

**AVALIAÇÃO DE MÉTODOS DE PREVISÃO DE PREÇOS PARA COMMODITIES
AGRÍCOLAS: UM ESTUDO APLICADO À SOJA
EVALUATION OF PRICE FORECASTING METHODS FOR AGRICULTURAL
COMMODITIES: A STUDY APPLIED TO SOYBEANS**

Autor: Fernanda Cigainski Lisbinski
Filiação: Universidade de São Paulo - USP/ESALQ
E-mail: lisbinskif@gmail.com

Autor: Felipe André Oliveira Freitas
Filiação: Universidade de São Paulo - USP/ESALQ
E-mail: felipeoliveirafreitas@usp.br

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

Resumo

O estudo teve como objetivo criar modelos de previsão para os preços do Indicador da Soja ESALQ/BM&FBOVESPA - Paranaguá, visando fornecer ferramentas para a tomada de decisão na compra e venda dessa commodity, de extrema importância para a economia brasileira. Foram utilizados modelos como ARIMA (*AutoRegressive Integrated Moving Average*), MLP (*Multi-Layer Perceptron*), LSTM (*Long Short-Term Memory*) e outros com técnicas de ensemble. Diante das estatísticas calculadas, o modelo MLP apresentou a melhor performance, com um MAPE (*Mean Absolute Percentage Error*) de 0,90%, correspondendo a 0,76 centavos do preço médio conforme indicado pelo MAE (*Mean Absolute Error*). Observa-se também que o modelo LSTM demonstrou superioridade em relação aos demais modelos apresentados, evidenciando que os modelos de *Machine e Deep Learning* foram mais eficazes do que o modelo tradicional. Concluiu-se que os modelos de *Neural Network* superaram o modelo tradicional ARIMA, atribuindo-se essa superioridade à capacidade das redes neurais de capturar padrões complexos e não-lineares nos dados, bem como de lidar de forma mais eficaz com a sazonalidade e a não-estacionariedade da série temporal.

Palavras-chave: Previsão de preços; Soja; Modelos de previsão; Neural Network; Deep Learning.

Abstract

The study aimed to create forecasting models for the prices of the ESALQ/BM&FBOVESPA Soybean Indicator - Paranaguá, aiming to provide decision-making tools for buying and selling this commodity, which is of extreme importance to the Brazilian economy. Models such as ARIMA (*AutoRegressive Integrated Moving Average*), MLP (*Multi-Layer Perceptron*), LSTM (*Long Short-Term Memory*), and others with ensemble techniques were utilized. Based on the calculated statistics, the MLP model exhibited the best performance, with a MAPE (*Mean Absolute Percentage Error*) of 0.90%, corresponding to 0.76 cents of the average price as indicated by the MAE (*Mean Absolute Error*). It is also observed that the LSTM model demonstrated superiority over the other models presented, highlighting that the Machine and Deep Learning models were more effective than the traditional model. It was concluded that the Neural Network models outperformed the traditional ARIMA model, attributing this superiority to the neural networks' ability to capture complex and non-linear patterns in the data, as well as to deal more effectively with the seasonality and non-stationarity of the time series.

Key words: Price forecasting; Soybean; Forecasting models; Neural Network; Deep Learning.

Classificação JEL: Q16; C53; C61; C63.

1. Introdução

A previsão de preços é uma preocupação central para os diversos agentes envolvidos no mercado agrícola, como corretores, hedgers, processadores e produtores. Para os especuladores, essa previsão é vital na busca por lucratividade, enquanto para os produtores, é essencial na definição dos preços de venda antes da produção. Além disso, os processadores e exportadores dependem da previsão para atender às demandas do mercado (XU; ZHANG, 2022). Dada a notável volatilidade dos preços agrícolas (TIMMERMAN, 2006; XU, 2017), esses valores exercem uma influência significativa sobre as decisões econômicas, impactando diretamente

na alocação de recursos e no bem-estar econômico (XU, 2019), especialmente em países cujas economias são altamente dependentes do setor primário.

Prever os preços das *commodities* é uma tarefa desafiadora, marcada pela volatilidade, incerteza e complexidade inerentes ao mercado. A disponibilidade de informações precisas de previsão de preços desempenha um papel crucial na tomada de decisões relacionadas à gestão dos suprimentos agrícolas e no comportamento de compra dos consumidores. Além das variações sazonais na produção, influenciadas por condições climáticas como regime de chuvas e intensidade luminosa, os preços são influenciados por uma série de fatores, incluindo as preferências dos consumidores, estratégias comerciais, aspectos culturais e comportamento dos fornecedores. Embora esses fatores humanos não sejam necessariamente sazonais, eles podem ser modelados como relações entre eventos passados e o estado atual (KURUMATANI, 2020).

O estudo e desenvolvimento de modelos de previsão de preços de *commodities* têm um impacto significativo no cenário político-econômico nacional. Ao aprimorar a precisão das previsões dos preços futuros das *commodities* agrícolas, esses modelos oferecem maior segurança para investidores, produtores agrícolas e formuladores de políticas, permitindo que os departamentos governamentais elaborem regulamentações e políticas agrícolas mais eficazes, reduzindo riscos e promovendo uma gestão mais eficiente dos recursos (ZHAO, 2021). Yashavanth et al. (2017) também enfatizam que os preços dos produtos agrícolas são frequentemente afetados por eventos naturais imprevisíveis, como secas, enchentes e surtos de pragas e doenças, gerando considerável risco e incerteza no processo de modelagem e previsão de preços. As previsões resultantes são valiosas para agricultores, governos, indústrias do agronegócio e outras partes interessadas, fornecendo informações técnicas e orientando as decisões de mercado no setor agrícola.

No Brasil, o setor agrícola desempenha um papel crucial na economia. De acordo com dados do Centro de Estudos Avançados em Economia Aplicada (CEPEA/USP-ESALQ) e da Confederação Nacional da Agricultura (CNA) (2023), em 2022, o Produto Interno Bruto (PIB) do agronegócio contribuiu com 25% para o PIB total da economia brasileira e empregou cerca de 19,35% da população do país (CEPEA, 2023). Dentro do agronegócio, o segmento agrícola representou a maior parte, com 72,7% do valor do PIB, enquanto o segmento pecuário foi responsável por 28,3%. Esses números destacam a significativa importância do setor agrícola brasileiro como um motor essencial para o crescimento econômico e a geração de emprego e renda.

O setor agrícola brasileiro é caracterizado por sua diversidade, abrangendo uma ampla variedade de culturas e produtos. Uma das áreas de destaque é a produção de grãos, na qual o Brasil se posiciona como o quarto maior produtor global, ficando atrás apenas dos Estados Unidos, China e Índia (FOOD AND AGRICULTURE ORGANIZATION - FAO, 2023). Na safra 2021/22, a produção de grãos atingiu 272,69 milhões de toneladas, com uma área plantada de 74,5 milhões de hectares (COMPANHIA NACIONAL DE ABASTECIMENTO - CONAB, 2023). Dentre os grãos produzidos, a soja se destacou, representando 46% da produção total (CONAB, 2023). Além disso, dados da FAO (2023) indicam que em 2022, o Brasil se tornou o maior produtor mundial de soja.

Na literatura, são encontradas várias metodologias para a construção de modelos de previsão de preços, incluindo a abordagem tradicional de Box e Jenkins (1970), os modelos Autoregressive Integrated Moving Average (ARIMA) e os modelos de Vector Autoregression (VAR). Além disso, as técnicas de *machine learning*, como as Redes Neurais Artificiais (ANN), têm se destacado devido à capacidade de aprender padrões complexos dos dados e gerar previsões precisas, mesmo em cenários de alta volatilidade. As ANN são amplamente utilizadas em diversos campos, como classificação de dados, previsão, otimização, reconhecimento de padrões e regressão, devido à sua flexibilidade e capacidade de lidar com relações não lineares

nos dados, tornando-as uma ferramenta valiosa na modelagem e previsão de preços de *commodities* agrícolas (BAKAY; AĞBULUT, 2021).

Diante disso, o objetivo geral deste estudo é criar modelos de previsão para os preços do Indicador da Soja ESALQ/BM&FBOVESPA - Paranaguá, que sirvam como ferramentas na tomada de decisão para compra e venda dessa *commodity*. Para alcançar esse objetivo, os objetivos específicos incluem: identificar e reunir dados relevantes, como dados históricos de preços, produção e demanda, avaliando suas relações e como influenciam os preços; estimar modelos de previsão usando técnicas tradicionais e de redes neurais, incluindo o modelo LSTM; avaliar o desempenho dos modelos usando índices de seleção e critérios estatísticos de validação; e fazer ajustes nos modelos conforme necessário, oferecendo modelos úteis para a tomada de decisões sobre os preços das *commodities* agrícolas.

