

PREVISÃO DO PREÇO DA ARROBA DO BOI GORDO: UM ESTUDO DE CASO USANDO ARIMA E GRU

BEEF CATTLE PRICE FORECASTING: A CASE STUDY USING ARIMA AND GRU

José Airton Azevedo dos Santos

RESUMO

Este trabalho teve como objetivo avaliar a eficácia de modelos de séries temporais, ARIMA (*Auto-Regressive Integrated Moving Average*) e GRU (*Gated Recurrent Unit*), na previsão do preço da arroba do boi gordo (US\$/@). A base de dados, disponibilizada pelo Centro de Estudos Avançados em Economia Aplicada (CEPEA), apresenta uma série, de preços mensal, compreendida ente Janeiro/1998 e Maio/2021, totalizando 281 observações. Modelos de previsão, baseados em Redes Neurais GRU e ARIMA, foram implementados, na linguagem Python. Resultados obtidos, dos dois modelos, foram comparados por meio da métrica MAPE (*Mean Absolute Percentage Error*). Verificou-se, para um horizonte de curto prazo, que os dois modelos de previsão fornecem estimativas confiáveis para o preço da arroba do boi gordo, especialmente o modelo GRU que apresentou o menor MAPE.

Palavras-chave: Modelos de previsão. Redes neurais. Séries temporais.

ABSTRACT

This paper aims to evaluate the effectiveness of time series models, ARIMA (*Auto-Regressive Integrated Moving Average*) and GRU (*Gated Recurrent Unit*), in predicting the price of beef cattle arroba (US\$/@). Database from Advanced Studies in Applied Economics Center (ASAEC), presents a series of monthly prices between January/1998 and May/2021, totaling 281 observations. Forecast models, based on GRU and ARIMA Neural Networks, were implemented in Python language. Results obtained from the two models were compared using MAPE metric (*Mean Absolute Percentage Error*). Results showed that, within short-term period, the two forecast models provided reliable estimates for the price of beef cattle arroba, especially the GRU model, which presented the lowest MAPE.

Key Words: Forecast models. Neural networks, Time series.

¹Doutor em Engenharia Elétrica, Programa de Pós-Graduação em Tecnologias Computacionais para o Agronegócio, Universidade Tecnológica Federal do Paraná, Avenida Brasil 4232, CEP 85884-000, Medianeira, Paraná, airton@utfpr.edu.br.

1 INTRODUÇÃO

A bovinocultura nacional ocupa um lugar de destaque, na economia mundial, tanto pelo tamanho do rebanho e da quantidade de abates anuais, quanto pelo volume de exportações de carne bovina. A pecuária, uma das principais responsáveis pela produção de proteína animal, fornece também subprodutos para indústria de calçados, vestuário, entre outros (CORREA et al., 2014). Entretanto, a pecuária está ligada a fatores que apresentam alto risco, tais como: clima, ciclo dos produtos, fragilidade, doenças, flutuações de preços, que podem comprometer a rentabilidade de todo o setor (BRESSAN; LIMA, 2002; BOECHAT, 2013; TAVARES et al., 2020).

A previsão de preços é um instrumento de grande importância, no auxílio ao processo de tomada de decisão, para produtores, compradores e investidores em geral. Especialmente no que se refere ao agronegócio, que é caracterizado pela exposição a distúrbios irregulares (ARAUJO et al., 2012).

Diversas técnicas de previsão, de séries temporais, podem auxiliar agentes envolvidos, no processo de tomada de decisão, em atividades que necessitem de avaliação de políticas, planejamento e redução de incertezas (BRESSAN; LIMA, 2002). Dentre estas, técnicas de previsão, tem-se o modelo ARIMA (Autorregressivo Integrado de Médias Móveis) e as RNAs (Redes Neurais Artificiais). Estes modelos têm sido utilizados, com muito sucesso, em tarefas de predição de séries temporais (CRISTALDO, 2018).

O modelo ARIMA, desenvolvido nos anos 1970 por George Box e Gwilym Jenkins, é muito utilizado em previsões de séries temporais. Este modelo, baseado no ajuste de valores observados, visa reduzir para próximo de zero a diferença entre os valores previstos e observados (SATO, 2013).

