



# ANÁLISE DE DESEMPENHO DOS MODELOS CNN - STACKED LSTM E SARIMA NA PREVISÃO DO PREÇO DO AÇÚCAR

Dalila Natalia Costa da Luz<sup>1</sup>, José Airton Azevedo dos Santos<sup>2</sup>, Débora Hungaro Micheletti<sup>3</sup>, Aldino Normelio Brun Polo<sup>4</sup>, Laercio Mantovani Frare<sup>5</sup>, Mara Heloisa Neves Olsen Scaliante<sup>6</sup>

<sup>1</sup>Mestranda do Programa de Pós-graduação em Tecnologias computacionais para o Agronegócio (PPGTCA), Campus Medianeira-PR, Universidade Tecnológica Federal do Paraná - UTFPR.

[dalilanataliacostadaluz@gmail.com](mailto:dalilanataliacostadaluz@gmail.com)

<sup>2</sup>Professor do Programa de Pós-graduação em Tecnologias computacionais para o Agronegócio (PPGTCA), Campus Medianeira-PR, Universidade Tecnológica Federal do Paraná - UTFPR.

[airton@utfpr.edu.br](mailto:airton@utfpr.edu.br)

<sup>3</sup>Doutoranda do Programa de Pós-graduação em Engenharia Química (PEQ), Campus Maringá-PR, na Universidade Estadual de Maringá - UEM

[deborahungaromicheletti@gmail.com](mailto:deborahungaromicheletti@gmail.com)

<sup>4</sup>Doutorando do Programa de Pós-graduação em Engenharia Química (PEQ), Campus Maringá-PR, na Universidade Estadual de Maringá - UEM

[aldinonbpolo@gmail.com](mailto:aldinonbpolo@gmail.com)

<sup>5</sup>Professor do Programa de Pós-graduação em Tecnologias computacionais para o Agronegócio (PPGTCA), Campus Medianeira-PR, Universidade Tecnológica Federal do Paraná - UTFPR.

[laercio@utfpr.edu.br](mailto:laercio@utfpr.edu.br)

<sup>6</sup>Professora no Programa de Pós-graduação em Engenharia Química (PEQ), Campus Maringá-PR, na Universidade Estadual de Maringá - UEM

[mhnoscaliente2@uem.br](mailto:mhnoscaliente2@uem.br)

## RESUMO

As ferramentas de previsão de preços de *commodities*, como o açúcar, são fundamentais para auxiliar os diversos agentes do setor agroindustrial em seus processos de tomada de decisão. Assim, este trabalho teve como objetivo comparar os desempenhos, na previsão do preço do açúcar, da rede CNN - *stacked LSTM* (*Convolutional Neural Network – stacked Long Short-Term Memory*) e do modelo SARIMA (*Seasonal Autoregressive Integrated Moving Average*). A base de dados, fornecida pelo *World Bank*, apresenta uma série histórica do preço do açúcar, no período entre ago/2004 e abr/2024, com 237 observações. Os resultados de previsão dos dois modelos foram comparados por meio das métricas: MAE, RMSE e MAPE. Observou-se, dos resultados de previsão para um horizonte de seis meses, que o modelo CNN - *stacked LSTM* apresentou, com relação ao modelo SARIMA, um melhor desempenho (MAE=0,098 R\$/kg, RMSE=0,116 R\$/kg e MAPE=3,938%).

**PALAVRAS-CHAVE:** Python; Redes híbridas; Séries temporais.

## 1 INTRODUÇÃO

O açúcar, quimicamente conhecido como sacarose em sua forma cristalizada, transcende sua função como adoçante para se estabelecer como uma das mais significativas *commodities* agrícolas do mercado global. Extraído majoritariamente da cana-de-açúcar e da beterraba, este insumo é um pilar para a indústria alimentícia mundial e um ativo econômico vital para diversas nações produtoras (NOVACANA, 2024). Este também influencia diretamente o setor de energia por meio da produção de biocombustíveis, como o etanol. Neste contexto, o Brasil é o maior produtor mundial de açúcar, apresentando 25% da produção global. Em 2023, foi registrada a exportação de 31,38 milhões de toneladas da *commodity*, um aumento de 15,1% em relação ao ano anterior (NOVACANA, 2024; ADVFN, 2020).

