



# Modelagem e Previsão da Receita do ICMS no Pará: Uma Abordagem Comparativa entre SARIMA e Perceptron Multicamadas

Nicolý Rodrigues Da Costa<sup>1</sup>, Vinicius Duarte Lima<sup>2</sup>,  
<sup>1</sup> UFPA, E-mail: Nicolýrodrigues1111@gmail.com <sup>2</sup> UFPA, E-mail: vinicius@ufpa.br



## Introdução

O ICMS, Imposto sobre Circulação de Mercadorias e Serviços, é um imposto estadual que incide sobre a circulação de mercadorias, a prestação de serviços de transporte interestadual e intermunicipal e a comunicação (SEFA). O ICMS é a principal fonte de arrecadação própria dos estados (ASCOM). No ano de 2023, o ICMS representou 61,95% da receita própria do estado do Pará, conforme indicado pela Secretaria da Fazenda do Estado do Pará.

Fazer a previsão de ICMS permite aos estados planejar melhor suas finanças, garantindo que haja recursos suficientes para cumprir com suas obrigações e executar políticas públicas. Além disso, uma previsão precisa ajuda na identificação de tendências econômicas, permitindo ajustes nas políticas fiscais e na adoção de medidas que possam estimular a economia local (Sabbag, 2010). A arrecadação total de ICMS do Estado foi distribuída em 18 segmentos, tais como Combustível, Comércio Atacadista, Comércio Varejista, Energia Elétrica, entre outros (Anuário, 2015)

Neste trabalho, será realizada uma análise comparativa entre a metodologia univariada de Box-Jenkins, utilizando o modelo SARIMA, e o modelo de redes neurais Perceptron Multicamadas. A análise será conduzida com base na série temporal do ICMS total arrecadado e na soma dos segmentos econômicos arrecadados no estado do Pará, considerando ambas as abordagens.

Figura 1- Série histórica ajustada da arrecadação de ICMS, abril de 2017 a dezembro de 2023.