As redes neurais artificiais (RNAs), surgidas na década de 80, são sistemas, de processamento paralelo distribuído, inspirados no funcionamento do cérebro humano, capazes de lidar com problemas não-lineares e complexos. As RNAs podem memorizar, analisar e processar um grande número de dados obtidos de um experimento (SEBASTIAN, 2016; ABRAHAM et al., 2019, BASTIANI et al., 2018; HAYKIN, 2001; PINHEIRO et al., 2020).

Diversos trabalhos utilizaram métodos de previsão de preços para o boi gordo. Dentre eles, podem-se citar os trabalhos de: Araujo et al. (2012) que avaliaram a eficácia de modelos, de séries temporais ARIMA e SARIMA, na previsão do preço do boi gordo em Campos dos Goitacazes-RJ. DILLY e FARIA (2018) demonstraram a eficácia, do modelo ARIMA, na previsão do preço do boi gordo no estado do Tocantins. OLIVEIRA (2018) comparou o

desempenho preditivo dos modelos de previsão, de redes neurais e de Suavização Exponencial, em dados do preço da arroba do boi gordo futuro, entre janeiro de 2010 até dezembro de 2015. Já SOARES (2016) analisou a oferta e o preço do boi gordo, no mercado brasileiro, por meio de modelos SARIMA e de Regressão Múltipla.

Embora a carne bovina apresente um importante papel, no aspecto econômico, são poucos os trabalhos que utilizam redes neurais recorrentes na previsão do seu preço, principalmente utilizando a rede neural GRU. Em grande parte, são utilizadas, na previsão do preço do boi gordo, abordagens tradicionais como os modelos de Suavização Exponencial e ARIMA.

Neste contexto, este trabalho teve como objetivo avaliar a eficácia de modelos, de séries temporais, ARIMA (*Auto-Regressive Integrated Moving Average*) e GRU (*Gated Recurrent Unit*) na previsão do preço da arroba do boi gordo (US\$/@), no período entre janeiro de 2021 e maio de 2021.

2 MATERIAL E MÉTODOS

2.1 Arima

Os modelos ARIMA são amplamente utilizados na previsão de variáveis econômicas (CERETTA et al., 2010). Em certos casos, o nome ARIMA e *Box-Jenkins* são utilizados como sinônimos (SATO, 2013). A metodologia *Box-Jenkins* parte do princípio de que modelos de previsão podem ser implementados a partir de informações contidas nos dados. Portanto, o comportamento da variável, em estudo, é que responderá por sua dinâmica futura (FELIPE, 2012; COSTA, 2012)

Modelos ARIMA, são geralmente denotados como ARIMA (p, d, q), onde p denota o número de termos autorregressivos, d é o número de vezes que a série precisa ser diferenciada para torná-la estacionária e q denota o número de termos de médias móveis (GUJARATI; PORTER, 2008). A expressão matemática do modelo ARIMA é dada por (ARÊDES, 2008):

$$\Delta y_t = \emptyset + \sum_{i=1;p} \beta_i \Delta_{t-i} + u_t + \sum_{i=1;q} \alpha_i u_{t-i} \quad 1$$

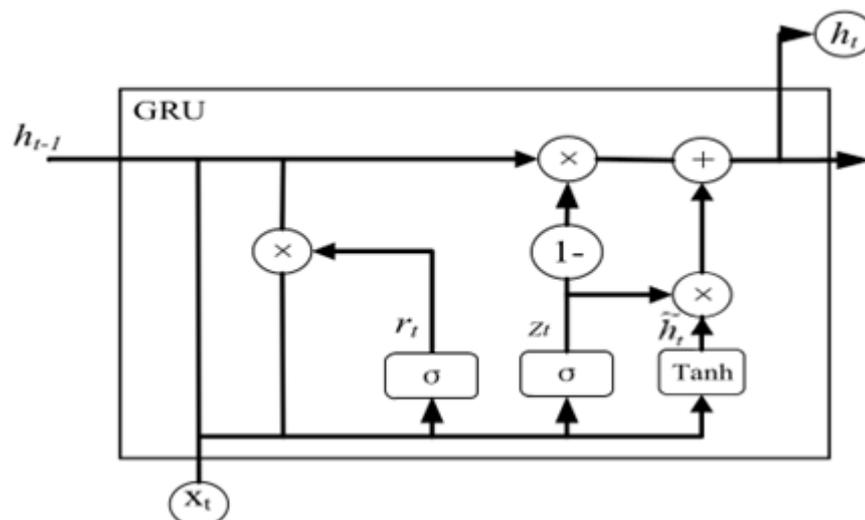
Onde: $\Delta_{(t-i)}$ é o operador de diferenças, u o termo de erro e \emptyset , β e α são parâmetros do modelo.

