

## PREVISÃO DA PRODUÇÃO DE MANDIOCA: UM ESTUDO DE CASO UTILIZANDO ARIMA E NEURALPROPHET

### CASSAVA PRODUCTION FORECASTING: A CASE STUDY USING ARIMA AND NEURALPROPHET

Aldino Normelio Brun Polo<sup>1</sup>      Maria Luiza Barco Catto Kossar<sup>2</sup>      André Sandmann<sup>3</sup>  
José Airton Azevedo dos Santos<sup>4</sup>      Deborah Catharine Assis Leite<sup>4</sup>

#### RESUMO

A mandioca, tubérculo brasileiro rico em carboidratos, cálcio, potássio e vitaminas, possui diversos benefícios para a saúde, tais como combater o excesso de radicais livres e fortalecer o sistema imunológico. Sua produção é de grande importância no cenário nacional, atingindo aproximadamente 18 milhões de toneladas anuais. Pode ser comercializada *in natura* ou na forma de farinha e fécula. Neste contexto, este trabalho tem como objetivo prever, por meio de modelos ARIMA (*Autoregressive Integrated Moving Average*) e NeuralProphet, a produção de mandioca no Brasil. O ARIMA é um algoritmo comumente usado para previsões de séries temporais. Já o NeuralProphet é uma ferramenta, de previsão de séries temporais, baseada no Prophet e nas Redes Neurais Artificiais. A base de dados, disponibilizada pelo IPEA (Instituto de Pesquisa Econômica Aplicada), apresenta uma série histórica, da produção de mandioca, no período entre 1920 e 2021. As métricas RMSE e MAPE foram utilizadas para avaliar o desempenho dos modelos de previsão. Os resultados obtidos, por meio deste estudo, mostraram uma boa correspondência entre os valores reais e previstos pelo modelo NeuralProphet.

**Palavras-chave:** Agricultura. Aprendizado de máquina. Aprendizado profundo. Redes neurais.

#### ABSTRACT

Cassava, a Brazilian tuber rich in carbohydrates, calcium, potassium and vitamins, has several health benefits, such as combating excess of free radicals and strengthening the immune system. Its production is of great importance on the national scene, reaching approximately 18 millions tons annually. It can be sold fresh or in the form of flour and starch. This paper aims to predict, using ARIMA (*Autoregressive Integrated Moving Average*) and NeuralProphet models, cassava production in Brazil. ARIMA is a commonly used algorithm for time series forecasting. NeuralProphet is a time series forecasting tool based on Prophet and Artificial Neural Networks. Database, made available by IPEA (Institute for Applied Economic Research), presents a historical series of cassava production in the period between 1920 and 2021. The RMSE and MAPE metrics were used to evaluate the performance of forecast models. Results showed a good correspondence between the real values and those predicted in NeuralProphet model.

**Key Words:** Agriculture. Deep learning. Machine learning. Neural networks.

<sup>1</sup> Mestrando em Tecnologias Computacionais para o Agronegócio. Universidade Tecnológica Federal do Paraná, UTFPR. Estrada para Boa Esperança, km 4, Comunidade São Cristóvão, CEP. 85660000 - Dois Vizinhos, PR - Brasil

<sup>2</sup> Mestranda em Tecnologias Computacionais para o Agronegócio. Universidade Tecnológica Federal do Paraná, UTFPR

<sup>3</sup> Doutor em Engenharia Agrícola. Universidade Federal de Campina Grande, UFCG, Brasil.

<sup>4</sup> Professor(a) da Universidade Tecnológica Federal do Paraná

## 1 INTRODUÇÃO

A mandioca (*Manihot esculenta Crantz*), também conhecida como macaxeira e aipim, é uma planta nativa da América do Sul. É uma fonte rica em carboidratos e expande suas características nutricionais ao cálcio, potássio e vitaminas, o que conseqüentemente, quando inclusa na alimentação, traz benefícios à saúde (Alves *et al.*, 2022; Souza *et al.*, 2014).

