

Tópicos especiais de previsão de inflação com aprendizado de máquina

Este boxe apresenta uma discussão metodológica sobre o uso de técnicas de aprendizado de máquina (ML) na previsão da inflação brasileira no Banco Central do Brasil (BC), destacando sua capacidade de lidar com bases de dados extensas, não linearidades e informações irrelevantes. Com base em Araujo e Gaglianone (2023), o estudo amplia o horizonte de previsão para até 24 meses, atualiza a base de dados macroeconômica e financeira e utiliza o algoritmo EM para tratar dados faltantes. São empregados diversos métodos de ML e modelos econométricos tradicionais, combinados de forma a minimizar o erro quadrático médio em exercícios pseudo fora da amostra. Os resultados mostram que as previsões com ML são competitivas em relação às expectativas do Focus, especialmente nos curto e médio prazos, e que a combinação de modelos e a inclusão da inflação implícita aumentam o poder preditivo, reforçando a utilidade desse arcabouço no monitoramento e na análise prospectiva da inflação no Brasil.

Previsões de inflação são essenciais para decisões econômicas, e particularmente desafiadoras em economias emergentes. Neste contexto, métodos de aprendizado de máquina (*machine learning*, ML) oferecem uma alternativa promissora por capturarem padrões não lineares e filtrarem informações irrelevantes. Este boxe tem como objetivo apresentar uma breve discussão metodológica de tópicos especiais na implementação de previsões de inflação com técnicas de ML no BC.¹

A metodologia baseia-se no estudo de Araujo e Gaglianone (2023), que investiga o desempenho de uma ampla gama de métodos de ML e modelos econométricos tradicionais na previsão da inflação brasileira. As principais inovações deste boxe em relação ao referido estudo são: (i) uso sistemático de um algoritmo EM (*Expectation Maximization*) para tratamento de dados recentes faltantes (*ragged edge*); (ii) revisão e atualização da base de dados macroeconômica e financeira;² (iii) extensão do horizonte de previsão até 24 meses à frente; e (iv) utilização de ML na construção de previsões de inflação desagregada, com foco nos três principais grupos da inflação de preços livres do Índice Nacional de Preços ao Consumidor Amplo (IPCA) (alimentação no domicílio, bens industriais e serviços).

Na construção da previsão de inflação com ML utilizam-se diferentes classes de modelos e métodos de previsão. Entre os métodos de aprendizado de máquina destacam-se: regressões penalizadas (*ridge*, *lasso*, *adaptive lasso* e *elastic net*), métodos baseados em árvores (*random forest*, *quantile regression forest* e *XGBoost*) e redes neurais recorrentes (RNN). Modelos econométricos tradicionais, tais como Vetor autorregressivo (VAR), modelos de fatores e curvas de Phillips, além de séries de inflação implícita e diversas combinações de modelos, também são considerados. Para maiores detalhes, vide Araujo e Gaglianone (2023).

A maior parte das previsões é construída com a abordagem *direct forecast*, na qual estima-se cada um dos modelos separadamente para cada horizonte de previsão.³ A seleção dos melhores métodos em cada horizonte é feita com base no Erro Quadrático Médio (EQM), sendo que a trajetória final de ML é dada pela média das três melhores previsões, de acordo com um exercício recursivo de previsão pseudo fora da amostra.⁴ Para horizontes superiores a 12 meses, na qual predomina a média histórica da inflação (comportamento típico de

1/ Este boxe detalha procedimentos e ferramentas de modelos satélites que integram o ciclo de análises e projeções do Copom, conforme descrito em BC (2023, p.61).

2/ Como a coleta de dados de diferentes fontes é um processo dinâmico, algumas variáveis foram acrescentadas e outras removidas da base de dados original (e.g., no caso de séries descontinuadas, que não estão mais sendo atualizadas pela fonte).

3/ Exceções são os modelos iterados do tipo VAR e ARMA, que tem uma estimação única para todos os horizontes de previsão.