2.2 GRU

A rede neural recorrente, *Recurrent Neural Network* (RNN), é um tipo, de rede neural artificial, que usa dados sequenciais ou de séries temporais. As RNNs são algoritmos de aprendizagem profunda (*deep learning*) comumente usados em tarefas que envolvem entradas sequenciais, como fala, linguagem, séries temporais, entre outras. A rede GRU (*Gated Recurrent Unit*) é um tipo de RNN capaz de aprender dependências de longo prazo (SANTOS; CHAUCOSKI, 2020).

A topologia de um neurônio, de uma rede GRU, é apresentada na Figura 1. Cada neurônio contém duas portas, uma porta de redefinição z_t (*reset gate*) e um portão de atualização r_t (*update gate*). O transporte das informações de estados ocultos anteriores para o estado atual é controlado por meio da porta de atualização. Quanto maior o valor da porta de atualização, mais informações são preservadas dos estados anteriores. Já a porta de redefinição controla até que ponto as informações dos estados anteriores são ignoradas. Quanto menor o valor desta porta, mais informações são ignoradas. As dependências de curto e longo prazo são associadas a ativação das portas de redefinição e atualização, respectivamente (JIA et al., 2020; SPANCERSKI; SANTOS, 2021).

Figura 1 – Estrutura de um neurônio GRU.

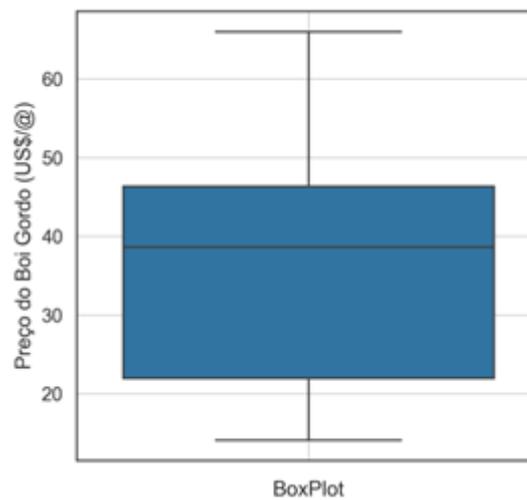


Fonte: JIA et al. (2018).

2.3 Base de dados

Para previsão do preço da arroba do boi gordo (Indicador do boi gordo CEPEA/B3: valor por arroba de 15 kg (US\$/@)) utilizou-se uma base de dados, com 281 meses (Jan/1998 - Mai/2021), obtida do Centro de Estudos Avançados em Economia Aplicada (CEPEA) (CEPEA, 2021). Os dados obtidos, da base de dados, já estavam limpos e sem a presença de *outliers*. Na Figura 2 apresenta-se o *boxplot* dos dados.

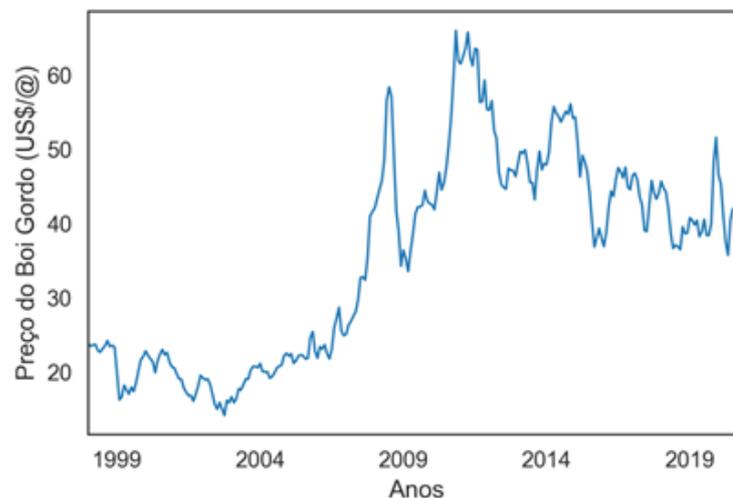
Figura 2 – *Boxplot* dos dados.



