

---

## Previsão do preço do milho: uma comparação entre os modelos SARIMA e LSTM

### Maize price forecast: comparison between SARIMA and LSTM models

---

#### **José Airton Azevedo dos Santos**

ORCID: <https://orcid.org/0000-0002-2568-5734>  
Universidade Tecnológica Federal do Paraná, Brasil  
E-mail: [airton@utfpr.edu.br](mailto:airton@utfpr.edu.br)

#### **André Sandmann**

ORCID: [orcid.org/0000-0001-8662-4321](https://orcid.org/0000-0001-8662-4321)  
Universidade Tecnológica Federal do Paraná, Brasil  
E-mail: [sandmann@utfpr.edu.br](mailto:sandmann@utfpr.edu.br)

#### **Vanessa Hlenka**

ORCID: <https://orcid.org/0009-0005-1208-439X>  
Universidade Tecnológica Federal do Paraná, Brasil  
E-mail: [vanessah@utfpr.edu.br](mailto:vanessah@utfpr.edu.br)

#### **Liliane Hellmann**

ORCID: <https://orcid.org/0000-0002-5346-9182>  
Universidade Tecnológica Federal do Paraná, Brasil  
E-mail: [lilianehellmann@utfpr.edu.br](mailto:lilianehellmann@utfpr.edu.br)

#### **Renato Hallal**

ORCID: <https://orcid.org/0000-0003-3039-4242>  
Universidade Tecnológica Federal do Paraná, Brasil  
E-mail: [renatohallal@utfpr.edu.br](mailto:renatohallal@utfpr.edu.br)

#### **Cidmar Ortiz dos Santos**

ORCID: <https://orcid.org/0000-0002-1727-3676>  
Universidade Tecnológica Federal do Paraná, Brasil  
E-mail: [cidmar@utfpr.edu.br](mailto:cidmar@utfpr.edu.br)

---

### RESUMO

O presente trabalho tem como objetivo comparar modelos de séries temporais, na sua forma univariada, na previsão do preço do milho. A base de dados, disponibilizada pela Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), apresenta uma série histórica do preço do milho, no estado do Rio Grande do Sul, no período entre 2011 e 2021. Modelos de previsão, baseados em Redes Neurais LSTM (Long Short-Term Memory) e SARIMA (Seasonal Autoregressive Integrated Moving Average), foram implementados na linguagem Python. Resultados obtidos, dos modelos de previsão, foram comparados por meio da métrica MAPE (Mean Absolute Percent Error). Verificou-se, para um horizonte de seis meses, que o modelo LSTM apresentou melhor desempenho.

**Palavras-chave:** Previsão; Modelo; Python; Séries temporais.

---

## ABSTRACT

The present work aims to compare time series models, in their univariate form, in predicting the price of maize. The database, made available by the Brazilian Agricultural Research Corporation (EMBRAPA), presents a historical series of maize prices, in the state of Rio Grande do Sul, in the period between 2011 and 2021. Forecast models, based on LSTM Neural Networks (Long Short-Term Memory) and SARIMA (Seasonal Autoregressive Integrated Moving Average) were implemented in the Python language. Results obtained from the forecast models were compared using MAPE (Mean Absolute Percent Error) metric. It was verified, for a horizon of six months, that the LSTM model presented better performance.

**Keywords:** Forecast; Model; Python; Time series.

---

## INTRODUÇÃO

O milho, com grande participação econômica no Brasil, é um dos cereais mais cultivados no mundo. Em termos de produção e área plantada, com relação aos demais grãos, fica somente atrás da soja. Pode ser considerado como base para produção de vários produtos. Tendo influência direta, por ser essencial na produção de rações, no preço das carnes, leite e seus derivados. Sua produção mundial se concentra principalmente em três países, Estado Unidos, China e Brasil (SANTOS, 2021; ATLAS, 2021).

O milho é produzido em todo o Brasil e consumido mundialmente. No Rio Grande do Sul o cultivo, em geral, é consorciado com outras culturas, como a da soja, possibilitando obter o produto por mais de uma safra anual. A área cultivada no estado foi de 783,30 mil hectares, com uma produção de 4,11 milhões de toneladas e um valor bruto de R\$ 3,17 bilhões (SARMENTO, 2021).

