

**AVALIAÇÃO DE MÉTODOS DE PREVISÃO DE PREÇOS PARA COMMODITIES
AGRÍCOLAS: EVIDÊNCIAS PARA SOJA E MILHO**
**EVALUATION OF PRICE FORECASTING METHODS FOR AGRICULTURAL
COMMODITIES: EVIDENCE FROM SOYBEANS AND CORN**

Fernanda Cigainski Lisbinski¹; Felipe André Oliveira Freitas²; Heloisa Lee Burnquist³

Filiação: Escola Superior de Agricultura “Luiz de Queiroz” (ESALQ – USP)

E-mail: lisbinskif@gmail.com¹; felipeoliveirafreitas@usp.br²; hlburnqu@usp.br³

Grupo de Trabalho (GT): GT01. Mercados agrícolas e comércio exterior

Resumo

Este estudo teve como objetivo desenvolver modelos de previsão para os Indicadores do milho ESALQ/BM&FBOVESPA e da soja CEPEA/ESALQ – Paranaguá que possam ser utilizados no processo de tomada de decisão para compra e venda dessas *commodities*. Foram adotadas abordagens tradicionais, como ARIMA, e técnicas de *Neural Networks*, como o LSTM para realizar previsões de preço de ambos os indicadores. A pesquisa avaliou o desempenho de diferentes modelos univariados, baseados apenas em séries temporais de preços, e multivariados, que incorporaram uma variável criada para captar a influência da oferta de soja e milho. Os resultados mostraram que o LSTM foi o modelo com melhor desempenho, superando o ARIMA em termos de acurácia, com menores erros de previsão, com MAE de 3,59, MAPE de 2,26 e RMSE de 3,89 para a soja e MAE de 0,48, MAPE de 0,82 e RMSE de 0,64 para o milho. Os modelos multivariados também se destacaram por capturar melhor as dinâmicas de mercado, entre eles o LSTM foi o que apresentou o melhor desempenho, com MAE de 0,65, MAPE de 0,46 e RMSE de 0,83 para a soja e MAE de 0,96, MAPE de 0,16 e RMSE de 0,13 para o milho. Concluiu-se que as *Neural Networks*, especialmente o LSTM, fornecem previsões mais precisas e robustas para previsão de preço mensal da soja e do milho.

Palavras-chave: Modelos Preditivos; Séries Temporais; Machine Learning, Inteligência Artificial, Neural Networks.

Abstract

This study aimed to develop forecasting models for the ESALQ/BM&FBOVESPA corn indicator and the CEPEA/ESALQ – Paranaguá soybean indicator, with the goal of supporting decision-making processes for buying and selling these commodities. Traditional approaches such as ARIMA and Neural Network techniques like LSTM were employed to forecast the prices of both indicators. The research evaluated the performance of different univariate models—based solely on historical price time series—and multivariate models, which incorporated a variable designed to capture the influence of soybean and corn supply. The results showed that LSTM was the best-performing model, outperforming ARIMA in terms of accuracy, with lower forecast errors: MAE of 3.59, MAPE of 2.26, and RMSE of 3.89 for soybeans; and MAE of 0.48, MAPE of 0.82, and RMSE of 0.64 for corn. Multivariate models also stood out by better capturing market dynamics, with LSTM once again delivering the best performance: MAE of 0.65, MAPE of 0.46, and RMSE of 0.83 for soybeans; and MAE of 0.96, MAPE of 0.16, and RMSE of 0.13 for corn. It was concluded that Neural Networks, particularly LSTM, provide more accurate and robust monthly price forecasts for soybeans and corn.

Key words: Predictive Models; Time Series; Machine Learning; Artificial Intelligence; Neural Networks.

1. Introdução

A previsão de preços é uma preocupação fundamental para os participantes do mercado no setor agrícola, tais como corretores, *hedgers*, processadores e produtores. É de extrema importância para os especuladores na geração de lucros, produtores na fixação dos preços de venda antes da produção e processadores e exportadores no atendimento das necessidades do mercado (Xu e Zhang, 2022). Os preços agrícolas apresentam volatilidade (Timmermann, 2006; Xu, 2017), exercem grandes influências na tomada de decisões, impactando na alocação de recursos e no bem-estar econômico (Xu, 2019), principalmente em países que apresentam uma economia impulsionada pelo setor primário, como o Brasil.

A previsão de preços auxilia também na gestão de suprimentos agrícolas e no comportamento de compra dos consumidores. Além da sazonalidade da produção, influenciada por fatores climáticos, os preços são impactados por preferências de usuários, estratégias comerciais, fatores culturais e comportamento dos fornecedores, que podem ser analisados com base em eventos passados (Kurumatani, 2020). Assim, o desenvolvimento de modelos de previsão de preços de *commodities* agrícolas é essencial para o cenário político-econômico, aumentando a segurança de investidores, produtores e formuladores de políticas, além de facilitar a criação de regulamentações agrícolas apropriadas (Zhao, 2021).

No Brasil, o setor agrícola desempenha um papel fundamental na economia do país. De acordo com o Centro de Estudos Avançados em Economia Aplicada (CEPEA/USP-ESALQ) e a Confederação Nacional da Agricultura (CNA) (2023), no ano de 2022, o Produto Interno Bruto (PIB) do agronegócio apresentou uma participação de 25% no PIB da economia brasileira e uma participação de 19,35% da população ocupada no Brasil (CEPEA, 2023). O ramo agrícola representou 72,7% do valor do PIB do agronegócio, enquanto o ramo pecuário foi responsável por 28,3%.

O grande destaque brasileiro é a produção de grãos. O Brasil é o quarto maior produtor mundial de grãos ficando atrás apenas dos Estados Unidos, China e Índia (*Food and Agriculture Organization* (FAO), (2023a). Na safra 2021/22 a produção de grãos no Brasil foi de 272,69 milhões de toneladas e uma área plantada de 74,5 milhões de hectares (Companhia Nacional de Abastecimento (CONAB), 2023). Os destaques são para a soja que representou 46% da produção total de grãos, milho com 41,19%, arroz com 3,96% e o trigo com 3,87% (CONAB, 2023).

Neste contexto, é possível encontrar na literatura diversas metodologias utilizadas para a construção de modelos de previsão de preços. A mais tradicional é a desenvolvida por Box e Jenkins (1976), a qual pode ser resumida em quatro etapas para a construção de um modelo de série temporal, a identificação, a estimação, o diagnóstico e a previsão. Contudo, diante aos avanços científicos quanto ao desenvolvimento de modelos e abordagens estatística para previsões do preço, paralelamente ao massivo aumento do aporte computacional nas últimas décadas, abordagens de Machine Learning (ML), como Deep Learning (DL) e Neural Network (NN) vem se apresentando como vantajosas e promissoras na melhoria dos modelos de predição. Tais técnicas apresentam grande eficiência e acurácia de predição, sendo mais favoráveis principalmente em cenários onde a série temporal apresenta comportamentos complexos e dinâmicos. É importante ressaltar que tais modelos podem trazer consigo aspectos relevantes de técnicas e metodologias de previsão já consolidadas em sua composição, além de ser amplamente adotados em análises que envolvem grandes quantidades de dados por sua eficiência (Hossain et al., 2018; Nabipour et al. 2020).

No que diz respeito aos modelos de previsão, além da metodologia tradicional desenvolvida por Box e Jenkins (1976), como os modelos ARIMA (*Autoregressive Integrated Moving Average*), abordagens como as *Artificial Neural Networks* (ANN) são capazes de identificar padrões mais complexos nos dados e produzir previsões precisas, mesmo em cenários de alta volatilidade. Este é um dos algoritmos de machine learning mais utilizados, abrangendo vários campos de classificação de dados, previsão, otimização, reconhecimento de modelo e regressão (Bakay e Ağbulut, 2021). Além do modelo ANN, existem os modelos de *Recurrent Neural Network* (RNN) que são projetados para lidar com dados sequenciais, onde as informações de entradas anteriores são usadas para influenciar a saída atual, tornando-o adequado para séries temporais e dados dependentes do tempo. Sua arquitetura permite manter "memória" ao longo da sequência, o que o diferencia das NN tradicionais (Goodfellow et al., 2016).

A dificuldade em desenvolver modelos de previsão para os preços da soja e do milho, principais *commodities* produzidas no Brasil, devido a volatilidade do preço dos mesmos e a sensibilidade à fatores como clima, políticas governamentais e a demanda global é o que justifica a escolha do tema abordado neste estudo. Essa dificuldade, influencia os produtores e exportadores agrícolas em seu processo de tomada de decisão sobre o plantio, a colheita e a comercialização. Assim, a importância deste trabalho, consiste em possibilitar aos produtores, a indústria, aos corretores e demais interessados o acesso a estimativas informativas, capaz de auxiliá-los na gestão de seus negócios, diminuindo desperdícios e maximizando seus lucros. Por fim, esse trabalho torna-se relevante para a literatura por utilizar métodos estatísticos de séries temporais baseados em modelos de *machine learning* que apresentam aplicação recente e são caracterizados pela sua maior complexidade nas estimações e melhor adesão aos comportamentos sazonais, ruidosos e não estacionários.