Fonte: Autoria própria (2021).

O gráfico da série histórica é apresentado na Figura 3.

Figura 3 – Gráfico ilustrativo da série temporal do preço do boi gordo.



Fonte: Autoria própria (2021).

Pode-se observar, da Figura 3, que a série apresenta, ao longo do período estudado, oscilações no preço da arroba do boi gordo. Segundo Tavares et al. (2020), a alteração no clima é um fator determinante nas oscilações do preço da arroba do boi gordo. Essa alteração impacta na qualidade das pastagens e gera consequências para o setor pecuário.

2.4 Etapas do trabalho

Inicialmente, neste trabalho, realizou-se a análise exploratória dos dados, obtidos do Centro de Estudos Avançados em Economia Aplicada, com a intenção de obter informações importantes sobre os mesmos. Em seguida, realizou-se a etapa de modelagem. Nesta etapa foram desenvolvidos os modelos GRU e ARIMA.

Finalmente, na última etapa, analisou-se a precisão dos modelos GRU e ARIMA, por meio do Erro Percentual Absoluto Médio (MAPE), no período entre janeiro de 2021 e maio de 2021.

2.5 Recursos

Para implementação dos modelos utilizou-se o *Jupyter Notebook*. Os modelos foram implementados por meio da biblioteca Keras. Keras é uma biblioteca para prototipagem rápida e fácil de redes neurais. Roda como *frontend* em *TensorFlow* ou Theano. *TensorFlow* é uma plataforma de código aberto para computação numérica e aprendizado de máquina. Foram também utilizadas as bibliotecas *Scikit-learn*, Pandas e Matplotlib (PINHEIRO et al., 2020; SANTOS; SPANCERSKI, 2021; SANTOS; CHAUCOSKI, 2020).

Um notebook com o sistema Ubuntu 18.04 foi utilizado para executar os modelos. O notebook conta com um processador Intel I7, 2,27GHz e 16GB de memória RAM.

2.6 Métricas

Os modelos implementados, neste trabalho, foram avaliados pelas seguintes métricas (CANKURT; SUBASI, 2015; SANTOS; SPANCERSKI, 2021).

- Coeficiente de Determinação (r^2): A qualidade de ajuste de um modelo pode ser avaliada pelo coeficiente de determinação. Este coeficiente indica quanto o modelo foi capaz de explicar os dados coletados. Quanto mais próximo de um, maior a qualidade dos valores preditos.

- Raiz Quadrada do Erro Médio Quadrático (RMSE): Penaliza mais os erros maiores, permite avaliar a qualidade de um previsor em relação aos dados. Quanto mais próximo de zero, maior a qualidade dos valores preditos.

- Erro Médio Absoluto (MAE): O valor do MAE representa o desvio médio entre o valor observado e o predito. Quanto mais próximo de zero, maior a qualidade dos valores preditos.

Na Tabela 1 apresentam-se as equações das métricas r^2 , RMSE e MAE.

Tabela 1 – Equações: r^2 , RMSE e MAE.

Sigla	r^2	RMSE	MAE
Equação	$1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2}$	$\sqrt{\frac{1}{n} \sum_{i=1}^n ((y_i - \hat{y})^2)}$	$\frac{1}{n} \sum_{i=1}^n y_i - \hat{y} $

Fonte: Cankurt e Subasi (2015).

Onde: y_i é o valor real, \hat{y} é o valor predito, \bar{y} o valor médio e n é o número de observações.

3 RESULTADOS E DISCUSSÃO

Inicialmente, neste trabalho, realizou-se uma análise descritiva dos dados (TABELA 2).

Tabela 2 – Análise descritiva dos dados.

