

**MODELOS PREDITIVOS PARA PREÇOS DE AÇÚCAR E ETANOL:  
COMPARAÇÃO ENTRE ABORDAGENS TRADICIONAIS E TÉCNICAS DE  
APRENDIZADO DE MÁQUINA**  
*PREDICTIVE MODELS FOR SUGAR AND ETHANOL PRICES: A COMPARISON  
BETWEEN TRADITIONAL APPROACHES AND MACHINE LEARNING  
TECHNIQUES*

**Autor(es):** Fernanda Cigainski Lisbinski<sup>1</sup>; Felipe André Oliveira Freitas<sup>2</sup>; Heloisa Lee Burnquist<sup>3</sup>

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

**E-mail:** lisbinskif@gmail.com<sup>1</sup>; felipeoliveirafreitas@usp.br<sup>2</sup>; hlburnqu@usp.br<sup>3</sup>

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

**Resumo**

Os modelos de previsão de preços de commodities são uma ferramenta capaz proporcionar cenários que oferecem maior segurança a investidores, produtores agrícolas entre outros, que são relacionados a cadeia de produção do açúcar e de biocombustíveis. Contudo, a dinâmica comercial bem como os fatores que influenciam a formação do preço de commodities estão em constante evolução, o que requer atualizações constantes das modelagens preditivas à medida que novas informações relevantes se tornam disponíveis, bem como a própria evolução de metodologias de previsão caminham. Nesse sentido, este estudo teve como objetivo desenvolver modelos de previsão de preços de açúcar e etanol dos estados de São Paulo e Alagoas. Foram utilizados os indicadores do Açúcar Cristal Branco e do Etanol Hidratado Combustível CEPEA/ESALQ do estado de São Paulo e de Alagoas Mercado Interno. Utilizou-se os modelos ARIMA e VAR como base para comparação com modelos de aprendizado de máquina e redes neurais, incluindo o MLP, LSTM e Transformer, avaliando seus respectivos indicadores de acurácia como RMSE e MAPE. Os resultados mostraram que as técnicas de aprendizado de máquina, especialmente LSTM e Transformer, foram mais acuradas na previsão de séries temporais com comportamento não linear, como para os preços do açúcar em São Paulo e os preços do etanol em ambos os mercados. Enquanto o modelo MLP e o ARIMA apresentaram melhor acurácia para séries temporais com comportamento linear, como para o açúcar de Alagoas. . Em geral, os modelos de aprendizado de máquina foram precisos, proporcionando uma ferramenta de suporte à decisão estratégica no mercado sucroenergético.

**Palavras-chave:** Modelos de Previsão; Mercado Sucroenergético; Redes Neurais; LSTM; Transformer.

**Abstract**

Commodity price forecasting models are valuable tools that can provide scenarios offering greater security for investors, agricultural producers, and other stakeholders involved in the sugar and biofuels production chain. However, both the commercial dynamics and the factors influencing commodity price formation are constantly evolving, which requires frequent updates to predictive models as new relevant data becomes available, along with the ongoing advancement of forecasting methodologies. In this context, the objective of this study was to develop price forecasting models for sugar and ethanol in the states of São Paulo and Alagoas. The analysis used CEPEA/ESALQ price indicators for White Crystal Sugar and Hydrous Fuel Ethanol in the São Paulo market, as well as internal market data from Alagoas. ARIMA and VAR models were used as baselines for comparison with machine learning and neural network models, including MLP, LSTM, and Transformer, and their respective accuracy metrics such as RMSE and MAPE were evaluated. The results showed that machine learning techniques—particularly LSTM and Transformer—were more accurate in forecasting time series with nonlinear behavior, such as sugar prices in São Paulo and ethanol prices in both markets. In contrast, MLP and ARIMA models performed better with time series that displayed linear behavior, such as sugar prices in Alagoas. Overall, machine learning models proved to be accurate and effective, offering a robust decision-support tool for strategic planning in the sugar-energy sector.

**Key words:** Forecasting Models; Sugar-Energy Market; Neural Networks; LSTM; Transformer.

**1. Introdução**

A previsão de preços é uma questão central para os participantes do mercado agrícola, incluindo corretores, hedges, processadores e produtores. Ela é essencial para especuladores na busca por lucros, para produtores que precisam fixar preços de venda antes da produção, e para

processadores e exportadores que buscam atender às demandas do mercado (Xu e Zhang, 2022). Os preços agrícolas, caracterizados por sua volatilidade (Xu, 2017), desempenham um papel vital nas decisões estratégicas, influenciando a alocação de recursos e o bem-estar econômico, especialmente em países onde a economia é fortemente dependente do setor primário (Xu, 2019).

Fornecer informações precisas sobre a previsão de preços pode auxiliar na tomada de decisões estratégicas para a gestão de suprimentos agrícolas e otimização do comportamento de compra dos consumidores. Além da sazonalidade associada à produção, influenciada por variáveis climáticas como regime de chuvas e intensidade luminosa, os preços são também afetados por fatores como preferências dos consumidores, estratégias comerciais, elementos culturais e comportamento dos fornecedores. Esses fatores humanos, embora não apresentem necessariamente sazonalidade, podem ser analisados em termos de relações entre eventos passados e a situação atual (Kurumatani, 2020). Portanto, o estudo e desenvolvimento de modelos de previsão de preços de commodities emerge como uma ferramenta essencial ao proporcionar maior segurança a investidores, produtores agrícolas e formuladores de políticas, mitigando riscos e permitindo que órgãos governamentais desenvolvam regulamentações e políticas agrícolas mais adequadas (Zhao, 2021).

Além disso, esses modelos exigem constantes atualizações à medida que novas informações relevantes se tornam disponíveis e as metodologias evoluem. As características dos preços das commodities agrícolas são frequentemente ruidosas e não estacionárias, o que impõe desafios significativos à previsão (Xiong et al., 2015) e, muitas vezes, resulta em estimativas menos eficientes. Ademais, os preços das commodities são influenciados por uma ampla gama de fatores externos, como clima, política e economia global (Baffes e Nagle, 2022), tornando difícil prever com precisão como esses elementos irão evoluir no futuro.

Neste contexto, identificar e compreender o conjunto de fatores que determinam os preços de mercado, especialmente no caso das *commodities* agroindustriais, é uma tarefa ainda mais complexa, dada a multiplicidade de dimensões envolvidas. Ao longo do tempo, os preços refletem uma combinação intrincada de mudanças associadas a fatores sazonais, cíclicos, tendências e fatores irregulares. A regularidade mais comum observada nos preços agrícolas é o padrão sazonal de variações. Normalmente, os preços de commodities agrícolas que podem ser estocadas, como o açúcar e o etanol, tendem a ser mais baixos durante o período de safra da cana (no Centro-Sul do Brasil, de abril a novembro, e no Norte-Nordeste, de outubro a março) e aumentam à medida que a produção avança, atingindo picos ao final da safra. Produtos semi-perenes, como a cana-de-açúcar, também exibem comportamentos cíclicos, que variam em duração e amplitude, respondendo a altos preços que, por sua vez, estimulam a expansão da produção, ajustando o mercado (Baffes e Nagle, 2022).

Diante disso, o objetivo desta pesquisa consiste em desenvolver modelos de previsão para o preço do açúcar e do etanol no mercado à vista brasileiro para os mercados de São Paulo e Alagoas, comparando metodologias tradicionais e modernas para identificar as mais eficientes, proporcionando previsões mais precisas que auxiliem na tomada de decisões.

A escolha do tema deste trabalho justifica-se pela complexidade de desenvolver modelos de previsão para os preços do açúcar e do etanol, dada sua alta volatilidade e sensibilidade a fatores como clima, políticas governamentais e demanda global. Essas dificuldades impactam diretamente o processo de tomada de decisão dos produtores e exportadores em relação ao plantio, colheita, produção e comercialização do açúcar. Além disso, a previsão de preços dessas commodities é essencial para a segurança jurídica no setor sucroenergético, proporcionando estabilidade e previsibilidade nas atividades econômicas. Com previsões confiáveis, produtores, comercializadores e processadores podem tomar decisões informadas, mitigando riscos de flutuações inesperadas nos preços. Isso facilita a

negociação de contratos, reduzindo disputas legais e promovendo um ambiente de negócios mais estável (Carpio, 2019).

Adicionalmente, previsões precisas são essenciais para a criação de políticas públicas que promovam a estabilidade do mercado. Pois elas são importantes na gestão de riscos, permitindo o uso eficaz de instrumentos financeiros e, consequentemente, incentivando financiamentos e investimentos de longo prazo (Geman, 2009). Além disso, governos e organizações podem usar essas previsões para implementar programas de assistência alimentar mais eficazes, alocando recursos para atender às populações mais vulneráveis (Barrett, 2021).

Destaca-se ainda, que a previsão de preços de açúcar e etanol é essencial para o setor de biocombustíveis, uma vez que ambos os produtos utilizam a cana-de-açúcar como matéria-prima, alternando esse mix conforme o preço. Previsões confiáveis ajudam a formular políticas energéticas e ambientais mais eficazes, promovendo práticas sustentáveis e incentivando o uso de energias renováveis (Goldemberg et al., 2018). Além disso, a pesquisa apoia os Objetivos de Desenvolvimento Sustentável (ODS) das Nações Unidas, especialmente os relacionados à energia limpa e acessível (ODS 7) e ação contra a mudança global do clima (ODS 13) (Nações Unidas - ONU, 2024). Ao integrar essas práticas sustentáveis, a pesquisa contribui para a implementação de políticas que promovem a sustentabilidade no setor sucroenergético, alinhando-se com os esforços globais para um futuro mais sustentável.