Nesse contexto, a hipótese dessa pesquisa é a de que modelos de previsão utilizando técnicas de NNs podem ser mais precisos do que os modelos tradicionais, como ARIMA. Além disso, supõe-se que os modelos de previsão multivariados, que incorporam dados históricos de preços, produção e demanda, tendem a ser mais precisos do que os modelos univariados, que utilizam apenas uma série temporal. Diante disso, o objetivo deste estudo foi desenvolver modelos de previsão para os Indicadores do milho ESALQ/BM&FBOVESPA e da soja CEPEA/ESALQ – Paranaguá que possam ser utilizados no processo de tomada de decisão para compra e venda dessas *commodities*. Os objetivos específicos incluem: estimar modelos de previsão tradicionais, como ARIMA; aplicar técnicas de NNs, incluindo o modelo LSTM que é um modelo de RNN, para prever Indicadores do milho ESALQ/BM&FBOVESPA e da soja CEPEA/ESALQ – Paranaguá considerando a série histórica de mensal de janeiro de 2013 a julho de 2024, para um horizonte de previsão de 2 meses; estimar modelos univariados e multivariados, verificando a precisão de cada abordagem na previsão desses indicadores agrícolas, para um horizonte de previsão de 2 meses; avaliar o desempenho dos diferentes modelos utilizando índices de seleção, como AIC e BIC, além de critérios estatísticos de validação de acurácia, para um horizonte de previsão de 2 meses; e realizar ajustes nos modelos conforme necessário, para oferecer ferramentas eficazes que apoiem decisões de negócios no mercado dessas *commodities* agrícolas.

2. Materiais e Métodos

2.1 Levantamento de dados

Os modelos foram estimados no formato univariado e multivariado, visando analisar quais os melhores modelos de estimação para cada indicador. Diante disso, os dados foram extraídos de diversas fontes, com frequência mensal. Previamente os dados foram tratados e realizou-se uma análise exploratória a fim de verificar a estrutura dos dados, possíveis quebras e comportamentos incomuns como *outliers*. As séries históricas utilizadas compreenderam os anos de janeiro de 2013 a julho de 2024. O ano de início foi escolhido devido à disponibilidade das séries que foram analisadas. Na Tabela 1, apresentam-se as variáveis que foram utilizadas no estudo.

Tabela 1. Variáveis utilizadas nesse estudo

Soja				Milho			
Variável		Fonte		Variável		Fonte	
Indicador	da	Soja	Cepea	Indicador	do	Milho	Cepea
ESALQ/BM&FBOVESPA-			USP/ESALQ	ESALQ/BM&FBOVESPA			USP/ESALQ
Paranaguá			(2024a)				(2024b)
Estimativas de Safra			Conab (2023)	Estimativas de Safra			Conab (2023)

Importação de soja	MDIC (2024)	Importação de milho	MDIC (2024)
Estoque inicial de soja	Conab (2023)	Estoque inicial de milho	Conab (2023)

Fonte: Elaborados pelos autores, 2024.

Nota: A Nomenclatura Comum do Mercosul (NCM) utilizada para coletar as importações de soja do Ministério da Indústria, Comércio Exterior e Serviços (MDIC) foram: 12011000; 12019000. Para as importações de milho foram 10051000; 10059010; 10059090.

O Indicador da Soja ESALQ/BM&FBOVESPA-Paranaguá e o Indicador do Milho ESALQ/BM&FBOVESPA foram ajustados pela inflação utilizando o IGP-DI (Índice Geral de Preços – Disponibilidade Interna), calculado pela Fundação Getúlio Vargas (FGV). Esse procedimento, conhecido como deflacionamento, permite que os valores das *commodities* sejam comparados ao longo do tempo em termos reais, eliminando o efeito da inflação e proporcionando uma análise mais precisa do comportamento dos preços (Mankiw, 2021).

Utilizando as variáveis da Tabela 1, foram construídas as variáveis de oferta de soja e milho as quais irão compor o modelo multivariado, conforme a eq. (1):

$$Oferta_t = Estimativa\ de\ safra_t + Importação_t + Estoque\ Inicial_t \quad (1)$$

Posteriormente, foi realizada a transformação pela Z-score da variável, que é uma técnica usada para estabilizar a variância em séries temporais e garantir que os dados estejam em uma escala comparável. A Z-score é obtida a partir da eq. (2):

$$Z = \frac{X - \mu}{\sigma} \quad (2)$$

onde Z é o valor da série normalizado (Z-score), X é o valor original da série temporal, μ é a média da série temporal, σ é o desvio padrão da série temporal. Assim, os modelos de previsão multivariados incluíram o indicador de preço da commodity como variável predita, além de uma variável adicional que representa a oferta pelo produto.

Antes de realizar as estimativas de previsão, foi necessário executar alguns testes preliminares. O primeiro foi o teste de raiz unitária, utilizado para verificar a estacionariedade das variáveis. Neste caso, foi aplicado o teste KPSS (Kwiatkowski-Phillips-Schmidt-Shin) de Kwiatkowski *et al.* (1992), no qual a não rejeição da hipótese nula indica a estacionariedade da série. A presença de raiz unitária pode sugerir uma relação econométrica espúria, o que exige correção.

Além disso, foi essencial realizar uma análise da sazonalidade da série temporal. Essa análise é feita a partir da observação do comportamento histórico da série e da aplicação da decomposição *Seasonal and Trend decomposition using Loess* (STL). A decomposição STL separa a série em seus componentes de tendência, sazonalidade e ruído, permitindo uma visão clara dos padrões sazonais recorrentes. Essa análise é particularmente importante em séries temporais de *commodities* agrícolas, pois permite capturar com maior precisão as flutuações periódicas, garantindo que os modelos de previsão, como ARIMA e RNN, possam representar adequadamente essas variações sazonais e melhorar a qualidade das previsões.

2.2 O modelo ARIMA

O modelo ARIMA é um modelo de série temporal que faz uso apenas da variável dependente e suas respectivas defasagens. Isso significa que o modelo não incorpora nenhuma relação econômica de outras variáveis para explicar a variável dependente. Em vez disso, o modelo estima a regressão da variável dependente em função de suas próprias defasagens (Box e Jenkins, 1976). Um modelo ARIMA consiste na combinação de uma parte autorregressiva (Autoregressive (AR), a média móvel (*Moving Average* (MA) e o diferencial. Cada modelo ARIMA é especificado por seus valores *p* (a ordem do modelo AR), *d* (o grau de diferenciação

não sazonal) e q (a ordem do modelo MA) (Box, *et al.*, 2015). Diante disso, Morettin (2011), descreve o modelo ARMA (p, q) conforme eq. (3):

$$\Delta^d X_t = \theta_0 + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (3)$$

Onde $\Delta^d X_t$ representa a diferenciação da série original para remover tendências; $\theta_0 = \mu(1 - \phi_1 - \dots - \phi_p)$, representa uma constante no modelo estimado; $\phi_1 \dots \phi_p$ são os parâmetros que ajustam valores passados de X_t do instante imediatamente anterior até o mais distante representado por p . Os valores de ε , o termo de erro da série, e ε_t é uma proporção não controlável do modelo chamado ruído branco. Os parâmetros $\theta_1 \dots \theta_q$ possibilitam escrever a série em função dos choques passados. Além disso, no geral pressupõe-se $\varepsilon_t \sim N(\mu, \sigma^2)$.

Para os modelos ARIMA, seguiu-se o método Box e Jenkins (1976) o qual consiste em um processo composto de uma série de etapas para determinar a ordem de estimação do modelo. Isso inclui a análise das funções de autocorrelação (ACF) e autocorrelação parcial (PACF) para identificar os parâmetros p, d e q do modelo. Para o diagnóstico do modelo será utilizado os critérios de AIC, BIC e SBC, onde o melhor modelo será aquele que apresentar o menor valor.

Além disso, os modelos ARIMA, podem incluir a sazonalidade, resultando em um modelo SARIMA, que é essencial para lidar com séries temporais que apresentam padrões cíclicos repetitivos. O modelo é representado como $ARIMA(p, d, q)(P, D, Q)_m$, onde P, D, Q são os parâmetros sazonais que controlam a autoregressão, diferenciação e média móvel sazonal, respectivamente, e m é o período da sazonalidade (como 12 para dados mensais com sazonalidade anual) (Box *et al.*, 2015).

No formato multivariado, utilizou-se o modelo *AutoRegressive Integrated Moving Average with Exogenous variables* (ARIMAX) é uma extensão do modelo ARIMA, incorporando a variável de oferta como exógena, influenciando a variável dependente. Ele é útil quando se deseja modelar séries temporais com padrões autoregressivos e de média móvel, enquanto se considera o impacto de fatores externos (variáveis explicativas) na previsão (Hyndman e Athanasopoulos, 2018).

Por fim, destaca-se que neste trabalho também foi utilizado o modelo ARIMA/XGBoost, o qual combina a abordagem clássica de séries temporais com uma técnica avançada de *machine learning*. Inicialmente, o modelo ARIMA é utilizado para capturar padrões lineares da série temporal e gerar previsões preliminares. Em seguida, os erros (resíduos) desse modelo são utilizados como entrada para o XGBoost, um algoritmo baseado em árvores de decisão que consegue identificar padrões não lineares e interações entre variáveis. Esse modelo híbrido permite que os componentes lineares sejam explicados pelo ARIMA, enquanto o XGBoost aprende relações complexas nos resíduos, resultando em previsões mais precisas e robustas. Tal abordagem é especialmente útil em cenários onde a série temporal possui influências externas e comportamentos não capturados por modelos puramente estatísticos (Wang e Guo, 2020).

2.3 O modelo Erro, Tendência e Sazonalidade (ETS)

O modelo ETS é uma abordagem estatística utilizada para previsão de séries temporais. Desenvolvido a partir dos métodos exponenciais tradicionais, o modelo ETS fornece uma estrutura sistemática para a modelagem de séries temporais ao considerar explicitamente três componentes principais: erro (E), tendência (T) e sazonalidade (S) (Hyndman e Athanasopoulos, 2018).