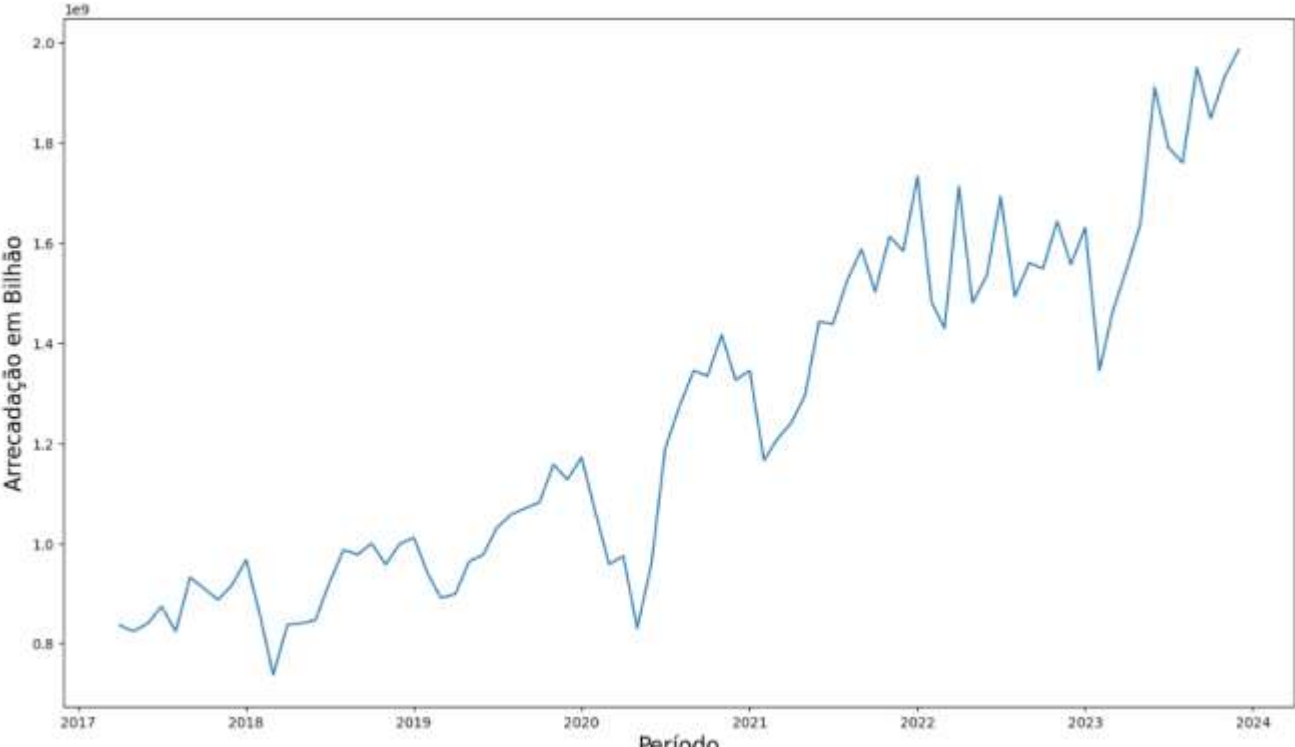
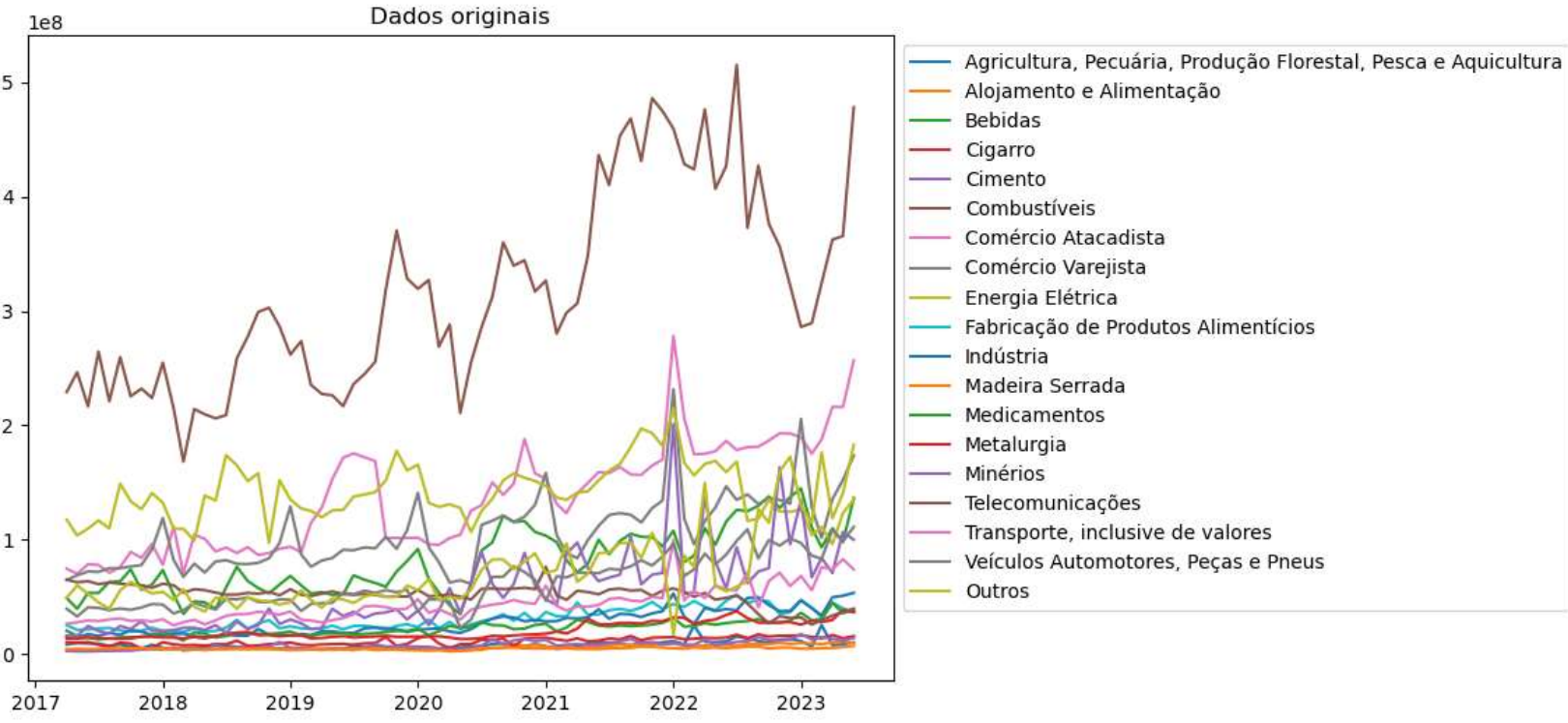


