

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
FACULDADE DE CIÊNCIAS ECONÔMICAS  
DEPARTAMENTO DE ECONOMIA E RELAÇÕES INTERNACIONAIS**

**GUILHERME DA SILVA FRAINER**

**PREVISÃO DA INFLAÇÃO BRASILEIRA COM MÉTODOS DE APRENDIZADO DE  
MÁQUINA**

**Porto Alegre**

**2024**

**GUILHERME DA SILVA FRAINER**

**PREVISÃO DA INFLAÇÃO BRASILEIRA COM MÉTODOS DE APRENDIZADO DE  
MÁQUINA**

Trabalho de conclusão submetido ao Curso de Graduação em Ciências Econômicas da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título Bacharel em Economia.

Orientador: Prof. Dr. Marcelo Savino Portugal

**Porto Alegre**

**2024**

## CIP - Catalogação na Publicação

Frainer, Guilherme da Silva

Previsão da inflação brasileira com métodos de  
aprendizado de máquina / Guilherme da Silva Frainer.

-- 2024.

64 f.

Orientador: Marcelo Savino Portugal.

Trabalho de conclusão de curso (Graduação) --  
Universidade Federal do Rio Grande do Sul, Faculdade  
de Ciências Econômicas, Curso de Ciências Econômicas,  
Porto Alegre, BR-RS, 2024.

1. Previsão de inflação. 2. Aprendizado de máquina.  
3. Econometria. I. Portugal, Marcelo Savino, orient.  
II. Título.

**GUILHERME DA SILVA FRAINER**

**PREVISÃO DA INFLAÇÃO BRASILEIRA COM MÉTODOS DE APRENDIZADO DE  
MÁQUINA**

Trabalho de conclusão submetido ao Curso de Graduação em Ciências Econômicas da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título Bacharel em Economia.

Aprovada em: Porto Alegre, \_\_\_\_\_ de \_\_\_\_\_ de 2024.

BANCA EXAMINADORA:

---

Prof. Dr. Marcelo Savino Portugal – Orientador  
UFRGS

---

Prof. Dr. Sabino da Silva Porto Júnior  
UFRGS

---

Prof. Dr. Hudson da Silva Torrent  
UFRGS

## **AGRADECIMENTOS**

Agradeço a Deus por me ter provido da perseverança necessária para concluir este trabalho e, de modo mais geral, o curso de Ciências Econômicas.

Agradeço aos meus pais, Mônica e Antônio, por terem imbuído em mim a importância da educação e do esforço. Ademais, agradeço pelo afeto ao longo da minha criação.

Agradeço ao meu orientador, não só pelo auxílio com o trabalho, mas principalmente pelos bons dois anos em que tive o prazer de ser seu bolsista. Acredito que o contato frequente com ele tenha me exposto a conteúdos mais exigentes que os que geralmente vemos na graduação, bem como a alguns orientandos extremamente inteligentes, tornando-me um profissional e acadêmico melhor.

Agradeço a Frederico Sager pela ajuda com os códigos e pela disponibilidade em responder às minhas dúvidas.

Agradeço aos professores da UFRGS que contribuíram para a minha formação. Menciono aqui Álvaro Krüger Ramos, Arthur Maria Ferreira Neto, João Plínio Juchem Neto, Marcelo Savino Portugal e Thaís Waideman Niquito, mas não limito meus agradecimentos a eles.

Agradeço ao contribuinte.

“We are drowning in information and starving for knowledge.” (Rutherford D. Rogers)

## RESUMO

A previsão da inflação é importante para agentes econômicos, como firmas, trabalhadores e investidores, e para as autoridades monetárias. Num contexto dos avanços da inteligência artificial (IA) e da grande disponibilidade de dados (*big data*), este trabalho objetiva testar a capacidade preditiva de alguns modelos de aprendizado de máquina (regressão *ridge*, *LASSO*, *Random Forest* e *Gradient Boosting*) e compará-los ao *ARIMA*. Para tanto, a variável escolhida foi o Índice Nacional de Preços ao Consumidor Amplo (IPCA). Os modelos foram treinados com os dados de janeiro de 2004 a fevereiro de 2017, e testaram-se as suas previsões para o período de março de 2017 a setembro de 2023. Além do IPCA, os modelos foram treinados com outras 40 séries temporais. Calcularam-se algumas métricas de erro para as previsões e compararam-se as previsões dos modelos de aprendizado de máquina às do *ARIMA* por meio do teste de Diebold-Mariano. Como resultado, encontrou-se que os modelos de aprendizado de máquina (com a exceção da regressão *ridge*) apresentaram resultados superiores (isto é, com menor erro) aos do *ARIMA* para quase todos os horizontes temporais. Contudo, os resultados não foram significativamente superiores aos do *ARIMA* para um nível de significância de 5% para o teste de Diebold-Mariano na maioria dos casos. Com isso, concluiu-se que, apesar de modelos de IA obterem bons resultados, ainda não é possível descartar o *ARIMA* num contexto de previsão de inflação. Além disso, destacaram-se as cinco variáveis mais selecionadas por cada modelo, e vê-se que há certa estabilidade entre horizontes e até mesmo entre modelos. Destaca-se a preferência por variáveis de atividade econômica, de moeda, de incerteza, e confiança e pelo próprio IPCA.

**Palavras-chave:** Previsão de inflação. Aprendizado de máquina. Econometria.

## **ABSTRACT**

Inflation forecasting is important for economic agents such as firms, workers, and investors, as well as for monetary authorities. In the context of advances in artificial intelligence (AI) and the extensive availability of data (big data), this study aims to test the predictive capacity of some machine learning models (ridge regression, LASSO, Random Forest, and Gradient Boosting) and compare them to ARIMA. For this purpose, the chosen variable was the National Consumer Price Index (IPCA) of Brazil. The models were trained with data from January 2004 to February 2017, and their predictions were tested for the period from March 2017 to September 2023. In addition to IPCA, the models were trained with 40 other time series. Some error metrics were calculated for the predictions, and the machine learning model predictions were compared to ARIMA using the Diebold-Mariano test. As a result, it was found that machine learning models (with the exception of ridge regression) yielded superior results (i.e., lower error) compared to ARIMA for almost all time horizons. However, the results were not significantly superior to ARIMA at a 5% significance level for the Diebold-Mariano test in most cases. Therefore, it was concluded that, despite AI models achieving good results, it is not yet possible to discard ARIMA in the context of inflation forecasting. Additionally, the five most selected variables by each model were highlighted, and there is a certain stability observed among horizons and even among models. The top predictors chosen included economic activity, currency, uncertainty and confidence variables, as well as the IPCA itself.

**Keywords:** Inflation forecasting. Machine learning. Econometrics.



## LISTA DE FIGURAS

Figura 1 – Funcionamento de um neurônio artificial	19
Figura 2 - Ilustração de neurônio	20
Figura 3 - Ilustração de RNA com quatro camadas	22
Figura 4 - RNA semelhante à regressão linear	25
Figura 5 - Ilustração da restrição dos parâmetros em RR e LASSO	30
Figura 6 - Exemplo de árvore de decisão	32
Figura 7 - Divisão resultante da árvore na figura 1	33
Figura 8 - Ilustração do processo de bootstrap	35
Figura 9 - Ilustração do processo de redução do sktime	42
Figura 10 - Estratégia de janela crescente	45

## **LISTA DE GRÁFICOS**

Gráfico 1 - Variação % mensal do IPCA de 2004 a set/2023	44
Gráfico 2 – IPCA real e previsões para $h = 1$	47
Gráfico 3 - IPCA real e previsões para $h = 2$	48
Gráfico 4 - IPCA real e previsões para $h = 3$	49
Gráfico 5 - IPCA real e previsões para $h = 6$	50
Gráfico 6 - IPCA real e previsões para $h = 12$	50

## **LISTA DE TABELAS**

Tabela 1 - Métricas de erro para os cinco horizontes de previsão	49
Tabela 2 - Valores-p para o teste de Diebold-Mariano	51
Tabela 3 - Cinco variáveis de maior peso para $h = 1$ por modelo	52
Tabela 4 - Cinco variáveis de maior peso para $h = 2$ por modelo	53
Tabela 5 - Cinco variáveis de maior peso para $h = 3$ por modelo	54
Tabela 6 - Cinco variáveis de maior peso para $h = 6$ por modelo	55
Tabela 7 - Cinco variáveis de maior peso para $h = 12$ por modelo	56

## LISTA DE ABREVIATURAS E SIGLAS

<i>AR</i>	<i>Autoregressive</i>
<i>ARIMA</i>	<i>Autoregressive Integrated Moving Average</i>
<i>ARMA</i>	<i>Autoregressive Moving Average</i>
BCB	Banco Central do Brasil
<i>CSFE</i>	<i>Cumulative Squared Forecast Errors</i>
<i>DL</i>	<i>Deep Learning</i>
DM	Diebold-Mariano
EMA	Erro Médio Absoluto
EPAM	Erro Percentual Absoluto Médio
EPE	Empresa de Pesquisa Energética
EQM	Erro Quadrático Médio
FGV	Fundação Getúlio Vargas
<i>FRED</i>	<i>Federal Reserve Economic Data</i>
IA	Inteligência Artificial
IBGE	Instituto Brasileiro de Geografia e Estatística
IPCA	Índice Nacional de Preços ao Consumidor Amplo
KPSS	Kwiatkowski–Phillips–Schmidt–Shin
<i>LASSO</i>	<i>Least Absolute Shrinkage and Selection Operator</i>
<i>LGBM</i>	<i>Light Gradient Boosting Machine</i>
<i>MA</i>	<i>Moving Average</i>
<i>ML</i>	<i>Machine Learning</i>
PIB	Produto Interno Bruto
<i>RF</i>	<i>Random Forest</i>
RNA	Rede Neural Artificial
RR	Regressão <i>Ridge</i>
SQR	Soma dos Quadrados dos Resíduos
STQ	Soma Total dos Quadrados

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>15</b>
<b>2</b>	<b>APRENDIZADO DE MÁQUINA: UMA REVISÃO</b> .....	<b>18</b>
2.1	INTELIGÊNCIA ARTIFICIAL.....	18
2.1.1	Redes Neurais Artificiais.....	18
2.1.2	O Cérebro Humano .....	19
2.2	APRENDIZADO.....	20
2.3	APRENDIZADO DE MÁQUINA E ECONOMETRIA .....	24
<b>3</b>	<b>METODOLOGIA</b> .....	<b>28</b>
3.1	MODELOS.....	28
3.1.1	Regressão <i>Ridge</i> (RR).....	28
3.1.2	<i>Least Absolute Shrinkage and Selection Operator</i> (LASSO) .....	29
3.1.3	<i>Random Forest</i> (RF).....	31
3.1.4	Gradient Boosting .....	36
3.1.5	ARIMA .....	38
3.2	MÉTRICAS DE COMPARAÇÃO .....	40
3.3	FERRAMENTAS UTILIZADAS .....	41
<b>4</b>	<b>DADOS</b> .....	<b>44</b>
<b>5</b>	<b>RESULTADOS</b> .....	<b>47</b>
5.1	RESULTADOS DOS MODELOS .....	47
5.2	VARIÁVEIS SELECIONADAS .....	51
<b>6</b>	<b>CONCLUSÃO</b> .....	<b>57</b>
	<b>REFERÊNCIAS</b> .....	<b>59</b>
	<b>APÊNDICE A – LISTA DE VARIÁVEIS UTILIZADAS NAS PREVISÕES</b> .....	<b>63</b>
	<b>APÊNDICE B – GRÁFICOS DE CSFE POR MODELO E POR HORIZONTE</b> .....	<b>64</b>

## 1 INTRODUÇÃO

A previsão da inflação é um assunto de grande relevância. Para os agentes econômicos, é importante para a formação de suas expectativas, o que frequentemente se manifesta na indexação a algum índice (como o Índice Nacional de Preços ao Consumidor Amplo – IPCA) ou a criação de alguma métrica de correção em contratos de emprego e de aluguel (*vide* a indexação generalizada no Brasil até o Plano Real). Para investidores e instituições financeiras, é relevante para o cálculo do retorno real de seus investimentos bem como para a precificação de ativos. Por fim, a previsão da inflação é relevante para banqueiros centrais pois auxilia na ancoragem de expectativas e, conseqüentemente, na condução de política monetária.

