

# XVI ECOPAR

## Encontro de Economia Paranaense

V International Meeting on Economic Theory and Applied Economics

II Jornada Internacional de Comunicação Científica

### ARIMA, XGBOOST E TIMEGPT: UM ESTUDO DE MODELOS ALTERNATIVOS PARA A PREVISÃO DE PREÇOS DA CESTA BÁSICA DE FRANCISCO BELTRÃO (PR)

ARIMA, XGBOOST AND TIMEGPT: A STUDY OF ALTERNATIVE MODELS FOR FORECASTING THE PRICE OF THE BASIC FOOD BASKET IN FRANCISCO BELTRÃO (PR)

ARIMA, XGBOOST Y TIMEGPT: ESTUDIO DE MODELOS ALTERNATIVOS PARA PREDECIR PRECIOS DE LA CESTA BÁSICA DE FRANCISCO BELTRÃO (PR)

Andressa Carla Henrique<sup>1</sup>  
Hugo Chu<sup>2</sup>  
José Maria Ramos<sup>3</sup>  
Renan Oliveira Regis<sup>4</sup>

Área Temática: Métodos Matemáticos e Quantitativos  
JEL Code : C22

**Resumo:** Este estudo investiga e compara diferentes modelos aplicados a séries temporais para previsão de preços da cesta básica na cidade de Francisco Beltrão, Paraná. Três abordagens foram consideradas: modelos econométricos ARIMA, técnicas de *machine learning* com o algoritmo *XGBoost* e inteligência artificial com o modelo *TimeGPT*. A análise comparativa baseou-se em métricas de desempenho, como o erro médio quadrático (MSE), utilizando dados históricos sobre os preços dos itens da cesta básica de Francisco Beltrão (PR). Os resultados indicam que o modelo de inteligência artificial *TimeGPT* obteve o melhor desempenho, seguido pelo *XGBoost*. Esses achados destacam a importância de escolher a abordagem de modelagem adequada para prever os preços da cesta básica, fornecendo insights valiosos para formuladores de políticas, gestores de empresas e consumidores na região.

**Palavras-chave:** Cesta básica; Previsão de preços; Séries temporais.

**Abstract:** This study compares the prediction of the price of the basic food basket in Francisco Beltrão (Paraná) using three different forecasting methods, namely, a conventional univariate ARIMA model, a machine learning technique based on the *XGBoost* algorithm, and a generative model called *TimeGPT*. Using the mean squared error (MSE) and the Mean Absolute Error (MAE) as the performance metrics, the results indicate that *TimeGPT* achieved the best performance, followed by *XGBoost*. These findings highlight the importance of choosing the appropriate forecasting approach to predict the price of food baskets, providing valuable insights for policymakers, business managers, and consumers in the region.

<sup>1</sup> Unioeste; Brasil; ORCID; [andressaach@gmail.com](mailto:andressaach@gmail.com)

<sup>2</sup> Unioeste; Brasil; ORCID; [hgchunwei@yahoo.com.br](mailto:hgchunwei@yahoo.com.br)

<sup>3</sup> Unioeste; Brasil; ORCID; [jose.ramos@unioeste.br](mailto:jose.ramos@unioeste.br)

<sup>4</sup> Unioeste; Brasil; ORCID; [renan.regis@unioeste.br](mailto:renan.regis@unioeste.br)



# XVI ECOPAR

## Encontro de Economia Paranaense

V International Meeting on Economic Theory and Applied Economics

II Jornada Internacional de Comunicação Científica

**Key-words:** Basic basket; Price forecast; Time series.

**Resumen:** Este estudio investiga y compara diferentes modelos aplicados a series de tiempo para pronosticar precios de la canasta alimentaria en la ciudad de Francisco Beltrão, Paraná. Se consideraron tres enfoques: modelos econométricos ARIMA, técnicas de aprendizaje automático con el algoritmo XGBoost e inteligencia artificial con el modelo TimeGPT. El análisis comparativo se basó en métricas de desempeño, como el error cuadrático medio (MSE), utilizando datos históricos sobre los precios de los artículos de la canasta básica alimentaria en Francisco Beltrão (PR). Los resultados indican que el modelo de inteligencia artificial TimeGPT logró el mejor rendimiento, seguido de XGBoost. Estos hallazgos resaltan la importancia de elegir el enfoque de modelado adecuado para predecir los precios de la canasta alimentaria, proporcionando información valiosa para los formuladores de políticas, los gerentes comerciales y los consumidores de la región.