Um dos principais problemas das *commodities* é a flutuação de preços provocada por uma complexa interação de fatores climatológicos, dinâmicas de mercado e eventos conjunturais, que resultam na incerteza de renda para agricultores, armazenadores, atacadistas, exportadores e agroindústrias. Neste contexto, ferramentas de previsão de preços surgem como instrumentos estratégicos, capacitando os agentes do setor a tomar



decisões mais assertivas sobre o planejamento de safras, o momento ideal de comercialização e o gerenciamento de riscos, visando a maximização dos lucros. Para a cana-de-açúcar, a previsão de preços tem o potencial de auxiliar toda a cadeia produtiva, desde o planejamento das safras pelos produtores, até o processamento da *commodity* pelas agroindústrias responsáveis (Spancerski; Santos, 2021).

A literatura científica tem explorado diversos modelos de séries temporais para abordar os desafios do setor sucroenergético. Estudos como o de Spancerski e Santos (2021) validaram o uso de modelos de *machine learning*, como Redes Neurais MLP (*Multilayer Perception*) e SMOReg (*Sequential Minimal Optimization*), do software WEKA, onde verificaram estimativas confiáveis para previsão a curto prazo do preço do açúcar. Em outra frente, pesquisas se concentraram na previsão de *produção*, como a de Agwingi et al. (2024), que utilizaram os modelos *Local Polynomial Regression* (LPR), *Ridge Regression* (RR) e *Penalized Least Regression* (PLS) e demonstraram que os desempenhos dos três modelos foram comparáveis ao desempenho do tradicional modelo ARIMA (*Autoregressive Moving Average*), superando-o na previsão da produção de açúcar.

Souza et al. (2022) utilizaram do modelo ARIMA para previsão da produção de cana-de-açúcar e subprodutos (açúcar, etanol anidro e etanol hidratado) na macrorregião centro-sul brasileira, tendo resultado de previsão considerados satisfatórios, apesar dos erros médios absolutos em porcentagem (MAPE) serem relativamente altos (13.6%). Enquanto isso, o estudo de Saleen et al. (2023) apresentou a utilização do modelo ARIMA para fazer a previsão da produção anual da cana-de-açúcar no Paquistão, onde puderam observar que existia uma tendência de crescimento da produção de açúcar no país.

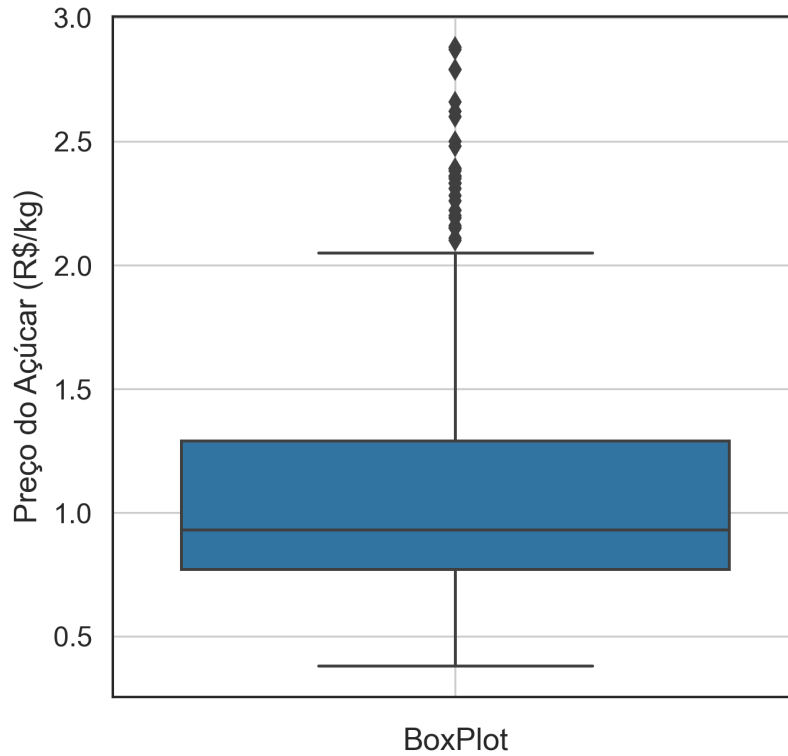
Neste contexto, este trabalho teve como objetivo comparar os desempenhos na previsão do preço do açúcar da rede neural CNN - *stacked* LSTM e do modelo SARIMA com a finalidade de determinar qual das ferramentas oferece maior precisão no gerenciamento de risco e a tomada de decisão estratégica no setor, mitigando incertezas causadas por flutuações de preços devido a fatores climatológicos e mercadológicos.