O milho, sendo uma cultura de alta produtividade, está sujeito a variações no seu preço. Portanto, técnicas de previsão, de séries temporais univariadas, podem ser úteis, a produtores e compradores, no momento de produzir e negociar (SANTOS, 2021).

Dentre as técnicas utilizadas, para previsão de séries temporais, estão as Redes Neurais Artificiais (RNAs) e o modelo ARIMA (Autoregressivo Integrado de Médias Móveis) com a Característica Sazonal da Série (SARIMA). A RNA, segundo Haykin (2001), é um sistema de processamento paralelo, composto por unidades simples, com capacidade de armazenar conhecimento e disponibilizá-lo para o futuro. Já o modelo ARIMA, proposto por Box e Jenkins na década de 1970, é um modelo estatístico muito utilizado em análise de séries temporais (STENGHELE et al., 2020).

Diversos trabalhos utilizaram métodos de previsão, de séries temporais, na análise do preço do milho. Santos (2021) realizou previsões do preço do milho, para o estado do Paraná, por meio de redes neurais BLSTM e MLP. Cas (2017) aplicou um modelo ARIMA na previsão do preço da *commodity* milho. Tibulo e Carli (2010) realizaram a previsão do preço do milho, no estado do Rio Grande do Sul, por meio de modelos de séries temporais ARIMA e Alisamento Exponencial de Holt-Winters. Já Marchezan e Souza (2010) utilizaram modelos SARIMA para previsão do preço do milho no estado do Rio Grande do Sul.

Apesar da importância econômica que o milho exerce, para o estado do Rio Grande do Sul e para o Brasil, muito poucos são os trabalhos que utilizam redes neurais

recorrentes LSTM na previsão do preço do milho. Geralmente, são utilizados modelos tradicionais como os modelos ARIMA.

Neste contexto, este trabalho tem como objetivo comparar modelos, de séries temporais na sua forma univariada, para previsão do preço do milho no estado do Rio Grande do Sul.

## METODOLOGIA

Nesta seção aborda-se a metodologia utilizada para realizar essa pesquisa, detalhando os métodos propostos para realização dos experimentos, bem como as técnicas utilizadas.

## SARIMA

O modelo Autorregressivo Integrado de Médias Móveis (ARIMA) é um dos mais difundidos métodos estatísticos de análise de séries temporais. Foi proposto por Box e Jenkins na década de 1970 e conduziu a uma nova geração de ferramentas de previsão (STENGHELE et al., 2020; WALTER et al., 2013; MONTGOMERY et al., 2008, BOX; JENKINS, 1976).

O modelo ARIMA com a Característica Sazonal da Série é definido como SARIMA(p,d,q)(P,D,Q)<sub>s</sub>, onde (p,d,q) é a parte não sazonal do modelo e (P,D,Q)<sub>s</sub> é a parte sazonal. A definição do modelo é dada por (SANTOS; CHAUCOSKI, 2020; SANTOS; CHAUCOSKI, 2022).

$$\phi(B)\phi(B^s)\Delta^d\Delta_s^D X_t = \theta(B)\theta(B^s)\varepsilon_t \quad 1$$

Onde:

$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  é o operador autorregressivo estacionário de ordem p;

$\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$  é o operador de médias móveis invertível de ordem q;

$\phi(B^s) = 1 - \phi_1 B^s - \dots - \phi_p B^{sp}$  é o operador autorregressivo sazonal de ordem P;

$\theta(B^s) = 1 - \theta_1 B^s - \dots - \theta_q B^{sq}$  é o operador de médias móveis sazonal de ordem Q, invertível;

$\Delta^d = (1 - B)^d$  - d indicando o número de diferenças;

$\Delta_s = (1 - B)^s$  - é o operador diferença sazonal;

$\Delta_s^D = (1 - B^s)^D$  - D indicando o número de diferenças sazonais;

$\varepsilon_t$  – representa o ruído branco com média zero.

