



Universidade de Brasília  
Departamento de Estatística

## Análise Comparativa de Modelos Estatísticos para Previsão do IPCA

Ana Clara Barbosa de França

Relatório apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Brasília  
2025

**Ana Clara Barbosa de França**

**Análise Comparativa de Modelos Estatísticos para Previsão do IPCA**

Orientador: Prof. Helton Saulo Bezerra dos Santos

Relatório apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília  
2025**

---

# Agradecimentos

Agradeço imensamente aos meus pais por todo o suporte, compreensão e tudo o que fizeram para que eu chegasse aqui. Agradeço também à minha irmã, Mariana, que tantas vezes, sem nem saber, me distraiu e encheu esses anos de momentos divertidos.

Agradeço ao meu namorado, Felipe, um dos grandes presentes que a graduação me trouxe, por todos os passeios pela UnB, lanches no subsolo, caronas e muito mais; você esteve ao meu lado e tornou essa caminhada inesquecível.

Aos meus amigos, João, Vinícius e Davi, seja pelas risadas compartilhadas, desespere conjunto em semana de provas ou madrugadas fazendo trabalho, com certeza o curso não teria sido o mesmo sem vocês.

Às minhas amigas, Eduarda, Camillinha e Grazi, nós permanecemos firmes da escola até aqui e que possamos continuar assim. É muito bom saber que tenho vocês ao meu lado.

Ao meu orientador, Prof. Helton, pela paciência, orientação e ajuda. Seu auxílio foi imprescindível e sou muito grata pela oportunidade.

---

# Resumo

Este estudo visa comparar a performance de diferentes modelos de séries temporais e redes neurais para a previsão do Índice Nacional de Preços ao Consumidor Amplo (IPCA). Este trabalho utiliza um conjunto de dados composto pela própria série histórica do IPCA e 123 outras covariáveis socioeconômicas que abrangem o período de janeiro de 2011 a setembro de 2024. A validação cruzada foi empregada e a performance dos modelos para diversos horizontes foi mensurada por meio de métricas como o MAE e o RMSE. A validação de uma parte das previsões para dados fora da amostra também foi possível, utilizando índices reais de outubro de 2024 a maio de 2025. A análise do desempenho evidenciou que os modelos que incorporam variáveis exógenas apresentaram desempenho superior. Para o ano de 2025, a projeção média do IPCA estimada pelos modelos foi de aproximadamente 5%. Para os anos subsequentes, 2026 e 2027, as projeções se mantêm ligeiramente superiores a 4%. Em geral, os resultados mostram que há uma boa convergência entre as previsões dos modelos e as expectativas do mercado no período 2025-2027.

Palavras-chave: IPCA, Séries Temporais, Redes Neurais, Previsão, NNAR, ANN, SARIMAX.

# Lista de Tabelas

1	RMSE dos modelos para diferentes horizontes (meses). . . . .	21
2	MAE dos modelos para diferentes horizontes (meses). . . . .	22
3	Previsões e métricas de erro para o índice (out/2024 a mai/2025). . . . .	23
4	Previsões e métricas de erro para a variação acumulada ao ano(out/2024 a mai/2025). . . . .	25
5	Previsões dos modelos para a variação acumulada do IPCA (dezembro de 2025 a 2027). . . . .	26
6	Variáveis selecionadas pelo LASSO . . . . .	30

## **Lista de Figuras**

1	Arquitetura de uma Rede Neural adaptada de Haykin (1994) . . . . .	11
2	Validação cruzada de Hyndman e Athanasopoulos (2018) . . . . .	15
3	Evolução do IPCA (Jan/2011 - Mai/2025) . . . . .	18
4	Variação mensal do IPCA (Fevereiro/2011 a maio/2025) . . . . .	18
5	RMSE por horizonte de previsão mensal. . . . .	21
6	MAE por horizonte de previsão mensal. . . . .	22
7	Comparação das previsões com o índice real - out/2024-mai/2025 . . . . .	24
8	Comparação das previsões para a variação acumulada ao ano - out/2024-mai/2025. . . . .	25

## Sumário

<b>1 Introdução . . . . .</b>	<b>8</b>
<b>2 Descrição dos modelos . . . . .</b>	<b>10</b>
2.1 SARIMAX/SARIMA . . . . .	10
2.2 RNA . . . . .	10
2.3 NNAR . . . . .	11
<b>3 Detalhes de implementação e dos dados . . . . .</b>	<b>13</b>
3.1 Regularização LASSO . . . . .	13
3.2 Validação Cruzada . . . . .	14
3.3 Conjunto de dados . . . . .	15
3.4 Implementação dos modelos . . . . .	16
<b>4 Resultados das previsões . . . . .</b>	<b>17</b>
4.1 Análise descritiva . . . . .	17
4.2 Seleção de covariáveis . . . . .	18
4.3 Resultados da validação cruzada . . . . .	20
4.4 Comparação com valores reais . . . . .	22
4.5 Previsões para 2025–2027 . . . . .	26
<b>5 Considerações finais . . . . .</b>	<b>27</b>
<b>Referências . . . . .</b>	<b>28</b>
<b>Apêndice . . . . .</b>	<b>30</b>
<b>A Variáveis selecionadas pelo LASSO . . . . .</b>	<b>30</b>

# 1 Introdução

A inflação é um indicador que mede a variação no valor de bens e serviços em um determinado período e região (DORNBUSCH; FISCHER; STARTZ, 2013). No Brasil, a inflação é acompanhada principalmente através do Índice Nacional de Preços ao Consumidor Amplo (IPCA), que serve como o termômetro oficial da inflação no país e será o foco deste trabalho.

O IPCA é um tipo de índice de Laspeyres (IBGE, 2024) utilizado para medir a variação dos preços de um conjunto de bens e serviços consumidos por famílias com renda mensal entre 1 e 40 salários mínimos. Esse conjunto é composto por nove categorias básicas, incluindo alimentação, transporte, educação, comunicação, saúde, vestuário e habitação. O cálculo do IPCA é realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), sendo divulgado tanto em valores absolutos quanto acumulados ao longo do tempo.

As variações nos preços dos bens e serviços impactam diretamente o poder de compra da população, exigindo uma maior quantidade de dinheiro para manter o mesmo padrão de consumo. Nesse contexto, a previsão do IPCA assume grande importância, pois influencia diretamente a definição da taxa básica de juros do país (Selic), além de orientar o planejamento de políticas públicas. O governo utiliza essas projeções para estimar o crescimento da arrecadação tributária, os gastos públicos e para realizar correções de benefícios sociais, como o salário mínimo e as aposentadorias. Assim, previsões acuradas da inflação podem contribuir significativamente para a estabilidade econômica do país.

Nesse cenário, destaca-se o papel do Banco Central do Brasil (BCB), que publica semanalmente estatísticas de expectativas de mercado para as principais variáveis macroeconômicas, incluindo o IPCA, por meio do Relatório Focus. Essas expectativas são insumos essenciais para as decisões de política monetária e permitem que empresas e cidadãos tenham acesso ao que os agentes de mercado estão projetando, constituindo uma importante ferramenta para o planejamento de ações no curto, médio e longo prazo (BCB, 2025).

A coleta dessas informações ocorre por meio do Sistema de Expectativas de Mercado, desenvolvido pelo BCB. Este sistema reúne projeções fornecidas por uma ampla gama de instituições participantes. Em junho de 2025, o sistema contava com 159 instituições ativas, representando um espectro diversificado do mercado financeiro e corporativo nacional.