4/ A previsão de inflação com técnicas de ML discutida neste boxe não usa nenhuma informação das expectativas Focus, com o objetivo de gerar uma previsão alternativa ao consenso de mercado, construída apenas com dados macroeconômicos e financeiros, e combinando métodos recentes de ML com modelos econométricos tradicionais de séries temporais.

modelos de séries temporais), utiliza-se uma combinação ponderada do tipo CLS (*Constrained Least Squares*) de métodos pré-selecionados.⁵

Neste contexto, além da previsão da inflação do IPCA, pode-se construir a previsão da inflação de preços livres do IPCA por duas abordagens: (i) diretamente, considerando esta variável como variável dependente em cada modelo; ou (ii) a partir de previsões desagregadas dos três principais grupos que constituem os preços livres, ponderando-se tais previsões pelos pesos correspondentes de cada componente na estrutura do índice.⁶

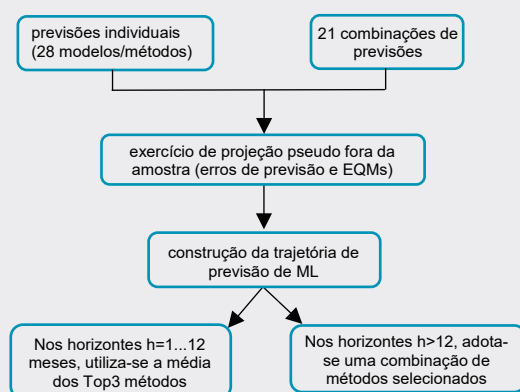
A base de dados utilizada inclui 137 séries macroeconômicas e financeiras, na frequência mensal, desde janeiro/2004 até outubro/2025,⁷ além de séries de inflação implícita construídas a partir de dados de títulos públicos federais.⁸ As séries são transformadas individualmente para apresentarem estacionariedade (conforme o teste KPSS), sendo os dados faltantes preenchidos via filtro de Kalman, por meio de uma versão do algoritmo EM capaz de lidar com bases de dados com centenas de variáveis.⁹

Desta forma, as principais etapas para a geração de previsões com ML envolvem:

- a) *Pré-processamento de dados*: com atualização da base de dados a partir de diferentes fontes, preenchimento de dados faltantes com o algoritmo EM, transformação das séries temporais para garantir estacionariedade (e.g., primeira diferença da série);
- b) *Processamento/estimação de modelos*: estimação de modelos individuais e dos pesos de combinações de modelos, construção das previsões de inflação dentro e fora da amostra, para horizontes até 24 meses, cálculo dos erros de previsão e dos erros quadráticos médios em cada horizonte; e
- c) *Pós-processamento dos resultados*: identificação dos melhores métodos, construção da trajetória de ML fora da amostra, cálculo das previsões de inflação acumulada em doze meses e comparação final com as respectivas medianas das expectativas do Focus.

O Gráfico 1 apresenta um fluxograma com os principais passos para a construção de uma trajetória final de previsão de inflação com ML, utilizando os 50 métodos de previsão discutidos em Araujo e Gaglianone (2023).¹⁰

Gráfico 1 – Exercício de Previsão de Inflação com ML



5/ Tais como: ARMA, *Random Walk* (RW-AO), *Quantile Regression Forest* (QRF), BEI (*Breakeven Inflation*) e Curva de Phillips híbrida. Vide Araujo e Gaglianone (2023) para maiores detalhes.

6/ Uma das vantagens de se construir trajetórias de inflação desagregada é a capacidade de antecipar reversões de tendência em períodos de choques específicos (como, por exemplo, em alimentos ou serviços).

7/ O conjunto de informação deste estudo considera a data-base de 28/11/2025.

8/ As inflações implícitas são construídas com a metodologia descrita em Val e Araujo (2019).

9/ Schafer (1997) propõe uma extensão do algoritmo EM para dados Gaussianos multivariados. Para maiores detalhes, vide também Little e Rubin (2002).

10/ Com exceção do Focus, que não faz parte do conjunto de métodos de previsão deste box. Tal fonte de expectativas é usada neste estudo apenas para fins comparativos.