A literatura apresenta diversas metodologias que são utilizadas para a construção de modelos de previsão de preços, a mais tradicional é a desenvolvido 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. A partir dessa metodologia autores como Amrouk e Heckeleei (2020), Tatarintsev et al. (2021) e Sahinli (2021) utilizaram modelos tradicionais como Modelo Bayesiano Médio (BMA), ARIMA, VAR (Vector AutoRegression), e modelos de volatilidade como ARCH (AutoRegressive Conditional Heteroskedasticity), GARCH (Generalized AutoRegressive Conditional Heteroskedasticity) e EGARCH (Exponential Generalized AutoRegressive Conditional Heteroskedasticity) para a previsão de preços do açúcar, demonstrando alta capacidade preditiva desses modelos.

Contudo, diante de avanços na modelagem e abordagens estatísticas para previsões do preço, paralelamente ao massivo aumento do aporte computacional nas últimas décadas, incluindo abordagens de Machine Learning (ML) e Deep Learning (DL), como a Neural Network (NN), vêm se apresentando como vantajosas e promissoras na melhoria dos modelos de predição. Autores como Silva et al. (2019), Xu e Zhang (2022), Santos et al. (2021), Zhang et al. (2021) e Palazzi et al. (2023) utilizaram modelos ML para a previsão dos preços de açúcar e etanol, mostrando que tais técnicas apresentam grande eficiência e acurácia de predição, sendo mais favoráveis principalmente em cenários onde a série temporal apresenta comportamentos complexos e dinâmicos (Hossain et al., 2018; Nabipour et al. 2020).

Nesse sentido, destaca-se que o objetivo deste trabalho é buscar a máxima precisão possível ao desenvolver modelos de previsão que ofereçam cenários balizadores. Esses cenários, quando acompanhados de outras informações relevantes, podem fornecer uma base sólida para a tomada de decisões assertivas. Ao oferecer uma visão abrangente das possíveis variações de preços, os modelos de previsão ajudam produtores, comerciantes e formuladores de políticas a entender melhor as tendências de mercado e a preparar estratégias de mitigação de riscos.

## **2. Os Fatores Determinantes dos Preços no Mercado Sucroenergético Brasileiro**

O mercado de açúcar e etanol no Brasil é um dos mais importantes do mundo, sendo o país o maior produtor e exportador global de açúcar e o segundo maior produtor de etanol, atrás apenas dos Estados Unidos. O mercado sucroenergético brasileiro é tradicionalmente dividido

em duas grandes regiões produtoras: Centro-Sul e Norte-Nordeste. O Centro-Sul, que inclui estados como São Paulo, Goiás, Minas Gerais e Paraná, é responsável por cerca de 90% da produção nacional de cana-de-açúcar, destacando-se como o principal polo produtor de açúcar e etanol do país. Já o Norte-Nordeste, com estados como Alagoas, Pernambuco e Paraíba, tem uma produção menor, mas de grande importância histórica e econômica, especialmente para o mercado interno de açúcar e o abastecimento regional. Essa divisão regional reflete não apenas diferenças na escala de produção, mas também nas características climáticas, econômicas e tecnológicas, que influenciam a dinâmica do mercado em cada área (Conab, 2024; Unica, 2024).

O Brasil possui uma infraestrutura altamente desenvolvida para a produção de açúcar e etanol. A safra da cana-de-açúcar, que vai de abril a novembro, com comercialização durante o ano todo, impacta diretamente a oferta e os preços de açúcar e etanol, com picos de produção resultando em maior disponibilidade e preços mais baixos durante esse período (UNICA, 2024). São Paulo, o principal estado produtor, é responsável por aproximadamente 60% da produção nacional de açúcar e etanol. O Norte-Nordeste, é composto por Alagoas, Pernambuco e Paraíba, sendo que Alagoas se destaca como o principal estado produtor, especialmente relevante para a produção de açúcar destinado ao mercado interno e externo (MAPA, 2023). Os preços dos produtos diferem entre as regiões, na região Norte-nordeste o preço é maior do que na região Centro-sul. De acordo com Guilhoto et al. (2014) e Pereira (2017) a falta de mecanização e as condições menos favoráveis de infraestrutura no Nordeste elevam os custos de insumos e mão de obra, impactando diretamente os preços dos produtos dessa região. Além disso, a produtividade no Centro-Sul, especialmente em São Paulo, é maior, diluindo custos fixos e permitindo preços mais baixos na comparação com o Nordeste, onde a produção é mais limitada e os custos são menos diluídos.

No Brasil, a interação entre açúcar e etanol é outro fator que impacta diretamente os preços, com ambos os produtos competindo pela cana-de-açúcar. Quanto maior a razão preço açúcar-etanol, maior a parcela de cana-de-açúcar destinada ao açúcar (Antunes, 2019). Assim, as usinas de açúcar no Brasil possuem flexibilidade para alternar entre a produção de açúcar e etanol, dependendo da rentabilidade de cada produto. Quando o preço do açúcar é mais favorável, a produção pode ser direcionada para ele, aumentando a oferta no mercado internacional. Dessa forma, na safra 23/24 houve um aumento na produção de açúcar de 25,5% em relação à safra anterior enquanto a produção de etanol apresentou um aumento de 16,4% (Unica, 2024). Com isso, até outubro de 2024 foram exportadas 32,04 milhões de toneladas de açúcar, um novo recorde de exportações brasileiras do produto. Assim, essa capacidade de alternar entre os produtos, aliada à posição dominante do Brasil no mercado mundial de açúcar, indica que as decisões das usinas afetam diretamente a disponibilidade de açúcar para exportação, influenciando, assim, os preços internacionais do produto.

Outro fator que impacta o mercado sucroenergético brasileiro é o mercado internacional. O Brasil é o maior produtor de cana de açúcar do mundo, evidenciado pela produção de 610,8 milhões de toneladas de cana na safra 2022/23, representando um crescimento de 4,38% em relação à safra anterior (2021/22). Além disso, o país mantém sua posição como o maior produtor e exportador mundial de açúcar, alcançando 36,8 milhões de toneladas na safra de 2022/23 (Conab, 2024). Em 2023, o complexo sucroenergético brasileiro exportou US\$ 17,38 bilhões, representando 10,44% das exportações do agronegócio tendo como principais destinos Ásia, África e Oriente Médio (Mapa/Agrostat, 2023). O etanol, por sua vez, apresentou uma participação pequena no comércio internacional, em 2023, foram exportados 2,56 bilhões de litros de etanol, um aumento de 2,5% em relação ao ano anterior. Os principais destinos do etanol brasileiro incluem a Coreia do Sul, que representou 30% das exportações, seguida pela Índia e Arábia Saudita, que responderam por 23% e 18%, respectivamente. A Europa,

especialmente através dos portos dos Países Baixos, também foi um destino significativo para o etanol, absorvendo 23% do total exportado (MDIC, 2024). Além disso, é amplamente utilizado no mercado interno como combustível alternativo, devido à política de mistura obrigatória de etanol anidro na gasolina, que varia entre 18% e 27,5%, conforme regulamentação do governo federal (ANP, 2024).

Entre 2020 e 2022, o Brasil foi responsável por 38,8% da produção mundial de açúcar, uma leve queda em relação aos 39,5% registrados entre 2017 e 2019 (FAO, 2024). Em termos de exportações, o Brasil desempenha um papel ainda mais significativo: em 2023, foi responsável por 83% do valor das exportações mundiais de açúcar bruto, totalizando US\$ 13,35 bilhões (UnComtrade, 2024). A posição dominante do Brasil no mercado mundial implica que qualquer choque de oferta no país, seja por questões climáticas ou econômicas, afeta diretamente os preços internacionais e os fluxos comerciais de açúcar. Assim, os preços globais tendem a seguir as variações no custo marginal de produção no Brasil, reforçando a importância estratégica do país no comércio internacional do produto.

O mercado sucroenergético brasileiro também é impactado por políticas governamentais que visam tanto a regulação da produção quanto a promoção de energias renováveis. Atualmente a Política de Mistura de Etanol, que estabelece uma porcentagem obrigatória de etanol anidro misturado à gasolina (variando entre 18% e 27,5%), é uma das principais políticas que afeta o mercado (Ministério de Minas e Energia (MME), 2022). Propostas recentes como a Lei do “Combustível do Futuro”, sugerem aumentar essa mistura para até 35%, o que pode aumentar a demanda por etanol, mas diminuir a oferta de açúcar disponível para consumo e exportação (MME, 2023). Além disso, há o programa RenovaBio que incentiva a produção de biocombustíveis, como o etanol, com o objetivo de reduzir as emissões de carbono no setor de transporte, promovendo eficiência energética e sustentabilidade por meio de créditos de descarbonização (CBIOS) (ANP, 2024). Essas políticas buscam alinhar o setor a metas de sustentabilidade e segurança energética, impactando diretamente o equilíbrio entre a produção de açúcar e etanol no Brasil e consequentemente o preço desses produtos.