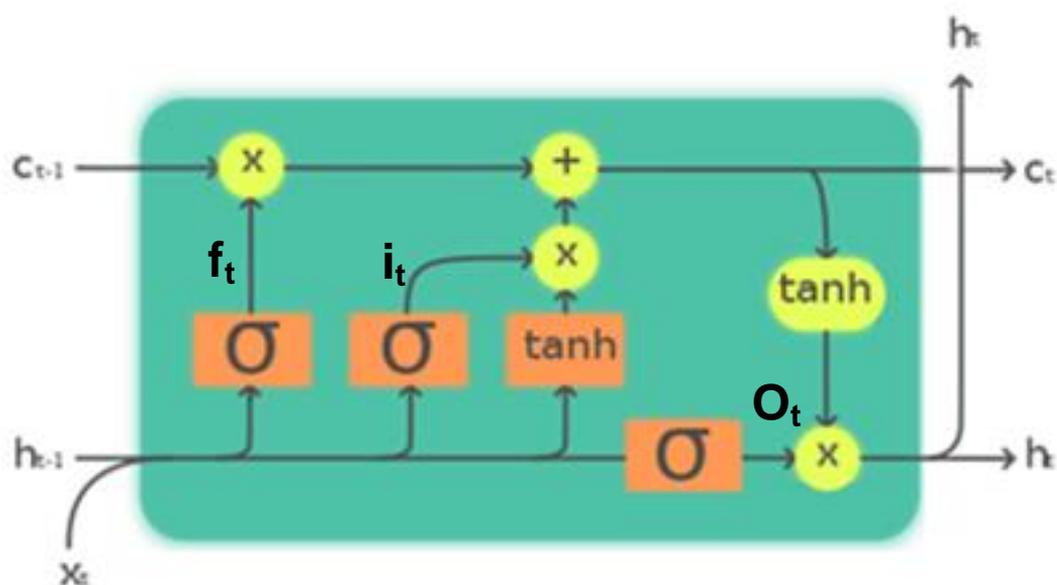
Os parâmetros do modelo, implementado neste trabalho, foram selecionados automaticamente por meio do pacote pmdarima do python, com base no critério de minimização dos AICs (*Akaike Information Criterion*) (CHEN, 2020). Segundo Faruk (2010) depois que um modelo apropriado é escolhido e seus parâmetros estimados, a metodologia Box-Jenkins requer o exame dos resíduos do modelo, para verificar se o modelo é adequado para a série (SANTOS; CHAUCOSKI, 2022).

## Rede LSTM

A rede LSTM é um tipo de Rede Neural Artificial projetada para modelar sequências de longo prazo. Sendo muito utilizada em tarefas que envolvam dados sequenciais, como as séries temporais (SANTOS; CHAUCOSKI, 2020).

Um neurônio LSTM (célula) trabalha com uma sequência de entrada  $x_t$  (Figura 1) e três portas: uma porta de entrada ( $i_t$  - *Update Gate*) que permite adicionar novos elementos a memória; uma porta de esquecimento ( $f_t$  - *Forget Gate*) que permite eliminar elementos da memória e uma porta de saída ( $O_t$  - *Output Gate*) que permite atualizar o estado da célula ( $C_t$ ). O fluxo de dados da rede neural é gerenciado por estas três portas (SPANCERSKI; SANTOS, 2021; GRAVES, 2014, JIA; WANG, 2020).

Figura 1 – Célula LSTM.



Fonte: Adaptado de Chevalier (2022)

Para modelar a rede neural LSTM realizou-se, neste trabalho, três etapas (Figura 2):

1. Coleta e análise de dados: Inicialmente, na primeira etapa, realizou-se a coleta e a análise exploratória dos dados obtidos da EMBRAPA;
2. Treinamento e Validação: Na segunda etapa, treinamento e validação, foi selecionado, por meio da métrica MAPE, o melhor modelo LSTM;
3. Teste: Na última etapa, etapa de teste, testou-se o modelo para dados que não participaram da etapa de treinamento e validação (Janeiro, Fevereiro, Março, Abril, Maio e Junho de 2021).

