STATE UNIVERSITY OF MARINGÁ
TECHNOLOGY CENTER
PRODUCTION ENGINEERING DEPARTMENT
POSTGRADUATE PROGRAM IN PRODUCTION ENGINEERING

LUCAS GROGENSKI MELOCA

Recurrent Neural Networks and Transformer-based models for multi-step prediction of agricultural commodity prices.

Maringá
2023

LUCAS GROGENSKI MELOCA

Recurrent Neural Networks and Transformer-based models for multi-step prediction of agricultural commodity prices.

Dissertation presented to the Postgraduate Program in Production Engineering of the Department of Production Engineering, Technology Center of the State University of Maringá, as a partial requirement for obtaining the title of Master's in Production Engineering. Area of concentration: Production Engineering.

Advisor: Prof. Dr. Ademir Aparecido Constantino
Co-advisor: Prof. Dr. Rodrigo Clemente Thom de Souza

Maringá
2023

Dados Internacionais de Catalogação-na-Publicação (CIP)
(Biblioteca Central - UEM, Maringá - PR, Brasil)

Meloca, Lucas Grogenski

M528r Recurrent neural networks and transformer-based models for multi-step prediction of agricultural commodity prices / Lucas Grogenski Meloca. -- Maringá, PR, 2023.
96 f.: il. color., figs., tabs.

Orientador: Prof. Dr. Ademir Aparecido Constantino.
Coorientador: Prof. Dr. Rodrigo Clemente Thom de Souza.
Dissertação (Mestrado) - Universidade Estadual de Maringá, Centro de Tecnologia, Departamento de Engenharia de Produção, Programa de Pós-Graduação em Engenharia de Produção, 2023.

1. Redes neurais recorrentes. 2. Modelos baseados em *Transformers*. 3. Previsão de séries temporais. 4. *Commodities* agrícolas. 5. Otimização de hiperparâmetros. I. Constantino, Ademir Aparecido, orient. II. Souza, Rodrigo Clemente Thom de, coorient. III. Universidade Estadual de Maringá. Centro de Tecnologia. Departamento de Engenharia de Produção. Programa de Pós-Graduação em Engenharia de Produção. IV. Título.

CDD 23.ed. 005.1

Elaine Cristina Soares Lira - CRB-9/1202

FOLHA DE APROVAÇÃO

LUCAS GROGENSKI MELOCA

Redes Neurais Recorrentes e modelos baseados em Transformers para predição multi-step de preços de commodities agrícolas

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção do Centro de Tecnologia da Universidade Estadual de Maringá, como requisito parcial para obtenção do título de Mestre em Engenharia de Produção pela Banca Examinadora composta pelos membros:

BANCA EXAMINADORA

Prof. Dr. Ademir Aparecido Constantino
Presidente Orientador
Universidade Estadual de Maringá – PGP/UEM

Prof. Dr. Rafael Henrique Palma Lima
Membro examinadora interno
Universidade Estadual de Maringá – PGP/UEM

Prof. Dr. Carlos Henrique Wachholz de Souza
Membro examinador externo
Universidade Federal do Paraná - UFPR

Prof. Dr. Rodrigo Clemente Thom de Souza
Universidade Federal do Paraná – UFPR
Coorientador

Aprovada em: 11 de agosto de 2023.
Local da defesa: Sala de Projeção, Bloco 19, *campus* da Universidade Estadual de Maringá.

To my loving parents, Pedro and Lucileide, and my incredible sister, Ana Clara, who have been unwavering pillars of support throughout my academic journey. Your constant presence, unwavering belief in my abilities, and tireless efforts to ensure my education has surpassed any words I can muster.

ACKNOWLEDGMENTS

First and foremost, I offer my heartfelt gratitude to God for His unwavering love and blessings throughout this journey.

To my remarkable parents, Pedro and Lucileide, words cannot express my profound appreciation for the exceptional education and solid values you have instilled in me. Your unwavering support and guidance have been the bedrock of my achievements. I owe everything I have in life to you.

To my dear sister, Ana Clara, you are the driving force behind my perseverance and the source of my boundless joy. Your unwavering belief in me has been a constant source of inspiration.

To my godparents, Marlene and Miguel, thank you for never giving up on me and for assuming the roles of second parents. Your unwavering support and guidance have played an integral role in shaping my journey.

To my beloved grandparents, your presence and unwavering emphasis on the importance of family have been a source of strength. Your unwavering belief in me has fueled my determination.

To my siblings, Edson, José, and Zélia, thank you for teaching me the values of humility and simplicity. Your unwavering support and camaraderie have been invaluable.

To my advisor and co-supervisor, Ademir Aparecido Constantino and Rodrigo Clemente Thom de Souza, I am immensely grateful for your guidance and support from the very beginning. Your expertise and mentorship have been instrumental in shaping me into a professional.

To my dear friends, Anderson Feola, Igor Cedran, and William Paschoalin, your unwavering support and assistance throughout every stage of my master's degree have been invaluable. I am deeply grateful for your unwavering support and friendship.

To my friend and teacher, André Luiz Justi, thank you for your unwavering support in my pursuit of an academic career. Your guidance and encouragement have been transformative.

To all the individuals who have shown love and affection during my learning process, your encouragement and support have fueled my determination.

To my classmates, thank you for walking alongside me on this journey of friendship and professional growth. Your companionship has made this experience truly meaningful.

To my esteemed teachers, I am forever indebted to you for your dedication in imparting knowledge. The gift of education you have bestowed upon me will forever be treasured.

To the Graduate Program in Production Engineering, thank you for providing unwavering support in the pursuit of knowledge. Your commitment to excellence has been invaluable. To Elisandra, the PGP secretary, thank you for your attentiveness and patience. Your support has been truly appreciated.

To the program's coordination, thank you for your unwavering goodwill and wholehearted support of the students. Your dedication has been remarkable.

To the State University of Maringá, thank you for providing a nurturing environment that has encapsulated the essence of knowledge in our hearts. I am grateful for the opportunities bestowed upon me.

To CAPES/Araucária Foundation, I extend my sincere appreciation for the financial support throughout my research journey in this postgraduate program. Your generosity has made this pursuit possible.

Lastly, to those whose names may not have been mentioned but have contributed to this journey in various ways, I extend my heartfelt thanks. Your support and contributions have not gone unnoticed.

EPIGRAPH

Trust in the Lord with all your heart and lean not on your own understanding; in all your ways submit to him, and he will make your paths straight.

(Proverbs 3:5-6)

Recurrent Neural Networks and Transformer-based models for multi-step prediction of agricultural commodity prices.

ABSTRACT

This dissertation explores the use of Recurrent Neural Networks (RNNs) and Transformer-based models for multi-step prediction of agricultural commodity prices. The study begins with a systematic literature review, providing a comprehensive understanding of the field and identifying gaps in research. Two experimental papers are then presented, focusing on the application of RNNs, specifically Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory Networks (BiLSTM) models, as well as a Transformer-based model for predicting agricultural commodity prices. The commodities of particular interest are soybeans, corn, cattle, and sugar, especially soybeans, the target of the analysis. The study investigates the performance of the Transformer model in forecasting agricultural commodity prices, comparing it with RNNs, both in short and long-term forecasting horizons. The findings highlight the importance of hyperparameter optimization, unbiased model evaluation, and the selection of a suitable forecast horizon that minimizes errors while providing useful information to farmers. Overall, this research contributes to advancing knowledge in the field of RNN models for predicting agricultural commodity prices. It underscores the significance of optimization techniques and the versatility of Transformer models across different domains.

Keywords: Recurrent Neural Networks. Transformer-based Models. Time Series Forecasting. Agricultural Commodities. Hyperparameter Optimization

Redes Neurais Recorrentes e modelos baseados em Transformers para predição *multi-step* de preços de commodities agrícolas

RESUMO

Essa dissertação explora o uso de Redes Neurais Recorrentes (RNRs) e modelos baseados em *Transformers* para a previsão de múltiplos passos dos preços de *commodities* agrícolas. O estudo começa com uma revisão sistemática da literatura, fornecendo uma compreensão abrangente do campo e identificando lacunas na pesquisa. Dois artigos experimentais são então apresentados, focando na aplicação de RNRs, especificamente os modelos *Long Short-Term Memory* (*LSTM*) e *Bidirectional Long Short-Term Memory Networks* (*BiLSTM*), assim como um modelo baseado em *Transformer* para prever preços de *commodities* agrícolas. As *commodities* de interesse particular são soja, milho, gado e açúcar, especialmente a soja, o alvo da análise. O estudo investiga a eficácia do modelo *Transformer* na previsão de preços de commodities agrícolas, comparando-o com as RNRs, tanto em horizontes de previsão de curto prazo quanto de longo prazo. Os resultados destacam a importância da otimização de hiperparâmetros, avaliação imparcial do modelo e seleção de um horizonte de previsão adequado que minimize erros ao mesmo tempo em que fornece informações úteis aos agricultores. No geral, essa pesquisa contribui para avançar o conhecimento no campo dos modelos de RNR para prever preços de *commodities* agrícolas. Ela destaca a importância das técnicas de otimização e a versatilidade dos modelos baseados em *Transformers* em diferentes domínios.

Palavras-chave: Redes Neurais Recorrentes. Modelos baseados em *Transformers*. Previsão de séries temporais. *Commodities* agrícolas. Otimização de hiperparâmetros.

LIST OF FIGURES

LIST OF TABLES

SUMARY

1

INTRODUCTION

Time series forecasting is an area of study that is of great interest in various academic disciplines and finds applications in various sectors such as energy (MANOHAR *et al.*, 2020; QIAO; YANG, 2020), environment (DE MELO *et al.*, 2019; JUNG *et al.*, 2020), industry (CANIZO *et al.*, 2019; WANG *et al.*, 2019), health (CHAMBON *et al.*, 2018; DA SILVA *et al.*, 2021), agriculture (DREES *et al.*, 2021; JUNG *et al.*, 2020), and economics (BALAJI; RAM; NAIR, 2018; YAN; OUYANG, 2018). In the financial context, time series forecasting plays a crucial role for economic agents by providing essential information for efficient analysis and decision-making. This applies not only to exchange rates (ESCUDERO; ALCOCER; PAREDES, 2021; MIDDI; MIDDI, 2022), stock market (GOEL; SINGH, 2022; MOHSIN KABIR *et al.*, 2022), but also to commodity price forecasting (DIVISEKARA; JAYASINGHE; KUMARI, 2021; JU; HUNG; CHEN, 2020).

In the financial market, time series forecasting is fundamental for economic agents, as knowledge about the behavior and future values of these series is essential for efficient analysis and decision-making (MENSI *et al.*, 2013). This also applies to commodities, which are products traded in futures markets, such as stock exchanges, where buyers and sellers establish contracts with pre-determined prices and quantities. The contracts established in these transactions influence fluctuations in commodity prices over time. Historical data of these prices form the so-called time series, which are collections of values that describe the evolution of prices over time (MORETTIN; TOLOI, 2006).

Due to the nature of commodity trading, the prices of these products exhibit complex stochastic processes due to market volatility and macroeconomic influences (VERÍSSIMO; XAVIER, 2014). Forecasting these prices is complex due to factors such as seasonality and volatility. Furthermore, considering that each financial series may have a particular behavior, the difficulty is increased (CARVALHO; PAVAN; HASEGAWA, 2020).

Additionally, agricultural commodities are affected by additional factors such as scale of production in the field, varied demand, and protectionist and incentive policies (CHAMBERS; BAILEY, 1996; PETERSON; TOMEK, 2005; ROBERTS; SCHLENKER, 2013). This scenario makes the analysis and forecasting of agricultural commodity prices even more critical and difficult for producers and professionals in the financial market, who base their decisions on the trend behavior of these price curves.

In this context, a promising approach for forecasting agricultural commodity prices' time series is the use of deep learning (DL) models, such as recurrent neural networks (RNNs), and Transformer-based models. These models are capable of capturing complex temporal dependencies and dealing with the stochastic and volatile characteristics of this data. By applying these techniques, it is expected to obtain more accurate forecasts and contribute to a more informed analysis and more efficient decision-making in the agricultural commodity market.

1.1 OBJECTIVES OF THE DISSERTATION

The primary objective of this dissertation is to employ our proposed methodology for the application of advanced time series forecasting techniques to enhance the accuracy of agricultural commodity price predictions. This study leverages methodologies such as RNRs and Transformer-based models to achieve this goal. To achieve this objective, this dissertation proposes the following specific objectives:

- Conduct a systematic literature review on time series forecasting techniques, focusing specifically on agricultural commodities;
- Compare the performance of RNNs such as Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory (BiLSTM), and Transformer-based models in forecasting agricultural commodity price time series;
- Evaluate the predictive capability of these models, taking into consideration hyperparameter optimization and arbitrariness; and

- Provide insights and recommendations to improve the techniques for forecasting agricultural commodity price time series, considering the peculiarities of this market.

These objectives focus on reviewing existing literature, performing comparative analysis of advanced forecasting techniques, understanding the specific factors that influence agricultural commodity price forecasting, and proposing improvements in these techniques. The study aims to contribute to enhancing the accuracy of predictions in this context, considering the volatility and complexity of agricultural commodity time series data, as well as providing guidance to enhance the forecasting approach in this area.

1.2 RELEVANCE OF THE DISSERTATION

This study bears significant academic relevance within the realm of predicting agricultural commodity price time series. Through a meticulous review of literature and a comparative evaluation of Recurrent Neural Networks (RNNs) and Transformer-based models, this research aims to enrich the precision of forecasting in this intricate domain. The volatile and complex nature of time series data pertaining to agricultural commodity prices necessitates rigorous exploration and scrutiny. This study aligns with the broader academic context, resonating with prior research endeavors in this sphere (KURUMATANI, 2020; PAUL; GARAI, 2021; ZHANG et al., 2020). The findings garnered herein can serve as a pivotal reference point for forthcoming investigations, thereby establishing a robust foundation for the evolution and enhancement of forecasting methodologies applied within this specific field.

Considering these aspects, it can be stated that this dissertation holds academic relevance. Its results have the potential to generate significant impacts in the agricultural sector, providing more accurate and efficient tools for forecasting agricultural commodity prices.

1.3 STRUCTURE OF THE DISSERTATION

Since the 1970s, American researchers have criticized the traditional form of thesis and dissertation writing, highlighting its limitation in reaching and disseminating knowledge within the scientific community. This conventional approach restricts the academic document to a limited audience, such as doctoral and master's students, and examination boards. Typically, these dissertations follow a linear structure with an introduction, literature review, description of methods, presentation of results, and a conclusion. On the other hand, the literature proposes an alternative style called multi-paper, which consists of a set of papers ready for publication (BOOTE; BEILE, 2005; DUKE; BECK, 1999).

This dissertation was written in the multi-paper format as it represents an alternative that promotes the sharing of knowledge. In addition to providing experience and expedience in publication, this format enhances the visibility of the research to a broader scientific-academic community (BOOTE; BEILE, 2005; DUKE; BECK, 1999).

Therefore, it consists of a dissertation composed of papers that meet the requirements established by this resolution. All the papers presented here are part of the candidate's research project. Importantly, article 3 is currently in the peer review phase and has not been published. Therefore, it's essential to acknowledge that the version utilized in this dissertation is based on the submitted version, and potential differences could emerge in the final published iteration. Table 1 shows the relationships between the papers and the specific objectives. The structure of the dissertation is shown in Figure 1 and discussed throughout this chapter.

Table 1 — Relationships between the papers and specific objectives

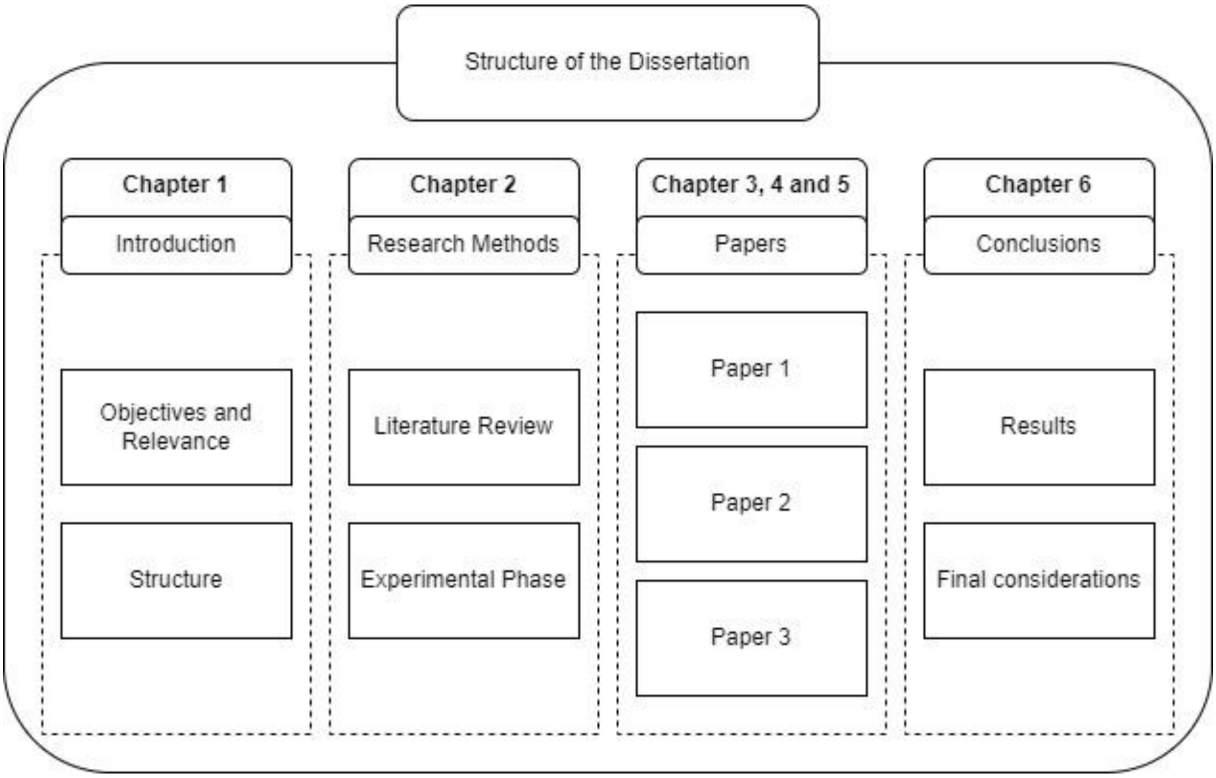
<table>
<tr><th>Dissertation main objective</th><th>Dissertation specific objective</th><th>Papers (P) titles</th><th>Research methods used</th><th>Main results / Contribution</th><th>Submitted to / Published in</th></tr>
<tr><td rowspan="4">Application of time series forecasting techniques with proposed methodology to enhance the accuracy of agricultural commodity price predictions</td><td>Conduct a systematic literature review on time series forecasting techniques, focusing specifically on agricultural commodities</td><td>P1 — A Systematic Review of Recent Literature on Deep Learning Applied to Commodities Time Series Prediction</td><td>Systematic literature review</td><td>• Identification of prevalent prediction techniques, performance metrics, and distribution of commodities in the research landscape.</td><td>SIMPEP 2022 (Published)</td></tr>
<tr><td>Compare the performance of models in forecasting agricultural commodity price time series</td><td>P2 — Avaliação do Desempenho de Técnicas Baseadas em Redes Neurais Recorrentes para Previsão em Séries Temporais de Commodities Agrícolas</td><td>Experimental Article</td><td>• Comparative evaluation of RNNs architectures for forecasting agricultural commodity prices.</td><td>SIMPEP 2022 (Published)</td></tr>
<tr><td>Hyperparameter optimization application</td><td rowspan="2">P3 — Optimized recurrent and Transformer-based models for multi-step forecasting of agricultural commodity prices</td><td rowspan="2">Experimental Article</td><td rowspan="2">• Application of the Transformer model for agricultural time series forecasting;

• Evaluation of the error ratio and the forecast horizon; and

• Application of hyperparameter optimization in time series forecasting techniques.</td><td rowspan="2">Computers and Electronics in Agriculture (Submitted)</td></tr>
<tr><td>Evaluate the performance of forecasting models across varying time horizons for agricultural commodity price predictions</td></tr>
</table>

Source: Author.

Figure 1 — Structure of the dissertation organized into chapters.



Source: Author.

After this introduction, in which the objectives of this dissertation were presented, Chapter 2 describes the research methods and procedures adopted. It also explains the decisions made regarding the stages of dissertation development and how these stages are interconnected. Additionally, the logic and criteria used in the selection of RNNs and Transformer-based models are presented, with a particular focus on agricultural commodity price time series data as the object of analysis. Chapter 2 concludes with the presentation of decisions regarding data collection and analysis procedures adopted in each stage of the literature review and analysis, as well as the procedures for conducting experiments. Each stage of the research project culminates in the production of a research paper as a contribution. In this chapter, the relationships between the papers that compose this dissertation are also discussed, as well as how each paper contributes to the others and to the dissertation as a whole.

Chapter 3 identifies and analyzes, through a systematic literature review, forecasting techniques, data, and performance metrics used in forecasting commodity time series (content of the first paper). It reviews the literature on forecasting techniques, data, and performance metrics used in forecasting commodity time series, contributing to the current understanding of the field. By following systematic research criteria and conducting careful investigation of the selected studies, the work provided a comprehensive overview of existing knowledge.

Chapter 4 evaluated the predictive capability of RNNs in modeling the price series of the three main agricultural commodities exported by Brazil in 2021: soybeans, beef cattle, and sugar (content of the second paper). The RNN models were implemented without undergoing proper hyperparameter optimization. This implies that the models were trained and evaluated using default hyperparameters or arbitrary settings, without conducting a systematic search for the optimal combination. The experiments were conducted without employing an appropriate hyperparameter optimization technique, resulting in outcomes that will subsequently be contrasted with those of experiments involving meticulous optimization. Thus, the study provided an evaluation of the performance of these RNNs models in their basic form, without the benefit of advanced optimization.

Chapter 5 used two RNNs (LSTM and BiLSTM) and a Transformer-based model to predict the time series of corn and soybeans, considering a double cropping system (content of the third paper). The models were optimized using a grid search algorithm, allowing to find the best combination of hyperparameters. The results provided an assessment of the performance and predictive capability of these models for the agricultural commodities in question, based on common evaluation metrics.

Chapter 6 explores the theoretical and practical contributions of this dissertation's development. It highlights findings from literature analysis, experiments using non-optimized DL models, and results from experiments employing grid search algorithms to optimize model hyperparameters. Moreover, it offers valuable insights to enhance predictions, furthering research in the field. Lastly, the chapter concludes the dissertation by summarizing key experiment points in agricultural commodity time series forecasting. It analyzes existing techniques, identifies literature gaps, suggests future research directions, and compares performance across forecasting approaches.

2

RESEARCH METHOD

This chapter details the procedures and strategies employed in this work, organized into three main topics: the systematic literature review, the first part of the experimental work, and the second part of the experimental work.

The systematic literature review was conducted to identify the extent and nature of available studies on the forecasting of commodity time series. Inclusion and exclusion criteria were applied for the selection of papers, limiting the research to the application of models in forecasting these time series. The critical review of the selected papers synthesized the content in this area, providing direction for further improvements.

The initial segment of the experimental phase focused on paper 2, and the subsequent part of this phase was conducted using paper 3.

2.1 LITERATURE REVIEW

The research methodology adopted in this systematic literature review was based on Kitchenham (2004), adapted to the specific topic. This methodology allows for the systematic selection of review elements, declaration of the strategies employed, identification of primary studies, data synthesis, and analysis of evidence to address the research questions.