A especificação do modelo ETS pode ser representada pela combinação de diferentes formas desses três componentes, resultando em variações como ETS (A,A,N), onde “A” indica um componente aditivo e “N” representa a ausência do respectivo componente. O modelo ETS pode assumir tendências aditivas ou multiplicativas e considerar a sazonalidade de forma semelhante. Neste trabalho o modelo ETS assumiu tendências multiplicativas e considerou a

tendência e a sazonalidade no modelo. Por fim, ressalta-se que a formulação geral do ETS segue a estrutura dos métodos de suavização exponencial, que ajustam os valores previstos com base em pesos decrescentes ao longo do tempo (Hyndman e Athanasopoulos, 2018).

2.4 O modelo *PROPHET*

O modelo *Prophet* é uma abordagem baseada em decomposição para previsão de séries temporais, projetada para lidar com padrões sazonais, feriados e tendências não lineares. Ele é utilizado devido à sua facilidade de implementação e robustez na modelagem de dados não lineares (Taylor e Letham, 2018). O *Prophet* assume que uma série temporal pode ser decomposta em três componentes principais, conforme eq. (4).

$$y(t) = g(t) + s(t) + h(t) + \epsilon(t) \quad (4)$$

onde, $y(t)$ é a série a ser prevista; $g(t)$ representa a tendência, modelada como um crescimento linear ou logístico com pontos de mudança; $s(t)$ é a sazonalidade, capturada por funções periódicas; $h(t)$ corresponde a efeitos de feriados ou eventos especiais; $\epsilon(t)$ é um termo de erro.

Uma das principais vantagens desse modelo é sua capacidade de lidar automaticamente com lacunas nos dados, ruído e variações estruturais, tornando-o adequado para aplicações em negócios, economia e ciência de dados (Taylor e Letham, 2018).

2.5 O modelo Multi-Layer Perceptrons Neural Network (MLP)

O modelo MLP é uma ANN. Um modelo ANN pode conter várias camadas de nós, no entanto, três camadas são as mais comuns para previsão. A primeira camada representa as variáveis explicativas, chamada de *inputs*; a segunda camada é a oculta, onde é realizado o processamento das informações de entrada; e a última camada consiste na variável prevista, os *outputs* (Torres-Pruñón et al. 2022; Zhang et al., 1998).

A MLP é composta por uma ou mais camadas ocultas entre as camadas de entrada e saída de uma rede. Assim, uma MLP apresenta boa capacidade de aprender o padrão de entrada e produzir uma saída que pode ser superior a NN de camada única.

Uma forma de NNs frequentemente utilizada para previsão é o *feedforward multilayer perceptron*. Nesse modelo a previsão de um passo à frente $\hat{y}_{t+1}$ é calculado usando entradas que são observações defasadas da série temporal ou outras variáveis explicativas. O I denota o número de entradas p_i da NN. Esse modelo pode ser definido a partir da eq. (5):

$$\hat{y}_{t+1} = \beta_0 + \sum_{h=1}^H \beta_h g \left(\gamma_{0i} + \sum_{i=1}^I \gamma_{hi} p_i \right) \quad (5)$$

onde, $\omega = (\beta, \gamma)$ são os pesos da rede com $\beta = [\beta_1, \dots, \beta_H]$, $\gamma = [\gamma_{11}, \dots, \gamma_{HI}]$ para a saída e as camadas ocultas, respectivamente. O β_0 e γ_{0i} são os vieses de cada neurônio, que para cada neurônio agem de forma semelhante à interceptação em uma regressão. H é o número de nós ocultos na rede e $g(\cdot)$ é uma função de transferência não linear, que normalmente é uma função logística sigmoide ou uma função tangente hiperbólica. As NNs podem modelar interações entre entradas, se houver. As saídas dos nós ocultos conectam-se a um nó de saída que gera a previsão. O nó de saída é normalmente linear (Crone e Kourentzes, 2010; Kourentzes et al., 2019). Nesta pesquisa utilizou-se 2 camadas ocultas para a estimação do modelo MLP.

2.6 O modelo Long Short-Term Memory (LSTM)

O modelo LSTM foi desenvolvido por Hochreiter e Schmidhuber (1997), é um tipo de RNN frequentemente usada em pesquisa de dados sequenciais, utilizando a técnica de *deep learning*. O LSTM é um modelo de *deep learning*, e mantém a informação de acordo com a importância do conteúdo, que pode ser mantido ou eliminado com a ajuda do portão de entrada e do portão de esquecimento, respectivamente. Normalmente a arquitetura do modelo contém três portões: o portão de entrada, o portão de saída e o portão de esquecimento (Hochreiter e Schmidhuber, 1997).

No modelo LSTM as portas de entrada e saída de uma célula de memória controlam os fluxos de entrada e saída usados para ativar a célula. Quando um valor auto recorrente deixa de ser relevante, um portão de esquecimento é aplicado para redefini-lo. A porta de esquecimento multiplica os valores de 0 e 1 com uma célula de memória para remover ou manter o valor para a próxima etapa (Swapna *et al.*, 2018). Este método permite que a rede se lembre de coisas por muito tempo. O recurso mais relevante do LSTM é a sua capacidade aprimorada de registrar dependências duradouras e dados de séries temporais. Ao alterar o estado da célula e o estado da célula da memória atual, esses portões trabalham juntos para regular como a célula de memória irá comportar (Pathinarupothi, *et al.*, 2017). A Figura 1 demonstra graficamente o comportamento de um modelo LSTM.

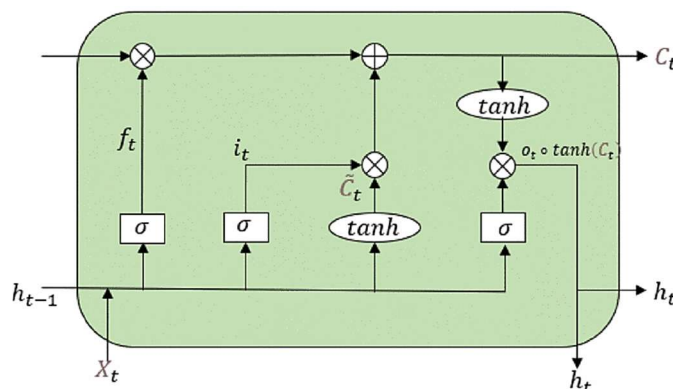


Figura 1. Diagrama do Modelo LSTM

Fonte: Govind e Babu, 2023

As portas têm a seguinte fórmula dada a série temporal de entrada X_t e o número de unidades ocultas como h , conforme eq. (8), (9), (10), (11), (12) e (13):

$$f_t = \sigma(X_t U_f + h_{t-1} W_f + b_f) \quad (7)$$

$$i_t = \sigma(X_t U_i + h_{t-1} W_i + b_i) \quad (8)$$

$$o_t = \sigma(X_t U_o + h_{t-1} W_o + b_o) \quad (9)$$

$$\check{C}_t = \tanh(X_t U_c + h_{t-1} W_c + b_c) \quad (10)$$

$$C_t = f_t * C_{t-1} + i_t * \check{C}_t \quad (11)$$

$$h_t = o_t * \tanh(C_t) \quad (12)$$

onde σ é a função de ativação sigmoide; f_t , i_t e o_t são as portas de esquecimento, de entrada e de saída, respectivamente. U_f , U_i , U_o e W_f , W_i , W_o são os fatores de pesos e b_f , b_i , b_o são os fatores de viés do portão de esquecimento, de entrada e de saída, respectivamente. C_t é o estado da célula. U_c e W_c são métricas específicas de peso e b_c é a variável de viés do estado da célula. C_t é baseado em dados de saída da célula de memória C_{t-1} e o passado do tempo atual $\check{C}_t$. O $\tanh$ é uma função de ativação (Zhang *et al.*, 2019; Govind e Babu, 2023). Nesta pesquisa utilizou-se 3 camadas ocultas para a estimação do modelo LSTM.

2.7 Critérios de seleção e métricas de desempenho

Os critérios de desempenho de predição dos modelos incluíram quatro métricas principais: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), e Root Mean Square Error (RMSE). A descrição das fórmulas desses indicadores está descrita no Tabela 2, conforme eq. (11), (12) e (13).

Tabela 2. Descrição dos critérios de desempenho de predição dos modelos

Indicador	Descrição
$MAE = \frac{1}{n} \sum_{t=1}^n y_i - \hat{y}_i \quad (11)$	onde n é o número de observações, y_i e $\hat{y}_i$ representam os valores de preços reais e previstos no tempo $t(t = 1, \dots, n)$, respectivamente. O MAE é uma medida de erro que é calculada a partir da média das diferenças absolutas entre os valores previstos e os valores reais. O MAE é expresso nas mesmas unidades que os dados originais, o que significa que ele fornece informações sobre a distância média entre as previsões e os valores reais (Govind e Babu, 2023).
$MAPE = \frac{1}{n} \sum_{t=1}^n \left \frac{y_i - \hat{y}_i}{y_i} \right \quad (12)$	O MAPE mede a magnitude média do erro em um conjunto de previsões e mostra os desvios em porcentagem (Vancsura <i>et al.</i> , 2023). O MAPE é uma medida de quão longe, em média, os valores previstos estão dos valores reais em termos percentuais.
$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_i - \hat{y}_i)^2} \quad (13)$	O RMSE mostra uma estimativa dos resíduos entre os valores reais e previstos (Vancsura <i>et al.</i> , 2023). O RMSE é uma medida de quão longe, em média, os valores previstos estão dos valores reais em termos absolutos.

Fonte: Elaborados pelos autores, 2024.