## 2 MATERIAIS E MÉTODOS

Para a previsão do preço do açúcar, foi realizada a análise do desempenho dos modelos CNN - *stacked* LSTM e do modelo SARIMA a partir da coleta e preparação dos dados, realizando ajustes de valores e tratamento de dados necessárias. Por fim, os modelos foram treinados e validados, e então verificou-se as métricas no conjunto de dados de teste.

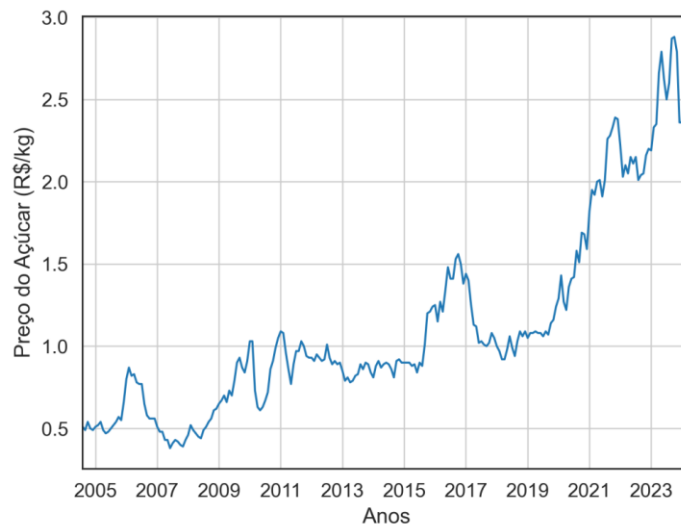
### 2.1 BASE DE DADOS

Foi realizada a previsão do preço do açúcar utilizando uma base de dados com 237 meses (ago/2004 à abr/2024). A base de dados do preço do açúcar (R\$/kg) foi obtida a partir o *World Bank* (INDEXMUNDI, 2024). O Gráfico 1 apresenta o *boxplot* da base de dados, que apresenta alguns *outliers*.



**Gráfico 1:** *Boxplot* dos dados  
**Fonte:** Dados da pesquisa

No Gráfico 2 é apresentada a série histórica do preço do açúcar, onde foi verificado uma tendência ao aumento no preço do açúcar durante o período em estudo.



**Gráfico 2:** Série histórica do preço do açúcar  
**Fonte:** Dados da pesquisa

A partir do Gráfico 3, foram verificados os seguintes períodos:

- 2004 – 2006: Aumento significativo do preço devido à alta do preço do petróleo. Quando o petróleo é caro, o etanol se torna mais atraente. Como o açúcar é uma matéria-prima para o etanol, em países como o Brasil, isso pode aumentar os preços.