No mercado financeiro, a previsão de inflação é especialmente interessante devido à existência de títulos de renda fixa indexados à inflação (como as NTN-Bs do Tesouro Nacional). Esses instrumentos pós-fixados<sup>1</sup> pagam a variação de um determinado índice de preços (como o IPCA) mais alguma porcentagem de retorno. Por coexistirem no mercado com outros títulos, como os pré-fixados, o preço dessas obrigações acaba indicando uma previsão implícita da inflação por parte do mercado. Por exemplo, se um determinado título de um ano pós-fixado à inflação está retornando IPCA + 6% e está sendo vendido pelo mesmo preço que um título pré-fixado com o mesmo valor de face e duração, mas que retorna 12% ao ano, a previsão implícita do mercado é de que o IPCA será de 6% ao longo do próximo ano. Assim, o investidor que tiver uma previsão de inflação mais precisa que a do mercado poderá ter um ganho de arbitragem, fato que reforça a relevância prática deste trabalho.

Ademais, a discussão de qual o melhor método para prever a inflação é por si só um tópico de importância teórica.

Desse modo, há um grande incentivo para melhorar nossas capacidades preditivas da inflação. A universalização das tecnologias da informação e sua aplicação em diversos contextos (inclusive econômicos) possibilita a coleta de quantidades

---

<sup>1</sup> Títulos pós-fixados são aqueles em que a remuneração está atrelada a alguma taxa desconhecida no presente, como a Selic ou algum índice de inflação. Em contrapartida, pré-fixados são remunerados a uma taxa já explicitada no presente, como 12% ao ano.

colossais de dados, num fenômeno conhecido como *big data*. Ao mesmo tempo, as melhorias em *hardware* e na área da inteligência artificial (IA) permitem a aplicação de modelos de previsão cada vez mais sofisticados. Conseqüentemente o ferramental disponível a econométristas se encontra consideravelmente expandido.

Portanto, este trabalho objetiva testar essas novas ferramentas na previsão de inflação e comparar sua capacidade preditiva ao *ARIMA (Autoregressive Integrated Moving Average)*, modelo já amplamente conhecido por econométristas. Comparamos o desempenho dos modelos por algumas estatísticas de erro e por meio do teste de Diebold-Mariano individualmente contra o *ARIMA*, visando observar se esses modelos superam significativamente o *ARIMA* nas suas previsões.

Experimentos semelhantes já foram realizados. Para a inflação brasileira, Medeiros, Vasconcelos e Freitas (2016) analisaram a capacidade preditiva do *Least Absolute Shrinkage and Selection Operator (LASSO)* utilizando o *ARIMA* como referencial, e encontraram que o *LASSO* apresentava previsões com menores erros para horizontes de tempo mais curtos, mas perdia para o *ARIMA* em horizontes mais longos. Num trabalho mais recente, Araújo e Gaglianone (2023) compararam 50 métodos de previsão, inclusive métodos de IA e modelos econométricos. Para os autores, dentre os modelos com as melhores previsões estavam modelos baseados em árvores (como o *Random Forest* e o *Gradient Boosting* estudados neste trabalho). Ademais, encontraram-se evidências de não-linearidades na inflação brasileira, o que ajuda a explicar o bom desempenho desses modelos.

No contexto americano, Medeiros *et al.* (2021) comparam alguns modelos de aprendizado de máquina a modelos econométricos como o *Random Walk* e o modelo autorregressivo. Como resultado, os modelos de IA com muitas covariáveis se mostraram sistematicamente mais precisos de modelos univariados. Ademais, os autores destacam o *Random Forest*, “*which dominated all other models in several cases*” (Medeiros *et al.*, 2021). Os autores atribuem o bom desempenho do modelo às não-linearidades existentes entre a inflação e outras variáveis macroeconômicas e à sua capacidade de seleção de variáveis.

O capítulo 2 contextualiza a área de IA comparando o funcionamento de alguns modelos ao cérebro humano e a modelos econométricos. O capítulo 3 explica

individualmente o funcionamento dos modelos de aprendizado de máquina que queremos testar bem como o do *ARIMA*, além de apresentar as métricas de erro que utilizamos para a comparação, o teste de Diebold-Mariano e as ferramentas utilizadas para a execução dos modelos. O capítulo 4 trata dos dados utilizados, tanto da variável endógena (variação mensal do IPCA) quanto das exógenas, bem como das manipulações realizadas sobre eles (imputação, estacionarização e normalização). O capítulo 5 apresenta os resultados na forma de gráficos e de tabelas e o capítulo 6 conclui.



## 2 APRENDIZADO DE MÁQUINA: UMA REVISÃO

Nesta seção, conceituaremos o aprendizado de máquina, comparando-o à econometria tradicional. Ademais, exploraremos trabalhos anteriores que tratam do seu uso para a previsão de variáveis econômicas.

Para exemplificar inteligência artificial e compará-la com modelos econométricos tradicionais escolhemos tratar de redes neurais artificiais (RNAs). Apesar de não testarmos nenhum algoritmo deste tipo para prever a inflação, as redes neurais são excelentes para ilustrar toda a motivação por trás da área da IA devido à sua semelhança ao funcionamento do cérebro humano. Além disso, como veremos adiante, sua relação com métodos estatísticos e econométricos é relativamente clara quando descritas matematicamente.

### 2.1 INTELIGÊNCIA ARTIFICIAL

Sendo o aprendizado de máquina uma área da inteligência artificial (What is Machine Learning?, [c2023]), é cabível em primeiro lugar descrevermos este termo. Para Sargent (2023, p.11), inteligência artificial são “*computer programs that are designed to do some of the ‘intelligent’ things that creative people like Galileo, Darwin and Kepler have done*”. No artigo, Sargent (2023) relata como algumas das mentes mais notórias da história da humanidade conseguiram construir teorias de profundo impacto partindo de grandes amostras de dados, reduzindo suas dimensões e extraindo um princípio que poderia ser generalizado.

Portanto, algumas das primeiras tentativas de criar uma inteligência artificial buscaram replicar o funcionamento do cérebro humano. Percebe-se isso ao vermos os sistemas especialistas, IAs que intentam reproduzir o comportamento de algum humano especializado em um determinado assunto. Eles operam a partir de uma base de conhecimento construída por especialistas humanos, e podem substituí-los mesmo quando utilizados por um leigo (Rougas, 2020).

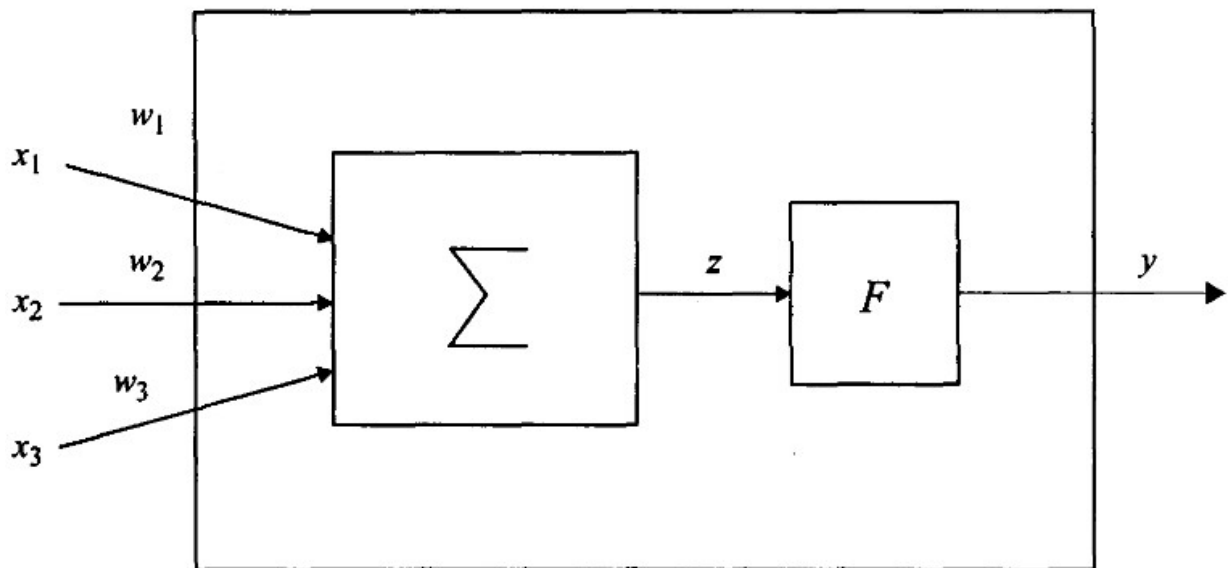
#### 2.1.1 Redes Neurais Artificiais

Outra corrente da área da IA que almeja reproduzir os mecanismos do cérebro humano são as RNAs. Segundo Portugal (1995), em se tratando do pensar, devemos fazê-lo à moda da natureza, isto é, para criar uma inteligência artificial, devemos antes

criar um cérebro artificial. Assim, as RNAs buscam recriar a operação dos neurônios humanos. A imagem abaixo ilustra o funcionamento de um neurônio artificial.

O neurônio recebe entradas ou “sinais” ( $x$ ), calcula a sua média ponderada ( $z$ ) e, alimenta esta média a uma função de ativação ( $F$ ) e produz um resultado ( $y$ ). Esse resultado posteriormente pode ser alimentado a outros neurônios.

