidden states of the LSTM network and the input sequence.

Mathematically, the attention mechanism can be described as follows:

Let $h_1, h_2, ..., h_N$ be the hidden states of the LSTM encoder corresponding to the input sequence, and s_{t-1} be the previous hidden state of the LSTM decoder at time step $t-1$. The attention weight α_{tn} for the n-th input element at time step t is calculated using a scoring function e_{tn}:

$$e_{tn} = \text{score}(s_{t-1}, h_n) \tag{6}$$

$$\alpha_{tn} = \frac{\exp(e_{tn})}{\sum_{n'} \exp(e_{tn'})} \tag{7}$$

where score is a function that computes the compatibility between the decoder state s_{t-1} and the encoder state h_n. Common choices for the scoring function include dot product, additive, and multiplicative forms.

The attention weights α_{tn} represent the importance of each input element in influencing the prediction at time step t. These weights are then used to compute a context vector c_t, which is a weighted sum of the encoder states:

$$c_t = \sum_n \alpha_{tn} h_n \tag{8}$$

The context vector c_t provides the LSTM decoder with relevant information from the input sequence for making accurate predictions at each time step.

C. Time Series Forecasting

Hyndman and Athanasopoulos (2018) [12] presents time series forecasting as a fundamental task in predictive analytics, involving the prediction of future values based on past observations. Traditional methods, such as autoregressive models and moving averages, have long been employed for time series forecasting. However, recent advancements in deep learning, particularly RNNs and their variants, have shown promise in capturing complex temporal patterns in time series data [12].

D. Feature Engineering in Machine Learning

Guyon and Elisseeff (2006) [13] present an introduction to variable and feature selection, emphasizing the role feature engineering plays in machine learning. This process involves selecting, transforming, and creating input features to enhance model performance. Effective feature engineering can significantly impact the predictive accuracy and generalization ability of machine learning models. Various techniques, including normalization, scaling, and dimensionality reduction, are employed to preprocess input data and extract informative features [13].

References

[1] Y. H. Gu, et al., ”Forecasting agricultural commodity prices using dual input attention LSTM,” *Agriculture*, vol. 12, no. 2, p. 256, 2022.

[2] H. Ouyang, X. Wei, and Q. Wu, ”Agricultural commodity futures prices prediction via long-and short-term time series network,” *Journal of Applied Economics*, vol. 22, no. 1, pp. 468-483, 2019.

[3] R. Ly, F. Traore, and K. Dia, ”Forecasting commodity prices using long-short-term memory neural networks,” *Intl Food Policy Res Inst*, 2021.

[4] S. Suman, et al., ”Commodity Price Prediction for making informed Decisions while trading using Long Short-Term Memory (LSTM) Algorithm,” in *2022 5th International Conference on Contemporary Computing and Informatics (IC3I)*, IEEE, 2022, pp. 406-411.

[5] A. J. Fofanah, ”Machine learning model approaches for price prediction in coffee market using linear regression, XGB, and LSTM techniques,” *International Journal of Scientific Research in Science and Technology*, no. 6, 2021.

[6] L. Setiyani and W. H. Utomo, ”Arabica Coffee Price Prediction Using the Long Short Term Memory Network (LSTM) Algorithm,” *Scientific Journal of Informatics*, vol. 10, no. 3, pp. 287-296, 2023.

[7] K. Mekala, et al., ”Coffee Price Prediction: An Application of CNN-BLSTM Neural Networks,” in *2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, IEEE, 2023, pp. 1-7.

[8] C. Deina, et al., ”A methodology for coffee price forecasting based on extreme learning machines,” *Information Processing in Agriculture*, vol. 9, no. 4, pp. 556-565, 2022.

[9] R. R. Novanda, et al., ”A comparison of various forecasting techniques for coffee prices,” in *Journal of Physics: Conference Series*, IOP Publishing, 2018, p. 012119.

[10] S. Hochreiter and J. Schmidhuber, ”Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, Nov. 1997.

[11] D. Bahdanau, et al., ”End-to-end attention-based large vocabulary speech recognition,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2016, pp. 4945-4949.

[12] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*. OTexts, 2018.

[13] I. Guyon and A. Elisseeff, ”An introduction to variable and feature selection,” *Journal of Machine Learning Research*, vol. 3, pp. 1157-1182, 2006.