**Palabras-clave:** Canasta básicavea; Previsión de precios; Series de tiempo.

### Introdução

A previsão de preços da cesta básica em Francisco Beltrão, Paraná, é crucial para garantir o acesso a alimentos essenciais a preços acessíveis, afetando diretamente a qualidade de vida das famílias. Este estudo compara três abordagens de modelagem: ARIMA, *Machine Learning* com *XGBoost* e Inteligência Artificial com *TimeGPT*, visando identificar a mais precisa.

O ARIMA, conhecido por sua interpretabilidade, captura padrões temporais simples e complexos em séries históricas através de autorregressão, diferenciação e média móvel. Em contraste, o *Machine Learning*, como o *XGBoost*, destaca-se pela capacidade de lidar com dados complexos e não lineares, superando as limitações de linearidade do ARIMA. Por outro lado, o *TimeGPT* utiliza inteligência artificial baseada em aprendizado por transferência, adaptando-se especificamente à previsão de séries temporais com dados sequenciais.

A pesquisa visa não apenas comparar a precisão das previsões usando métricas como RMSE e MAE, mas também entender a interpretabilidade, complexidade computacional e robustez de cada abordagem. Ao analisar padrões sazonais, tendências e influências externas nos preços da cesta básica, os resultados oferecem insights valiosos para formuladores de políticas, gestores e consumidores em Francisco Beltrão.

### Procedimentos Metodológicos

O presente estudo buscou comparar métodos alternativos de previsão do preço da cesta básica da cidade de Francisco Beltrão (PR). Os dados utilizados foram coletados mensalmente pelo Grupo de Pesquisa em Economia, Agricultura e Desenvolvimento (GPEAD) da UNIOESTE, campus de Francisco Beltrão, e abrangeu o período de janeiro de 2014 a janeiro de 2024. Os preços dos itens da cesta básica foram ajustados pelo Índice de Preços ao Consumidor Amplo (IPCA).

Os dados foram divididos em conjuntos de treinamento e testados como parte da metodologia adotada. O conjunto de treinamento é utilizado para ajustar as configurações do modelo, enquanto o



# XVI ECOPAR

## Encontro de Economia Paranaense

V International Meeting on Economic Theory and Applied Economics

II Jornada Internacional de Comunicação Científica

conjunto de teste é reservado para avaliar o desempenho do modelo em dados não visíveis, simulando sua aplicação em situações reais. De acordo com Müller e Guido (2016), é fundamental dividir os dados em conjuntos de treinamento e teste para avaliar o desempenho do modelo em dados não vistos. Os modelos de previsão utilizados foram o ARIMA, *XGBoost* e *TimeGPT*. O ARIMA é uma técnica estatística clássica para modelagem de séries temporais, incorporando componentes auto regressivos, de diferenciação e de médias móveis. Por sua vez, o *XGBoost* é um algoritmo de aprendizado de máquina baseado em árvores de decisão, eficaz em problemas de regressão temporal, enquanto o *TimeGPT* adapta a arquitetura *Transformer* para capturar padrões temporais complexos, usando mecanismos de atenção e codificação posicional.

Para avaliar o desempenho dos modelos, foram empregadas métricas como RMSE (Raiz do Erro Médio Quadrático) e MAE (Erro Médio Absoluto). O RMSE quantifica a diferença média entre os valores previstos e observados, sendo sensível à escala dos dados e penalizando erros maiores de forma mais significativa. Por outro lado, o MAE calcula a média das diferenças absolutas entre os valores previstos e observados, proporcionando uma visão mais equilibrada dos erros de previsão, sem considerar a direção dos erros.

A escolha dessas métricas foi estratégica, dado que o RMSE é ideal para casos em que erros grandes precisam ser penalizados de forma mais rigorosa, enquanto o MAE é preferível quando todos os erros devem ser tratados de maneira uniforme. Essas métricas foram aplicadas para comparar como cada modelo se saiu na previsão dos preços da cesta básica, oferecendo uma avaliação abrangente e precisa do desempenho de cada abordagem analítica.