A divulgação pública e regular dessas expectativas fortalece a transparência e a credibilidade da política monetária brasileira, além de proporcionar previsibilidade aos agentes econômicos, contribuindo para o funcionamento adequado dos mercados financeiros e para a ancoragem das expectativas de inflação dentro do regime de metas de inflação vigente no Brasil.

Dada a grande relevância da previsão do IPCA e os avanços no desenvolvimento de modelos de *machine learning* (ML), que se destacam em relação aos métodos tradicionais por sua capacidade de lidar com Big Data, diversos estudos têm sido conduzidos para comparar o desempenho dessas abordagens. Entre eles, Santos e Nolau (2022) discutem o uso de diferentes técnicas de previsão e avaliam sua performance conforme o horizonte temporal, seja de curto ou longo prazo. Além disso, o trabalho de Araujo e Gaglianone (2023) analisa 50 métodos de previsão, abrangendo tanto técnicas de ML quanto modelos econométricos tradicionais, com resultados favoráveis ao desempenho dos modelos baseados em ML. Fora do Brasil, estudos como o de Chu e Qureshi (2023) já realizam comparações de modelos na previsão do PIB dos Estados Unidos, evidenciando o crescente interesse global pelo uso de diferentes abordagens preditivas em macroeconomia.

O tema central deste trabalho é a aplicação de técnicas estatísticas avançadas na análise e previsão do IPCA. Serão explorados modelos de séries temporais e redes neurais artificiais, permitindo uma abordagem abrangente e comparativa. Particularmente, serão utilizados as seguintes abordagens: modelos sazonal autorregressivo integrado de médias móveis com termos exógenos (SARIMAX) (BOX et al., 2015; TARSITANO; AMERISE, 2017) e o caso particular, SARIMA, redes neurais artificiais (RNA) (HAYKIN, 1994), e redes neurais autorregressivas (NNAR) (HYNDMAN; ATHANASOPOULOS, 2018). A escolha deste tema justifica-se pela relevância de previsões precisas na condução de políticas econômicas e pelo interesse em avaliar a eficácia de diferentes modelagens estatísticas. O estudo será fundamentado em dados reais e recentes, proporcionando uma análise prática.

O trabalho está estruturado em mais quatro capítulos além desta introdução. No Capítulo 2, a descrição dos modelos é apresentada. No Capítulo 3, tanto a base de dados quanto detalhes de implementação são descritos. No Capítulo 4, os resultados obtidos são discutidos e, por fim, no Capítulo 5, a conclusão do trabalho é apresentada.

## 2 Descrição dos modelos

Neste capítulo são apresentados os modelos a serem utilizados na previsão da série considerada.

### 2.1 SARIMAX/SARIMA

O modelo SARIMAX é uma ampliação do SARIMA que incorpora variáveis exógenas (também chamadas de covariáveis). A forma geral do modelo pode ser expressa por (BOX et al., 2015; TARSITANO; AMERISE, 2017):

$$\phi_p(L)\Omega_P(L^s)(1-L)^d(1-L^s)^D y_t = \theta_q(L)\Psi_Q(L^s)\varepsilon_t + \mathbf{x}_t^\top \boldsymbol{\beta}, \quad (2.1.1)$$

em que  $y_t$  representa a série temporal no instante  $t$ , e  $\varepsilon_t$  é um termo de erro (ruído branco). O operador de defasagem  $L$  é definido como  $L^l y_t = y_{t-l}$ . O vetor de covariáveis é dado por  $\mathbf{x}_t^\top = (x_{t1}, \dots, x_{tk})^\top$ , com os respectivos coeficientes reunidos em  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)^\top$ . O parâmetro  $s$  indica a periodicidade sazonal, enquanto  $d$  e  $D$  são as ordens de diferenciação regular e sazonal, utilizadas para estacionarizar a série.

Os polinômios autorregressivos são definidos como  $\phi_p(L) = 1 - \phi_1 L - \dots - \phi_p L^p$  (componente não sazonal) e  $\Omega_P(L^s) = 1 - \varphi_1 L^s - \dots - \varphi_P L^{sP}$  (componente sazonal). Os polinômios de médias móveis são dados por  $\theta_q(L) = 1 + \theta_1 L + \dots + \theta_q L^q$  e  $\Psi_Q(L^s) = 1 + \vartheta_1 L^s + \dots + \vartheta_Q L^{sQ}$ . As ordens  $p$  e  $P$  correspondem às ordens autorregressivas não sazonais e sazonais, respectivamente, enquanto  $q$  e  $Q$  se referem às ordens das médias móveis. O operador de diferenciação  $(1-L)^d(1-L^s)^D$  remove as raízes unitárias, assegurando que a série transformada  $(1-L)^d(1-L^s)^D y_t$  seja estacionária.

### 2.2 RNA

RNAs são técnicas computacionais baseadas em modelos matemáticos cujos algoritmos são inspirados no funcionamento do cérebro humano. Essas redes recebem dados de entrada e são treinadas para reconhecer padrões, permitindo a previsão de resultados para novos dados (HAYKIN, 1994). Uma RNA é composta por uma camada de entrada (*input*), uma ou mais camadas ocultas (*hidden layers*) e uma camada de saída (*output*). As conexões entre os neurônios dessas camadas possuem pesos associados, que determinam a influência de cada entrada no processamento da rede. Os valores de entrada são

multiplicados pelos respectivos pesos, somados, e o resultado é passado por uma função de ativação, que introduz não linearidade ao modelo; ver Figura 1.

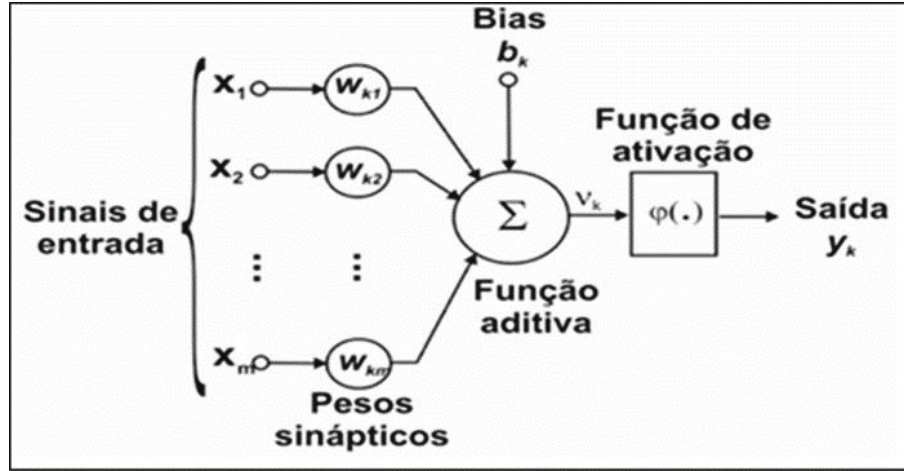


Figura 1: Arquitetura de uma Rede Neural adaptada de Haykin (1994)

Matematicamente, o processamento de uma rede neural pode ser descrito pelas equações:

$$v_k = \sum_{j=0}^m w_{kj} x_j,$$

$$y_k = \varphi(v_k),$$

em que  $x_0, x_1, x_2, \dots, x_m$  os valores de entrada, com  $x_0 = 1$ , e  $w_{k0}, w_{k1}, w_{k2}, \dots, w_{km}$  os pesos sinápticos correspondentes, em que  $w_{k0} = b_k$  representa o viés,  $\varphi$  é a função de ativação, e  $y_k$  é a saída estimada pelo neurônio da camada seguinte.