Figura 2- Série histórica ajustada da arrecadação dos segmentos econômicos, abril de 2017 a dezembro de 2023.



## Metodologia

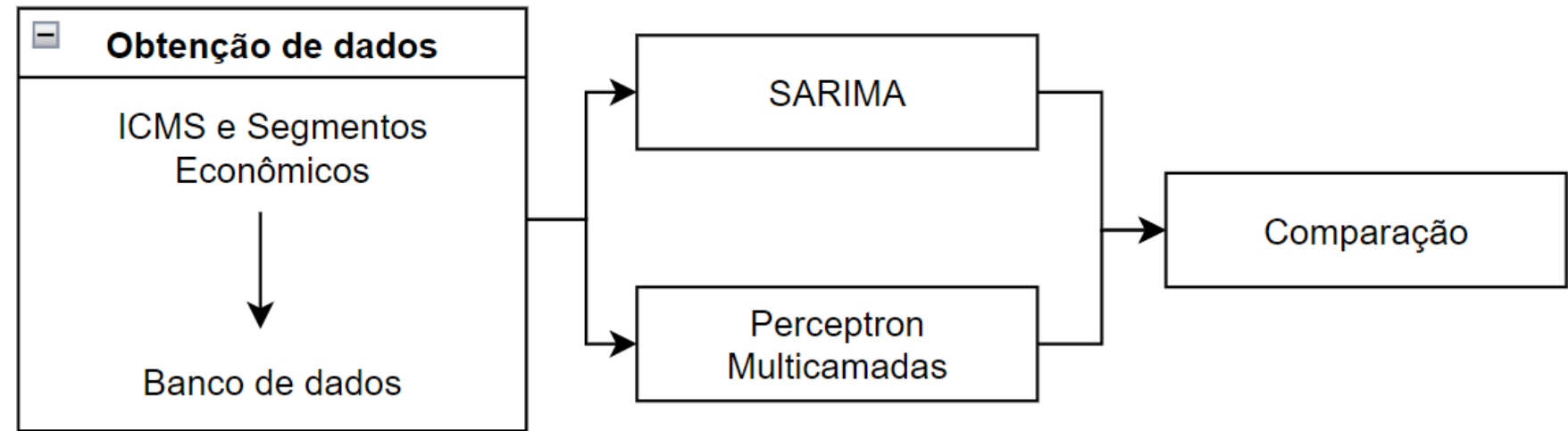
Na Figura 3, é ilustrado o fluxo geral da metodologia adotada neste estudo. Inicialmente, os dados foram extraídos do site da Secretaria da Fazenda (SEFA) do Pará, através de sua Diretoria de Arrecadação e Informações Fazendária (DAIF), no endereço eletrônico (<http://antigo.sefa.pa.gov.br/index.php/orientacoes/manual-de-atendimento/beneficio-fiscal/12783-boletim-mensal-de-arrecadacao>).

Foram baixados os dados de arrecadação total de ICMS do estado do Pará, bem como a arrecadação de cada um dos 18 segmentos econômicos do estado, abrangendo o período de abril de 2017 a dezembro de 2023, totalizando 81 observações mensais.