Figura 1 – Funcionamento de um neurônio artificial



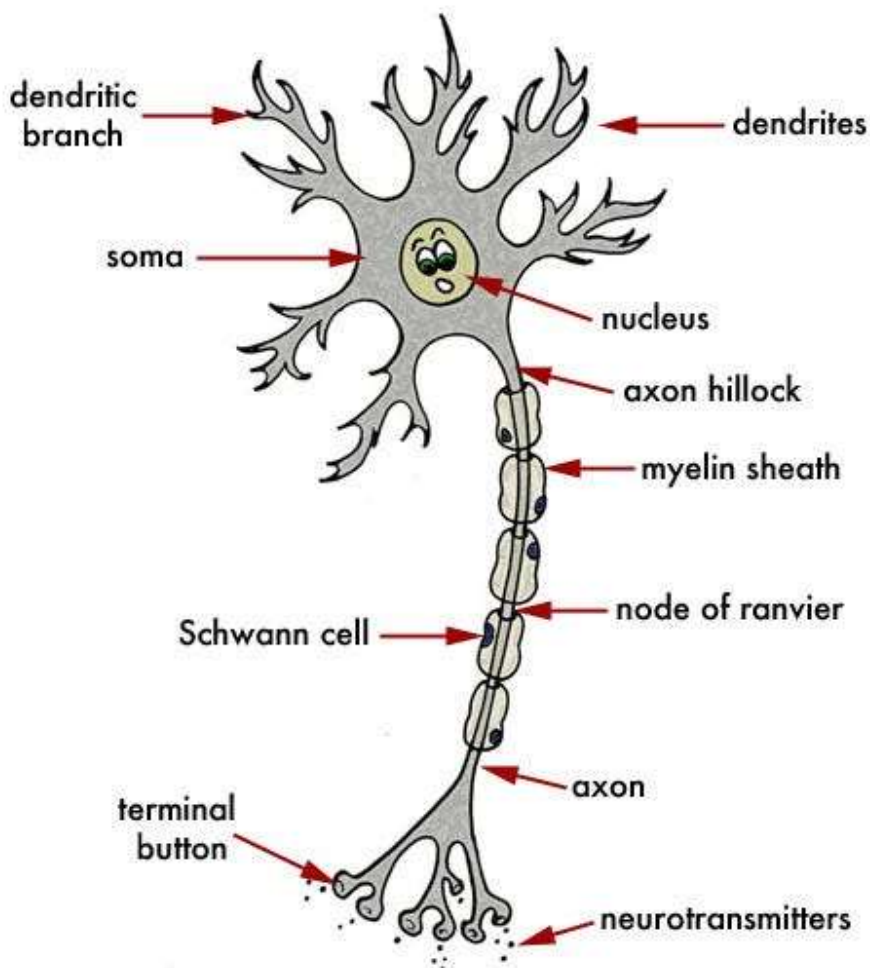
Fonte: Portugal (1995)

### 2.1.2 O Cérebro Humano

O neurônio humano funciona de maneira semelhante ao neurônio artificial que inspirou. Para o paralelo com o neurônio artificial, três componentes são relevantes: os dendritos, o corpo celular e o axônio. Os dendritos captam os estímulos que o neurônio recebe, de modo semelhante aos pesos  $w$  na Figura 1. O corpo celular junta todos os sinais e os envia para o axônio, que fica encarregado de os repassar a outras células (Uzunian; Birner, 2013). A Figura 2 indica esses componentes num neurônio. Acredita-se que o equivalente à função do neurônio artificial (o ponto em que se decide se o estímulo será passado adiante) se localize na base do axônio (Clark; Goldberg; Rudy, 2009). Os estímulos são repassados segundo a chamada *lei do tudo ou nada*: são repassados sempre na mesma intensidade, e só se os estímulos que o neurônio recebe atingem certo ponto, o *limiar de excitação*. Caso não atinjam esse limiar, o neurônio não

emite impulso algum (Uzunian; Birner, 2013). Interligando cerca de 100 bilhões desses neurônios (von Bartheld; Bahney; Herculano-Houzel, 2016), o cérebro humano se assemelha às RNAs que tentam simulá-lo.

Figura 2 - Ilustração de neurônio



Fonte: Panchal (2018)

## 2.2 APRENDIZADO

Consoante Carbonell (1989), o aprendizado, no contexto de inteligência artificial, como a capacidade de realizar tarefas previamente impossíveis ou de aprimorar a habilidade de fazer antigas tarefas (seja mais rapidamente, mais precisamente).

A expressão “aprendizado de máquina” compreende uma ampla variedade de métodos, mas estes podem ser divididos em duas grandes categorias: aprendizado *supervisionado* e *não-supervisionado*. O primeiro caso ocorre quando, para cada

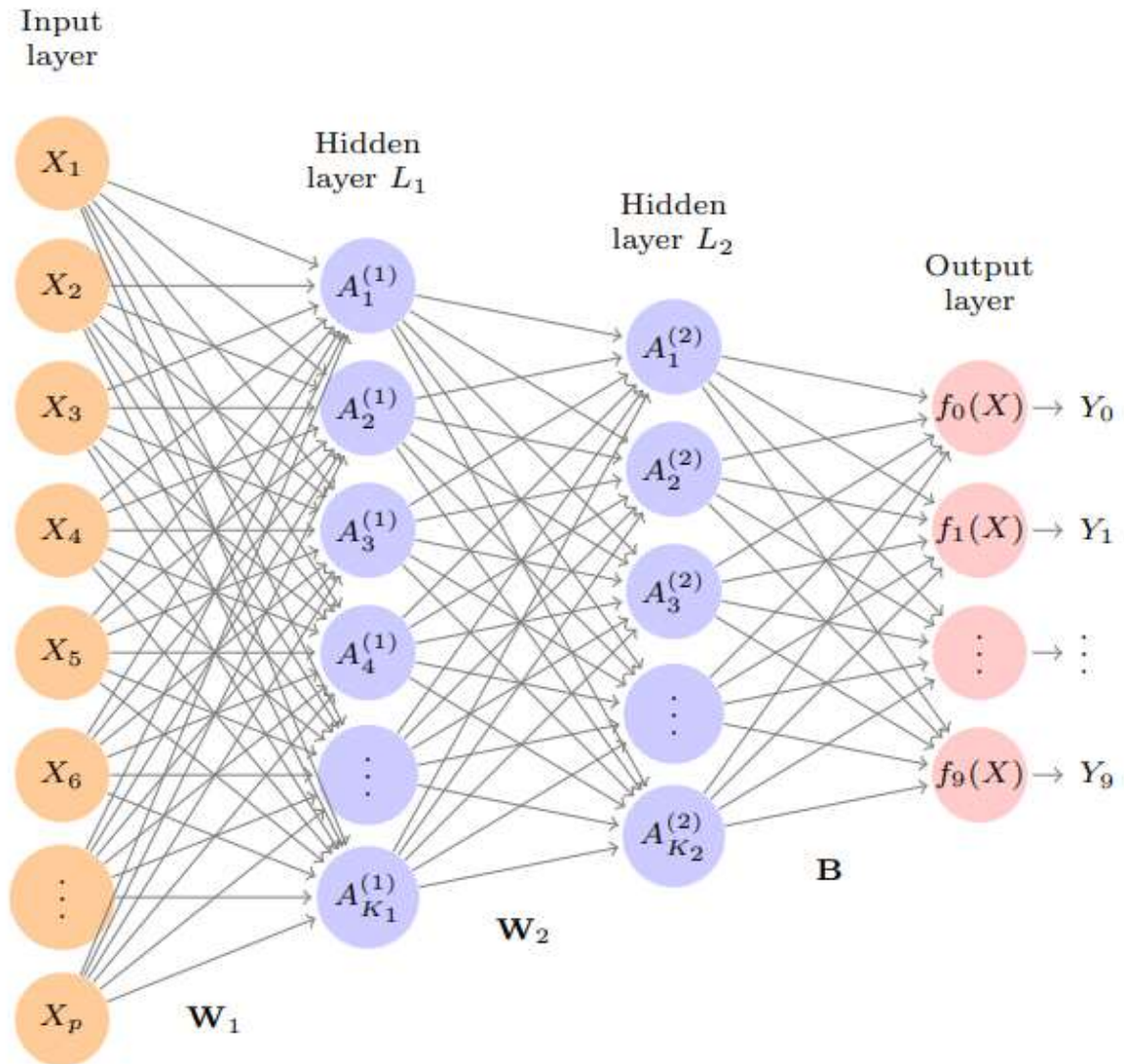
observação  $x_i$ , temos uma resposta  $y_i$ . Desse modo, o modelo relaciona os preditores  $x$  à resposta  $y$ . Aqui se encaixa a clássica regressão linear, mas também métodos como regressão Ridge, *LASSO* (ambos abordados neste trabalho), RNAs e outros métodos mais sofisticados (como máquinas de vetores de suporte). O aprendizado supervisionado é geralmente voltado à previsão ou à inferência (James *et al.*, 2023d).

Por outro lado, temos aprendizado não-supervisionado quando não há uma resposta correspondente para as observações. Os algoritmos que se enquadram nesta categoria almejam reconhecer padrões nos dados e determinar categorias de classificação (Jung; Patnam; Ter-Martirosyan, 2018). Exemplos de uso deste tipo de método são o agrupamento de imagens e a classificação de documentos. Este trabalho não tratará deste tipo de método.

Cabe explicar aqui como funciona o algoritmo de aprendizado das RNAs citadas anteriormente, pois ilustram bem o processo de aprendizado supervisionado. Há também RNAs que operam por aprendizado não-supervisionado, todavia este trabalho trata apenas de algoritmos de aprendizado supervisionado, então focaremos a exposição neste tipo.

Como o nome sugere, RNAs nada mais são que redes de neurônios como os apresentados acima. A rede recebe observações  $X$  e saídas  $Y$ , tendo que conectá-las de algum modo. Para tanto, os diversos neurônios são conectados entre si, formando diversas camadas. Desse modo, a rede possui uma camada de entrada, que é a primeira a receber  $X$ , e uma camada de saída, que produz  $Y$ . Quaisquer camadas entre essas duas são chamadas de “camadas escondidas” no jargão de redes neurais. A figura abaixo ilustra uma rede neural com quatro camadas, uma de entrada, uma de saída e duas ocultas.

Figura 3 - Ilustração de RNA com quatro camadas



Fonte: James *et al.* (2023a)

O aprendizado ocorre por meio da comparação entre a saída da rede e o valor real de  $Y$  correspondente às entradas. Se a saída produzida pela rede a partir das entradas for igual à saída alvo, não há modificações. Contudo, se houver diferença, os pesos dos neurônios (representados pelo vetor  $w$  na Figura 1) são alterados segundo alguma regra. O processo é repetido até se alcançar um erro tão pequeno quanto o desejado (Portugal; Fernandes, 1996).

White (1992) descreve detalhadamente o processo de aprendizado de RNAs conhecido como “retropropagação” (*backpropagation*). Matematicamente, o neurônio  $j$  de uma camada escondida emite o sinal descrito por:

$$h_j = G \left( \sum_{i=0}^r \gamma_{ji} x_i \right)$$

Em que  $G$  é a função de ativação (geralmente de formato sigmoide) e  $\gamma_{ji}$  é o peso da entrada  $i$  para o neurônio  $j$ . De modo análogo, podemos descrever a saída da rede como:

$$o = G \left( \sum_{j=0}^q \beta_j G \left( \sum_{i=0}^r \gamma_{ji} x_i \right) \right) = f(X, \theta)$$

Em que  $\beta$  são os pesos dos neurônios escondidos na saída da rede. A expressão mais à direita é apenas uma abreviação conveniente, pois depende apenas dos pesos e das entradas. Hornik *et al.* (1989<sup>2</sup>, *apud* WHITE, 1992, p. 83) demonstra que essa função de saída pode aproximar qualquer função que provavelmente será encontrada, contanto que o número de neurônios escondidos  $q$  seja suficientemente grande. Essa propriedade de “aproximação universal” viabiliza a aplicação do conceito em diversas áreas, inclusive a previsão de séries temporais (White, 1992).