A mandioca pode ser comercializada *in natura* ou na forma de farinha e fécula. A fécula de mandioca é utilizada, por várias indústrias, na fabricação de comprimidos, pasta de dente, embutidos, papel, cola, etc. Sendo o estado do Paraná o maior produtor de fécula de mandioca do país (AES, 2020).

No ano de 2021 a produção de mandioca, segundo o IBGE (Instituto Brasileiro de Geografia e Estatística), foi de aproximadamente 18 mil toneladas, com um valor de produção de 12 milhões de reais, ocupando a sétima posição no ranking de cultivos no Brasil. No ano de 2022, segundo o mesmo instituto de pesquisa, a produção de mandioca apresentou resultados semelhantes ao ano de 2021, com 18,2 mil toneladas produzidas (IBGE, 2023).

A aplicação de modelos, para prever a produção de mandioca, oferece, aos gestores da cadeia produtiva, uma visão sistêmica do passado e do presente, fornecendo uma tendência para o futuro, auxiliando-os nos processos de tomada de decisão (Proença *et al.*, 2017). Os modelos de previsão ARIMA e NeuralProphet podem ser utilizados como ferramentas de apoio no processo de tomada de decisão.

O modelo ARIMA, proposto por Box e Jenkins na década de 1970, é um modelo estatístico muito utilizado em análises de séries temporais (Santos; Chaucoski, 2022). Já o NeuralProphet é um *framework* híbrido, baseado no Prophet e nas Redes Neurais Artificiais, também muito utilizado para análise de séries temporais (Triebe *et al.*, 2021).

Diversos trabalhos utilizaram métodos de previsão em aplicações voltadas ao mercado da mandioca. Ikuemonisan e Akinbola (2021) utilizaram o método ARIMA para previsão, na Nigéria, da produção da mandioca. Choosuk e Kengpol (2016) utilizaram Redes Neurais e Regressão Linear Múltipla para prever, na Tailândia, a produção de mandioca. Pentean *et al.* (2012) utilizaram o método de Holt-Winters para prever o preço da mandioca no Mato Grosso do Sul. Já Proença *et al.* (2017) utilizaram ARIMA e Holt-Winters para avaliar o cenário do cultivo da mandioca no Brasil.

Neste contexto, este trabalho tem como objetivo prever, por meio de modelos ARIMA e NeuralProphet, a produção de mandioca no Brasil.

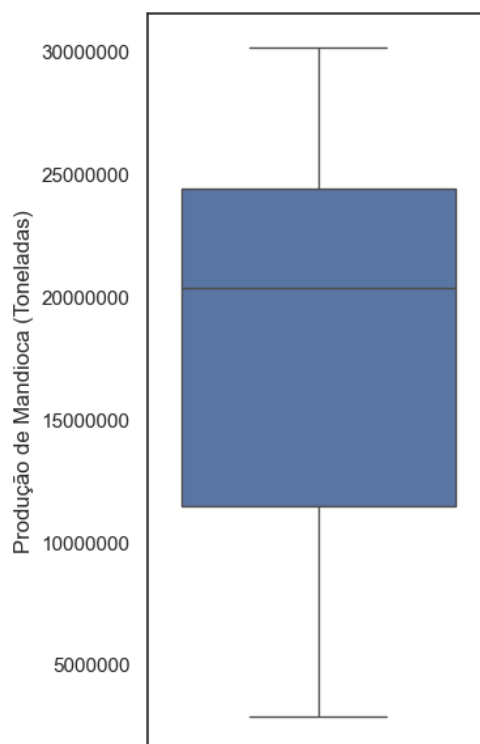
## 2 MATERIAL E MÉTODOS

Nesta seção aborda-se a base de dados, os modelos utilizados para realizar essa pesquisa, bem como as métricas utilizadas.

### 2.1 Base de dados

Para a previsão da produção de mandioca, em toneladas ( $t$ ), utilizou-se uma base de dados com 102 safras (1920 – 2021), obtida do Instituto de Pesquisa Econômica Aplicada (IPEA, 2023). Os dados coletados, no site do IPEA, já estavam limpos e sem a presença de *outliers*, como é demonstrado, na Figura 1, por meio do gráfico *Boxplot*.