Para avaliar a precisão do modelo, a base de dados foi dividida entre um conjunto de dados para treino e outro conjunto de dados para validação, o método é chamado de *cross validation*. Existem vários métodos de *cross validation* que podem ser usados para séries temporais, o mais comum, e o que foi utilizado neste trabalho é o método *k-fold cross-validation*. Neste método, todos os dados disponíveis são particionados aleatoriamente em k conjuntos, posteriormente, todo o procedimento de treinamento ou ajuste do modelo, bem como o cálculo do erro é realizado k vezes, sendo cada conjunto uma vez utilizado como conjunto de teste, e os demais conjuntos utilizados para a construção do modelo. Assim, o método adquire k realizações independentes da medida de erro, e todos os dados são usados tanto para treinamento quanto para teste. De modo que, a média do k medidas de erro obtidas produzirá uma medida de erro geral que, normalmente, é mais robusta do que as medidas individuais (Arlot e Celisse, 2010; Bergmeir e Benítez, 2012). Diante disso, o conjunto de treinamento é usado para ajustar o modelo, enquanto o conjunto de validação foi usado para avaliar a precisão do modelo.

Além disso, foram considerados uma previsão de 2 passos à frente, ou seja, uma previsão de até 2 meses. Os menores valores dessas estatísticas definirão o modelo mais adequado.

3. Resultados e Discussão

3.1 Panorama da produção e comercialização de soja e milho no Brasil

Na safra 2021/2022 o Brasil foi o maior produtor mundial de soja, com uma produção de 125,54 milhões de toneladas (CONAB, 2023), o equivalente a cerca de 35% dos 360,14 milhões de toneladas de soja produzidos em todo o mundo neste período (United States Department of Agriculture (USDA), 2023), correspondendo a uma parte da oferta mundial do produto. A área plantada na safra 2021/22 no país foi de 41,49 milhões de hectares, cerca de

48,20% da área cultivada de lavouras temporárias, e a produtividade atingiu a média de 3.525 kg/ha. A produção da cultura concentra-se principalmente nas regiões Centro-Oeste e Sul, que representaram, na safra 2021/22, 54,26% e 18,63% do total produzido no Brasil, respectivamente (Produção Agrícola Municipal - Instituto Brasileiro de Geografia e Estatística (IBGE), 2023; CONAB, 2023).

A soja tem uma grande importância na economia brasileira, sendo um dos principais impulsionadores do agronegócio no país. De acordo com dados divulgados pelo Cepea em parceria com a Associação Brasileira das Indústrias de Óleos Vegetais (ABIOVE) (Barros *et al.*, 2023) o PIB da cadeia produtiva da soja e do biodiesel (grão, óleo, farelo, proteína de soja, glicerol e biodiesel) registrou um valor de R\$ 673,7 bilhões em 2022, o que representou cerca de 27% de todo o PIB do agronegócio nacional. Em 2010, esta participação representava apenas 9%. Entre os anos de 2010 e 2022, o PIB dessa cadeia apresentou um crescimento de 58%. Ressalta-se que nesse mesmo período, o agronegócio expandiu 8% e a economia 12%. No ano de 2022, essa cadeia produtiva gerou 2,05 milhões de ocupações, 80% a mais no comparativo com o ano de 2012 quando se deu início à série (Barros *et al.*, 2023), demonstrando a dimensão dessa cultura no Brasil.

Com relação às exportações, representando o lado da demanda, destaca-se que o principal destino de embarque da soja em grão brasileira é a China. Em 2022, a China absorveu 68% das exportações do grão, destacam-se ainda Espanha (4,20%), Tailândia (3,59%) e Irã (2,86%). A região Centro-oeste exporta 49% dessa quantidade, o estado do Mato Grosso é responsável por 31,46% do total exportado pelo Brasil, seguido dos estados de Goiás (12,74%) e Mato Grosso do Sul (4,54%). A região Sul ocupa a segunda posição como maior exportadora com 15% desse volume, o estado do Rio Grande do Sul é responsável por 6,83% das exportações totais de soja em grão brasileira, seguido do Paraná (6,4%) e de Santa Catarina (1,30%) (MDIC, 2023).

O Brasil foi o terceiro maior produtor mundial de milho no ano de 2021, ficando atrás dos Estados Unidos e da China (FAO, 2023a), demonstrando seu grande potencial de oferta no mercado global. Na safra 2021/2022 o Brasil produziu 113 milhões de toneladas (CONAB, 2023), o equivalente a cerca de 10% dos 1129,88 milhões de toneladas de milho produzidos em todo o mundo neste período (USDA, 2023). A área plantada na safra 2021/22 no país foi de 21,58 milhões de hectares, cerca de 25% da área cultivada de lavouras temporárias, e a produtividade atingiu a média de 5.242 kg/ha. A produção da cultura concentra-se principalmente nas regiões Centro-Oeste e Sul, que representaram, na safra 2021/22, 56,8% e 19% do total produzido no Brasil, respectivamente (IBGE, 2023; CONAB, 2023).

Para a Empresa Brasileira de Pesquisa Agropecuária (Embrapa) (2021) a importância econômica do milho pode ser caracterizada pelas suas várias formas de utilização, contemplando desde a alimentação animal até a indústria de alta tecnologia. Assim, o uso do milho em grão na forma de alimentação animal representa a maior parcela do consumo desse cereal, cerca de 70% no mundo. Ressalta-se que nos Estados Unidos, por exemplo, cerca de 50% são destinados a essa finalidade, enquanto no Brasil esse percentual varia de 60% a 80%, tendo dependência da fonte da estimativa e de ano para ano. Destaca-se ainda, que apesar da alimentação humana não apresentar uma participação muito grande no uso de milho em grão, ele se constitui como um fator alimentício de grande importância em regiões de baixa renda, como no Nordeste do Brasil, onde o milho é a fonte de energia para muitas pessoas que vivem no Semiárido sendo muitas vezes a ração diária de alimentação (Embrapa, 2021), o que demonstra a importância dessa cultura para a economia brasileira.

Quanto às exportações, os principais destinos do milho brasileiro em 2022 foram o Irã (15,22%), Japão (11,4%), Espanha (11,25%), Egito (9,5%) e Colômbia (5,65%). A região Centro-oeste exporta 72,3% dessa quantidade, o estado do Mato Grosso é responsável por

56,8% do total exportado pelo Brasil, seguido dos estados de Mato Grosso do Sul (7,71%) e Goiás (7,69%). A região Nordeste ocupa a segunda posição como maior exportadora com 8,13% desse volume, com destaque para o estado do Maranhão que é responsável por 4,77% das exportações totais brasileira de milho (MDIC, 2023).

Segundo Sanches e Sales (2022), atualmente, o Brasil não exporta grande quantidade de milho para a China (2,68%). No entanto, o país pode se tornar um importante parceiro comercial nos próximos anos, visto que recentemente o país asiático manifestou interesse pelo produto nacional. No entanto, para que as exportações ocorram, é necessário que as negociações em relação a aspectos fitossanitários e regulação de eventos biotecnológicos da produção nacional avancem. Segundo dados do USDA (2022a), as importações chinesas passaram de 7,6 milhões de toneladas na safra 2019/20 para 29,5 milhões de toneladas na safra 2020/21 e 23 milhões de toneladas na safra 2021/22, com uma perspectiva de importar 18 milhões de toneladas na safra 2022/23, se consolidando como principal importador mundial de milho nos últimos anos (USDA, 2022a).

Portanto, de acordo com a teoria econômica, o preço de um bem é determinado pela interação entre a oferta e a demanda no mercado. No caso da soja, por exemplo, o Brasil foi o maior produtor mundial na safra 2021/2022, contribuindo para a oferta global do produto, o que influencia diretamente o preço da soja no mercado internacional. Ao mesmo tempo, a demanda elevada de países como a China, que absorveu 68% das exportações de soja em grão do Brasil em 2022, reforça a importância de entender a elasticidade da demanda e como ela pode afetar os preços. Da mesma forma, no caso do milho, a oferta brasileira de 113 milhões de toneladas na safra 2021/2022, associada a uma demanda global crescente, especialmente de mercados como o Irã e o Japão, segue os princípios da teoria de oferta e demanda, onde variações na produção ou nas necessidades dos consumidores impactam diretamente os preços de mercado.

3.3 Análise das séries da soja e do milho

A Figura 2 oferece uma visão geral do comportamento temporal ao longo dos anos. Com relação a análise da soja, Figura 2(a), é possível visualizar que entre 2015 e 2019, os preços da soja apresentaram relativa estabilidade, com uma leve tendência de alta no final do período. Esse comportamento foi influenciado pela expansão da produção global, especialmente no Brasil e nos Estados Unidos, os dois maiores produtores mundiais. No entanto, as tensões comerciais entre os Estados Unidos e a China, iniciadas em 2018, trouxeram incertezas ao mercado global, criando variações temporárias nos preços devido à especulação sobre possíveis restrições comerciais (USDA, 2022b). Apesar disso, a oferta foi suficiente para atender à demanda crescente, mantendo os preços controlados.

Já em 2020, a pandemia de COVID-19 causou uma mudança nas cadeias de suprimento globais, afetando a produção e a logística de exportação de soja. Esse cenário impulsionou os preços, já que muitos países aumentaram a demanda por alimentos e rações em meio às incertezas globais (FAO, 2020). Além disso, condições climáticas adversas, como secas nas principais regiões produtoras, como o Brasil, reduziram a oferta, o que intensificou a pressão sobre os preços (FAO, 2021a). Durante 2020 e 2021, os preços da soja atingiram níveis recordes, impulsionados pela combinação de aumento da demanda e dificuldades na oferta.