Cabe ressaltar que a previsão de ML gerada conforme o fluxograma acima é condicional a um conjunto de informação disponível numa data-base específica (com séries observadas até essa data ou preenchidas via algoritmo EM) e não condicional a cenários, como ocorre geralmente em modelos macroeconômicos estruturais ou semiestruturais.¹¹

A Tabela 1 apresenta as variáveis mais relevantes na previsão de inflação mensal do IPCA, identificadas conforme métodos selecionados de ML para a data-base em questão. De modo geral, os resultados mostram que, para horizontes curtos, a dinâmica inercial da inflação permanece central, com destaque para medidas defasadas de inflação, e variáveis relacionadas à atividade econômica (tais como: consumo de energia elétrica e exportações agrícolas), selecionadas de forma recorrente pelos diversos métodos de ML considerados.

À medida que o horizonte de previsão se alonga, observa-se certa diversificação do conjunto de variáveis relevantes para a previsão de inflação, com maior presença de indicadores fiscais, financeiros e do setor externo. Desta forma, a combinação entre informações inerciais da inflação, condições macroeconômicas domésticas e choques externos mostra-se relevante para explicar a trajetória da inflação nos diferentes horizontes analisados neste estudo.

Tabela 1 – Variáveis relevantes para previsão de inflação com métodos selecionados

Horizonte de previsão (meses)	Top5 variáveis de métodos selecionados de ML
1	Commodity BR, Commodity Metal BR, Núcleo IPCA DW, eletricidade comercial, PIB, juros 1ano, IPC-Fipe, IPCA difusão, IPCA, IPCA livres, juros real 1, 2 e 5 anos
2	Commodity Metal BR, Núcleo IPCA DW, eletricidade comercial, taxa de câmbio, PIB, juros 1 e 2 anos, IPC-Fipe, IPCA difusão, IPCA bens industriais, IPCA livres, juros real 1 e 2 anos, VIX
3	Commodity Agricultura BR, eletricidade comercial, exportação agricultura, PIB, importação outros, juros 1 e 2 anos, IPCA livres, juros real 2 e 5 anos, VIX
6	Eletricidade comercial, preços de exportação, INCC, prod. ind. não duráveis, juros 1, 2 e 5 anos, IPCA serviços, resultado primário/PIB, juros real 1 e 2 anos, US Treasury 3m
9	Utilização de capacidade da indústria, eletricidade total, PIB, IBC-BR, IGP-DI, IGP-M, INCC, Selic, IPC-BR, juros real 1, 2 e 5 anos
12	CRB, CRB alimentos, eletricidade comercial, PIB, IGP-10, IGP-DI, IGP-M, IPCA livres, IPCA não comercializáveis, resultado primário/PIB, juros real 1, 2 e 5 anos, depósitos de poupança, US Treasury 3m

Obs.: A lista de variáveis selecionadas em cada horizonte, considerando a data-base 28/11/2025, mostra as Top5 séries dos seguintes métodos: lasso, lasso adaptativo, ridge, elastic net, random forest, xgboost.

A Tabela 2 ilustra a seleção de métodos de previsão de inflação mensal do IPCA em determinados horizontes (h), conforme os respectivos erros quadráticos médios. Cabe destacar o bom desempenho da inflação implícita (BEI, *breakeven inflation*) escolhida como método Top1 para h=1 mês, refletindo a qualidade do seu conteúdo informacional, além da existência de inúmeras combinações de modelos e abordagens não lineares dentre os melhores modelos, em linha com os resultados de Araujo e Gaglianone (2023).

Tabela 2 – Melhores métodos de previsão da inflação do IPCA (% a.m.)

Horizonte de previsão (meses)	Top1	Top2	Top3
1	BEI	Comb2 CLS	Comb2 Adalasso
2	Comb1 Adalasso	Comb2 Adalasso	Comb2 GR
3	Comb2 Média	Comb2 Mediana	Comb1 Adalasso
6	BEI	Comb2 Média	Comb2 Mediana
9	FM3	Comb2 Média	Comb2 Mediana
12	RF	QRF	Disag RF

Obs.: As siglas indicam os métodos de previsão: BEI (Breakeven Inflation), CLS (Constrained Least Squares), Adalasso (Lasso Adaptativo), GR (Granger-Ramanathan), FM3 (Factor Model 3), RF (Random Forest), QRF (Quantile Regression Forest), Disag RF (Random Forest desagregado). Comb1 é uma combinação de previsões dos modelos 1 a 28 da Tabela 1 de Araujo e Gaglianone (2023). Comb2 é uma combinação via model confidence set de Hansen et al. (2011), usando os modelos 1 a 28 da mesma tabela. Data-base considerada: 28/11/2025.