Dada essa descrição de uma RNA, o processo de aprendizado por retropropagação consiste em atualizar os pesos  $\theta$  segundo a seguinte fórmula:

$$\theta_t = \theta_{t-1} + \eta \nabla f(X_t, \theta_{t-1}) (Y_t - f(X_t, \theta_{t-1})), t = 1, 2, \dots,$$

Em que  $\eta$  é a taxa de aprendizado e  $\nabla f$  é o vetor gradiente em função dos pesos  $\theta$ . Desse modo, o erro (descrito aqui por  $Y_t - f(X_t, \theta_{t-1})$ ) se “propaga de volta”, alterando os pesos de cada neurônio.

As RNAs receberam mais atenção nos últimos anos, mas dessa vez com o nome de aprendizado profundo (*deep learning*) (James *et al.*, 2023a), que nada mais são que redes neurais com múltiplas camadas ocultas. Ocorre que as RNAs deste tipo treinadas com poucos dados acabam obtendo resultados piores que modelos mais tradicionais

---

<sup>2</sup> Hornik, K. M., M. Stinchcombe and H. White (1989) Multilayer feedforward networks are universal approximators. *Neural Networks*, 2, 359-66.

(London *et al.*, 2017). O sucesso recente deste tipo de algoritmo se deve ao aumento do poder computacional, isto é, melhorias no *hardware* especializado (Asokan, 2023), e ao advento do *big data*, ou seja, a ampla disponibilidade de dados.

Já houve tentativas de utilizar modelos de *DL* para a previsão da inflação. Theoharidis, Guillén e Lopes (2023) compararam RNAs com outros modelos de aprendizado de máquina, como LASSO (*Least Absolute Shrinkage and Selection Operator*) e RF (*Random Forest*), e modelos econométricos, como RW (*Random Walk*) e ARIMA, e encontraram que as redes neurais os superaram na previsão da inflação americana. Contudo, devido ao poder computacional e abundância de dados necessários para rodar esse tipo de algoritmo, este trabalho não almeja replicar esses resultados.

### 2.3 APRENDIZADO DE MÁQUINA E ECONOMETRIA

Quanto à relação entre aprendizado de máquina e econometria tradicional, White (1992, p. 81) tem a dizer:

Neural networks are enticing, statistics are not. Despite this difference or perhaps because of it, people often ask me: “So, this neural network stuff and this back-propagation thing. Is it really a new and amazing way of finding patterns in data? How does it fit in with the standard ways that statisticians have cooked up for looking at data, like regression analysis?”

De fato, retropropagação não foi uma criação da ciência da computação, mas uma aplicação de um método estatístico proposto em 1951 por Herbert Robbins e Sutton Monro, chamado de “aproximação estocástica”. O lado bom é que a literatura estatística deste assunto pode agora ser aplicada no contexto de retropropagação. Por exemplo, essa literatura prova que, dadas certas condições, retropropagação encontra pesos que convergem para o vetor  $\theta^*$  que resolve a equação (White, 1992):

$$E(\nabla f(X_t, \theta)[Y_t - f(X_t, \theta)]) = 0$$

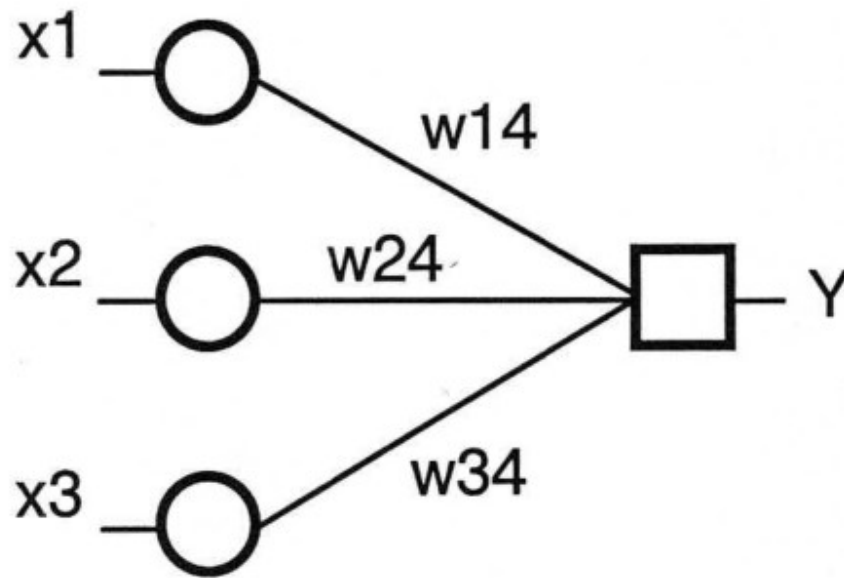
Uma solução para essa equação também é uma solução local para o problema de mínimos quadrados:

$$\min_{\theta} E \left[ (Y_t - f(X_t, \theta))^2 \right]$$

Em que  $X$  são as variáveis exógenas e  $\theta$  são os parâmetros. A semelhança não é mera coincidência: funções de saída de redes neurais correspondem a específicas famílias de curvas de regressão (White, 1992). Em verdade, os valores de entrada de uma RNA nada mais são que as variáveis exógenas de mínimos quadrados, e a saída

da rede é apenas o valor previsto da variável endógena. Com efeito, uma RNA de apenas uma camada é idêntica ao modelo de regressão linear (Portugal; Fernandes, 1996).

Figura 4 - RNA semelhante à regressão linear



Fonte: Portugal e Fernandes (1996)

A figura acima ilustra o caso de uma RNA simples, sem camadas ocultas. Neste caso, os pesos dos neurônios ( $w_{14}$ ,  $w_{24}$  e  $w_{34}$ ) nada mais são que os parâmetros do modelo de regressão linear. Desse modo, retropropagação e regressão não linear podem ser vistos como diferentes abordagens para resolver o problema de mínimos quadrados (White, 1992). De fato, Athey (2019) postula que há coincidência entre diversos métodos utilizados há décadas na Estatística e nas Ciências Sociais e que foram adotados na área do aprendizado de máquina, bem como métodos ainda não reaproveitados.

Todavia, ainda há diferenças importantes entre as duas áreas. Consoante Athey (2019), nos modelos de IA, a comparação e seleção entre diferentes modelos (*tuning*, ou ajuste) faz parte do algoritmo, logrando em melhor desempenho por parte dos modelos e também permitindo aos pesquisadores descreverem sistematicamente o processo de escolha do modelo. Em bibliotecas modernas de aprendizado de máquina como o *scikit-learn* (usado neste trabalho), os parâmetros de diversos modelos (como a penalidade de



regularização do *LASSO*) são escolhidos automaticamente por validação cruzada já como parte do algoritmo. A validação cruzada é usada na Economia, mas é uma parte fundamental do processo na área do aprendizado de máquina.

Outro motivo pelo qual podemos segregar as duas áreas é o fato de que, até o passado recente, os métodos de *ML* (*machine learning*) não eram usados em publicações de Ciências Sociais, apesar de serem amplamente utilizados na literatura de *ML*, excetuando-se apenas a *RR* e o *LASSO* (Athey, 2019).

As duas áreas também se diferenciam nos resultados. Segundo Medeiros *et al.* (2021), para previsões do *consumer price index* americano, a redução de erros quadráticos médios pode chegar até 30%. Para os autores, os modelos de aprendizado de máquina superam as alternativas univariadas, principalmente no período de 2001 a 2015, em que a inflação foi mais volátil que nos anos 90. Dentre os modelos de *ML* testados no trabalho, se destaca o *Random Forest*. Consoante os autores, sua superioridade se deve tanto à sua capacidade de selecionar variáveis quanto às não-linearidades presentes na inflação. Estas não-linearidades advêm, por exemplo, da relação entre emprego e inflação, pois depende da ociosidade na economia. Outras fontes de não-linearidades são a incerteza e o piso igual a zero para taxas de juros nominais. Araújo e Gaglianone (2023) também reforçam a importância das não-linearidades na inflação, dessa vez no caso brasileiro.

Outrossim, devido à sua capacidade de lidarem com uma grande quantidade de dados, há um incentivo para utilizar métodos de *ML* na macroeconomia aplicada (Araújo; Gaglianone, 2023). Segundo Medeiros *et al.* (2021), “*Big data’ and ML methods are not passing fads, and investigating whether the combination of these two methods is able to provide more accurate forecasts is of paramount importance*”. Num contexto de avanço do poder computacional e de maior disponibilidade de dados (*big data*), faz sentido compararmos modelos de aprendizado de máquina com modelos econométricos clássicos univariados (ou com poucas variáveis).

Nesse sentido, apesar de teoricamente fazerem a mesma coisa, os modelos de *ML* se diferenciam da econometria tradicional por aproveitar melhor esses avanços. Contudo, é importante ressaltar que, apesar do jargão de “inteligência” e “aprendizado”, fundamentalmente o procedimento é o mesmo: a minimização de alguma métrica de erro

(num contexto de regressão, geralmente o erro quadrado). A grande contribuição do aprendizado de máquina é permitir a escolha de uma nova gama de funções não-lineares para relacionar as entradas e saídas (variáveis exógenas e endógenas).

De qualquer modo, as semelhanças entre aprendizado de máquina e os métodos estatísticos com os quais já estamos familiarizados levam alguns a questionarem se esta área do conhecimento deveria ter recebido um nome diferente de Estatística, ou se há alguma diferença entre IA e aprendizado de máquina (Athey, 2019). Como disse White (1992, p. 87): “*Neural network learning procedures are inherently statistical techniques*”. A diferença entre a Estatística e o *ML* parece estar principalmente na terminologia e nos agentes que os aplicam, não necessariamente nos métodos ou na teoria.

### 3 METODOLOGIA

Nesta seção, apresentamos os modelos de aprendizado de máquina utilizados neste trabalho, bem como o modelo de referência (*benchmark*), ARIMA. Ademais, listamos as métricas de avaliação utilizadas para comparar os modelos. Por fim, introduzimos as principais bibliotecas escolhidas para rodar os algoritmos.

#### 3.1 MODELOS

##### 3.1.1 Regressão Ridge (RR)

Sabe-se que a regressão linear estima seus parâmetros encontrando os valores que minimizam

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

Introduzida por Hoerl e Kennard (1970), a Regressão Ridge é semelhante. O valor a ser minimizado é

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

em que  $\lambda \geq 0$  é um parâmetro de ajuste (*tuning parameter*). A nova adição à fórmula é chamada de “penalidade de regularização” (*shrinkage penalty*) e os parâmetros  $(\beta_1, \dots, \beta_p)$  serão menores quanto menor for a penalidade. Desse modo, a RR favorece parâmetros menores, encolhendo-os (*shrinking*). Nesse sentido, a escolha de  $\lambda$  é fundamental, pois determina a importância relativa dos dois termos da fórmula na escolha dos coeficientes.