### Resultados e discussão

Partindo do princípio de que “O trabalho empírico baseado nos dados de séries temporais supõe que a série temporal subjacente seja estacionária.” (Gujarati e Porter, 2008, p. 731). Um teste de raiz unitária criado por Dickey e Fuller (1979) foi aplicado sob a hipótese nula de que a série é não estacionária, comparando uma estatística de teste (t) com um valor crítico em uma distribuição conhecida para determinar se rejeitamos a hipótese nula. Se a estatística de teste for menor que o valor crítico, podemos rejeitar a hipótese nula e concluir que a série temporal é estacionária. O p-value de 0,94, maior que p-value base de 0,05, e a comparação da estatística t (maior que os valores críticos de 1%, 5% e 10%) nos levou a concluir que a série em questão não é estacionária. Portanto, foi necessário a aplicar a técnica de diferenciação, que consiste em tomar as primeiras diferenças da série de modo que seus resíduos se tornarão estacionários. Após tomar as duas primeiras diferenças verificou-se que a série tornou estacionária e está apropriada para a modelagem de séries temporais. A figura 1 apresenta um comparativo do desempenho dos modelos baseado na métrica RMSE. O RMSE é uma das métricas mais comuns e amplamente utilizadas na avaliação de modelos de previsão. Um RMSE menor indica um melhor ajuste do modelo aos dados observados. No contexto da previsão do valor da cesta básica, o RMSE nos fornece uma medida da precisão das previsões em relação aos valores reais.

Figura 1 - Desempenho dos modelos baseado na métrica RMSE para previsão do valor da cesta básica de Francisco Beltrão-PR para o horizonte de 12 meses a partir de fevereiro/2023