Este trabalho aborda a complexidade no desenvolvimento de modelos de previsão para os preços da soja, dada sua alta volatilidade e sensibilidade a fatores como condições climáticas, políticas governamentais e demanda global. Essa instabilidade dificulta a tomada de decisão para produtores e exportadores agrícolas em relação ao plantio, colheita e comercialização dessas *commodities*. O objetivo é fornecer uma ferramenta informativa que auxilie na gestão de negócios, contribuindo para a redução de desperdícios e maximização de lucros. Além disso, destaca-se a relevância deste estudo para a economia brasileira. O uso de métodos estatísticos de séries temporais baseados em modelos de *machine learning* oferece uma contribuição significativa, adaptando-se melhor aos comportamentos dos preços da soja. Assim, este trabalho visa preencher uma lacuna na literatura ao explorar e aplicar abordagens inovadoras para a previsão de preços da soja.

Diante da necessidade de previsões precisas no mercado agrícola, é crucial fornecer modelos eficazes que utilizem metodologias atualizadas, além das tradicionais, para auxiliar agricultores, investidores e outros participantes do mercado na tomada de decisões. Nesse contexto, a hipótese de pesquisa sugere que modelos de previsão empregando técnicas de *machine learning* podem ser mais precisos do que os modelos tradicionais, como ARIMA. Essas abordagens visam maximizar os ganhos e atender às demandas do mercado, tanto dentro como fora da propriedade agrícola.

2. Caracterização da Produção da Soja Brasileira

Na safra 2021/2022, o Brasil alcançou o título de maior produtor mundial de soja, com uma impressionante produção de 125.549,8 milhões de toneladas (CONAB, 2022). Isso corresponde a aproximadamente 37,72% do total global de 360,14 milhões de toneladas de soja produzidas durante a safra (USDA, 2023). A área plantada no país durante a safra 2020/21 foi de 39,2 milhões de hectares, representando cerca de 47,73% da área total cultivada com lavouras temporárias, enquanto a produtividade média alcançou 3.026 kg/ha (PAM-IBGE, 2023; CONAB, 2023).

De acordo com dados da Abiove e da Produção Agrícola Municipal (ABIOVE, 2023; PAM-IBGE, 2023), a produção de soja no Brasil registrou um crescimento médio de 6,6% ao ano durante o período entre os anos civis de 2010 e 2021. Isso culminou em um recorde de 139,39 milhões de toneladas no ano civil de 2021, representando um aumento de quase 50% em comparação com a quantidade produzida em 2010 (Figura 1).

A produção da cultura da soja concentra-se principalmente nas regiões Centro-Oeste e Sul do Brasil, que representaram, em 2021, 46,48% e 30,87% do total produzido no país, respectivamente. Embora a região Sul tenha apresentado crescimento ao longo do período analisado, sua participação diminuiu de 38% em 2010 para 30,87% em 2021. As estimativas para 2022 indicam uma redução significativa da produção nessa região, caindo para 18,64% da produção nacional devido a uma estiagem que afetou o estado do Rio Grande do Sul, onde a produção de soja caiu cerca de 54,3% conforme relatórios do Departamento de Economia e

Estatística vinculado à Secretaria de Planejamento, Governança e Gestão (DEE/SPGG, 2022). Diante desse cenário, verifica-se o potencial de crescimento das demais regiões em relação à cultura da soja.

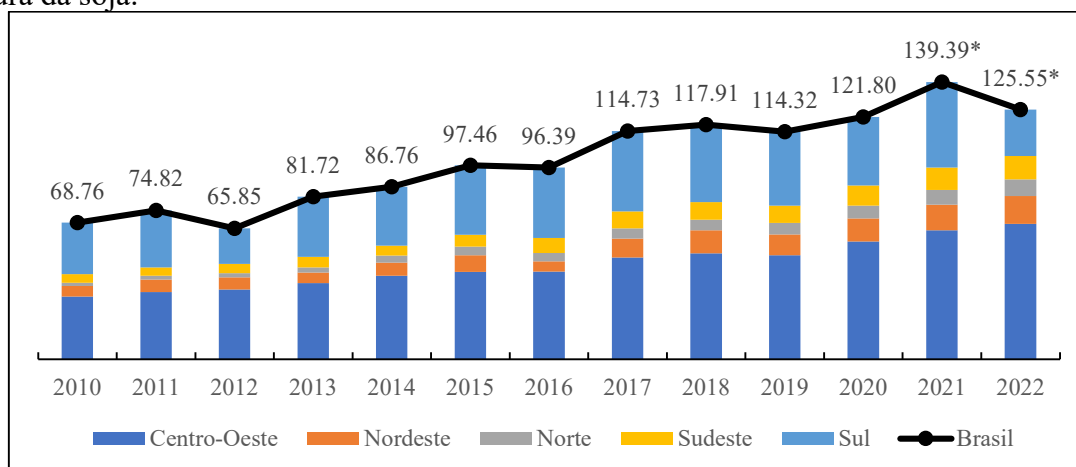


Figura 1 – Produção de Soja em grãos nas regiões brasileiras entre 2010 e 2022 (Milhões de toneladas)

Nota: De 2010 a 2020 dados da PAM-IBGE; 2021* e 2022* estimativas e projeções CONAB

Fonte: Elaborado pelos autores com base nos dados de (PAM-IBGE, 2023) e (CONAB, 2023)

As variações na quantidade produzida de soja no Brasil podem ser atribuídas a diversos fatores, sendo os mais significativos relacionados às condições climáticas e às flutuações de preço no mercado internacional. Em 2012, uma queda de aproximadamente 12% na produção brasileira em comparação com o ano anterior foi observada, fenômeno que pode ser atribuído à estiagem que afetou as lavouras da América do Sul durante o período de desenvolvimento do grão, causada pelo fenômeno climático conhecido como La Niña. Posteriormente, os preços da soja nos mercados interno e externo aumentaram, incentivando os produtores a aumentarem a produção em cerca de 24% em 2013 (CEPEA, 2012).

Em 2016, houve uma leve queda na produção de soja (-1%), influenciada pela diminuição dos preços e pela redução do processamento do grão pelas indústrias brasileiras, com destaque para o estado do Mato Grosso. Mesmo com a valorização do óleo de soja nesse ano, algumas indústrias processadoras encerraram suas atividades em outras regiões do país devido à baixa demanda por farelo, resultante da redução na produção de suínos e da expectativa de diminuição na produção de frango. Ressalta-se que, para cada tonelada de grão processada, cerca de 78% são transformadas em farelo de soja, enquanto apenas cerca de 19% resultam em óleo (CEPEA, 2016).

Em 2019, foi registrada uma retração de 3% na produção de soja em relação ao ano anterior, atribuída principalmente ao déficit hídrico que afetou as regiões Nordeste e Centro-Oeste (CEPEA, 2019). Em 2021, o estado de Mato Grosso se destacou como o principal produtor brasileiro, concentrando 26,19% da produção nacional de soja, o equivalente a 3,5 milhões de toneladas, seguido pelo Rio Grande do Sul (15,13%) e Paraná (14,23%) (PAM-IBGE, 2023). Quanto às variações na produtividade, observa-se um crescimento ao longo da última década na produção brasileira de soja.

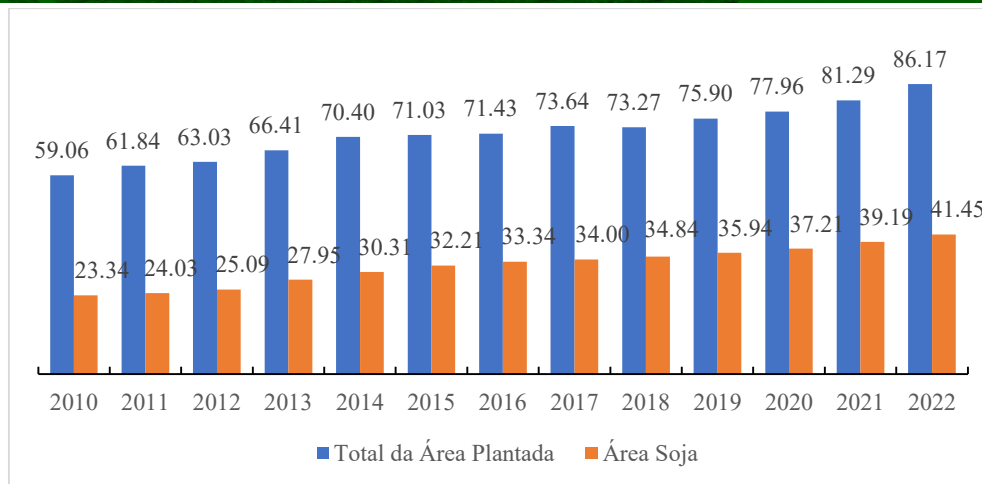


Figura 2 – Área plantada de soja total Brasileira entre 2010 e 2022 (Milhões de hectares)

Nota: De 2010 a 2021 dados da PAM-IBGE; 2022* estimativas e projeções CONAB

Fonte: Elaborado pelos autores com base nos dados de (PAM-IBGE, 2023) e (CONAB, 2023)

Conforme Figura 2, observa-se um aumento significativo da área cultivada de lavouras temporárias, cerca de 37,64% em 2021 em comparação com 2010, sendo que aproximadamente 48,21% dessa área é destinada ao cultivo de soja. Em 2021, cinco estados localizados na Região Centro-Sul do Brasil - Mato Grosso (27%), Rio Grande do Sul (15,6%), Paraná (14%), Goiás (9,84%) e Mato Grosso do Sul (8,62%) - concentraram mais de 75% da área cultivada com soja, totalizando 29,32 milhões de hectares, e 75% da produção nacional, alcançando 100 milhões de toneladas (PAM-IBGE, 2023).

