

## Previsão de preço de ovos, no estado do Paraná, por meio de modelos SARIMA e PROPHET

*Eggs price forecast, in the state of Paraná, using SARIMA and PROPHET models*

José Airton Azevedo dos Santos<sup>1</sup>, Yasmin Chaucoski<sup>2</sup>

### RESUMO

O presente trabalho tem como objetivo avaliar a eficácia de modelos de séries temporais, na sua forma univariada, na previsão do preço médio de 30 dúzias de ovos, tipo grande, recebido pelo agricultor no estado do Paraná. A base de dados, disponibilizada pelo Instituto de Pesquisa Econômica Aplicada (IPEA), apresenta uma série histórica do preço, de 30 dúzias de ovos, no período entre 2006 e 2021. Modelos de previsão, baseados nas metodologias PROPHET e SARIMA (*Seasonal Autoregressive Integrated Moving Average*), foram implementados na linguagem Python. Resultados obtidos, dos dois modelos, foram comparados. Verificou-se, para um horizonte de doze meses, que o modelo SARIMA apresentou melhor desempenho, em relação as métricas RMSE e MAPE, que o modelo PROPHET.

**Palavras-chave:** Ovo; Previsão; Séries temporais.

### ABSTRACT

The present work aims to evaluate the effectiveness of time series models, in their univariate form, in forecasting the average price of 30 dozen eggs, large type, received by the farmer in the state of Paraná. The database, made available by the Institute for Applied Economic Research (IPEA), presents a historical series of prices, of 30 dozen eggs, in the period between 2006 and 2021. Forecast models based on PROPHET and SARIMA (*Seasonal Autoregressive Integrated Moving Average*) methodologies were implemented in Python language. Results obtained from the two models were compared. It was verified, for a period of twelve months, that the SARIMA model presented better performance, in relation to the RMSE and MAPE metrics, than the PROPHET model.

**Keywords:** Egg; Forecast; Time series.

<sup>1</sup> Doutor em Engenharia Elétrica. Docente do Programa de Pós-Graduação em Tecnologias Computacionais para o Agronegócio (PPGTCA). Universidade Tecnológica Federal do Paraná (UTFPR). E-mail: [airton@utfpr.edu.br](mailto:airton@utfpr.edu.br). <https://orcid.org/0000-0002-2568-5734>

<sup>2</sup> Discente. Engenharia Elétrica. Universidade Tecnológica Federal do Paraná (UTFPR). E-mail: [chaucoski@alunos.utfpr.edu.br](mailto:chaucoski@alunos.utfpr.edu.br). <http://orcid.org/0000-0002-0155-5919>

## 1. INTRODUÇÃO

A avicultura de postura transformou-se ao longo dos anos, em um empreendimento de grande rentabilidade. Atualmente, o país é um dos maiores produtores de ovos do mundo, produzindo em 2020, devido a pandemia do coronavírus, quatro bilhões de dúzias, uma alta de 3% em relação a 2019 (AVISITE, 2021).

O ovo, uma das mais importantes fontes de proteína de origem animal, é considerado como um alimento de alto valor nutritivo. É também considerado como uma proteína, em comparação com as carnes, de valor mais acessível (OLIVEIRA et al., 2017).

O preço do ovo, no mercado, apresenta grandes flutuações durante o ano, dificultando na organização, programação e comercialização da produção. Portanto, uma estimativa futura, do preço do ovo, pode ser útil, a produtores e compradores, no momento de organizar, produzir e negociar (SANTOS, 2021; OLIVEIRA et al., 2017).

Técnicas de análise, de séries temporais, estão sendo muito utilizadas para realizar análise de mercado e formular planos e estratégias. Dentre estas técnicas tem-se a ferramenta PROPHET e o modelo ARIMA (Autoregressivo Integrado de Médias Móveis) com a Característica Sazonal da Série (SARIMA). O PROPHET, desenvolvido pelo Facebook, implementa modelos, de séries temporais, utilizando ferramentas de previsão avançadas (TAYLOR; LETHAM, 2018). 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 avicultura. Dentre eles, podem-se citar os trabalhos de: Pinheiro et al. (2020) que aplicaram Redes LSTMs na previsão do preço do quilo de frango congelado e resfriado. Bastiani et al. (2018) que aplicaram Redes MLPs para predição e classificação, de variáveis produtivas, das aves produzidas por uma cooperativa localizada na região oeste paranaense. Tápia et al. (2000) utilizou modelos ARIMA e redes neurais artificiais na previsão de preços de ovos. Já Rojas et al. (2019) utilizou a metodologia Box-Jenkins na previsão do preço de ovos brancos no México.