Por fim, as condições climáticas têm um grande impacto na produção de cana-de-açúcar no Brasil, e eventos extremos, como estiagens prolongadas, geadas e queimadas, podem afetar gravemente a oferta do produto, pressionando os preços no mercado sucroenergético. A seca atual, intensificada pelo fenômeno climático El Niño, tem causado perdas consideráveis nas lavouras de cana-de-açúcar em várias regiões produtoras, especialmente no Sudeste (World Economic Forum, 2023/USDA, 2023). Outro fator climático relevante são as queimadas, que, embora utilizadas historicamente para facilitar a colheita manual de cana, também trazem impactos ambientais e produtivos negativos. As queimadas podem comprometer a qualidade da cana e causar volatilidade na oferta. Recentemente, o Brasil tem adotado políticas para proibir as queimadas em áreas mecanizadas, no entanto o número de focos de incêndio na região sudeste, principal produtora de cana-de-açúcar, até 12 de setembro de 2024, aumentou em 78%, a nível de Brasil esse aumento foi de 27,8% (Instituto Nacional De Pesquisas Espaciais – INPE, 2024). Esses fatores climáticos levam a uma menor oferta de açúcar, que, quando combinada com uma demanda global estável ou crescente, pressiona os preços para cima. Além disso, a menor oferta de cana também afeta a produção de etanol, o que pode aumentar ainda mais os preços em ambos os mercados, já que as usinas brasileiras alternam entre a produção de açúcar e etanol conforme a rentabilidade (USDA, 2024).

Assim, no mercado interno, essas flutuações de preço influenciam diretamente a produção industrial, beneficiando os produtores de cana com melhores margens de lucro e incentivando o aumento da produção. Para os consumidores, o preço do açúcar afeta o custo de diversos alimentos, já que o açúcar é um ingrediente amplamente utilizado na indústria alimentícia. Em contextos de alta nos preços do açúcar, pode ocorrer um aumento no custo de

alimentos processados e produtos que dependem da cana-de-açúcar, afetando a acessibilidade para os consumidores, especialmente em países em desenvolvimento, como o Brasil (World Bank, 2013; Pantoja et al., 2016).

### 3. Metodologia

#### 3.1 O Modelo ARIMA

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 ARIMA ( $p, q$ ) com a seguinte estrutura:

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

onde  $X_t$  é o indicador de preço para açúcar ou etanol hidratado;  $d$  representa o número de diferenciações necessárias para tornar a série estacionária;  $\Delta^d X_t$  representa a diferenciação da série original para remover tendências;  $\theta_0 = \mu(1 - \phi_1 - \dots - \phi_p)$ , 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  $\varepsilon$ , o termo de erro da série, e  $\varepsilon_t$  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  $\varepsilon_t$  é considerado como tendo distribuição normal independente e identicamente distribuída com média zero e variância constante. O Modelo ARIMA utilizado nesta pesquisa segue o método Box e Jenkins.

#### 3.2 O Modelo de Artificial Neural Networks (ANN)

O algoritmo de ANN consiste em três camadas: 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-Pruñonosa et al. 2022; Zhang et al, 1998).

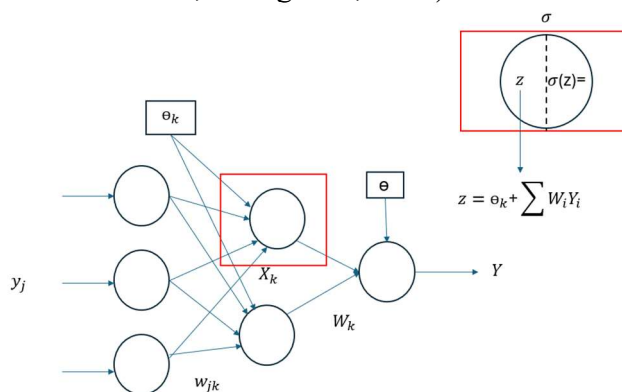


Figura 1 - *Artificial Neural Networks* para previsão de séries temporais.

Fonte: adaptado de Monge e Lazcano (2022).

Na Figura 1 apresenta-se a expressão gráfica de um modelo de ANN para previsão. Onde,  $Y_j$  representa o indicador de preço do açúcar ou do etanol, sendo um 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 NN

com um neurônio de saída  $Y$  representa a saída do modelo e é determinado pela fórmula  $Y = f(\sum_{j=1}^n W_k X_k + \Theta)$ .  $\sigma$  é a função de ativação. Por fim,  $\Theta_k (k = 1, 2, 3, \dots, k)$  é o valor de viés dos nós da camada oculta e  $\Theta$  é o valor de viés da camada de saída.

### 3.3 O Modelo *Multi-Layer Perceptrons Neural Network*

O MLP é composto por uma ou mais camadas ocultas entre as camadas de entrada e saída de uma rede. Assim, uma MLP apresenta melhor capacidade de aprender o padrão de entrada e produzir uma saída melhor do que a NN 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 dos indicadores de preços do açúcar ou do etanol. O  $I$  denota o número de entradas  $p_i$  da NN. Esse modelo pode ser definido a partir da seguinte equação:

$$\hat{y}_{t+1} = \beta_0 + \sum_{h=1}^H \beta_h g \left( \gamma_{0i} + \sum_{i=1}^I \gamma_{hi} p_i \right). \quad (2)$$

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  $\beta_0$  e  $\gamma_{0i}$  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 (Kourentzes e Crone, 2010; Kourentzes et al., 2014).

### 3.4 O Modelo *Long Short-Term Memory* (LSTM)

O modelo LSTM trata-se de um modelo de um modelo de RNN, utilizado para previsão de dados sequenciais e provou ser extremamente eficaz. Ele funciona da seguinte maneira: 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, também chamada de camada sigmoide, 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, algo que as tradicionais RNNs simples não conseguem (Kumar et al., 2019). 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 se irá comportar (Pathinarupothi, et al., 2017). A Figura 2 ilustra graficamente o comportamento de um modelo LSTM.

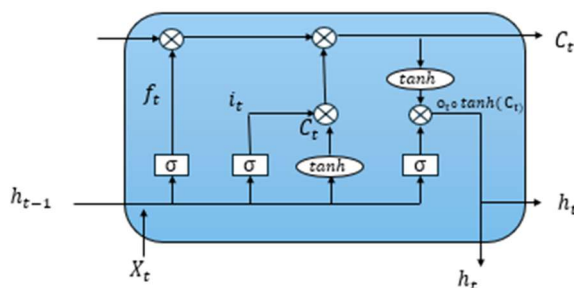


Figura 2 - Diagrama do Modelo LSTM

Fonte: adaptado de Govind e Babu, 2023.

As portas têm as seguintes fórmulas, dadas a série temporal de entrada, indicadores de preços de açúcar ou etanol,  $X_t$  e o número de unidades ocultas como  $h_t$ :

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

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

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

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

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

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

Onde  $\sigma$  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., 2019; Govind e Babu, 2023).

A ideia central dos modelos LSTMs tem como chave o estado da célula, representado pela linha horizontal que passa pela parte superior do diagrama. O estado da célula é semelhante a uma esteira transportadora, percorrendo toda a cadeia, com apenas algumas pequenas interações lineares. Diante disso, o modelo LSTM tem a capacidade de remover ou adicionar informações ao estado da célula, reguladas pelas portas ou portões. O portão de esquecimento ( $f_t$ ), decide quais partes do estado anterior ( $C_{t-1}$ ) devem ser descartadas ou mantidas, usando uma função sigmoide para avaliar a relevância dos dados. O portão de entrada ( $i_t$ ), seleciona quais novas informações, processadas por uma tangente hiperbólica ( $\tanh$ ), devem ser adicionadas ao estado da célula, combinando-as com os dados filtrados pelo portão de esquecimento. Por fim, o portão de saída ( $o_t$ ), controla quais informações do estado da célula ( $C_t$ ) serão usadas na saída atual ( $h_t$ ), aplicando uma sigmoide seguida por ( $\tanh$ ). Esses portões trabalham juntos para gerenciar a memória da célula ao longo do tempo (Hochreiter e Schmidhuber, 1997; Gers et al., 2000).

### 3.5 O Transformer Model

Os modelos Transformer são projetados para usar mecanismos de atenção/autoatenção para capturar padrões e memória de longo prazo nos dados, como dependências entre pontos de dados em uma série temporal ou informações contextuais em um histórico de preços. A maior diferença entre o mecanismo de atenção e os modelos tradicionais de RNN ou LSTM é que o mecanismo de atenção foca diretamente em partes específicas da sequência, em vez de tratá-las igualmente de acordo com a ordem. Isso permite que o modelo capture informações em posições muito iniciais na sequência (Vaswani et al., 2017).

Na Figura 3 apresenta-se a arquitetura de um modelo Tranformer. A Input Embedding se refere aos dados de entrada que neste trabalho são os dados dos indicadores de preços do açúcar e etanol, eles são convertidos em vetores numéricos (embeddings) que representam as características de cada ponto no tempo. O Positional Encoding é usado para uma codificação posicional garantindo que a sequência temporal dos dados seja considerada no modelo. Esses valores são adicionados aos embeddings para incorporar informações sobre a ordem cronológica. No Bloco Transformer, o modelo é dividido em encoder e decoder, mas com foco específico na previsão de valores futuros. O Encoder (lado esquerdo), processa a sequência de dados históricos para extrair padrões e relações contextuais entre os valores em diferentes momentos. Cada bloco ( $N \times$ ) contém: Multi-Head Attention que foca em diferentes partes do histórico, identificando padrões em várias escalas temporais simultaneamente; Add & Norm que normaliza os valores para estabilizar os cálculos e combina os resultados da atenção com o estado anterior; e, o Feed Forward que é a NN que processa cada ponto temporal separadamente, identificando características adicionais.