Foi feita a substituição de dois valores em duas variáveis (ICMS total e o segmento econômico Minério) que estavam muito acima da média devido ao Refis (programa de incentivo à regularização de débitos que permite o pagamento de dívidas de ICMS). Para substituir os valores de janeiro de 2022 e novembro de 2022, subtraíu-se o valor total do Refis (encontrado no boletim mensal da SEFA) do ICMS total. Para o segmento Minério, a substituição foi feita utilizando médias móveis.

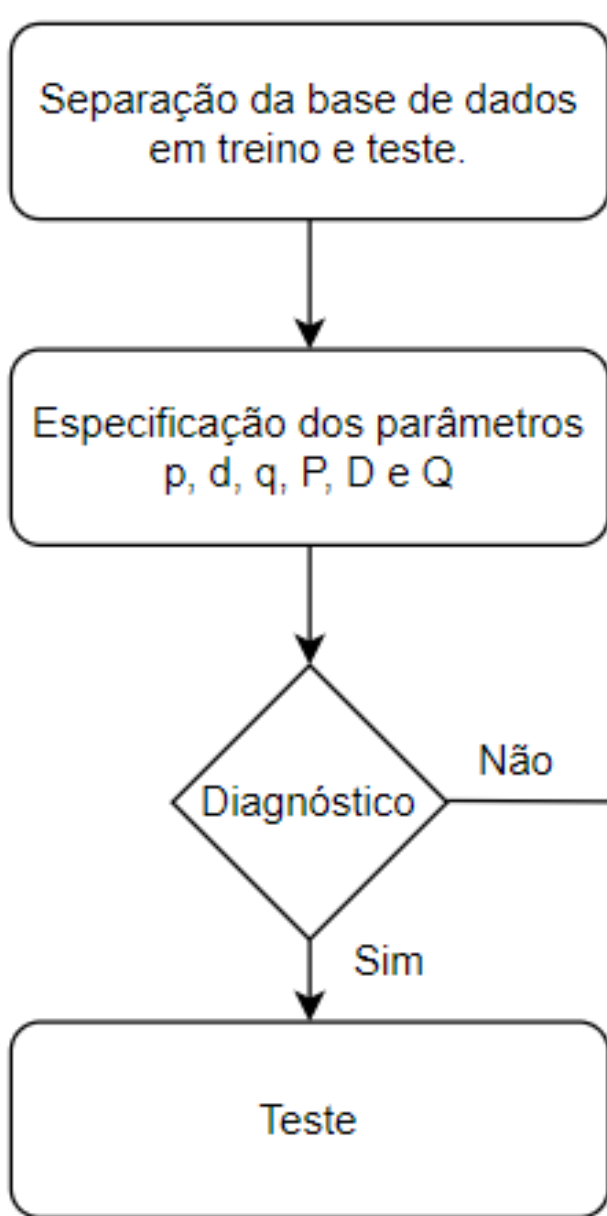
Foi realizada a modelagem para cada uma das séries temporais, utilizando o modelo SARIMA (Figura 4) e o modelo Perceptron Multicamadas (Figura 5). Para cada um dos modelos, as previsões dos segmentos econômicos foram somadas e comparadas com a previsão da série de ICMS total, a fim de verificar qual abordagem era mais eficaz. Em seguida, foram comparados os dois modelos utilizados com base nessas previsões.

Figura 3- Fluxograma ilustrando a metodologia geral de modelagem e comparação entre os modelos.



## Modelo AutoRegressivo Integrado de Média Móvel com Sazonalidade (SARIMA)

Figura 4- Fluxograma ilustrando a metodologia SARIMA.



Para a implementação do modelo SARIMA, seguimos a metodologia Box-Jenkins. Inicialmente, procedemos à divisão dos dados em conjuntos de treinamento e teste, conjunto de treino abrangeu o período de abril de 2017 a junho de 2023, totalizando 75 amostras, enquanto o conjunto de teste compreendeu o período de julho de 2023 a dezembro de 2023, totalizando 6 amostras.