# XVI ECOPAR

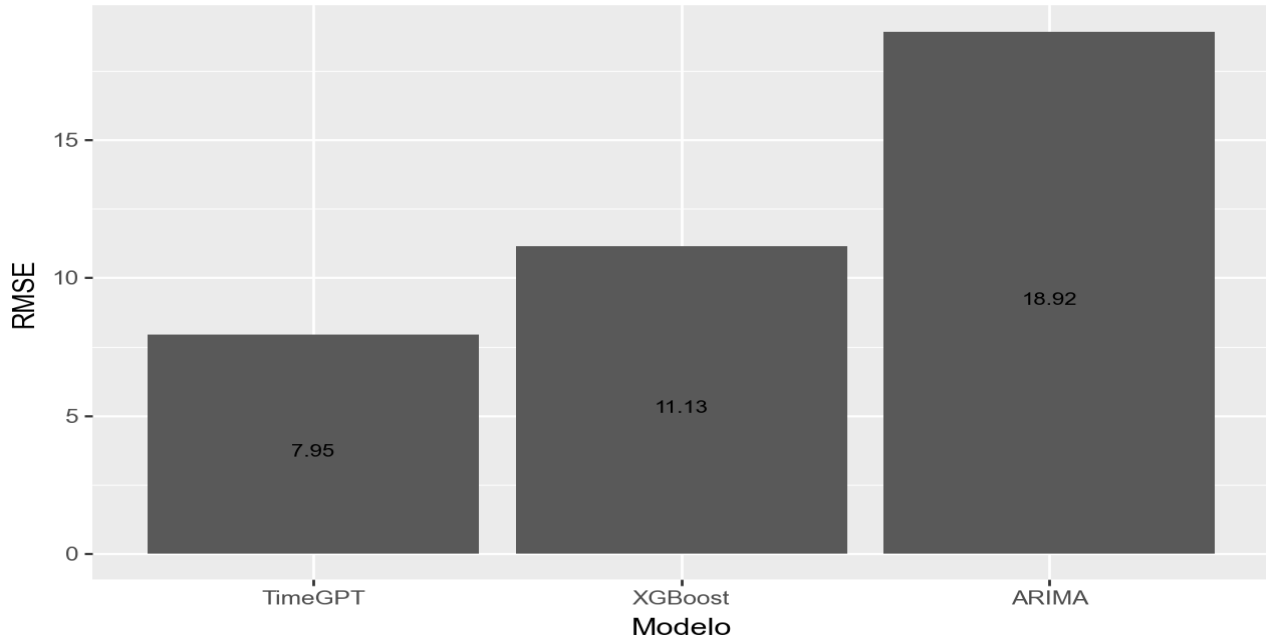
## Encontro de Economia Paranaense

### V International Meeting on Economic Theory and Applied Economics

### II Jornada Internacional de Comunicação Científica

Desempenho de modelos na previsão do valor da cesta básica - RMSE

Menor = melhor



Nota: valores para amostra de 12 meses até 01/2024

Fonte: Elaboração própria utilizando a biblioteca *Plotnine* disponível na linguagem *Python*.

Como pode-se ver através da Figura 1, o modelo *TimeGPT* apresentou o menor RMSE (7.95), indicando que suas previsões têm, em média, os menores erros quadráticos. Na sequência, o modelo *XGBoost* com um RMSE de 11.13, maior que o de *TimeGPT*, mas menor que o de *ARIMA*. Por fim, o modelo que apresentou o maior RMSE foi o *ARIMA* (14.41), sugerindo que seus erros de previsão são os maiores entre os três modelos.

A Figura 2, apresenta um comparativo dos modelos utilizando a métrica MAE, a média das diferenças absolutas entre os valores preditos pelo modelo e os valores observados, oferecendo uma medida direta da magnitude dos erros de previsão.

# XVI ECOPAR

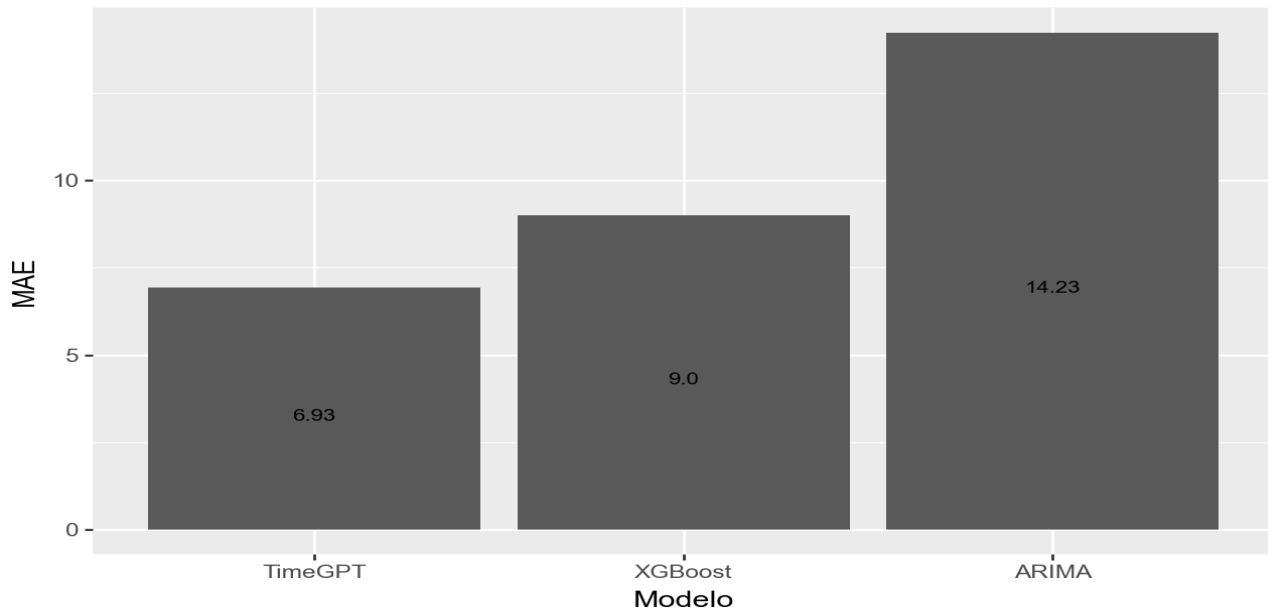
Encontro de Economia Paranaense

V International Meeting on Economic Theory and Applied Economics

II Jornada Internacional de Comunicação Científica

Figura 2 - Desempenho dos modelos baseado na métrica MAE para previsão do valor da cesta básica de Francisco Beltrão-PR para o horizonte de 12 meses a partir de fevereiro/2023

Desempenho de modelos na previsão do valor da cesta básica - MAE  
Menor = melhor



Nota: valores para amostra de 12 meses até 01/2024

Fonte: Elaboração própria utilizando a biblioteca *Plotnine* disponível na linguagem *Python*.