Resumo Descritivo	Valores
Média (US\$/@)	36,07
Mínimo (US\$/@)	14,13
Máximo (US\$/@)	65,96
Desvio Padrão (US\$/@)	14,02
Coefficiente de Variação (%)	38,87

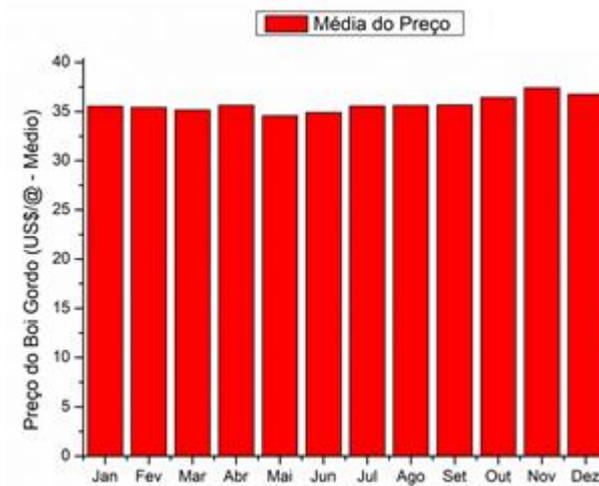
Fonte: Autoria própria (2021).

Pode-se observar, dos dados apresentados na Tabela 2, que o preço, para o período em estudo, ficou em média 36,07 US\$/@. Apresentando, neste período, preços mínimo e máximo de 14,13 US\$/@ e 65,96 US\$/@, respectivamente.

Observa-se também, da Tabela 2, que o coeficiente de variação é 38,87%. Considerado alto, o que indica variabilidade dos dados (PIMENTEL, 2009).

O custo médio mensal do preço da arroba do boi gordo, durante o período estudado, é apresentado na Figura 4. Pode-se notar, por meio desta figura, um custo médio mínimo de 34,57 US\$/@ no mês Maio e um custo médio máximo de 37,96 US\$/@ para o mês de Novembro.

Figura 4 – Preço médio da arroba do boi gordo.

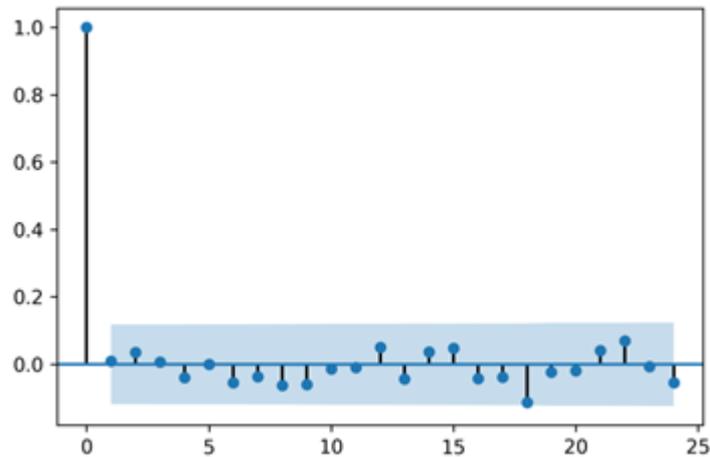


Fonte: Autoria própria (2021).

3.1 Modelos

ARIMA: Inicialmente, identificou-se o modelo ARIMA(1,1,1), que obteve, dos modelos testados, o menor valor do critério de AKAIKE (AIC). Na sequência, verificou-se a normalidade e a autocorrelação dos resíduos (FIGURA 5). Obteve-se, do teste de normalidade (*Jarque-Bera normality test*), um p-valor de 0,08, o que revela a não rejeição da hipótese nula de normalidade dos resíduos. Observa-se, do correlograma, que os resíduos não são autocorrelacionados, pois os coeficientes de autocorrelação são estatisticamente iguais a zero, isto é não ultrapassam os limites de confiança.