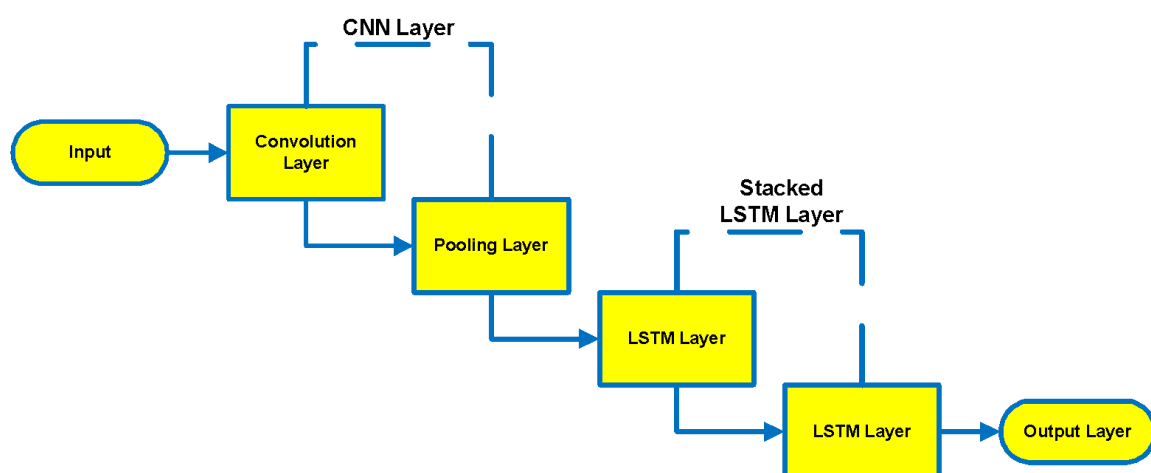
- 2008: Queda durante a crise financeira global, com redução da demanda por *commodities*.
- 2009-2011: Preços do açúcar atingiram 1,09 R\$/kg em 2011. Tendo como motivos as colheitas ruins no Brasil e a importação da Índia, tradicional exportadora, pressionando o mercado.
- 2012-2015: Excedentes e quedas de preço. Devido as boas safras no Brasil, Índia e Tailândia.
- 2016-2019: Volatilidade nos preços. O *El Niño* causou secas na Tailândia em 2016. Crise política e econômica, no Brasil, afetaram investimentos no setor sucroenergético. Tarifas comerciais EUA-China e subsídios indianos, entre 2018 e 2019, geraram incertezas.
- 2020-2023: Pandemia e Choques Geopolíticos. Queda de preços, em 2020, devido à COVID-19. Recuperação rápida dos preços, em 2021, com geadas no Brasil (perdas de 20% na safra) e alta do petróleo. A guerra Rússia-Ucrânia a partir de 2022 afetou os mercados de energia, aumentando a demanda por etanol e, portanto, aumentando os preços do açúcar.

## 2.2 MÉTODO DE FRAGMENTAÇÃO E SELEÇÃO DE HIPERPARÂMETROS

A partir disso, para criar os conjuntos de dados de treinamento e validação, foi utilizado o método *holdout*, onde 70% dos dados foram utilizados para treinamento e 30% para validação. A seleção de hiperparâmetros do modelo CNN - *stacked* LSTM foi realizada por meio da biblioteca Optuna, que busca e encontra a melhor combinação de hiperparâmetros por meio de algoritmos de otimização *bayesiana* (Deng et al., 2024).

## 2.3 REDE CNN-LSTM:

A arquitetura de uma rede CNN - *stacked* LSTM é formada pela combinação das redes neurais CNN e duas camadas LSTM (Gráfico 3). A rede CNN, composta por uma camada convolucional (*Convolution Layer*) e uma camada de pooling (*Pooling Layer*), é utilizada para extrair características dos dados de entrada. A rede neural LSTM, usualmente utilizada para tarefas que possuem entradas sequenciais, usa essas características para fazer previsões (Zha et al, 2022; Santos, 2022; Livieris et al., 2020, Lu et al., 2020).



**Gráfico 3:** Estrutura da CNN - *stacked* LSTM

Fonte: Dados da pesquisa



## 2.4 SARIMA:

O modelo ARIMA, proposto Box-Jenkins na década de 1970, é um modelo muito utilizado para análise de séries temporais. O modelo SARIMA, forma multiplicativa do modelo ARIMA, é utilizado para estudar séries contendo padrões sazonais. O modelo SARIMA é definido pela Equação 1 (Santos; Chaukoski, 2020). Todos os modelos SARIMA foram verificados utilizando o critério de informação de Akaike (AIC).

$$SARIMA((p, d, q)(P, D, Q)_s)$$

Onde:  $(p, d, q)$  e  $(P, D, Q)$  são as partes não sazonal e sazonal do modelo e  $p$  é a ordem do modelo autorregressivo não sazonal,  $q$  é a ordem do modelo de média móvel não sazonal,  $P$  é a ordem do modelo autorregressivo sazonal,  $Q$  é a ordem do modelo de média móvel sazonal,  $d$  é o número de diferenças não sazonais,  $D$  é o número de diferenças sazonais e  $s$  a duração sazonal.