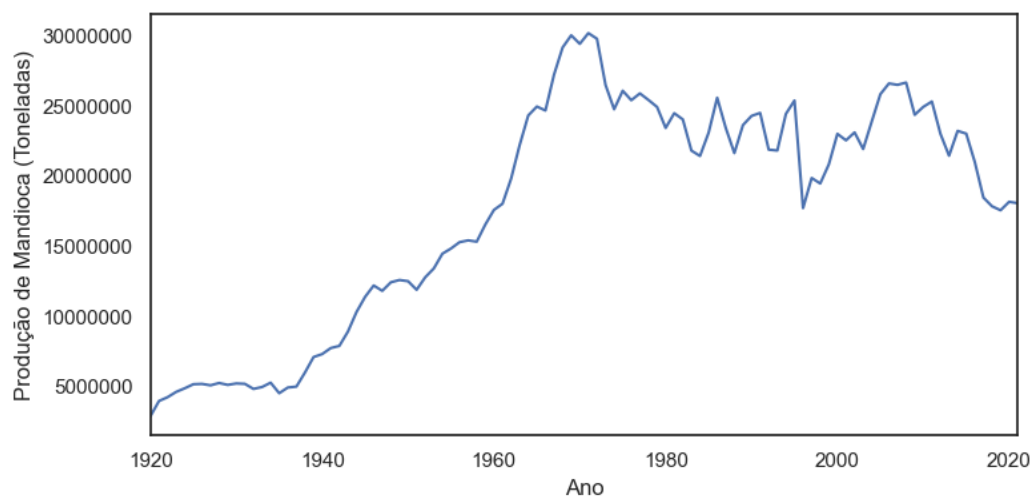
Figura 1 - *Boxplot* do conjunto de dados.



Fonte: Autoria própria (2024).

Na Figura 2 apresenta-se o gráfico da série histórica anual da produção de mandioca.

Figura 2 - Série histórica da produção de mandioca no Brasil.



Fonte: Autoria própria (2024).

Na Tabela 1 apresentam-se os cinco primeiros registros da base de dados de produção de mandioca.

Tabela 1 - Cinco primeiros registros da base de dados.

<b>Produção de Mandioca (t)</b>
2.899.000
3.987.000
4.258.000
4.636.000
4.893.000

Fonte: Autoria própria (2023).

Neste trabalho a base de dados foi dividida com 82 dados para o treinamento dos algoritmos, 10 para validação e 10 para teste.

## 2.2 ARIMA

O modelo ARIMA foi desenvolvido por Box e Jenkins na década de 1970. Este modelo possui três partes: a auto regressiva (AR), a de integração (I) e a média móvel (MA). O modelo ARIMA apresenta bom desempenho para previsões simples e de curto prazo (Gujarati; Porter, 2011; Proença *et al.*, 2017).

Modelos ARIMA são geralmente denotados como  $ARIMA(p,d,q)$ . Onde  $p$  é parte auto regressiva do conjunto de dados,  $d$  a de integração não estacionária e  $q$  a parte da média móvel (PROENÇA *et al.*, 2017). A expressão matemática do modelo ARIMA é dada por:

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

Onde:  $y$  é a variável dependente,  $\Delta_{(t-i)}$  é o operador de diferenças,  $u$  o termo de erro e  $\phi$ ,  $\beta$  e  $\alpha$  são parâmetros do modelo.

### 2.3 NeuralProphet

O NeuralProphet é um *framework* híbrido, baseado no Prophet e em Redes Neurais Artificiais, utilizado para prever dados de séries temporais. O Prophet é uma ferramenta desenvolvida pelo Facebook. O NeuralProphet utiliza redes neurais para superar as limitações do Prophet, em lidar com dados complexos (Triebe *et al.*, 2021).