The objective of this systematic literature review is to identify the most important contributions in the development of Artificial Neural Network (ANN) models for predicting commodity time series. To achieve this, a search protocol was established, starting with the formulation of Research Questions (RQs). The literature search was conducted following the problem formulation, followed by data collection, analysis, and synthesis.

The literature search was performed in two databases, Scopus and ScienceDirect. The keywords were selected to obtain specific results, yet broad enough to not limit the number of results.

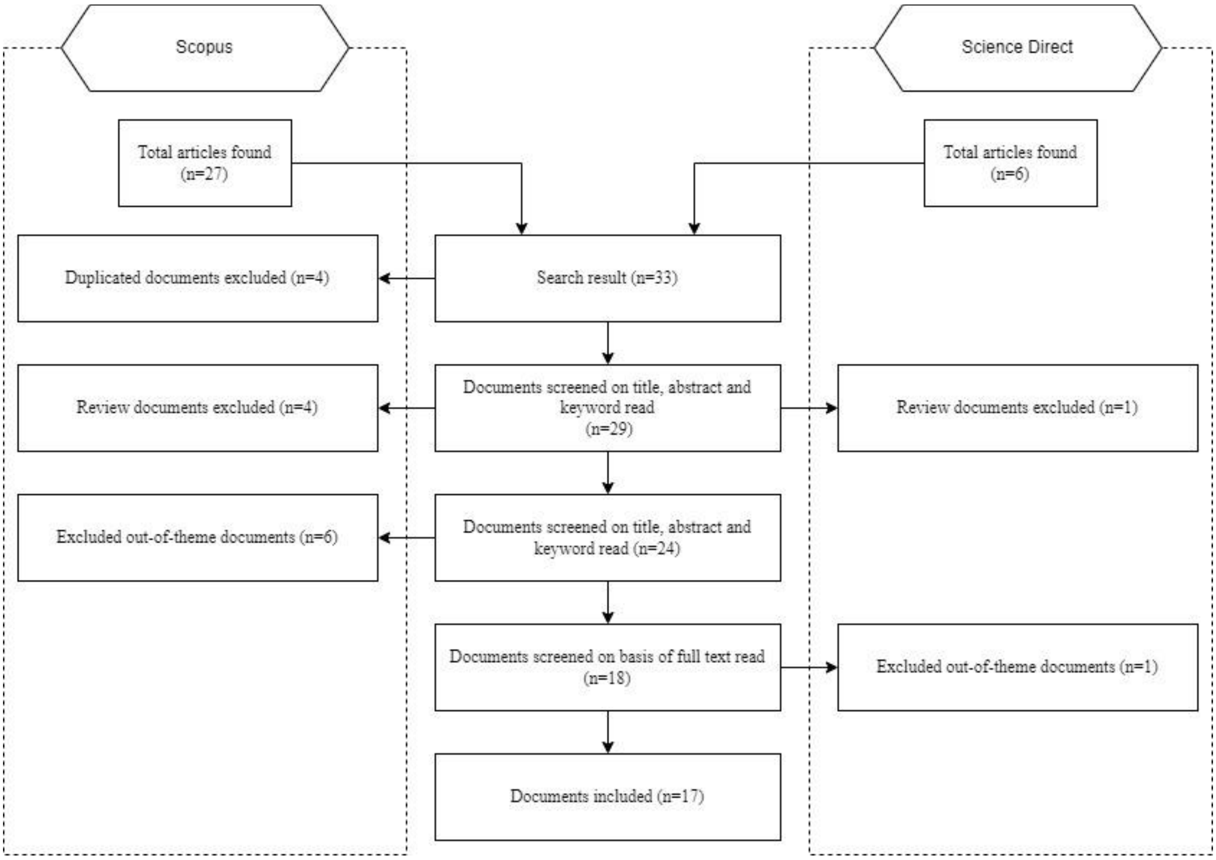
Following a comprehensive search across databases for articles incorporating the terms outlined in the search protocol within their titles, abstracts, and keywords, the outcomes were meticulously examined and synthesized to extract pertinent data for the literature review. The identified articles underwent scrutiny based on exclusion and inclusion criteria, encompassing factors like language and duplicates, ultimately yielding 15 articles, 5 conference papers, and a book chapter.

By conducting a partial reading of the titles, abstracts, and keywords of each publication, 11 documents that did not address topics related to the desired scope or did not analyze commodity data were excluded. During the full-text reading, one article was excluded as it presented a different concept of commodities.

At the end of the exclusion and inclusion process, 17 documents remained, which were used for literature analysis and investigation of research trends on the addressed topic.

A more detailed list of excluded publications by database can be found in Figure 2.

Figure 2 — Flowchart of the document selection process for the systematic review.

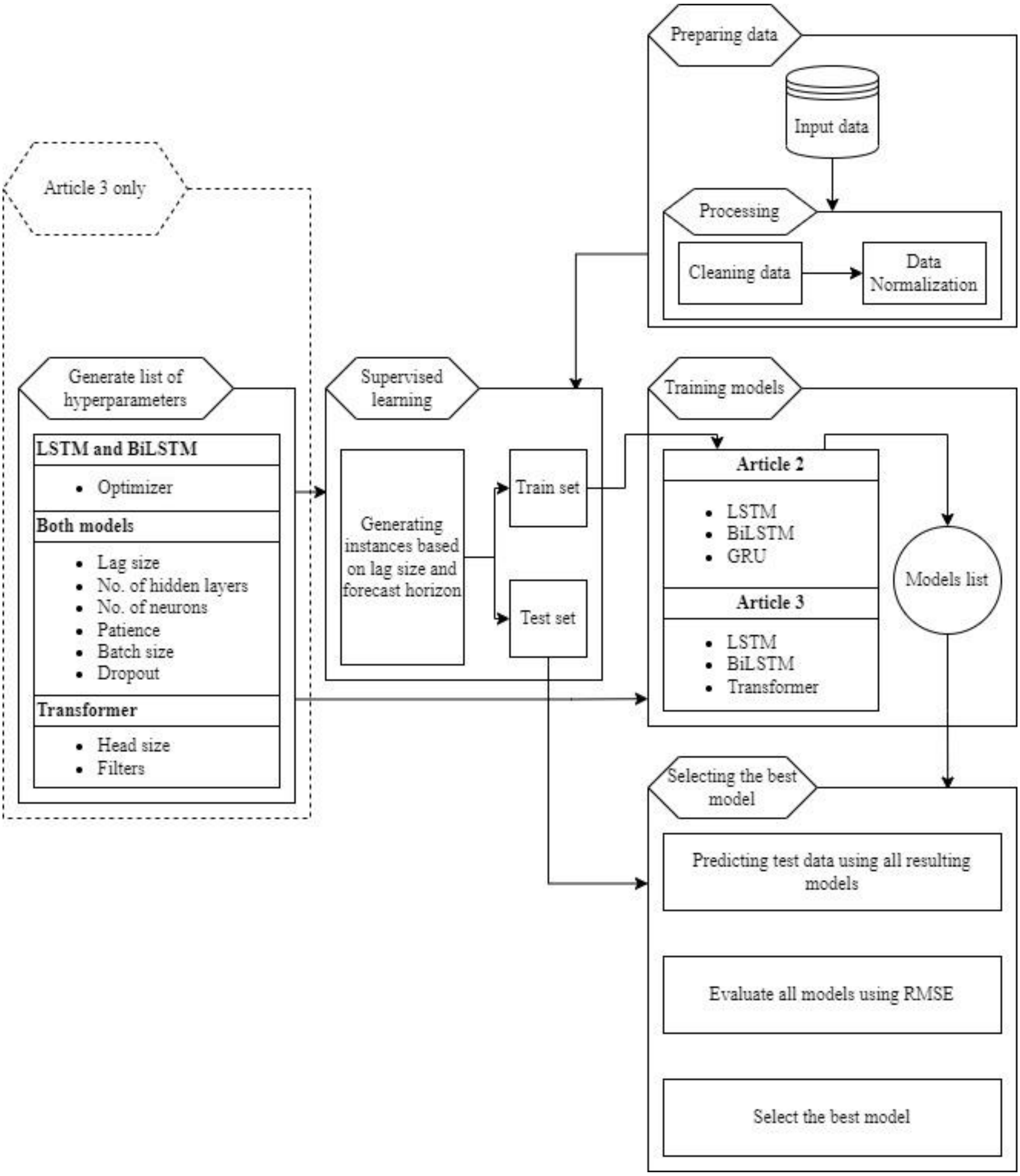


Source: Author.

2.2 EXPERIMENTAL PHASE

In this dissertation, four agricultural commodities were selected: cattle, sugar, corn, and soybean. Cattle and sugar are used only in paper 2, while corn is exclusively used in paper 3, and soybean is used in both. The data were obtained from the Center for Advanced Studies in Applied Economics (CEPEA) website, belonging to the Luiz de Queiroz College of Agriculture (ESALQ) at the University of São Paulo (USP). Each commodity is described in detail in their respective papers where they are used. The following section describes the steps involved in the methodology, summarized in Figure 1.

Figure 3 — The proposed methodology for price forecasting used in papers 2 and 3.



Source: Author (inspired by Abbasimehr, Shabani e Yousefi (2020))

2.2.1Data Preparation

Before using the data for forecasting, it is necessary to preprocess them, including cleaning and normalizing the time series. Cleaning involves removing noise and handling missing values using appropriate techniques. Normalization is performed to facilitate the learning process of the network by transforming the data into a smaller scale. In this study, we used the MinMaxScaler method from Scikit-learn [1]for normalization.

2.2.2Creation of a Hyperparameter Combination Collection

Hyperparameter optimization techniques aim to find the ideal set of user-defined hyperparameters. In this study, we will use the following hyperparameters: lag size, forecast horizon, number of hidden layers, number of neurons, epoch size, patience, batch size, dropout, optimizer, head size, and filters. The last two hyperparameters are exclusive to the Transformer model, while the optimizer is exclusive to the recurrent models, considering the architecture used in this dissertation.

2.2.3Data transformation into supervised learning

In this step, the time series data is restructured into instances with input features and an output feature. These instances are divided into training and testing sets. The forecast horizon and lag size are important in determining the data formatting. The forecast horizon indicates the number of future steps to be predicted, while the lag size represents the number of past steps used as input. In one-step time series forecasting, the model predicts the value of the next step based on historical data. In multi-step forecasting, the model aims to predict multiple future steps. The data is formatted according to these scenarios, where each input-output pair represents a single step or a sequence of past and future steps.

2.2.4Partitioning instances into training and testing Sets

The instances are divided into a training set and a testing set. The training set is used to train the models and create a prediction model. The testing set is used to evaluate the predictive performance of the model using various performance metrics.

[1] https://scikit-learn.org/

3

PAPER 1 — SYSTEMATIC REVIEW: DL FOR COMMODITY TIME SERIES

UMA REVISÃO SISTEMÁTICA DA LITERATURA RECENTE SOBRE APRENDIZADO PROFUNDO APLICADO À PREVISÃO DE SÉRIES TEMPORAIS DE *COMMODITIES*

1º LUCAS GROGENSKI MELOCA - pg403414@uem.br
UNIVERSIDADE ESTADUAL DE MARINGÁ – UEM

2º ADEMIR APARECIDO CONSTANTINO - ademir@din.uem.br
UNIVERSIDADE ESTADUAL DE MARINGÁ – UEM

3º RODRIGO CLEMENTE THOM DE SOUZA - thom@ufpr.br
UNIVERSIDADE FEDERAL DO PARANÁ - UFPR

ÁREA: *3. PESQUISA OPERACIONAL*
SUBÁREA: *3.7 - INTELIGÊNCIA COMPUTACIONAL*

RESUMO: *A PREVISÃO DE SÉRIES TEMPORAIS É UMA FERRAMENTA IMPORTANTE PARA APOIAR O MERCADO DE COMMODITIES, FORNECENDO INFORMAÇÕES SOBRE A DINÂMICA DE PREÇOS E REDUZINDO OS RISCOS DAS NEGOCIAÇÕES. OS MODELOS DE APRENDIZADO PROFUNDO GANHARAM ATUALMENTE ESPAÇO NO CAMPO ACADÊMICO GRAÇAS A SEUS RESULTADOS SUPERIORES AOS MÉTODOS TRADICIONAIS DE APRENDIZAGEM DE MÁQUINAS. NO ENTANTO, POUCAS REVISÕES TRATAM EXCLUSIVAMENTE DE TÉCNICAS DE APRENDIZAGEM PROFUNDA VOLTADAS PARA O MERCADO DE COMMODITIES. NESTE ESTUDO, FIZEMOS UMA REVISÃO SISTEMÁTICA DA LITERATURA PARA IDENTIFICAR TÉCNICAS DE PREVISÃO, DADOS E MÉTRICAS DE DESEMPENHO UTILIZADAS NA PREVISÃO DE SÉRIES TEMPORAIS DE COMMODITIES. COM BASE NOS CRITÉRIOS DE PESQUISA SELECIONADOS, PROCURAMOS RESPONDER AS QUESTÕES DE PESQUISA LEVANTADAS SOBRE O TEMA E FORNECER SUGESTÕES PARA PESQUISAS FUTURAS ATRAVÉS DE UMA INVESTIGAÇÃO CUIDADOSA DOS ESTUDOS SELECIONADOS. DE ACORDO COM NOSSA ANÁLISE, COMMODITIES AGRÍCOLAS E ENERGÉTICAS SÃO OS PRINCIPAIS DADOS DOS ESTUDOS. CONVOLUTIONAL NEURAL NETWORKS (CNN) E LONG SHORTTERM MEMORY NETWORKS (LSTM) SÃO OS ALGORITMOS DE APRENDIZADO PROFUNDO MAIS UTILIZADOS NESTES ESTUDOS. AS MÉTRICAS MAIS FREQUENTES SÃO MEAN ABSOLUTE PERCENT ERROR (MAPE), MEAN ABSOLUTE ERROR (MAE), ROOT MEAN SQUARE ERROR (RMSE) E MEAN SQUARED ERROR (MSE).*

PALAVRAS-CHAVES: REDE NEURAL ARTIFICIAL; APRENDIZADO PROFUNDO; PREVISÃO DE SÉRIE TEMPORAL; *COMMODITY*.

A SYSTEMATIC REVIEW OF RECENT LITERATURE ON DEEP LEARNING APPLIED TO COMMODITIES TIME SERIES PREDICTION

ABSTRACT: *TIME SERIES FORECASTING IS AN IMPORTANT TOOL TO SUPPORT THE COMMODITIES MARKET, PROVIDING INFORMATION ON PRICE DYNAMICS AND REDUCING THE RISKS OF TRADES. DEEP LEARNING (DL) MODELS HAVE CURRENTLY GAINED SPACE IN THE ACADEMIC FIELD THANKS TO THEIR SUPERIOR RESULTS COMPARED TO TRADITIONAL MACHINE LEARNING (ML) METHODS. HOWEVER, FEW REVIEWS DEAL EXCLUSIVELY WITH DEEP LEARNING TECHNIQUES FOR THE COMMODITIES MARKET. IN THIS STUDY, WE PERFORMED A SYSTEMATIC LITERATURE REVIEW (SLR) TO IDENTIFY FORECASTING TECHNIQUES, DATA, AND PERFORMANCE METRICS USED IN COMMODITY TIME SERIES FORECASTING. BASED ON THE SELECTED RESEARCH CRITERIA, WE SOUGHT TO ANSWER THE RESEARCH QUESTIONS RAISED ON THE TOPIC AND PROVIDE SUGGESTIONS FOR FUTURE RESEARCH THROUGH A CAREFUL INVESTIGATION OF THE SELECTED STUDIES. ACCORDING TO OUR ANALYSIS, ENERGY AND AGRICULTURAL COMMODITIES ARE THE MAIN DATA IN THE STUDIES. CONVOLUTIONAL NEURAL NETWORKS (CNN) AND LONG SHORTTERM MEMORY NETWORKS (LSTM) ARE THE DEEP LEARNING ALGORITHMS MOST USED IN THESE STUDIES. THE MOST FREQUENT METRICS ARE MEAN ABSOLUTE PERCENT ERROR (MAPE), MEAN ABSOLUTE ERROR (MAE), ROOT MEAN SQUARE ERROR (RMSE) AND MEAN SQUARED ERROR (MSE).*

KEYWORDS: *ARTIFICIAL NEURAL NETWORK; DEEP LEARNING; TIME SERIES PREDICTION; COMMODITY.*

1. INTRODUCTION

Forecasting time series has been a challenge and a subject of interest in several areas of study, such as energy and fuels (MANOHAR *et al.*, 2020), environment (JUNG *et al.*, 2020), industry (LUO *et al.*, 2019), health (DA SILVA *et al.*, 2021) agriculture (YAO *et al.*, 2021), finance (YAN; OUYANG, 2018) and even regarding the commodities market (SOUZA *et al.*, 2016; SOUZA, 2008).

Commodities are merchandise and comprise a series of characteristics that enable them to be distinguished, however, their operation occurs similarly to raw materials. Its production comes in scale and must present uniform qualities and characteristics, even under storage and stocking conditions, and their differentiation does not occur as a function of its production origin. Generally speaking, these goods can be understood in large groups, but their classification is superficial in the literature. Their global negotiability, over time, forms price series or, also known, time series.

Forecasting the dynamics of time series arising from commodities is a relevant issue for reducing investment risks by producers and buyers, stock market professionals, and all those who are somehow connected to this field. However, commodity data present relatively complex stochastic processes, which lack specific techniques for their prediction. Several studies have been published with computational tools that present a relatively satisfactory performance, such as the case of Deep Learning (DL) and Machine Learning models (ML) (RISTANOSKI; LIU; BAILEY, 2013; ZHANG *et al.*, 2014). Traditional approaches may not be as effective specifically when there is time or space behavior in the system; on the other hand, the DL approach can automatically extract features and present a better result in understanding systems (CLEMENTS; FRANSES; SWANSON, 2004; REN *et al.*, 2021).

This article presents a Systematic Literature Review (SLR) on models and techniques used in forecasting commodity time series. The review describes the modeling techniques, which sets of commodities and performance indicators are used in the articles. The rest of this article is organized as follows: a description of the methods applied to the literature review is presented in section 2. Section 3 will show the results obtained through the content of the selected literature. Finally, section 4 covers discussions and conclusions with implications for future practice and research.

2. METHOD

The research methodology of this SLR was based on Kitchenham (2004), adapted for the present theme. This methodology enables the elements of the review to be systematically selected, always declaring the strategies used, identifying primary studies, synthesizing data, and analyzing the evidence to answer the research questions.

The objective of this SLR is to identify the most important contributions in developing Artificial Neural Networks (ANNs) models for commodities' time series forecasting. Thus, a rigorous protocol was defined, starting with the elaboration of the Research Questions (RQs). The literature search began after formulating the problem, followed by the collection, analysis, and synthesis of data. In planning the problem, the main theme was then conceptualized, and the revision scope was architected through the questions below:

RQ1. What are the emerging Neural Network models for commodities time series forecasting?

RQ2. What are the most used Neural Network models for comparison?

RQ3. What are the main performance measurement metrics used in the literature?

RQ4. What are the main commodities in the literature?

The literature search begins by selecting databases. Thus, two databases were selected for this study, Scopus, and ScienceDirect. No time restriction was set for the publication years of the articles.

The searches encompassed both databases, limited to title, abstract, and keywords. With the search string, [("Neural Network" OR "Deep Learning") AND ("Time Series Forecasting" OR "Time Series Prediction") AND commodities], a total of 33 results were found, with 27 articles from Scopus and the other 6 from Science Direct.

After selecting relevant studies, they underwent some exclusion and inclusion criteria described below. Only documents written in English were accepted. From the total, 4 were excluded for presenting duplicates, leaving 29 documents for the next phases. A partial reading of the papers was performed and another 11 documents were eliminated, many of them for not performing analysis on data from commodities. During the process of reading the full text, it was noted that one of the articles did not match the scope, so it was excluded. As a result, 17 documents remained. They were used to carry out literature analysis and investigation of research trends on the topic addressed.

3. ANALYSIS AND DISCUSSION

This section presents the content analysis and discussion of the literature. All relevant data from the selected studies are extracted and eventually, data were synthesized in response to the research questions.

3.1 COMMODITIES

The expression commodity was diffused in the economic market by constituting the concept of goods produced on a large scale and assuming its characteristics. The literature is superficial and quite diffuse about a standardization regarding the classification of commodities, making an analysis focused on groups difficult. Countries often adopt internal classifications, as is the case of the United States, which adopts the Advanced Technology Products (ATP) and The North American Industry Classification System (NAICS). The Standard International Trade Classification (SITC), proposed by the United Nations, relating to commodities entering foreign trade, is designed to provide the aggregates of commodities needed for economic analysis purposes and to facilitate international comparison of trade data per commodity. The Interagency Task Force of the International Merchandise Trade Statistics recommends SITC for use in analyzing the international trade in goods by interested countries and international organizations.

In this study, we will adopt the SITC for classification, as it is an international standardization of products and commodities. SITC Revision 4 was accepted by the United Nations Statistical Commission at its 37th session in 2006 and is currently in effect. The SITC classification is divided into 10 sections and 2970 basic headings (items). In this article, we will adopt a grouping of the SITC sections into six main categories: “Food, drinks, and tobacco” (Sections 0 and 1 - including live animals), “Raw materials” (Sections 2 and 4), “Energy products” (Section 3), “Chemicals” (Section 5), “Machinery and transport equipment” (Section 7) and “Other manufactured goods” (Sections 6, 8 and 9), as proposed by Statistics Explained, an official Eurostat website featuring statistical topics. Items that could be inserted in sections 1, 4, 5, 7, and 8 were not identified for the present study.

Items belonging to the “Energy products” group appeared 48% of the time, while the other classes, “Raw materials”, “Other manufactured goods” and “Food, drinks and tobacco”, had 7%, 10%, and 35%, respectively.

The distribution of agricultural commodities is heterogeneous, showing great diversity as a whole. These goods range from grains such as soybeans (ABRAHAM *et al.*, 2020; LI *et* *al.*, 2020), corn (JU; HUNG; CHEN, 2020), and canola (SHAHWAN; ODENING, 2007). In addition to these, more varied goods such as animal products (RAUDYS; ZLIOBAITE, 2005; SHAHWAN; ODENING, 2007) and vegetables (ZHIYUAN; VAN KHOA; BOON, 2017) can be found. According to the United Nations, the world population is constantly growing and the projection is that the world will reach 9.8 billion people in 2050, besides this the rapid growth of the emerging countries and the industrialized countries generate a high global demand for food, especially in animal products, fruits, and vegetables (FUKASE; MARTIN, 2020).

In the case of energy commodities, which correspond to 48%, a total of five commodities were identified, natural gas (LIU *et al.*, 2021; PANELLA; LIPARULO; PROIETTI, 2014), coal (PANELLA; LIPARULO; PROIETTI, 2014), fuels (LUO *et al.*, 2019), oil, and electricity(LUO *et al.*, 2019; MEMARZADEH; KEYNIA, 2021; PANELLA; LIPARULO; PROIETTI, 2014; VEGA-MÁRQUEZ *et al.*, 2021). Natural gas, coal, and fuel account for 28% of the "Energy products" category, while electricity and oil together represent 72%.

In summary, crude oil is highlighted for being present in 21% of all commodities and constitutes 43% of energy commodities. There is great commercial movement around oil, currently, approximately 65,000 thousand barrels per day were traded in 2020, according to the US Energy Information Administration (USEIA) and BP Statistical Review of World Energy.

The "Raw materials" category had a raw material of mineral origin and an agricultural one. Groundnut oil (SINGH; MISHRA, 2015) does not fit into the "Food, drinks, and tobacco" class, as it undergoes an additional extraction process, so oils of vegetable or animal origin are included in this new classification. The second item is zinc (LIU *et al.*, 2021), still in its raw state.