A vantagem da RR sobre a regressão linear se deve ao dilema entre viés e variância. À medida que  $\lambda$  cresce, a flexibilidade da regressão diminui, aumentando o viés, mas reduzindo a variância. Essa troca pode ser vantajosa, particularmente quando há poucas observações e muitos parâmetros.

Computacionalmente, a RR não é muito mais complicada que a regressão linear. É possível demonstrar que as computações necessárias para encontrar os coeficientes

para todos os valores de  $\lambda$  são quase idênticas às feitas para ajustar um modelo com mínimos quadrados (James *et al.*, 2023b).

### 3.1.2 *Least Absolute Shrinkage and Selection Operator (LASSO)*

A principal desvantagem da RR é que nenhum dos parâmetros chegarão a zero (a não ser que  $\lambda = \infty$ ). Nesse sentido, o *LASSO*, introduzido por Tibshirani (1996), se apresenta como uma boa alternativa. Seus coeficientes minimizam

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

Vê-se que as fórmulas da RR e do *LASSO* são semelhantes. A diferença está no fato de que a *penalidade de regularização* é a soma dos valores absolutos dos parâmetros, não de seus quadrados. No jargão estatístico, diz-se que a RR usa uma penalidade  $\ell_2$ , enquanto o *LASSO* utiliza uma penalidade  $\ell_1$ . Na prática, isso significa que alguns coeficientes podem chegar a zero quando  $\lambda$  é suficientemente grande. Desse modo, o *LASSO* acaba selecionando variáveis, o que confere a seus modelos uma interpretação mais fácil.

O *LASSO* foi escolhido por apresentar bons resultados na literatura de previsão da inflação. Em Medeiros *et al.* (2016), o modelo apresenta bons resultados para previsões de curto prazo, mas apresenta resultados inferiores (apesar de semelhantes) ao *ARIMA* em horizontes mais longos.

Há também formulações alternativas para RR e *LASSO*, apresentadas abaixo respectivamente (James *et al.*, 2023b):

$$\min_{\beta} \left\{ \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 \right\} \text{ s. a. } \sum_{j=1}^p \beta_j^2 \leq s$$

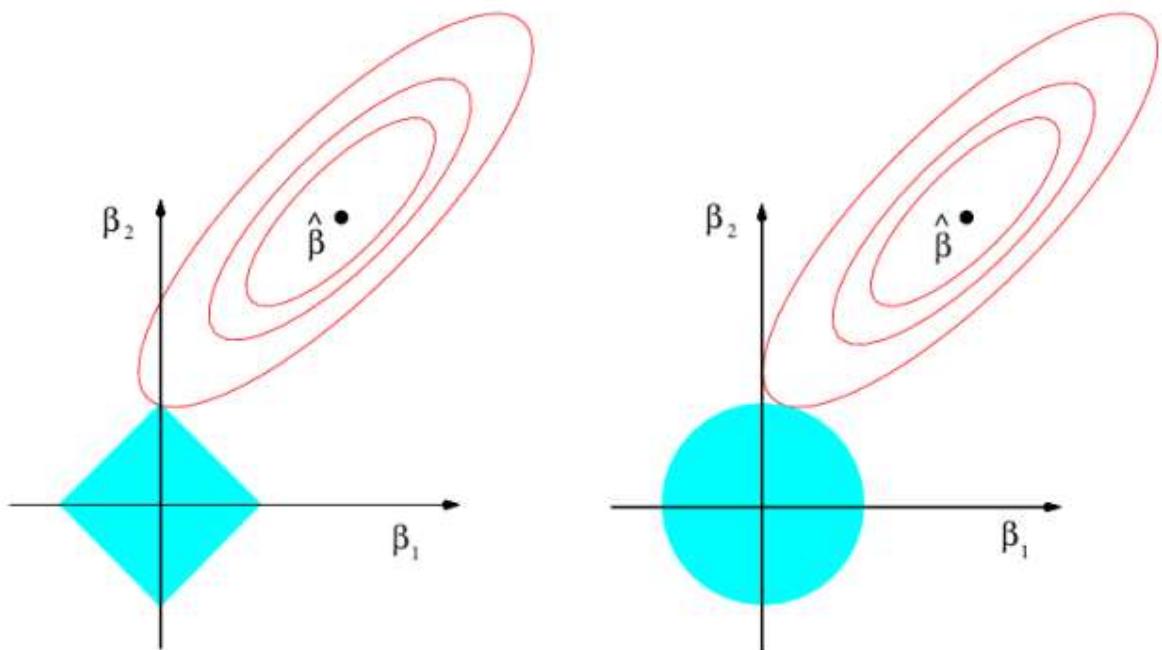
e

$$\min_{\beta} \left\{ \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 \right\} \text{ s. a. } \sum_{j=1}^p |\beta_j| \leq s$$

Isto é, para cada valor de  $\lambda$ , existe um  $s$  para o qual as formulações de ambos os modelos estimarão os mesmos coeficientes. É conveniente pensar em  $s$  como um “orçamento” para os parâmetros. Quando esse orçamento é grande, os parâmetros

podem ser maiores. De fato, se o orçamento for suficientemente grande, tanto *LASSO* quanto *RR* estimarão os mesmos parâmetros que mínimos quadrados ordinários. Do mesmo modo, quando o orçamento é pequeno, assim também serão os parâmetros estimados. A figura abaixo oferece uma intuição visual para a restrição.

Figura 5 - Ilustração da restrição dos parâmetros em *RR* e *LASSO*



Fonte: James *et al.* (2023b)

A figura ilustra uma situação em que  $p = 2$ . A solução por mínimos quadrados ordinários está indicada com  $\hat{\beta}$ , e as linhas vermelhas são isolinhas de SQR. Quanto mais as elipses se afastam de  $\hat{\beta}$ , maiores as SQR. O quadrilátero azul representa a região da restrição do *LASSO* ( $|\beta_1| + |\beta_2| \leq s$ ), enquanto o círculo azul representa a restrição da *RR* ( $\beta_1^2 + \beta_2^2 \leq s$ ). Os parâmetros em ambos os modelos serão escolhidos quando o primeiro ponto de uma isolinha se encontrar com a região de restrição. Devido ao formato circular de sua região de restrição, a isolinha da *RR* dificilmente se encontrará com a região de restrição sobre um eixo (onde um dos estimadores é igual a zero). Contudo, devido à presença de quinas em sua região de restrição, isso ocorre com mais frequência. Desse modo, a figura apresenta uma intuição geométrica do porquê do *LASSO* ser mais seletivo com as suas variáveis.

Devido à penalidade de regularização, as variáveis precisam estar na mesma escala, pois ela põe o mesmo peso em todos os coeficientes. Portanto, para rodarmos tanto o *LASSO* como a *RR* precisamos antes normalizar os dados, isto é, transformá-los de modo que fiquem com média igual a zero e desvio-padrão igual a um. Posteriormente, para utilizarmos as previsões, devemos desnormalizá-las (Ding, 2023).

Nas aplicações do *LASSO* e da *RR* neste trabalho, utilizaram-se as classes do *scikit-learn* que realizam a seleção do parâmetro de ajuste  $\lambda$  por meio de validação cruzada.

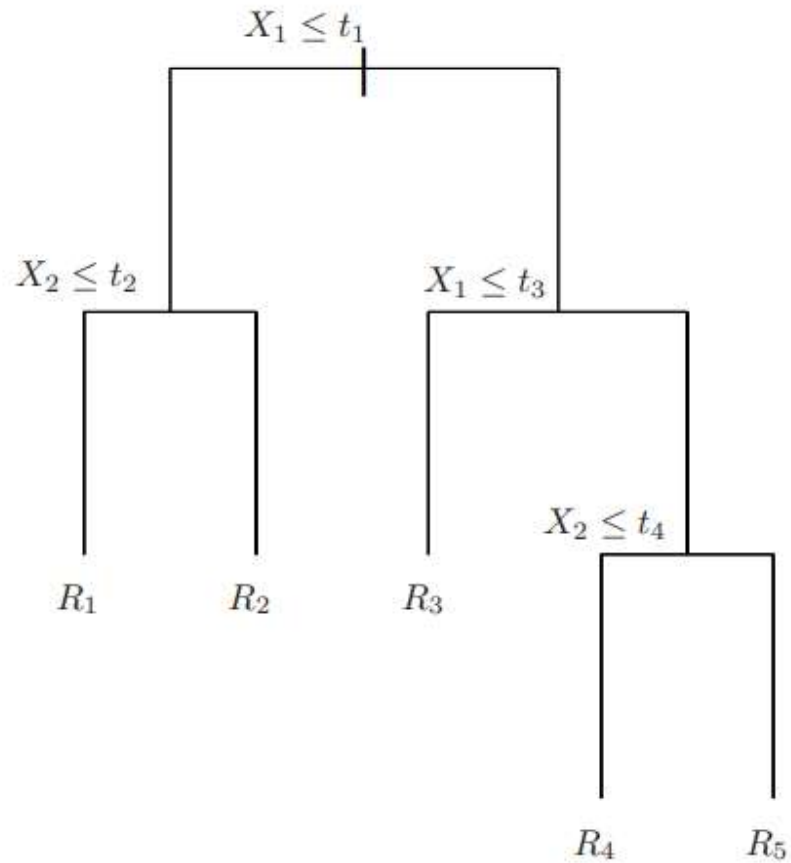
### 3.1.3 *Random Forest (RF)*

O *Random Forest* é um modelo de *aprendizado conjunto (ensemble learning)* introduzido por Breiman (2001). Chamam-se de aprendizado conjunto os modelos que juntam vários *aprendizes fracos (weak learner)*, modelos simples que obtêm resultados marginalmente melhores que a sorte. Ao juntar vários aprendizes fracos, modelos de aprendizado conjunto conseguem melhorar a qualidade de suas previsões. No caso do *Random Forest*, os aprendizes fracos são as árvores de decisão (de onde vem o “*forest*” no nome). Desse modo, a próxima seção é dedicada a descrever as árvores de decisão, pois é necessário entendê-las para compreender o *RF*. Além disso, a seção seguinte é dedicada ao *bagging*, outro elemento fundamental do *RF*.

#### 3.1.3.1 Árvores de decisão

Árvores de decisão são grafos com nodos que se bifurcam com base em algum critério, dividindo a base de dados. Nos contextos de regressão e classificação, foram introduzidas pela primeira vez por Breiman (1998). A figura abaixo ilustra uma árvore de decisão.