## 2.5 MÉTRICAS

O desempenho dos modelos, implementados neste trabalho, foram avaliados pelas métricas Raiz Quadrada do Erro Médio Quadrático (RMSE), Erro Médio Absoluto (MAE) e Erro Percentual Absoluto Médio (MAPE). Na Tabela 1 são apresentadas as equações das métricas MAPE, RMSE e MAE (Oliveira; Santos, 2024):

**Tabela 1:** Equações MAPE, RMSE e MAE

Sigla	MAPE (%)	RMSE (R\$/kg)	MAE (R\$/kg)
Equação	$\frac{1}{n} \sum_{i=1}^n  (y_i - \hat{y}_i)/y_i  \times 100$	$\sqrt{\frac{1}{n} \sum_{i=1}^n ((y_i - \hat{y})^2)}$	$\frac{1}{n} \sum_{i=1}^n  y_i - \hat{y} $

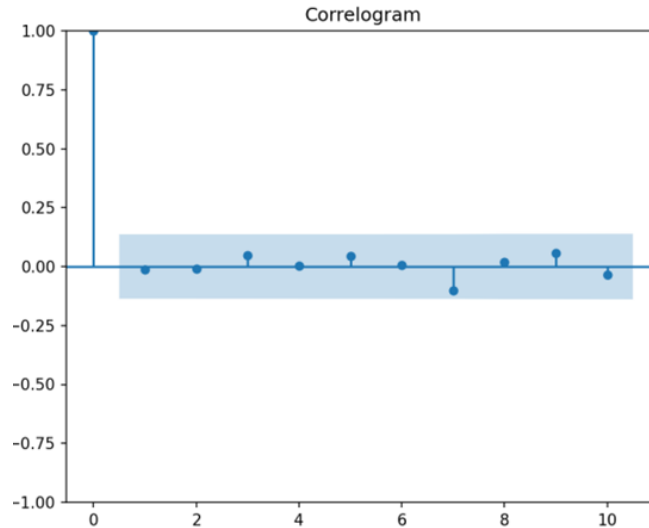
Fonte: Dados da pesquisa

Onde:  $y_i$  é o valor real do período  $i$ ,  $\hat{y}_i$  é a previsão para o período  $i$  e  $n$  é o número de observações. Quanto mais próximo de zero, os valores de MAE, RMSE e MAPE, maior a precisão dos modelos (Lu et al., 2020).

## 3 RESULTADOS E DISCUSSÕES

### 3.1 ANÁLISE DOS RESULTADOS DA SARIMA:

O modelo SARIMA (2,1,0) (3,2,1)<sub>12</sub> foi identificado, de forma automática, por meio da biblioteca pmdarima do Python. Esta identificação foi baseada no teste de AKAIKE (AIC). Verificou-se, para o modelo SARIMA, a normalidade e autocorrelação dos resíduos. Sendo que a normalidade dos resíduos foi verificada por meio do teste de normalidade Jarque-Bera, que obteve um p-valor de 0.26. Observou-se também, por meio do Correlograma (Gráfico 4), que os resíduos não estão autocorrelacionados, devido aos coeficientes de correlação não ultrapassarem os limites de confiança (Santos; Chaukoski, 2022).



**Gráfico 4:** Correlograma dos resíduos  
**Fonte:** Dados da pesquisa

### 3.2 TREINAMENTO E VALIDAÇÃO DOS MODELOS

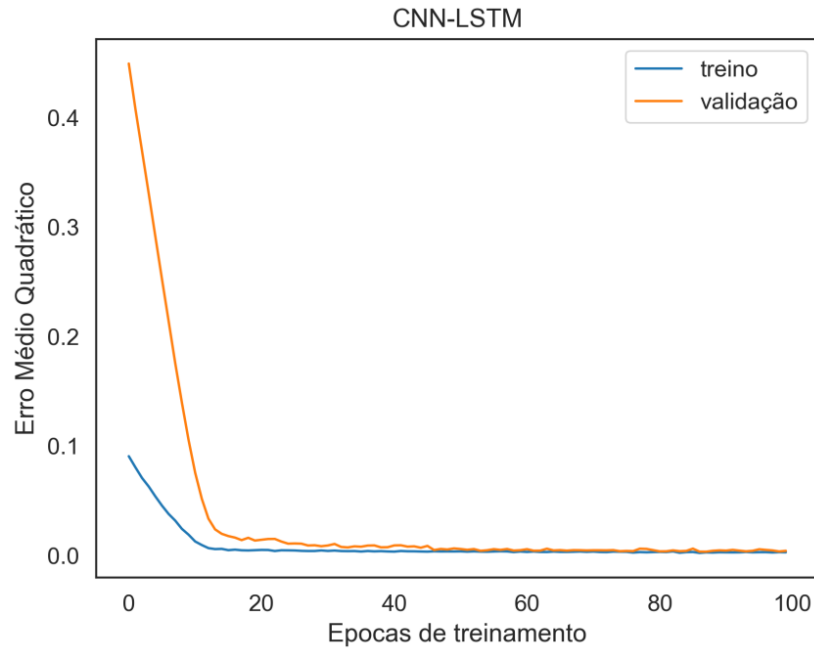
Nesta etapa, com o objetivo de encontrar os melhores modelos de previsão, vários hiperparâmetros foram otimizados por meio do Optuna. Na Tabela 2 apresentam-se os valores dos hiperparâmetros encontrados pelo Optuna.