The last class belongs to various manufactured commodities, which is the case for steel (TORBAT; KHASHEI; BIJARI, 2018). The other two items belong to the last section, however, they have placed in the "Other manufactured goods" category, as they do not fall under the other SITC classifications. Luo *et al*. (2019) presented a financial commodity originating from ten aerospace companies and two other energy commodities. Usually, precious metals have their classification, but the gold studied by (LIU *et al.*, 2021) is a precious (non-monetary) metal, which receives another classification in the SITC.

3.2 MODELS

Studies on this topic work with mathematical optimizations whose purpose is pattern recognition, and then model or predict future results based on past information. Many authors implement original, adapted, or hybrid models, and part of the literature regarding this will be discussed in this section. The models and data used by the authors and which ones were used for comparison are summarized in **Table 1**.

Table 1. Trend forecasting using commodities time series data.

Paper	Model	Comparisons	Data
Abraham *et al*. (2020)	Non-linear Autoregressive Network with External Input (NARX)	ANN and classical functions	Soybean
Huang and Wu (2018)	Deep (or, Hierarchical) Multiple Kernel Learning (DMKL)	Auto-regressive Integrated Moving Average (ARIMA), Feed-forward and Generalized Neural Network (FFNN and GRNN)	Crude oil
Ju, Hung and Chen (2020)	Convolutional Neural Networks with Market Profiles (CNN-MP)	Long Short-Term Memory Networks (LSTM)	Corn
Li *et al*. (2020)	Multimodal-VAE with LST-Prediction (DP-MAELS)	CNN, RNN, LSTM, ARIMA and Support Vector Machines - Vector Autoregression (SVM-VAR)	Crude oil
Lin and Sun (2020)	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise with Based Multilayer Gated Recurrent Unit Networks (CEEMDAN-ML-GRU)	Naive forecasts, ARIMA, Least Squares Support Vector Machine (LSSVR), ANN, Recurrent Neural Network (RNN), LSTM and some hybrid models.	Crude oil
Liu *et al*. (2021)	Variational Mode Decomposition with ANN (VMD-ANN)	Hybrid models: Support Vector Regression (SVR), ANN, Empirical Mode Decomposition (EMD), CEEMDAN and VMD	Crude oil, natural gas, zinc and gold
Luo *et al*. (2019)	Evolving Recurrent Interval Type-2 Intuitionistic Fuzzy Neural Network (eRIT2IFNN)	Fuzzy Inference System (FIS), Intuitionistic-FIS (IFIS), Interval Type-2 Fuzzy Logic System (IT2FLS) and IT2- Atanassov -IFLS (IT2AIFLS)	Electric energy, fuel and stocks
Memarzadeh and Keynia (2021)	Hybrid model	Newly published papers in this field	Electric energy
Mohapatra *et al*. (2020)	Non-linear Autoregressive Network (NARnet)	GP, SC, Adaptive Neuro-FIS (ANFIS) and Group Method of Data Handling (GMDH)	Crude oil
Panella, Liparulo and Proietti (2014)	Higher-Order Neuro-Fuzzy Inference System (HONFIS)	Autoregressive Moving Average with Generalized Autoregressive Conditional Heteroscedasticity (ARMA-GARCH), GARCH, Radial Basis Function (RBF), Mixture of Gaussian (MoG) and ANFIS	Coal, natural gas, crude oil and electric energy
Raudys and Zliobaite (2005)	MultiLayer-Perceptron (MLP) SingleLayer-Perceptron (SLP)	Previous results	Pork bellies
Shahwan and Odening (2007)	ARIMA-Elman Neural Network (ARIMA-ENN)	Single ENN, ARIMA and ENN hybrids	Hog and canola

Paper	Model	Comparisons	Data
Singh and Mishra (2015)	ANNs	ARIMA	Groundnut oil
Torbat, Khashei and Bijari (2018)	ARIMA-Probabilistic Neural Networks (ARIMA-PNN)	ARIMA, MLP and Fuzzy-ARIMA	Steel
Vega-Márquez *et al*. (2021)	LSTM, CNN e Temporal Convolutional Network (TCN)	MLP, Tree and Random Forests (RF)	Electric energy
Wu *et al*. (2021)	CNN (Text Mining) and VAR	Backpropagation Neural Networks (BPNN), SVM, Multiple Linear Regression (MLR), RNN and LSTM	Crude oil
Zhiyuan, Van Khoa and Boon (2017)	DE-ANN	ARIMA and MLP	Fruits, meat and vegetables

Abraham *et al*. (2020) indicates that ANN is the best approach to predicting soybean harvest area and production, while classical linear function shows more effective efficiency to predict soybean yield.

The model proposed by Huang and Wu (2018) reduced the prediction errors. Abstractly extracting data using a directed acyclic graph is a good strategy for dealing with the complex dynamics of oil prices.

Ju, Hung and Chen (2020) transforms time series data into grayscale images using MP and CNN to image recognition. According to the authors, in the aspect of profitability comparison, the sequential MP performed better than the LSTM networks, but the stacked profile did not.

Li *et al*. (2020) model proved to be a promising model for forecasting multivariate time series, according to the experimental results for the respective data.

Lin and Sun (2020) experimental results show that the proposed method goes beyond other methods. It is important to note that the method proposed by the authors is mainly applied to short-term forecasts, so only daily data is used.

Liu *et al*. (2021) has proposed a model that is superior to other methods compared for his data, so it is a methodology that can be applied to other similar forecasting fields

Luo *et al*. (2019) added fuzzy rules and the strategy of eliminating reduced or invalid rules to the model, improving the efficient implementation of the proposed system. In general, the model achieves better or comparatively good generalization performance compared to similar studies in the literature.

Memarzadeh and Keynia (2021) developed a hybrid forecasting model based on Wavelet Transform, Feature Selection, and time series prediction with LSTM. The results of the proposed method have shown its capability in forecasting electric load and price.

The tests done by Mohapatra *et al*. (2020) on the results of the considered benchmark data set, it is confirmed that NARnet, has shown better performance in predicting population growth, production, and consumption of petroleum oil for six countries.

Panella, Liparulo and Proietti (2014) proposed a variation of ANFIS, the proposed approach was successful in its ability to estimate daily prices over a long-time horizon and to reproduce the same probability distribution for the various series.

Unlike other studies presented during the review, Raudys and Zliobaite (2005) took a different approach in assuming strategies to reduce error, consequently increasing the accuracy of predictions. During the problem, MLP is used to map the data in a low dimension space, and the SLP as a classifier.

According to the Shahwan and Odening (2007), the potential gain from ANN and hybrid models seems to depend on the characteristics of the time series and how their use justifies the relatively high setup costs.

Singh and Mishra (2015) has proposed a feed-forward neural network with only one hidden layer method. The results presented by the authors showed that ANNs performed considerably better than ARIMA models, justified by the chaotic nature of the data, behavior that cannot be fully captured by the linear ARIMA model.

Torbat, Khashei and Bijari (2018) explores the unique advantages of non-linear probabilistic classifiers in situations where data are incomplete. The empirical results indicate that the proposed model is an effective solution in incomplete consumption data.

Vega-Márquez *et al*. (2021) used DL techniques to make an effective forecast of electricity prices. Specifically, LSTM performed best in the case of the quarantine period, CNN for the normal period, and Trees for the fraud period.

Wu *et al*. (2021) has proposed a model that collects news headlines about oil, released by popular websites. CNN model is applied to extract textual information about these headlines. VAR method is used to select the order in which the appropriate CNN output and the oil data history. The text resource and oil data history are injected into various prevision techniques. The results demonstrate that social media information contributes to the forecasting of oil price, production, and consumption. However, the forecasting accuracies of oil inventory are unaffected.

Zhiyuan, Van Khoa and Boon (2017) proposed a hybrid model that took advantage of the optimization algorithm. The learning operation of the model was more advanced compared to the MLP.

3.3 PERFORMANCE METRICS

The most common metrics used to assess the performance of the prediction models within the commodities time series prediction literature will be presented in this section and shown in **Table 2**, and afterward, a discussion is carried out around the topic.

Table 2. Performance metrics used by documents.

Paper	MAE	MSE	MAPE	NMSE	R	RMSE	Theil's U	Others
Abraham *et al*. (2020)	X	X			X			
Huang and Wu (2018)	X	X	X			X		
Ju, Hung and Chen (2020)								X
Li *et al*. (2020)			X		X		X	X
Lin and Sun (2020)	X		X			X		X
Liu *et al*. (2021)			X			X		X
Luo *et al*. (2019)	X		X			X		
Memarzadeh and Keynia (2021)	X		X			X		X
Mohapatra *et al*. (2020)				X				
Panella, Liparulo and Proietti (2014)		X		X				X
Raudys and Zliobaite (2005)								X
Shahwan and Odening (2007)			X				X	
Singh and Mishra (2015)		X	X			X		X
Torbat, Khashei and Bijari (2018)	X					X		
Vega-Márquez *et al*. (2021)	X		X					X
Wu *et al*. (2021)	X		X			X		
Zhiyuan, Van Khoa and Boon (2017)	X	X	X					

The use of these metrics varies for each author, and their use is subjective. One metric is generally not considered to be better than others, but we can still say that some are used more frequently. We can verify the higher frequency of some metrics in the selected literature. The most observed is the Mean Absolute Percent Error (MAPE), followed by the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Squared Error (MSE).

The MAPE metric is most common among documents, appearing in 65% of them. MAPE is generally useful for an overview of the mean error. In statistics, the MAE is a measure of errors between paired observations that express the same phenomenon, very similar to the previous one. The RMSE obtains the error of all calculated values, and unlike MAPE, this calculated error is not a percentage, but the numerical values that represent the dimension of the mean error. Another metric called MSE is one of the main ones used.

It is often noticed that most authors use more than one metric to represent the mean errors obtained. There are some fewer common metrics which serve specific situations and with specific objectives, with it being up to the authors' subjectivity to decide which ones to use. Some documents addressed some unusual metrics, or that had a single appearance, framed in **Table 2** as "Others", as described below. Ju, Hung and Chen (2020) worked with accuracy and profitability to measure the feasibility of their methodology. Liu *et al*. (2021) used the MAPE and RMSE metrics together with Directional Statistics (D stat). Similarly, Vega-Márquez *et al*. (2021) combined several metrics, including MAE, MAPE, Weighted Average Percentage Error (WAPE), Time (T), and Mean Absolute Scaled Error (MASE). Raudys and Zliobaite (2005) chose to only use the Generalization Error. Other techniques were also included in the "Others" class, such as Absolute Mean Error (AME) (SINGH; MISHRA, 2015), Diebold-Mariano (DM) (LIN; SUN, 2020), Noise-to-Signal Ratio (NSR) (PANELLA; LIPARULO; PROIETTI, 2014), Relative Absolute Error (RAE) (LI *et al.*, 2020), Root Relative Squared Error (RSE) (LI *et al.*, 2020) and variance (VAR) (MEMARZADEH; KEYNIA, 2021).

4. CONCLUSION

This study applied a protocol for an RSL, respecting its rigor and replicability, resulting in 17 documents. These papers served as a basis to extract information about prediction techniques, data, and performance metrics used by the authors and grant insights for future research. The analysis in subsection 3.2 answered the first question (RQ1). The authors' choices are fuzzy as they each choose a model to adapt, however, the LSTM and CNN models share common authors who have chosen these neural networks to use in their studies (JU; HUNG; CHEN, 2020; LI *et al.*, 2020; MEMARZADEH; KEYNIA, 2021; VEGA-MÁRQUEZ *et al.*, 2021; WU *et al.*, 2021). The second question, RQ2, can be answered by the same analysis, in which ARIMA is still a baseline in time series forecasting, meanwhile MLP and LSTM have also been used for comparisons.

After the analysis performed in subsection 3.3, we can conclude that the RQ3 response can be summarized in four of the most frequent main metrics in the literature, namely MAPE, MAE, RMSE, and MSE, respectively. Other techniques are used less frequently, as is the case of NMSE, R, and Theil's U. The answer to the last question (analyzed in subsection 3.1), RQ4, showed that the categories "Energy products" and "Food, drinks, and tobacco" have almost 75% of the share among the other classes.

This review found some limitations, as the concept of the commodity is often misused and not consistently inserted in documents. Only articles that contained the word “commodities” were included due to the systematic review, leaving articles that did not present this characteristic out of scope, even though they worked with this type of data. Another limitation found was the lack of standardization in the classification of commodities.

Many models that can still be explored, with great potential in forecasting commodity time series. The models presented in this review could be tested on other types of commodities, or be modified as proposed improvements. Exploring data on agricultural and energetic commodities is an emerging theme in the coming years due to the scenario of food security and sustainable development.

5. ACKNOWLEDGMENTS

We would also like to thank CAPES, Araucária Foundation, and CNPq (grant number 315475/2018-8) for their financial support.

REFERÊNCIAS

ABRAHAM, E. R.; MENDES DOS REIS, J. G.; VENDRAMETTO, O.; OLIVEIRA COSTA NETO, P. L. de; CARLO TOLOI, R.; SOUZA, A. E. de; OLIVEIRA MORAIS, M. de. Time Series Prediction with Artificial Neural Networks: An Analysis Using Brazilian Soybean Production. **Agriculture**, v. 10, n. 10, p. 475, 2020. Disponível em: https://doi.org/10.3390/agriculture10100475

CLEMENTS, M. P.; FRANSES, P. H.; SWANSON, N. R. Forecasting economic and financial time-series with non-linear models. **International Journal of Forecasting**, v. 20, n. 2, p. 169–183, 2004. Disponível em: https://doi.org/10.1016/j.ijforecast.2003.10.004

DA SILVA, D. B.; SCHMIDT, D.; DA COSTA, C. A.; DA ROSA RIGHI, R.; ESKOFIER, B. DeepSigns: A predictive model based on Deep Learning for the early detection of patient health deterioration. **Expert Systems with Applications**, v. 165, p. 113905, 2021. Disponível em: https://doi.org/10.1016/j.eswa.2020.113905

FUKASE, E.; MARTIN, W. Economic growth, convergence, and world food demand and supply. **World Development**, v. 132, p. 104954, 2020. Disponível em: https://doi.org/10.1016/j.worlddev.2020.104954

HUANG, S.-C.; WU, C.-F. Energy Commodity Price Forecasting with Deep Multiple Kernel Learning. **Energies**, v. 11, n. 11, p. 3029, 2018. Disponível em: https://doi.org/10.3390/en11113029

JU, C.-B.; HUNG, M.-C.; CHEN, A.-P. Market Profile with Convolutional Neural Networks: Learning the Structure of Price Activities. *In*: 2020, **2020 International Symposium on Computer, Consumer and Control (IS3C)**. : IEEE, 2020. p. 454–457.Disponível em: https://doi.org/10.1109/IS3C50286.2020.00123

JUNG, D.-H.; KIM, H. S.; JHIN, C.; KIM, H.-J.; PARK, S. H. Time-serial analysis of deep neural network models for prediction of climatic conditions inside a greenhouse. **Computers and Electronics in Agriculture**, v. 173, p. 105402, 2020. Disponível em: https://doi.org/10.1016/j.compag.2020.105402

KITCHENHAM, B. **Procedures for performing systematic reviewsKeele, UK, Keele University**. *[S. l.: s. n.]*.

LI, H.; CUI, Y.; WANG, S.; LIU, J.; QIN, J.; YANG, Y. Multivariate Financial Time-Series Prediction With Certified Robustness. **IEEE Access**, v. 8, p. 109133–109143, 2020. Disponível em: https://doi.org/10.1109/ACCESS.2020.3001287

LIN, H.; SUN, Q. Crude Oil Prices Forecasting: An Approach of Using CEEMDAN-Based Multi-Layer Gated Recurrent Unit Networks. **Energies**, v. 13, n. 7, p. 1543, 2020. Disponível em: https://doi.org/10.3390/en13071543

LIU, W.; WANG, C.; LI, Y.; LIU, Y.; HUANG, K. Ensemble forecasting for product futures prices using variational mode decomposition and artificial neural networks. **Chaos, Solitons & Fractals**, v. 146, p. 110822, 2021. Disponível em: https://doi.org/10.1016/j.chaos.2021.110822

LUO, C.; TAN, C.; WANG, X.; ZHENG, Y. An evolving recurrent interval type-2 intuitionistic fuzzy neural network for online learning and time series prediction. **Applied Soft Computing**, v. 78, p. 150–163, 2019. Disponível em: https://doi.org/10.1016/j.asoc.2019.02.032

MANOHAR, M.; KOLEY, E.; GHOSH, S.; MOHANTA, D. K.; BANSAL, R. C. Spatio-temporal information based protection scheme for PV integrated microgrid under solar irradiance intermittency using deep convolutional neural network. **International Journal of Electrical Power & Energy Systems**, v. 116, p. 105576, 2020. Disponível em: https://doi.org/10.1016/j.ijepes.2019.105576

MEMARZADEH, G.; KEYNIA, F. Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm. **Electric Power Systems Research**, v. 192, p. 106995, 2021. Disponível em: https://doi.org/10.1016/j.epsr.2020.106995

MOHAPATRA, S. K.; KAMILLA, S. K.; SWARNKAR, T.; PATRA, G. R. Forecasting World Petroleum Fuel Crisis by Nonlinear Autoregressive Network. *In*: **Advances in Intelligent Systems and Computing**. *[S. l.: s. n.]*. v. 1030, p. 67–76. Disponível em: https://doi.org/10.1007/978-981-13-9330-3_7

PANELLA, M.; LIPARULO, L.; PROIETTI, A. A higher-order fuzzy neural network for modeling financial time series. *In*: 2014, **2014 International Joint Conference on Neural Networks (IJCNN)**. : IEEE, 2014. p. 3066–3073.Disponível em: https://doi.org/10.1109/IJCNN.2014.6889574

RAUDYS, S.; ZLIOBAITE, I. Prediction of Commodity Prices in Rapidly Changing Environments. *In*: **Lecture Notes in Computer Science**. *[S. l.: s. n.]*. v. 3686, p. 154–163. Disponível em: https://doi.org/10.1007/11551188_17

REN, X.; LI, X.; REN, K.; SONG, J.; XU, Z.; DENG, K.; WANG, X. Deep Learning-Based Weather Prediction: A Survey. **Big Data Research**, v. 23, p. 100178, 2021. Disponível em: https://doi.org/10.1016/j.bdr.2020.100178

RISTANOSKI, G.; LIU, W.; BAILEY, J. A time-dependent enhanced support vector machine for time series regression. *In*: 2013, New York, NY, USA. **Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining**. New York, NY, USA: ACM, 2013. p. 946–954.Disponível em: https://doi.org/10.1145/2487575.2487655

SHAHWAN, T.; ODENING, M. Forecasting Agricultural Commodity Prices using Hybrid Neural Networks. *In*: **Computational Intelligence in Economics and Finance**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. p. 63–74. Disponível em: https://doi.org/10.1007/978-3-540-72821-4_3

SINGH, A.; MISHRA, G. C. Application of box-jenkins method and artificial neural network procedure for time series forecasting of prices. **Statistics in Transition. New Series**, v. 16, n. 1, p. 83–96, 2015. Disponível em: https://doi.org/10.21307/stattrans-2015-005

SOUZA, R. C. T. de; GUEDES FILHO, O.; SANTOS, M. A. R. dos; COELHO, L. D. S. Monthly closing price forecasting of soybean grain in paraná using sarima modeling with intervention. **Journal of Geospatial Modelling**, v. 2, n. 1, p. 27, 2016. Disponível em: https://doi.org/10.22615/2526-1746-jgm-2.1-5887

SOUZA, R. **Previsão de séries temporais utilizando rede neural treinada por filtro de Kalman e evolução diferencial**. 2008. Dissertation (Master in Industrial and Systems Engineering) - Pontifical Catholic University of Paraná, Curitiba, Brazil, 2008.

TORBAT, S.; KHASHEI, M.; BIJARI, M. A hybrid probabilistic fuzzy ARIMA model for consumption forecasting in commodity markets. **Economic Analysis and Policy**, v. 58, p. 22–31, 2018. Disponível em: https://doi.org/10.1016/j.eap.2017.12.003

VEGA-MÁRQUEZ, B.; RUBIO-ESCUDERO, C.; NEPOMUCENO-CHAMORRO, I. A.; ARCOS-VARGAS, Á. Use of Deep Learning Architectures for Day-Ahead Electricity Price Forecasting over Different Time Periods in the Spanish Electricity Market. **Applied Sciences**, v. 11, n. 13, p. 6097, 2021. Disponível em: https://doi.org/10.3390/app11136097

WU, B.; WANG, L.; WANG, S.; ZENG, Y.-R. Forecasting the U.S. oil markets based on social media information during the COVID-19 pandemic. **Energy**, v. 226, p. 120403, 2021. Disponível em: https://doi.org/10.1016/j.energy.2021.120403

YAN, H.; OUYANG, H. Financial Time Series Prediction Based on Deep Learning. **Wireless Personal Communications**, v. 102, n. 2, p. 683–700, 2018. Disponível em: https://doi.org/10.1007/s11277-017-5086-2

YAO, J.-P.; WANG, Z.-Y.; DE OLIVEIRA, R. F.; WANG, Z.-Y.; HUANG, L. A deep learning method for the long-term prediction of plant electrical signals under salt stress to identify salt tolerance. **Computers and Electronics in Agriculture**, v. 190, p. 106435, 2021. Disponível em: https://doi.org/10.1016/j.compag.2021.106435

ZHANG, X.; ZHANG, T.; YOUNG, A. A.; LI, X. Applications and Comparisons of Four Time Series Models in Epidemiological Surveillance Data. **PLoS ONE**, v. 9, n. 2, p. e88075, 2014. Disponível em: https://doi.org/10.1371/journal.pone.0088075

ZHIYUAN, C.; VAN KHOA, L. D.; BOON, L. S. A Hybrid Model of Differential Evolution with Neural Network on Lag Time Selection for Agricultural Price Time Series Forecasting. *In*: **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**. *[S. l.: s. n.]*. v. 10645 LNCS, p. 155–167. Disponível em: https://doi.org/10.1007/978-3-319-70010-6_15

4

PAPER 2 — EVALUATION OF NN IN AGRICULTURAL COMMODITY FORECASTING

AVALIAÇÃO DO DESEMPENHO DE TÉCNICAS BASEADAS EM REDES NEURAIS RECORRENTES PARA PREVISÃO EM SÉRIES TEMPORAIS DE *COMMODITIES* AGRÍCOLAS

1º LUCAS GROGENSKI MELOCA - pg403414@uem.br
UNIVERSIDADE ESTADUAL DE MARINGÁ – UEM

2º ADEMIR APARECIDO CONSTANTINO - ademir@din.uem.br
UNIVERSIDADE ESTADUAL DE MARINGÁ – UEM

3º RODRIGO CLEMENTE THOM DE SOUZA - thom@ufpr.br
UNIVERSIDADE FEDERAL DO PARANÁ - UFPR

ÁREA: *3. PESQUISA OPERACIONAL*
SUBÁREA: *3.7 - INTELIGÊNCIA COMPUTACIONAL*

RESUMO: *A PREVISÃO DOS PREÇOS FUTUROS DE COMMODITIES AGRÍCOLAS FORNECE INFORMAÇÕES IMPORTANTES DE PREÇOS QUE SÃO UTILIZADAS NA TOMADA DE DECISÃO E REDUZ AS INCERTEZAS E RISCOS DO MERCADO AGRÍCOLA. TODAVIA, OS DADOS QUE FORMAM AS SÉRIES TEMPORAIS APRESENTAM PROCESSOS ESTOCÁSTICOS RELATIVAMENTE COMPLEXOS POR CONTA DA VOLATILIDADE, REFLEXO DO DINAMISMO MACROECONÔMICO DO MERCADO. AS SÉRIES TEMPORAIS SÃO AFETADAS POR DIVERSAS VARIÁVEIS EXÓGENAS DESCONHECIDAS E ALEATÓRIAS, TORNANDO-AS ELEMENTOS INCERTOS E COMPLEXOS QUE NECESSITAM DE TÉCNICAS NÃO LINEARES PARA SUA MODELAGEM. TORNA-SE DE GRANDE IMPORTÂNCIA O ESTUDO E A PROPOSIÇÃO DE NOVOS MÉTODOS QUE MELHOREM A CAPACIDADE DE PREVISÃO DE SÉRIES DE PREÇOS DE COMMODITIES DO AGRONEGÓCIO. ESTA PESQUISA TEM COMO OBJETIVO AVALIAR A CAPACIDADE PREDITIVA DE REDES NEURAIS RECORRENTES AO MODELAR AS SÉRIES DE PREÇOS DAS TRÊS PRINCIPAIS COMMODITIES AGRÍCOLAS EXPORTADAS PELO BRASIL NO ANO DE 2021, SENDO ELAS A SOJA, BOI GORDO E O AÇÚCAR. OS EXPERIMENTOS EMPÍRICOS INDICAM QUE O MODELO BILSTM APRESENTOU UM ERRO MÉDIO INFERIOR PARA AS SÉRIES DE PREÇOS DA SOJA E DO BOI GORDO, ENQUANTO QUE O MENOR ERRO MÉDIO PARA A SÉRIE DE PREÇOS DO AÇÚCAR FOI O MODELO LSTM. EM AMBOS OS CASOS É PERCEPTÍVEL QUE O ERRO AUMENTA PROPORCIONALMENTE AO HORIZONTE DE PREVISÃO.*

PALAVRAS-CHAVES: APRENDIZADO PROFUNDO*; REDES NEURAIS RECORRENTES; PREVISÃO DE SÉRIES TEMPORAIS*; *COMMODITIES* AGRÍCOLAS.