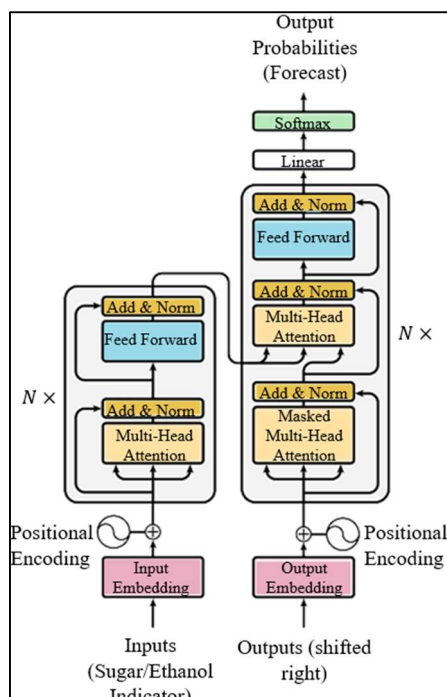


Figura 3 - Arquitetura de um modelo Transformer.

Fonte: Vaswani et al. (2017).

O Decoder (lado direito), utiliza os padrões extraídos pelo encoder para prever os valores futuros e inclui o Masked Multi-Head Attention, que impede que o modelo veja os valores futuros da sequência durante o treinamento, garantindo previsões autoregressivas. As Conexões estabelecidas com os resultados do encoder ajudam o decoder a contextualizar os valores futuros com base no histórico. O Feed Forward e Add & Norm seguem a estrutura do encoder. A Saída (Output Probabilities) é o último bloco do decoder e trata-se de uma previsão contínua, projetada para valores futuros dos indicadores de preços do açúcar e do etanol. Em vez de usar o softmax, o modelo pode usar uma camada linear para produzir diretamente os valores numéricos previstos (preço).

### 3.6 Indicadores de Avaliação dos Resultados

Para calcular o desempenho de predição dos diferentes modelos que serão utilizados cinco critérios o *Mean Absolute Error* (MAE), o *Mean Absolute Percentage Error* (MAPE), a *Root Mean Square Error* (RMSE), o *Mean Percentage Error* (MPE) e o *Mean Error* (ME). As fórmulas desses indicadores são apresentadas a seguir:

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_i - \hat{y}_i| \quad (9)$$

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.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (10)$$

O MAPE mede a magnitude média do erro em um conjunto de previsões e mostra os desvios em porcentagem.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (11)$$

O RMSE mostra uma estimativa dos resíduos entre os valores reais e previstos. O ME calcula a média dos erros de previsão, representando o viés médio nas previsões.

$$ME = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t) \quad (12)$$

Por fim, o MPE mede o erro médio em termos percentuais em relação aos valores reais.

$$MPE = \frac{1}{n} \sum_{t=1}^n \frac{y_t - \hat{y}_t}{y_t} \times 100 \quad (13)$$

Para avaliar a previsão dos modelos são considerados uma previsão de até 6 passos à frente. Assim, os menores valores dessas estatísticas definirão o melhor modelo.

### 3.7 Dados a serem utilizados

Os dados foram extraídos do Centro de Estudos Avançados em Economia Aplicada - Cepea 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*. Além disso, os preços foram ajustados pela inflação utilizando o Índice Geral de Preços – Disponibilidade Interna (IGP-DI). No Quadro 1, encontram-se a descrição das variáveis que foram utilizadas.

Quadro 1 – Variáveis dos modelos

| Variável                                                                     | Fonte             | Período          |
|------------------------------------------------------------------------------|-------------------|------------------|
| Indicador do Açúcar Cristal Branco<br>CEPEA/ESALQ - São Paulo                | Cepea – USP/ESALQ | 04/2012-07/2024  |
| Indicador Mensal do Açúcar CEPEA/ESALQ<br>Alagoas - Mercado Interno          | Cepea – USP/ESALQ | 04/2012-07/2024  |
| Indicador Semanal do Etanol Hidratado<br>Combustível CEPEA/ESALQ - São Paulo | Cepea – USP/ESALQ | 04/2009-07/20024 |
| Indicador Mensal do Etanol Hidratado<br>Combustível CEPEA/ESALQ - Alagoas    | Cepea – USP/ESALQ | 04/2009-07/20024 |

Fonte: elaborado pelos autores, 2024.

O ano de início dos dados analisados foi escolhido devido à disponibilidade das séries que foram utilizadas. As estimativas foram feitas para os estados de São Paulo, e Alagoas. Para a estimação do modelo foram realizados teste de estacionariedade, neste caso utilizou-se o Dickey-Fuller (ADF). Além disso, séries temporais de *commodities* apresentam sazonalidade, que será analisado a partir da análise gráfica da série dos indicadores analisados. A forma mais comum de se corrigir os componentes sazonais de uma série é com a inclusão de variáveis *dummies* ou com a aplicação da primeira diferença (Morettin, 2011; Enders, 2010).

Após a realização dos testes e, no caso do modelo ARIMA e VAR, da definição dos parâmetros, partiu-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; Bergmeir e 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.

Por fim, para a estimação dos modelos LSTM e Trasnformer foi utilizada a biblioteca “tensorflow.keras” do Python (Abadi et al., 2015); já para os modelos ARIMA, VAR e MLP utilizou-se o software R. Além disso, para o desenvolvimento dos modelos preditivos de ML, foram configuradas diferentes arquiteturas de NNs. No modelo MLP, foram utilizadas três camadas ocultas, enquanto no modelo LSTM foram empregadas quatro camadas e para o modelo Transformer foram configuradas cinco camadas ocultas no total, sendo 4 blocos Transformer e 1 camada na MLP.

## 4. Resultados e Discussões

### 4.1 Análise das Séries de Açúcar e Etanol Utilizados

A Figura 4 oferece uma visão geral do comportamento temporal ao longo dos anos das séries do Indicador do Açúcar Cristal Branco CEPEA/ESALQ - São Paulo (a) e do Indicador do Açúcar CEPEA/ESALQ Alagoas - Mercado Interno (b). No ano de 2016, houve um aumento significativo no preço do açúcar no mercado interno, impulsionado principalmente pelo reajuste no preço da gasolina autorizado em setembro do mesmo ano. Esse aumento tornou o etanol mais competitivo, elevando a sua demanda e redirecionando uma maior proporção da cana-de-açúcar para a produção do biocombustível em detrimento do açúcar. Paralelamente, a perspectiva de um déficit na oferta de açúcar para a safra 2015/16 contribuiu para fortalecer o preço da commodity. Além disso, o aumento das exportações, que passaram a oferecer remuneração superior às vendas no mercado doméstico, também exerceu pressão sobre os preços internos (Cepea, 2015).

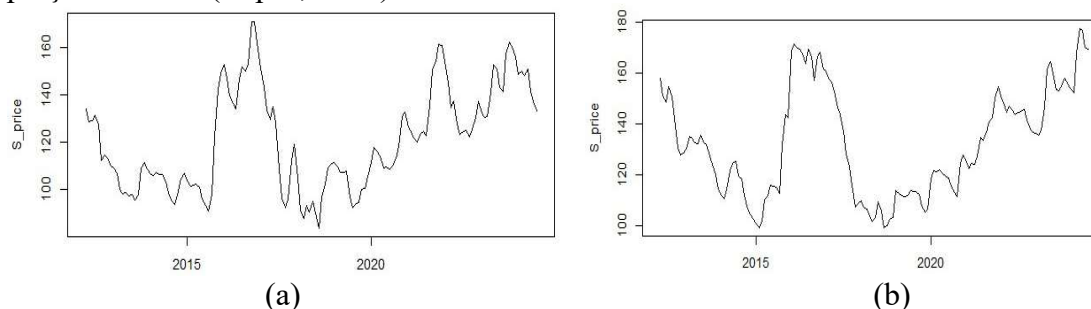


Figura 4 – Série de preços do (a) Indicador do Açúcar Cristal Branco CEPEA/ESALQ - São Paulo e (b) Indicador do Açúcar CEPEA/ESALQ Alagoas - Mercado Interno

Fonte: Cepea, 2024.

Nota: Valores deflacionados pelo IGP-DI

Posteriormente, os preços voltam a reduzir até agosto de 2018, essa redução ocorreu devido a maior oferta e a fraca demanda pelo produto o que pressionou os preços do açúcar nas usinas (CEPEA, 2018). No ano de 2021 ocorreu um novo aumento nos preços, principalmente no Indicador São Paulo, de acordo com a UDOP (2022) os valores foram sustentados pelo baixo estoque de açúcar cristal nas usinas. Nos últimos meses ocorreu um novo aumento dos preços do açúcar, devido ao clima seco no Brasil que contribuiu para o aumento das queimadas, de acordo com a ÚNICA (2024), pelo menos 231,83 mil hectares de lavouras de cana-de-açúcar foram atingidos pelos incêndios registrados no interior de São Paulo, destes, 132,04 mil hectares

são de áreas que ainda seriam colhidas. O restante, ou seja, 99,79 mil hectares, é referente a locais em que a cana-de-açúcar já havia sido colhida ou a lavouras de cana-planta (plantadas nesse ano para colheita no próximo ciclo agrícola). Por fim, é possível observar que os preços do Indicador Alagoas apresentam um preço superior ao indicador São Paulo, em 2024 essa o Indicador Alagoas foi 16% maior que o Indicador São Paulo.