**Tabela 2:** Parâmetros da rede CNN - *stacked* LSTM

Hiperparâmetro	Valores Encontrados
Epochs	100
<i>Batch</i>	10
<i>LSTM activation function</i>	tanh
<i>Convolution layer kernel size</i>	3
<i>Convolution layer filters</i>	128
<i>Numbers of hidden units in LSTM layer</i>	64/64
<i>Pooling layer pool size</i>	2
<i>Dropout</i>	0.307
<i>Learning rate</i>	$7.3 \times 10^{-5}$
<i>Optimizer</i>	RMSprop

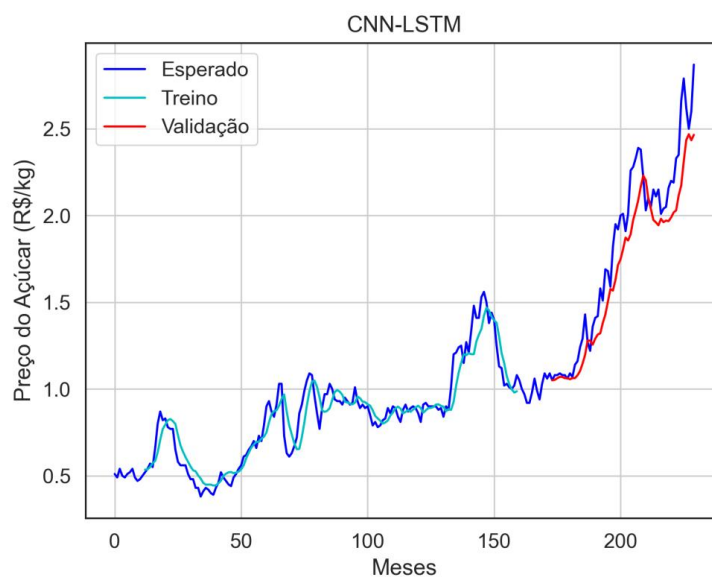
**Fonte:** Dados da pesquisa

A curva de aprendizagem do modelo CNN - *stacked* LSTM, com os hiperparâmetros apresentados na Tabela 2, é apresentada na Gráfico 5. Pode-se observar, por meio deste Gráfico, a boa estabilidade na convergência das curvas de treino e validação.



**Gráfico 5:** Curvas de aprendizagem de treino e validação  
**Fonte:** Dados da pesquisa

No Gráfico 6 apresentam-se os resultados das previsões para os conjuntos de treino e validação.



**Gráfico 6:** Preço do açúcar (Conjuntos de treino e validação)  
**Fonte:** Dados da pesquisa

### 3.3 TESTE DOS MODELOS

Na sequência, realizaram-se previsões, do preço do açúcar, para os meses, de novembro/2023, dezembro/2023, janeiro/2024, fevereiro/2024, março/2024 e abril/2024 (Conjunto de Teste). Estes seis meses não participaram da etapa de seleção dos hiperparâmetros (Tabela 3).