No período entre 2010 e 2020, o maior montante registrado de exportações de soja ocorreu em 2018, quando o Brasil alcançou um total exportado de 78 bilhões de toneladas. Destaca-se que nos anos de 2015 e 2016, houve uma retração nas exportações, resultante da queda dos preços das principais *commodities* agrícolas oferecidas pelo Brasil, como soja e carne. Essa queda de preços levou o país a diminuir seu faturamento com exportações em cerca de US\$ 25 bilhões no ano de 2015 (BRASIL, 2016).

Além da queda nos preços praticados em anos anteriores, outra razão que contribuiu para a redução das exportações em 2016 foi a retração nas vendas de soja por parte dos produtores, como uma resposta aos preços baixos observados no ano anterior (MAPA, 2016). Em 2019, o déficit hídrico, principalmente nas regiões Nordeste e Centro-Oeste, resultou em uma redução de 19,8% no valor exportado (CEPEA, 2019). Essa situação foi superada apenas em 2021, quando o país registrou um novo recorde de exportações do complexo soja, totalizando US\$ 48 bilhões. Desse montante, US\$ 38 bilhões correspondiam a soja em grão (80,5%), US\$ 7 bilhões a farelo de soja (15,30%) e US\$ 2 bilhões a óleo de soja (4,20%) (MIDIC, 2023) (Figura 3).

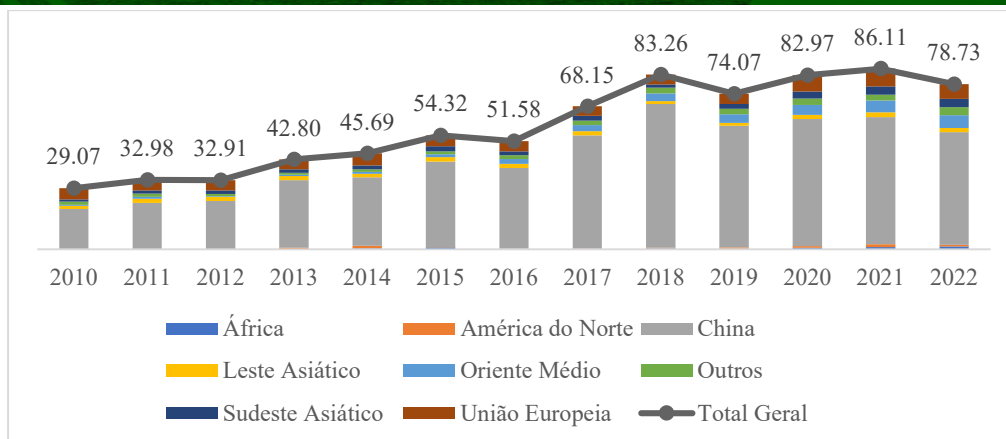


Figura 3 – Exportações de soja entre 2010 e 2022 (Milhões de Toneladas)

Fonte: Elaborado pelos autores com base nos dados de (MIDIC, 2023).

A China tem desempenhado um papel crucial como principal parceiro comercial do Brasil na comercialização de produtos do complexo soja brasileiro, sendo responsável por 59,09% das exportações brasileiras desse produto em 2022. Mantendo consistentemente a primeira posição no ranking de destino das exportações brasileiras do segmento ao longo da série analisada, a China tem aumentado sua participação desde o ano de 2013, representando mais da metade das exportações brasileiras do complexo soja. Em comparação, os demais países que compõem o ranking apresentam uma participação menor nas exportações brasileiras em relação à parcela destinada à China.

A região do Centro-Oeste lidera as exportações de soja brasileira, contribuindo, em 2022, com 48,95% do total, seguida pelo Sul, com 14,54%, Sudeste com 13,42%, Nordeste com 12,59% e Norte com 9,58%. Apesar de apresentar uma participação menor, o Nordeste demonstra um crescimento promissor, enquanto o Norte possui um considerável potencial inexplorado. O Sudeste se destaca pela tecnificação, enquanto o Sul é reconhecido pela tradição na produção. A categoria "Não Declarada" representa 0,92% das exportações. É fundamental ressaltar que a participação das regiões pode variar ao longo do tempo, influenciada por fatores como clima, preços internacionais e políticas públicas.

A alta quantidade de soja em grão exportada é atribuída, de acordo com Hirakuri e Lazzarotto (2014), ao incentivo proporcionado pela Lei Kandir, promulgada em setembro de 1996. Esta lei isenta a cobrança de ICMS sobre as exportações de *commodities* agrícolas in natura, o que beneficia significativamente as exportações de produtos de menor valor agregado, como o grão de soja, em detrimento de produtos de maior valor agregado, como o óleo de soja e o farelo. No ano de 2021, aproximadamente 62% da produção de soja em grão teve como destino o mercado externo, enquanto cerca de 34% foi utilizada na indústria de processamento.

Com relação aos preços (Tabela 1), o preço da soja em grão no mercado externo tem sido consistentemente superior ao interno desde 2010, o que tem impulsionado as exportações do produto. Em 2010, observa-se uma diferença nos preços da soja em grão entre as praças do mercado interno, com valores entre 8% e 18% inferiores aos preços do mercado externo. Entretanto, em 2021, essa diferença diminuiu, variando entre 3% e 5%. Já em 2022, esse diferencial aumentou novamente, ficando entre 10% e 25%. A Tabela 1 apresenta os preços da soja em diferentes praças de comercialização em relação aos preços do mercado externo.

Tabela 1 – Evolução da média anual dos Preços de 2010 a 202 por US\$/saca 60 kg e R\$/saca de 60 kg

Ano	Externo	Maringá PR	Mogiana SP	Passo Fundo RS	Rondonópolis MT
-----	---------	---------------	---------------	-------------------	--------------------

		(US\$/saca) - sem ICMS	R\$/saca - (sem ICMS)	(US\$/saca) - sem ICMS	R\$/saca - sem ICMS	(US\$/saca) - sem ICMS	R\$/saca - sem ICMS	(US\$/saca) - sem ICMS	R\$/saca - sem ICMS
2010	24,76	21,56	37,95	21,59	38,00	22,70	39,95	20,30	35,73
2011	30,05	26,67	44,67	26,35	44,13	27,58	46,19	25,04	41,95
2012	35,56	32,60	63,72	31,36	61,29	32,98	64,46	31,06	60,72
2013	32,16	30,20	65,25	29,27	63,23	30,74	66,41	28,37	61,30
2014	29,30	27,55	64,87	27,21	64,07	28,15	66,28	25,62	60,32
2015	22,36	20,74	69,24	20,50	68,44	21,54	71,90	19,14	63,89
2016	23,73	22,42	78,11	22,10	77,00	22,66	78,95	21,52	74,95
2017	22,91	20,99	67,01	20,37	65,03	21,30	67,99	19,56	62,46
2018	23,58	21,74	79,47	21,27	77,74	21,98	80,36	20,04	73,28
2019	21,27	19,79	78,08	19,30	76,16	20,24	79,87	18,80	74,21
2020	23,72	23,06	118,88	22,88	117,97	23,47	121,02	23,03	118,76
2021	32,01	31,01	167,29	30,91	166,76	31,15	168,07	30,49	164,50
2022	31,44	27,07	187,80	27,83	187,95	28,55	192,08	25,12	175,91

Fonte: Elaborado pelos autores com base nos dados de (ABIOVE, 2023).

Diante da desvalorização do real, o preço da saca de soja apresentou um aumento significativo ao longo da série analisada até o ano de 2022, atingindo o valor de R\$ 185,93 por saca de 60 kg. Esse valor representa um aumento nominal de 490% em comparação com o ano de 2010, quando o preço médio era de R\$ 37,91 por saca (ABIOVE, 2023).

Potanto, o Brasil mantém uma posição proeminente na produção e exportação de soja, influenciando o mercado global e sua economia interna. As variações na produção afetam os preços internos e externos da *commodity*, refletindo a dinâmica entre oferta, demanda e condições macroeconômicas globais. A expansão contínua da produção brasileira demonstra sua capacidade de atender às demandas do mercado internacional. É crucial que o país adote políticas que promovam o desenvolvimento sustentável da produção de soja para manter sua competitividade global e seu papel como líder nesse setor essencial para a economia nacional.