**Figura 2** – Etapas do trabalho.

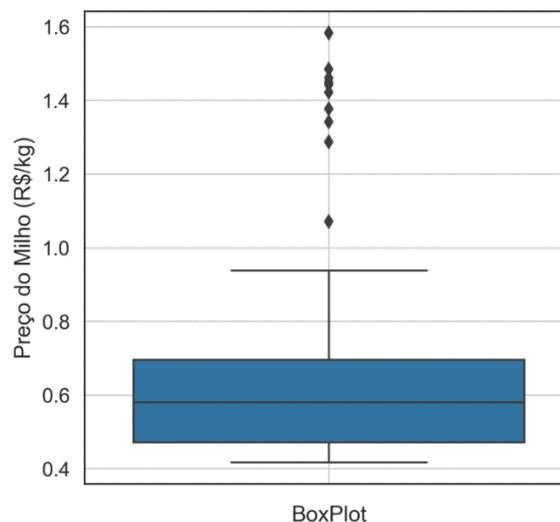


Fonte: Os Autores (2023)

## Base de dados

Para previsão do preço do milho, em R\$/kg, utilizou-se uma base de dados com 114 meses (Jan/2011 - Jun/2021) obtida da Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA). Os preços do milho, para os meses de Janeiro, Fevereiro, Março, Abril, Maio e Junho de 2021, foram retirados do conjunto de dados, para serem utilizados posteriormente para testes dos modelos LSTM e SARIMA. Na Figura 3 apresenta-se o *boxplot* dos dados.

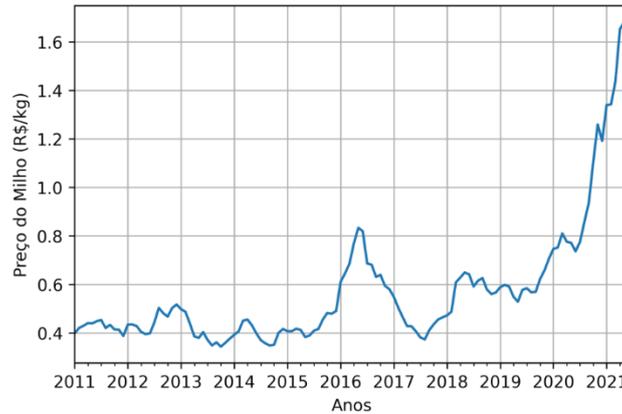
**Figura 3** – *Boxplot* – Preço do milho.



Fonte: Os Autores (2023)

A série histórica, obtida da EMBRAPA, é apresentada na Figura 4.

**Figura 3** – Série: preço do milho (período Jan/2011 - Jun/2021).



Fonte: Os Autores (2023)

Percebe-se, do gráfico apresentado na Figura 3, que os preços da soja mostraram, ao longo do período estudado, uma tendência de elevação dos valores.

### Método da média aritmética

Este método, Equação 2, considera a combinação de duas previsões por média aritmética (MARTINS; WERNER, 2014).

$$MArit = 0,5F_1 + 0,5F_2 \quad 2$$

Onde  $F_1$  e  $F_2$  são as previsões, LSTM e SARIMA, a serem combinadas.

### Métricas

Neste trabalho, os modelos obtidos foram avaliados por meio da métrica MAPE (*Mean Absolute Percent Error*). O MAPE (Equação 3) mede o tamanho do erro em termos percentuais (CANKURT; SUBASI, 2015; SANTOS; CHAUCOSKI, 2020).

$$MAPE = \frac{1}{n} \sum_{i=1}^n |(y_i - \hat{y}_i)/y_i| \times 100 = \frac{1}{n} \sum_{i=1}^n ERP_i \quad 3$$

Onde:  $y_i$  é o valor real do período  $i$ ,  $\hat{y}_i$  é a previsão para o período  $i$ ,  $ERP$  é o erro relativo percentual e  $n$  é o número de observações.

## RESULTADOS E DISCUSSÕES

Inicialmente, neste trabalho, realizou-se uma análise descritiva dos dados (Tabela 1).