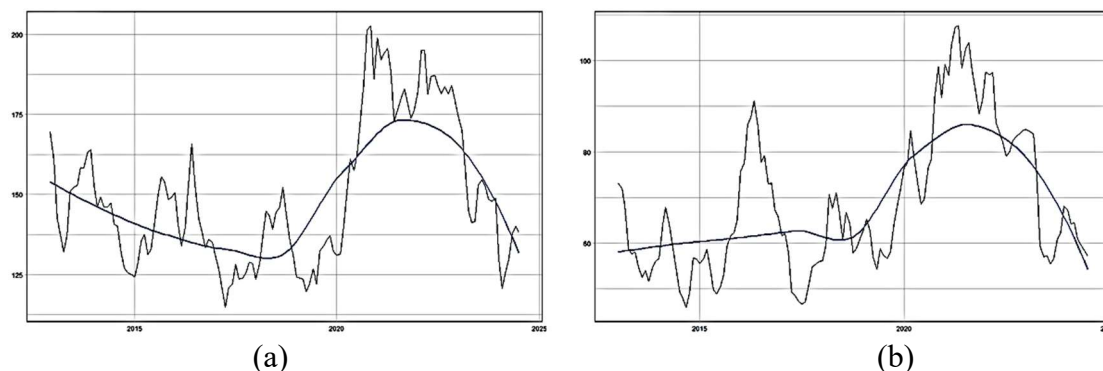


Figura 2. Indicador mensal real da Soja ESALQ/BM&FBOVESPA - Paranaguá (a) e Indicador mensal real do Milho ESALQ/BM&FBOVESPA (b)

Fonte: Elaborados pelos autores, 2024.

Nota: deflacionados pelo IGP-DI-FGV

Após o pico em 2021, os preços da soja começaram a recuar em 2022, conforme a normalização das cadeias de suprimento global começou a ocorrer. No entanto, a recuperação foi lenta devido às persistentes condições climáticas adversas em regiões produtoras importantes, o que manteve uma pressão sobre a oferta (*Organisation for Economic Co-operation and Development* (OECD), 2020). Além disso, os custos crescentes de insumos e fertilizantes, consequência dos impactos globais da covid-19 e conflitos na Ucrânia, mantiveram a volatilidade dos preços. No contexto de recuperação econômica pós-pandemia, os preços da soja ainda refletem a instabilidade no mercado agrícola global.

Assim como a soja, os preços do milho, Figura 2 (b), entre 2015 e 2019 permaneceram relativamente estáveis, com flutuações moderadas. A oferta global conseguiu acompanhar a crescente demanda por milho, tanto para consumo quanto para a produção de biocombustíveis, especialmente o etanol (OECD, 2020). No entanto, eventos climáticos, como secas nos Estados Unidos e na América do Sul, causaram picos temporários nos preços. A partir de 2018, a demanda por biocombustíveis também aumentou, impulsionando gradualmente os preços (FAO, 2021b).

Entre 2020 e 2021, a pandemia de COVID-19 afetou drasticamente o mercado de milho, com interrupções nas cadeias de suprimento, aumento dos custos de transporte e menor produção em regiões importantes, como os Estados Unidos (FAO, 2020). Esses fatores, somados à crescente demanda por milho para produção de rações e biocombustíveis, resultaram em um aumento acentuado dos preços. A seca severa que atingiu as principais áreas produtoras nos EUA e no Brasil em 2020 também impactou diretamente a oferta, o que elevou ainda mais os preços do milho (CONAB, 2023; USDA, 2023).

Em 2022, os preços do milho começaram a mostrar sinais de estabilização, embora permanecessem voláteis devido às incertezas climáticas e ao aumento dos custos de insumos agrícolas. A normalização das cadeias de suprimento globais e a recuperação gradual da produção nos principais países exportadores ajudaram a conter os aumentos extremos dos preços (FAO, 2021b). No entanto, a demanda por biocombustíveis e a contínua pressão dos altos custos de produção mantiveram os preços em níveis relativamente elevados, refletindo a persistente volatilidade no mercado global.

Na Figura 3 é possível verificar a análise das funções ACF e PACF dos indicadores de preço analisados, indicando que as séries podem ser não estacionárias, já que a ACF decresce gradualmente em ambos os casos, sugerindo a presença de uma tendência ou de um componente de raiz unitária. Esse comportamento indica que os valores atuais ainda estão correlacionados com os valores passados, sugerindo a necessidade de diferenciação para tornar a série

estacionária. A PACF, por sua vez, apresenta um pico no lag 1 nas duas séries, seguido por uma queda abrupta, o que é típico de um processo autoregressivo (AR) de ordem 1. Portanto, é necessário realizar o teste de raiz unitária para verificar a estacionariedade nas séries analisadas.

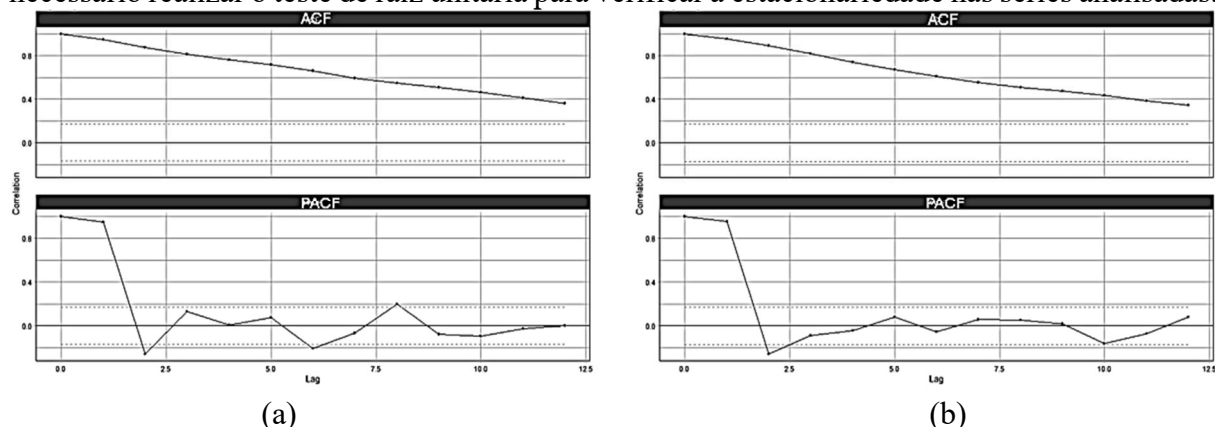


Figura 3. Funções de autocorrelação (ACF) e autocorrelação parcial (PACF) – (a) Indicador da Soja ESALQ/BM&FBOVESPA – Paranaguá e (b) Indicador do Milho ESALQ/BM&FBOVESPA

Fonte: Elaborados pelos autores, 2024.

Os resultados do teste KPSS, Tabela 3, indicaram que o Indicador da Soja ESALQ/BM&FBOVESPA – Paranaguá e o Indicador do Milho ESALQ/BM&FBOVESPA não são estacionários em sua forma original, sugerindo a presença de uma tendência ou raiz unitária. No entanto, após a primeira diferenciação as duas séries tornam-se estacionárias. Por outro lado, a oferta, variável que irá compor os modelos multivariados para captar o efeito da oferta e da demanda nos preços, demonstrou-se estacionária em nível, não necessitando de diferenciação.

Tabela 3. Teste de estacionariedade para as variáveis que compõem o modelo de previsão para a soja

Variável	KPSS	KPSS - Variável em primeira diferença
Indicador da Soja ESALQ/BM&FBOVESPA – Paranaguá	0,742***	0,114
Oferta	0,134	
Indicador Do Milho ESALQ/BM&FBOVESPA	1,04***	0,111
Oferta	0,0418	

Fonte: Elaborados pelos autores, 2024.

Nota: Valores críticos para KPSS: 10%: 0.119; 5%: 0.146; 2.5%: 0.176; 1%: 0.216

Na Figura 4 a análise dos dados de soja e milho revela uma sazonalidade clara, influenciada pelos ciclos de colheita e plantio. No caso da soja, observa-se uma relativa estabilidade nos primeiros meses do ano, com maior variação nos meses de outubro e novembro. Esses meses refletem o período de plantio, quando há incerteza sobre a oferta futura, o que resulta em um aumento dos preços. Em contraste, os preços tendem a ser mais baixos durante os meses de colheita, quando a oferta é abundante. Essa dinâmica também é evidente no comportamento trimestral, com o terceiro e quarto trimestre, que coincidem com o final do plantio e início do período de menor oferta, apresentando maior volatilidade.

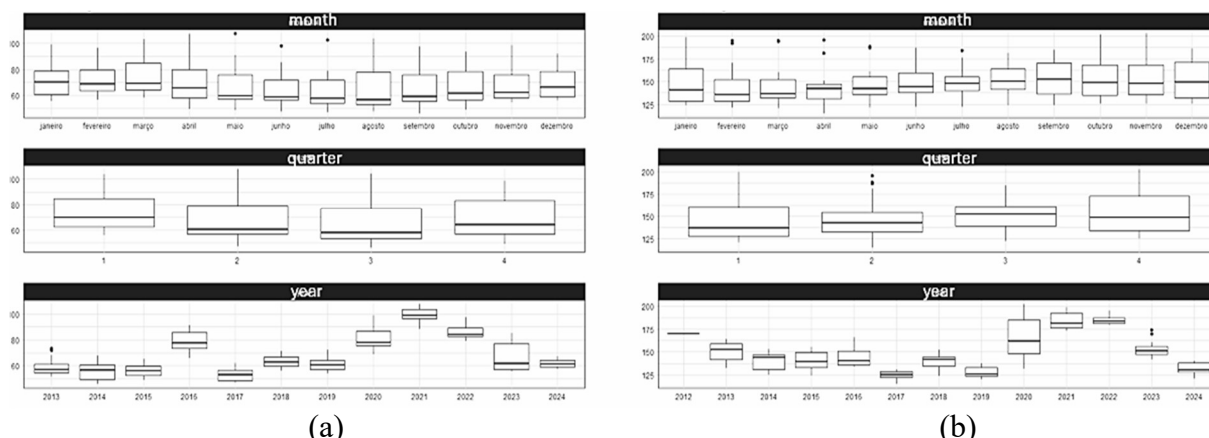


Figura 4. Sazonalidade do Indicador da Soja ESALQ/BM&FBOVESPA – Paranaguá (a) e do Indicador Do Milho ESALQ/BM&FBOVESPA (b)

Fonte: Elaborados pelos autores, 2024.