PERFORMANCE EVALUATION OF RECURRENT NEURAL NETWORK-BASED TECHNIQUES FOR AGRICULTURAL *COMMODITY* TIME SERIES FORECASTING

ABSTRACT: *FORECASTING FUTURE PRICES OF AGRICULTURAL COMMODITIES PROVIDES IMPORTANT PRICE INFORMATION THAT IS USED IN DECISION MAKING AND REDUCES THE UNCERTAINTIES AND RISKS OF THE AGRICULTURAL MARKET. HOWEVER, THE DATA THAT FORM THE TIME SERIES PRESENT RELATIVELY COMPLEX STOCHASTIC PROCESSES DUE TO VOLATILITY, A REFLECTION OF THE MACROECONOMIC DYNAMISM OF THE MARKET. TIME SERIES ARE AFFECTED BY SEVERAL UNKNOWN EXOGENOUS AND RANDOM VARIABLES, MAKING THEM UNCERTAIN AND COMPLEX ELEMENTS THAT REQUIRE NON-LINEAR TECHNIQUES FOR THEIR MODELING. IT IS OF GREAT IMPORTANCE TO STUDY AND PROPOSE NEW METHODS THAT IMPROVE THE ABILITY TO FORECAST AGRIBUSINESS COMMODITY PRICE SERIES. THIS RESEARCH AIMS TO EVALUATE THE PREDICTIVE ABILITY OF RECURRENT NEURAL NETWORKS WHEN MODELING THE PRICE SERIES OF THE THREE MAIN AGRICULTURAL COMMODITIES EXPORTED BY BRAZIL IN THE YEAR 2021, NAMELY SOYBEANS, FED CATTLE AND CRYSTAL SUGAR. THE EMPIRICAL EXPERIMENTS INDICATE THAT THE BILSTM MODEL PRESENTED A LOWER AVERAGE ERROR FOR THE PRICE SERIES OF SOYBEANS AND FED CATTLE, WHILE THE LOWEST AVERAGE ERROR FOR THE CRYSTAL SUGAR PRICE SERIES WAS THE LSTM MODEL. IN BOTH CASES IT IS NOTICEABLE THAT THE ERROR INCREASES PROPORTIONALLY TO THE FORECAST HORIZON.*

KEYWORDS: *DEEP LEARNING; RECURRENT NEURAL NETWORK; TIME SERIES FORECASTING; AGRICULTURAL COMMODITIES.*

1. INTRODUÇÃO

Segundo as Nações Unidas, a população mundial, em 2050, deverá chegar ao marco de 9,8 bilhões. A atual produção de alimentos seria insuficiente para suprir a necessidade causada pelo contingente populacional, devendo aumentar cerca de 60% para satisfazer a demanda por alimentos (ABRAHAM *et al.*, 2020; ALEXANDRATOS; BRUINSMA, 2012). De acordo com o Ministério da Agricultura Pecuária e Abastecimento (MAPA), cerca de 34% das exportações do agronegócio brasileiro, em 2021, teve a China como principal destino. A China criou estratégias para aumentar a sua segurança alimentar, diversificando seus canais de importação e suas estratégias de aquisição, dado que sua projeção populacional se concentre principalmente na área urbana (65,4%) e não terá condições de suprir o mercado interno de alimentos, abrindo oportunidades e sinais de alerta para o mercado brasileiro de *commodities* (EMBRAPA, 2020). Os principais produtos exportados para o mercado internacional estão a soja, o boi gordo e o açúcar.

De acordo com o MAPA, o complexo de soja participou de 40% das exportações brasileiras em 2021, cerca de USD 47,9 bilhões movimentados. A soja é a principal fonte global de proteína e é amplamente utilizada na criação de animais (JEGADEESAN; YU, 2020).

O constante aumento da renda per capita, causa pressão nas cadeias de fornecimento de alimentos de origem animal (FUKASE; MARTIN, 2020). No ano de 2021 a carne bovina foi o segundo produto mais exportado pelo Brasil, cerca de USD 7,9 bilhões movimentados.

Apesar da preocupação que gira em torno dos grandes produtores agrícolas, como o Brasil e os EUA, que utilizam suas áreas agrícolas para produzir biocombustíveis (RASK; RASK, 2011), a cana de açúcar ocupa apenas 2% de terras aráveis brasileiras, cerca de 8,9 milhões de hectares (DEFANTE; VILPOUX; SAUER, 2018), e a maioria é utilizada para a produção de açúcar em vez de etanol. O açúcar é o terceiro produto mais exportado, participando de 6,6% das exportações do agronegócio brasileiro em 2021.

O equilíbrio dos preços das *commodities* agrícolas é importante para manter a segurança alimentar (LANG; BARLING, 2012), no entanto, o aumento dos preços das *commodities* têm levantado preocupações em relação à diminuição geral na área de colheita agrícola global, no rendimento e na produção (FUGLIE, 2018), dado que esses produtos estão sujeitos aos impactos do clima na produtividade, a diminuição das taxas de rendimento das colheitas e a desaceleração dos investimentos em pesquisa agrícola. Uma boa previsão dos preços futuros de *commodities* agrícolas é crucial para fornecer informações de preços que servem de apoio à tomada de decisão ao mesmo tempo em que reduz a incerteza e os riscos dos mercados agrícolas e programas de seguro agrícola (OUYANG; WEI; WU, 2019; WANG *et al.*, 2017).

Portanto, este estudo visa estimar a série de preços das três principais *commodities* agrícolas exportadas pelo Brasil utilizando Redes Neurais Recorrentes (RNR) para comparar a capacidade preditiva dos modelos. Para isso, as séries de preços dos produtos foram coletadas e estão disponibilizados no endereço eletrônico do Centro de Estudos Avançados em Economia Aplicada (CEPEA).

2. REDES NEURAIS RECORRENTES

Os algoritmos de Aprendizado Profundo (AP), como as Redes Neurais Artificiais (RNA) e as RNR, são capazes de trabalhar com conjuntos de dados maiores e com maior processamento quando comparado com modelos tradicionais ou que não utilizam AP em sua arquitetura.

Apresentadas inicialmente por Rumelhart, Hinton e Williams (1986), as RNR são redes especializadas no processamento de longas sequências de valores e compartilhamento de parâmetros. As RNRs são flexíveis e têm sido utilizadas em diferentes problemas como reconhecimento de fala (GREFF *et al.*, 2017), modelagem de linguagem (BOWMAN *et al.*, 2016), traduções automáticas (CHO *et al.*, 2014), análise de sentimentos (CHEN *et al.*, 2017) e previsão de séries temporais (MEMARZADEH; KEYNIA, 2021; SALES *et al.*, 2018; SOUZA, 2008).

Esse tipo de rede possui um processamento temporal, o qual utiliza as informações anteriores para prever as saídas posteriores, e assim conta com a capacidade de tomar decisões e adaptar memórias baseados nesse aprendizado (CARVALHO *et al.*, 2012; STAUDEMEYER; MORRIS, 2019). Todo esse processo de execução de uma mesma tarefa em cada elemento de uma sequência é o motivo pelo qual essas redes são chamadas recorrentes, o que o possibilita utilizar essas informações em sequências arbitrariamente longas (BRITZ *et al.*, 2017). Exemplos de RNRs encontradas na literatura são *Long Short-Term Memory Network* (*LSTM*), *Bidirecional LSTM* (*BiLSTM*) e a *Gated Recurrent Unit* (*GRU*)

2.1 LONG SHORT-TERM MEMORY NETWORK

LSTM é uma RNR que substitui os neurônios ocultos por unidades *LSTM*. Introduzidas inicialmente por (HOCHREITER; SCHMIDHUBER, 1997) visava resolver o problema de gradiente que as RNR apresentavam ao processar uma sequência muito grande de dados. Esse problema surgiu a partir da procura de um equilíbrio entre uma boa memória de curto prazo e um método de aprendizado que fosse praticável.

O *LSTM* consegue armazenar informações de diversas entradas anteriores (*long-term memory*) enquanto mantém a maior relevância dos estados mais recentes. Isso ocorre pelo uso dos estados *cell state* e *hidden state*, responsáveis pela transferência de informação entre os neurônios. Além das entradas e das saídas, o neurônio é composto, no seu interior, por combinações entre funções de ativação, adições e produtos e, também, por operações, denominados de portões (do inglês, *gates*), sendo eles, *forget gate*, *input gate*, *cell gate* e *output gate*.

2.2 LSTM BIDIRECIONAL

Os modelos *LSTM*s bidirecionais profundos (SCHUSTER; PALIWAL, 1997) são versões modificadas dos modelos de *LSTM* em que se introduziu características das RNR bidirecionais, os quais dois *LSTM*s são aplicados aos dados de entrada.

O fluxo de troca de informações ocorre em ambos os sentidos, sendo assim, um *LSTM* é aplicado sobre a sequência de entrada (ou seja, *forward layer*) e a forma inversa da sequência de entrada é alimentada no modelo *LSTM* (ou seja, *backward layer*). Isto ocorre pela combinação de dois estados ocultos, que permitem que a informação venha tanto da camada de trás quanto da camada de frente.

2.3 GATED RECURRENT UNIT

Cho *et al*. (2014) introduziu o *GRU*, uma versão simplificada da rede *LSTM*. Os portões foram reduzidos, restando apenas o *reset gate* e o *update gate*, a união entre o *forget gate* e o *input gate*. A *update gate* tem o objetivo de regular a quantidade memória passada que será mantida, enquanto a *reset gate* define as possibilidades de combinações entre as memórias anteriores com as novas entradas da célula.

O modelo *GRU* passou a ter o *internal state* e o *hidden state* fundido em um só, o *cell state*. O *new state* é a memória da célula atual que passará a informação para a próxima célula. Em resumo, o *GRU* manteve a habilidade do *LSTM*, mas com uma estrutura interna muito mais simples, portanto, é mais rápido para se treinar, dado que menos cálculos são necessários (GULLI; PAL, 2017; WANG; RAJ; XING, 2017).

3. MATERIAIS E MÉTODOS

Todos os experimentos foram realizados utilizando laptop com processador Intel(R) Core(TM) i5-10210U CPU 1.60GHz 2.11 GHz e *RAM* (do inglês, *Random Access Memory*) instalada de 8,00 GB (utilizável: 7,82 GB). O sistema operacional utilizado foi o Windows 11 versão 21H2. Todos os códigos escritos utilizaram a linguagem de programação *Python* 3.8 através da IDE *Spyder* versão 5.1.5 e algumas bibliotecas: *Sklearn*[2], *Pandas*[3], *Numpy*[4], *Matplotlib*[5] e *Statsmodels*[6].

3.1 CONJUNTO DE DADOS

Neste trabalho, foram selecionadas as séries de preços das três principais *commodities* agrícolas exportadas no ano de 2021, sendo elas a soja, boi gordo e açúcar. As séries de preços do boi gordo e da soja envolvem 4 mil observações diárias, no período de 2006-2022. A série de preço do açúcar é composta por 2 mil observações diárias, no período de 2014-2022. Os dados estão disponíveis no endereço eletrônico do CEPEA, pertencente ao Departamento de Economia, Administração e Sociologia da Escola Superior de Agricultura “Luiz de Queiroz” (ESALQ), unidade da Universidade de São Paulo (USP).

A amostra da série de preço da soja inicia no dia 13 de março de 2006, do boi gordo dia 23 de fevereiro de 2006 e para a série do açúcar dia 26 de fevereiro de 2014. Todas as amostras das séries terminam no dia 30 de março de 2022. Os valores são coletados nos dias úteis de cada localidade e, para efeito dessa metodologia, considera-se dia útil o dia em que há negociação de derivativos na BM&FBOVESPA.

O indicador de preço CEPEA/ESALQ, é uma média aritmética dos preços para cada produto e utilizado como valor de referência por agentes e usuários que de alguma forma operam o mercado. Neste trabalho será utilizado o Indicador a Soja ESALQ/BM&FBOVESPA - Paranaguá, Indicador de Preços do Boi Gordo CEPEA/B3 e Indicador de Açúcar Cristal ESALQ/BVMF – Santos.

3.1.1Soja

O Indicador da Soja ESALQ/BM&FBOVESPA - Paranaguá é divulgado diariamente pelo CEPEA desde março de 2006 e tem a moeda/unidade de medida em dólares americanos por saca

[2] https://scikit-learn.org/
[3] https://pandas.pydata.org/
[4] https://numpy.org/
[5] https://matplotlib.org/
[6] https://www.statsmodels.org/

de 60 kg. O indicador é determinado pela média aritmética dos preços do produto posto no porto de Paranaguá, estado do Paraná, nas condições *DAP* (*Delivered at Place*) no pátio ou *FAS* (*Free Alongside Ship*) em armazéns/silos que efetuem carregamento de navios via corredor de exportação no porto de Paranaguá, livre de quaisquer encargos.

O indicador representa soja brasileira em grão a granel tipo exportação em concordância com a descrição do produto negociado no Contrato Futuro de Soja com Liquidação Financeira da BM&FBOVESPA.

3.1.2 Boi Gordo

O Indicador de Preços do Boi Gordo CEPEA/B3 é uma média diária ponderada de preços à vista do boi gordo no Estado de São Paulo. O preço é o valor em reais (posteriormente convertido e disponibilizado em dólares americanos) acertados entre comprador e vendedor, informado por Agente Colaborador do CEPEA, cotado por arroba de carcaça de boi gordo, livre de ICMS, para retirar em fazendas das regiões de origem onde está localizado o animal transacionado. O Estado de São Paulo está dividido pelo CEPEA em cinco regiões: Araçatuba, Presidente Prudente, São José do Rio Preto, Bauru e Vale do Paraíba. A divisão considera a representatividade do rebanho de bovinos divulgada pelo IBGE e estudo dos mercados regionais.

O Boi Gordo considerado para o Indicador são bovinos machos, com 16 arrobas líquidas ou mais de carcaça e idade máxima de 42 meses, de acordo com as especificações do contrato futuro de boi gordo da B3.

3.1.3 Açúcar

O Indicador de Açúcar Cristal ESALQ/BVMF – Santos é divulgado diariamente desde janeiro de 2013 e tem a moeda/unidade de medida em reais (posteriormente convertido e disponibilizado em dólares americanos) por saca de 50 kg. O indicador considera as vendas externas e visa captar movimentos no mercado provocados pela exportação do açúcar cristal, a qual envolve riscos diferenciados daqueles relativos às exportações do açúcar bruto.

O referencial para o preço são negócios de venda efetivados por unidades produtoras de açúcar do Centro-Sul cujo produto é exportado pelo porto de Santos para exportação na condição à vista ou em valor presente, sem impostos.

3.2 PREPARAÇÃO DOS CONJUNTOS DE DADOS

As séries temporais das *commodities* agrícolas selecionadas foram segmentadas em dois conjuntos, 80% voltados para o treinamento dos modelos e os 20% restantes para o teste. Para o treinamento da rede existe a necessidade de normalização dos valores dos dados de treinamento para uma faixa de valores, neste caso, entre 0 e 1. Essa etapa é importante em situações em que os valores apresentem uma magnitude muito grande pois alguns modelos podem ser afetados de tal maneira que ocorre a inviabilização do mesmo.

3.3 MODELOS DE PREVISÃO

Neste trabalho foram adotadas três arquiteturas neurais, *LSTM*, *BiLSTM* e *GRU*. Para a execução dos modelos de RNAs foi utilizado o pacote *Keras*[7], uma biblioteca que fornece blocos de construção altamente poderosos e abstratos para construir redes de AP e de aprendizado de máquina (do inglês, *Machine Learning* — *ML*), em que estes blocos utilizam o *TensorFlow*[8]. É uma ótima ferramenta para construir protótipos de ideias de forma ágil e possibilita a utilização de diversos otimizadores, sendo que o método escolhido foi o *Adam*[9] (KINGMA; BA, 2014).

Ambos os modelos têm 64 neurônios na camada de entrada, uma camada oculta incluindo 64 neurônios e 1 neurônio na camada de saída. Os modelos foram configurados com 100 épocas (iterações) para a finalidade do processo de estimação e tamanho de lote de 16.

Para evitar o *overfitting*, um problema comum quando há treinamento excessivo do modelo, foi adicionado aos três modelos uma parada antecipada no treinamento quando ocorrer perda na validação (SRIVASTAVA *et al.*, 2014). Caso o modelo não apresentar uma melhora na perda de validação, o treinamento é interrompido, neste caso, após 10 épocas (*patience = 10*). A função *dropout* descartou aleatoriamente 20% das unidades da rede.

As métricas dos erros de previsão ou acurácia da previsão são muito importantes para resolver problemas práticos, visto que, tipicamente são utilizadas para estimar a qualidade da predição dos métodos de previsão e servir de ferramenta para a tomada de decisão para apontar os modelos mais indicados, quando há comparações com mais de uma técnica (SHCHERBAKOV *et al.*, 2013). De acordo com uma revisão da literatura feita pelos autores, as técnicas mais

---

7 https://keras.io/
8 https://www.tensorflow.org/
9 Método gradiente descendente estocástico baseado em estimação adaptativa de momentos de primeira e segunda ordem.

utilizadas para mensuração do erro de técnicas de previsão utilizando os dados de séries temporais baseadas em *commodities* são *Mean Absolute Percent Error* (*MAPE*), *Mean Absolute Error* (*MAE*), *Root Mean Square Error* (*RMSE*) e *Mean Squared Error* (*MSE*), utilizadas neste trabalho.

4. RESULTADOS OBTIDOS

Nesta seção são apresentados os resultados experimentais dos modelos *GRU*, *LSTM* e *BiLSTM* na previsão das três principais *commodities* agrícolas exportadas pelo Brasil no ano de 2021, sendo elas a soja, boi gordo e o açúcar bruto. Na TABELA 1, os erros médios *MAE*, *RMSE*, *MSE* e *MAPE* são apresentados para os respectivos modelos e dados, destacando em negrito os menores erros para cada série.

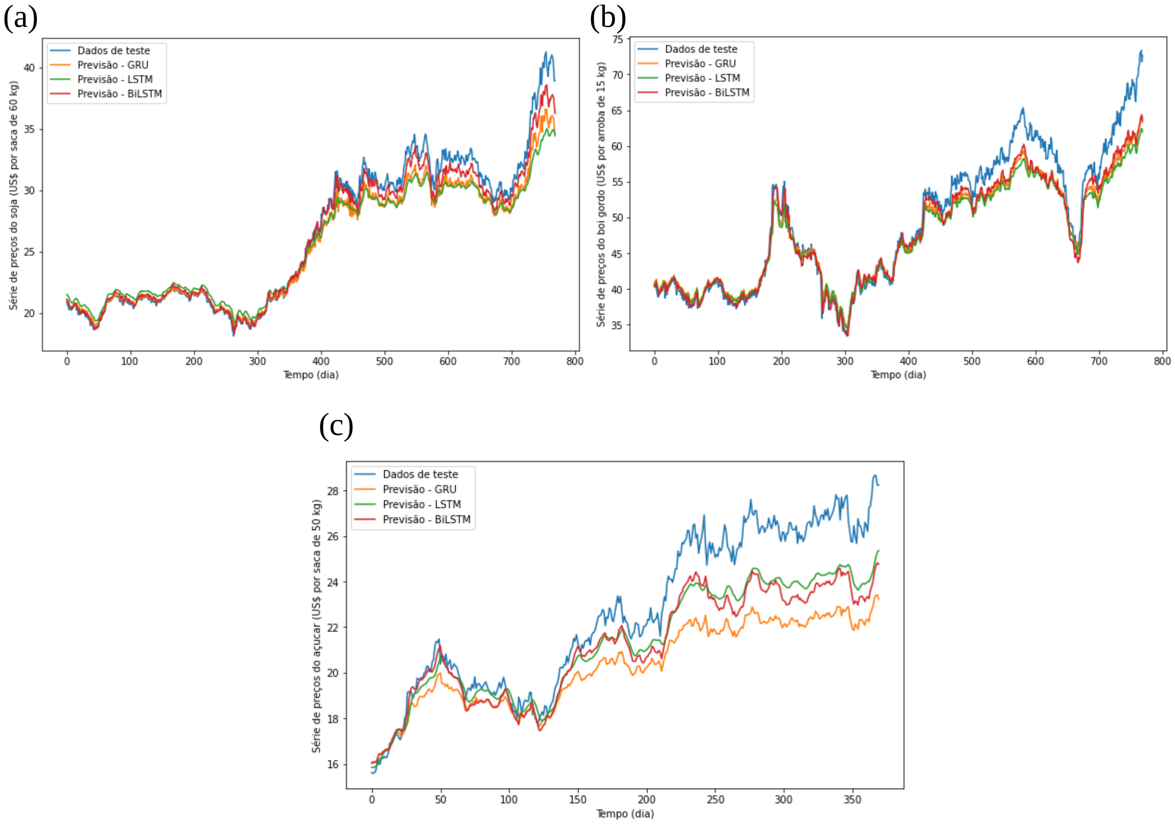
TABELA 1 — Erros apresentados pelas métricas MAE, RMSE, MSE e MAPE para os modelos e dados propostos.

	MAE	*RMSE*	*MSE*	*MAPE*
		Soja		
GRU	1,0510	1,5026	2,2577	3,3900
LSTM	1,3217	1,8002	3,2406	4,4535
BiLSTM	***0,5852***	***0,8433***	***0,7111***	***1,9604***
		Boi gordo		
GRU	2,0371	2,8297	8,0070	3,6928
LSTM	2,3283	3,2786	10,7490	4,1751
BiLSTM	***1,8165***	***2,5929***	***6,7231***	***3,2823***
		Açúcar		
GRU	2,2822	2,7922	7,7961	9,2454
LSTM	***1,3284***	***1,6628***	***2,7650***	***5,3765***
BiLSTM	1,4855	1,8830	3,5455	6,0036

Fonte: Autor (2022).