Apesar da importância econômica da avicultura de postura, para o Brasil, muito poucos são os trabalhos que utilizam técnicas de previsão de séries temporais na previsão do preço de ovos.

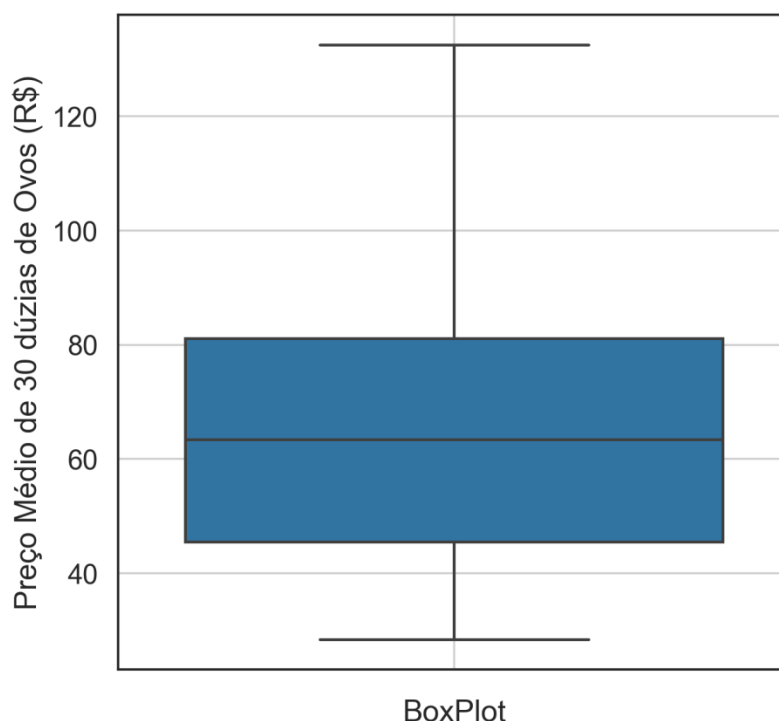
Neste contexto, este trabalho tem como objetivo avaliar a eficácia de modelos de séries temporais, na sua forma univariada, na previsão do preço médio de 30 dúzias de ovos, tipo grande, recebido pelo agricultor no estado do Paraná.

## 2. MATERIAIS E MÉTODOS

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

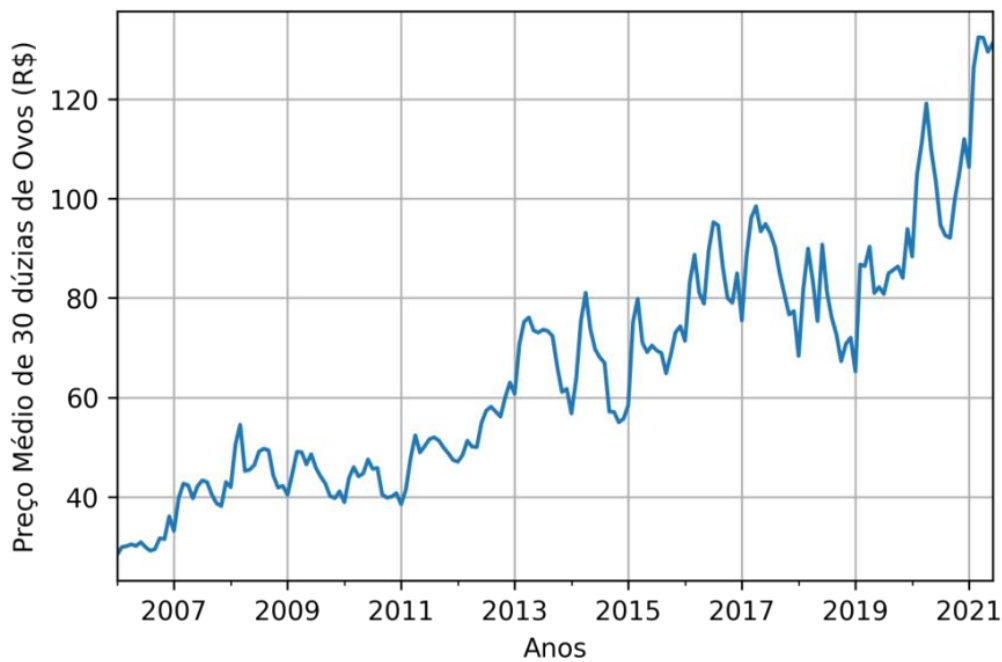
Base de Dados:

Para previsão do preço médio de 30 dúzias de ovos, tipo grande, recebido pelo agricultor, no estado do Paraná, utilizou-se uma base de dados com 186 meses (Jan/2006 - Jun/2021), fornecida pelo Instituto de Pesquisa Econômica Aplicada (IPEA) (IPEA, 2021). Os dados obtidos, da base de dados, já estavam limpos e sem a presença de *outliers* (Figura 1).



**Figura 1.** *Boxplot* dos dados  
Fonte: IPEA (2021).

A série histórica, fornecida pelo IPEA, é apresentada na Figura 2.



**Figura 2.** Série: preço de 30 dúzias de ovos (período Jan/2006 - Jun/2021)  
**Fonte:** IPEA (2021).

## PROPHET:

A ferramenta PROPHET, desenvolvida pelo Facebook, é uma ferramenta de previsão, de séries temporais, disponível em Python. PROPHET é uma ferramenta, de código aberto, que usa parâmetros intuitivos que são fáceis de ajustar. Utiliza modelos de séries temporais decompostos em três componentes principais: tendência, sazonalidade e feriados. O seu modelo de previsão é composto por (TAYLOR; LETHAM, 2018; PINTO, 2020; XIE et al., 2021):

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \quad 1$$

$g(t)$  – representa tendência dos modelos, se preocupa com as mudanças não periódicas nos valores;

$s(t)$  – representa a sazonalidade, descreve como os dados são afetados por fatores sazonais, como a época do ano;

$h(t)$  – representa os feriados, modela os efeitos de feriados, ou grandes eventos, que afetam as séries temporais;

$\epsilon_t$  – representa o erro, quaisquer alterações incomuns não acomodadas pelo modelo.

## SARIMA:

O modelo ARIMA (Autorregressivo Integrado de Média Móvel) é um dos mais populares métodos estatísticos de análise de séries temporais. Foi proposto por Box e Jenkins na década de 1970 e tem origem nos modelos autorregressivo (AR), médias móveis (MA) e da combinação dos modelos AR e MA (ARMA) (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), onde D é o termo de diferenciação sazonal do modelo. A definição do modelo é dada por (SANTOS; CHAUCOSKI, 2020):

$$\phi(B)\phi(B^s)\Delta^d\Delta_s^D X_t = \theta(B)\theta(B^s)\varepsilon_t \quad 2$$

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;

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

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

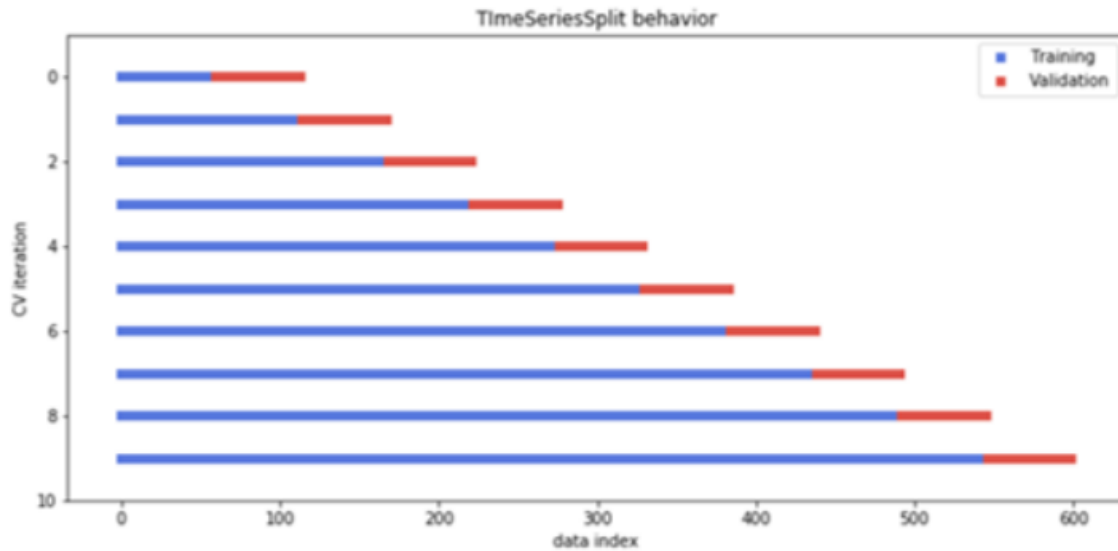
$\varepsilon_t$  – representa o ruído branco com média zero.