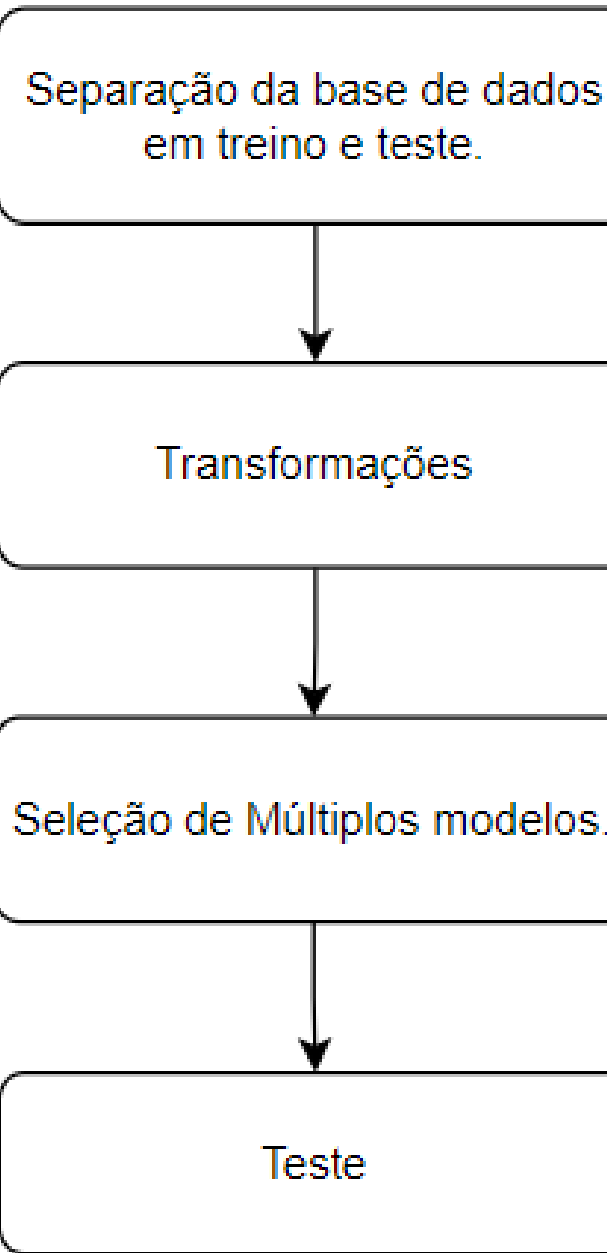
Para cada uma das séries temporais de treinamento, realizamos uma verificação de estacionariedade utilizando o teste de Dickey-Fuller e a análise do gráfico da Função de Autocorrelação (FAC). Além disso, foi investigada a presença de heterocedasticidade por meio do teste de média versus amplitude.

Para a especificação dos parâmetros, analisamos a Função de Autocorrelação (FAC) e a Função de Autocorrelação Parcial (FACP), selecionando um conjunto adequado de parâmetros, usando o período sazonal igual a 12. Estes foram aplicados em cada uma das séries, utilizando o Critério de Informação de Akaike (AIC) como métrica para minimização.

Na fase de diagnóstico, foi apresentada a correlação serial dos resíduos dos modelos pré-estabelecidos. Caso a correlação possua valores próximos de zero (dentro do intervalo de confiança), significa que o modelo consegue modelar as séries sem deixar uma relação significativa entre os resíduos. Se isso for confirmado, o modelo avança para a fase seguinte. Na fase seguinte, são realizadas as previsões das séries de treinamento, que são então comparadas com as séries de teste.

## Rede Neural Perceptron Multicamadas

Figura 5- Fluxograma ilustrando a metodologia da Rede Neural Perceptron Multicamadas.



Para a implementação da rede neural Perceptron Multicamadas, primeiramente, a base de dados foi dividida em conjuntos de treino e teste para todas as variáveis. A divisão foi feita da mesma forma que foi demonstrado no processo do modelo SARIMA. Após a divisão da base de dados, foram realizadas as devidas transformações no conjunto de treino, consistindo apenas na normalização dessas séries temporais.