Analisando o gráfico de comparação de desempenho dos modelos dado pela figura 2, pode-se dizer que o modelo *TimeGPT* se destaca como o modelo mais preciso tanto em termos de MAE quanto de RMSE, indicando que suas previsões do valor da cesta básica são mais confiáveis e têm menor variabilidade em comparação aos outros dois modelos. O *XGBoost* é o segundo melhor, mas ainda apresenta erros maiores do que *TimeGPT*. *ARIMA* tem o pior desempenho em ambas as métricas, indicando que este modelo tem maiores dificuldades em prever o valor da cesta básica com precisão. Portanto, com base nas métricas apresentadas, *TimeGPT* é o modelo preferido para prever o valor da cesta básica, seguido por *XGBoost* e, por último, *ARIMA*.

## Considerações Finais.

O estudo comparou diferentes modelos para prever os preços da cesta básica em Francisco Beltrão-PR, essenciais para o planejamento financeiro e acesso a alimentos. Foram avaliados três modelos: *ARIMA*, *XGBoost* e *TimeGPT*, utilizando métricas como Erro Médio Absoluto (MAE) e Raiz do Erro Médio Quadrático (RMSE). Os resultados evidenciaram que o modelo de inteligência artificial *TimeGPT*, teve o melhor desempenho, seguido pelo *XGBoost*, enquanto o *ARIMA* foi o menos preciso, confirmando a tendência de que modelos algorítmicos superam os tradicionais na previsão de preços.



# XVI ECOPAR

## Encontro de Economia Paranaense

V International Meeting on Economic Theory and Applied Economics

II Jornada Internacional de Comunicação Científica

Apesar dos avanços, a análise identificou limitações em cada abordagem. O ARIMA captura padrões temporais, mas lida mal com nuances nos dados. O *XGBoost* é flexível, mas exige mais recursos computacionais. O *TimeGPT* destaca-se por explorar relações temporais complexas, mas requer maior processamento.

Vale ressaltar que os resultados obtidos nesta pesquisa foram específicos para a amostra analisada e podem variar em outras condições. Portanto, a generalização dos resultados deve ser feita com cautela, e análises adicionais em diferentes cenários são necessárias para uma compreensão mais completa e robusta do tema.

Em suma, este estudo contribui para a compreensão da previsão de preços da cesta básica, oferecendo *insights* valiosos para formuladores de políticas e gestores. Para uma análise mais abrangente e robusta, é fundamental considerar diferentes contextos, especificações e metodologias.

### Referências

DICKEY, D. A., & FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. **Journal of the American Statistical Association**, [S. l.], p. 427-431, 2024.

GRUPO DE PESQUISA ECONOMIA, AGRICULTURA E DESENVOLVIMENTO. Cesta Básica de Alimentos de Dois Vizinhos, Francisco Beltrão e Pato Branco. Jan 2014 – Jan 2024. Disponível em: <https://www.unioeste.br/portal/determinacao-do-custo-da-cesta-basica-de-alimentos/francisco-beltrao/boletins>. Acesso em: 28 abr. 2024.

GUJARATI, D. N., PORTER, D. C. **Basic Econometrics**. 5. ed. New York, NY: The McGraw-Hill Companies, Inc., 2008.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). IPCA - Variação mensal (% [janeiro 2014 a janeiro 2024]): Recuperado de SIDRA - Sistema IBGE de Recuperação Automática. Disponível em: <https://sidra.ibge.gov.br/tabela/1737>. Acesso em: 28 abr. 2024.

MÜLLER, A. C. & GUIDO S. **Introduction to Machine Learning with Python: A Guide for Data Scientists**. [S. l.: s. n.], 2016.