O modelo *BiLSTM* apresentou um valor de erro menor em relação às outras metodologias ao prever as séries de preços da soja e do boi gordo, enquanto que o modelo *LSTM* obteve um melhor desempenho na previsão do preço do açúcar. Na FIGURA 1, são representados os valores reais e os previstos pelos modelos para as séries da soja (a), boi gordo (b) e açúcar (c), respectivamente. Observa-se um padrão bastante semelhante entre a linha dos valores preditos e valores verdadeiros da série.

FIGURA 1 — Resultados gráficos dos valores previstos pelos modelos GRU, LSTM e BiLSTM para a série de preços da (a) soja, (b) boi gordo e (c) açúcar.



Fonte: Autor (2022).

5. CONSIDERAÇÕES FINAIS

Neste trabalho foram testadas três arquiteturas neurais recorrentes para compará-las e apontar quais apresentaram um melhor resultado em estimar os preços futuros das *commodities* agrícolas mais exportadas pelo Brasil. Com isso, este estudo empírico teve como dados de entrada as séries de preço de três diferentes produtos: soja, boi gordo e açúcar. Os modelos não-lineares foram baseados em RNRs, entre os modelos não-lineares, temos o *GRU*, *LSTM*, e o *BiLSTM*.

As técnicas de RNA selecionadas para fazer parte deste estudo são utilizadas em estudos recentes na previsão de séries temporais financeiras, e apresentarem resultados bem sucedidos por conta de suas características não-lineares.

Os resultados obtidos a partir dos experimentos realizados neste trabalho indicam que o modelo *BiLSTM* retornou resultados mais satisfatórios na previsão das séries de preços da soja e do boi gordo, enquanto a técnica *LSTM* obteve o menor erro ao prever os dados de teste da série de preço do açúcar. Em todos os casos é possível notar que o erro aumenta proporcionalmente ao horizonte de previsão.

Futuros trabalhos podem melhorar a robustez dos experimentos realizados neste estudo. Uma das formas para isso é encontrar hiperparâmetros ideais que elevam a capacidade preditiva dos modelos, entretanto, identificar qual é a melhor configuração dos hiperparâmetros de uma rede pode ser um trabalho repetitivo e demanda muito recurso, tempo e esforço. Existem maneiras eficientes de alcançar resultados satisfatórios, como a aplicação de *Grid Search* ou *Smart Search*.

A técnica *Grid Search* tem por objetivo ajustar os hiperparâmetros na melhor configuração possível, substituindo o trabalho manual e exaustivo de testes, visto que esse trabalho é realizado utilizando estimativas realizadas computacionalmente. O *Smart Search* compreende os algoritmos de busca meta-heurísticas, são mais complexos e requerem mais recursos, no entanto os resultados tendem a ser melhores, colaborando para uma experimentação mais homogênea e justa entre os modelos, e assim, apresentando de fato os melhores modelos para os determinados conjuntos de dados.

REFERÊNCIAS

ABRAHAM, E. R.; MENDES DOS REIS, J. G.; VENDRAMETTO, O.; OLIVEIRA COSTA NETO, P. L. de; CARLO TOLOI, R.; SOUZA, A. E. de; OLIVEIRA MORAIS, M. de. Time Series Prediction with Artificial Neural Networks: An Analysis Using Brazilian Soybean Production. **Agriculture**, v. 10, n. 10, p. 475, 2020. Disponível em: https://doi.org/10.3390/agriculture10100475

ALEXANDRATOS, N.; BRUINSMA, J. World agriculture towards 2030/2050: the 2012 revision. 2012-06-11, ESA Working Papers 12-03, 2012. Disponível em: https://doi.org/10.22004/ag.econ.288998

BOWMAN, S. R.; VILNIS, L.; VINYALS, O.; DAI, A.; JOZEFOWICZ, R.; BENGIO, S. Generating Sentences from a Continuous Space. *In*: 2016, Berlin, Germany. **Proceedings of The 20th {SIGNLL} Conference on Computational Natural Language Learning**. Berlin, Germany: Association for Computational Linguistics, 2016. p. 10–21.Disponível em: https://doi.org/10.18653/v1/K16-1002

BRITZ, D.; GOLDIE, A.; LUONG, M.-T.; LE, Q. **Massive Exploration of Neural Machine Translation Architectures**. *[S. l.]*: arXiv, 2017. Disponível em: https://doi.org/10.48550/ARXIV.1703.03906

CARVALHO, E.; FERREIRA, B.; FERREIRA, M.; PEREIRA, G.; UEYAMA, J.; PESSIN, G. Uso de Redes Neurais Recorrentes para Localização de Agentes em Ambientes Internos. *In*: 2012, **Symposium on Knowledge Discovery, Mining and Learning**. *[S. l.: s. n.]*

CHEN, P.; SUN, Z.; BING, L.; YANG, W. Recurrent Attention Network on Memory for Aspect Sentiment Analysis. *In*: 2017, Copenhagen, Denmark. **Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing**. Copenhagen, Denmark: Association for Computational Linguistics, 2017. p. 452–461.Disponível em: https://doi.org/10.18653/v1/D17-1047

CHO, K.; VAN MERRIENBOER, B.; GULCEHRE, C.; BAHDANAU, D.; BOUGARES, F.; SCHWENK, H.; BENGIO, Y. **Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation**. *[S. l.]*: arXiv, 2014. Disponível em: https://doi.org/10.48550/ARXIV.1406.1078

DEFANTE, L.; VILPOUX, O.; SAUER, L. Rapid expansion of sugarcane crop for biofuels and influence on food production in the first producing region of Brazil. **Food Policy**, v. 79, 2018. Disponível em: https://doi.org/10.1016/j.foodpol.2018.06.005

EMBRAPA. **Em busca de segurança alimentar, China se prepara para a fase pós- Covid-19 e pode influenciar o agronegócio brasileiro**. *[S. l.: s. n.]*. Disponível em: https://www.embrapa.br/busca-de-noticias/-/noticia/53150898/em-busca-de-seguranca-alimentar-china-se-prepara-para-a-fase-pos--covid--19-e-pode-influenciar-o-agronegocio-brasileiro.

FUGLIE, K. Is agricultural productivity slowing? **Global Food Security**, v. 17, 2018. Disponível em: https://doi.org/10.1016/j.gfs.2018.05.001

FUKASE, E.; MARTIN, W. Economic growth, convergence, and world food demand and supply. **World Development**, v. 132, p. 104954, 2020. Disponível em: https://doi.org/https://doi.org/10.1016/j.worlddev.2020.104954

GREFF, K.; SRIVASTAVA, R. K.; KOUTNÍK, J.; STEUNEBRINK, B. R.; SCHMIDHUBER, J. LSTM: A Search Space Odyssey. **IEEE Transactions on Neural Networks and Learning Systems**, v. 28, n. 10, p. 2222–2232, 2017. Disponível em: https://doi.org/10.1109/TNNLS.2016.2582924

GULLI, A.; PAL, S. **Deep Learning with Keras**. *[S. l.]*: Packt Publishing, 2017.

HOCHREITER, S.; SCHMIDHUBER, J. Long Short-Term Memory. **Neural Computation**, v. 9, n. 8, p. 1735–1780, 1997. Disponível em: https://doi.org/10.1162/neco.1997.9.8.1735

JEGADEESAN, S.; YU, K. Food Grade Soybean Breeding, Current Status and Future Directions. *In*: *[S. l.: s. n.]*. Disponível em: https://doi.org/10.5772/intechopen.92069

KINGMA, D. P.; BA, J. **Adam: A Method for Stochastic Optimization**. *[S. l.]*: arXiv, 2014. Disponível em: https://doi.org/10.48550/ARXIV.1412.6980

LANG, T.; BARLING, D. Food security and food sustainability: reformulating the debate. **The Geographical Journal**, v. 178, n. 4, p. 313–326, 2012. Disponível em: https://doi.org/https://doi.org/10.1111/j.1475-4959.2012.00480.x

MEMARZADEH, G.; KEYNIA, F. Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm. **Electric Power Systems Research**, v. 192, p. 106995, 2021. Disponível em: https://doi.org/https://doi.org/10.1016/j.epsr.2020.106995

OUYANG, H.; WEI, X.; WU, Q. Agricultural commodity futures prices prediction via long- and short-term time series network. **Journal of Applied Economics**, v. 22, n. 1, p. 468–483, 2019. Disponível em: https://doi.org/10.1080/15140326.2019.1668664

RASK, K. J.; RASK, N. Economic development and food production–consumption balance: A growing global challenge. **Food Policy**, v. 36, n. 2, p. 186–196, 2011. Disponível em: https://doi.org/https://doi.org/10.1016/j.foodpol.2010.11.015

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, v. 323, n. 6088, p. 533–536, 1986. Disponível em: https://doi.org/10.1038/323533a0

SALES, J. P. de; TAVARES, M. S.; CUNHA, C. E. S. C.; SOUZA, R. B. de; FIRMINO, P. R. A. Previsão do desmatamento na amazônia legal via redes neurais artificiais. *In*: 2018, Bauru - SP. **XXV Simpósio de Engenharia de Produção**. Bauru - SP: UNESP, 2018. p. 13.

SCHUSTER, M.; PALIWAL, K. K. Bidirectional recurrent neural networks. **IEEE Transactions on Signal Processing**, v. 45, n. 11, p. 2673–2681, 1997. Disponível em: https://doi.org/10.1109/78.650093

SHCHERBAKOV, M. V.; BREBELS, A.; SHCHERBAKOVA, N. L.; TYUKOV, A. P.; JANOVSKY, T. A.; KAMAEV, V. A. evich. A survey of forecast error measures. **World**

Applied Sciences Journal, v. 24, n. 24, p. 171–176, 2013. Disponível em: https://doi.org/10.5829/idosi.wasj.2013.24.itmies.80032

SOUZA, R. **Previsão de séries temporais utilizando rede neural treinada por filtro de Kalman e evolução diferencial**. 2008. - Pontifícia Universidade Católica do Paraná (PUC), Curitiba, Brasil, 2008.

SRIVASTAVA, N.; HINTON, G.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. Dropout: a simple way to prevent neural networks from overfitting. **The journal of machine learning research**, v. 15, n. 1, p. 1929–1958, 2014.

STAUDEMEYER, R. C.; MORRIS, E. R. **Understanding LSTM -- a tutorial into Long Short-Term Memory Recurrent Neural Networks**. *[S. l.]*: arXiv, 2019. Disponível em: https://doi.org/10.48550/ARXIV.1909.09586

WANG, D.; YUE, C.; WEI, S.; LV, J. Performance analysis of four decomposition- ensemble models for one-day-ahead agricultural commodity futures price forecasting. **Algorithms**, v. 10, n. 3, 2017. Disponível em: https://doi.org/10.3390/a10030108

WANG, H.; RAJ, B.; XING, E. P. On the Origin of Deep Learning. **CoRR**, v. abs/1702.0, 2017.

5

PAPER 3 — OPTIMIZED MODELS FOR MULTI-STEP AGRICULTURAL COMMODITY FORECASTING

Optimized recurrent and Transformer-based models for multi-step forecasting of agricultural commodity prices

Lucas Grogenski Meloca[a,*], Rodrigo Clemente Thom de Souza[a,b,c], Ademir Aparecido Constantino[a,b]

[a]Production Engineering Graduate Program (PGP), State University of Maringa (UEM), Maringa, PR, Brazil
[b]Computer Science Graduate Program (PCC), State University of Maringa (UEM), Maringa, PR, Brazil
[c]Advanced Campus in Jandaia do Sul, Federal University of Parana (UFPR), Jandaia do Sul, PR, Brazil

*Corresponding Author
Email addresses: lucas.grogenskimeloca@gmail.com (Meloca, L. G.), thom@ufpr.br (Thom de Souza, R. C.), ademir@din.uem.br (Constantino, A. A.)
Full postal address: State University of Maringa, Av. Colombo, 5790 - Jd. Universitário
CEP 87020-900 - Maringá – PR – Brazil

Abstract
Farmers, particularly those in challenging climates, prioritize profitability and seek crops that can optimize their financial gains. Making informed decisions in agriculture is crucial to achieve satisfactory outcomes. Access to reliable information plays a vital role in formulating effective plans. Utilizing forecasting tools for agricultural commodity prices becomes essential in supporting decision-making processes, reducing uncertainty, and mitigating risks in agricultural markets and insurance programs. This study aims to employ two recurrent neural networks, long short-term memory (LSTM) and bidirectional-LSTM (BiLSTM), as well as a transformer-based model. These models will be optimized using a grid search algorithm to forecast the time series of two agricultural commodities, corn and soybean, cultivated during the dual cropping. The forecast horizon for this study extends up to 13 weeks. The Transformer model outperforms existing forecasting models, especially for soybean data, demonstrating its effectiveness in agricultural time series forecasting. It combines a convolutional neural network (CNN) with the Transformer model to provide accurate forecasts, benefiting rural producers by reducing risks and improving decision-making. Although the Transformer model may not

outperform other models in corn forecasting, its versatility and adaptability across domains are noteworthy.

Keywords: deep learning; recording neural networks; transformer; time series forecasting; agricultural commodities.

1. INTRODUCTION

Brazil currently ranks among the world's largest exporters of corn and holds the position as the primary producer and exporter of soybeans globally. According to Conab (National Supply Company) (Gomes et al., 2022), the volume of grain production for the 2021/22 harvest totals 271.2 million tons, marking a 5.6% increase compared to the previous harvest. Despite adverse weather conditions in certain producing regions, the current harvest is projected to reach record-breaking volumes. The significant growth in Brazilian production has played a crucial role in meeting the rising global demands.

The growth of Brazil as a major producer and exporter can be attributed to the technological advancements implemented over the past few decades (da Silveira et al., 2021; Sharma et al., 2022). In 2006, a team of three researchers, consisting of two Brazilians and one American, was awarded the World Food Prize for their work in correcting the acidic soils of the Brazilian Cerrado region. The region receives ample rainfall throughout the growing season, enabling the expansion of agricultural frontiers (da SILVA, 2012). Around the year 2000, there was a rapid increase in the planted area and production of soybeans and second-crop corn. During the early 2000s, researchers and farmers discovered the potential profitability of double cropping, where corn is grown after soybeans within the same crop year. This practice gained attention from farmers due to its profitability. Additionally, the government's efforts to develop corn and soybean cultivars adapted to the region's climate have resulted in significant yield increases, almost doubling over the past decade.

Although these technologies have opened new horizons for farmers, it necessitated making new choices. By selecting corn as the first crop, the opportunity to double the harvest is lost since corn takes longer to mature compared to soybeans (approximately 30 days longer, on average). This fact has led to a change in the decision-making process for Brazilian farmers. The decision-making process for Brazilian farmers resembled that of US farmers before the possibility of two annual harvests, as they based their decisions on the crops that would yield the highest profitability, taking logistical constraints into account. While other important factors need consideration, studies such as Miao et al. (2016) emphasize profitability as the primary determinant. With the possibility of growing two crops per year, Brazilian farmers in the Cerrado region become less price-sensitive because, irrespective of the price ratio, the best combination for double cropping is soybeans followed by corn.

The majority of the first corn crop is now grown in the southern region, which has a different dynamic in the Brazilian corn market. The climate in this region is not conducive to two crops per year (Nóia Júnior and Sentelhas, 2019). Additionally, this region is home to the country's significant livestock sector, and according to the Corn Industry Association (ABIMILHO)[10], the poultry and pork production chains are the primary consumers of corn in the Brazilian domestic market. Meanwhile, a substantial portion of the second corn crop is exported, utilizing the same transportation and port infrastructure as soybean exports, albeit a few months later. Consequently, local prices for the first corn crop are less influenced by international markets compared to the second corn crop.

Farmers, especially those situated in regions with unfavorable climates, seek crops that optimize profitability among other factors. Decision-making in the field requires assertiveness to ensure satisfactory results. It is essential for farmers to have access to reliable information that can aid in formulating adequate plans. The use of tools for forecasting time series of buying and selling agricultural commodities becomes crucial in providing price information to support decision-making processes while reducing uncertainty and risks in agricultural markets and agricultural insurance programs (Ouyang et al., 2020; Wang et al., 2017).

Therefore, the objective of this study is to utilize a grid search algorithm to optimize two recurrent neural networks (RNN), namely long short-term memory (LSTM) and bidirectional-LSTM (BiLSTM), and a Transformer-based solution, in a forecast horizon of up to 13 weeks, to compare the predictive capabilities of these models in estimating the price series of the two agricultural commodities used in dual cropping, namely soybeans and corn. The price series data for these commodities can be accessed at the Center for Advanced Studies in Applied Economics (CEPEA)[11] website. The incorporation of the grid search algorithm in the models aims to optimize the hyperparameters. Hyperparameters are values that control the learning process and ensure that the model can solve a problem in the best possible way, minimizing pre-defined losses and providing accurate results.

---

[10] http://www.abimilho.com.br/
[11] https://www.cepea.esalq.usp.br/

2. RELATED WORKS

The progress of artificial intelligence (AI) methodologies has made it easier to utilize them across a wide range of domains and applications. These applications encompass object detection (Kong et al., 2021), scene classification (Hung et al., 2021), transportation systems (Abduljabbar et al., 2019), and time series prediction (Abraham et al., 2020; Li et al., 2020; Sun and Guo, 2023). AI models have demonstrated promising results compared to other prediction methods due to their ability to learn, generalize, and establish meaningful connections among significant features.

Over the past few decades, the field of forecasting has evolved from relying on traditional statistical methods (e.g., ARIMA) (Ariyo et al., 2014; Xu et al., 2021; Zhang et al., 2016) and machine learning techniques (e.g., MLP) (Ak et al., 2016; Oliveira et al., 2014) to embracing deep learning-based solutions. Deep learning models, such as RNNs (Ayus et al., 2023) and temporal convolutional networks (K. Liu et al., 2021), have gained popularity. Unlike traditional AI techniques and statistical methods, which are characterized by shallow structures and limited learning capabilities, deep learning stands out as a powerful AI technique capable of addressing a wide range of applications. It offers high generalizability, optimal support for unsupervised learning, and the ability to handle large datasets and complex applications effectively, including prediction problems (Ayoobi et al., 2021).

RNNs belong to the family of artificial neural networks and are characterized by their recurrent architecture. These networks utilize previous outputs as inputs, influencing the computation of subsequent outputs. The presence of recurrent connections enables the integration of past information along with the regular input, allowing the network to retain memory and learn from previous events. The fundamental form of RNN, often referred to as Simple RNN or vanilla RNN, is adept at modeling temporal dependencies and is particularly well-suited for predicting sequential data (Parmezan et al., 2019). They can effectively process one or more sequential inputs, making them applicable in various machine learning contexts (Goodfellow et al., 2016).

For time series prediction, several deep learning models have been successfully applied, including BiLSTM (Murugesan et al., 2021), LSTM (Ouyang et al., 2019), and autoencoders (Gastli et al., 2021), which have demonstrated their effectiveness in solving this problem. The LSTM neural network (Hochreiter and Schmidhuber, 1997) is a type of RNN that is specifically designed to address the vanishing gradient problem in traditional RNNs. It is widely used in various applications such as natural language processing (NLP) (Palangi et al., 2016), speech recognition (Shashidhar et al., 2022), and time series forecasting (Elsworth and Güttel, 2020; Kumar et al., 2023; Sagheer and Kotb, 2019). A comprehensive description of how LSTM networks work in detail can be found in (Hochreiter & Schmidhuber, 1997).

BiLSTM (Schuster and Paliwal, 1997) is a variant of the popular RNN architecture called LSTM. Unlike traditional LSTMs that process sequential data in a unidirectional manner, BiLSTM simultaneously processes input sequences in both forward and backward directions. BiLSTMs have proven to be effective in various tasks that require understanding contextual information, such as speech recognition (Efanov et al., 2022), sentiment analysis (Xu et al., 2019), and time series forecasting (Pirani et al., 2022; Staffini, 2023). By leveraging information from both directions, BiLSTMs enhance the model's ability to capture long-term dependencies and make more accurate predictions.

Another recent development is the introduction of Transformers (Vaswani et al., 2017) as a novel methodology for handling sequence data, particularly in the context of sequence-to-sequence tasks. In 2017, Google introduced the Transformer model for tackling NLP tasks. Over the years, the Transformer has made significant strides in both image processing and NLP (Gillioz et al., 2020; Li et al., 2022; Yan et al., 2023), demonstrating its effectiveness in enhancing model performance.

Convolutional neural networks (CNNs) (He et al., 2016; Krizhevsky et al., 2017; Simonyan and Zisserman, 2015) are deep learning models specifically designed for processing grid-like structured data, such as images or time series, by extracting meaningful patterns and features through convolutional and pooling layers. Notably, researchers have proposed Transformer-like models that combine CNNs, capable of extracting local features, to solve various tasks, including time series forecasting (Y. Liu et al., 2021; WANG and ZHAO, 2023; Zhou et al., 2021). These models allow for capturing long-term dependencies in sequences and support parallel computing.

Compared to time series forecasting models based on CNNs (Bai et al., 2018; Stoller et al., 2019; van den Oord et al., 2016) and RNNs (Lai et al., 2018; Salinas et al., 2020; Yu et al., 2019), the Transformer's self-attention mechanism offers the advantage of modeling each aspect of input sequences (Li et al., 2019; Wu et al., 2021; Zhou et al., 2021, 2022). However, the self-attention mechanism of Transformers introduces additional computational costs and memory consumption as the input length increases.

3. METHODOLOGY

Agricultural commodity price time series exhibit nonlinearity and non-stationarity, which pose inherent challenges in developing accurate forecasting models. Therefore, the proposed method primarily aims to optimize the hyperparameters of LSTM, BiLSTM, and Transformer networks. The following section outlines the steps involved in the methodology.

3.1 PREPARING DATA

Before the data can be utilized for the forecasting task, it needs to undergo preparation, which includes cleaning and normalizing the time series.

Cleaning the data involves removing noise and handling missing values using appropriate techniques. Meanwhile, normalization is crucial for facilitating the learning process of the network. It involves transforming the original scale of the data into smaller values.

In this study, we employ the MinMaxScaler method from Scikit-learn[12] to perform the normalization operation. The time series, denoted as $x(t)$ and of length N, is represented by $\{x(t_i),\ i = 1, 2, \dots, N\}$. The Min-Max Normalization equation (1) is as follows:

$$x(t_i) = \frac{x(t_i) - min(x)}{max(x) - min(x)} \tag{1}$$

Where $min(x)$ is the minimum value of the time series and $max(x)$ is the maximum value. Normalizing the time series values $x(t)$ results in the data expressed as $\{x_{normalized}(t_i),\ i = 1, 2, \dots, N\}$.

3.2 CREATING A COLLECTION OF HYPERPARAMETER COMBINATIONS

Machine learning algorithms consist of parameters that need to be set prior to the learning process, known as hyperparameters. These hyperparameters directly impact the performance of the model (Goodfellow et al., 2016).

[12] https://scikit-learn.org/

Finding the optimal hyperparameters, which yield the best performance on selected evaluation metrics for both the training and test sets, is a time-consuming task that requires expertise and effort (Feurer and Hutter, 2019).

Hyperparameter optimization techniques aim to systematically search for the best set of user-defined hyperparameters that maximize performance. One commonly used approach is the grid search algorithm, which exhaustively searches through a predefined grid of hyperparameters to identify the best performing configuration (Reimers and Gurevych, 2017). However, grid search can be computationally demanding, especially when dealing with a large search space.