## 2.3 NNAR

NNARs são modelos híbridos que combinam a lógica dos modelos tradicionais de séries temporais com a flexibilidade não-linear das redes neurais artificiais. Enquanto os modelos autorregressivos convencionais assumem uma relação linear entre as observações passadas e o valor atual da série, as NNAR permitem capturar relações não-lineares e complexas entre essas variáveis. A ideia central consiste em alimentar uma rede neural com as defasagens (lags) da própria série como variáveis de entrada, de modo que a rede aprenda a função não-linear que relaciona essas observações passadas ao valor futuro da série.

Matematicamente, o modelo pode ser representado da seguinte forma:

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}) + \epsilon_t,$$

sendo  $y_t$ , o valor atual da série temporal,  $y_{t-1}, y_{t-2}, \dots, y_{t-p}$ , defasagens,  $f$ , representa a função não-linear e  $\epsilon_t$ , o termo de erro.

Nos modelos sazonais, a estrutura ajustada pode ser chamada de NNAR(p,P,k)[m], onde  $p$  representa o número de lags não sazonais usados como entrada,  $P$  o número de lags sazonais,  $k$  o número de neurônios na camada oculta e  $m$  o período da sazonalidade. Este modelo é análogo a um ARIMA(p,0,0)(P,0,0)[m], mas com a vantagem de permitir capturar relações não lineares nos dados por meio das funções de ativação da rede (HYNDMAN; ATHANASOPOULOS, 2018). Para dados sazonais, o modelo utiliza defasagens tanto de curto prazo quanto sazonais como variáveis explicativas, oferecendo maior flexibilidade na modelagem de padrões complexos.

### 3 Detalhes de implementação e dos dados

Este capítulo detalha a arquitetura metodológica e os recursos de dados empregados para a previsão do IPCA. Inicia-se com a apresentação da técnica de regularização LASSO para seleção de covariáveis em ambientes com alta dimensionalidade. Em seguida, descreve-se o procedimento de validação cruzada por janela deslizante, essencial para avaliar o desempenho de modelos de previsão. Subsequentemente, apresenta-se detalhes sobre o conjunto de dados macroeconômicos que compõem as covariáveis e, por fim, especificamos os detalhes da implementação computacional, incluindo o software, os pacotes e os horizontes de previsão adotados no estudo.

#### 3.1 Regularização LASSO

Em conjuntos de dados que contêm um grande número de variáveis explicativas, nem sempre todas são relevantes para explicar ou prever a variável de interesse. Nesse cenário, técnicas de seleção de variáveis buscam identificar um subconjunto de preditores que contribua de forma significativa para o modelo, melhorando sua interpretação, reduzindo o risco de sobreajuste e, consequentemente, aumentando seu desempenho preditivo.

O LASSO (Least Absolute Shrinkage and Selection Operator), proposto por Tibshirani (1996), trata-se de um método de regularização que modifica o processo tradicional de estimação dos coeficientes da regressão linear, incorporando uma penalização baseada na soma dos valores absolutos dos coeficientes. Em uma regressão linear tradicional, os coeficientes  $\beta$  são estimados buscando minimizar a soma dos quadrados dos resíduos:

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2,$$

no LASSO, essa função objetivo é modificada, passando a incluir um termo de penalização:

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p |\beta_j|,$$

em que  $y_i$  representa a variável resposta,  $x_{ij}$  são as variáveis preditoras,  $\beta_j$  são os coeficientes da regressão a serem estimados e  $\lambda$  é o parâmetro de regularização, que controla a intensidade da penalização aplicada. Para valores pequenos de  $\lambda$ , o resultado obtido se aproxima de uma regressão comum, devido à pouca penalização. À medida que  $\lambda$  aumenta, o número de coeficientes reduzidos a zero cresce, tornando o modelo mais sim-

ples e com menos variáveis. Dessa forma, método tem a capacidade de reduzir alguns coeficientes a zero, realizando assim a seleção automática de variáveis.

### **3.2 Validação Cruzada**

Para fins de comparação, avaliação de desempenho e capacidade de generalização dos modelos para novos dados, utiliza-se a técnica de validação cruzada.

Como o objeto de estudo envolve séries temporais, é essencial preservar a ordem cronológica dos dados durante o processo de validação. Nesse contexto, emprega-se o método de janela deslizante com origem fixa. Inicialmente, define-se um período de treinamento e, a cada iteração, o conjunto de treinamento é expandido com a inclusão de novas observações, sem a exclusão das mais antigas. O modelo é ajustado com o conjunto atualizado e, em seguida, realiza-se a previsão para o próximo ponto (ou período) da série. Esse processo é repetido continuamente, permitindo que, a cada nova iteração, uma previsão seja gerada e comparada com o valor observado. Ao final, os erros acumulados ao longo das iterações possibilitam o cálculo de métricas de avaliação do desempenho preditivo do modelo (HYNDMAN; ATHANASOPOULOS, 2018).

A Figura 2 ilustra o funcionamento da validação cruzada, na qual cada linha horizontal representa uma iteração do processo de validação. Os pontos azuis indicam as observações utilizadas no treinamento do modelo em cada etapa. A cada nova iteração, o conjunto de treinamento é ampliado com a inclusão de mais uma observação, enquanto os dados anteriores são sempre mantidos. Os pontos laranja representam os pontos de teste, ou seja, as observações imediatamente seguintes ao conjunto de treinamento, sobre as quais o modelo realiza a previsão e posteriormente, tem seu desempenho avaliado. Já os pontos cinzas indicam os dados ainda não utilizados, que serão incorporados nas iterações subsequentes. As setas ao final de cada linha destacam que o processo continua até que todos os dados disponíveis sejam utilizados para a validação. Esse procedimento permite simular um ambiente real de previsão ao longo do tempo, assegurando que as previsões sejam sempre feitas com base exclusivamente em dados passados, respeitando a estrutura temporal da série.

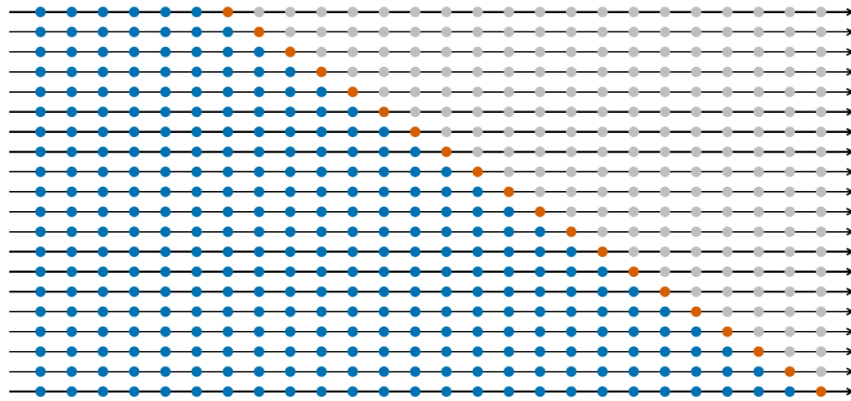


Figura 2: Validação cruzada de Hyndman e Athanasopoulos (2018)

### 3.3 Conjunto de dados

A base de dados utilizada no estudo é composta por 165 observações mensais para cada série temporal, cobrindo o intervalo entre janeiro de 2011 e setembro de 2024. A variável de interesse (número índice do IPCA) foi obtida através do próprio site do IBGE, onde os índices são disponibilizados mensalmente. O IPCA é disponibilizado em diferentes formas: número índice, representa o nível acumulado dos preços em relação ao período-base (no caso,  $\text{dez}/93 = 100$ ); variação mensal (%), quanto os preços variaram em relação ao mês imediatamente anterior; variação acumulada no ano (%), quanto os preços variaram desde o início do ano até o mês em questão; variação em 3, 6 e 12 meses (%), variações acumuladas ao longo desses períodos.