3. Metodologia

3.1 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; 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) com a seguinte estrutura:

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

onde $\theta_0 = \mu(1 - \phi_1 - \dots - \phi_p)$, representando 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. Em geral, cada ε_t é considerado como tendo distribuição normal independente e identicamente distribuída com média zero e variância constante

3.2 O Modelo de Artificial Neural Networks (ANN)

Os algoritmos de Redes Neurais Artificiais (ANNs) são uma simulação do cérebro humano, organizados em três camadas: entrada, oculta e saída. Cada camada é composta por

células neurais conectadas, formando uma rede interligada. Esses modelos possuem recursos poderosos para reconhecer e classificar padrões, tornando-os eficientes para previsão. Geralmente, as ANNs têm três camadas principais: entrada para variáveis explicativas, oculta para processamento e saída para a variável prevista. Elas podem ser aplicadas tanto para séries temporais univariadas quanto multivariadas, com modelos multivariados recomendados para maior precisão quando há dados confiáveis disponíveis (TORRES-PRUÑONOSA et al., 2022; ZHANG et al., 1998; ARMSTRONG, 2001). Assim, a estrutura de uma rede neural pode ser generalizada da seguinte forma:

$$I - (H_1, H_2, H_3 \dots H_N) - O \quad (2)$$

Onde I é o número de nós da camada de entrada; H_N é o número de neurônios na camada oculta n ; e O é o número de neurônios da camada de saída.

Na Figura 4 apresenta-se a expressão gráfica de um modelo de redes neurais para previsão. Onde, Y_j representa o vetor de entrada denotado por $Y_j = \{y_1, y_2, y_3, \dots, y_n\}$. O vetor de peso da conexão dos j nós da camada de entrada para os k nós da camada oculta é representado por $W_{jk} (j = 1, 2, 3 \dots n; k = 1, 2, 3 \dots, m)$, o vetor dos k neurônios na camada oculta é $X_k (k = 1, 2, 3, \dots m)$ determinado pela fórmula $X_k = f(\sum_{j=1}^n W_{jk} Y_j + \Theta_k)$, os pesos de conexão dos k nós da camada oculta para a camada de saída é representado por $W_k (k = 1, 2, 3, \dots, m)$ o vetor unitário de saída para a rede neural com um neurônio de saída é Y e é determinado pela fórmula $Y = f(\sum_{k=1}^m W_k X_k + \Theta)$. Por fim, $\Theta_k (k = 1, 2, 3, \dots, k)$ é o valor de viés dos nós da camada oculta e Θ é o valor de viés da camada de saída.

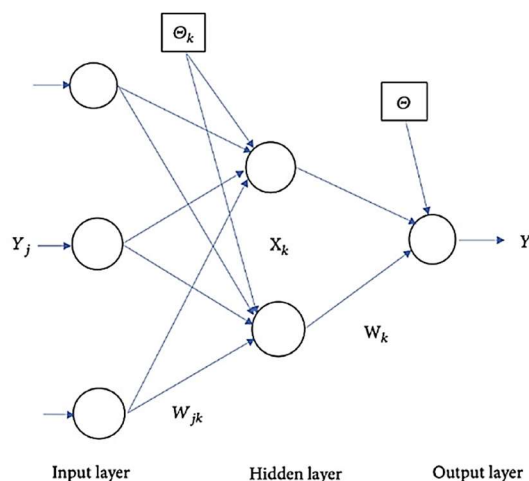


Figura 4 - *Artificial Neural Networks* para previsão de séries temporais.

Fonte: Monge e Lazcano (2022)

Os modelos de ANN são amplamente empregados na previsão de séries temporais econômicas devido à sua precisão em comparação com outros métodos de previsão (KAASTRA; BOYD, 1996; KHASHEI; BIJARI, 2010). Esses modelos são capazes de aprender de forma inteligente com as características dos dados, dispensando a necessidade de especificação prévia do modelo, o que os torna adequados para lidar com problemas como sazonalidade, não estacionariedade e padrões de não linearidade variados nas séries temporais.

3.2.1 O Multi-Layer Perceptrons Neural Network

A *Multi-Layer Perceptrons Neural Network* é composta por uma ou mais camadas ocultas entre as camadas de entrada e saída de uma rede. Assim, uma *Multi-Layer Perceptrons Neural Network* apresenta melhor capacidade de aprender o padrão de entrada e produzir uma saída melhor do que a *Neural Network* de camada única.

A forma mais utilizada de NNs 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 seguinte equação:

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

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. 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; KOURENTZES, 2010; KOURENTZES; BARROW; CRONE, 2014).

3.3 O Modelo *Long Short-Term Memory* (LSTM)

O modelo LSTM é uma variante das Redes Neurais Recorrentes (RNN) frequentemente utilizada em análise de dados sequenciais, especialmente no contexto de aprendizado profundo. Enquanto as RNNs são limitadas no aprendizado de afiliações de longo prazo, o LSTM supera essa restrição, permitindo a retenção e eliminação seletiva de informações por meio de portões específicos. Desenvolvido por Hochreiter e Schmidhuber (1997), o LSTM possui uma arquitetura composta por três portões principais: o de entrada, de saída e de esquecimento. Esses portões possibilitam o processamento eficiente de sequências de dados, mantendo informações relevantes de entradas anteriores por um período prolongado. Comparado às Redes Neurais Artificiais (ANN), o LSTM é capaz de processar toda a sequência de previsão, categorizando os dados de entrada em células de memória de curto e longo prazo (KOSHIYAMA; FIROOZY; TRELEAVEN, 2020; SUMATHI; PANEERSELVAM, 2010).

O funcionamento do LSTM é caracterizado pelo controle dos fluxos de entrada e saída por meio das portas de entrada e saída de uma célula de memória. Quando determinados valores deixam de ser relevantes, um portão de esquecimento é acionado para redefinir-los, multiplicando-os por valores de 0 e 1 para remover ou manter o valor para a próxima etapa. Isso permite que a rede se lembre de informações por longos períodos, uma capacidade que as RNNs tradicionais não possuem. O aspecto mais relevante do LSTM é sua capacidade aprimorada de registrar dependências de longo prazo em dados de séries temporais. Por meio da alteração do estado da célula e do estado da célula de memória atual, as portas de entrada, saída e esquecimento trabalham juntas para regular o comportamento da célula de memória. A Figura 5 ilustra graficamente o funcionamento de um modelo LSTM.

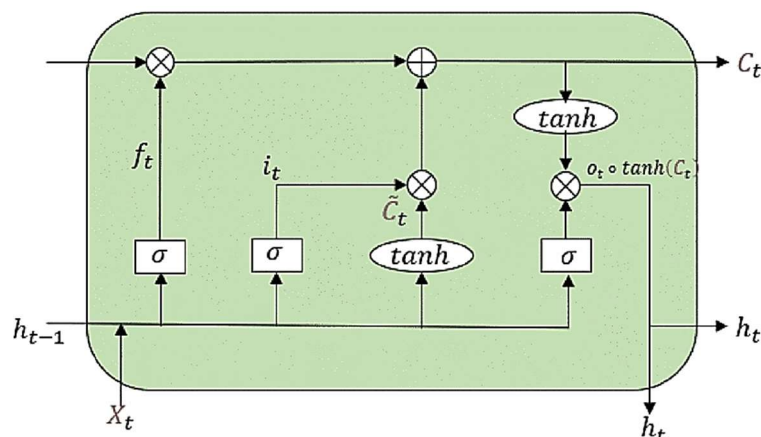


Figura 5 - Diagrama do Modelo LSTM

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 :

$$f_t = \sigma(X_t U_f + h_{t-1} W_f + b_f) \quad (3)$$

$$i_t = \sigma(X_t U_i + h_{t-1} W_i + b_i) \quad (4)$$

$$o_t = \sigma(X_t U_o + h_{t-1} W_o + b_o) \quad (5)$$

$$\tilde{C}_t = \tanh(X_t U_c + h_{t-1} W_c + b_c) \quad (6)$$

$$C_t = \sigma(f_t * C_{t-1} + i_t * \tilde{C}_t) \quad (7)$$

$$h_t = o_t * \tanh + C_t \quad (8)$$

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 $\tilde{C}_t$. O $\tanh$ é uma função de ativação (ZHANG et al., 2018; GOVIND; BABU, 2023).

3.4 Procedimento econométrico

Antes de realizar as estimativas de previsão, é fundamental conduzir alguns testes, como o teste de raiz unitária, para verificar a estacionariedade das variáveis. Nesse caso, será empregado o teste KPSS (Kwiatkowski-Phillips-Schmidt-Shin) proposto por Kwiatkowski et al. (1992), no qual a não rejeição da hipótese nula indica estacionariedade da série. Vale ressaltar que a presença de raiz unitária pode sugerir uma relação econométrica espúria.

Para os modelos ARIMA, é necessário seguir alguns passos para determinar a ordem de estimação do modelo. Assim, é necessário analisar a função de autocorrelação (*Autocorrelation Function* - ACF) e a função de autocorrelação parcial (*Partial Autocorrelation Function* (PACF), para verificar os parâmetros p , d e q do modelo a ser estimado. 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.