11/ A elaboração de um cenário multivariado para previsão condicional num contexto de *big data* não é uma tarefa trivial, e tal prática geralmente produz mais ruído do que sinal em termos de ganho de acurácia, principalmente em horizontes de previsão mais longos. Além disso, o prêmio de risco, em geral, é maior para a inflação implícita em horizontes longos, além da convergência à média incondicional da inflação nos modelos de séries temporais que compõe a previsão de ML.

A Tabela 3 compara o erro quadrático médio da previsão do Focus (mediana das expectativas) com as previsões de ML, incluindo ou não a inflação implícita (BEI) no rol de previsões. Os resultados confirmam que ML é uma abordagem competitiva em termos de capacidade preditiva no curto/médio prazos. Em particular, a inclusão da inflação implícita na composição da previsão de ML aumenta substancialmente a capacidade preditiva de ML em horizontes curtos. Desta forma, o nicho da previsão de inflação mensal com ML situa-se entre 2-3 até 9-12 meses, pois horizontes muito curtos são, geralmente, dominados pela previsão de especialistas (e.g., via julgamento ou coleta direta de preços), enquanto que horizontes mais longos (e.g., acima de 1 ano) não costumam ser informativos, com previsões se aproximando da média incondicional da inflação.

Tabela 3 – Erro Quadrático Médio (EQM) na previsão da inflação do IPCA (% a.m.)

Horizonte de previsão (meses)	Focus	ML	ML (sem BEI)
1	0,115	0,116	0,293
2	0,269	0,243	0,321
3	0,342	0,330	0,359
6	0,366	0,363	0,369
9	0,374	0,370	0,370
12	0,379	0,374	0,374
18	0,385	0,402	0,411
24	0,390	0,423	0,416

Obs.: Focus indica a mediana das expectativas de inflação, ML a previsão de inflação de Machine Learning, e BEI (Breakeven Inflation, inflação implícita). Data-base considerada: 28/11/2025.

Em suma, a previsão de inflação com métodos de aprendizado de máquina permite combinar informações de uma ampla base de dados com a flexibilidade e robustez de algoritmos modernos. Os resultados deste boxe ilustram como a combinação de previsões entre modelos distintos pode contribuir para o acompanhamento da inflação no Brasil.¹²

A integração desse arcabouço no processo de elaboração de projeções do BC tem se mostrado útil no monitoramento tempestivo de pressões inflacionárias em diferentes segmentos do IPCA. Essa iniciativa reforça o contínuo esforço de aprimoramento das ferramentas quantitativas utilizadas pelo Banco Central do Brasil.

Referências

ARAUJO, G.S., GAGLIANONE, W.P. (2023). "Machine Learning Methods for Inflation Forecasting in Brazil: new contenders versus classical models". *Latin American Journal of Central Banking* 4(2), 100087.

BANCO CENTRAL DO BRASIL (2023). "Sistema de análise e projeções do BC". Relatório de Inflação, Março/2023.

HANSEN, P.R., LUNDE, A., NASON, J.M. (2011). "The Model Confidence Set". *Econometrica* 79(2), 453-497.

LITTLE, R.J.A., RUBIN, D.B. (2002). *Statistical Analysis with Missing Data* (2a Ed.). New York: Wiley.

SCHAFER, J.L. (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall/CRC Press.

VAL, F.F., ARAUJO, G.S. (2019). "Breakeven inflation rate estimation: An alternative approach considering indexation lag and seasonality". Banco Central do Brasil. Trabalho para Discussão n. 493.

12/ As previsões de inflação discutidas neste boxe são meramente ilustrativas da metodologia proposta, e foram geradas por uma abordagem estatística e não estrutural, complementar aos demais modelos econômicos do BC.