Os dados utilizados como covariáveis (123, no total) consistem em séries temporais de indicadores econômicos, sociais e financeiros provenientes de bases de dados de fontes oficiais e institucionais brasileiras. Entre as instituições fornecedoras desses dados estão: Associação Brasileira de Supermercados (Abras), Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais (AMBIMA), Banco Central do Brasil, Confederação Nacional do Comércio de Bens, Serviços e Turismo (CNC), Companhia Nacional de Abastecimento (Conab), Confederação Nacional da Indústria (CNI), Eletrobras, Federação do Comércio de Bens, Serviços e Turismo do Estado de São Paulo (FecomercioSP), Fundação Getulio Vargas (FGV), Instituto Brasileiro de Geografia e Estatística (IBGE), Instituto de Pesquisa Econômica Aplicada (IPEA), Operador Nacional do Sistema Elétrico (ONS), e Fundação Instituto de Pesquisas Econômicas (FIPE).

### 3.4 Implementação dos modelos

As implementações, ajustes e previsões dos modelos foram realizadas no software estatístico R, versão 4.4.3. Para isso, foram utilizados pacotes como o `forecast` (HYNDMAN; KHANDAKAR, 2008), que oferece a função `auto.arima` para modelagem e previsão de séries temporais via modelos SARIMAX/SARIMA, e também a função `nnfor` (KOURENTZES, 2023), voltada especificamente à aplicação de redes neurais em séries temporais, como a NNAR; e o pacote `keras` (ALLAIRE; CHOLLET, 2024), interface que permite a criação, treinamento e ajuste de redes neurais, particularmente a RNA.

Todas as variáveis são sazonalmente ajustadas (quando necessário) por meio do método X-13ARIMA-SEATS (SAX; EDDELBUETTEL, 2018), e primeiras diferenças são empregadas para tornar as séries estacionárias.

A fim de comparar o desempenho dos modelos em diferentes horizontes de previsão, o método da validação cruzada por janela deslizante com origem fixa foi utilizado. Os modelos foram treinados iterativamente, adicionando uma observação a cada *loop* e realizando a previsão para horizontes de 1 a 12 meses à frente.

Os códigos e os conjuntos de dados utilizados neste estudo estão disponíveis publicamente no repositório GitHub: <https://github.com/clara-franca/previsao-ipca>.



## 4 Resultados das previsões

Este capítulo apresenta e analisa os resultados obtidos a partir dos modelos de previsão do IPCA. A análise começa com uma descrição da série temporal do IPCA. Em seguida, discute-se covariáveis selecionadas pelo método LASSO, que são fundamentais para os modelos SARIMAX, RNA e NNAR. Adiante, são apresentados os resultados da validação cruzada, utilizando as métricas MAE e RMSE para avaliar a acurácia dos modelos SARIMA, SARIMAX, RNA e NNAR em múltiplos horizontes de previsão. Posteriormente, a capacidade preditiva dos modelos é testada em um cenário de extrapolação fora da amostra, onde as previsões para o período de outubro de 2024 a maio de 2025 são comparadas com os valores reais divulgados pelo IBGE. Por fim, o capítulo examina as projeções de longo prazo para os anos de 2025 a 2027, contextualizando-as com as expectativas de mercado.

### 4.1 Análise descritiva

A previsão do IPCA tem como foco o número índice, que variou entre 3.248,2 e 6.997,15 ao longo do período analisado. Essa evolução pode ser observada na Figura 3, que apresenta a trajetória do índice de janeiro de 2011 a maio de 2025. No entanto, duas outras formas mensais, obtidas dos números índices, são consideradas para avaliação: a diferença absoluta em relação ao mês anterior e a variação percentual mensal.

A Figura 4 permite visualizar a evolução da variação mensal do índice ao longo do tempo, destacando possíveis tendências, sazonalidades e oscilações no comportamento dos preços. Essa abordagem facilita a compreensão das flutuações do IPCA e auxilia na interpretação de padrões inflacionários ao longo do período analisado. É possível notar que, durante o período, as maiores quedas do índice foram registradas em maio de 2020 e em julho de 2022.

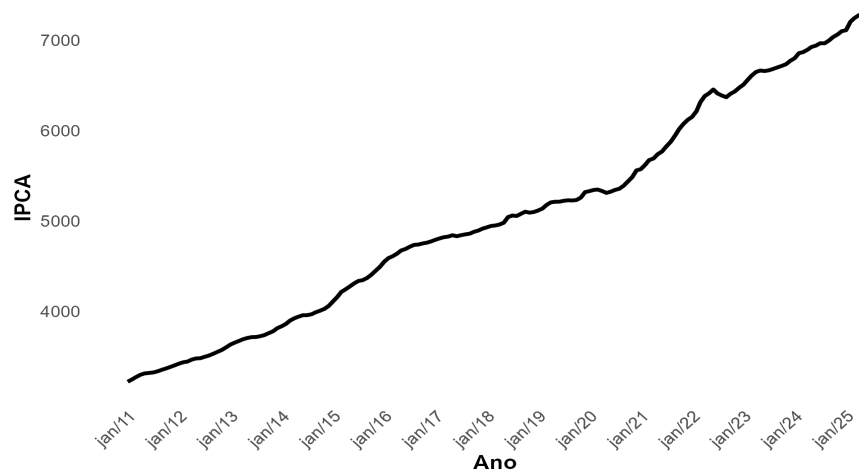


Figura 3: Evolução do IPCA (Jan/2011 - Mai/2025)

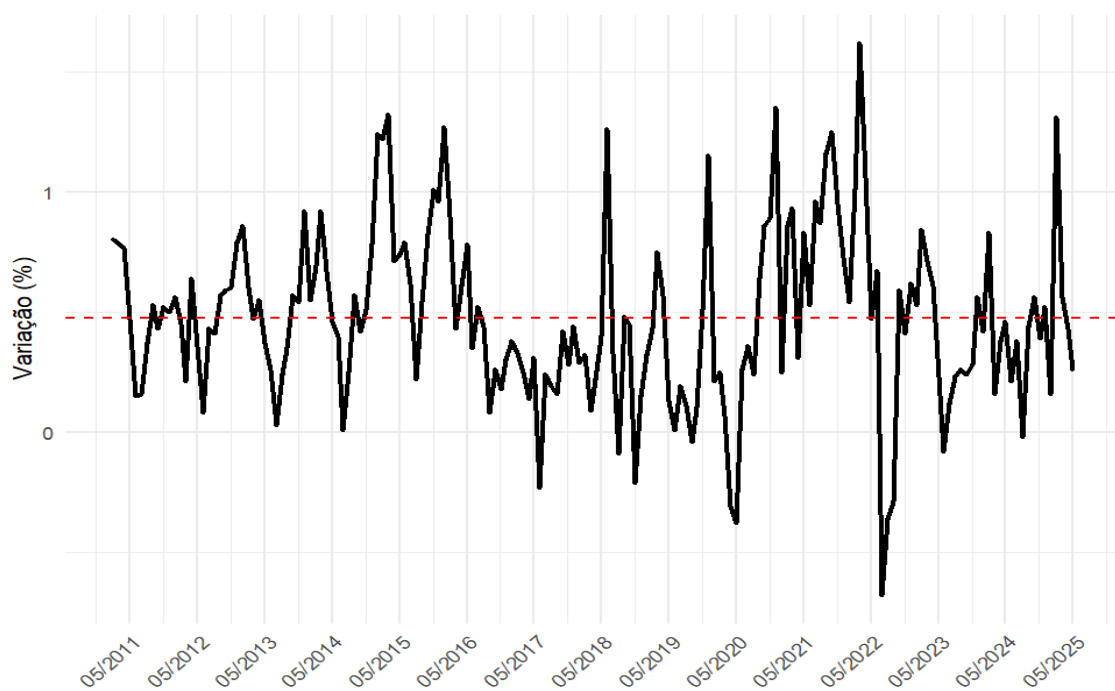


Figura 4: Variação mensal do IPCA (Fevereiro/2011 a maio/2025)