Após a realização dos testes e, no caso do modelo ARIMA, da definição dos parâmetros, parte-se para a validação dos modelos de previsão. Para isso, a base de dados será 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 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, será mais robusta do que as medidas individuais (STONE, 1974; ARLOT; CELISSE, 2010; BERGMEIR; BENÍTEZ, 2012). Assim, o conjunto de treinamento é usado para ajustar o modelo, enquanto o conjunto de validação é usado para avaliar a precisão do modelo.

3.5 Indicadores de Avaliação

Os critérios de desempenho de predição dos modelos incluem quatro métricas principais: Erro Absoluto Médio (MAE), Erro Percentual Médio Absoluto (MAPE) e Raiz do

Erro Quadrático Médio (RMSE). A descrição das fórmulas desses indicadores estão descritas no Quadro 1.

Quadro 1 – Descrição dos critérios de desempenho de predição dos modelos

Indicador	Descrição
$MAE = \frac{1}{n} \sum_{t=1}^n y_i - \hat{y}_i \quad (9)$	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; BABU, 2023).
$MAPE = \frac{1}{n} \sum_{t=1}^n \left \frac{y_i - \hat{y}_i}{y_i} \right \quad (10)$	O MAPE mede a magnitude média do erro em um conjunto de previsões e mostra os desvios em porcentagem (VANCSURA; TATAY; BAREITH, 2023). O MAPE é uma medida de quão longe, em média, os valores previstos estão dos valores reais em termos percentuais.
$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_i - \hat{y}_i)^2} \quad (11)$	O RMSE mostra uma estimativa dos resíduos entre os valores reais e previstos (VANCSURA; TATAY; BAREITH, 2023). O RMSE é uma medida de quão longe, em média, os valores previstos estão dos valores reais em termos absolutos.

Fonte: Elaborado pelos autores, 2023.

Para avaliar a previsão dos modelos são considerados uma previsão de até 90 passos à frente, ou seja, uma previsão de até 90 dias. Assim, os menores valores dessas estatísticas definirão o melhor modelo.

3.6 Levantamento de dados

Os modelos foram ajustados no formato univariado utilizando a série temporal do Indicador da Soja ESALQ/BM&FBOVESPA - Paranaguá, fornecida pelo CEPEA, com frequência diária. Antes de realizar a estimação do modelo de previsão, os dados foram pré-processados e uma análise exploratória foi conduzida para examinar a estrutura dos dados, identificar possíveis rupturas e detectar comportamentos atípicos, como *outliers*. As séries históricas utilizadas abrangem o período de 02 de janeiro de 2008, que marca o início da disponibilidade dos dados, até 29 de fevereiro de 2024, o último dado disponível até o momento da estimativa do modelo.

4. Análise e Apresentação dos Modelos de Previsão para o Indicador da Soja ESALQ/BM&FBOVESPA em Paranaguá

Iniciando pela análise da série de interesse nesta investigação, a Figura 6, oferece um panorama do comportamento temporal ao longo dos anos. É possível observar que em 2021, a série apresentou um comportamento incomum devido à pandemia de COVID-19 e uma variedade de influências nos mercados globais de *commodities* agrícolas, incluindo a soja. Após o impacto inicial da pandemia, políticas de estímulo econômico e programas de recuperação foram implementados pelos governos, impulsionando a atividade econômica global. Além disso, a rápida distribuição das vacinas contra a COVID-19 possibilitou a reabertura gradual das economias e o retorno à normalidade em vários setores, incluindo o agrícola. Esses eventos, combinados com condições climáticas favoráveis em algumas regiões produtoras e a demanda

continua por alimentos e biocombustíveis, provavelmente contribuíram para os movimentos atípicos nos preços da soja em 2021 (CONAB, 2021; MENEZES et al., 2023).



Figura 6 – Indicador da Soja ESALQ/BM&FBOVESPA – Paranaguá

Fonte: Elaborado pelos autores, 2023.

Os preços da soja alcançaram patamares significativamente elevados em 2021, impulsionados por uma combinação de fatores que afetaram tanto a oferta quanto a demanda global dessa importante *commodity* agrícola. O aumento na demanda, especialmente da China, principal importadora mundial de soja, devido ao crescimento econômico e à expansão da indústria de criação de suínos e aves, contribuiu para o aumento dos preços. Além disso, eventos climáticos adversos em alguns países produtores, como secas e inundações, reduziram a oferta global, exercendo pressão altista sobre os preços. O aumento dos custos de produção agrícola e as políticas de incentivo à produção também influenciaram os preços. Além disso, o crescente uso da soja em várias indústrias, como biocombustíveis e alimentos processados, contribuiu para a demanda robusta e para os preços elevados. Esses fatores combinados resultaram em uma conjuntura de mercado que impulsionou os preços da soja a patamares historicamente altos em 2021 (KRETER; SOUZA JR., 2021; CNA, 2021). Na Figura 7, é possível verificar o diagnóstico de anomalia, o qual aponta um maior número de anomalias no ano de 2021.

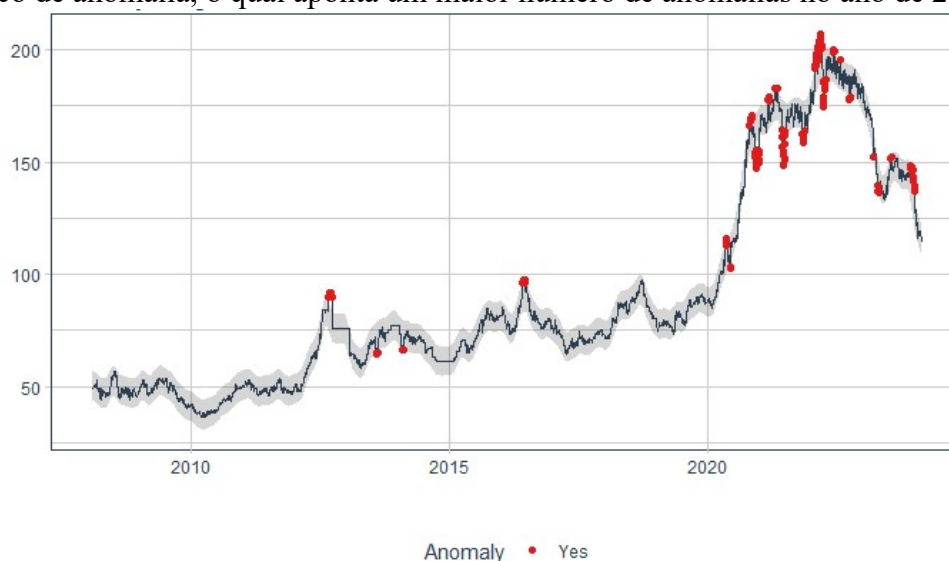


Figura 7 – Diagnostico de anomalias do Indicador da Soja ESALQ/BM&FBOVESPA – Paranaguá

Fonte: Elaborado pelos autores, 2023.

No entanto, apesar da queda observada nos preços em 2023 em comparação com os anos de 2021 e 2022, é interessante notar que os preços da *commodity* permanecem significativamente elevados em relação aos patamares anteriores à pandemia de covid-19. Esse fenômeno pode ser atribuído a uma série de fatores, incluindo a recuperação econômica gradual pós-pandemia, a retomada da demanda global, especialmente da China, e as condições climáticas que podem ter afetado a oferta e a produção agrícola em diferentes regiões produtoras de *commodities* (BASTOS; LEITE, 2024).

Na Figura 8, estão representados os gráficos da Função de Autocorrelação (ACF) e da Função de Autocorrelação Parcial (PACF), os quais fornecem informações sobre a ordem de autorregressão (AR) e média móvel (MA). Ao examinar o padrão da ACF, nota-se que os coeficientes não diminuem rapidamente em direção a zero, sugerindo uma semelhança com uma senoide amortecida. Esse padrão indica um primeiro indício de não estacionariedade na série, que será posteriormente confirmado por testes de raiz unitária. Além disso, é possível observar a presença de sazonalidade na série, aspecto que será analisado em detalhes adiante.