A arquitetura de redes neurais utilizada pelo NeuralProphet é a LSTM (*Long Short Term Memory*). A rede LSTM é uma rede recorrente projetada para lidar com dependências temporais de dados de séries temporais. Sua arquitetura é baseada em células de memória. Isso permite que a rede se lembre de informações anteriores, da série temporal, e as use para prever o futuro (SANTOS & CHAUCOSKI, 2022).

O NeuralProphet, ao utilizar a arquitetura LSTM, permite aos modelos apreender e descrever padrões complexos de dados de séries temporais. Isto permite que o *framework* NeuralProphet produza previsões precisas (Hidarto *et al.*, 2023; Costa, 2022).

### 2.4 Métricas

Neste trabalho, os modelos ARIMA e NeuralProphet foram avaliados por duas métricas: RMSE (*Root Mean Squared Error* - Raiz quadrada do erro-médio) e MAPE (*Mean Absolute Percentage Error* - Erro Percentual Absoluto Médio). O RMSE (Equação 2) permite que seja avaliado a qualidade de um previsor em relação aos dados. Essa métrica tem como vantagem a estabilidade da unidade na mesma escala que o dado original. Já o MAPE (Equação 3) mede o tamanho do erro em porcentagem (Santos; Chaucoski, 2022).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad 2$$

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad 3$$

Sendo:  $y_i$  o valor real do período  $i$ ,  $\hat{y}_i$  é o valor previsto para o período  $i$  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 de produção de mandioca.

<b>Parâmetros</b>	<b>Resultados</b>
Quantidade	102
Média ( $t$ )	17.678.643,76
Desvio padrão ( $t$ )	8.142.593,78
Coefficiente de variação (%)	46,06
Valor mínimo ( $t$ )	2.899.000
Valor máximo ( $t$ )	30.229.166

Fonte: Autoria própria (2023).

Observa-se, da Tabela 2, que os dados apresentam um coeficiente de variação de 46,06%, considerado alto, indicando variabilidade dos dados.

#### 3.1 Escolha e ajuste dos modelos

ARIMA:

Os parâmetros  $p$ ,  $d$ ,  $q$ , do modelo ARIMA, foram obtidos por meio da biblioteca pmdarima. Pmdarima é uma biblioteca estatística, do Python, para análise de séries temporais (SANTOS *et al.*, 2023). Definiu-se, utilizando-se esta biblioteca, os parâmetros  $p=0$ ,  $d=1$ ,  $q=0$ . Os valores destes parâmetros foram selecionados por meio do critério de AKAIKE (AIC).

NeuralProphet:

A ferramenta NeuralProphet possui vários parâmetros que influenciam nos resultados do modelo. Para fazer o ajuste desses parâmetros foi utilizado a técnica de otimização de Bayes. Esta técnica consiste na otimização de uma função objetivo que levaria muito tempo para

avaliar com técnicas tradicionais (FRAZIER, 2018). Sendo assim, os melhores parâmetros, para este trabalho, são apresentados na Tabela 3.

Tabela 3 - Parâmetros do modelo NeuralProphet.

<b>Parâmetro</b>	<b>Valor</b>
<i>n_changepoints</i>	18
<i>trend_reg</i>	4
<i>yearly_seasonality</i>	'False'
<i>n_forecasts</i>	8
<i>changepoints_range</i>	0,8396
<i>learning_rate</i>	0,0868

Fonte: Autoria própria (2024).

### 3.2 Avaliação dos modelos

Na sequência, realizaram-se previsões, da produção de mandioca, para o período de 2012 a 2021 (Conjunto de Teste) (Tabela 3).

Tabela 4 - Valores reais (IPEA) e valores preditos (ARIMA e NeuralProphet).

<b>Data</b>	<b>IPEA (t)</b>	<b>NeuralProphet (t)</b>	<b>ARIMA (t)</b>
2012	23.044.557	21.431.806	25.349.542
2013	21.484.218	21.211.092	25.349.542
2014	23.253.514	20.990.978	25.349.542
2015	23.059.704	20.770.866	25.349.542
2016	21.036.314	20.550.752	25.349.542
2017	18.501.645	20.330.036	25.349.542
2018	17.877.163	20.109.922	25.349.542
2019	17.593.186	19.889.810	25.349.542
2020	18.197.572	19.669.696	25.349.542
2021	18.098.115	19.448.979	25.349.542

Fonte: Autoria própria (2024).