Observa-se que, neste trabalho, os modelos foram implementados por meio da biblioteca `pmdarima`. Esta biblioteca seleciona os parâmetros  $p, d, q$  e  $P, D, Q$  automaticamente, com base no critério de minimização dos AICs (*Akaike Information Criterion*) (CHEN, 2020).

### Treinamento, validação e teste:

O processo de treino e validação, dos modelos, utilizou a técnica *Walk Forward Simulation*. Esta técnica é usada para separar as séries temporais em intervalos de tempos fixos, expandindo o conjunto de treino em cada iteração (Figura 3) (SIMÕES, 2021). O

conjunto de teste foi formado pelos preços relativos ao período entre os meses de Julho de 2020 e Junho de 2021, totalizando 12 meses.



**Figura 3.** Validação *Walk Foward*  
 Fonte: Simões (2021).

Métricas:

Neste trabalho, os modelos obtidos foram avaliados pelas métricas RSME (*Root Mean Squared Error*) e MAPE (*Mean Absolute Percent Error*). Estas métricas são amplamente empregadas em problemas de previsão de séries temporais. O RMSE permite avaliar a qualidade de um previsor em relação aos dados e o MAPE mede o tamanho do erro em termos percentuais.

As equações das métricas RMSE e MAPE são apresentadas na Tabela 1.

**Tabela 1.** Equações – RMSE e MAPE.

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

Fonte: Cankurt e Subasi (2015).

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.

Utilizou-se, para encontrar o melhor modelo SARIMA, o critério de AKAIKE (AIC). O método AIC é baseado no processo *Maximum Likelihood Estimation* (MLE). O melhor modelo é o modelo que tem o menor AIC. O valor do AIC é obtido por meio da equação (MASLIM; ARINANDA, 2020).

$$AIC = -2(\log - likelihood) + 2K \quad 3$$

Onde:  $k$  é o número de parâmetros estimados no modelo e *log-likelihood* é o tamanho do modelo (obtido de resultados estatísticos).