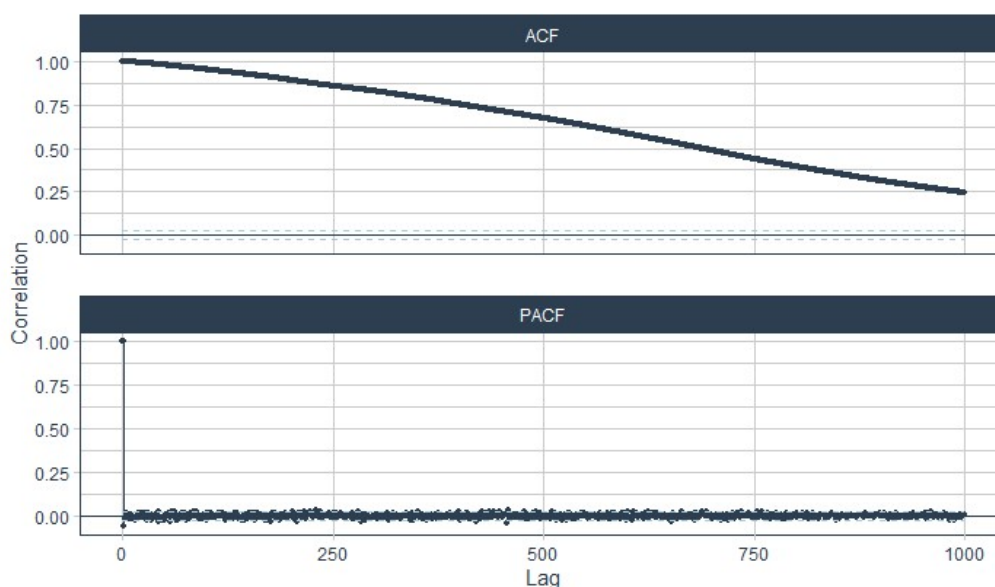


Figura 8 – ACF e PACF DO Indicador da Soja ESALQ/BM&FBOVESPA – Paranaguá
Fonte: Elaborado pelos autores, 2023.

O diagnóstico sugere evidências de sazonalidade e não-estacionariedade do Indicador da Soja ESALQ/BM&FBOVESPA – Paranaguá. No caso da sazonalidade, isso seria esperado pois a produção de soja segue padrões sazonais, com períodos específicos de plantio, crescimento e colheita. A não estacionariedade indica a necessidade de realizar transformações, como diferenciação, para torná-la estacionária e, assim, melhorar a eficácia dos modelos. Dessa forma, realizou-se os testes de estacionariedade da série para confirmar o que fora observado pela apresentação gráfica. Foi utilizado o teste KPSS para verificar a estacionariedade da série, conforme Tabela 2. O teste KPSS é considerado o mais completo em detectar diferentes formas de não estacionariedade relacionadas a tendências determinísticas.

Tabela 2 – Teste de Estacionariedade do Indicador Soja ESALQ/BM&FBOVESPA – Paranaguá

Teste de Raiz Unitária	Indicador da Soja ESALQ/BM&FBOVESPA – Paranaguá	Indicador da Soja ESALQ/BM&FBOVESPA – Paranaguá – em primeira diferença
KPSS	5.6707***	0.15336

Fonte: elaborado pelos autores, 2023.

Nota: valores críticos para o KPSS: 10%=0.347; 5%=0.463; 2,5%=0.574; 1%=0.739.

Após análise da série e aplicação das correções necessárias para torná-la estacionária, além do deflacionamento da mesma, avança-se para a modelagem econométrica. Para tratar a sazonalidade no modelo ARIMA, emprega-se o modelo SARIMA. Por outro lado, os modelos de rede neural têm a capacidade de reconhecer e lidar com a sazonalidade da série durante o treinamento. Para validar o modelo e calcular as estatísticas de erro, a base de dados é dividida entre treino e validação, utilizando o método *k-fold cross-validation*. Vale destacar que a série é diária, com 5873 observações, sendo essa dividida entre treino e validação dos dados. A partir disso, são calculadas as métricas de desempenho do modelo para uma projeção de 90 dias, conforme apresentado na Tabela 3.

Modelo	MAE	MAPE	RMSE
SARIMA	11,26	9,19	14,93
ARIMA W/ XGBOOST ERRORS	11,28	9,24	15,24
ETS	11,26	9,20	14,97
LM	21,00	16,79	24,68
LSTM	3,59	2,26	3,89
MLP	0,76	0,90	1,79

Nota: Para a estimação do SARIMA foi utilizada a seguinte ordem do modelo (0,1,1)(0,0,2); para o modelo ARIMA W/ XGBOOST ERRORS foi utilizada a ordem ((1,1,0).

Fonte: Elaborado pelos autores, 2024.

Além dos modelos LSTM, Multi-Layer Perceptron (MLP) e do ARIMA, foram utilizados diferentes modelos para prever os preços da série temporal do Indicador da Soja ESALQ/BM&FBOVESPA – Paranaguá. Entre os modelos empregados, destacam-se o ARIMA W/ XGboost Errors, que combina o ARIMA, um modelo estatístico clássico, com o XGBoost, um algoritmo de aprendizado de máquina, para capturar padrões temporais complexos e interações não lineares. O ETS (*Error, Trend, Seasonality*) é outro modelo utilizado, que descreve a série em termos de sua tendência, sazonalidade e erro. Além disso, um modelo linear (LM) foi empregado, oferecendo uma abordagem simples, mas útil, quando as relações entre as variáveis são aproximadamente lineares e de fácil interpretação. Cada modelo foi avaliado com base em métricas como Erro Absoluto Médio (MAE), Erro Percentual Médio Absoluto (MAPE) e Raiz do Erro Quadrático Médio (RMSE) para determinar sua eficácia na previsão dos preços da soja.

O modelo de *Neural Network* utilizado foi o *Multi-Layer Perceptron* (MLP), utilizando o pacote “nnfor” do software R. Para a estimativa do modelo LSTM foi utilizado o algoritmo “Keras” do software R. Para ambos a função de ativação utilizada foi a tangente hiperbólica a qual apresentou o melhor resultado, os pesos atribuídos por essa função variam entre -1 e 1, além disso, a arquitetura do MLP foi composta por sete camadas escondidas, enquanto para o LSTM seis camadas escondidas foram utilizadas. Dessa forma, diante das estatísticas calculadas o modelo que apresentou a melhor performance com os menores indicadores de erro foi o modelo MLP, com um erro percentual médio de predição de 0,90%, o que representa 0,76

centavos do preço médio, conforme apontado pelo MAE. Nota-se ainda que o modelo LSTM foi superior aos demais modelos apresentados, portanto, os modelos de *Machine* e *Deep Learning* utilizados foram superiores ao modelo tradicional.

5. Conclusão

O objetivo deste estudo foi criar modelos de previsão para os preços do Indicador da Soja ESALQ/BM&FBOVESPA - Paranaguá, que sirvam como ferramentas na tomada de decisão para compra e venda dessa *commodity*, que é de extrema importância para a economia brasileira. Para isso foram utilizados os modelos: ARIMA, MLP, LSTM e outros modelos com técnicas de *ensemble*.

Os resultados indicaram que os modelos de *Neural Network* foram superiores ao modelo tradicional ARIMA. Essa superioridade pode ser atribuída à capacidade das redes neurais de capturar padrões complexos e não-lineares nos dados, além de lidar de forma mais eficaz com a sazonalidade e a não-estacionariedade da série temporal. A flexibilidade, adaptabilidade e maior poder de representação dos modelos de *Neural Network* contribuíram para uma melhor performance na previsão dos preços da *commodity* agrícola.

Essa descoberta é crucial, pois destaca a importância de buscar constantemente modelos inovadores para aprimorar o processo de tomada de decisão na análise e previsão de séries temporais econômicas, especialmente em setores tão complexos e dinâmicos como o agronegócio. A habilidade desses modelos de aprendizado de máquina e deep learning em lidar com a complexidade dos dados e identificar padrões não-lineares pode oferecer valiosas vantagens estratégicas para tomadores de decisão, investidores e stakeholders do setor agrícola. Essas ferramentas podem contribuir significativamente para uma gestão mais eficiente de riscos, permitindo um planejamento estratégico mais preciso e uma otimização mais eficaz dos recursos disponíveis. Portanto, este estudo não só avança o conhecimento teórico sobre métodos de previsão de preços, mas também tem implicações práticas profundas e positivas para a indústria agrícola e para a economia em geral.

Apesar dos resultados promissores, é deve-se reconhecer limitações no estudo, como a dependência dos dados disponíveis e decisões metodológicas. Futuras pesquisas podem explorar abordagens avançadas, como modelos de *ensemble*, para melhorar a precisão das previsões. Esses modelos têm potencial para capturar uma variedade mais ampla de padrões nos dados e mitigar possíveis vieses. Além disso, a incorporação de informações adicionais, como dados de mídias sociais, pode melhorar ainda mais a precisão das previsões e fornecer informações mais abrangentes sobre os mercados agrícolas.

Referências

- AHMED, N. K.; et al. An empirical comparison of machine learning models for time series forecasting. **Econometric reviews**, 29(5-6), 594-621, 2010.
- ALIU, F.; KUČERA, J.; HAŠKOVÁ, S. Agricultural Commodities in the Context of the Russia-Ukraine War: Evidence from Corn, Wheat, Barley, and Sunflower Oil. **Forecasting**, 5(1), 351-373, 2023.
- ARLOT, S.; CELISSE, A. A survey of cross-validation procedures for model selection. 2010.
- ARMSTRONG, J. S. **Principles of forecasting: a handbook for researchers and practitioners**. Boston, MA: Kluwer Academic, 2001.
- BAKAY, M. S.; AĞBULUT, Ü. 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, 2021.
- BARROS, G.S.C.; CASTRO, N.R.; MACHADO, G.C.; ALMEIDA, F.M.S.; ALMEIDA, A.N. **Boletim Mercado de Trabalho do Agronegócio Brasileiro**. Centro de Estudos Avançados em Economia Aplicada (CEPEA). Piracicaba, 4º trimestre de 2022, 2023.