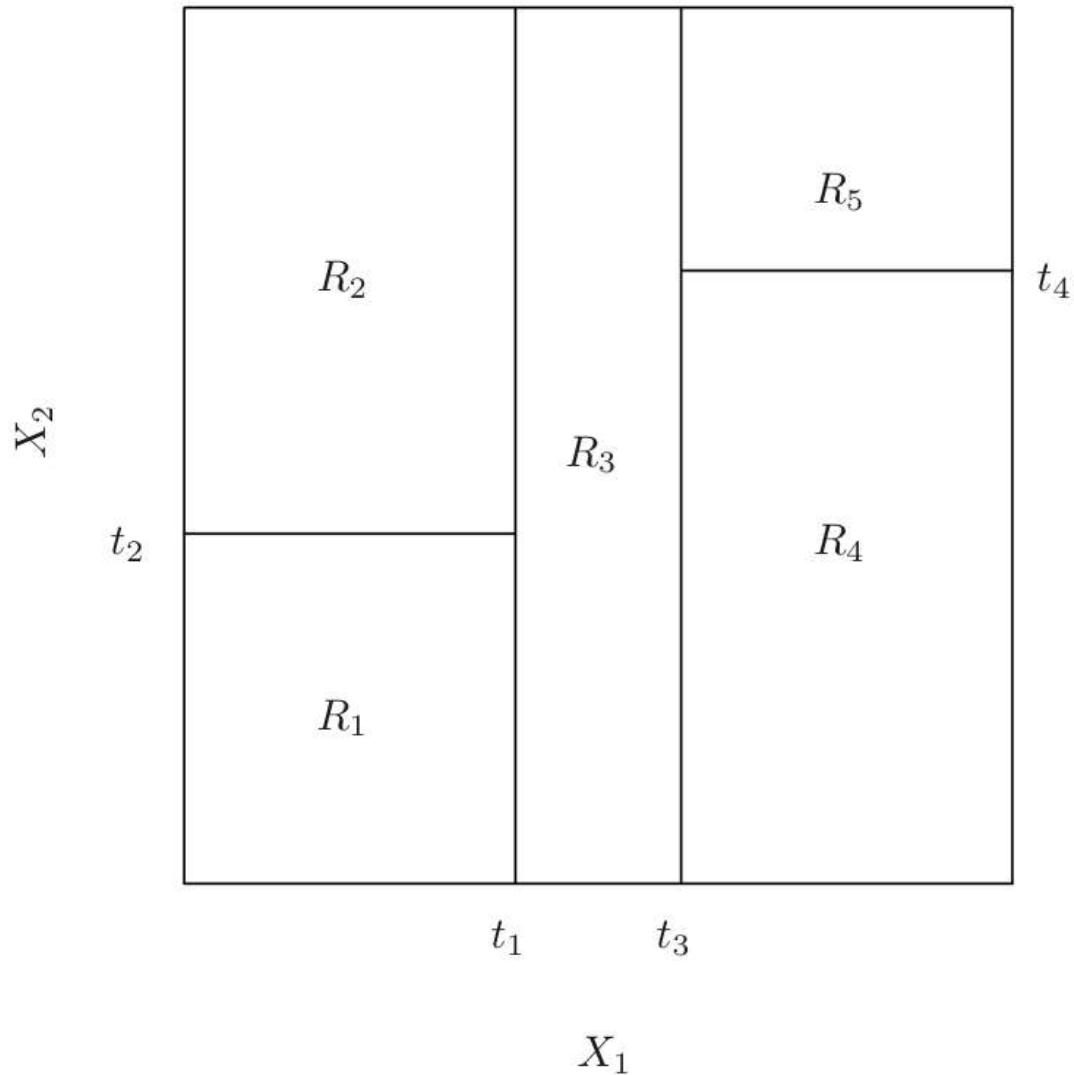
Figura 6 - Exemplo de árvore de decisão



Fonte: James *et al.* (2023e)

Ou seja, na figura acima, primeiramente a amostra é dividida em duas partes: observações em que  $X_1 \leq t_1$  e observações em que  $X_1 > t_1$ . Subsequentemente, o lado esquerdo da árvore é testado para  $X_2 \leq t_2$  e, o direito, para  $X_1 \leq t_3$ . Ao final do processo, a amostra fica segregada em cinco áreas,  $R_1, R_2, R_3, R_4$  e  $R_5$ . A figura abaixo ilustra essa separação.

Figura 7 - Divisão resultante da árvore na figura 1



Fonte: James *et al.* (2023e)

Num contexto de regressão, as previsões de uma árvore de decisão são as médias em cada um dos nós terminais, chamados de *folhas* (*leaf nodes*). Portanto, caso a árvore tivesse que prever o  $Y$  de uma nova observação em que  $X_1 \leq t_1$  e  $X_2 \leq t_2$ , o valor previsto seria a média das observações em  $R_1$ . Para classificação, a previsão é apenas a categoria majoritária na folha.

A escolha do melhor ponto de ramificação depende da aplicação da árvore: num contexto de classificação, pode-se utilizar a queda no índice de Gini. Num contexto de

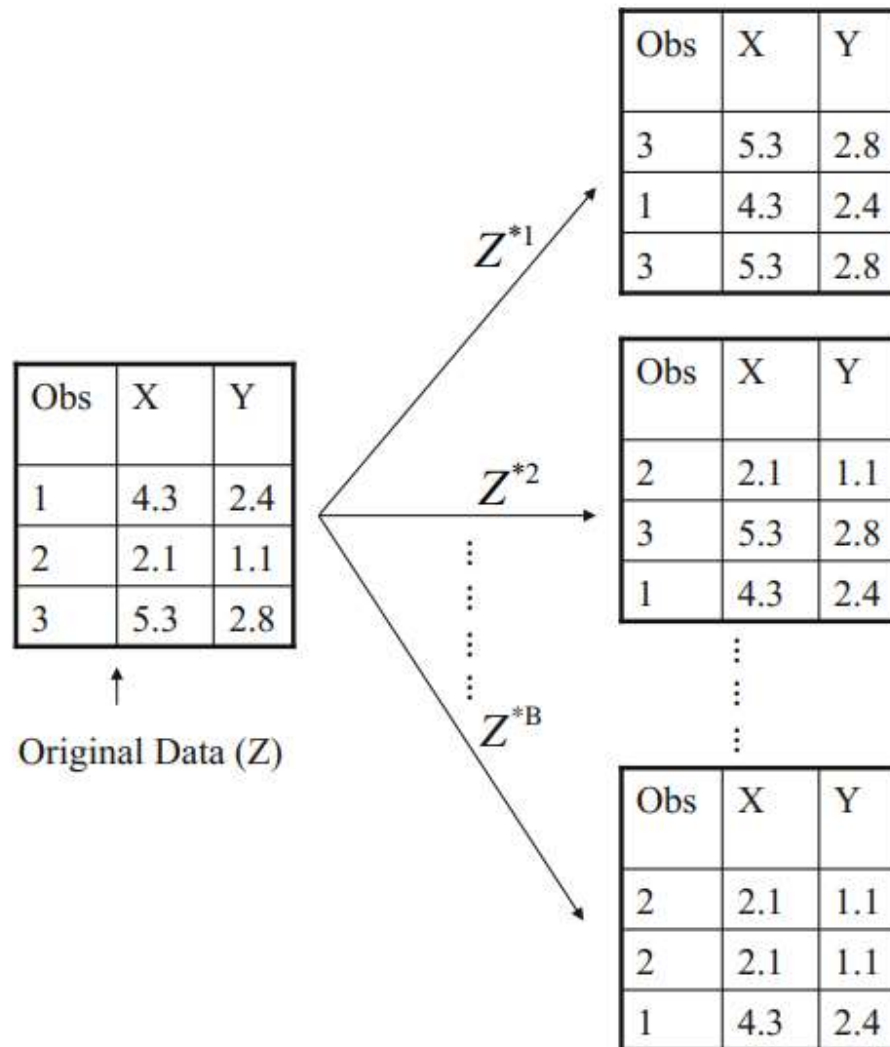


regressão, comumente usa-se a redução no erro quadrático ou no erro absoluto. Isto reforça a semelhança entre aprendizado de máquina e econometria abordada anteriormente neste trabalho.

Uma desvantagem de árvores de decisão é que elas podem ter variância elevada, isto é, uma mudança pequena nas observações utilizadas para estima-las pode levar a mudanças consideráveis na árvore final (James *et al.*, 2023e). Ademais, segundo Hastie, Tibshirani e Friedman (2017), “*Trees have one aspect that prevents them from being the ideal tool for predictive learning, namely inaccuracy.*”. Para solucionar esses problemas, frequentemente se utiliza o *bagging*, que reduz a variância da árvore final. Explicamos o *bagging* na seção seguinte.

#### 3.1.3.2 *Bagging*

*Bagging* (amalgama de *bootstrap* e *aggregation*) consiste em treinar modelos em amostras geradas a partir de um processo de *bootstrap*, que consiste em criar uma nova amostra escolhendo elementos da amostra original com reposição, de modo que a mesma observação pode aparecer mais de uma vez na nova amostra. A Figura 8 abaixo ilustra esse processo para uma amostra original de três observações. Após os modelos serem treinados nas amostras criadas, suas tira-se a média de suas previsões, resultando na previsão final do modelo. Esta última etapa é a “agregação” (James *et al.*, 2023c).

Figura 8 - Ilustração do processo de *bootstrap*

Fonte: James *et al.* (2023c)

Matematicamente, *bagging* se resume a tirar a média das previsões dentre diversas árvores de decisão treinadas em amostras geradas por *bootstrapping*. Isto é, a previsão gerada pelo processo pode ser descrita por (James *et al.*, 2023e):

$$\hat{f}_{bag}(x) = \frac{1}{B} \sum_{b=1}^B \hat{f}^{*b}(x)$$

Em que  $B$  é o número total de árvores (e de amostras criadas por *bootstrapping*) e  $\hat{f}^{*b}$  é uma árvore treinada na amostra  $b$ .

Apesar de reduzir a variância, as árvores criadas por um processo de *bagging* clássico geralmente serão correlacionadas, fazendo com que a redução em variância não seja tão grande. Para isso surge o *Random Forest*. O algoritmo funciona de maneira semelhante ao *bagging*, com uma pequena, mas significativa modificação: no momento em que a árvore deve criar uma nova ramificação, ela precisa escolher um dos  $p$  parâmetros para utilizar de critério. Contudo, se um dos parâmetros consistentemente gera ramificações melhores, várias (se não todas) as árvores o escolherão. Portanto, o *Random Forest* limita o número de parâmetros disponíveis a uma constante  $m$ . Obviamente, se  $m = p$ , estamos apenas fazendo o *bagging* padrão. Na prática, geralmente se opta por  $m = \sqrt{p}$ . Desse modo, o *RF* força as árvores a escolherem parâmetros diferentes, tornando-as menos correlacionadas e o modelo como um todo menos variável.

Por ser baseado em árvores de decisão, o *RF* é um modelo não-paramétrico. Na prática, isso significa que é impossível “abrir a equação” de um *RF* treinado, o que faz com que o modelo seja de difícil interpretação (James *et al.*, 2023e). As RNAs apresentadas no capítulo 2 sofrem do mesmo problema.