Nota: O primeiro bloco de gráficos mostra a sazonalidade mensal, o segundo aborda a sazonalidade quadrimestral, e o terceiro captura a sazonalidade anual

No mercado de milho, a sazonalidade é ainda mais pronunciada, os preços são mais baixos nos meses de junho e julho, período de colheita, devido ao aumento da oferta, enquanto nos meses de outubro e novembro, que correspondem ao início do plantio, os preços se elevam devido à incerteza sobre a produção futura. Nos trimestres, o quarto trimestre, que coincide com o plantio, apresenta as maiores flutuações nos preços. Além disso, as flutuações observadas a partir de 2020 nos preços de ambas as *commodities* podem ser atribuídas aos impactos da pandemia de COVID-19, que afetou as cadeias globais de suprimentos e intensificou a incerteza no mercado de soja e de milho.

3.4 Os modelos Univariados

Foram realizadas as estimativas dos modelos univariados para avaliar seu desempenho na previsão das séries temporais de preços da soja e do milho com dois passos à frente. Os resultados dos diferentes modelos podem ser observados nas Tabelas 4 e 5. A análise busca comparar a eficiência de um modelo tradicional, no caso o ARIMA e do ETS (Erro, Tendência e Sazonalidade), com modelos de *machine learning*, como MLP, LSTM e outros, a fim de verificar qual abordagem se mostra mais eficaz na previsão dessas séries temporais univariadas, onde a previsão é realizada com base no próprio histórico do indicador de preço.

Para a soja, os modelos ARIMA (0,1,1), ARIMA (0,1,1) combinado com XGBOOST, ETS, *Prophet*, MLP e LSTM foram comparados utilizando métricas como MAE, MAPE e RMSE. Embora a série de preços da soja apresente sazonalidade, o melhor modelo, de acordo com os critérios de AIC, SBC e BIC, foi o ARIMA (0,1,1). No entanto, os modelos de *deep learning*, como MLP e LSTM, apresentaram desempenho superior em termos de MAE e RMSE, especialmente o LSTM, que obteve os menores erros: MAE de 3,59, MAPE de 2,26 e RMSE de 3,89.

Tabela 4. Resultados dos modelos univariados estimados para a soja

Modelo	MAE	MAPE	RMSE
ARIMA (0,1,1)	6,53	1,65	10,55
ARIMA(0,1,1)/XGBOOST	6,53	1,69	10,58
ETS(M,A,A)	6,52	1,65	10,54
PROPHET	39,76	10,13	55,82
MLP	4,29	2,90	5,65

LSTM 3,59 2,26 3,89

Fonte: Elaborados pelos autores, 2024.

No caso do milho, os modelos avaliados foram os mesmos, e novamente o LSTM obteve o melhor desempenho, com os menores valores de MAE (0,48), MAPE (0,82) e RMSE (0,64). O modelo MLP também apresentou bons resultados, com um MAE de 2,91 e RMSE de 3,85. Por outro lado, o modelo ARIMA (0,1,1), apesar de seu bom desempenho para a soja, não se mostrou tão eficaz para o milho, apresentando valores consideravelmente maiores de erro em comparação com os modelos baseados em *machine learning*.

Tabela 5. Resultados dos modelos univariados estimados para o milho

Modelo	MAE	MAPE	RMSE
ARIMA(0,1,1)	6,33	10,04	7,51
ARIMA(0,1,1)/XGBOOST	9,10	14,64	9,92
ETS(M,N,N)	2,60	4,13	3,62
PROPHET	32,12	53,23	32,28
MLP	2,91	2,87	3,85
LSTM	0,48	0,82	0,64

Fonte: Elaborados pelos autores, 2024.

Para a visualização e a análise do comportamento das séries temporais previstas, foram gerados gráficos das previsões dos preços da soja e do milho utilizando o modelo LSTM, que se destacou como o melhor modelo nas comparações anteriores. As Figuras 5 e 6 apresentam os dados reais dos últimos 12 meses (junho de 2023 a julho de 2024) e as previsões para os últimos 2 meses (junho e julho de 2024), permitindo observar de forma clara o desempenho do modelo na captura das variações das séries temporais ao longo do período.

Na Figura 5, a linha azul representa os dados reais da série temporal do Indicador da Soja ESALQ/BM&FBOVESPA – Paranaguá, enquanto a linha verde escura representa os dados reais ou dados de treinamento para os dois meses preditos, a linha azul mais fina representa as previsões realizadas no período de treino e a linha verde mais fina representa a predição sobre os dados reais ou dados de teste. Observa-se uma leve distinção entre os dados de treinamento, as previsões realizadas com os dados de treinamento do modelo, e as previsões feitas com os dados de validação.

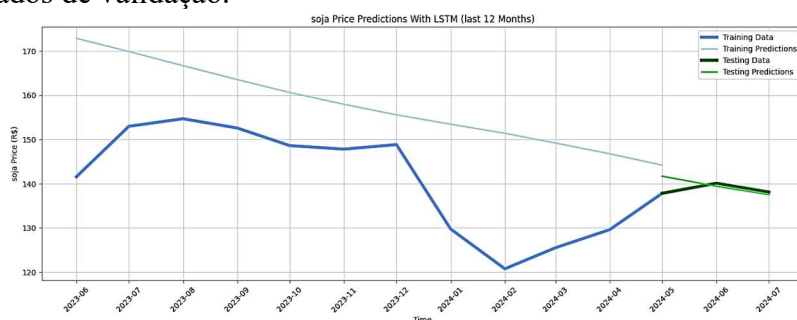


Figura 5. desempenho do modelo LSTM para previsão para o Indicador da Soja ESALQ/BM&FBOVESPA – Paranaguá (junho de 2023 a julho de 2024)

Fonte: Elaborados pelos autores, 2024.

As previsões do modelo LSTM para os dados de teste (linha verde) estão alinhadas com os preços reais da série, capturando a tendência de recuperação observada nos últimos meses do período analisado. A diferença entre as previsões e os dados reais é pequena, indicando um bom desempenho do modelo LSTM para essa série temporal.

Para o modelo do Milho, é possível verificar que a linha azul demonstra uma elevação no Indicador do Milho ESALQ/BM&FBOVESPA durante os meses finais de 2023, seguida por uma queda acentuada nos primeiros meses de 2024, conforme Figura 6. O modelo LSTM prevê essa queda de forma eficiente, como observado pela proximidade entre as previsões (linha verde) e os valores reais. Assim como no caso da soja, o LSTM consegue capturar bem as tendências principais da série de preços do milho.

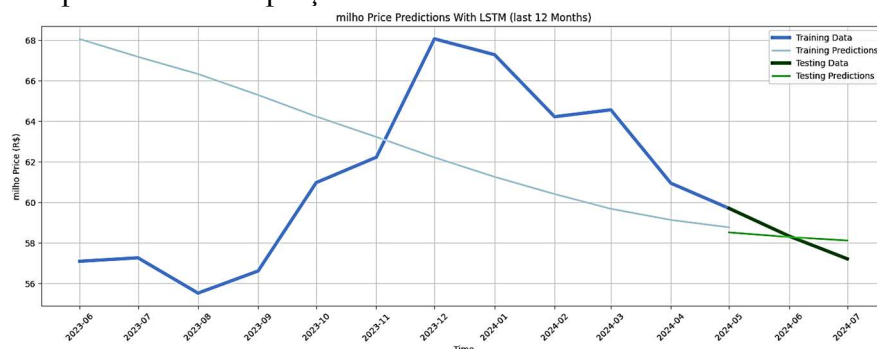


Figura 6. Desempenho do modelo LSTM para previsão para o Indicador do Milho ESALQ/BM&FBOVESPA (junho de 2023 a julho de 2024)

Fonte: Elaborados pelos autores, 2024.

A capacidade do LSTM em modelar tanto padrões sazonais quanto mudanças abruptas nos preços reforça sua superioridade frente a modelos tradicionais como o ARIMA, especialmente em séries temporais não lineares, como as de *commodities* agrícolas. As Figuras 5 e 6 corroboram com os resultados apresentados anteriormente neste trabalho, em que o LSTM obteve os menores valores de erro, tanto para a soja quanto para o milho.

3.5 Os modelos Multivariados

Os modelos multivariados consideram múltiplas variáveis explicativas simultaneamente, permitindo uma análise mais abrangente das dinâmicas que afetam a série temporal. Neste estudo, utilizou-se uma variável de oferta, além da série de preços que se deseja prever. O objetivo é capturar interações e correlações entre essas variáveis, que um modelo univariado não consegue captar. Essa abordagem tem o potencial de melhorar a precisão das previsões, ao incorporar fatores externos que influenciam os preços de maneira mais complexa e interdependente.