## 4.2 Seleção de covariáveis

As covariáveis selecionadas pelo método LASSO para compor os modelos SARIMAX, RNA e NNAR refletem os principais fatores que influenciam a dinâmica macroeconômica e, portanto, apresentam potencial explicativo relevante para a previsão do comportamento da série em estudo.

As variáveis explicativas foram selecionadas utilizando a função `cv.glmnet()` do pacote `glmnet`, com a série do IPCA como variável dependente e as covariáveis transformadas em matriz de design sem intercepto. O modelo foi ajustado com penalização do tipo LASSO e validado por validação cruzada, que identificou o valor ótimo de penalização,  $\lambda$ , igual a 1,64. As variáveis com coeficientes diferentes de zero nesse ponto foram consideradas as mais relevantes para explicar a dinâmica do IPCA.

Ao todo, foram selecionadas 31 variáveis, as quais estão listadas no Apêndice A.

Entre as covariáveis escolhidas, destacam-se aquelas associadas à política monetária e às condições de crédito, tais como a taxa de juros pré-fixada e diversos indicadores do mercado de crédito, incluindo concessões para pessoa jurídica, saldo da carteira de crédito e taxa média de juros. Essas covariáveis, fornecidas pela ANBIMA e pelo Banco Central do Brasil, são fundamentais para capturar os efeitos da política monetária sobre o consumo, o investimento e, conseqüentemente, sobre a inflação e o produto.

As expectativas de inflação, tanto para os próximos seis quanto para os doze meses, também figuram entre os preditores selecionados, refletindo a relevância das expectativas no processo de formação de preços e nas decisões de política econômica. Essas informações são obtidas por meio do Boletim Focus, publicado pelo Banco Central do Brasil, e representam a percepção dos agentes de mercado sobre a trajetória futura da inflação, sendo amplamente utilizadas como referência para a condução da política monetária.

No âmbito fiscal, foram selecionados indicadores como o déficit nominal e o resultado primário do setor público, além do endividamento do setor público não financeiro. Esses indicadores são essenciais para avaliar o impacto das contas públicas sobre a economia, especialmente em relação às taxas de juros e ao risco-país, indicador que mede a probabilidade de um país não cumprir com suas obrigações financeiras, especialmente em relação ao pagamento de sua dívida externa.

Os preços de produtos agropecuários no atacado, como feijão, soja, bovinos, entre outros, obtidos junto à Seab-PR, refletem pressões de custo no setor primário que podem ser repassadas ao consumidor final, contribuindo para variações na inflação, especialmente no curto prazo, por meio do impacto direto sobre os preços dos alimentos na cesta de consumo das famílias.

O conjunto de covariáveis selecionadas inclui ainda indicadores relacionados ao mercado de trabalho e à massa salarial, que são determinantes da demanda interna e do poder de compra das famílias.

Destacam-se também as covariáveis do setor energético, provenientes da Eletrobras, que contemplam as tarifas médias de energia elétrica no comércio, na indústria e no mercado em geral. Esses indicadores capturam choques de custo relevantes, com impactos diretos sobre os processos produtivos e o consumo das famílias. Além disso, a atividade do varejo é representada por covariáveis derivadas da Pesquisa Mensal de Comércio (PMC), do IBGE, as quais refletem a dinâmica da demanda por bens essenciais, funcionando como um importante termômetro do comportamento do consumo interno.

Além disso, a inclusão da taxa de câmbio efetiva real ressalta a relevância do setor externo e do câmbio como canal de transmissão de choques externos e fator de competitividade da economia.

Por fim, destaca-se a inclusão de uma variável indicadora de crise econômica, que sinaliza a ocorrência de períodos de maior instabilidade no país. Essa variável de natureza qualitativa tem como objetivo capturar efeitos estruturais e choques exógenos associados a cenários adversos, que podem alterar a dinâmica macroeconômica e impactar diretamente o comportamento das demais variáveis do modelo.

Em síntese, a combinação de covariáveis selecionadas é abrangente e coerente com os principais determinantes da inflação no Brasil.

### **4.3 Resultados da validação cruzada**

A comparação de desempenho entre os modelos foi realizada com base em duas métricas de erro: o erro médio absoluto (MAE) e a raiz do erro quadrático médio (RMSE).

As Tabelas 1, 2 e as Figuras 5 e 6 apresentam os valores das métricas MAE e RMSE obtidas por meio da validação cruzada por horizonte de previsão, permitindo avaliar a acurácia de cada método implementado ao longo dos diferentes períodos analisados. Para essa validação, o período inicial utilizado para treinamento foi de 120 observações, o que corresponde a aproximadamente 70% do conjunto de dados, garantindo uma base histórica suficiente para estimar os modelos e realizar previsões em janelas futuras.

Em ambas as métricas, observa-se o fraco desempenho do modelo SARIMA, que, por ser univariado, mostrou limitações na captura da dinâmica da série, sobretudo em horizontes mais longos. Por outro lado, os modelos que incorporam variáveis exógenas ou componentes não lineares apresentaram resultados superiores. O SARIMAX destacou-se especialmente nos primeiros horizontes, alcançando os menores valores de RMSE em 8 dos 12 períodos e de MAE em 3 deles, evidenciando os benefícios da inclusão de variáveis ex-

ternas. Já o modelo NNAR, que combina autorregressão e redes neurais com componentes sazonais, apresentou o menor MAE na maioria dos horizontes (9 de 12) e desempenho competitivo em termos de RMSE, superando os demais modelos em horizontes médios e longos. A Rede Neural tradicional demonstrou resultados estáveis, mas consistentemente inferiores aos de NNAR e SARIMAX, indicando que abordagens que exploram melhor a estrutura temporal dos dados tendem a ser mais eficazes para a previsão multi-horizonte. Assim, conclui-se que a inclusão de variáveis exógenas (SARIMAX) e a modelagem explícita da sazonalidade com não linearidade (NNAR) são estratégias promissoras para a previsão do índice em estudo.

Tabela 1: RMSE dos modelos para diferentes horizontes (meses).