O modelo *RF* foi escolhido por apresentar bons resultados na literatura de previsão de inflação. Isso resulta do fato do algoritmo capturar bem as não-linearidades da inflação brasileira (Araújo; Gaglianone, 2023). Segundo Medeiros *et al.* (2021), “*the method that deserves more attention is the random forest model, which dominates all other models.*”

### 3.1.4 Gradient Boosting

Esta seção descreve o *gradient boosting*, um algoritmo de aprendizado conjunto não-paramétrico semelhante ao *RF*. A primeira parte explica o que é *boosting* de modo geral, enquanto a segunda especifica para o contexto de *gradient boosting*. Apesar de ser possível aplicar *boosting* a diversos métodos de aprendizado de máquina (James *et al.*, 2023e), focamos nossa discussão para o âmbito de árvores de decisão.

#### 3.1.4.1 Boosting

*Boosting*, quando aplicado para árvores de decisão (que é o caso mais comum), funciona de modo semelhante ao *bagging*. Todavia, em vez de se criarem diversas árvores em diferentes amostras criadas pelo processo de *bootstrap*, as árvores são criadas sequencialmente com base nos resíduos das árvores anteriores (desse modo

utilizando a informação acumulada por elas). O algoritmo abaixo descreve o processo (James *et al.*, 2023e):

1. Seja  $\hat{f}(x) = 0$  e  $r_i = y_i$  para todo  $i$  no conjunto de treino.
2. Para  $b = 1, \dots, B$ :
  - a. Ajuste uma árvore  $\hat{f}^b$  com  $d$  nós de ramificação ( $d + 1$  nós terminais) aos dados  $(X, r)$
  - b. Atualize  $\hat{f}$  adicionando-lhe uma versão reduzida da nova árvore:
 
$$\hat{f}(x) \leftarrow \hat{f}(x) + \lambda \hat{f}^b(x)$$
  - c. Atualize os resíduos:

$$r_i \leftarrow r_i - \lambda \hat{f}^b(x_i)$$

3. Devolva o modelo resultante:

$$\hat{f}(x) = \sum_{b=1}^B \lambda \hat{f}^b(x)$$

O algoritmo possui três parâmetros de ajuste: o número de árvores  $B$ , o parâmetro de redução  $\lambda$  (por vezes chamado de *taxa de aprendizado*) e o número de ramificações  $d$  para cada árvore. Em diversos modelos de *boosting*, favorecem-se árvores menores<sup>3</sup>, até mesmo com  $d = 1$ . Neste caso, diz-se que a árvore é um “toco” (*stump*).

#### 3.1.4.2 Gradient boosting

Introduzido por Friedman (2002), *gradient boosting* é uma extensão do método de *boosting*. O processo geral é o mesmo, com algumas diferenças importantes:

1. Os resíduos são calculados pelo gradiente<sup>4</sup> de uma função de perda diferenciável  $L(y, f(x))$  do seguinte modo:

$$r_i = - \left[ \frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]$$

2. A função de perda geralmente toma a forma abaixo, o que faz com que os resíduos recebam o nome de *pseudo-resíduos*:

---

<sup>3</sup> Não é necessariamente o caso do *LGBM*, algoritmo escolhido para este trabalho. Esta implementação de *gradient boosting* favorece árvores mais profundas, mas também funciona com tocos.

<sup>4</sup> É deste gradiente que vem o nome do algoritmo

$$L(y, f(x)) = \frac{1}{2} \sum_{i=1}^n (y_i - f(x_i))^2$$

3.  $\hat{f}(x)$  é inicializado do seguinte modo:

$$\hat{f}(x) = \arg \min_{\hat{y}} \sum_{i=1}^n L(y_i, \hat{y})$$

4. Nas iterações seguintes,  $\hat{y}$  é computado dessa maneira:

$$\hat{y}_t = \arg \min_{\hat{y}} \sum_{i=1}^n L(y_i, \hat{y}_{t-1} + \lambda \hat{y})$$

Em que  $t$  indica a iteração atual.

### 3.1.5 ARIMA

Para averiguar a eficácia dos modelos apresentados anteriormente, precisamos de algum referencial (*benchmark*). Nesse sentido, o modelo *ARIMA* (do inglês para “modelo autorregressivo integrado de médias móveis”) é uma boa opção devido à sua ampla utilização na literatura.

Modelos autorregressivos univariados são aqueles em que a variável dependente de uma série de tempo é determinada apenas pelos seus próprios valores passados ou com os resíduos (passados ou correntes) da série temporal. Nessa categoria se enquadram os modelos autorregressivos (*AR*), de médias móveis (*MA*), autorregressivos de médias móveis (*ARMA*) e processos integrados mistos (*ARIMA*) (MATOS, 2000<sup>5</sup>, *apud* PADILHA *et al.*, 2013). Desse modo, o *ARIMA* se diferencia dos outros métodos propostos neste trabalho por não se utilizar de variáveis independentes.

O modelo autorregressivo prevê a variável com base nos  $p$  valores passados da série e um termo de erro. Isto é, ele reproduz a correlação intertemporal observada nos dados. Nesse sentido, dada uma série temporal  $X$ , a previsão em  $t$  segundo o método *AR* de ordem  $p$  é descrita por

$$X_t = \varphi_0 + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + e_t$$

---

<sup>5</sup> MATOS, O.C. de. *Econometria Básica: teoria e aplicações*. 3 ed. rev. e ampl. São Paulo: Atlas, 2000.

$$\left(1 - \sum_{i=1}^p \varphi_i L^i\right) X_t = \varphi_0 + e_t$$

em que  $\varphi$  são os parâmetros do modelo,  $e$  é um erro normalmente distribuído e  $L$  é o operador de atraso.

Já o modelo de médias móveis prevê uma determinada variável a partir dos últimos  $q$  termos de erro na série temporal (Arce; Mahía, 2003). Assim, a previsão em  $t$  segundo o método MA é dada por

$$X_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

$$X_t = \mu + \left(1 - \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t$$

em que  $\mu$  é um termo independente,  $\varepsilon$  são os erros e  $\theta$ , os parâmetros.

Juntando ambos, obtemos um modelo *ARMA*( $p$ ,  $q$ )

$$X_t = \mu + \varphi_0 + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} + \dots - \theta_q \varepsilon_{t-q}.$$

$$\left(1 - \sum_{i=1}^p \varphi_i L^i\right) X_t = \mu + \left(1 - \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t$$

Ou em notação polinomial:

$$\varphi_p(L) X_t = \mu + \theta_q(L) \varepsilon_t$$

No entanto, o modelo *AR* (e, conseqüentemente, o *ARMA*) requer uma série estacionária, tornando-se mister alguma transformação no caso de séries não-estacionárias. Para tanto, é comum diferenciar-se a série. Desse modo, obtemos um processo *ARIMA*( $p$ ,  $d$ ,  $q$ ), em que  $d$  denota a ordem de integração (número de vezes que a série foi diferenciada). Assim, o modelo *ARIMA*( $p$ ,  $d$ ,  $q$ ) tem a seguinte forma (Bayer; Souza, 2010):

$$\left(1 - \sum_{i=1}^p \varphi_i L^i\right) (1 - L)^d X_t = \mu + \left(1 - \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t$$

Ou em notação polinomial:

$$\varphi_p(L) (1 - L)^d X_t = \mu + \theta_q(L) \varepsilon_t$$

Também é possível estender o modelo *ARIMA* para incorporar sazonalidades, sendo o modelo resultante frequentemente chamado de *SARIMA* (*Seasonal ARIMA*). Este consiste apenas em estender as características do *ARIMA* para um determinado

intervalo regular (trimestres, anos). Desse modo, um processo  $SARIMA(p, d, q)(P, D, Q)_s$  é descrito por:

$$\varphi_p(L)\Phi_P(L^s)(1-L^s)^D(1-L)^dX_t = \mu + \theta_q(L)\Theta_Q(L^s)\varepsilon_t$$

Em que P, D e Q respectivamente denotam os componentes AR, de integração e de médias móveis sazonais e s denota o período do intervalo regular a ser estudado. A via de exemplo, no caso deste trabalho, em que tratamos da inflação mensal, s deve ser igual a 12 para capturarmos os efeitos anuais. Isso é útil pois diversos fenômenos econômicos possuem padrões sazonais, como é o caso do desemprego, que cresce nos primeiros meses do ano (Carneiro, 2023).

Segundo Stock e Watson (2010), dificilmente um modelo multivariado supera um modelo univariado sistematicamente. Contudo, essa percepção ignora os fenômenos do aprendizado de máquina e da *big data* na Economia (Medeiros *et al.*, 2021).

### 3.2 MÉTRICAS DE COMPARAÇÃO

Podemos nos utilizar de diversas métricas para compararmos a eficácia dos modelos. Seu cálculo geralmente se dá pela diferença entre o valor real observado ( $y$ ) e o valor predito pelo modelo ( $\hat{y}$ ). Utilizaremos o Erro Médio Absoluto (EMA) e o Erro Quadrado Médio (EQM), cujas formulações se encontram abaixo:

$$EMA(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

$$EQM(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Além disso, também utilizaremos o Erro Percentual Absoluto Médio (EPAM):

$$EPAM(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Para facilitar a visualização de quando um determinado modelo está superando o referencial, utilizaremos o *cumulative squared forecast errors* (CSFE) como apresentado por Caldeira e Torrent (2017):

$$CSFE_{m,t} = \sum_{t=1}^T \left[ (\hat{y}_{t+h|t,bench} - y_{t+h})^2 - (\hat{y}_{t+h|t,m} - y_{t+h})^2 \right]$$

Em que  $\hat{y}_{t+h|t,bench}$  é a previsão do modelo de referência (*benchmark*) em  $t$  para  $h$  períodos à frente,  $y_{t+h}$  é o valor real em  $t + h$  e  $\hat{y}_{t+h|t,m}$  é a previsão do modelo  $m$  em  $t$  para  $h$  períodos à frente. O *CSFE* será uma série crescente quando o modelo superar o *benchmark*, e decrescente caso este supere aquele.

Dentre as métricas que utilizaremos, destaca-se o teste de Diebold-Mariano (DM). Introduzido por Diebold e Mariano (1995), a hipótese nula deste teste é de que os valores esperados dos erros de duas séries de previsões são iguais. Isto é, seja  $e_A$  o erro associado à série de previsões  $A$ ,  $e_B$  o erro associado à série de previsões  $B$  e  $L$  uma função de perda (geralmente o quadrado ou o valor absoluto do erro). Temos que:

$$d_{AB} = L(e_A) - L(e_B)$$

E as hipóteses são:

$$H_0: E(d_{AB}) = 0$$

$$H_1: E(d_{AB}) \neq 0$$

É um teste  $Z$  assintótico, extremamente simples e extensível (Diebold, 2015). Portanto, neste trabalho utilizamos uma versão modificada do teste, introduzida por Lago *et al.* (2021), que disponibilizaram sua implementação do teste na biblioteca *epftoolbox* de Python. A diferença é apenas que a hipótese nula é de que os erros da série de previsões  $B$  são maiores que os de  $A$ . Isto é, as hipóteses são:

$$H_0: E(d_{AB}) \leq 0$$

$$H_1: E(d_{AB}) \geq 0$$

Transformando o teste DM em um teste unilateral. A estatística de teste então é calculada por:

$$DM = \sqrt{N} \frac{\hat{\mu}}{\hat{\sigma}}$$

Em que  $\hat{\mu}$  e  $\hat{\sigma}$  são, respectivamente, a média e o desvio padrão da série  $d$ , e  $N$  é a duração do período de teste fora da amostra.