- BERGMEIR, C.; BENÍTEZ, J. M. On the use of cross-validation for time series predictor evaluation. **Information Sciences**, v. 191, p. 192-213, 2012.
- BOX, G. E. P. et al. **Time series analysis: forecasting and control**. John Wiley & Sons, 2015.
- BOX, G. E. P., JENKINS, G. M.; REINSEL, G. C. **Time series analysis: forecasting and control**. 3ª ed. New York: Wiley, 1994.
- BOX, G. E. P.; JENKINS, G. M. **Time Series Analysis: Forecasting and Control**, Holden Day, San Francisco, 1976.
- BRASIL. Senado Federal. Queda nas ‘commodities’ reduz em US\$ 25 bi exportações do Brasil. Disponível em < <https://www2.senado.leg.br/bdsf/bitstream/handle/id/513776/noticia.html?sequence=1> > Acesso em 07 de mar. de 2023.
- CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA (CEPEA) E ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS DE ÓLEOS VEGETAIS (ABIOVE). **PIB da Cadeia de Soja e do Biodiesel**. 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 de ago. de 2023.
- CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA (CEPEA) E CONFEDERAÇÃO NACIONAL DA AGRICULTURA E PECUÁRIA (CNA). **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 de ago. de 2023.
- CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA (CEPEA). MILHO/CEPEA: INCERTEZAS SOBRE EMBARQUES E DÓLAR ENFRAQUECEM PREÇO INTERNO, 2013. Disponível em < <https://www.cepea.esalq.usp.br/br/diarias-de-mercado/milho-cepea-incertezas-sobre-embarques-e-dolar-enfraquecem-preco-interno.aspx> > acessado em 07 de out. de 2022.
- CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA (CEPEA). MILHO/CEPEA: SOB INFLUÊNCIA DE CHICAGO, PREÇOS DESPENCAM EM PARANAGUÁ, 2012. Disponível em: < <https://www.cepea.esalq.usp.br/br/diarias-de-mercado/milho-cepea-sob-influencia-de-chicago-precos-despencam-em-paranagua.aspx> > acessado em 07 de out. de 2022.
- CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA (CEPEA). MILHO/RETRO 2018: CLIMA PREJUDICA PRODUÇÃO EM 2018, PREÇO INTERNO SOBE E EXPORTAÇÃO CAI, 2019. Disponível em < <https://www.cepea.esalq.usp.br/br/releases/milho-retro-2018-clima-prejudica-producao-em-2018-preco-interno-sobe-e-exportacao-cai.aspx> > acessado em 07 de out. de 2022.
- CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA (CEPEA-ESALQ/USP). Mercado de Trabalho/Cepea: Número de Trabalhadores no Setor Sucro Cai, mas Qualidade dos Empregos Cresce, 2018. Disponível em < <https://www.cepea.esalq.usp.br/br/releases/mercado-de-trabalho-cepea-numero-de-trabalhadores-no-setor-sucro-cai-mas-qualidade-dos-empregos-cresce.aspx> > acessado em 10 de ago. de 2023.
- CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA-CEPEA . INDICADOR DA SOJA ESALQ/BM&FBOVESPA - PARANAGUÁ. Disponível em: < <https://www.cepea.esalq.usp.br/br/indicador/soja.aspx> >. Acesso em: 25 mar. 2024.
- CHEN, Q., et al. Empirical mode decomposition based long short-term memory neural network forecasting model for the short-term metro passenger flow. **PloS one**, 14(9), e0222365, 2019.

COMPANHIA NACIONAL DE ABASTECIMENTO (CONAB). Boletim da Safra de Grãos, 2023. Disponível em < <https://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos> >. Acesso em 13 de ago. de 2023.

COMPANHIA NACIONAL DE ABASTECIMENTO (CONAB). Conjuntura da Agropecuária, SOJA – Agosto/2021. Disponível em < https://www.conab.gov.br/info-agro/analises-do-mercado-agropecuario-e-extrativista/analise-regional-do-mercado-agropecuario/analise-regional-mg-soja/item/download/39114_0897b95d6388335ff1ca42aa6a6dd912#:~:text=O%20m%C3%AAs%20de%20Agosto%20fechou,mesmo%20per%C3%ADodo%20ao%20ano%20anterior >. Acesso em 13 de ago. de 2023.

CONFEDERAÇÃO NACIONAL DA AGRICULTURA E PECUÁRIA (CNA). Panorama do Agro, edição 22 de 2023. Disponível em < https://www.cnabrazil.org.br/storage/arquivos/files/dtec.panorama-agro-ed22_12-a-16junho2023.v2.pdf >. Acesso em 13 de ago. de 2023.

CRONE, S. F.; KOURENTZES, N. Feature selection for time series prediction—A combined filter and wrapper approach for neural networks. **Neurocomputing**, v. 73, n. 10-12, p. 1923-1936, 2010.

DUAN, Q., et al. Forecasting model and validation for aquatic product price based on time series GA-SVR. **Transactions of the Chinese Society of Agricultural Engineering**, 33(1), 308-314, 2017.

DURBIN, J.; WATSON, G. S. Testing for serial correlation in least squares regression. III. **Biometrika**, v. 58, n. 1, p. 1-19, 1971.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA – EMBRAPA. Milho: Importância Socioeconômica, 2021. 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).) > acessado em 10 de ago. de 2023.

FOOD AND AGRICULTURE ORGANIZATION (FAO). **Crops and livestock products**, 2023. Disponível em < <https://www.fao.org/faostat/en/#data/QCL/visualize> >. Acesso em 13 de ago. de 2023.

GOVIND, G. R.; BABU, A. S. Hybrid Deep Learning Model to Forecast Crude Oil Price. In: **2023 International Conference on Inventive Computation Technologies (ICICT)**. IEEE, 2023. p. 19-23.

GOWTHAMAN, et al. A Review of Deep Learning Models for Price Prediction in Agricultural Commodities. **Economic Affairs**, 68(1), 561-568, 2023.

GRANGER, C. W. J. Investigating causal relations by econometric models and cross-spectral methods. **Econometrica: journal of the Econometric Society**, p. 424-438, 1969.

HAMILTON, J. D. **Time Series Analysis**. Princeton University Press, 1994.

HASSOUN, M. H. **Fundamentals of artificial neural networks**. MIT press, 1995.

HIRAKURI, M. H.; LAZZAROTTO, J. J. O agronegócio da soja nos contextos mundial e brasileiro. **Embrapa Soja-Documents (INFOTECA-E)**, 2014.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural computation**, v. 9, n. 8, p. 1735-1780, 1997.

HOSSAIN, M. A. et al. Hybrid deep learning model for stock price prediction. In: **IEEE Symposium Series on Computational Intelligence, SSCI, Bangalore, India**. p. 1837-1844, 2018.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. 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 > acessado em 10 de ago. de 2023.

JADHAV, V.; CHINNAPPA, REDDY BV; GADDI, G. M. Application of ARIMA model for forecasting agricultural prices. **J. Agr. Sci. Tech.** Vol. 19: 981-992, 2017.

JAVANMARD, M. E.; GHADERI, S. F. A hybrid model with applying machine learning algorithms and optimization model to forecast greenhouse gas emissions with energy market data. **Sustainable Cities and Society**, v. 82, p. 103886, 2022.

JESUS, G. S. D.; et al. LSTM Applied for Commodity Futures Price Forecast. **Preprints** 2022, 2022110278. <https://doi.org/10.20944/preprints202211.0278.v1>

KAASTRA, I.; BOYD, M. Designing a neural network for forecasting financial and economic time series. **Neurocomputing**, v. 10, n. 3, p. 215-236, 1996.

KARIA, A. A.; BUJANG, I. Progress accuracy of CPO price prediction: Evidence from ARMA family and artificial neural network approach. **International Research Journal of Finance and Economics**, 64(64), 66-79, 2011.

KENNEDY, P. **A Guide to Econometrics**. 2. ed. Cambridge: The MIT Press, 1991.

KHASHEI, M.; BIJARI, M. An artificial neural network (p, d, q) model for timeseries forecasting. **Expert Systems with applications**, v. 37, n. 1, p. 479-489, 2010.

KOSHIYAMA, A.; FIROOZYE, N.; TRELEAVEN, P. Algorithms in future capital markets. **Available at SSRN 3527511**, 2020.

KUMAR, A. G.; SINDHU, M. R.; KUMAR, S. S. Deep neural network based hierarchical control of residential microgrid using LSTM. In: **TENCON 2019-2019 IEEE Region 10 Conference (TENCON)**. IEEE, 2019. p. 2129-2134.