Horizonte	SARIMA	SARIMAX	Rede Neural	NNAR
1	26,826	23,680	30,339	24,120
2	34,383	25,770	30,697	27,432
3	37,513	26,033	30,661	27,858
4	38,417	25,938	28,311	26,183
5	38,852	24,919	27,105	25,502
6	36,637	24,124	28,921	26,367
7	33,952	23,581	28,417	27,579
8	31,316	24,701	25,204	24,096
9	29,668	24,573	27,728	23,624
10	27,867	23,167	24,668	22,770
11	28,399	23,671	25,241	23,244
12	29,693	24,131	24,888	25,731

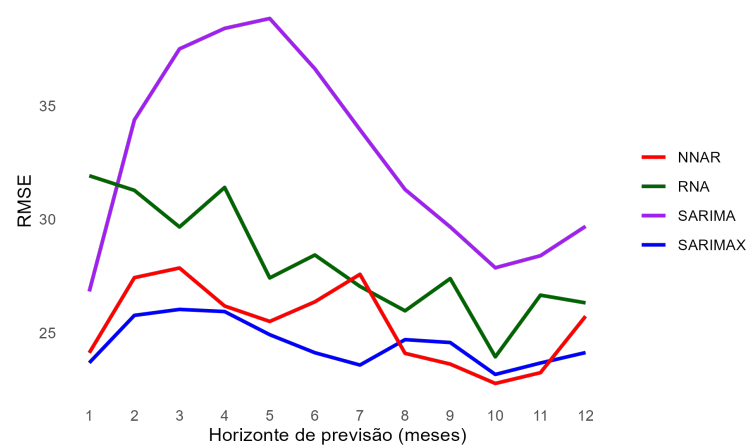


Figura 5: RMSE por horizonte de previsão mensal.

Tabela 2: MAE dos modelos para diferentes horizontes (meses).

Horizonte	SARIMA	SARIMAX	Rede Neural	NNAR
1	18,412	19,547	23,405	17,526
2	25,788	21,316	24,715	18,254
3	27,830	20,775	24,019	19,821
4	28,755	20,002	20,477	18,851
5	30,307	19,444	20,204	17,042
6	29,516	18,532	22,868	19,731
7	27,628	18,096	22,730	20,849
8	24,490	19,204	19,437	17,969
9	22,537	19,081	22,427	17,523
10	20,956	17,578	18,786	16,999
11	21,724	17,936	18,963	17,255
12	22,226	18,434	19,172	19,622

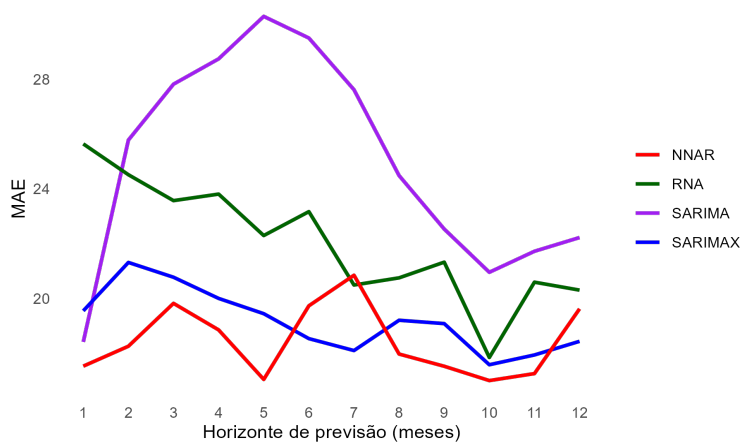


Figura 6: MAE por horizonte de previsão mensal.

#### 4.4 Comparação com valores reais

Uma vez analisado o desempenho dos modelos via validação cruzada, é interessante avaliá-los em um contexto de extrapolação, ou seja, fora da amostra. Uma vez que a base original cobre o intervalo entre janeiro de 2011 e setembro de 2024, é possível fazer previsões fora da amostra, isto é, no período de outubro de 2024 a maio de 2025, e contrastá-las com os valores reais divulgados pelo IBGE. Para essa finalidade, as projeções das covariáveis foram geradas para o período de outubro de 2024 a maio de 2025. A

escolha do modelo mais adequado para cada série, entre as opções SARIMA e ETS, foi determinada por um processo de validação cruzada.

A Tabela 3 e a Figura 7 apresentam as previsões de outubro de 2024 a maio de 2025, confrontando-as com os valores observados. Os resultados confirmam os achados da validação cruzada, destacando a superioridade dos modelos NNAR e SARIMAX. O NNAR demonstrou a maior aderência aos dados reais, seguido de perto pelo SARIMAX. Em contrapartida, os modelos RNA e, principalmente, o SARIMA, apresentaram desvios mais significativos, com este último mostrando uma divergência crescente que evidencia as limitações de uma abordagem univariada para a série em questão.

Tabela 3: Previsões e métricas de erro para o índice (out/2024 a mai/2025).

Período	Índice	NNAR	SARIMAX	RNA	SARIMA
Out/2024	7036,33	7030,051	7034,981	7033,704	7024,480
Nov/2024	7063,77	7071,474	7070,249	7074,061	7049,949
Dez/2024	7100,50	7121,482	7107,845	7111,274	7074,375
Jan/2025	7111,86	7165,581	7142,578	7141,861	7098,217
Fev/2025	7205,03	7212,748	7178,310	7177,511	7121,731
Mar/2025	7245,38	7249,610	7209,209	7211,033	7145,061
Abr/2025	7276,54	7282,040	7238,992	7242,594	7168,289
Mai/2025	7295,46	7320,932	7267,104	7275,689	7191,459
MAE	—	16,36	24,08	24,67	68,92
RMSE	—	19,83	27,27	27,47	79,23

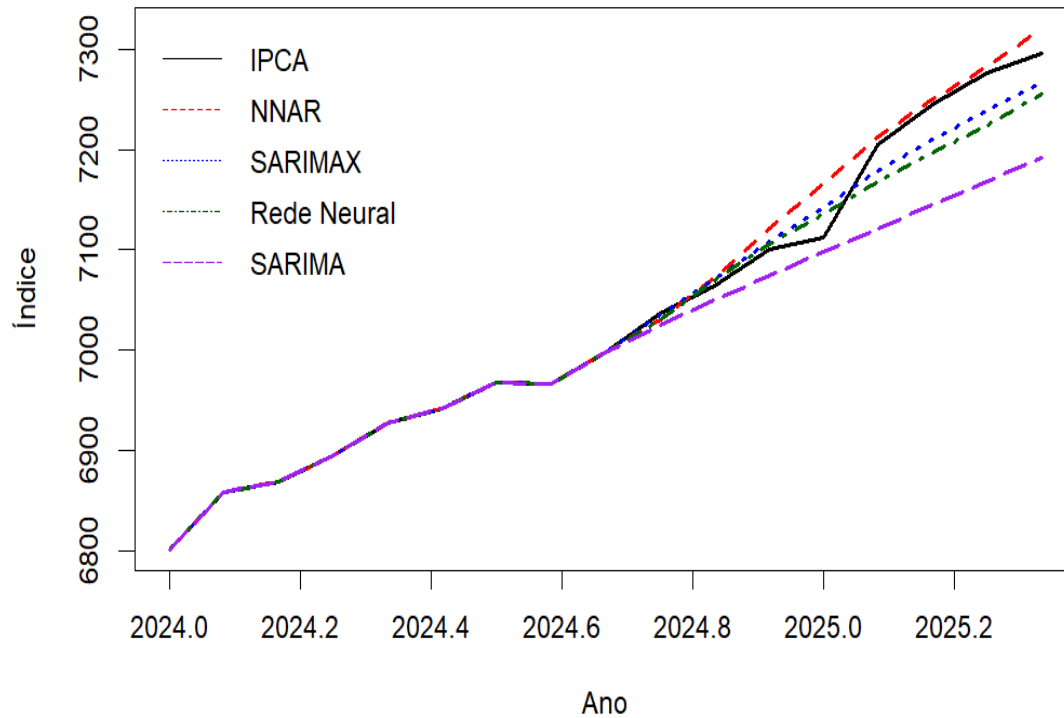


Figura 7: Comparação das previsões com o índice real - out/2024-mai/2025

Outra forma de apresentar as previsões é pela variação acumulada no ano do IPCA. Nesse sentido, a Tabela 4 e a Figura 8 apresentam os resultados das previsões para a variação acumulada do IPCA segundo os modelos considerados. Os resultados reafirmam a hierarquia de desempenho observada anteriormente. Os modelos NNAR e SARIMAX continuam sendo os mais acurados, com o primeiro se destacando. Em contraste, a RNA e, principalmente, o modelo univariado SARIMA, apresentam previsões mais distantes dos valores reais observados.