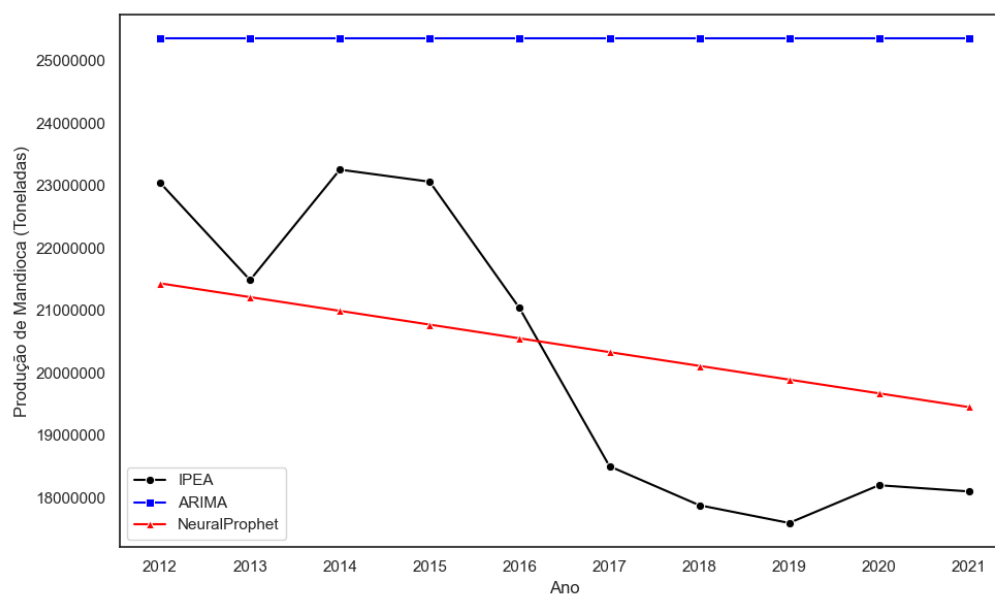
Os resultados das métricas, RMSE e MAPE, para o Conjunto de Teste são mostrados na Tabela 5 e na Figura 3.

Tabela 5 - Valores das métricas para os modelos ARIMA e NeuralProphet.

<b>Métrica</b>	<b>ARIMA</b>	<b>NeuralProphet</b>
RMSE (t)	5.613.549	1.756.811
MAPE (%)	26,97	8,12

Fonte: Autoria própria (2024).

Figura 3 - Produção de mandioca no período de 2012 a 2021 (IPEA, ARIMA e NeuralProphet).



Fonte: Autoria própria (2024).

Através das Tabelas 4 e 5 e da Figura 3, nota-se que os valores preditos, pelo modelo NeuralProphet, estão próximo aos valores reais fornecidos pelo IPEA, apresentando taxas de erro de 1.756.811 toneladas (RMSE) e 8,12% (MAPE). Já o modelo ARIMA apresentou resultados insatisfatórios, com 5.613.549 toneladas (RMSE) e 26,97% (MAPE).

## 4 CONCLUSÕES

Buscou-se avaliar, neste trabalho, modelos de previsão aplicados a produção de mandioca no Brasil. Para isso, foram utilizados os algoritmos ARIMA e NeuralProphet. Esses algoritmos foram implementados na linguagem Python. Os dados, utilizados para fazer as previsões, foram obtidos do Instituto de Pesquisa Econômica Aplicada (IPEA).

Ambos os modelos são de fácil implementação na linguagem Python. Um fator positivo para o modelo NeuralProphet é que ele combina redes neurais artificiais com métodos estatísticos tradicionais para fazer previsões.

Concluiu-se, por meios dos experimentos realizados neste trabalho, que a proximidade entre valores preditos e reais demonstram a boa capacidade de generalização, para um horizonte de 10 anos, do modelo NeuralProphet.