Na etapa seguinte, foi realizada a seleção dos parâmetros a serem utilizados para cada uma das séries. Para isso, foi criada uma função que implementa uma rede neural com uma camada de entrada, duas camadas densas ocultas e uma camada densa de saída. Os parâmetros testados incluíram diferentes funções de ativação (ReLU, tanh e linear) e o número de neurônios nas camadas ocultas (10 e 30). As escolhas foram feitas com base na minimização do erro absoluto médio no conjunto de validação, a partir do ajuste no conjunto de treino.

Alguns parâmetros foram mantidos constantes durante os testes para avaliar seu impacto, incluindo o otimizador (SGD), a função de perda (erro quadrático médio) e a métrica de avaliação (accuracy). O número de épocas variou entre 100 e 1000, com um tamanho de lote (batch size) de 32 e uma divisão de validação (validation split) de 0,1. Após determinar os melhores parâmetros, a previsão foi realizada no conjunto de teste para uma futura comparação.

## Resultados e discussões

Os resultados da aplicação dos modelos SARIMA e Perceptron Multicamadas são apresentados nas Figuras 6 e 7, respectivamente. Nas figuras 6(a) e 7(a), a soma das previsões de cada segmento econômico é comparada com os dados de teste do ICMS. Nas figuras 6(b) e 7(b), a previsão total do ICMS é comparada com os dados de teste do ICMS. Na Tabela 1, são apresentadas as métricas de erro para cada um desses quatro resultados, incluindo o Erro Percentual Absoluto Médio (MAPE), o Erro Absoluto Médio (MAE) e a Raiz do Erro Quadrático Médio (RMSE).

Figura 6 - Soma das previsões das séries temporais de segmentos e previsão da série temporal de ICMS, usando SARIMA.