Figura 5 – Correlograma dos resíduos



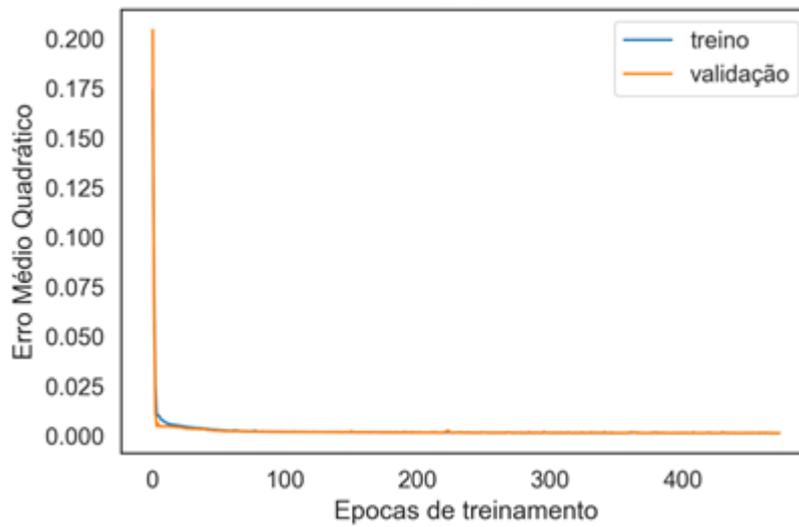
Fonte: Autoria própria (2021).

GRU: Para criar os subconjuntos de dados, de treinamento e validação, foram usadas 276 observações da base de dados do CEPEA. Nota-se que os preços relativos aos meses de Janeiro, Fevereiro, Março, Abril e Maio de 2021 foram retirados do conjunto de dados, para serem utilizados posteriormente para testes dos modelos (conjunto de teste). Neste trabalho utilizou-se o método de fragmentação de *Houldout* onde a base de dados foi dividida com 70% (193) dos dados para treinamento dos algoritmos e 30% (83) para validação. O modelo GRU com melhor desempenho, obtido por meio das métricas, utilizou o algoritmo de otimização Adam com os seguintes hiperparâmetros: GRU *neurons* = 12, *batch* = 10, *learning rate* = 0.001 e *activate* = relu. Os resultados das métricas, para este modelo, foram $r^2=0,77$, RMSE=2,52 e MAE=2,05.

Como critério de parada, para a rede GRU, utilizou-se o método conhecido como Método de Parada Antecipada (*Earling Stopping*). Segundo Silva (2010), este método ajuda a projetar uma rede neural com bom poder de generalização. Neste contexto, definiu-se neste trabalho, como critério de parada do treinamento, a função *EarlyStopping()* com o parâmetro *patience*=33. O parâmetro *patience* indica o número de épocas, após a qual nenhuma melhoria foi observada.

Na Figura 6, apresenta-se a curva de aprendizagem para o preço da arroba do boi gordo. Pode-se observar, por meio da figura, a estabilidade na convergência da curva de treino e validação.

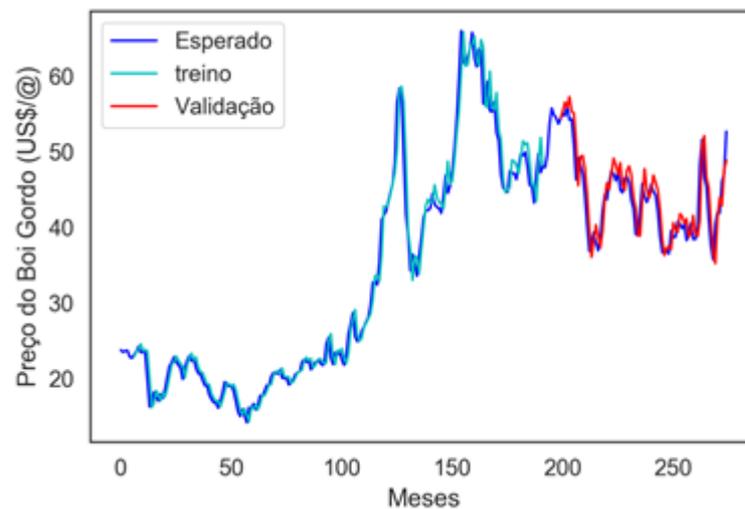
Figura 6 – Curva de aprendizagem – treino e validação (GRU)



Fonte: Autoria própria (2021).

Na Figura 7 apresentam-se as previsões para os conjuntos de treinamento e validação do modelo GRU.

Figura 7 – Previsão de treino e validação – GRU



Fonte: Autoria própria (2021).