To perform grid search, predefined values are assigned to each hyperparameter, and a list of parameter combinations is generated. In this study, we will focus on optimizing the hyperparameters for LSTM, BiLSTM, or Transformer models. In this paper, we will use the following hyperparameters:

Lag size (LSTM, BiLSTM and Transformer): The lag size parameter has a significant impact on the performance of time series forecasting (Ribeiro et al., 2011). Thus, it is crucial to assess the model's performance using different lag sizes.

Forecast horizon (LSTM, BiLSTM and Transformer): The forecast horizon in neural networks refers to the time period for which predictions or forecasts are made. It represents the future time steps that the model aims to predict based on the available historical data.

Number of hidden layers: The number of hidden layers in neural networks refers to the intermediate layers between the input and output layers. These hidden layers play a crucial role in capturing complex patterns and relationships in the data (Hermans and Schrauwen, 2013; Utgoff and Stracuzzi, 2002). Increasing the number of hidden layers allows the neural network to learn more intricate representations of the input data, potentially improving its ability to solve complex tasks. However, adding more hidden layers also increases the model's complexity and can lead to overfitting if not properly regularized. Therefore, determining an appropriate number of hidden layers is a crucial decision in designing effective neural network architectures.

Number of neurons (LSTM, BiLSTM and Transformer): Determining the optimal number of neurons in a neural network is not a straightforward task. The number of neurons directly affects the model's performance. If the number is too small, the network may not store enough information for accurate predictions. Conversely, an excessive number of neurons can lead to poor generalization and overfitting. For the Transformer model, we focus on the number of MLP units used in the architecture.

Epoch size (LSTM, BiLSTM and Transformer): this hyperparameter defines the number of iterations over all training instances (Brownlee, 2018). During the training process, the neural network learns from a training dataset, adjusting its weights and biases to minimize the training error. However, if the network is trained for too long, it may start to overfit the training data, meaning it becomes too specialized and fails to generalize well to new, unseen data.

Patience (LSTM, BiLSTM and Transformer): It involves monitoring the network's performance on a separate validation dataset during training. If the performance on the validation dataset fails to improve for a certain number of consecutive iterations, the training process is stopped. By doing so, early stopping helps strike a balance between capturing important patterns in the training data and avoiding overfitting. It leads to a more generalizable model that performs well on unseen data. Early stopping is an effective method to improve the efficiency and effectiveness of neural network training (Muhammad et al., 2022).

Batch size (LSTM, BiLSTM and Transformer): The batch size refers to the number of samples processed before the model's weights are updated (Brownlee, 2018). Finding an optimal value for the batch size is crucial for achieving good model performance.

Dropout rate (LSTM, BiLSTM and Transformer): The dropout rate involves randomly excluding a fraction of hidden neurons during training to reduce sensitivity to individual neuron weights. The dropout rate influences the model's performance by enhancing its generalizability (Srivastava et al., 2014).

Optimizer (LSTM and BiLSTM): Optimizers play a crucial role in training networks by adjusting the model's parameters. Some commonly used optimizers include Stochastic Gradient Descent (SGD), Adam, Adamax, AdaGrad, and Nadam. These optimizers help efficiently train networks by updating parameters, adapting the learning rate, and handling challenges in RNN training. Experimentation is often needed to determine the most suitable optimizer.

Head size (Transformer only): In summary, the head size in a Transformer model refers to the number of parallel self-attention heads used within each layer. It enables the model to capture diverse relationships and dependencies in the input data, leading to richer representations and improved performance on NLP tasks. Increasing the head size allows for more parallel processing and capturing finer-grained dependencies but comes with a higher computational cost (Tenney et al., 2019).

Filters (Transformer only): in a CNN model determines the number of learnable filters or convolutional kernels that are applied to the input data. Increasing the number of filters allows the model to capture more diverse and complex features. However, the optimal number of filters depends on the specific task and available computational resources. The reason for this application is that our Transformer model incorporates CNN layers (Krizhevsky et al., 2017).

3.3 TRANSFORMING DATA INTO SUPERVISED LEARNING

In this stage, we begin by restructuring the time series data into a collection of instances, with predefined input features and an output feature. These instances are then divided into train and test sets.

3.3.1 Generating instances based on lag size and forecast horizon

The forecast horizon and input size play a crucial role in determining the data formatting for time series forecasting. The forecast horizon refers to the number of future time steps for which predictions are made, while the input size represents the number of past time steps used as input to make those predictions.

In one-step time series forecasting, the model predicts the value of the next time step based on the historical data up to the current time step. Here, the input size is typically one, as the model only needs the immediate past data point to generate the next prediction. This format is suitable for short-term predictions or when the focus is on immediate future values.

On the other hand, in multi-step time series forecasting, the model aims to predict multiple future time steps ahead. The forecast horizon is typically greater than one, indicating the number of time steps to be predicted into the future. In this case, the input size becomes crucial as the model needs a sufficient number of past time steps to capture the patterns and dependencies leading up to the future predictions. The input size is usually larger than one, and it can vary depending on the complexity and patterns present in the time series.

When formatting the data for one-step forecasting, each input-output pair consists of a single input time step and its corresponding target output. In multi-step forecasting, each input-output pair involves a sequence of past time steps as the input and a sequence of future time steps as the target output. This reshaping task is necessary to utilize the LSTM, BiLSTM and Transformer model effectively. Using the lag L and forecast horizon h, the instances are created as follows (2):

$$x_{input} = \begin{bmatrix} x(t_1) & x(t_2) & \cdots & x(t_L) \\ x(t_2) & x(t_3) & & x(t_{L+1}) \\ \cdots & \cdots & & \cdots \\ x(t_{N-L}) & x(t_{N-L+1}) & \ldots & x(t_{N-1}) \end{bmatrix} \tag{2}$$

The associated output value can be represented in the following manner (3):

$$x_{output} = \begin{bmatrix} x(t_{L+h}) \\ x(t_{L+1+h}) \\ \vdots \\ x(t_{N+h}) \end{bmatrix} \tag{3}$$

Note that in our data formatting, we adopt a different approach. The output values we consider are the h values following the last lag L value of each instance. Although the format resembles the one-step format, the values included fall within a predetermined forecast horizon.

3.3.2 Splitting instances into train and test sets

At this point, we split the instances into two distinct parts: the train set and the test set. The train set is utilized to train the LSTM, BiLSTM, and Transformer network and create a forecasting model. The test set is employed to evaluate the predictive performance of the model using various performance metrics.

3.4 MODEL CONFIGURATION AND TRAINING

Since our study encompasses two models, one primarily utilizing a multilayer LSTM and BiLSTM architecture and the other employing a Transformer architecture with CNN layers, we aim to offer a comprehensive description of the structure of each model.

3.4.1 Specification of stacked LSTM

Stacked LSTM refers to the use of multiple LSTM layers. Each LSTM layer contains a series of memory cells that store and process information over time. Stacking multiple LSTM layers allows for the extraction of more complex temporal patterns and dependencies in sequential data.

In a stacked LSTM, the information flows from one LSTM cell to another through a process called the forward pass. During the forward pass, the input sequence is fed into the first LSTM layer, and the output of each LSTM cell in that layer is passed as input to the corresponding cells in the subsequent layer.

The flow of information from one LSTM cell to another within a layer is facilitated by three key components: the input gate, the forget gate, and the output gate. These gates regulate the flow of information and control the memory operations within each LSTM cell.

The input gate determines which parts of the input sequence should be stored in the cell's memory. The forget gate decides which information from the previous cell state should be discarded. The output gate controls which parts of the cell's memory should be used to generate the output.

By stacking multiple LSTM layers, each layer can learn different levels of abstraction from the input sequence. The lower layers capture low-level temporal patterns, while the higher layers capture more complex patterns that are built upon the representations learned by the lower layers.

3.4.2Transformer with CNN

The integration of CNN layers within the Transformer architecture for time series forecasting has gained significant attention in recent research. This combination leverages the strengths of both CNN and Transformer models to improve the predictive performance of time series forecasting tasks.

CNN layers excel at capturing local patterns and extracting relevant features from input data (Krizhevsky et al., 2017). By incorporating CNN layers into the Transformer architecture, the model can effectively capture spatial dependencies within the time series data. This is particularly beneficial when dealing with time series that exhibit spatial patterns or contain spatially related information.

The CNN layers in the Transformer architecture allow for efficient feature extraction from the input sequence, providing the model with a more comprehensive representation of the underlying patterns and trends in the time series (Wang and Zhao, 2023). This enhanced representation can lead to improved forecasting accuracy and the ability to capture complex relationships between different time steps.

Furthermore, the combination of CNN and Transformer models can effectively handle both short-term and long-term dependencies in the time series data (Zhou et al., 2020). While the CNN layers focus on capturing local patterns and short-term dependencies, the self-attention mechanism in the Transformer model enables the modeling of long-term dependencies and captures global relationships within the sequence.

3.4.3Modeling

With the exception of the lag size and forecast horizon parameter, the remaining hyperparameters play a crucial role in configuring and training the respective networks. These hyperparameters include the number of hidden layers, epoch size, batch size, number of neurons, dropout rate, patience, optimizer, head size, and filters. It is important to note that the head size and filter hyperparameters are specific to the Transformer network, while the optimizer is exclusively used for the LSTM and BiLSTM networks.

To determine the optimal values for these hyperparameters, the grid search method is employed in the previous step. Subsequently, in this stage, for each combination of hyperparameters, an LSTM, BiLSTM, and Transformer network is created and trained using the training set. This systematic approach allows for comprehensive exploration of various hyperparameter configurations, enabling the selection of the most suitable settings for each network.

3.5 SELECTING THE BEST MODEL

In this step, we evaluate the performance of each generated candidate model using the RMSE (Root Mean Squared Error) metric. The RMSE is calculated based on the denormalized input and predicted output data (Shcherbakov et al., 2013). RMSE (4) is defined by:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \quad (4)$$

Where y_t is the value taken by the explained variable at time t, $\hat{y}_t$ is the value of the explained variable calculated by the model and n is the observations number. Finally, we select the best-performing model based on this evaluation.

4. EXPERIMENTS

Three different RNN architectures, along with a Transformer-based model, were utilized for this project. The implementation of these models was carried out using Keras[13] (Chollet, 2015), a popular deep learning library. In addition, several essential libraries such as Scikit-learn, Pandas[14], Numpy[15], and Matplotlib[16] were employed to perform various tasks related to data formatting.

4.1 DATA DESCRIPTION

A common starting point, February 2006, was selected to align the studies. This date was chosen because it marked the availability of all the series and came after the beginning of the double corn crop, which caused significant fluctuations in the series. For the analysis, a total of 4,000 observations were collected for both commodities.

The Brazilian prices considered are the average daily prices at the Port of Paranaguá for soybeans and the average daily prices in Campinas, São Paulo state, for corn. The price for the Campinas-SP region represents transactions involving purchases, sales, and offers to buy and sell the product originating from any region of the country and placed in any of the 162 municipalities. These price data are sourced from the CEPEA, which is a research center affiliated with the Department of Economics, Administration, and Sociology of the University of São Paulo.

Data from COMEX STAT[17] reveals that the Port of Paranaguá is the second-largest exporter of soybeans, exporting around 15 million tons, following the Port of Santos, which exports approximately 20 million tons. The distinction between these two ports is that while Santos exclusively exports grains from the Midwest and Southeast regions, Paranaguá exports grains from these regions as well as grains originating in the South. This is why we utilize the Port of Paranaguá location for Brazilian soybean prices.

In addition to the aforementioned reasons, these data sets were chosen due to their significant representation in exports and their importance in the Brazilian agribusiness sector.

[13] https://keras.io/
[14] https://pandas.pydata.org/
[15] https://numpy.org/
[16] https://matplotlib.org/
[17] http://comexstat.mdic.gov.br/

Furthermore, the availability of historical series for these price indicators allows for public access and consultation, making them suitable for analysis and research purposes.

4.2 PREPARING DATA

The purpose of data preparation is to ensure the availability of a high-quality dataset that is suitable for modeling and aligns with our objectives. The time series utilized in this study do not contain any missing values. Moreover, the data has not undergone any noise reduction or smoothing techniques in order to preserve its original characteristics, thus ensuring that the model is trained on data that closely reflects real-world patterns.

Since many prediction techniques exhibit improved performance when applied to normalized data, we performed data normalization using the min-max normalization algorithm (Han et al., 2022). By applying this normalization technique, we scaled the data to a standardized range, allowing for more effective comparisons and analysis within the model. This normalization process aids in enhancing the predictive capabilities of the model and ensures that the data is in a suitable format for further analysis and modeling.

4.3 GENERATING A LIST CONTAINING COMBINATIONS OF HYPERPARAMETERS

In this stage, we identify a range of values for each hyperparameter that will be utilized in configuring and training the model. The values for each hyperparameter are outlined in Table 1, which represents the search space for the hyperparameters. It is worth noting that the epoch size was not included as part of the hyperparameter list, but rather set as a constant value of 200 epochs. This decision was made because the model will only utilize this number of epochs if the early stopping criteria specified by the patience parameter is not met, reaching the maximum limit of 200 epochs. Additionally, when the patience value is set to zero (disabled), the model will train for the full 200 epochs without any early stopping. This straightforward configuration allows us to investigate the potential impact of patience on the performance of the model.

Table 1 — A range of values was defined for each hyperparameter.

Hyperparameters	Values
Lag size	[30, 60, 90, 120]
Forecast horizon	[1, 7, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77, 84, 91]
Nº hidden layers	[2, 4, 8]
Nº of neurons	[64, 128, 256, 512]
Paience	[“off”, 10, 20]
Batch size	[20, 32, 40, 64, 80, 128]
Dropout	[0.1, 0.2, 0.25]
Optimizer	[“Adam”, “Adamax”, “Nadam”]
Head size	[2, 4, 8]
Filters	[2, 4, 8]

4.4 TRANSFORMING DATA INTO SUPERVISED LEARNING

A lag size was established to convert the time series into instances in the input-output format. In this study, the grid search method was employed with lag sizes of 30, 60, 90, and 120. The forecast horizons chosen were 1 and 91. Subsequently, the best-performing model was further evaluated with new forecast horizons, evenly spaced at intervals of 7 observations. This step aimed to observe the error behavior across different horizons. The selected test horizons included 1, 7, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77, 84, and 91 observations. Each combination of lag size and prediction horizon resulted in a distinct instance. For instance, when the lag size was set to 5 and the forecast horizon to 3, a subset of the created instances is presented in Table 2. Following the transformation, the instances were divided into a test set and a training set, with the test set comprising 20% of all instances in all experiments.

Table 2 — Generation of a Subset of Instances.

	Input instances with lag size=5					Output with forecast horizon=3
	t1	t2	t3	t4	t5	t8
Instance 1	31.65	32.25	31.77	32.64	31.42	33.56
Instance 2	32.25	31.77	32.64	31.42	32.48	32.87
Instance 3	31.77	32.64	31.42	32.48	32.48	33.13
Instance 4	32.64	31.42	32.48	32.48	33.56	33.61
Instance 5	31.42	32.48	32.48	33.56	32.87	33.46
Instance 6	32.48	32.48	33.56	32.87	33.13	33.20
Instance 7	32.48	33.56	32.87	33.13	33.61	32.77
Instance 8	33.56	32.87	33.13	33.61	33.46	32.81
Instance 9	32.87	33.13	33.61	33.46	33.20	32.17
Instance 10	33.13	33.61	33.46	33.20	32.77	31.90

5. RESULTS DISCUSSION

To identify the optimal forecasting model for agricultural commodity price series, a comparison was conducted among the LSTM, BiLSTM, and Transformer methods. The grid search method was applied to all three methods to optimize their respective hyperparameters and determine the most suitable configuration for the models. The structure of the results section is as follows: Initially, we provide the RMSE values computed for each model and forecast horizon, encompassing both time series. Next, we conduct statistical tests to compare the forecast performance results of each predictor. Finally, we illustrate the forecasting performance by presenting the actual and predicted daily data generated by each model. Additionally, we assess the models' performance across the forecast horizons.

5.1 PERFORMANCE RESULTS

The evaluation of the models was based on RMSE, which enabled the selection of the most favorable model. The hyperparameters for the best model with a short forecast horizon (forecast horizon=1) are presented in Table 3, while the hyperparameters for the best model with a long forecast horizon (forecast horizon=91) are provided in Table 4.

Table 3 — The best hyperparameters of the models acquired for a forecast horizon of 1.

	Soybean			Corn		
Hyperparameters	LSTM	BiLSTM	Tranformer	LSTM	BiLSTM	Tranformer
Lag size	60	60	60	90	30	60
Nº hidden layers	2	2	4	2	2	4
Nº of neurons	64	256	128	64	64	256
Paience	10	10	"off"	10	10	"off"
Batch size	40	80	32	80	80	64
Dropout	0.1	0.1	0.25	0.1	0.1	0.25
Optimizer	Adam	Nadam	—	Adamax	Adamax	—
Head size	—	—	128	—	—	256
Filters	—	—	4	—	—	4

Table 4 — The best hyperparameters of the models acquired for a forecast horizon of 91.

	Soybean			Corn		
Hyperparameters	LSTM	BiLSTM	Tranformer	LSTM	BiLSTM	Tranformer
Lag size	60	90	60	90	120	60
Nº hidden layers	2	4	4	2	4	4
Nº of neurons	256	256	128	128	256	256
Paience	"off"	20	20	10	10	10
Batch size	64	128	64	128	64	128
Dropout	0.2	0.2	0.2	0.2	0.2	0.25
Optimizer	Adam	Adam	—	Adam	Adam	—
Head size	—	—	256	—	—	128
Filters	—	—	4	—	—	4

5.2 STATISTICAL COMPARISONS OF THE PERFORMANCE OF THE APPLIED MODELS

Table 5 presents a depiction of the effectiveness of implementing the approaches on the agricultural commodity price series data, considering both long (91 days) and short (1 day) forecast horizons.

Table 5 — Performance test results.

	Soybean			Corn		
Forecast horizon	LSTM	BiLSTM	Tranformer	LSTM	BiLSTM	Tranformer
Short-term	**0.3695**	0.3720	0.4696	**0.1729**	0.1732	0.2347
Long-term	4.2911	4.4433	**3.8540**	**3.0725**	3.3903	3.1550

The error series generated by the models did not exhibit a normal behavior based on the results of the Kolmogorov-Smirnov and Shapiro-Wilk tests. Consequently, the validity of the outcomes obtained from any parametric test becomes questionable. To address this issue, a non-parametric Kruskall-Wallis test was employed. The Kruskal-Wallis test (Kruskal and Wallis, 1952) is a nonparametric statistical test that assesses differences between three or more independently sampled groups concerning a single continuous variable that does not follow a normal distribution. It extends the Mann-Whitney U-test (Wilcoxon rank-sum test) typically used for two groups, making it a more generalized version of the Mann-Whitney U-test and the nonparametric counterpart of one-way ANOVA (McKight and Najab, 2010).

In the Kruskal-Wallis analysis, H0 represents the null hypothesis that the mean RMSE errors of all methods are equal, while H1 denotes the alternative hypothesis where at least one mean differs from the others. The tests were performed separately for the corn and soybean time series. For the corn series, the null hypothesis is not rejected as the p-value (0.805) exceeds the significance level of 0.05. Hence, no significant differences exist between the error series produced by the models in the corn time series. In contrast, the null hypothesis is rejected for the error series generated by the models predicting the soybean time series, given that the p-value (9.66E-11) is less than 0.05. When H0 is rejected, post hoc tests are necessary to compare the equality of each pair of means under the null hypothesis. In this study, we employed the Kruskal-Wallis rank-sum test for the post hoc tests. Thus, the Kruskal-Wallis rank-sum test was utilized for pairwise comparisons, as illustrated in Table 6. According to the results, the Transformer model exhibits statistically superior performance compared to all other models. Interestingly, the LSTM and BiLSTM methods share the second position.

Table 6 — Kruskal-Wallis post hoc test results for soybean dataset.

Compared Factors	Difference Observed	Critical Difference	Difference
BiLSTM - LSTM	23.625	109.899	No
BiLSTM - Transformer	176.523	109.899	Yes
LSTM - Transformer	200.148	108.593	Yes

5.3 PERFORMANCE ACROSS THE FORECAST HORIZONS

In the Fig. 1, we present the RMSE comparisons of the three models at different forecast horizons to determine if the Transformer model can capture the dependencies of longer time inputs compared to the LSTM and BiLSTM models, specifically using the soybean series. For both datasets, the forecast horizon sizes are set as 7, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77, 84, and 91, which correspond to {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, and 13 weeks, respectively}.

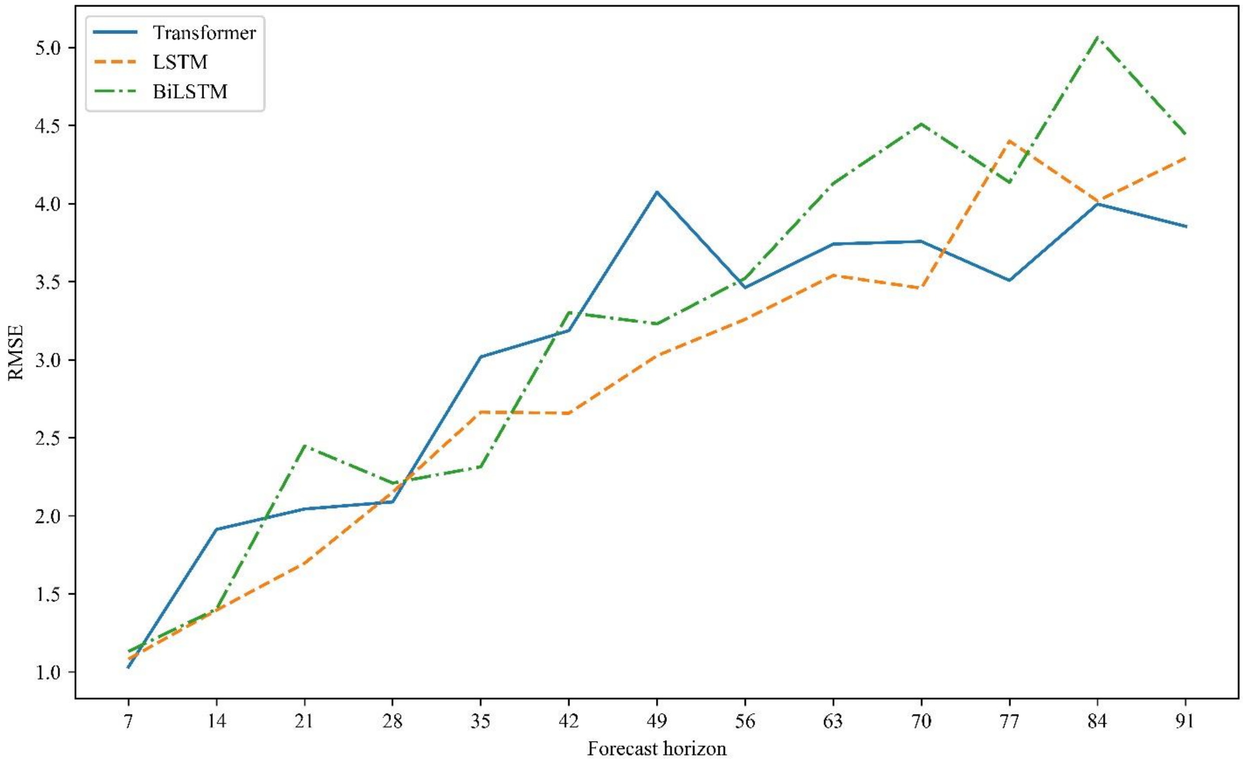


Fig. 1 — The RMSE results (Y-axis) of models for soybean data are plotted against different forecast horizon sizes (X-axis) for both long-term forecasting and short-term forecasting.