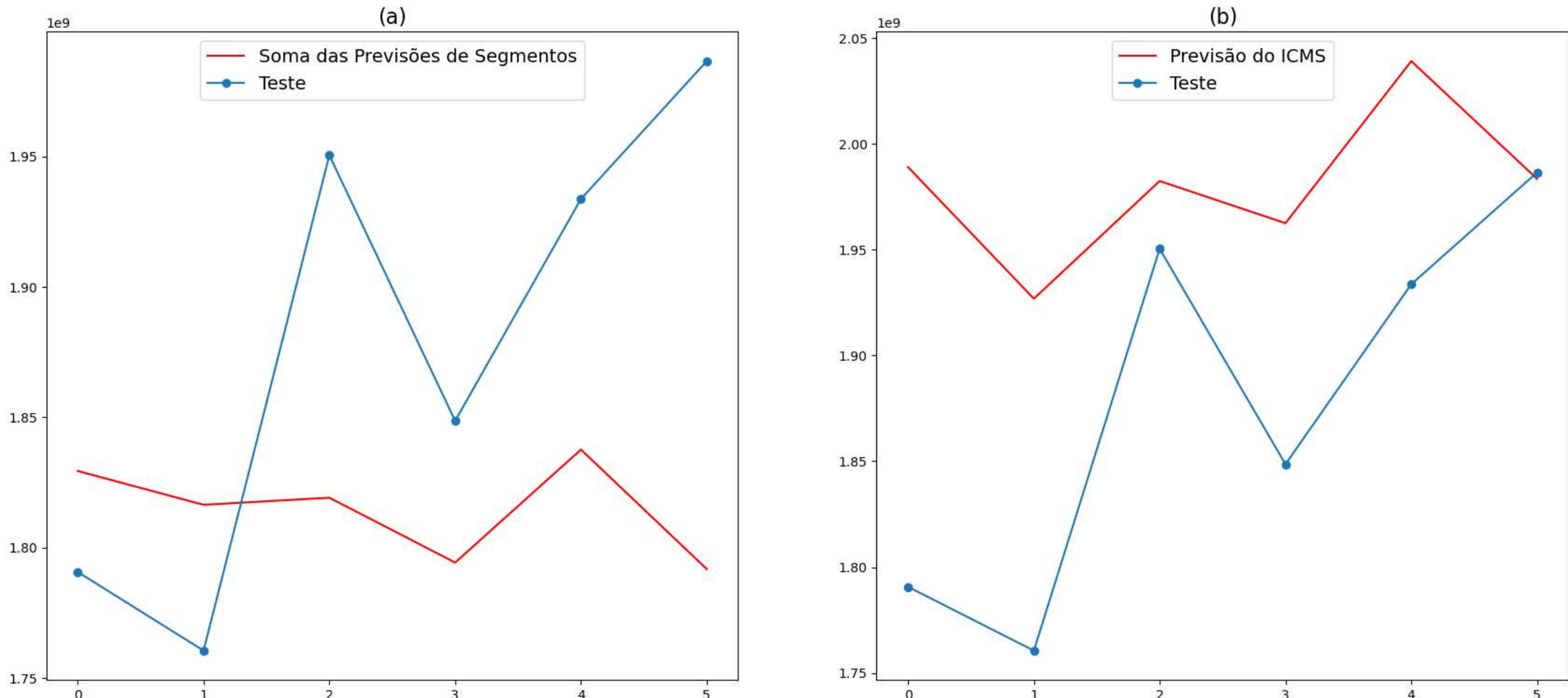


Figura 7 - Soma das previsões das séries temporais de segmentos e previsão da série temporal de ICMS, usando Perceptron Multicamadas.