### 3. RESULTADOS E DISCUSSÃO

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

**Tabela 2.** Análise descritiva.

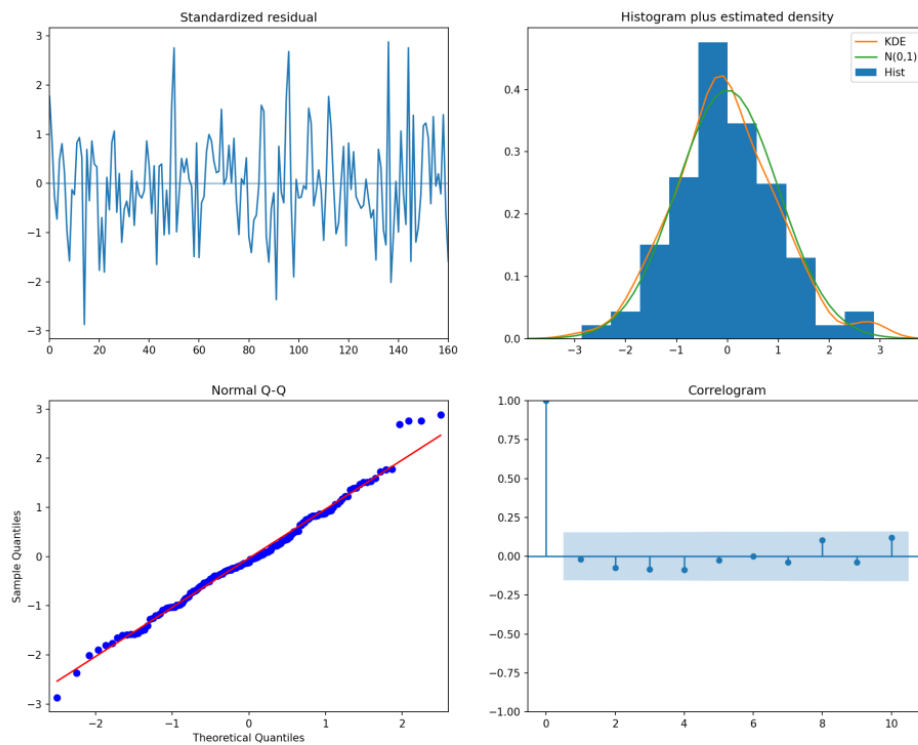
Parâmetro analisado	Resultado
Média (R\$)	56,15
Mínimo (R\$)	20,54
Máximo (R\$)	132,45
Desvio Padrão (R\$)	25,37
Coeficiente de Variação (%)	45,18

Pode-se observar, dos dados apresentados na Tabela 2, que o preço ficou, para o período em estudo, em média de R\$ 56,15. Apresentando, neste período, preços mínimo e máximo de R\$ 20,54 e R\$ 132,45, respectivamente. Observa-se também um alto coeficiente de variação dos dados (45,18%), o que indica variabilidade dos dados.

SARIMA:

Inicialmente, identificou-se, de maneira automática, por meio do pacote *pmdarima*, o modelo SARIMA(0,1,0)(4,1,1)<sub>12</sub>, 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 4). Obteve-se, do teste de normalidade (*Jarque-Bera normality test*), um p-valor de 0,2, o que revela a não rejeição da hipótese nula de normalidade dos resíduos. Observa-se do correlograma, apresentado na Figura 4, que os resíduos não são

autocorrelacionados, pois os coeficientes de autocorrelação dos resíduos são estatisticamente iguais a zero, isto é, não ultrapassam os limites de confiança.



**Figura 4.** Gráficos do modelo Sarima

Os teste estatísticos realizados, bem como o diagnóstico dos resíduos validaram o modelo SARIMA(0,1,0)(4,1,1)<sub>12</sub>.

PROPHET:

O modelo PROPHET contém um grande número de parâmetros que podem modificar o resultado do modelo. Observa-se que, neste trabalho, utilizou-se a técnica *Grid Search*. Esta técnica testa todas as combinações de parâmetros exaustivamente. Selecionando os parâmetros que obtiverem o menor erro. O melhor modelo encontrado, neste trabalho, utiliza os parâmetros apresentados na Tabela 3.



**Tabela 3.** Parâmetros do modelo PROPHET.

Parâmetros
interval_width=0.95
growth='linear'
daily_seasonality='auto'
weekly_seasonality='auto'
yearly_seasonality='auto'
seasonality_mode='multiplicative',
seasonality_prior_scale =0.1
changeoint_prior_scale=0.001,
n_changeoints=25

Na Tabela 4 apresentam-se, para os meses de Julho de 2020 a Junho de 2021, os dados observados e preditos pelos modelos SARIMA e PROPHET.

**Tabela 4.** Dados observado e preditos pelos modelos (conjunto de teste) (R\$).

Mês	IPEA	SARIMA	PROPHET
jul/20	94,653	103,0297	110,5334
ago/20	92,6	94,39031	112,2932
set/20	92,121	88,27534	106,0567
out/20	99,825	89,86824	103,472
nov/20	105,357	98,29334	101,5682
dez/20	111,995	109,4175	105,945
jan/21	106,369	105,8088	99,916
fev/21	126,429	123,8923	115,374
mar/21	132,498	136,6154	125,3175
abr/21	132,38	129,9591	126,6506
mai/21	129,563	125,2002	117,096
jun/21	131,219	135,6317	122,9532