Tabela 4: Previsões e métricas de erro para a variação acumulada ao ano(out/2024 a mai/2025).

Período	Variação	NNAR	SARIMAX	RNA	SARIMA
Out/2024	3,88	3,79	3,86	3,79	3,71
Nov/2024	4,29	4,40	4,38	4,38	4,08
Dez/2024	4,83	5,14	4,94	4,91	4,45
Jan/2025	0,16	0,62	0,49	0,43	0,34
Fev/2025	1,47	1,28	0,99	0,87	0,67
Mar/2025	2,04	1,80	1,43	1,28	1,00
Abr/2025	2,48	2,25	1,85	1,67	1,33
Mai/2025	2,75	2,80	2,24	2,10	1,66
MAE	—	0,21	0,35	0,42	0,63
RMSE	—	0,24	0,42	0,51	0,75

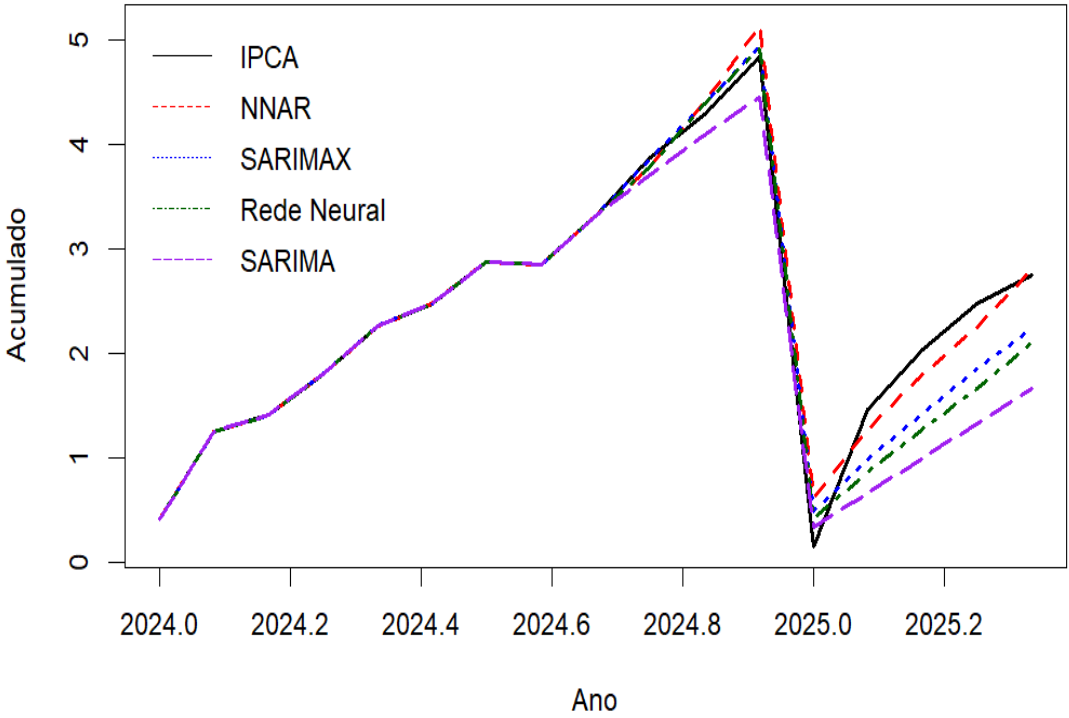


Figura 8: Comparação das previsões para a variação acumulada ao ano - out/2024-mai/2025.

## 4.5 Previsões para 2025–2027

Além da avaliação do desempenho histórico e recente dos modelos, é relevante examinar as tendências apontadas para os próximos anos em termos de inflação. As projeções das covariáveis foram geradas para o período de outubro de 2024 a dezembro de 2027, e a escolha do modelo mais adequado para cada série, entre as opções SARIMA e ETS, foi determinada por um processo de validação cruzada.

A Tabela 5 apresenta as previsões acumuladas para o IPCA nos anos de 2025 a 2027, conforme estimativas geradas pelos modelos considerados neste estudo. Observa-se que, para dezembro de 2025, a projeção média dos modelos é de aproximadamente 5%, com destaque para o modelo NNAR, que projeta um valor mais elevado (6,46%), enquanto o SARIMA apresenta a menor estimativa (3,94%). Para 2026 e 2027, as projeções médias convergem para 4,43% e 4,32%, respectivamente, sinalizando uma trajetória de melhora da inflação ao longo do período.

As projeções da Tabela 5 podem ser comparadas às expectativas do *Relatório Focus* do Banco Central do Brasil de 27 de junho de 2025 (<https://www.bcb.gov.br/content/focus/focus/R20250627.pdf>), que indicam taxas esperadas de 5,20% para 2025, 4,50% para 2026 e 4,00% para 2027. Assim, de maneira geral, há uma boa convergência entre as previsões dos modelos e as expectativas do mercado. Apesar dessa proximidade nas médias, é possível observar alguma heterogeneidade entre os modelos individuais. Por exemplo, o modelo NNAR projeta valores consistentemente mais altos ao longo do período, enquanto o SARIMA tende a gerar estimativas mais conservadoras. O uso de dados mais recentes, isto é, de outubro de 2024 a maio de 2025, pode ajudar a melhorar as projeções. No entanto, esse exercício de extrapolação mostra a capacidade dos modelos considerados em fornecer projeções em linha com as expectativas do mercado.

Tabela 5: Previsões dos modelos para a variação acumulada do IPCA (dezembro de 2025 a 2027).

Período	NNAR	SARIMAX	RNA	SARIMA	Média
Dez/2025	6,46	4,77	4,85	3,94	5,00
Dez/2026	5,61	4,25	4,10	3,77	4,43
Dez/2027	5,27	4,21	4,18	3,63	4,32

## 5 Considerações finais

Este trabalho teve como objetivo principal comparar a performance de diferentes modelos na previsão do Índice Nacional de Preços ao Consumidor Amplo (IPCA). Para isso, foram aplicadas abordagens clássicas de séries temporais, como o SARIMA, sua extensão com variáveis exógenas (SARIMAX), e métodos de aprendizado profundo, como Redes Neurais Artificiais (RNA) e Redes Neurais Autorregressivas (NNAR). Os resultados, obtidos via validação cruzada e comparação com dados reais, evidenciaram que os modelos que incorporam covariáveis, notadamente o SARIMAX e o NNAR, apresentaram desempenho superior. As projeções de longo prazo indicaram uma inflação de aproximadamente 5% para 2025, com uma tendência de desaceleração para pouco mais de 4% em 2026 e 2027, valores que convergem com as expectativas de mercado. Para estudos futuros, recomenda-se a investigação de arquiteturas baseadas em aprendizado profundo (deep learning), bem como a aplicação de outros métodos tradicionais de modelagem de séries temporais. Além disso, a implementação de técnicas de seleção dinâmica de covariáveis ao longo do tempo pode oferecer ganhos em termos de desempenho preditivo e interpretação dos modelos.