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.

KWIATKOWSKI, D. et al. 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, 1992.

MAKRIDAKIS, S.; WHEELWRIGHT, S. C.; HYNDMAN, R. J. **Forecasting methods and applications**. Wiley-India. New Delhi, 2008.

MENHAJ, M. H.; KAVOOSI-KALASHAMI, M. Developing a hybrid forecasting system for agricultural commodity prices (case study: Thailand rice free on board price). **Ciência Rural**, v. 52, p. e20201128, 2022.

MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO. Secretaria de Relações Internacionais do Agronegócio. Balança Comercial do Agronegócio – Dezembro/2016. Disponível em: <https://www.gov.br/agricultura/pt-br/assuntos/noticias/com-aumento-36-6-exportacao-de-acucar-e-destaque-na-balanca-comercial-do-2016/nota-dezembro-2016.docx>. Acesso em: 25 dez. 2023.

MINISTÉRIO DA INDÚSTRIA, COMÉRCIO EXTERIOR E SERVIÇOS – MDIC. COMEXSTAT: Exportação e Importação Geral. Disponível em < <http://MIDIC.mdic.gov.br/pt/geral> > Acessado em 07 de ago. de 2023.

MINISTÉRIO DA INDÚSTRIA, COMÉRCIO EXTERIOR E SERVIÇOS – MDIC. Exportação e Importação Geral. Disponível em < <http://MIDIC.mdic.gov.br/pt/geral> > Acessado em 07 de out. de 2022.

MITRA, D.; PAUL, R. K. Hybrid time-series models for forecasting agricultural commodity prices. **Model Assisted Statistics and Applications**, 12(3), 255-264, 2017.

MONGE, M.; LAZCANO, A. Commodity prices after COVID-19: Persistence and time trends. **Risks**, v. 10, n. 6, p. 128, 2022.

MORETTIN, P. A. **Econometria financeira: um curso em séries temporais financeiras**. 2ª ed. Editora Blucher, 2011.

- NABIPOUR, M. et al. 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, 2020.
- NASSAR, L. et al. Deep Learning based approach for fresh produce market price prediction. In: **2020 International Joint Conference on Neural Networks (IJCNN)**. IEEE. p. 1-7, 2020.
- ONOUR, I.; SERGI, B. S. Modeling and Forecasting Volatility in the Global Food Commodity Prices (Modelování a Prognózování Volatility Globálních cen Potravinářských Komodit). **Agricultural Economics-Czech**, v. 57, n. 3, p. 132-139, 2011.
- ORTU, M. et al. On technical trading and social media indicators for cryptocurrency price classification through deep learning. **Expert Systems with Applications**, v. 198, p. 116804, 2022.
- PATHINARUPOTHI, R. K. et al. 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)**. IEEE, 2017. p. 293-296.
- PAUL, R. K. et al. Machine learning techniques for forecasting agricultural prices: A case of brinjal in Odisha, India. **Plos one**, v. 17, n. 7, p. e0270553, 2022.
- PINDYCK, R. S.; RUBINFELD, D. L. **Econometria: modelos e previsões**. Rio de Janeiro: Elsevier, 2004.
- SANCHES, A.; SALES, C. Expressivo Aumento nas Exportações Brasileiras de Milho impõe Mudanças na Dinâmica do Mercado Nacional, 2022. 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> > acessado em 10 de out. de 2022.
- SANUSI, O. I., et al. Forecasting agricultural commodity price using different models: a case study of widely consumed grains in Nigeria. **Agricultural and Resource Economics: International Scientific E-Journal**, v. 8, n. 1868-2022-778, p. 124-140, 2022.
- SECRETARIA DE PLANEJAMENTO, GOVERNANÇA E GESTÃO DEPARTAMENTO DE ECONOMIA E ESTATÍSTICA. Indicadores. Disponível em < <https://dee.rs.gov.br/indicadores> > acessado em 10 de ago. de 2022.
- SIMS, C. A. Macroeconomics and reality. **Econometrica: journal of the Econometric Society**, v. 48, p. 1-48, 1980.
- STONE, M. Cross-validatory choice and assessment of statistical predictions. **Journal of the royal statistical society: Series B (Methodological)**, v. 36, n. 2, p. 111-133, 1974.
- SUMATHI, S.; PANEERSELVAM, S. **Computational intelligence paradigms: theory & applications using MATLAB**. crc Press, 2010.
- SWAPNA, G.; KP, S.; VINAYAKUMAR, R. Automated detection of diabetes using CNN and CNN-LSTM network and heart rate signals. **Procedia computer science**, v. 132, p. 1253-1262, 2018.
- TIMMERMAN, A. Chapter 4 forecast combinations. **Handbook of Economic Forecasting**, v. 1, p. 135.
- TORRES-PRUÑONOSA, J. et al. How on Earth Did Spanish Banking Sell the Housing Stock? **Sage Open**, v. 12, n. 1, p. 21582440221079916, 2022.
- UNITED STATES DEPARTMENT OF AGRICULTURE – USDA. World Agricultural Production, 2023. Disponível em < <https://www.fas.usda.gov/data/world-agricultural-production> > acessado em 10 de ago. de 2023.
- UNITED STATES DEPARTMENT OF AGRICULTURE, USDA. Corn Exports in 2021, 2022. Disponível em < <https://www.fas.usda.gov/commodities/corn> > acessado em 10 de ago. de 2023.

- VANCSURA, L.; TATAY, T.; BAREITH, T. Evaluating the Effectiveness of Modern Forecasting Models in Predicting Commodity Futures Prices in Volatile Economic Times. **Risks**, v. 11, n. 2, p. 27, 2023.
- VIDAL, M. F. Açúcar: cenário mundial e situação de produção no Brasil e no Nordeste brasileiro. **Escritório Técnico de Estudos Econômicos do Nordeste – ETENE**. Ano 7, Nº 215, Mar. de 2022.
- WOOLDRIDGE, J. F. **Introdução à Econometria: uma abordagem moderna**. 3ª ed. São Paulo, SP: Cengage Learning, 2019.
- XU, X. Contemporaneous and granger causality among us corn cash and futures prices. **European Review of Agricultural Economics**, v. 46, n. 4, p. 663-695, 2019.
- XU, X. Contemporaneous causal orderings of US corn cash prices through directed acyclic graphs. **Empirical Economics**, v. 52, p. 731-758, 2017.
- XU, X.; ZHANG, Y. 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, 2022.
- YANG, X. et al. Evaluation of short-term freeway speed prediction based on periodic analysis using statistical models and machine learning models. **Journal of Advanced Transportation**, v. 2020, 2020.
- YASHAVANTH, B. S., et al. Forecasting prices of coffee seeds using Vector Autoregressive Time Series Model. **Indian Journal of Agricultural Sciences**, 87(6): 754–8, 2017.
- ZHAO, H. Futures price prediction of agricultural products based on machine learning. **Neural computing and applications**, v. 33, p. 837-850, 2021.
- ZHU, Q., et al. Time Series Analysis Based on Informer Algorithms: A Survey. **Symmetry**, 15(4), p.951, 2023.
- ZIVOT, E.; WANG, J. **Modeling financial time series with S-PLUS**. New York: Springer, 2006.
- FRANCO, Cleiton; FIGUEIREDO, ADRIANO MARCOS RODRIGUES. OS EFEITOS DA COVID-19 NOS CUSTOS DE PRODUÇÃO E LUCRATIVIDADE DAS CULTURAS DE MILHO E SOJA NO MATO GROSSO. In: **20th CONTECSI-INTERNATIONAL CONFERENCE ON INFORMATION SYSTEMS AND TECHNOLOGY MANAGEMENT VIRTUAL**. 2024.
- KRETER, A. C. E SOUZA JR. J. R. C. INSTITUTO DE PESQUISA ECONÔMICA APLICADA (IPEA). Carta de Conjuntura, Mercados e Preços Agropecuários, agosto de 2021. Disponível em: <<https://www.ipea.gov.br/cartadeconjuntura/index.php/2021/08/mercados-e-precos-agropecuarios-2/>>. Acesso em: 25 dez. 2023.
- CONFEDERAÇÃO DA AGRICULTURA E PECUÁRIA DO BRASIL (CNA). "PIB do Agronegócio Alcança Participação de 26,6% no PIB Brasileiro em 2020". Disponível em: <<https://www.cnabrazil.org.br/publicacoes/pib-do-agronegocio-alcanca-participacao-de-26-6-no-pib-brasileiro-em-2020>>. Acesso em: 25 dez. 2023.
- BASTOS, E. K.X.; LEITE, C. R. G. INSTITUTO DE PESQUISA ECONÔMICA APLICADA (IPEA). Preços de Commodities. Disponível em: <<https://www.ipea.gov.br/cartadeconjuntura/index.php/tag/precos-de-commodities/>>. Acesso em: 25 dez. 2023.