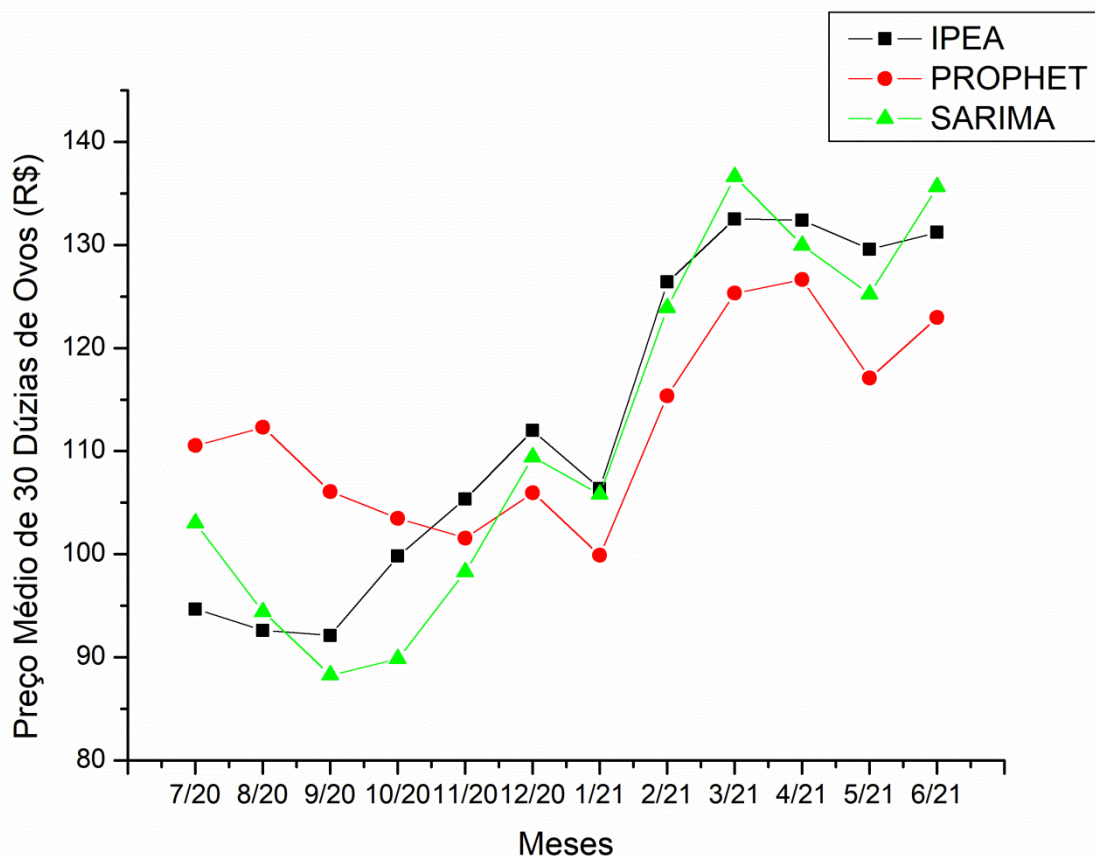
Na Tabela 5 apresentam-se os resultados das métricas para o conjunto de teste.

**Tabela 5.** Métricas

Métrica	SARIMA	PROPHET
RMSE	5,09	10,68
MAPE	4,01	8,85

Os resultados das previsões dos preços, em termos gráficos, são apresentados na Figura 5.

**Figura 5.** Preços médios de 30 dúzia de ovos (IPEA, PROPHET e SARIMA).



Por meio dos resultados apresentados, nas Tabelas 4 e 5 e na Figura 5, conclui-se que os resultados das previsões, dos dois modelos, estão próximos aos fornecidos pelo IPEA. Contudo, o modelo SARIMA apresenta erros (MAPE e RMSE) menores que o modelo PROPHET. Segundo Chen et al. (2020) os modelos ARIMA tem muito boa habilidade para lidar com pequenas quantidades de dados.

#### 4. CONSIDERAÇÕES FINAIS

Neste trabalho apresentou-se uma aplicação, de modelos de séries temporais, para previsão do preço médio de 30 dúzias de ovos, utilizando as metodologias PROPHET e SARIMA (*Seasonal Autoregressive Integrated Moving Average*), implementados por meio

da linguagem Python. Os modelos foram construídos baseados nos dados disponibilizados pelo Instituto de Pesquisa Econômica Aplicada - IPEA.