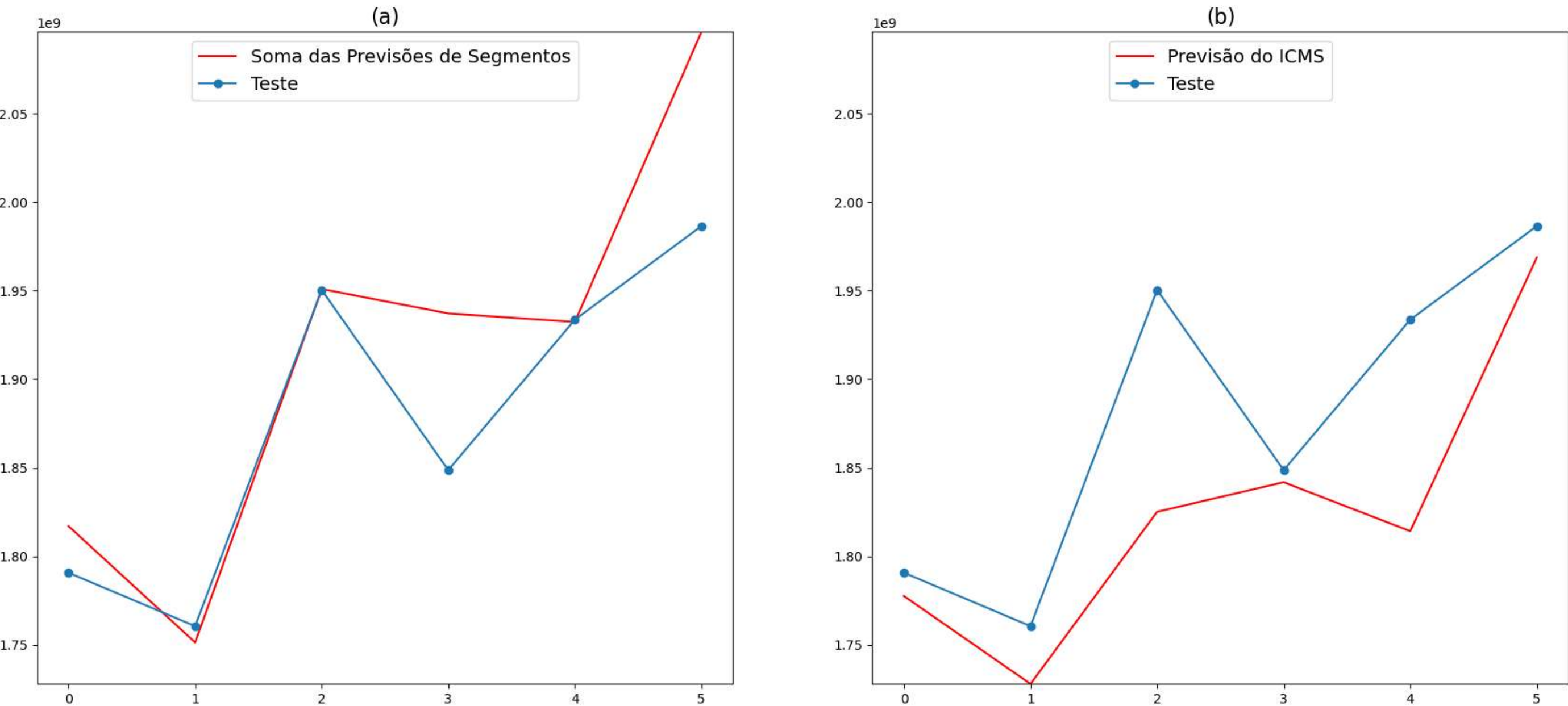


Tabela 1 – Métricas de erros usado na série de teste e de previsão nos dois modelos e nas duas técnicas de previsões demonstrados nas figuras 6 e 7.

Modelo	Série Temporal	MAPE	RMSE	MAE
SARIMA	Segmentos	4,96	109523645,2	95180577,52
	ICMS	8,35	168340504,5	165814371,2
Perceptron Multicamadas	Segmentos	2,06	58684236,14	39273033,21
	ICMS	2,74	72556630,52	52536399,92

De acordo com os resultados apresentados, o modelo que obteve os menores valores nas métricas de erro foi a rede neural Perceptron Multicamadas. A técnica que apresentou o melhor desempenho, segundo essas métricas, foi a soma das previsões dos segmentos econômicos. Portanto, com base nos modelos e técnicas avaliados, a combinação mais eficaz é a soma das previsões dos segmentos econômicos utilizando o modelo de rede neural Perceptron Multicamadas.

## Conclusões

Este estudo comparou o modelo univariado SARIMA e o modelo de redes neurais Perceptron Multicamadas na previsão da arrecadação do ICMS no Pará. Os resultados indicaram que o Perceptron Multicamadas apresentou um menor valor nas métricas de erro comparado ao SARIMA. A técnica que utilizou a soma das previsões dos segmentos econômicos no modelo de redes neurais foi a mais eficaz.

Esses achados sugerem que modelos de redes neurais são mais eficazes para prever a arrecadação de ICMS devido à sua capacidade de capturar padrões não lineares e interações complexas. Recomenda-se explorar mais o uso de redes neurais e outros métodos de aprendizado de máquina na previsão de arrecadação de impostos. Integrar variáveis macroeconômicas pode aumentar ainda mais a precisão das previsões. Este estudo destaca a importância de abordagens inovadoras na gestão financeira pública e na formulação de políticas econômicas.

Código do trabalho disponível em: <https://github.com/nicolycosta>

## Referências bibliográficas

ASCOM. Pará tem alíquota modal do ICMS de 19% a partir de 16/03/23. Disponível em: <http://antigo.sefa.pa.gov.br/noticias/20714-para-tem-aliquota-modal-do-icms-de-19-a-partir-de-16-03-23>. Acesso em: 22 mar. 2024.

SABBAG, Eduardo. Manual de direito tributário. 2. ed. São Paulo: Saraiva, 2010.

SECRETARIA DA FAZENDA DO ESTADO DO PARÁ. Boletim mensal de arrecadação. Disponível em: <http://antigo.sefa.pa.gov.br/index.php/orientacoes/manual-de-atendimento/beneficio-fiscal/12783-boletim-mensal-de-arrecadacao>. Acesso em: 02 fev. 2024.

SECRETARIA DA FAZENDA DO PARÁ. ICMS. Disponível em: <https://www.sefa.pa.gov.br/internal/services/ifc>. Acesso em: 17 mar. 2024.