**Tabela 1** – Análise descritiva.

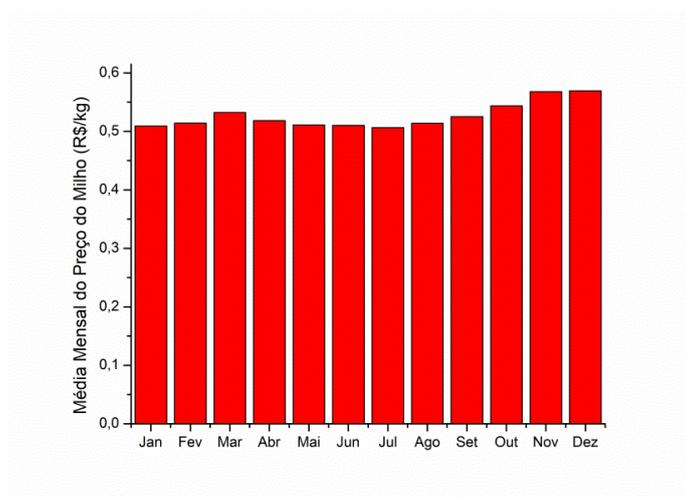
Parâmetro analisado	Resultado
Média (R\$/kg)	0,6521
Mínimo (R\$/kg)	0,417
Máximo (R\$/kg)	1,583
Desvio Padrão (R\$/kg)	0,2628
Coefficiente de Variação (%)	40,3

Fonte: Os Autores (2023)

Pode-se observar, dos dados apresentados na Tabela 1, que o preço ficou, para o período em estudo, em média de 0,6521 R\$/kg. Apresentando, neste período, consumos mínimo e máximo de 0,417 R\$/kg e 1,583 R\$/kg, respectivamente. Observa-se também, da Tabela 3, um alto coeficiente de variação (40,3%), o que indica variabilidade nos dados.

A quantidade média do preço do milho, durante o período estudado, é apresentado na Figura 4. Pode-se notar, por meio desta figura, que os preços médios são maiores para os dois últimos meses do ano. Observou-se também um preço mínimo de 0,5063 (R\$/kg) no mês Julho e um preço médio máximo de 0,5693 (R\$/kg) para o mês de Dezembro.

**Figura 4** – Média Mensal do preço do milho.

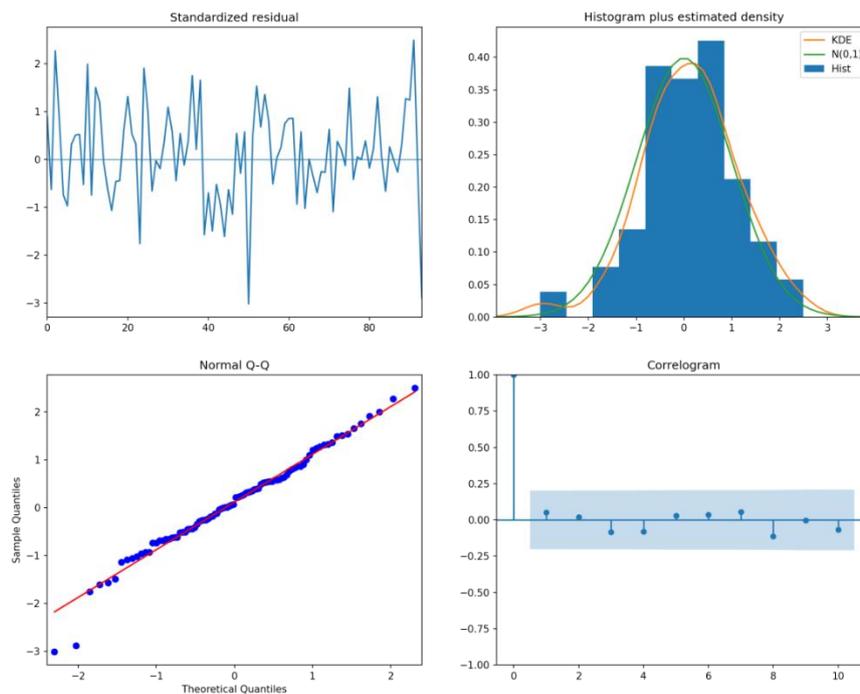


Fonte: Os Autores (2023)