Estimaram-se, neste trabalho, os valores referentes aos meses de Julho de 2020 a Junho de 2021. Observou-se, por meio dos dados previstos, que os modelos apresentaram resultados próximos aos fornecidos pelo Instituto de Pesquisa Econômica Aplicada- IPEA. As taxas de erro do modelo de previsão SARIMA foi de 5,09 (RMSE) e 4,01% (MAPE), enquanto as taxas de erro do modelo PROPHET foi 10.68 (RMSE) e 8,85% (MAPE). O que demonstram a boa capacidade de generalização, para um horizonte de doze meses, dos modelos implementados neste trabalho. Contudo, o modelo SARIMA apresentou taxas de erro menores que o modelo PROPHET.

Finalmente, observa-se que as previsões do comportamento, do preço médio de 30 dúzias de ovos, podem auxiliar, produtores e compradores, na avaliação do setor de avicultura de postura, para um melhor entendimento do mercado e possibilitar maior segurança em seus planejamentos e negociações.

## REFERÊNCIAS

- AVISITE. **Produção de ovos cresce 3% em 2020 e atinge nível recorde**. Disponível em: <<https://www.avisite.com.br/index.php?page=noticias&id=23315>>. Acesso em: 28 jul. 2021.
- 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, 2018.
- 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.
- CHEN, Y; KANG, Y; WANG, Z. Probabilistic forecasting with temporal convolucional neural network. **Neurocomputing**, v. 399, 2020.
- IPEA. **Preço médio recebido pelo agricultor: ovo grande (30 dúzias - Paraná)**. Disponível em: < <http://www.ipeadata.gov.br/Default.aspx>>. Acesso em: 6 jul. 2021.
- MASLIM, M.; ARINANDA, K. Motorcycle parts sales forecasting using auto-Regressive Integrated moving average model. **International Journal of Computer Theory and Engineering**, v. 12, 2020.
- MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. **Introduction to time series analysis and forecasting**. New Jersey: John Wiley& Sons, 2008.
- OLIVEIRA, L.; BASTIANI, M.; TONIN, P. C. Desenvolvimento de planilhas eletrônicas para automatizar dados da produção de ovos, visando eliminar a etapa de anotação em planilha

de papéis. **In:** Congresso Brasileiro de Engenharia de Produção (ConBrepro). Ponta Grossa. 2017.

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.

PINTO, C. T. **Análise e projeção comportamental de clientes na área de venda de cimento e derivados**. Dissertação (Mestrado) – Universidade do porto, Porto, 2020.

ROJAS, L. S.; SÁNCHEZ, R. C. G.; NATA, R. G.; CORONADO, O. A. A.; ESTRADA, A. G. Box-jekins methodology to forecast White egg prices paid to the producer in mexico. **CAB Direct**, v. 53, 2019.

SANTOS, J. A. S. Aplicação de redes neurais artificiais na previsão do preço do milho no estado do paraná. **Revista Engenharia e Tecnologia**, v. 13, 2021.

SANTOS, J. A. A.; CHAUCOSKI, Y. Previsão do consumo de energia elétrica na região sudeste: um estudo de caso usando SARIMA e LSTM. **CEREUS**, v. 12, 2020.

SIMÕES, M. **Machine Learning na previsão de Séries Temporais**. Disponível em: <<https://www.bi4all.pt/noticias/blog/machine-learning-na-previsao-de-series-temporais/>>. Acesso em: 11 ago. 2021.

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

TÁPIA, M.; ROISENBERG, M.; BARRETO, J. M. Aplicação de redes neurais na previsão de preço de ovos. **In:** Congresso e Mostra de Agroinformática (InfoAgro). Ponta Grossa. 2000.

TAYLOR, S. J.; LETHAM, B. Forecasting at scale. **American Statistic**, v. 72, 2018.

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.

XIE, C.; HAOYU, W.; WENWEN, Y; et al. Trend analysis and forecast of daily reported incidence of hand, foot and mouth disease in Hubei, China by Prophet model. **Scientific Reports**, 2021.