As depicted in Fig. 1, as the forecast horizon size increases, the RMSE of the Transformer model demonstrates a lower accumulation of errors starting from week 8 onwards. However, this pattern is not observed in the LSTM and BiLSTM models.

When comparing the corn time series, all models exhibit a similar increase in errors (Fig. 2), which is consistent with the earlier conducted statistical tests. In most cases, the performance of the models fluctuates or deteriorates as the forecast horizons extend.

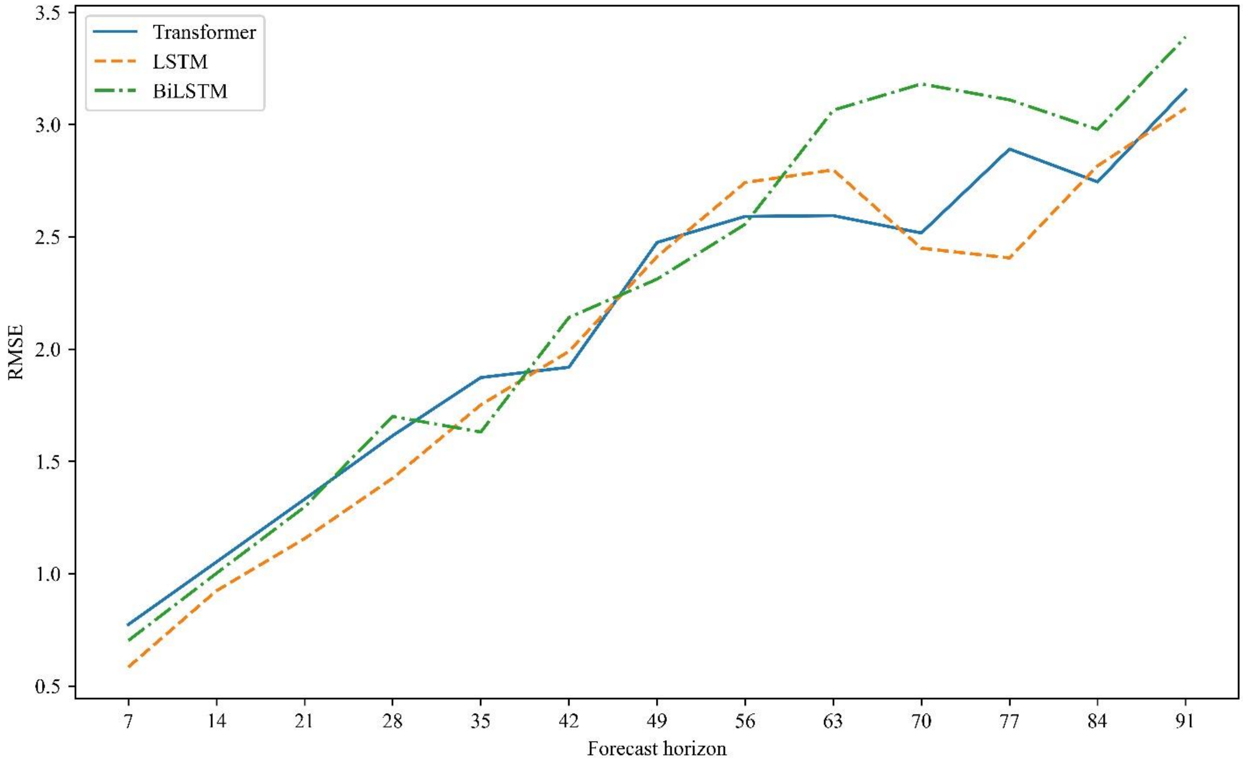


Fig. 2 — The RMSE results (Y-axis) of models for corn data are plotted against different forecast horizon sizes (X-axis) for both long-term forecasting and short-term forecasting

In Fig. 3 and Fig. 4, we present the predicted values of all models, along with their best hyperparameter settings, compared to the actual values of the corn and soybean time series, respectively. When considering a forecast horizon of 91 days, we observe a discernible trend in the values, occasionally accompanied by slight deviations. Notably, the LSTM and BiLSTM models exhibit similar performance, while the Transformer model shows a comparatively distinct pattern. This distinction becomes more evident in Fig. 4, which corresponds to the corn data.

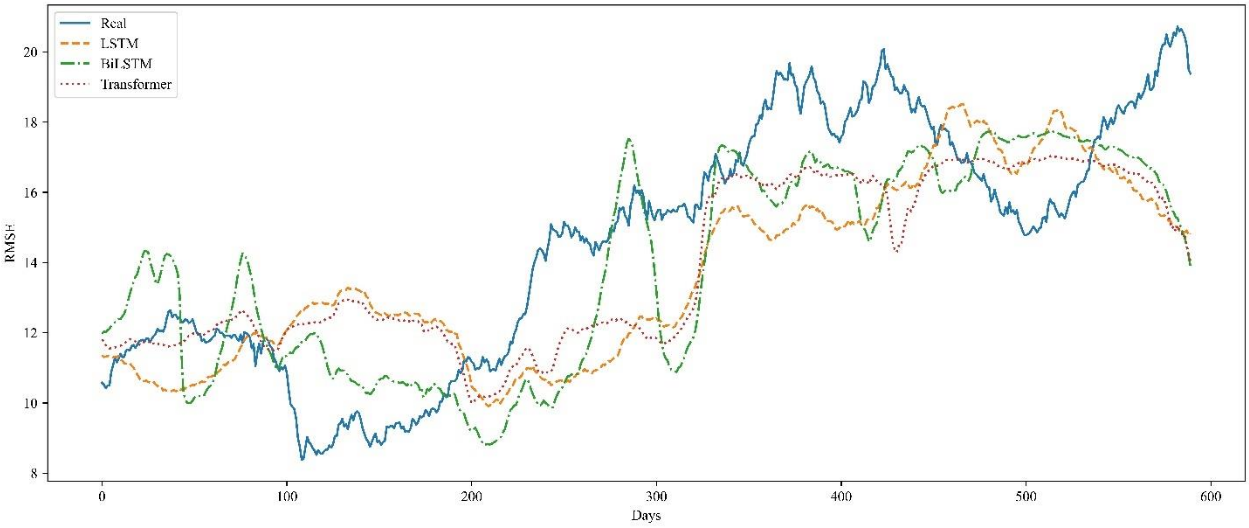


Fig. 3 — Corn actual data vs prediction using Transformer, LSTM and BiLSTM models.

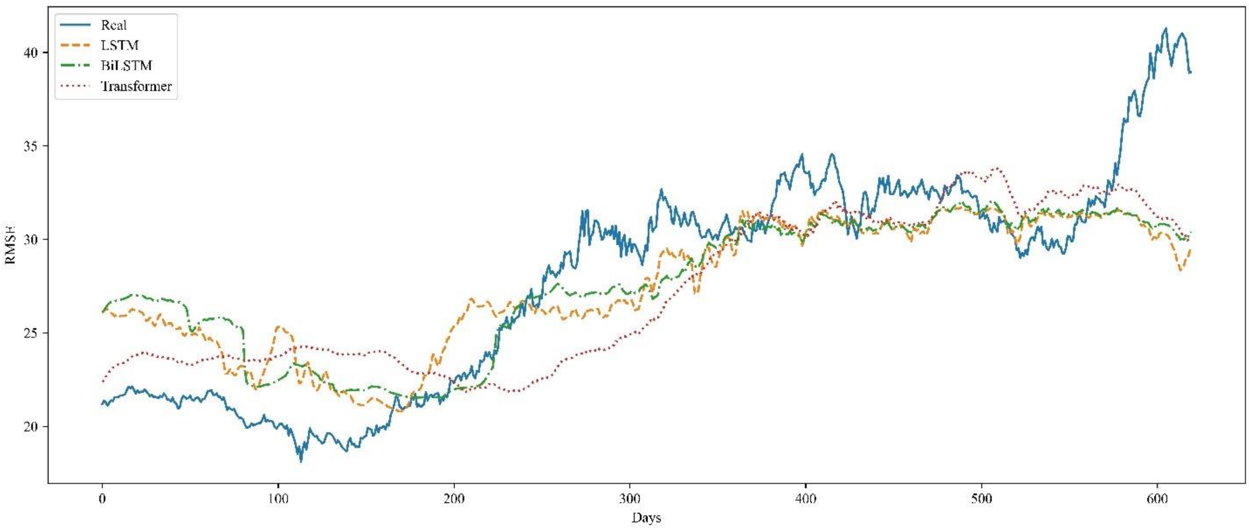


Fig. 4 — Soybean actual data vs prediction using Transformer, LSTM and BiLSTM models.

Moving on to Fig. 5, we introduce a violin graph, a statistical representation of numerical data. It resembles a box plot but incorporates a rotated kernel density for additional insights. This graph provides information about the range of errors and offers a visual representation of error size density. Analyzing the soybean data, we observe a significant variation in error amplitudes across the models. However, the Transformer model compensates for this by yielding a considerable number of errors below 2. The mean error is indicated by a prominent dot at the center of each graph, and we note that the mean error of the Transformer model is notably lower than that of the other models.

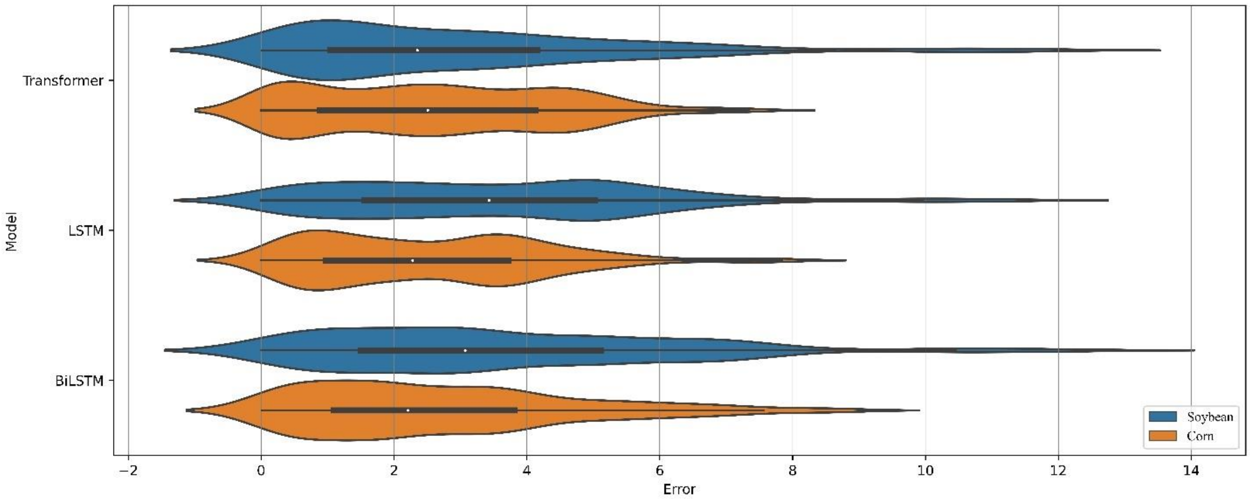


Fig. 5 — Violin plot showing distribution, mean, and range of model prediction errors.

Regarding the corn data, the error amplitudes are also relatively small compared to the amplitudes observed in the soybean data. Notably, the mean error for the Transformer model closely aligns with the mean error of the soybean data and is similar to the mean error of the LSTM and BiLSTM models. This graphical evidence confirms the lack of significant difference suggested by the Kruskal-Wallis statistical test.

Fig. 6 and Fig. 7 illustrate the actual data for corn and soybeans, respectively, alongside the predicted values generated by the Transformer model at two different forecast horizons: 1 day and 91 days. The figures clearly demonstrate that as the forecast horizon increases, the Transformer model exhibits larger errors, making it challenging to achieve accurate forecasts. This highlights the crucial role that the forecast horizon plays in the accuracy of time series forecast models.

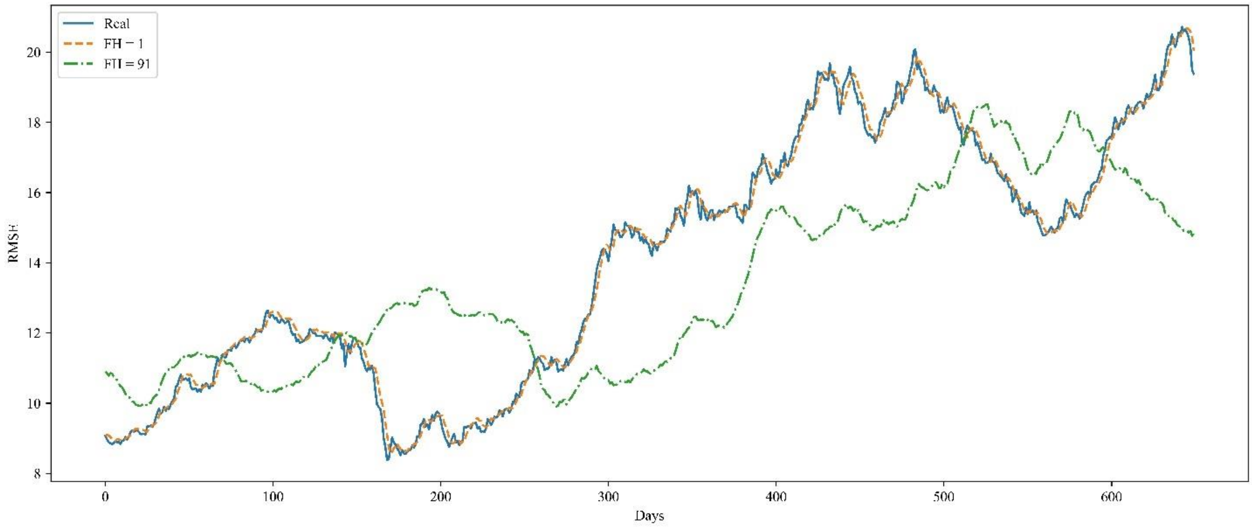


Fig. 6 — Corn actual data vs prediction using Transformer at two different forecast horizons (FH=1 and FH=91).

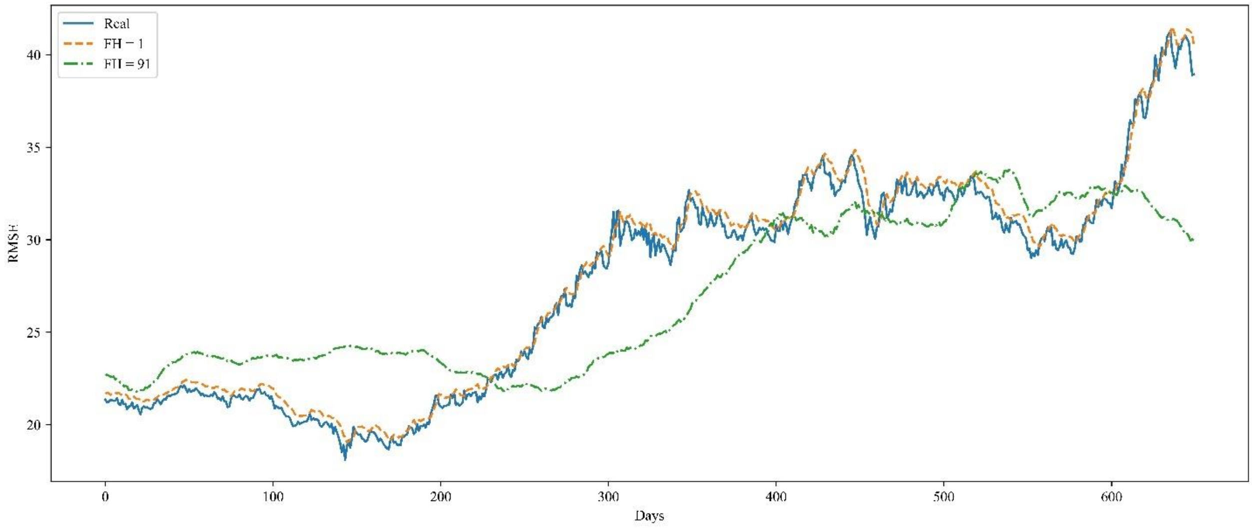


Fig. 7 — Soybean actual data vs prediction using Transformer at two different forecast horizons (FH=1 and FH=91).

Short-term forecasts, which are closer to the observed data and recent patterns, generally yield higher accuracy. However, as the forecast horizon extends, several challenges arise. These challenges include capturing complex patterns, dealing with error accumulation over time, managing uncertainty associated with long-term forecasts, and maintaining model sensitivity to changing dynamics. It is important to consider these factors when selecting an appropriate model and forecasting strategy.

Since the other models yielded similar results, only the comparison with the Transformer model is presented graphically. This avoids redundancy and unnecessary repetition of similar outcomes, streamlining the presentation of results.

6. CONCLUSION

The Transformer model demonstrates superior performance compared to current state-of-the-art forecasting models, particularly in the context of soybean data. These impressive results highlight the effectiveness of combining a CNN with the Transformer model to address the challenges of forecasting agricultural time series. By providing more accurate forecasts, this model offers significant benefits to rural producers, reducing their risks and enhancing decision-making capabilities in the field.

While the Transformer model may not exhibit a significantly higher prediction power than other models in the corn time series, it still holds its ground remarkably well considering it was not originally designed specifically for the forecasting problem. This underscores its versatility and adaptability in diverse domains.

An important observation from the study is that all models experienced an increase in error as the forecast horizon extended (Zeng et al., 2022). Consequently, it becomes crucial to identify an optimal forecast horizon that minimizes errors while still delivering the necessary information for farmers to make informed decisions.

The success of the Transformer model opens up a wealth of possibilities and research avenues in the field. The methodology employed in this study can be extended to other time series datasets and longer forecast horizons. Further modifications to the model can be explored to enhance its forecasting potential and strive towards surpassing state-of-the-art models.

Additionally, there are several favorable aspects that warrant further exploration. For instance, investigating the potential of performing multivariate forecasting by incorporating a selection of relevant features could lead to more comprehensive insights. Furthermore, studying the computational cost of Transformers is important, as although the model demonstrates good predictive power, it is necessary to address its relatively higher computational requirements compared to LSTM and BiLSTM models.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Lucas G. Meloca: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Rodrigo C. Thom de Souza:** Conceptualization, Methodology, Resources, Writing - Review & Editing, Supervision, Project administration, Funding acquisition. **Ademir Aparecido Constantino:** Resources, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

ACKNOWLEDGEMENTS

This work was supported by the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) - Financing Code 001; and by the Araucária Foundation and CNPq (grant number 315475/2018-8).

REFERENCES

Abduljabbar, R., Dia, H., Liyanage, S., Bagloee, S.A., 2019. Applications of artificial intelligence in transport: An overview. Sustainability (Switzerland) 11. https://doi.org/10.3390/su11010189

Abraham, E.R., Dos Reis, J.G.M., Vendrametto, O., Neto, P.L.O.C., Toloi, R.C., de Souza, A.E., Morais, M.O., 2020. Time series prediction with artificial neural networks: An analysis using Brazilian soybean production. Agriculture (Switzerland) 10, 1–18. https://doi.org/10.3390/agriculture10100475

Ak, R., Fink, O., Zio, E., 2016. Two Machine Learning Approaches for Short-Term Wind Speed Time-Series Prediction. IEEE Trans Neural Netw Learn Syst 27, 1734–1747. https://doi.org/10.1109/TNNLS.2015.2418739

Ariyo, A.A., Adewumi, A.O., Ayo, C.K., 2014. Stock Price Prediction Using the ARIMA Model, in: 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation. IEEE, pp. 106–112. https://doi.org/10.1109/UKSim.2014.67

Ayoobi, N., Sharifrazi, D., Alizadehsani, R., Shoeibi, A., Gorriz, J.M., Moosaei, H., Khosravi, A., Nahavandi, S., Gholamzadeh Chofreh, A., Goni, F.A., Klemeš, J.J., Mosavi, A., 2021. Time series forecasting of new cases and new deaths rate for COVID-19 using deep learning methods. Results Phys 27, 104495.

https://doi.org/10.1016/j.rinp.2021.104495

Ayus, I., Natarajan, N., Gupta, D., 2023. Comparison of machine learning and deep learning techniques for the prediction of air pollution: a case study from China. Asian Journal of Atmospheric Environment 17. https://doi.org/10.1007/s44273-023-00005-w

Bai, S., Kolter, J.Z., Koltun, V., 2018. Convolutional Sequence Modeling Revisited, in: International Conference on Learning Representations.

Brownlee, J., 2018. What is the Difference Between a Batch and an Epoch in a Neural Network. Machine Learning Mastery 20.

Chollet, F., 2015. keras.

da SILVA, C.M., 2012. De um Dust Bowl paulista à busca de fertilidade no cerrado: a trajetória do IRI Research Institute e as pesquisas em ciências do solo no Brasil (1951-1963). Revista Brasileira de História da Ciência 5, 146–155.

da Silveira, F., Lermen, F.H., Amaral, F.G., 2021. An overview of agriculture 4.0 development: Systematic review of descriptions, technologies, barriers, advantages, and disadvantages. Comput Electron Agric. https://doi.org/10.1016/j.compag.2021.106405

Efanov, D., Aleksandrov, P., Karapetyants, N., 2022. The BiLSTM-based synthesized speech recognition. Procedia Comput Sci 213, 415–421. https://doi.org/10.1016/j.procs.2022.11.086

Elsworth, S., Güttel, S., 2020. Time Series Forecasting Using LSTM Networks: A Symbolic Approach.

Feurer, M., Hutter, F., 2019. Hyperparameter Optimization, in: Hutter, F., Kotthoff, L., Vanschoren, J. (Eds.), Automated Machine Learning: Methods, Systems, Challenges. Springer International Publishing, Cham, pp. 3–33. https://doi.org/10.1007/978-3-030-05318-5_1

Gastli, M.S., Nassar, L., Karray, F., 2021. Deep Learning Models for Strawberry Yield and Price Forecasting Using Satellite Images, in: 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, pp. 1790–1796. https://doi.org/10.1109/SMC52423.2021.9658728

Gillioz, A., Casas, J., Mugellini, E., Khaled, O.A., 2020. Overview of the Transformer-based Models for NLP Tasks, in: 2020 15th Conference on Computer Science and Information Systems (FedCSIS). pp. 179–183. https://doi.org/10.15439/2020F20

Gomes, C.E., Couglan Hilter, O., Cardoso, S., Pereira De Oliveira, E., Juarez Batista De Oliveira, F., Gomes, L., Marco, S., Garcia, A., Chaves, M., Gama De Macêdo, M.H., Gontijo, E.C., Arthur, F., Lima, S., Santos, R. Dos, Tarsis, S., De Oliveira Piffer, R., Barros De Sousa, F., Mara, R., Neves, R., Kaefer, A.L., Valnier, A., Santos De Azevedo, A., Pan, A.A., 2022. Acompanhamento da safra brasileira de grãos v.9 – Safra 2021/22, n.12 - Décimo segundo levantamento. Brasília.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep learning. MIT press.

Han, J., Pei, J., Tong, H., 2022. Data mining: concepts and techniques. Morgan kaufmann.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778. https://doi.org/10.1109/CVPR.2016.90

Hermans, M., Schrauwen, B., 2013. Training and Analysing Deep Recurrent Neural Networks, in: Burges, C.J., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc.

Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. Neural Comput 9, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hung, S.-C., Wu, H.-C., Tseng, M.-H., 2021. Integrating image quality enhancement methods and deep learning techniques for remote sensing scene classification. Applied Sciences (Switzerland) 11. https://doi.org/10.3390/app112411659

Kong, J., Wang, H., Wang, X., Jin, X., Fang, X., Lin, S., 2021. Multi-stream hybrid architecture based on cross-level fusion strategy for fine-grained crop species recognition in precision agriculture. Comput Electron Agric 185. https://doi.org/10.1016/j.compag.2021.106134

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. Imagenet classification with deep convolutional neural networks. Commun ACM 60, 84–90.