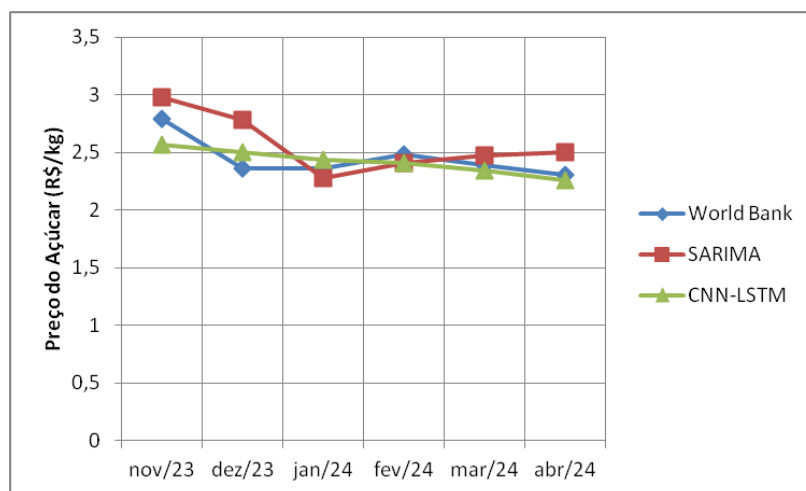


**Tabela 3:** Dados observados (*Word Bank*) e preditos (*SARIMA* e *CNN - stacked LSTM*)

Meses	World Bank	SARIMA	CNN - stacked LSTM
nov/23	2,79	2,98	2,57
dez/23	2,36	2,78	2,5
jan/24	2,36	2,28	2,44
fev/24	2,48	2,41	2,41
mar/24	2,39	2,47	2,34
abr/24	2,31	2,5	2,26

Fonte: Dados da pesquisa

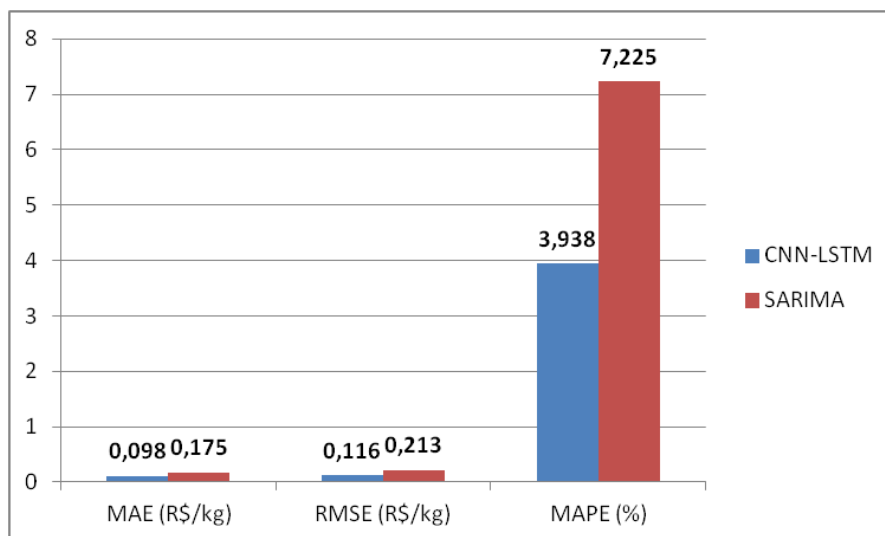
Os resultados das previsões, em termos gráficos, são apresentados no Gráfico 7.



**Gráfico 7:** Resultados das previsões – Conjunto de Teste

Fonte: Dados da pesquisa

Os resultados das métricas MAE, RMSE e MAPE, para o Conjunto de Teste, são apresentados no Gráfico 8.



**Gráfico 8:** Resultados das previsões – Conjunto de Teste

Fonte: Dados da pesquisa



Observa-se, dos resultados das métricas apresentados no Gráfico 8, que o modelo CNN - *stacked* LSTM obteve, em relação ao modelo SARIMA, menores valores de MAE, RMSE e MAPE. Nota-se também, do Gráfico 7, um melhor ajuste da curva CNN - *stacked* LSTM com a curva dos valores observados.