## SARIMA

Inicialmente, por meio do pacote pmdarima do Python, identificou-se que o modelo SARIMA(0,2,1)(0,1,2)<sub>12</sub> obteve, entre os modelos testados, o menor valor do critério de AKAIKE (AIC). Na sequência, verificou-se a normalidade e a autocorrelação dos resíduos. Para o teste de normalidade obteve-se, do *Jarque-Bera normality test*, um p-valor de 0,16, o que indica a não rejeição da hipótese nula de normalidade dos resíduos. Pode-se observar, por meio do correlograma apresentado na Figura 5, que os resíduos não são autocorrelacionados, isto é, não ultrapassam os limites de confiança (SANTOS; CHAUCOSKI, 2020) .

**Figura 5** – Gráficos do modelo Sarima.



Fonte: Os Autores (2023)

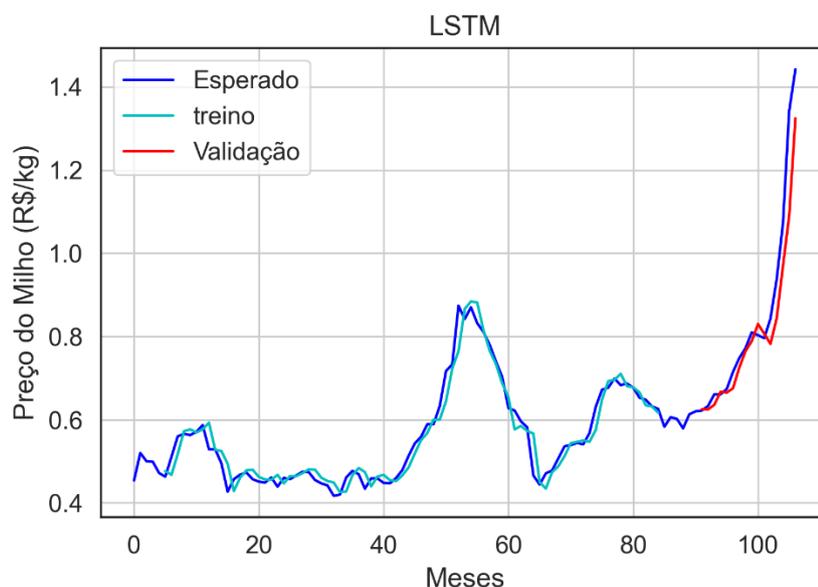
## LSTM

Neste trabalho implementou-se os modelos LSTMs por meio da biblioteca Keras, rodando como *frontend* em TensorFlow. As redes neurais foram treinadas com 86 amostras (80%) e testadas com 22 amostras (20%). O melhor modelo, encontrado

por meio da métrica MAPE, utilizou o algoritmo de otimização *Adam* com os seguintes hiperparâmetros: *look\_back* = 5, *cells* = 8, *batch* = 12, *learning rate* = 0.001 e *activate* = *relu*. O resultado da métrica MAPE, para este modelo, foi MAPE=3,76%.

Pode-se também notar, por meio da Figura 6, a boa aderência dos dados previstos com os dados reais para o modelo LSTM.

**Figura 6** – Rede LSTM – treino e validação.



Fonte: Os Autores (2023)

Na Tabela 2 apresentam-se os dados observados (EMBRAPA), preditos pelos modelos (LSTM e SARIMA), a combinação das previsões por média aritmética (MARit), os Erros Percentuais (ERP) e o Erro Percentual Médio Absoluto (MAPE) para os meses que não participaram da etapa de treinamento e validação (Conjunto de Teste).