Já na Figura 5 observa-se uma visão geral do comportamento temporal ao longo dos anos das séries do Indicador do Etanol Hidratado Combustível CEPEA/ESALQ - São Paulo (a) e do Indicador do Etanol Hidratado Combustível CEPEA/ESALQ - Alagoas (b). As séries apresentam um comportamento bastante sazonal com preços mais elevados no período entre safras: na região Centro-sul o período de safra é de abril a novembro e na região Norte-nordeste é de setembro a março (Picoli et al., 2014). Em 2011, os preços do etanol registraram um aumento, influenciados por uma menor oferta do produto e um crescimento na demanda. Esse aumento na procura foi impulsionado pelo crescimento da frota de veículos no Brasil, especialmente de carros *flex*, que podem utilizar tanto etanol quanto gasolina (Banco Nacional de Desenvolvimento Econômico e Social – BNDES, 2012). Em 2016, o aumento nos preços foi motivado principalmente pela sua menor oferta. A relação entre os preços médios do etanol hidratado e da gasolina C manteve-se desfavorável ao biocombustível em vários estados brasileiros (UDOP, 2017).

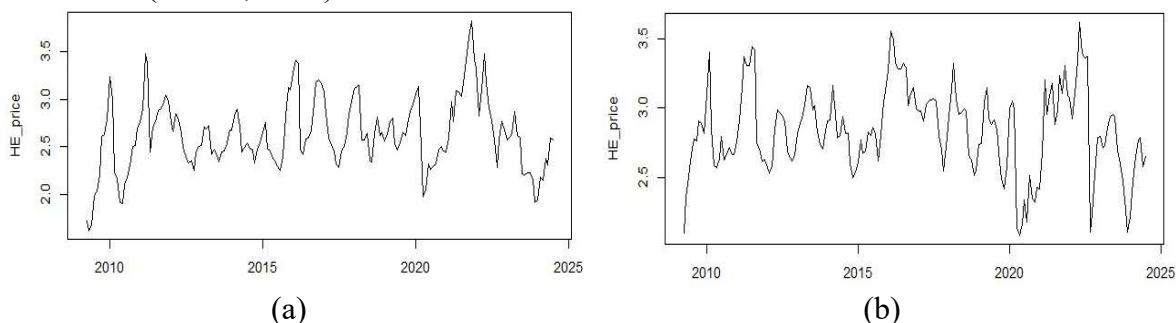


Figura 5 – Série de preços do (a) Indicador do Etanol Hidratado Combustível CEPEA/ESALQ - São Paulo (b) Indicador do Etanol Hidratado Combustível CEPEA/ESALQ - Alagoas

Fonte: Cepea, 2024.

Nota: Valores deflacionados pelo IGP-DI

Em 2020, os preços do etanol sofreram uma queda significativa, atingindo os menores valores históricos do produto. Essa redução foi causada principalmente pela retração no consumo, que apresentou uma diminuição de 16,7% no Brasil nos primeiros seis meses do ano. De janeiro a junho, o volume consumido foi de 8,96 bilhões de litros, em comparação com 10,76 bilhões no mesmo período de 2019. Já 2021/22, os preços do etanol atingiram recordes históricos, impulsionados principalmente pelas valorizações no mercado internacional do petróleo e pela sequência de reajustes positivos nos preços da gasolina nas refinarias brasileiras pela Petrobras. Esses fatores aumentaram a competitividade do etanol como alternativa à gasolina, intensificando a elevação dos preços dos biocombustíveis, com maior impacto observado na região Nordeste (UDOP, 2023). Após uma queda nos preços em 2023, os valores do etanol têm apresentado recuperação, motivadas pela nova Lei do “Combustível do Futuro” que aumenta o percentual da mistura do etanol anidro na gasolina e por fatores climáticos como a estiagem e as altas temperaturas que impactam a produção da cana (UDOP, 2024; Brasil, 2024). Destaca-se que os preços do etanol Norte-nordeste, em 2024, foram 12,12% maiores do que na região Centro-sul.

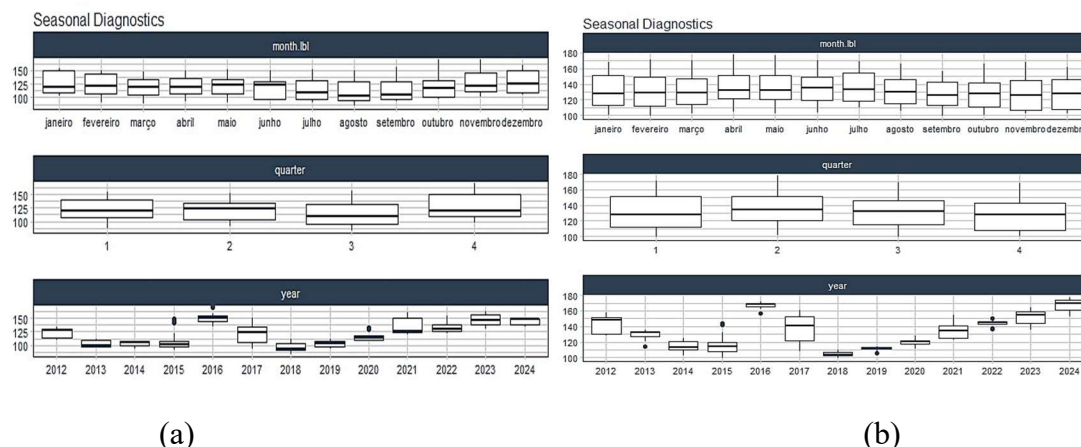


Figura 6 – Análise da sazonalidade Mensal, quadrimestral e anual do (a) Indicador do Açúcar Cristal Branco CEPEA/ESALQ - São Paulo e do (b) Indicador do Açúcar CEPEA/ESALQ Alagoas - Mercado Interno  
Fonte: Cepea, 2024.

Na Figura 6 a análise dos dados do Indicador do Açúcar Cristal Branco CEPEA/ESALQ - São Paulo (a) e do Indicador do Açúcar CEPEA/ESALQ Alagoas - Mercado Interno (b) revela uma sazonalidade clara, influenciada pelos ciclos de colheita da cana. No caso da região Centro-sul, a colheita ocorre de abril a novembro e na região Norte-nordeste o período é de setembro a março, observa-se que no período entre safras, os preços tendem a apresentar menor estabilidade e a se elevarem. Em contraste, os preços tendem a ser mais baixos e menos voláteis nos meses de colheita, quando a oferta é mais abundante. Isso também fica evidente no comportamento trimestral, onde no terceiro e quarto trimestre (final da safra) o Indicador São Paulo apresenta maior variação dos preços; enquanto o Indicador Alagoas apresenta maior variação no primeiro e quarto trimestre, isso também está relacionado com a maior demanda pelo produto ao final do ano. No comportamento anual, é possível observar picos nos anos de 2016, 2021 e 2023.

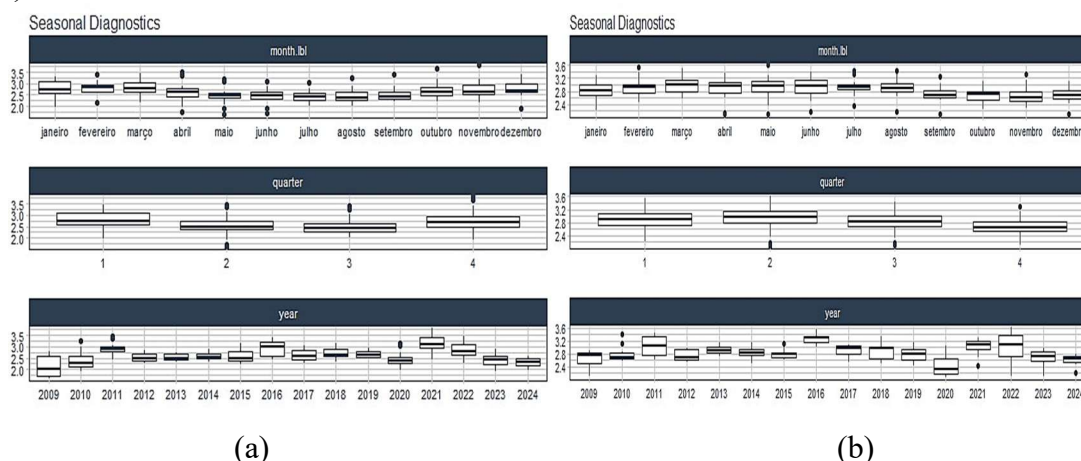


Figura 7 – Análise da sazonalidade Mensal, quadrimestral e anual do (a) Indicador do Etanol Hidratado Combustível CEPEA/ESALQ - São Paulo e (b) Indicador do Etanol Hidratado Combustível CEPEA/ESALQ - Alagoas  
Fonte: Cepea, 2024.