#### 4 CONSIDERAÇÕES FINAIS

Para selecionar os melhores modelos, optou-se por realizar a seleção de hiperparâmetros por meio da biblioteca Optuna. Na sequência, com os melhores modelos, estimaram-se os valores referentes aos meses de novembro de 2023 à abril de 2024 ( 6 meses). Observou-se, por meio dos dados previstos, que os modelos CNN - *stacked* LSTM e SARIMA apresentaram resultados próximos aos obtidos do *Word Bank*. As taxas de erros do modelo de previsão SARIMA foram de 0,175R\$/kg (MAE), 0,213R\$/kg (RMSE) e 7,225% (MAPE), enquanto as taxas de erro do modelo CNN - *stacked* LSTM foram de 0,098R\$/kg (MAE), 0,116R\$/kg (RMSE) e 3,938% (MAPE). Os resultados das métricas demonstram a boa capacidade de generalização, para um horizonte de seis meses, dos modelos implementados neste trabalho. Contudo, o modelo CNN - *stacked* LSTM apresentou um melhor desempenho de previsão do preço do açúcar.

Assim, a comparação do desempenho dos modelos CNN - *stacked* LSTM e SARIMA na previsão do preço da *commodity* açúcar entre o período de ago/2004 e abr/2024, com 237 observações, demonstrou que o modelo CNN - *stacked* LSTM é válido e pode auxiliar, agentes do setor açucareiro, em processos de tomada de decisão.

#### REFERÊNCIAS

ADVFN **Introdução ao açúcar**. Disponível em:

<https://br.advfn.com/investimentos/commodities/acucar/introducao>. 2020. Acesso em: 30 ago. de 2024.

AGWINGI, W. O.; KIHARA, P.; KARANJAH, A. Functional Time Series Models for K-Step Prediction of Sugar Production: A Case Study of Eswatini, Kenya, and South Africa. **Asian Journal of Probability and Statistics**. V. 26, 2024.

DENG, Z.; ZHOU, D.; KANG, Z.; DONG, H. Deep learning-based dynamic forecasting method and application for ultra-deep fractured reservoir production. **Frontiers in Energy Research**, 2024.

IndexMundi Preço do Açúcar. (2020). Disponível em:

<https://www.indexmundi.com/pt/preços-demercado/?mercadoria=açúcar&meses=180&moeda=brl> Acesso em: 30 ago. de 2024.

LIVIERIS, I. E.; PINTELAS, E.; PINTELAS, P. A CNN–LSTM model for gold price time-series forecasting. **Neural Computing and Applications**, 2020.

LU, W.; LI, J.; SUN, A. A CNN-LSTM-based model to forecast stock prices. **Complexity**, 2020.

NOVA CANA. **Brasil exporta volume recorde de açúcar em 2023, com 31,38 milhões de toneladas**. Disponível em: < <https://www.novacana.com/noticias/exportacao-brasileira-acucar-cresce-74-9-volume-dezembro-2022-080124>>. Acesso em: 28 jul. 2024.

OLIVEIRA, L.; SANTOS; J. A. A. Previsão do preço futuro do óleo de girassol:uma



abordagem baseada no modelo CNN-Attention. **Revista Cereus**, 2024.

SALEEM, M. K.; NAZIR, P.; NAZEER, F. Forecasting the Production of Centrifugal Sugar in Pakistan using ARIMA Modelling. **Journal of Statistics**, v. 27, 2023.

SANTOS, J. A. Previsão do preço do café arábica: uma aplicação de redes neurais CNN-BLSTM. **Research, Society and Development**, v. 11, 2022.

SANTOS, J. A. A.; CHAUCOSKI, Y. Previsão de preço de ovos, no estado do Paraná, por meio de modelos SARIMA e PROPHET. **Revista Cereus**, v. 14, 2022.

SANTOS, J. A. A.; CHAUCOSKI, Y. Previsão do Consumo de Energia Elétrica na Região Sudeste: Um Estudo de Caso Usando SARIMA e LSTM. **Revista Cereus**, v. 12, 2020.

SOUZA, C. C. et al. Previsão de produção de cana-de-açúcar e seus derivados, através de ARIMA, no período de 2020 a 2025, na macrorregião Centro-Sul do Brasil. **Research, Society and Development**, v. 11, n. 4, 2022.

SPANCERSKI, J. S.; SANTOS, J. A. A. Previsão do preço do açúcar: um estudo de caso usando o software WEKA. **Revista Engenharia e Tecnologia**, v. 13, n. 1, 2021.

ZHA, W.; LIU, Y.; WAN, Y. Forecasting monthly gas field production based on the CNN-LSTM model. **Energy**, 2022.