Para avaliar o desempenho dos modelos multivariados, foram comparadas três abordagens com previsão de 2 passos à frente: o modelo tradicional ARIMA (com regressora exógena), o modelo de NN MLP e o modelo de RNN LSTM, tendo em vista que esses foram os melhores desempenhos na abordagem univariada. Conforme Tabela 6, o modelo LSTM apresentou o melhor desempenho, com um RMSE de 0,826625, MAE de 0,649124 e MAPE de 0,465911, superando tanto o MLP quanto o ARIMA em todas as métricas. Isso evidencia a capacidade do LSTM em capturar melhor os padrões complexos e não lineares da série, com erros menores. O MLP também demonstrou bom desempenho, mas ficou abaixo do LSTM, com um RMSE de 0,905379 e MAPE de 0,623299. O modelo ARIMA, por outro lado, apresentou os piores resultados, com erros mais altos em todas as métricas, indicando sua limitação em comparação aos modelos de *deep learning*.

Tabela 6. Resultados dos modelos multivariados estimados para a soja

Modelo	RMSE	MAE	MAPE
MLP	0,905379	0,770353	0,623299
ARIMA	1,15207	1,079209	0,778377

LSTM 0,826625 0,649124 0,465911

Fonte: Elaborados pelos autores, 2024.

Para o milho, o comportamento dos modelos seguiu uma tendência semelhante. O LSTM novamente se destacou como o modelo mais eficiente, com um RMSE de apenas 0,126916, MAE de 0,096063 e MAPE de 0,164409. O modelo MLP teve um desempenho razoável, mas inferior ao LSTM, com um RMSE de 2,516618 e MAPE de 4,543783. O ARIMA, apresentou o pior desempenho, refletindo uma menor capacidade de prever adequadamente as flutuações nos preços do milho, conforme Tabela 7.

Tabela 7. Resultados dos modelos multivariados estimados para o milho

Modelo	RMSE	MAE	MAPE
MLP	2,516618	2,512216	4,543783
ARIMA	6,868907	5,244793	7,918968
LSTM	0,126916	0,096063	0,164409

Fonte: Elaborados pelos autores, 2024.

Esses resultados destacam a eficácia dos modelos de *deep learning*, como LSTM e MLP, em lidar com a complexidade e não linearidade das séries temporais multivariadas, particularmente no caso de *commodities* agrícolas, onde diversas variáveis interagem de maneira complexa para influenciar os preços. O desempenho inferior do ARIMA reforça a limitação dos modelos tradicionais em comparação às abordagens mais modernas baseadas em NN, sugerindo um novo caminho para a busca do aprimoramento de modelos de previsão de preço.

Para demonstrar o comportamento dos modelos de previsão, as Figura 7 e 8 fornecem uma visualização clara das previsões feitas pelo modelo LSTM para as séries temporais de preços da soja e do milho, permitindo comparar os dados reais de treinamento e teste com as previsões geradas pelo modelo, demonstrando a eficácia do LSTM em capturar as variações de preço ao longo do tempo. As Figuras 8 e 9 apresentam os dados reais dos últimos 12 meses (junho de 2023 a julho de 2024) e as previsões para os últimos 2 meses (junho e julho de 2024), permitindo observar de forma clara o desempenho do modelo na captura das variações das séries temporais ao longo do período.

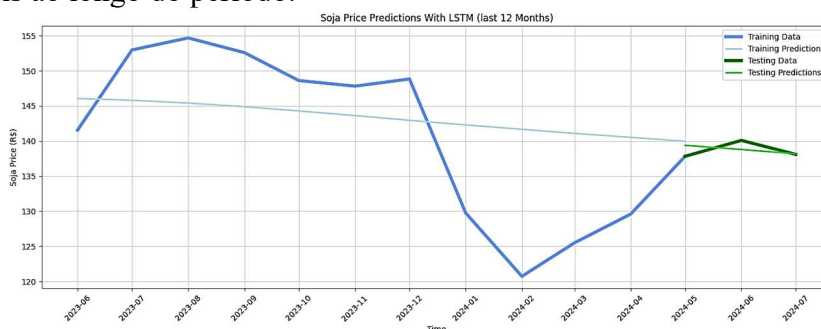


Figura 7. Desempenho do modelo LSTM para a previsão para o Indicador da Soja ESALQ/BM&FBOVESPA – Paranaguá (junho de 2023 a julho de 2024)

Fonte: Elaborados pelos autores, 2024.

Observa-se que, após uma queda acentuada nos preços durante os últimos meses de 2023, os valores começam a se recuperar no início de 2024. O modelo LSTM acompanha essa tendência de recuperação de forma bastante precisa, conforme demonstrado pela proximidade entre as previsões (linha verde mais fina) e os preços reais (linha verde). Isso indica que o

modelo foi capaz de aprender os padrões da série temporal e aplicá-los de maneira eficaz às previsões futuras.

Já para a série temporal do preço do milho, nota-se um comportamento de subida nos preços até o final de 2023, seguido de uma queda consistente ao longo dos primeiros meses de 2024. Assim como no caso da soja, o modelo LSTM captura bem essa tendência de queda, com as previsões (linha verde mais fina) se aproximando bastante, em cima da linha dos dados reais da série (linha verde). O modelo apresenta um desempenho robusto ao prever a trajetória de declínio, reafirmando sua superioridade em relação aos modelos tradicionais para esse tipo de previsão.



Figura 8. Desempenho do modelo LSTM para a previsão para o Indicador Do Milho ESALQ/BM&FBOVESPA (junho de 2023 a julho de 2024)

Fonte: Elaborados pelos autores, 2024.

As Figuras 7 e 8 reforçam o desempenho superior do LSTM na modelagem de séries temporais não lineares e complexas, especialmente de *commodities* agrícolas. O LSTM se destaca por sua capacidade de lidar com padrões não lineares e sazonalidades, proporcionando previsões mais precisas que os modelos tradicionais. Além disso, o uso de um modelo multivariado, incorporando variáveis como a oferta, permite ao LSTM capturar melhor essas interações, melhorando a acurácia das previsões e oferecendo uma análise mais robusta.

4. Conclusão

Os resultados mostraram que os modelos de NN, especialmente o LSTM, superaram os modelos tradicionais, como ARIMA, em termos de precisão e menor erro. O LSTM obteve os menores erros tanto para a previsão dos preços do milho quanto da soja, demonstrando superioridade na captura de padrões complexos e não lineares presentes nas séries temporais analisadas. Nos modelos multivariados, que incluíram uma variável para captar a oferta, o LSTM também apresentou um desempenho melhor que os outros modelos.

As conclusões da pesquisa confirmam as hipóteses levantadas. Primeiro, as técnicas de NN, como o LSTM, provaram ser mais precisas do que os modelos tradicionais, como o ARIMA. Em segundo lugar, os modelos multivariados, que incorporou uma variável para captar a oferta, foram mais eficazes do que os modelos univariados, que utilizam apenas séries temporais de preços, validando assim a importância de incluir múltiplos fatores explicativos nas previsões de preços. Assim, a principal contribuição deste estudo foi fornecer ferramentas mais precisas para a tomada de decisões no mercado agrícola, ao integrar variáveis multivariadas, o que aprimorou a acurácia das previsões e ampliou a aplicabilidade prática dos modelos para os agentes do setor.

Por fim, uma limitação deste estudo foi o uso de dados mensais, o que pode ter limitado a captura de algumas flutuações de curto prazo. Trabalhos futuros podem explorar a inclusão de variáveis adicionais, como fatores climáticos e dados diários, bem como a aplicação de

modelos de previsão em cenários de alta volatilidade, para aumentar a robustez das previsões e sua aplicabilidade no mercado real.