Na Figura 7 a análise dos dados do Indicador do Etanol Hidratado Combustível CEPEA/ESALQ - São Paulo (a) e Indicador do Etanol Hidratado Combustível CEPEA/ESALQ - Alagoas (b) também demonstra um comportamento sazonal, influenciada pelos ciclos de

colheita da cana, com preços mais voláteis no período entre safras e menos voláteis nos meses de colheita. Da mesma forma, no comportamento trimestral, o primeiro e quarto trimestre (período entre safra) o Indicador São Paulo apresenta maior variação dos preços; enquanto o Indicador Alagoas apresenta maior variação no primeiro e segundo trimestre. No comportamento anual, é possível observar picos nos anos de 2011, 2016, 2021 e 2022.

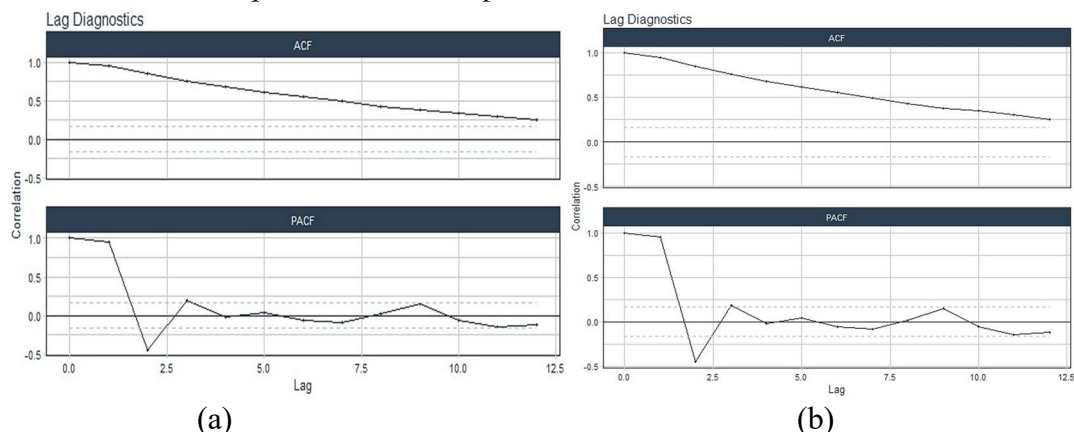


Figura 8 – Análise da Autocorrelation Function (ACF) e da Partial Autocorrelation Function (PACF) do (a) Indicador do Açúcar Cristal Branco CEPEA/ESALQ - São Paulo e do (b) Indicador do Açúcar CEPEA/ESALQ Alagoas - Mercado Interno  
Fonte: elaborado pelos autores, 2024.

Na Figura 8 é possível verificar a análise das funções ACF e PACF do Indicador do Açúcar Cristal Branco CEPEA/ESALQ - São Paulo (a) e do Indicador do Açúcar CEPEA/ESALQ Alagoas - Mercado Interno (b), 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.

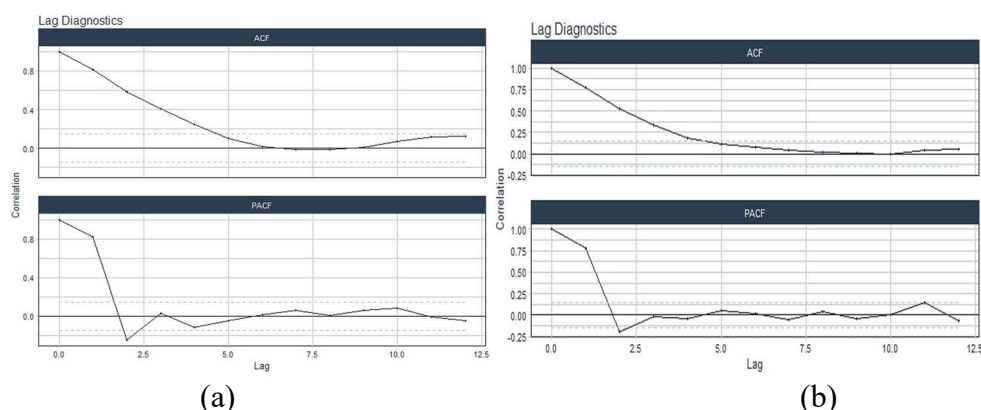


Figura 9 – Análise da Autocorrelation Function (ACF) e da Partial Autocorrelation Function (PACF) do (a) Indicador do Etanol Hidratado Combustível CEPEA/ESALQ - São Paulo e (b) Indicador do Etanol Hidratado Combustível CEPEA/ESALQ - Alagoas  
Fonte: elaborado pelos autores, 2024.

Na Figura 9 é possível verificar a análise das funções ACF e PACF do Indicador do Etanol Hidratado Combustível CEPEA/ESALQ - São Paulo (a) e Indicador do Etanol Hidratado Combustível CEPEA/ESALQ - Alagoas (b), indicando que as séries são estacionárias, já que a ACF não decresce gradualmente em ambos os casos, o que sugere a ausência de tendência ou de um componente de raiz unitária. A PACF, por sua vez, apresenta um pico no “lag” 1 nas duas séries, seguido por uma queda, indicando um processo autoregressivo (AR) de, no mínimo, ordem 1.

Tabela 1 – Testes Dickey-Fuller para Raiz Unitária das Séries Utilizadas

| Variável                                                          | Teste em Nível | Teste em Primeira Diferença |
|-------------------------------------------------------------------|----------------|-----------------------------|
| Indicador do Açúcar Cristal Branco CEPEA/ESALQ - São Paulo        | -1.297         | -8.846                      |
| Indicador do Açúcar CEPEA/ESALQ Alagoas - Mercado Interno         | -1.949         | -7.574                      |
| Indicador do Etanol Hidratado Combustível CEPEA/ESALQ - São Paulo | -4.480         | -11.770                     |
| Indicador do Etanol Hidratado Combustível CEPEA/ESALQ - Alagoas   | -4.933         | -12.327                     |

Fonte: elaborado pelos autores, 2024.

Nota: estatísticas do teste: 1%: -3.483; 5% -2.885; 10%: -2.575.

Os resultados do teste de Dickey-Fuller, Tabela 1, indicaram que os Indicadores de Preço do açúcar não são estacionários em sua forma original, indicando a presença de uma tendência ou raiz unitária. No entanto, após a primeira diferenciação todas as séries tornam-se estacionárias.

#### 4.2 Resultados dos Modelos Estimados

Foram realizadas as estimativas dos modelos para avaliar seu desempenho na previsão das séries temporais dos indicadores de açúcar e etanol utilizando 6 passos à frente. Os resultados dos diferentes modelos podem ser observados nas Tabelas 2 e 3. A análise busca comparar a eficiência de um modelo tradicional, no caso o ARIMA, com modelos de ML, como MLP, LSTM e Transformer, 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 o Indicador do Açúcar Cristal Branco CEPEA/ESALQ - São Paulo, as estatísticas mostraram que o modelo LSTM foi superior, seguido do ARIMA (1,1,1) (1,0,2) que também apresentou um bom desempenho. Já para o Indicador do Açúcar CEPEA/ESALQ Alagoas - Mercado Interno, o modelo MLP e o modelo ARIMA (1,1,1) (0,0,1), demonstraram-se os melhores modelos para previsão.

Tabela 2 - Resultados dos modelos univariados estimados para os Indicadores de açúcar

| Indicador do Açúcar Cristal Branco CEPEA/ESALQ - São Paulo |          |         |         |         |          |
|------------------------------------------------------------|----------|---------|---------|---------|----------|
| Modelo                                                     | ME       | RMSE    | MAE     | MPE (%) | MAPE (%) |
| ARIMA(1,1,1)(1,0,2)                                        | 0.1023   | 5.4681  | 4.1046  | 0.0386  | 3.4321   |
| MLP                                                        | -0.0002  | 5.3386  | 4.1350  | 0.1308  | 3.4938   |
| LSTM                                                       | 0.1092   | 2.5587  | 1.9262  | 0.0803  | 1.3322   |
| Transformer                                                | -11.7772 | 12.0614 | 11.7772 | -8.7021 | 8.7021   |
| Indicador do Açúcar CEPEA/ESALQ Alagoas - Mercado Interno  |          |         |         |         |          |
| ARIMA(1,1,1)(0,0,1)                                        | -0.0054  | 4.8976  | 3.3685  | 0.0351  | 2.5458   |
| MLP                                                        | -0.0005  | 3.3956  | 2.3166  | 0.0675  | 1.8565   |
| LSTM                                                       | -9.1053  | 12.8481 | 12.170  | -5.7301 | 7.6281   |
| Transformer                                                | -9.3711  | 11.3748 | 9.3711  | -5.8110 | 5.8110   |

Fonte: elaborado pelos autores, 2024.

Com relação ao Indicador do Etanol Hidratado Combustível CEPEA/ESALQ - São Paulo, as estatísticas mostraram que os modelos de ML foram superiores ao modelo tradicional ARIMA, o melhor desempenho demonstrado foi do LSTM, seguido do Transformer que também apresentou um bom desempenho. Para o Indicador do Etanol Hidratado Combustível CEPEA/ESALQ – Alagoas o modelo LSTM demonstrou-se superior também, seguido do modelo MLP.