Kruskal, W.H., Wallis, W.A., 1952. Use of ranks in one-criterion variance analysis. J Am Stat Assoc 47, 583–621.

Kumar, B., Sunil, Yadav, N., 2023. A novel hybrid model combining βSARMA and LSTM for time series forecasting. Appl Soft Comput 134, 110019. https://doi.org/10.1016/j.asoc.2023.110019

Lai, G., Chang, W.-C., Yang, Y., Liu, H., 2018. Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks, in: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR ’18. Association for Computing Machinery, New York, NY, USA, pp. 95–104. https://doi.org/10.1145/3209978.3210006

Li, H., Cui, Y., Wang, S., Liu, J., Qin, J., Yang, Y., 2020. Multivariate Financial Time-Series Prediction With Certified Robustness. IEEE Access 8, 109133–109143.

https://doi.org/10.1109/ACCESS.2020.3001287

Li, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y.-X., Yan, X., 2019. Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting, in: Wallach, H., Larochelle, H., Beygelzimer, A., d Alché-Buc, F., Fox, E., Garnett, R. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc.

Li, X., Chen, X., Yang, J., Li, S., 2022. Transformer helps identify kiwifruit diseases in complex natural environments. Comput Electron Agric 200, 107258. https://doi.org/10.1016/j.compag.2022.107258

Liu, K., Ke, F., Huang, X., Yu, R., Lin, F., Wu, Y., Ng, D.W.K., 2021. DeepBAN: A Temporal Convolution-Based Communication Framework for Dynamic WBANs. IEEE Transactions on Communications 69, 6675–6690. https://doi.org/10.1109/TCOMM.2021.3094581

Liu, Y., Wu, Y.-H., Sun, G., Zhang, L., Chhatkuli, A., Van Gool, L., 2021. Vision Transformers with Hierarchical Attention.

McKight, P.E., Najab, J., 2010. Kruskal-Wallis Test, in: The Corsini Encyclopedia of Psychology. Wiley, pp. 1–1. https://doi.org/10.1002/9780470479216.corpsy0491

Miao, R., Khanna, M., Huang, H., 2016. Responsiveness of Crop Yield and Acreage to Prices and Climate. Am J Agric Econ 98, 191–211. https://doi.org/https://doi.org/10.1093/ajae/aav025

Muhammad, A.R., Utomo, H.P., Hidayatullah, P., Syakrani, N., 2022. Early Stopping Effectiveness for YOLOv4. Journal of Information Systems Engineering and Business Intelligence 8, 11–20.

Murugesan, R., Mishra, E., Krishnan, A., 2021. Deep Learning Based Models: Basic LSTM, Bi LSTM, Stacked LSTM, CNN LSTM and Conv LSTM to Forecast Agricultural Commodities Prices. https://doi.org/10.21203/rs.3.rs-740568/v1

Nóia Júnior, R. de S., Sentelhas, P.C., 2019. Soybean-maize off-season double crop system in Brazil as affected by El Niño Southern Oscillation phases. Agric Syst 173, 254–267. https://doi.org/10.1016/j.agsy.2019.03.012

Oliveira, T.P., Barbar, J.S., Soares, A.S., 2014. Multilayer perceptron and stacked autoencoder for Internet traffic prediction, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, pp. 61–71. https://doi.org/10.1007/978-3-662-44917-2_6

Ouyang, H., Wei, X., Wu, Q., 2020. Discovery and Prediction of Stock Index Pattern via Three-Stage Architecture of TICC, TPA-LSTM and Multivariate LSTM-FCNs. IEEE Access 8, 123683–123700. https://doi.org/10.1109/ACCESS.2020.3005994

Ouyang, H., Wei, X., Wu, Q., 2019. Agricultural commodity futures prices prediction via long- and short-term time series network. J Appl Econ 22, 468–483. https://doi.org/10.1080/15140326.2019.1668664

Palangi, H., Deng, L., Shen, Y., Gao, J., He, X., Chen, J., Song, X., Ward, R., 2016. Deep Sentence embedding using long short-term memory networks: Analysis and application to information retrieval. IEEE/ACM Trans Audio Speech Lang Process 24, 694–707. https://doi.org/10.1109/TASLP.2016.2520371

Parmezan, A.R.S., Souza, V.M.A., Batista, G.E.A.P.A., 2019. Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. Inf Sci (N Y) 484, 302–337. https://doi.org/10.1016/j.ins.2019.01.076

Pirani, M., Thakkar, P., Jivrani, P., Bohara, M.H., Garg, D., 2022. A Comparative Analysis of ARIMA, GRU, LSTM and BiLSTM on Financial Time Series Forecasting, in: 2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE). IEEE, pp. 1–6. https://doi.org/10.1109/ICDCECE53908.2022.9793213

Reimers, N., Gurevych, I., 2017. Optimal hyperparameters for deep lstm-networks for sequence labeling tasks. arXiv preprint arXiv:1707.06799.

Ribeiro, G.H.T., Neto, P.S.G. d. M., Cavalcanti, G.D.C., Tsang, I.R., 2011. Lag selection for time series forecasting using Particle Swarm Optimization, in: The 2011 International Joint Conference on Neural Networks. pp. 2437–2444. https://doi.org/10.1109/IJCNN.2011.6033535

Sagheer, A., Kotb, M., 2019. Time series forecasting of petroleum production using deep LSTM recurrent networks. Neurocomputing 323, 203–213. https://doi.org/10.1016/j.neucom.2018.09.082

Salinas, D., Flunkert, V., Gasthaus, J., Januschowski, T., 2020. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. Int J Forecast 36, 1181–1191. https://doi.org/https://doi.org/10.1016/j.ijforecast.2019.07.001

Schuster, M., Paliwal, K.K., 1997. Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing 45, 2673–2681. https://doi.org/10.1109/78.650093

Sharma, V., Tripathi, A.K., Mittal, H., 2022. Technological revolutions in smart farming:

Current trends, challenges & future directions. Comput Electron Agric. https://doi.org/10.1016/j.compag.2022.107217

Shashidhar, R., Patilkulkarni, S., Puneeth, S.B., 2022. Combining audio and visual speech recognition using LSTM and deep convolutional neural network. International Journal of Information Technology 14, 3425–3436. https://doi.org/10.1007/s41870-022-00907-y

Shcherbakov, M., Brebels, A., Shcherbakova, N.L., Tyukov, A., Janovsky, T.A., Kamaev, V.A., 2013. A survey of forecast error measures. World Appl Sci J 24, 171–176. https://doi.org/10.5829/idosi.wasj.2013.24.itmies.80032

Simonyan, K., Zisserman, A., 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research 15, 1929–1958.

Staffini, A., 2023. A CNN–BiLSTM Architecture for Macroeconomic Time Series Forecasting, in: ITISE 2023. MDPI, Basel Switzerland, p. 33. https://doi.org/10.3390/engproc2023039033

Stoller, D., Tian, M., Ewert, S., Dixon, S., 2019. Seq-U-Net: A One-Dimensional Causal U-Net for Efficient Sequence Modelling.

Sun, J., Guo, W., 2023. Time Series Prediction Based on Time Attention Mechanism and LSTM Neural Network, in: 2023 IEEE International Conference on Integrated Circuits and Communication Systems, ICICACS 2023. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICICACS57338.2023.10099498

Tenney, I., Das, D., Pavlick, E., 2019. BERT rediscovers the classical NLP pipeline. arXiv preprint arXiv:1905.05950.

Utgoff, P.E., Stracuzzi, D.J., 2002. Many-Layered Learning. Neural Comput 14, 2497–2529. https://doi.org/10.1162/08997660260293319

van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., Kavukcuoglu, K., 2016. WaveNet A Generative Model for Raw Audio.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need, in: Guyon, I., Fergus, R., Wallach, H., Wallach, H., Guyon, I., Vishwanathan, S.V.N., U, von L., Garnett, R., Vishwanathan, S.V.N., Bengio, S., Fergus, R. (Eds.), Advances in Neural Information Processing Systems. Neural information processing systems foundation, pp. 5999–6009.

Wang, D., Yue, C., Wei, S., Lv, J., 2017. Performance analysis of four decomposition-ensemble models for one-day-ahead agricultural commodity futures price forecasting. Algorithms 10. https://doi.org/10.3390/a10030108

WANG, N., ZHAO, X., 2023. Time Series Forecasting Based on Convolution Transformer. IEICE Trans Inf Syst 106, 976–985.

Wang, N., Zhao, X., 2023. Time Series Forecasting Based on Convolution Transformer. IEICE Trans Inf Syst E106.D, 976–985. https://doi.org/10.1587/transinf.2022EDP7136

Wu, H., Xu, J., Wang, J., Long, M., 2021. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting, in: Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P.S., Vaughan, J.W. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc., pp. 22419–22430.

Xu, F., Du, Y.-A., Chen, H., Zhu, J.-M., 2021. Prediction of fish migration caused by ocean warming based on SARIMA model. Complexity 2021. https://doi.org/10.1155/2021/5553935

Xu, G., Meng, Y., Qiu, X., Yu, Z., Wu, X., 2019. Sentiment Analysis of Comment Texts Based on BiLSTM. IEEE Access 7, 51522–51532. https://doi.org/10.1109/ACCESS.2019.2909919

Yan, C., Li, Z., Zhang, Z., Sun, Y., Wang, Y., Xin, Q., 2023. High-resolution mapping of paddy rice fields from unmanned airborne vehicle images using enhanced-TransUnet. Comput Electron Agric 210, 107867. https://doi.org/10.1016/j.compag.2023.107867

Yu, R., Zheng, S., Anandkumar, A., Yue, Y., 2019. Long-term Forecasting using Higher Order Tensor RNNs.

Zeng, A., Chen, M., Zhang, L., Xu, Q., 2022. Are Transformers Effective for Time Series Forecasting?

Zhang, G., Zhang, X., Feng, H., 2016. Forecasting financial time series using a methodology based on autoregressive integrated moving average and Taylor expansion. Expert Syst 33, 501–516. https://doi.org/10.1111/exsy.12164

Zhou, H., Zhang, Shanghang, Peng, J., Zhang, Shuai, Li, J., Xiong, H., Zhang, W., 2021. Informer: Beyond efficient transformer for long sequence time-series forecasting, in: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 11106–11115.

Zhou, H., Zhang, Shanghang, Peng, J., Zhang, Shuai, Li, J., Xiong, H., Zhang, W., 2020. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting.

Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L., Jin, R., 2022. FEDformer: Frequency Enhanced Decomposed Transformer for Long-term Series Forecasting, in: Chaudhuri, K., Jegelka, S., Song, L., Szepesvari, C., Niu, G., Sabato, S. (Eds.), Proceedings of the 39th International Conference on Machine Learning, Proceedings of Machine Learning Research. PMLR, pp. 27268–27286.

6

CONCLUSIONS

The objective of this study was to map and analyze the state-of-the-art in the development of RNN models for forecasting agricultural commodity time series, with a primary focus on the soybean crop. Initially, a systematic review was conducted, providing an important contribution to the field by consolidating and presenting relevant information on the topic. This information served as a basis for subsequent studies and guided the methodological choices adopted in this work.

Two experimental papers were selected for detailed analysis, in which RNN models, specifically LSTM and BiLSTM, were used. In the first paper, LSTM and BiLSTM models were applied to forecast soybean price time series, resulting in RMSE of 1.8002 and 0.8433, respectively. It is important to note that the hyperparameter values were arbitrarily selected without the aid of optimization algorithms. This lack of optimization directly impacted the interpretation of the prediction errors for each model, making it impossible to make any claims about which model could have achieved the best result.

The results presented in the second paper indicated that the BiLSTM model was potentially closer to the ideal hyperparameter values for the dataset compared to the LSTM model. However, this statement is hypothetical since the lack of optimization in the models of the first paper obscured the true potential of both models. Upon optimizing the hyperparameters, it was found that the LSTM model achieved an RMSE of 0.3695, and the BiLSTM model obtained 0.3720. The BiLSTM model showed an improvement of approximately 56% compared to the average error without optimization, while the LSTM model had a reduction of almost 80%, reaching a statistically equivalent error average to the BiLSTM model with a 95% confidence level. These results highlight the importance of using optimization algorithms as unfair judgment could occur without them, masking the potential of both models.

Hyperparameter optimization allows adjusting the model settings to better fit the data and capture relevant patterns in the time series. This results in more accurate and reliable predictions, providing valuable information for farmers to make informed decisions regarding the cultivation and marketing of agricultural commodities.

It is relevant to note that the experiments in the first paper were limited to one-step predictions, i.e., predicting only the next value in the time series. However, the results from the second experimental paper showed that the average error increases proportionally with the forecast horizon. This means that models without proper optimization may generate even larger errors as the forecast horizon is extended, rendering impractical predictions aimed at guiding decision-making.

In addition to the LSTM and BiLSTM models, this study also evaluated the performance of the Transformer model for forecasting agricultural commodity time series, including soybean cultivation data. The results highlighted that the Transformer model proved to be a superior option compared to recurrent models, especially for longer forecast horizons. This brings significant benefits to farmers as the use of the Transformer model reduces risks and enhances decision-making in the field.

Although the Transformer model may not present a significant advantage in forecasting the corn time series compared to other models, it is important to emphasize its adaptability across different domains, even if not originally designed for forecasting. This highlights the versatility of the Transformer model as a forecasting tool.

A relevant finding of this study is that all evaluated models showed an increase in error as the forecast horizon was extended. This result was expected but it is important to highlight the challenges faced in this domain. Therefore, it is crucial to find an optimal forecast horizon that minimizes errors while providing valuable information for farmers to make informed decisions.

In summary, this study contributed to advancing knowledge about the development of ANN models for forecasting agricultural commodity time series, with an emphasis on soybean cultivation. The results emphasized the importance of hyperparameter optimization, demonstrated the superiority of the Transformer model under certain conditions, and underscored the need to select an appropriate forecast horizon. These findings provide valuable insights for researchers and professionals involved in the field of agricultural time series forecasting.

In conclusion, this study contributed to advancing the knowledge on the development of RNN models for agricultural commodity time series forecasting. Hyperparameter optimization was identified as a crucial element to enhance model performance. Furthermore, the Transformer model stood out as a superior option under certain conditions, highlighting its utility in the field of agricultural time series forecasting.

As final considerations, it is recommended that future research further explore hyperparameter optimization in ANN models, not limited to agricultural commodity forecasting alone. Additionally, it is important to investigate the generalization of models to longer forecast horizons and other financial assets to better understand their adaptability in different contexts.

As an additional recommendation, the exploration of adding new variables that assist models in forecasting, introducing feature selection in conjunction with hyperparameter optimization, is suggested to further enhance the accuracy and efficiency of time series forecasting models. This allows for the identification of the most relevant variables for the forecasting process, eliminating those that have little impact or may introduce noise into the models. This can result in more robust and reliable models.

These advancements will contribute to improving agricultural time series forecasts and provide valuable insights for farmers to make more informed and strategic decisions in the cultivation and marketing of agricultural commodities.

REFERENCES

ABBASIMEHR, H.; SHABANI, M.; YOUSEFI, M. An optimized model using LSTM network for demand forecasting. **Computers and Industrial Engineering**, v. 143, 2020. Disponível em: https://doi.org/10.1016/j.cie.2020.106435

BALAJI, A. J.; RAM, D. S. H.; NAIR, B. B. Applicability of deep learning models for stock price forecasting an empirical study on BANKEX data. **Procedia computer science**, v. 143, p. 947–953, 2018.

BOOTE, D. N.; BEILE, P. Scholars Before Researchers: On the Centrality of the Dissertation Literature Review in Research Preparation. **Educational Researcher**, v. 34, n. 6, p. 3–15, 2005. Disponível em: https://doi.org/10.3102/0013189X034006003

CANIZO, M.; TRIGUERO, I.; CONDE, A.; ONIEVA, E. Multi-head CNN--RNN for multi-time series anomaly detection: An industrial case study. **Neurocomputing**, v. 363, p. 246–260, 2019.

CARVALHO, J. C. de; PAVAN, L. S.; HASEGAWA, M. M. Transmissões de volatilidade de preços entre Commodities agrícolas brasileiras. **Revista de Economia e Sociologia Rural**, v. 58, 2020.

CHAMBERS, M. J.; BAILEY, R. E. A Theory of Commodity Price Fluctuations. **Journal of Political Economy**, v. 104, n. 5, p. 924–957, 1996. Disponível em: https://doi.org/10.1086/262047

CHAMBON, S.; GALTIER, M. N.; ARNAL, P. J.; WAINRIB, G.; GRAMFORT, A. A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series. **IEEE Transactions on Neural Systems and Rehabilitation Engineering**, v. 26, n. 4, p. 758–769, 2018.

DA SILVA, D. B.; SCHMIDT, D.; DA COSTA, C. A.; DA ROSA RIGHI, R.; ESKOFIER, B. DeepSigns: A predictive model based on Deep Learning for the early detection of patient health deterioration. **Expert Systems with Applications**, v. 165, p. 113905, 2021.

DE MELO, G. A.; SUGIMOTO, D. N.; TASINAFFO, P. M.; SANTOS, A. H. M.; CUNHA, A. M.; DIAS, L. A. V. A new approach to river flow forecasting: LSTM and GRU multivariate models. **IEEE Latin America Transactions**, v. 17, n. 12, p. 1978–1986, 2019.

DIVISEKARA, R. W.; JAYASINGHE, G. J. M. S. R.; KUMARI, K. W. S. N. Forecasting the red lentils commodity market price using SARIMA models. **SN Business & Economics**, v. 1, n. 1, p. 20, 2021. Disponível em: https://doi.org/10.1007/s43546-020-00020-x

DREES, L.; JUNKER-FROHN, L. V; KIERDORF, J.; ROSCHER, R. Temporal prediction and evaluation of Brassica growth in the field using conditional generative adversarial networks. **Computers and Electronics in Agriculture**, v. 190, 2021. Disponível em: https://doi.org/10.1016/j.compag.2021.106415

DUKE, N. K.; BECK, S. W. Research News And Comment: Education Should Consider Alternative Formats for the Dissertation. **Educational Researcher**, v. 28, n. 3, p. 31–36, 1999. Disponível em: https://doi.org/10.3102/0013189X028003031

ESCUDERO, P.; ALCOCER, W.; PAREDES, J. Recurrent Neural Networks and ARIMA Models for Euro/Dollar Exchange Rate Forecasting. **Applied Sciences**, v. 11, n. 12, p. 5658, 2021. Disponível em: https://doi.org/10.3390/app11125658

GOEL, H.; SINGH, N. P. Dynamic prediction of Indian stock market: an artificial neural network approach. **International Journal of Ethics and Systems**, v. 38, n. 1, p. 35–46, 2022. Disponível em: https://doi.org/10.1108/IJOES-11-2020-0184

JU, C.-B.; HUNG, M.-C.; CHEN, A.-P. Market Profile with Convolutional Neural Networks: Learning the Structure of Price Activities. *In*: 2020, **2020 International Symposium on Computer, Consumer and Control (IS3C)**. : IEEE, 2020. p. 454–457.Disponível em: https://doi.org/10.1109/IS3C50286.2020.00123

JUNG, C. F. **Metodologia para pesquisa e desenvolvimento: aplicada a novas tecnologias, produtos e processos**. *[S. l.]*: Axcel Books, 2004.

JUNG, D.-H.; KIM, H. S.; JHIN, C.; KIM, H.-J.; PARK, S. H. Time-serial analysis of deep neural network models for prediction of climatic conditions inside a greenhouse. **Computers and Electronics in Agriculture**, v. 173, 2020. Disponível em: https://doi.org/10.1016/j.compag.2020.105402

KITCHENHAM, B. Procedures for performing systematic reviews. **Keele, UK, Keele University**, v. 33, n. 2004, p. 1–26, 2004.

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

MANOHAR, M.; KOLEY, E.; GHOSH, S.; MOHANTA, D. K.; BANSAL, R. C. Spatio-temporal information based protection scheme for PV integrated microgrid under solar irradiance intermittency using deep convolutional neural network. **International Journal of Electrical Power & Energy Systems**, v. 116, p. 105576, 2020.

MENSI, W.; BELJID, M.; BOUBAKER, A.; MANAGI, S. Correlations and volatility spillovers across commodity and stock markets: Linking energies, food, and gold. **Economic Modelling**, v. 32, p. 15–22, 2013. Disponível em:

https://doi.org/https://doi.org/10.1016/j.econmod.2013.01.023

MIDDI, A. R.; MIDDI, V. S. R. Currency Exchange Rate Prediction Using Multi-layer Perceptron. *In*: KUMAR A. SENATORE S., G. V. K. (org.). **Lecture Notes in Electrical Engineering**. *[S. l.]*: Springer Science and Business Media Deutschland GmbH, 2022. v. 783, p. 231–237. *E-book*. Disponível em: https://doi.org/10.1007/978-981-16-3690-5_20

MOHSIN KABIR, Md.; LIMA, A. A.; MRIDHA, M. F.; ABDUL HAMID, Md.; MONOWAR, M. M. Forecasting Closing Price of Stock Market Using LSTM Network: An Analysis from the Perspective of Dhaka Stock Exchange. *In*: **Lecture Notes on Data Engineering and Communications Technologies**. *[S. l.]*: Springer Science and Business Media Deutschland GmbH, 2022. v. 95, p. 289–299. *E-book*. Disponível em: https://doi.org/10.1007/978-981-16-6636-0_23

MORETTIN, P. A.; TOLOI, C. M. Análise de Séries Temporais—2ª Edição Revista e Ampliada. **ABE–Projeto Fisher, Editora Edgar Blücher**, 2006.

PAUL, R. K.; GARAI, S. Performance comparison of wavelets-based machine learning technique for forecasting agricultural commodity prices. **Soft Computing**, v. 25, n. 20, p. 12857–12873, 2021. Disponível em: https://doi.org/10.1007/s00500-021-06087-4

PETERSON, H. H.; TOMEK, W. G. How much of commodity price behavior can a rational expectations storage model explain? **Agricultural Economics**, v. 33, n. 3, p. 289–303, 2005. Disponível em: https://doi.org/https://doi.org/10.1111/j.1574-0864.2005.00068.x

QIAO, W.; YANG, Z. Forecast the electricity price of US using a wavelet transform-based hybrid model. **Energy**, v. 193, p. 116704, 2020.

ROBERTS, M. J.; SCHLENKER, W. Identifying Supply and Demand Elasticities of Agricultural Commodities: Implications for the US Ethanol Mandate. **American Economic Review**, v. 103, n. 6, p. 2265–2295, 2013. Disponível em: https://doi.org/10.1257/aer.103.6.2265

VERÍSSIMO, M. P.; XAVIER, C. L. Tipos de commodities, taxa de câmbio e crescimento econômico: evidências da maldição dos recursos naturais para o Brasil. **Revista de Economia Contemporânea**, v. 18, n. 2, p. 267–295, 2014.

WANG, S.; CHEN, J.; WANG, H.; ZHANG, D. Degradation evaluation of slewing bearing using HMM and improved GRU. **Measurement**, v. 146, p. 385–395, 2019.

YAN, H.; OUYANG, H. Financial time series prediction based on deep learning. **Wireless Personal Communications**, v. 102, n. 2, p. 683–700, 2018.

ZHANG, D.; CHEN, S.; LIWEN, L.; XIA, Q. Forecasting Agricultural Commodity Prices Using Model Selection Framework With Time Series Features and Forecast Horizons. **IEEE Access**, v. 8, p. 28197–28209, 2020. Disponível em: https://doi.org/10.1109/ACCESS.2020.2971591