Referências

- Aggarwal, C. 2015. Mining Text Data. In: Data Mining. Springer, Cham. https://doi.org/10.1007/978-3-319-14142-8_13
- Arlot, S.; Celisse, A. 2010. A survey of cross-validation procedures for model selection.
- Associação Brasileira das Indústrias de Óleos Vegetais [ABIOVE]. 2024. Estatísticas da cadeia da soja – Mensal. ABIOVE. Disponível em: <https://abiove.org.br/estatisticas-cadeia-da-soja-mensal/>. Acesso em: 10 ago. 2024.
- Bakay, M. S.; Ağbulut, Ü. 2021. Electricity production based forecasting of greenhouse gas emissions in Turkey with deep learning, support vector machine and artificial neural network algorithms. Journal of Cleaner Production, v. 285, p. 125-324.
- Barros, G. S. C.; Castro, N. R.; Machado, G. C.; Almeida, F. M. S.; Almeida, A. N. 2023. Boletim Mercado de Trabalho do Agronegócio Brasileiro. Centro de Estudos Avançados em Economia Aplicada [CEPEA]. Disponível em: <https://www.cepea.esalq.usp.br/br/releases/cepea-abiove-cadeia-da-soja-e-do-biodiesel-representou-27-do-pib-do-agronegocio-e-gerou-2-milhoes-de-empregos-em-2022.aspx#:~:text=Cepea%2C%2010%2F05%2F2023,o%20PIB%20do%20agroneg%C3%B3cio%20nacional>. Acesso em: 13 ago. 2023.
- Bergmeir, C.; Benítez, J. M. 2012. On the use of cross-validation for time series predictor evaluation. Information Sciences, v. 191, p. 192-213.
- Box, G. E. P.; Jenkins, G. M., Reinsel, G. C., Ljung, G. M. 2015. Time Series Analysis: Forecasting and Control. John Wiley & Sons, Hoboken, NJ, USA.
- Box, G. E. P.; Jenkins, G. M. 1976. Time Series Analysis: Forecasting and Control. Holden Day, San Francisco, CA, USA.
- Centro de Estudos Avançados em Economia Aplicada [CEPEA]. 2024a. Indicador da soja ESALQ/BM&FBOVESPA – Paraná. Disponível em: <https://www.cepea.esalq.usp.br/br/indicador/soja.aspx>. Acesso em: 15 out. 2024.
- Centro de Estudos Avançados em Economia Aplicada [CEPEA]. 2024b. Indicador do milho ESALQ/BM&FBOVESPA. Disponível em: <https://www.cepea.esalq.usp.br/br/indicador/milho.aspx>. Acesso em: 15 out. 2024.
- Centro de Estudos Avançados em Economia Aplicada [CEPEA]; Confederação Nacional Da Agricultura E Pecuária [CNA]. 2023. PIB do agronegócio brasileiro de 1996 a 2021. Disponível em: <https://www.cepea.esalq.usp.br/br/pib-do-agronegocio-brasileiro.aspx>. Acesso em: 13 ago. 2023.
- Companhia Nacional de Abastecimento [CONAB]. 2023. Boletim da Safra de Grãos. Disponível em: <https://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos>. Acesso em: 13 ago. 2023.
- Crone, S. F.; Kourentzes, N. 2010. Feature selection for time series prediction—A combined filter and wrapper approach for neural networks. Neurocomputing, v. 73, n. 10-12, p. 1923-1936.
- Empresa Brasileira de Pesquisa Agropecuária [EMBRAPA]. 2021. Milho: Importância Socioeconômica. Disponível em: [https://www.embrapa.br/en/agencia-de-informacao-tecnologica/cultivos/milho/pre-producao/socioeconomia/importancia-socioeconomica#:~:text=Na%20realidade%2C%20o%20uso%20do,para%20ano%20\(Figura%201\)](https://www.embrapa.br/en/agencia-de-informacao-tecnologica/cultivos/milho/pre-producao/socioeconomia/importancia-socioeconomica#:~:text=Na%20realidade%2C%20o%20uso%20do,para%20ano%20(Figura%201)). Acesso em: 10 ago. 2023.
- Food and Agriculture Organization [FAO]. 2023. Crops and livestock products. Disponível em: <https://www.fao.org/faostat/en/#data/QCL/visualize>. Acesso em: 13 ago. 2023.

- Food and Agriculture Organization of the United Nations [FAO]. 2020. Impact of COVID-19 on global food security. Rome: Food and Agriculture Organization of the United Nations. Disponível em: <https://openknowledge.fao.org/server/api/core/bitstreams/0b6045e6-ebc8-4569-b383-eafdc51b27f8/content>. Acesso em: 15 out. 2024.
- Food and Agriculture Organization of the United Nations [FAO]. 2021a. The State of Food Security and Nutrition in the World 2021: transforming food systems for food security, improved nutrition and affordable healthy diets for all. Rome: Food and Agriculture Organization of the United Nations [FAO]. World Food and Agriculture - Statistical Yearbook 2021b. Disponível em: <https://openknowledge.fao.org/server/api/core/bitstreams/522c9fe3-0fe2-47ea-8aac-f85bb6507776/content>. Acesso em: 15 out. 2024.
- Goodfellow, I.; Bengio, Y.; Courville, A. 2016. Deep Learning. MIT Press, Cambridge, MA, USA.
- Govind, G. R., Babu, A. S. 2023. Hybrid deep learning model to forecast crude oil price. In 2023 International Conference on Inventive Computation Technologies (ICICT) (pp. 19–23). IEEE.
- Hochreiter, S.; Schmidhuber, J. 1997. Long short-term memory. Neural Computation, v. 9, n. 8, p. 1735-1780.
- Hossain, M. A. Karim, R., Thulasiram, R., Bruce, N. D., Wang, Y. 2018. Hybrid deep learning model for stock price prediction. In: IEEE Symposium Series on Computational Intelligence, SSCI, Bangalore, India, p. 1837-1844.
- Hyndman, R. J., Athanasopoulos, G. 2018. Forecasting: Principles and Practice. OTexts.
- Instituto Brasileiro De Geografia E Estatística [IBGE]. 2023. Pesquisa Agrícola Municipal – PAM, 2023. Disponível em: <https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria/9117-producao-agricola-municipal-culturas-temporarias-e-permanentes.html>. Acesso em: 10 ago. 2023.
- Kourentzes, N.; Barrow, D.; Petropoulos, F. 2019. Another look at forecast selection and combination: Evidence from forecast pooling. International Journal of Production Economics, v. 209, p. 226-235.
- Kurumatani, K. 2020. Time series forecasting of agricultural product prices based on recurrent neural networks and its evaluation method. SN Applied Sciences, 2(8), 1434.
- Kwiatkowski, D. Phillips, P. C., Schmidt, P., Shin, Y. 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? Journal of Econometrics, v. 54, n. 1-3, p. 159-178.
- Mankiw, N. G. 2021. Principles of Economics. Cengage Learning, Boston, MA, USA.
- Ministério da Indústria, Comércio Exterior e Serviços [MDIC]. 2023. COMEXSTAT: Exportação e Importação Geral. Disponível em: <http://MIDIC.mdic.gov.br/pt/geral>. Acesso em: 07 ago. 2023.
- Morettin, P. A. 2011. Econometria financeira: um curso em séries temporais financeiras. 2ª ed. Editora Blucher, São Paulo, SP, Brasil.
- Nabipour, M., Nayyeri, P., Jabani, H., Shahab, S., Mosavi, A. 2020. Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; a comparative analysis. IEEE Access, v. 8, p. 150199-150212.
- Organisation for Economic Co-operation and Development [OECD]. 2020. Agricultural trade and markets. Disponível em: <https://www.oecd.org/en/topics/policy-issues/agricultural-trade-and-markets.html>. Acesso em: 15 out. 2024.
- Pathinarupothi, R. K., Vinaykumar, R., Rangan, E., Gopalakrishnan, E., Soman, K. P. 2017. Instantaneous heart rate as a robust feature for sleep apnea severity detection using deep

- learning. In 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI) (pp. 293–296). IEEE.
- Sanches, A.; Sales, C. 2022. Expressivo aumento nas exportações brasileiras de milho impõe mudanças na dinâmica do mercado nacional. Disponível em: <https://www.cepea.esalq.usp.br/br/opiniaoe-cepea/expressivo-aumento-nas-exportacoes-brasileiras-de-milho-impoe-mudancas-na-dinamica-do-mercado-nacional.aspx>. Acesso em: 10 out. 2022.
- Swapna, G.; Kp, S.; Vinayakumar, R. 2018. Automated detection of diabetes using CNN and CNN-LSTM network and heart rate signals. *Procedia Computer Science*, v. 132, p. 1253-1262.
- Taylor, J. W., Snyder, R. D. 2012. Forecasting intraday time series with multiple seasonal cycles using parsimonious seasonal exponential smoothing. *Omega*, 40(6), 748-757.
- Taylor, S. J., Letham, B. 2018. Forecasting at scale. *The American Statistician*, 72(1), 37-45.
- Timmermann, A. 2006. Forecast combinations. In: *Handbook of economic forecasting*, v. 1, p. 135-196, Springer, New York, NY, USA.
- Torres-Pruñonosa, J., García-Estévez, P., Raya, J. M., Prado-Román, C. 2022. How on Earth Did Spanish Banking Sell the Housing Stock? *Sage Open*, v. 12, n. 1, p. 21582440221079916.
- United States Department of Agriculture [USDA]. 2022a. Corn Exports in 2021. Disponível em: <https://www.fas.usda.gov/commodities/corn>. Acesso em: 10 ago. 2023.
- United States Department of Agriculture [USDA]. 2022b. Corn yield in the United States for 2021. Disponível em: <https://www.nass.usda.gov/Newsroom/archive/2022/01-12-2022.php#:~:text=Corn%20yield%20in%20the%20United,%2C%20up%205%25%20from%202020>. Acesso em: 15 out. 2024.
- United States Department of Agriculture [USDA]. 2023. World Agricultural Production. Disponível em: <https://www.fas.usda.gov/data/world-agricultural-production>. Acesso em: 10 ago. 2023.
- Vancsura, L.; Tatay, T.; Bareith, T. 2023. Evaluating the effectiveness of modern forecasting models in predicting commodity futures prices in volatile economic times. *Risks*, v. 11, n. 2, p. 27.
- Wang, Y., Guo, Y. 2020. Forecasting method of stock market volatility in time series data based on mixed model of ARIMA and XGBoost. *China Communications*, 17(3), 205-221.
- Xu, X. 2017. Contemporaneous causal orderings of US corn cash prices through directed acyclic graphs. *Empirical Economics*, v. 52, p. 731-758.
- Xu, X. 2019. Contemporaneous and Granger causality among US corn cash and futures prices. *European Review of Agricultural Economics*, v. 46, n. 4, p. 663-695.
- Xu, X.; Zhang, Y. 2022. Soybean and soybean oil price forecasting through the nonlinear autoregressive neural network (NARNN) and NARNN with exogenous inputs (NARNN-X). *Intelligent Systems with Applications*, v. 13, p. 200061.
- Zhang, G.; Patuwo, B. E.; Hu, M. Y. 1998. Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, v. 14, n. 1, p. 35-62.
- Zhang, Y. J., Yao, T., He, L. Y., & Ripple, R. 2019. Volatility forecasting of crude oil market: Can the regime switching GARCH model beat the single-regime GARCH models? *International Review of Economics & Finance*, v. 59, p. 302-317.
- Zhao, H. 2021. Futures price prediction of agricultural products based on machine learning. *Neural Computing and Applications*, v. 33, p. 837-850.