Tabela 3 - Resultados dos modelos univariados estimados para os Indicadores de açúcar

| Indicador do Etanol Hidratado Combustível CEPEA/ESALQ - São Paulo |         |        |        |         |          |
|-------------------------------------------------------------------|---------|--------|--------|---------|----------|
| Modelo                                                            | ME      | RMSE   | MAE    | MPE (%) | MAPE (%) |
| ARIMA(4,1,1)(1,0,1)                                               | 0.02522 | 0.1997 | 0.1457 | 0.53675 | 5.1900   |
| MLP                                                               | -0.0001 | 0.1934 | 0.1393 | 0.24652 | 5.5117   |
| LSTM                                                              | -0.0071 | 0.0720 | 0.0615 | -0.2712 | 2.6008   |
| Transformer                                                       | 0.0823  | 0.1241 | 0.1052 | 3.4531  | 4.4319   |
| Indicador do Etanol Hidratado Combustível CEPEA/ESALQ - Alagoas   |         |        |        |         |          |
| ARIMA(1,1,4)(1,0,1)                                               | 0.0130  | 0.1833 | 0.1285 | 0.0580  | 4.6364   |
| MLP                                                               | -0.0001 | 0.1818 | 0.1210 | 0.0362  | 4.3856   |
| LSTM                                                              | 0.0138  | 0.1069 | 0.0921 | 0.5173  | 3.4609   |
| Transformer                                                       | -0.1660 | 0.1924 | 0.1660 | -6.5923 | 6.5923   |

Fonte: elaborado pelos autores, 2024.

Nota: para a escolha do modelo ARIMA mais adequado, utilizou-se os critérios de AIC, SBC e BIC.

Portanto, os modelos de ML demonstraram, em geral, capacidade de capturar as complexidades das séries temporais, apresentando erros de previsões que indicam uma boa acurácia dos modelos. No entanto, vale destacar que a escolha do modelo ideal depende das características específicas de cada série temporal. Para o açúcar de São Paulo, por exemplo, o LSTM se destacou, enquanto para o açúcar de Alagoas, o MLP e o ARIMA apresentaram melhor desempenho.

## 5. Conclusão

Este estudo buscou desenvolver modelos de previsão de preços de açúcar e etanol dos estados de São Paulo e Alagoas, utilizando os indicadores do Açúcar Cristal Branco e do Etanol Hidratado Combustível CEPEA/ESALQ do estado de São Paulo e de Alagoas Mercado Interno, visando oferecer ferramentas mais precisas para a tomada de decisões no mercado sucroenergético. Foram adotadas modelagens tradicionais, como o ARIMA, e técnicas avançadas de NN, incluindo LSTM e o Transformer, para avaliar qual abordagem oferece maior precisão nas previsões.

Os resultados mostraram que no caso do açúcar em São Paulo, o LSTM mostrou-se mais eficiente, conseguindo modelar ciclos sazonais, tendências de longo prazo e eventos de falha que modelos alternativos, como o ARIMA, não conseguiram modelar com a mesma precisão. Por outro lado, para séries de preços com tendência mais linear, como os preços do açúcar em Alagoas, o modelo MLP e mesmo o ARIMA apresentaram resultados mais eficazes. Para os indicadores de preços do etanol, os resultados das previsões do LSTM e do Transformer também foram muito bons, com erros baixos de previsão e capacidade de prever mudanças no comportamento ao longo dos períodos analisados.

Para o meio acadêmico, estes resultados contribuem ao expandir a compreensão do comportamento dos preços agrícolas e energéticos e a aplicabilidade de abordagens modernas de previsão. Para os fabricantes e a indústria, melhorar a precisão das previsões pode apoiar decisões mais estratégicas, como o planejamento da produção, a gestão de inventários e a negociação de contratos futuros. Para os consumidores, a previsibilidade dos preços pode traduzir-se numa maior transparência e estabilidade do mercado. Outros interessados, como os

decisores públicos e as instituições financeiras, também se beneficiam destes modelos, que fornecem ferramentas poderosas para a análise econômica e mitigação de riscos.

A principal limitação deste estudo é a ausência de variáveis externas relevantes, como políticas de subsídios, clima e câmbio, o que pode reduzir a capacidade preditiva dos modelos. Futuras pesquisas devem considerar a inclusão desses fatores, bem como explorar modelos híbridos que combinem abordagens tradicionais e aprendizado de máquina. Além disso, aplicar esses modelos a outros mercados e produtos pode ampliar seu uso e validar sua eficácia em diferentes contextos.

## Bibliografia

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jozefowicz, R., Jia, Y., Kaiser, L., Kudlur, M., ... Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems*. Software disponível em <https://www.tensorflow.org>.
- Amrouk, E., & Heckelei, T. (2020). Forecasting international sugar prices: A Bayesian model average analysis. *Sugar Tech*.
- Agência Nacional do Petróleo, Gás Natural e Biocombustíveis. (2024). *Etanol*. <https://www.gov.br/anp/pt-br/assuntos/producao-e-fornecimento-de-biocombustiveis/etanol>
- Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP). (2024). *Renovabio: Política Nacional de Biocombustíveis*. <https://www.gov.br/anp/pt-br/assuntos/renovabio>.
- Antunes, F. A., et al. (2019). Biofuel production from sugarcane in Brazil. In *Sugarcane Biofuels: Status, Potential, and Prospects of the Sweet Crop to Fuel the World* (pp. 99–121).
- Baffes, J., & Nagle, P. (Eds.). (2022). *Commodity markets: evolution, challenges, and policies*. World Bank Publications.
- Banco Nacional de Desenvolvimento Econômico e Social (BNDES). (2011). *O déficit de produção de etanol no Brasil*. [https://web.bndes.gov.br/bib/jspui/bitstream/1408/1521/1/A%20set.35\\_O%20d%C3%A9ficit%20de%20produ%C3%A7%C3%A3o%20de%20etanol%20no%20Brasil\\_P.pdf](https://web.bndes.gov.br/bib/jspui/bitstream/1408/1521/1/A%20set.35_O%20d%C3%A9ficit%20de%20produ%C3%A7%C3%A3o%20de%20etanol%20no%20Brasil_P.pdf)
- Barrett, C. B. (2021). Overcoming global food security challenges through science and solidarity. *American Journal of Agricultural Economics*, 103(2), 422–447.
- Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191, 192–213.
- Box, G. E. P., et al. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.
- Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: forecasting and control*. Holden Day, San Francisco.
- Brasil. (2024). *Lei n.º 14.993, de 12 de novembro de 2024*. [https://www.planalto.gov.br/ccivil\\_03/\\_ato2023-2026/2024/lei/L14993.htm](https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2024/lei/L14993.htm)
- Carpio, L. G. T. (2019). The effects of oil price volatility on ethanol, gasoline, and sugar price forecasts. *Energy*, 181, 1012–1022.
- Centro de Estudos Avançados em Economia Aplicada (CEPEA). (2017). *Etanol retro 2017: preço médio da safra 2017/18 fica 8,5% abaixo do da 2016/17*. <https://www.cepea.esalq.usp.br/br/releases/etanol-retro-2017-preco-medio-da-safra-2017-18-fica-8-5-abaixo-do-da-2016-17.aspx>
- Centro de Estudos Avançados em Economia Aplicada (CEPEA). (2024). *Açúcar Cepea: Indicador dispara em outubro e aumento chega a 35%*. <https://www.cepea.esalq.usp.br/br/diarias-de-mercado/acucar-cepea-indicador-dispara-em-outubro-e-aumento-chega-a-35.aspx>
- Centro de Estudos Avançados em Economia Aplicada (CEPEA). (2018). *Açúcar retro 2018: produção no BR cai, mas preços interno e externo não reagem*.

<https://www.cepea.esalq.usp.br/br/releases/acucar-retro-2018-producao-no-br-cai-mas-precos-interno-e-externo-nao-reagem.aspx>

Centro de Estudos Avançados em Economia Aplicada (CEPEA). (2024). *Indicador do açúcar*.

<https://www.cepea.esalq.usp.br/br/indicador/acucar.aspx>

Centro de Estudos Avançados em Economia Aplicada (CEPEA). (2024). *Indicador do etanol*.

<https://www.cepea.esalq.usp.br/br/indicador/etanol.aspx>

Companhia Nacional de Abastecimento (CONAB). (2022). *Acompanhamento da safra brasileira de cana-de-açúcar*. Brasília: CONAB.

Companhia Nacional de Abastecimento (CONAB). (2024). *Produção de cana-de-açúcar*.

Disponível em: <https://www.conab.gov.br/info-agro/safras/cana>

Crone, S. F., & Kourentzes, N. (2010). Feature selection for time series prediction: A combined filter and wrapper approach for neural networks. *Neurocomputing*, 73(10–12), 1923–1936.

Enders, W. (2010). *Applied Econometric Time Series* (3rd ed.). John Wiley & Sons.

Food and Agriculture Organization of the United Nations (FAO). (2024). *FAOSTAT: banco de dados estatísticos*. <https://www.fao.org/faostat/en/#data>

United States Department of Agriculture (USDA). (2024). *Produção de commodities*.

<https://fas.usda.gov/data/production/commodity/0612000>

United States Department of Agriculture (USDA). (2023). *Brazil Sugar Annual 9*.

<https://fas.usda.gov/data/brazil-sugar-annual-9>

Geman, H. (Ed.). (2009). *Risk Management in Commodity Markets: From Shipping to Agriculturals and Energy*. John Wiley & Sons.

Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. *Neural Computation*, 12(10), 2451–2471.

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.

Guilhoto, J., Marjotta-Maistro, M., Mendonça de Barros, A., & Istake, M. (2014). Emprego e mecanização na colheita da cana-de-açúcar: diferenças regionais. *MPRA Paper*, 38325.

Hinton, G. (2015). Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.

Hossain, M. A., et al. (2018). Hybrid deep learning model for stock price prediction. In *IEEE Symposium Series on Computational Intelligence, SSCI* (pp. 1837–1844). Bangalore, Índia.

Instituto Nacional de Pesquisas Espaciais (INPE). (2024). *TerraBrasilis: situação atual das queimadas*. [https://terrabrasilis.dpi.inpe.br/queimadas/situacao-atual/situacao\\_atual/](https://terrabrasilis.dpi.inpe.br/queimadas/situacao-atual/situacao_atual/).

Kourentzes, N., Barrow, D., & Petropoulos, F. (2019). Another look at forecast selection and combination: evidence from forecast pooling. *International Journal of Production Economics*, 209, 226–235. <https://doi.org/10.1016/j.ijpe.2018.02.016>

Kumar, A. G., Sindhu, M. R., & Kumar, S. S. (2019). Deep neural network based hierarchical control of residential microgrid using LSTM. In *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)* (pp. 2129–2134). IEEE.

Kumar, U., & Jain, V. K. (2010). ARIMA forecasting of ambient air pollutants (O<sub>3</sub>, NO, NO<sub>2</sub> and CO). *Stochastic Environmental Research and Risk Assessment*, 24, 751–760.

Ministério da Agricultura, Pecuária e Abastecimento (MAPA). (2024). *Agrostat: indicadores agropecuários*. <https://mapa-indicadores.agricultura.gov.br/publico/extensions/Agrostat/Agrostat.html>

<https://mapa-indicadores.agricultura.gov.br/publico/extensions/Agrostat/Agrostat.html>

Ministério da Indústria, Comércio Exterior e Serviços (MDIC). (2024). *ComexStat: Comércio Exterior*. <https://comexstat.mdic.gov.br/pt/home>

- Ministério de Minas e Energia (MME). (2022). *CNPE passa a ter competência para fixar teor de etanol anidro na gasolina*. <https://www.gov.br/mme/pt-br/assuntos/noticias/cnpe-passa-a-ter-competencia-para-fixar-teor-de-etanol-anidro-na-gasolina>
- Ministério de Minas e Energia (MME). (2023). *PL Combustível do Futuro propõe alteração do limite máximo da mistura de etanol à gasolina de 27% para 30%*. <https://www.gov.br/mme/pt-br/assuntos/noticias/pl-combustivel-do-futuro-propoe-alteracao-do-limite-maximo-da-mistura-de-etanol-a-gasolina-de-27-para-30>
- Monge, M., & Lazcano, A. (2022). Commodity prices after COVID-19: persistence and time trends. *Risks*, 10(6), 128. <https://doi.org/10.3390/risks10060128>.
- Morettin, P. A. (2011). *Econometria financeira: um curso em séries temporais financeiras* (2ª ed.). São Paulo: Editora Blucher.
- Nabipour, M., et al. (2020). Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data: A comparative analysis. *IEEE Access*, 8, 150199–150212.
- Organização das Nações Unidas (ONU). (2024). *Objetivos de Desenvolvimento Sustentável*. <https://brasil.un.org/pt-br/sdgs>
- Palazzi, R. B., Maçaira, P., Meira, E., & Klotzle, M. C. (2023). Forecasting commodity prices in Brazil through hybrid SSA-complex seasonality models. *Production*.
- Pantoja, D. E. L., Samanez, C. P. M., Castro, J. G., & Aiube, F. A. L. (2016). Economic appraisal of production flexibility in different regions of the Brazilian sugar and ethanol industry. *Revista Brasileira de Gestão de Negócios*, 18, 226–244.
- Pathinarupothi, R. K., et al. (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.
- Pereira, N. A. (2017). Variáveis de custos de produção da cana-de-açúcar e suas diferenças entre as regiões produtoras. *Revista em Agronegócio e Meio Ambiente*, 10(3), 757–774.
- Picoli, J. F., Gomes, R. T., & Matsuura, M. I. da S. F. (2014). Sistemas de produção de cana-de-açúcar: avaliação de desempenho ambiental. In *Congresso Interinstitucional de Iniciação Científica*, 8., Campinas. Anais... Campinas: Instituto Agrônomo.
- Sahinli, M. A. (2021). Predicting and analyzing Turkish sugar price with ARCH, GARCH, EGARCH, and ARIMA methods. *Scientific Papers-Series Management Economic Engineering in Agriculture and Rural Development*.
- Silva, N., Siqueira, I., Okida, S., Stevan, S. L., & Siqueira, H. (2019). Neural networks for predicting prices of sugarcane derivatives. *Sugar Tech*, 21, 514–523.
- Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2), 111–133.
- Swapna, G., KP, S., & Vinayakumar, R. (2018). Automated detection of diabetes using CNN and CNN-LSTM network and heart rate signals. *Procedia Computer Science*, 132, 1253–1262.
- Tatarintsev, M., Korchagin, S., Nikitin, P., Gorokhova, R., Bystrenina, I., & Serdechnyy, D. (2021). Analysis of the forecast price as a factor of sustainable development of agriculture. *Agronomy-Basel*.
- Torres-Pruñonosa, J., et al. (2022). How on Earth Did Spanish Banking Sell the Housing Stock? *SAGE Open*, 12(1), 21582440221079916.
- UN Comtrade. (2024). *TradeFlow: Banco de Dados de Comércio Internacional*. <https://comtradeplus.un.org/TradeFlow>
- União da Indústria de Cana-de-Açúcar (UNICA). (2024). *Consumo de etanol cai 16,7% em 2020 no Brasil*. <https://unica.com.br/noticias/consumo-de-etanol-cai-167-em-2020-no-brasil/>.

- União da Indústria de Cana-de-Açúcar (UNICA). (2024). *Safra 2023/2024 termina como a maior da história*. <https://unica.com.br/noticias/safra-2023-2024-termina-como-a-maior-da-historia/>
- União dos Produtores de Bioenergia (UDOP). (2017). EPE analisa a conjuntura dos biocombustíveis e seus impactos em 2016. <https://www.udop.com.br/noticia/2017/06/14/epe-analisa-a-conjuntura-dos-biocombustiveis-e-seus-impactos-em-2016.html>.
- União dos Produtores de Bioenergia (UDOP). (2023). Preço da gasolina sobe mais R\$ 0,23 por litro e já é o maior desde julho de 2022. <https://www.udop.com.br/noticia/2023/08/28/preco-da-gasolina-sobe-mais-r-0-23-por-litro-e-ja-e-o-maior-desde-julho-de-2022.html>
- União dos Produtores de Bioenergia (UDOP). (2024). Produção de cana-de-açúcar pode cair 20% nos próximos 10 anos devido a mudanças climáticas. <https://www.udop.com.br/noticia/2024/09/02/producao-de-cana-de-acucar-pode-cair-20-nos-proximos-10-anos-devido-a-mudancas-climaticas.html>.
- UNICA – União da Indústria de Cana-de-Açúcar. (2022). *Relatório Anual 2022*. São Paulo: UNICA.
- UNICA – União da Indústria de Cana-de-Açúcar. UnicaData. (2024). *Listagem de Safras*. <https://unicadata.com.br/listagem.php?idMn=63>
- Vaswani, A., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.
- World Economic Forum (WEF). (2023). Preços globais do açúcar e o fenômeno El Niño. <https://www.weforum.org/agenda/2023/11/global-sugar-prices-el-nino/>.
- World Bank. (2013). *Prices and Rural Poverty*. Disponível em: <https://documents1.worldbank.org/curated/en/612501468154488885/pdf/666770PUB00PUB0Prices0Rural0Poverty.pdf>. Acesso em: 23 de novembro de 2024.
- Xiong, T., et al. (2015). A combination method for interval forecasting of agricultural commodity futures prices. *Knowledge-Based Systems*, 77, 92–102.
- Xu, X. (2017). Contemporaneous causal orderings of US corn cash prices through directed acyclic graphs. *Empirical Economics*, 52, 731–758.
- Xu, X. (2019). Contemporaneous and Granger causality among US corn cash and futures prices. *European Review of Agricultural Economics*, 46(4), 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*, 13, 200061.
- Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14(1), 35–62.
- Zhang, J. L., Meng, Y. M., Wei, J., Chen, J., & Qin, J. (2021). A novel hybrid deep learning model for sugar price forecasting based on time series decomposition. *Mathematical Problems in Engineering*.
- Zhang, T. T., & Tang, Z. P. (2024). Agricultural commodity futures prices prediction based on a new hybrid forecasting model combining quadratic decomposition technology and LSTM model. *Frontiers in Sustainable Food Systems*.
- Zhang, Y. J., et al. (2019). Volatility forecasting of crude oil market: Can the regime switching GARCH model beat the single-regime GARCH models? *International Review of Economics & Finance*, 59, 302–317.
- Zhao, H. (2021). Futures price prediction of agricultural products based on machine learning. *Neural Computing and Applications*, 33, 837–850.