Na Tabela 3 apresentam-se os dados, observados e preditos, para os meses de Janeiro, Fevereiro, Março, Abril e Maio de 2021 (Conjunto de teste).

Tabela 3 – Previsões do conjunto de teste (US\$/@).

Mês	CEPEA	ARIMA	GRU
jan/21	53,99	54,93	53,04
fev/21	55,65	54,5	54,04
mar/21	54,85	54,83	55,75
abr/21	56,79	54,95	57,38
mai/21	58,66	55,13	58,32

Fonte: Autoria própria (2021).

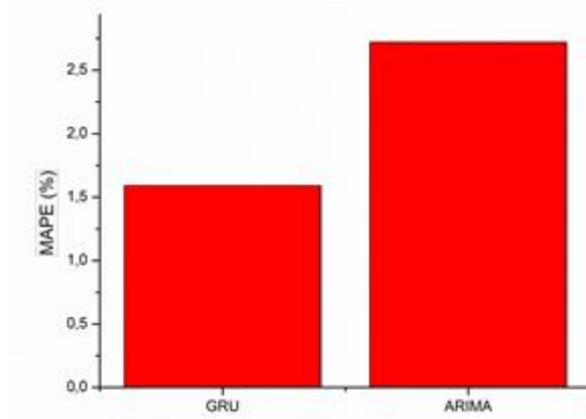
Na sequência, para comparar a precisão das previsões dos modelos GRU e ARIMA, no conjunto de teste, utilizou-se o Erro Percentual Absoluto Médio (MAPE). O MAPE exprime a porcentagem obtida pela divisão da diferença entre os valores real (y_i) e predito (\hat{y}) pelo valor real. Quanto menor, o valor do MAPE, mais preciso é o modelo de previsão (CANKURT; SUBASI, 2015).

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}}{y_i} \right| \quad 2$$

Onde n é o número total de previsões.

Na Figura 8 apresentam-se os resultados do MAPE para os dois modelos.

Figura 8 –MAPE : Conjunto de teste.



Fonte: Autoria própria (2021).

Por meio dos resultados apresentados, na Tabela 3, conclui-se que os resultados das previsões, dos dois modelos, estão próximos aos fornecidos pelo CEPEA. Contudo, como apresentado na Figura 8, o modelo GRU apresenta um erro percentual absoluto médio menor que o modelo ARIMA (MAPE-GRU= 1,59% e MAPE-ARIMA= 2,72%).

4 CONCLUSÕES

Neste trabalho apresentou-se uma aplicação, de modelos de séries temporais, para previsão da arroba do boi gordo. A série de preços da arroba do boi gordo, no período entre 1998 e 2021, foi fornecida pelo Centro de Estudos Avançados em Economia Aplicada (CEPEA). Os modelos, implementados por meio da linguagem Python, passaram pelas fases de: preparação de dados, definição das estruturas, estimativas, avaliação dos resultados e validação.

Observou-se, para os meses de Janeiro, fevereiro, Março, Abril e Maio de 2021, que as previsões foram bem precisas e as diferenças entre valores reais e preditos foram pequenas. Portanto, a proximidade entre valores preditos e reais demonstram a boa capacidade de generalização, para um horizonte de curto prazo, dos modelos implementados neste trabalho. Contudo, o modelo GRU apresentou erro percentual absoluto médio menor que o modelo ARIMA.

Por fim, pode-se observar que ambos os modelos são válidos e podem auxiliar na minimização dos riscos e incertezas do processo de tomada de decisão de agentes que transacionam esta mercadoria.

REFERÊNCIAS

ARAÚJO, S. J.; AREDES, A. F.; SANTOS V. F. Previsão do boi gordo com modelos ARIMA e SARIMA. **Revista Eletrônica de Economia da Universidade Estadual de Goiás**, v. 8, p.27-44, 2012.

ARÊDES, A. F.; PEREIRA, M. W. G. Potencialidade da utilização de modelos de séries temporais na previsão do preço do trigo. **Revista de Economia Agrícola**, v. 55, n. 1, p. 63-76, Jan./jun. 2008.

ABRAHAM, B. **Statistical methods for forecasting**. New York: Wiley & Sons, 2019.