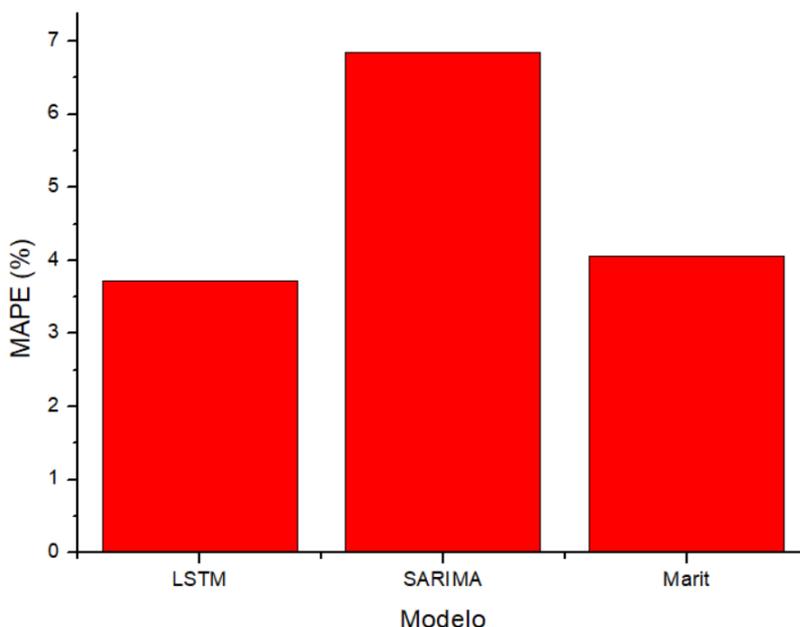
**Tabela 2** – Previsão do preço do milho (R\$/kg), MAPE (%) e ERP (%).

Mês	EMBRAPA	LSTM	SARIMA	Marit	ERP_LSTM	ERP_SARIMA	ERP_Marit
jan/21	1,37	1,35	1,31	1,33	1,46	4,38	2,77
fev/21	1,42	1,37	1,37	1,37	3,52	3,52	3,32
mar/21	1,44	1,4	1,51	1,45	2,78	4,86	1,14
abr/21	1,46	1,43	1,6	1,51	2,05	9,59	4,09
mai/21	1,58	1,5	1,62	1,56	5,06	2,53	1,17
jun/21	1,48	1,59	1,72	1,65	7,43	16,22	11,84
<b>MAPE (%)</b>					<b>3,72</b>	<b>6,85</b>	<b>4,06</b>

Fonte: Os Autores (2023)

As previsões e suas combinações, obtidas no período de teste, possibilitaram, por meio da métrica MAPE, a avaliação da acurácia. Segundo Werner et al. (2020) a medida de acurácia MAPE (Erro Percentual Médio Absoluto) é a mais popular dentre as métricas. Na Figura 7 apresenta-se o MAPE para os dois modelos e para a média aritmética.

**Figura 7** – Erro Percentual Médio Absoluto (MAPE).



Fonte: Os Autores (2023)

Por meio dos resultados apresentados, na Tabela 3 e na Figura 7, conclui-se que os resultados das previsões estão próximos aos fornecidos pela EMBRAPA. Entretanto, o modelo LSTM apresentou, em relação ao modelo SARIMA e o método da média aritmética, maior acurácia (MAPE=3,72%). Observa-se também, da Figura 7, que o método da média aritmética apresentou maior acurácia que o modelo SARIMA.

## CONCLUSÃO

Buscou-se, por meio deste estudo, determinar o comportamento futuro do preço do milho, no estado do Rio Grande do Sul, a partir da dinâmica do passado. Para isto utilizou-se a série histórica do preço do milho, no período entre 2011 e 2021, fornecida pela Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA). Modelos de previsão

LSTM (*Long Short-Term Memory*) e SARIMA (*Seasonal Autoregressive Integrated Moving Average*) foram implementados na linguagem Python.

Inicialmente, neste trabalho, validaram-se os modelos LSTM e SARIMA. Na sequência estimaram-se os valores referentes aos meses de janeiro de 2021 a junho de 2021 (Conjunto de Teste), totalizando seis meses de previsões. Observou-se, por meio dos dados previstos, que os modelos apresentaram resultados próximos aos obtidos da EMBRAPA. Portanto, a proximidade entre valores preditos e reais demonstram a boa capacidade de generalização, para um horizonte de curto prazo, dos modelos LSTM e SARIMA implementados neste trabalho. Contudo, o modelo LSTM apresentou maior acurácia, apresentando um Erro Percentual Médio Absoluto menor que o modelo SARIMA e o método da média aritmética (MAPE-LSTM= 3,72% MAPE-MARit= 4,06% e MAPE-SARIMA= 6,85%).