## Referências

- ALLAIRE, J.; CHOLLET, F. *keras: R Interface to 'Keras'*. [S.l.], 2024. R package version 2.15.0. Disponível em: <https://CRAN.R-project.org/package=keras>.
- ARAÚJO, G. S.; GAGLIANONE, W. P. Machine learning methods for inflation forecasting in brazil: New contenders versus classical models. *Latin American Journal of Central Banking*, v. 4, n. 2, p. 100087, 2023. ISSN 2666-1438. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2666143823000042>.
- BCB. *Relatório Focus*. 2025. Disponível em: <https://www.bcb.gov.br/controleinflacao/relatoriofocus>.
- BOX, G. E. et al. *Time series analysis: forecasting and control*. [S.l.]: John Wiley & Sons, 2015.
- CHU, B.; QURESHI, S. Comparing out-of-sample performance of machine learning methods to forecast us gdp growth. *Computational Economics*, Springer, v. 62, n. 4, p. 1567–1609, 2023.
- DORNBUSCH, R.; FISCHER, S.; STARTZ, R. *Macroeconomia - 11ed*. Bookman Editora, 2013. ISBN 9788580551853. Disponível em: <https://books.google.com.br/books?id=MB44AgAAQBAJ>.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. <http://www.deeplearningbook.org>.
- HAYKIN, S. *Neural networks: a comprehensive foundation*. [S.l.]: Prentice Hall PTR, 1994.
- HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting: principles and practice*. [S.l.]: OTexts, 2018.
- HYNDMAN, R. J.; KHANDAKAR, Y. Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, v. 27, n. 3, p. 1–22, 2008.
- IBGE. *IPCA - Índice Nacional de Preços ao Consumidor Amplo*. 2024. <https://metadados.ibge.gov.br/consulta/estatisticos/operacoes-estatisticas/IA>. Accessed: 2024-12-29.
- KOURENTZES, N. *nnfor: Time Series Forecasting with Neural Networks*. [S.l.], 2023. R package version 0.9.9. Disponível em: <https://CRAN.R-project.org/package=nnfor>.
- SANTOS, F. L.; NOLAU, I. Td 2814 - previsão de inflação : análise preliminar de desempenho de técnicas de machine learning. *Texto para Discussão*, Instituto de Pesquisa Economica Aplicada - IPEA, p. 1–31, 11 2022.
- SAX, C.; EDDERBUETTEL, D. Seasonal adjustment by x-13arima-seats in r. *Journal of Statistical Software*, v. 87, n. 11, p. 1–17, 2018. Disponível em: <https://doi.org/10.18637/jss.v087.i11>.

TARSITANO, A.; AMERISE, I. L. Short-term load forecasting using a two-stage sarimax model. *Energy*, v. 133, p. 108–114, 2017. ISSN 0360-5442. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0360544217308848>.

TIBSHIRANI, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, v. 58, n. 1, p. 267–288, 12 1996. ISSN 0035-9246. Disponível em: <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>.

## Apêndice

### A Variáveis selecionadas pelo LASSO

Tabela 6: Variáveis selecionadas pelo LASSO

Variável	Descrição	Fonte
ANBIMA12_TJTLN112	Taxa de juros pré fixada - estrutura a termo - LTN - 1 mês	Anbima
BM12_CCAPJ12	Operações de crédito - concessões - pessoa jurídica	Bacen/Notas Im- prensa/Moeda
BM12_CRDCAPJ12	Operações de crédito - concessões - recursos direcionados - pessoa jurídica	Bacen/Notas Im- prensa/Moeda
BM12_CS12	Operações de crédito - saldo da carteira de crédito - total	Bacen/Notas Im- prensa/Moeda
BM12_CTJ12	Operações de crédito - taxa média de juros - total	Bacen/Notas Im- prensa/Moeda
BM12_DTSPY12	DSLP - total - setor público não-financeiro	Bacen/Notas Im- prensa/F. Púb.
BM12_IPCAEXP1212	Expectativa média de Inflação - IPCA - taxa acumulada para os próximos doze meses	Bacen/Boletim/Ativ. Ec.
BM12_IPCAEXP612	Expectativa média de Inflação - IPCA - taxa anualizada para os próximos seis meses	Bacen/Boletim/Ativ. Ec.
BM12_M0NY12	Base monetária restrita - M0 - fim de período	Bacen/Notas Im- prensa/Moeda
BM12_NFSPJNYS12	NFSP - setor público - juros nominais - sem desvalorização cambial - fluxo acumulado em 12meses	Bacen/Notas Im- prensa/F. Púb.
BM12_NFSPPYS12	NFSP - setor público - resultado primário - sem desvalorização cambial - fluxo acumulado em 12 meses	Bacen/Notas Im- prensa/F. Púb.
CNI12_REMN12	Indicadores Industriais - massa salarial - indústria - índice real (média 2006 = 100)	CNI

Variável	Descrição	Fonte
CRISE_A	Indicador binário de ocorrência de crise econômica	IPEA
DERAL12_ATBCAT12	Preço médio - atacado - bovinos - carcaça traseira - kg - PR	Seab-PR
DERAL12_ATFEC12	Preço médio - atacado - feijão de cor - 30 kg - PR	Seab-PR
DERAL12_ATFRR12	Preço médio - atacado - frango resfriado - kg - PR	Seab-PR
DERAL12_ATFSO12	Preço médio - atacado - soja (em farelo) - tonelada - PR	Seab-PR
DERAL12_ATOLR12	Preço médio - atacado - óleo de soja - refinado - 20 latas - PR	Seab-PR
DERAL12_PRCCO12	Preço médio - produtor - café - kg - PR	Seab-PR
DERAL12_PRFRV12	Preço médio - produtor - frutas e verduras - kg - PR	Seab-PR
DERAL12_PRLECO12	Preço médio - produtor - leite de vaca - litro - PR	Seab-PR
DERAL12_PRSO12	Preço médio - produtor - soja em grão - 60 kg - PR	Seab-PR
ELETRO12_CEETCOM12	Energia elétrica - consumo - comércio - tarifa média por MWh	Eletrobras
ELETRO12_CEETIND12	Energia elétrica - consumo - indústria - tarifa média por MWh	Eletrobras
ELETRO12_CEETT12	Energia elétrica - consumo - tarifa média por MWh	Eletrobras
GAC12_TCERXTINPC12	Taxa de câmbio - efetiva real - INPC - exportações - índice (média 2010 = 100)	IPEA
PMC12_VRFARMN12	Vendas reais - varejo - artigos farmacêuticos, médicos, ortopédicos, de perfumaria e cosméticos - índice (média 2022 = 100)	IBGE/PMC
PNADC12_ESPUBCC12	Pessoas ocupadas - empregado no setor público com carteira assinada (inclusive servidor estatutário e militar)	IBGE/PNAD Contínua

Variável	Descrição	Fonte
PNADC12.MRRTH12	Massa de rendimento real de todos os trabalhos habituais mensais	IBGE/PNAD Contínua
PNADC12.MRTEM12	Massa de rendimento de todos os trabalhos efetivos mensais - mensalizada	IPEA
PNADC12.NOCUP12	Nível da ocupação	IBGE/PNAD Contínua