BASTIANI, M.; SANTOS, J. A. A.; SCHMIDT, C. A P.; SEPULVEDA, G. P. L. Application of data mining algorithms in the management of the broiler production. **Geintec**, v. 8, p. 4574-4587, 2018.

BOECHAT, A. M. F. ANÁLISE DO COMPORTAMENTO DOS PREÇOS DO BOI GORDO E DO BOI MAGRO ENTRE 2000 e 2012. **Revista de Economia e Agronegócio**, v. 11, p. 419-437, 2013.

BRESSAN, A. A.; LIMA Modelos de previsão de preços aplicados aos contratos futuros de boi gordo na BM&F. **Revista Nova Economia**, v. 12, p.117-140, 2002.

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

CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA (CEPEA). **Preço do frango CEPEA/ESALQ - Estado SP**. Disponível em:

<<https://www.cepea.esalq.usp.br/br/indicador/frango.aspx>>. Acesso em: 5 jul. 2021.

CERETTA, P. S.; RIGHI, M. B.; SCHELENDER, S. G. Previsão do preço da soja: uma comparação entre os modelos ARIMA e redes neurais artificiais. **Informações Econômica**, v. 40, n. 9, set., p. 15-27, 2010.

CORREA, A. L.; LEÃO, I. A.; SOARES, L. A.; SOUZA, W. A. Avaliação dos Preços do Boi Gordo no Estado de Goiás: Análise da Trajetória de 2008 a 2012. **Revista rama**, v. 7, n.3, p. 613-632, 2014.

COSTA, W. L. S. Tendências do preço da arroba do boi gordo, uso de séries temporais e modelo ARIMA no SPSS. **Cadernos do Leste**, v. 12, n.12, p. 72-81, 2012.

CRISTALDO, M. F. **Aplicação de inteligência artificial à previsão de cheias de pequenas bacias**. 2018. [Tese]. Universidade Anhanguera-UNIDER, Campo Grande, 2018.

DILLY, K. G.; FARIA, V. A. Projeção da bovinocultura no estado do Tocantins utilizando o modelo ARIMA. **Revista Integração Universitária**, 2018.

FELIPE, J. J. Aplicação de modelos ARIMA em séries de soja no norte do estado do Paraná. **Tekhne e Logos**, v. 3, 2012

GUJARATI, D. N.; PORTER, D. C. Basic econometrics. 5 ed. McGraw-Hill New York, 2008.

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.

OLIVEIRA, G. L. **Uma comparação entre modelos de previsão de preços do boi gordo paulista**. 2018. [Dissertação]. Universidade de São Paulo (USP), Campo Grande, 2018.

PINHEIRO, T. C., SANTOS, J. A. A., PASA, L. A. Gestão da produção de frangos de corte por meio de redes neurais artificiais, **Revista Holos**, 2020.

PIMENTEL, F. **Curso de estatística experimental**. Piracicaba: ESALQ, 2009.

SATO, R. C. Gerenciamento de doenças utilizando séries temporais com o modelo ARIMA. **Einstein**, v. 11, n. 1, jan./mar. 2013.

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, n. 4, p.93-104, 2020.

SANTOS, J. A. A.; SPANCERSKI, J. S. Previsão da produtividade de arroz: uma aplicação de redes neurais LSTM. **Revista CEREUS**, v. 13, n. 1, 2021.

SPANCERSKI, J. S.; SANTOS, J. A. A. Aplicação de redes neurais recorrentes na previsão da geração eólica. **Revista CEREUS**, v. 13, n. 1, 2021.

SEBASTIAN, S. Performance evaluation by artificial neural network using WEKA. **International Research Journal of Engineering and Technology**, v. 3, 2016.

SILVA, M. P. **Aplicação de redes neurais artificiais no diagnóstico de falha de turbina a gás**. [Tese]. Rio de Janeiro: Pontifícia Universidade Católica do Rio de Janeiro, 2010.

TAVARES, E. B.; QUINTANILHA, H. T.; RODRIGUES, V. D. V. Análise de cointegração entre preço do boi gordo e taxa de câmbio no período 2000 a 2018. **Research, Society and Development**, v. 9, 2020.