Conclui-se, para uma previsão de curto prazo, que o modelo LSTM, implementado neste trabalho, pode ser utilizado nos processos de tomada de decisão, de produtores e compradores, nos momentos de produzir e negociar a *commodity* milho.

## REFERÊNCIAS

ATLAS, S. E. **O Rio Grande do Sul é atualmente o sexto maior produtor de milho em grãos do Brasil**. Disponível em: < <https://atlassocioeconomico.rs.gov.br/milho>>. Acesso em: 28 jul. 2021.

BOX, G. E. P.; JENKINS, G. M. **Time series analysis forecasting and control**. San Francisco: Holden-Day, 1976.

CANKURT, S.; SUBASI, A. Comparasion of linear regression and neural network models forecasting tourist arrivals to turkey. **Eurasian Journal of Science & Engineering**, 2015.

CAS, C. G Aplicação do modelo arima para previsão do preço da commodity milho. **GEPROS**, n. 1, p. 263-279, 2018.

CHEN, Y; KANG, Y; WANG, Z. Probabilistic forecasting with temporal convolutional neural network. *Neurocomputing*, v. 399, 2020.

EMBRAPA. Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA). **Milho**. Disponível em: <<https://www.embrapa.br/suinos-e-aves/cias/precos>>. Acesso em: 05 abr. 2021.

FARUK, D. O. A hybrid neural network and ARIMA model for water quality time series prediction. **Engineering Applications of Artificial Intelligence**, V. 23, 2010.

GRAVES, A. Towards end-to-end speech recognition with recurrent neural networks. *Proceedings [...]. 31st International Conference on Machine Learning (ICML-14)*, Beijing, China, 2014.

HAYKIN, S. **Neural networks: a comprehensive foundation**. New Delhi: Pearson Prentice Hall, 2001.

JIA, P.; LIU, H.; WANG, P. Research on a Mine Gas Concentration Forecasting Model Based on a GRU Network. **IEEE Access**, 2020.

MARCHEZAN, A.; SOUZA A. M. Previsão do preço dos principais grãos produzidos no Rio Grande do sul. **Ciência Rural**, v 40, n. 11, 2010.

MARTINS, V. L. M.; WERNER, L. Comparação de previsões individuais e suas combinações: um estudo com séries industriais. **Production**, v. 24, 2014.

MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. **Introduction to time series analysis and forecasting**. New Jersey: John Wiley & Sons, 2008.

SANTOS, J. A. A. Aplicação de redes neurais artificiais na previsão do preço do milho no estado do Paraná. **Revista de Engenharia e Tecnologia**, v. 13, p. 124-134, 2021.

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.

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.; SPANCERSKI. J. S. Previsão da produtividade de arroz: uma aplicação de redes neurais recorrentes. **Revista CEREUS**, v. 13, p. 163-175, 2021.

SARMENTO, S. **Milho: Palmeira e o maio produtor do RS**. Disponível em: <<https://www.folhadonoroeste.com.br/noticias/lista/autor/?autor=5> >. Acesso em: 27 jul. 2021.

SPANCERSKI. J. S.; SANTOS, J. A. A. Aplicação de redes neurais recorrentes na previsão da geração eólica. **Revista CEREUS**, v. 13, p. 217-229, 2021.

STENGHELE, G. ; SANTOS, J. A. A.; MIURA, F. S.; PASA, L. A.; SCHMIDT, C. A. Previsão do consumo de energia elétrica na região sudeste: um estudo de caso usando SARIMA. **Revista FSA**, v. 17, 2020.

TIBULO, C.; CARLI, V. Previsão do preço do milho, através de series temporais. **Scientia Plena**, v. 10, n. 10, 2014.

WALTER, O. M. F. C.; HENNING, E.; MORO, G., SAMOHYL, R. W. Aplicação de um modelo SARIMA na previsão de vendas de motocicletas. **Exacta – EP**, v. 11, 2013.

WERNER, L.; BINOGNIN, C. ARAUJO, C. W. Análise de técnicas de previsão: um estudo de caso para o volume de ações da Petrobras. **Brazilian Journal of Development**, v. 6, n. 1, p.1103-1115, 2020