Por fim, para trabalhos futuros, recomenda-se o desenvolvimento de modelos multivariados com a adição de variáveis exógenas para realizar a previsão da produção de mandioca.

## REFERÊNCIAS

AES. **Indústria paranaense produz 70% da fécula de mandioca do País**. 2020. Disponível em: < <https://www.aen.pr.gov.br/Noticia/Industria-paranaense-produz-70-da-fecula-de-mandioca-do-Pais>>. Acesso em: 1 mar. 2024.

ALVES, G. C. *et al.* A Importância Sociocultural da Mandioca no Brasil. In: Encontro Latino Americano de Iniciação Científica. **Anais...** João Pessoa, Universidade do Vale do Paraíba. 2022.

CHOOSUK, N.; KENGPOL, A. An application of forecasting models for the supply and demand management of cassava products. **Applied Science and Engineering Progress**, 2016, , v. 9, n. 3, 175-187, 2016.

COSTA, R. L. C. (2022). Convolutional-LSTM networks and generalization in forecasting of household photovoltaic generation. **Engineering Applications of Artificial Intelligence**, v. 116, 2022.

FRAZIER, P. I. **A Tutorial on Bayesian Optimization**. 2018. Disponível em: <<https://arxiv.org/abs/1807.02811>>. Acesso em: 20 mar. 2024.

GUJARATI, D. N.; PORTER, D. C. **Econometria Básica**. 5.ed. Porto Alegre: AMGH, 2011.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Produção Agropecuária no Brasil**. 2023. Disponível em: < <https://www.ibge.gov.br/explica/producao-agropecuaria/mandioca/br>>. Acesso em: 18 abr. 2023.

IPEA – INSTITUTO DE PESQUISA ECONÔMICA APLICADA. **Produção de Mandioca no Brasil**. 2021. Disponível em: < <http://www.ipeadata.gov.br/Default.aspx>>. Acesso em: 19 abr. 2023.

IKUEMONISAN, E. S.; AKINBOLA, A. E. ARIMA Forecasts of Cassava Production Indicators and its Implication for Future Food Supply in Nigeria. **Tarım Ekonomisi Araştırmaları**, v. 7, n. 1, 14-30, 2021.

HIDARTO, D.; HENDRATA, F.; HARIADI, M. The application of Neural Prophet Time Series in predicting rice stock at Rice Stores. **Journal of Computer Networks, Architecture and High Performance Computing**, v. 5, n. 2, 2023.

PENTEAN, G. P. *et al.* O. O. A Previsão do preço da mandioca no extremo sul do Mato Grosso do Sul. In: XXXII Encontro Nacional de Engenharia de Produção. **Anais...** Bento Gonçalves, 2012.

PROENÇA, G. G.; SCHMIDT, C. A. P.; SANTOS, J. A. A. Construção de modelos estatísticos baseados na avaliação de séries temporais históricas da cultura da mandioca no Brasil. **Agroalimentaria**, v. 23, n. 45, 2017.

SANTOS, J. A. A.; CHAUCOSKI, Y. Aplicação de modelos de redes neurais artificiais na previsão do preço do alumínio. **ABCustos**, v. 16, n. 2, 31-49, 2021.

\_\_\_\_\_. Previsão de preço de ovos, no estado do Paraná, por meio de modelos SARIMA e PROPHET. **Revista Cereus**, v. 14, n. 1, 2022.

SANTOS, J. A. A. *et al.*. Previsão do preço do milho: uma comparação entre os modelos SARIMA e LSTM. **Peer Review**, v. 5, n. 18, 537-549, 2023.

SOUZA, R. F. *et al.* Análise econômica no cultivo da mandioca. **Revista Verde de Agroecologia e Desenvolvimento Sustentável**, v. 9, n. 2, p. 345-354, 2014.

TRIEBE, O. **NeuralProphet: Explainable Forecasting at Scale**. 2021. Disponível em: <<http://arxiv.org/abs/2111.15397>>. Acesso em: 19 abr. 2023