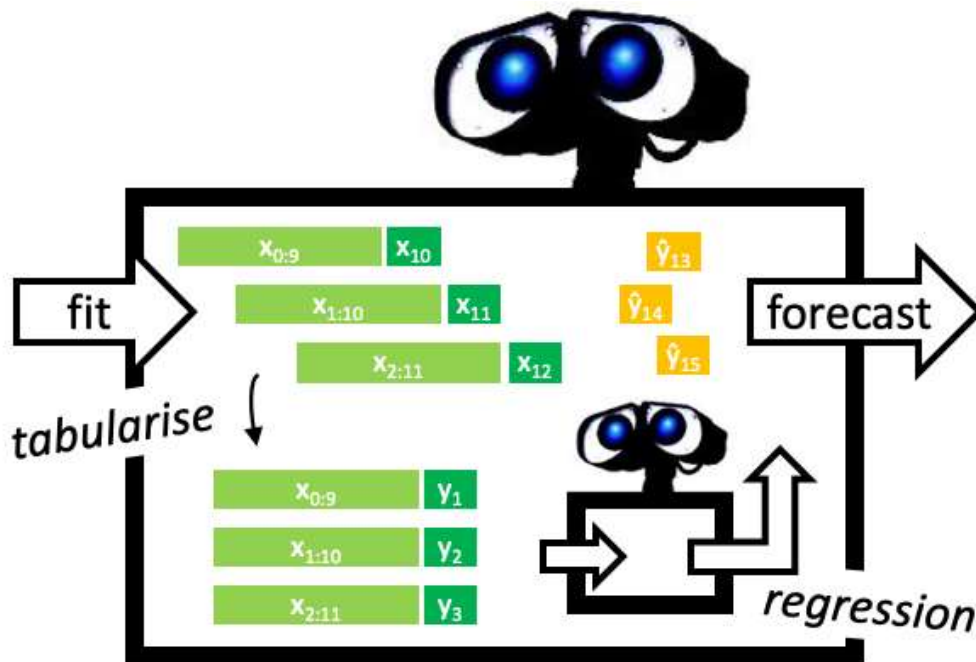
### 3.3 FERRAMENTAS UTILIZADAS

Para executar os modelos, utilizou-se a biblioteca *sktime* da linguagem de programação de *Python*, criada por Löning e Király (2020). Esta biblioteca apresenta uma interface para outras bibliotecas de aprendizado de máquina, dentre as quais se destaca o *scikit-learn*. Desse modo, além de oferecer diversas funções e classes que facilitam o



trato de séries temporais, o *sktime* permite que modifiquemos os modelos tabulares do *scikit-learn* para que fiquem aptos a lidar com séries temporais por meio de um processo chamado “redução”. A imagem abaixo ilustra esse processo:

Figura 9 - Ilustração do processo de redução do sktime



Fonte: SKTIME (2023)

É possível utilizar os modelos do *scikit-learn* para problemas de séries temporais, porém adaptá-los a tal situação necessitaria muito código manual, facilitando a inserção de erros. Além disso, há outros problemas possíveis: a separação da amostra em seções de treino e de teste é mais complicada num cenário de previsões temporais; os estimadores tabulares do *scikit-learn* não possuem parâmetros comuns a previsões de séries temporais (como *lags* e o tamanho da janela apresentada ao estimador) e há complicações para gerar previsões para mais de um passo (Lenczuk, 2021). Desse modo, o processo de redução do *sktime* consiste em converter um problema de previsão num problema de regressão tabular, permitindo assim o uso dos algoritmos do *scikit-learn* e de toda a família de modelos de aprendizado de máquina que usam a mesma API. Para tanto, os dados da série temporal são separados em janelas de tamanho fixo e “empilhadas”, resultando numa matriz de valores atrasados em formato tabular, o que permite o uso de qualquer algoritmo tabular de regressão (Löning; Király, 2020).

O *sktime* disponibiliza diversas estratégias de previsão. Neste trabalho, utilizamos a estratégia “recursiva”, que consiste em utilizar a última janela como entrada para o regressor ajustado ao problema para gerar a primeira previsão. Para a previsão de múltiplos valores à frente, o algoritmo recebe suas últimas previsões recursivamente (Löning; Király, 2020).

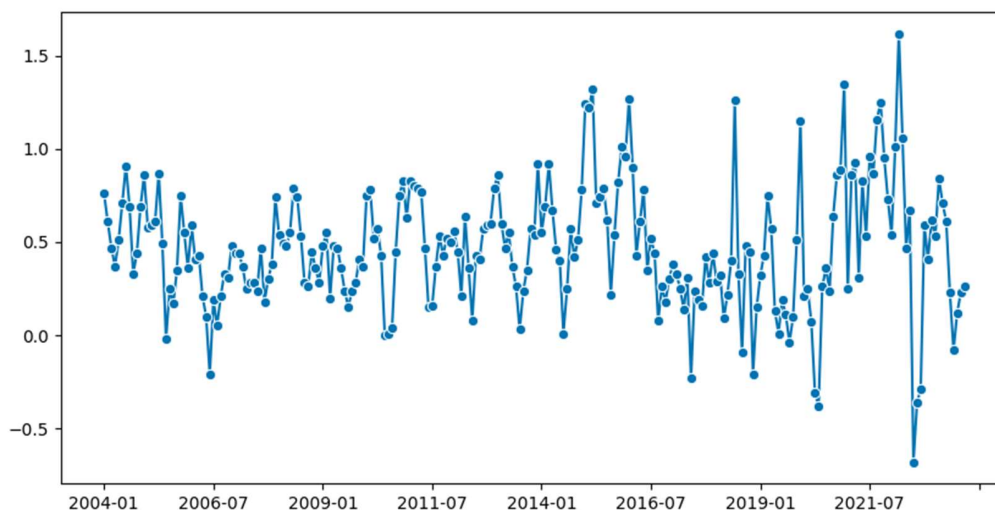
Para o algoritmo de *gradiente boosting*, utilizou-se o *Light Gradient Boosting Machine (LGBM)* da Microsoft, que promete ser mais rápido e mais preciso que o *XGBoost*. O *LightGBM* apresenta a mesma API que os modelos do *scikit-learn*, podendo ser utilizado com as ferramentas do *sktime* após redução.

Para o *ARIMA*, utilizou-se a classe *AutoARIMA* da biblioteca *sktime*, que importa da biblioteca *pmdarima*. Esta, por sua vez, procura trazer para o *Python* a funcionalidade do *auto.arima* do *R*. O modelo selecionado pela classe para todos os horizontes de previsão foi de ordem  $(1, 0, 0)$   $(0, 0, 1)$ .

#### 4 DADOS

Sendo este um trabalho sobre previsão de inflação, utilizamos a variação mensal do IPCA, publicado pelo IBGE, como alvo de nossas previsões. Inspirando-nos em Araújo e Gaglianone (2023), o período de análise é de janeiro de 2004 a setembro de 2023 ( $T = 237$ ), pois assim evitamos a estabilização do Plano Real e o problema de diferentes regimes de política apontado por Machado e Portugal (2014). O gráfico abaixo ilustra a série no período de interesse.

Gráfico 1 - Variação % mensal do IPCA de 2004 a set/2023



Fonte: Elaboração do autor com base em dados do IBGE (2023).

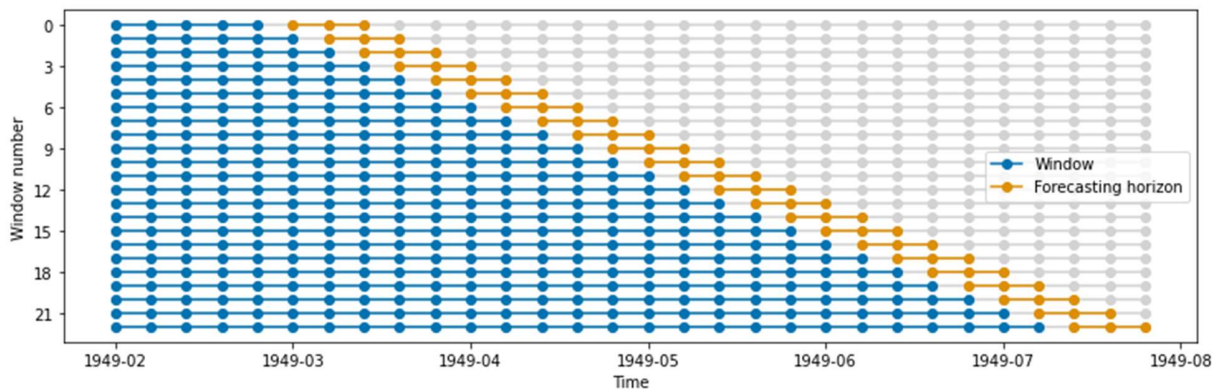
Utilizou-se uma ampla gama de variáveis para as previsões ( $n = 41$ ), incluindo variáveis macroeconômicas (desemprego, PIB, juros), financeiras (Ibovespa), de comércio exterior (câmbio, preços de *commodities*) e da economia real (produção industrial, consumo energético), bem como índices de confiança e de incerteza (Baker; Bloom; Davis, 2016). As principais fontes incluem IBGE, BCB e FGV. Apresentamos no Apêndice A uma lista com as variáveis selecionadas e suas fontes.

A série foi dividida em dois períodos, um para treinar os modelos e outro para testar suas previsões. A divisão foi de exatamente dois terços para treino e um para testes, de modo que os modelos são primeiramente ajustados para o período de janeiro de 2004 a

fevereiro de 2017 ( $T = 158$ ) e testados no período de março de 2017 a setembro de 2023 ( $T = 79$ ).

Para alimentar os modelos, utilizou-se a estratégia de *janela crescente* (*expanding window*), ilustrada na figura abaixo. O modelo é ajustado para uma determinada janela inicial e realiza a sua primeira previsão. Posteriormente, para as próximas previsões, novos dados vão sendo adicionados à janela de treino, e o modelo é reajustado recursivamente para incorporá-los, como em Araújo e Gaglianone (2023).

Figura 10 - Estratégia de janela crescente



Fonte: Sktime (2023)

As previsões foram realizadas para os períodos  $h = 1, 2, 3, 6$  e  $12$ . Seguindo a estratégia de redução apresentada anteriormente, os modelos foram alimentados com uma janela de 12 períodos, isto é, com as variáveis em  $t$  e 11 defasagens.

Para garantir a estacionariedade das séries, adotou-se um procedimento semelhante ao de Araújo e Gaglianone (2023). Consideraram-se três possibilidades para as séries: (1) sem transformação; (2)  $\Delta x_t$ ; (3)  $\Delta^2 x_t$ . A transformação realizada em cada série está evidenciada no Apêndice A. Após as transformações, realizou-se o teste Kwiatkowski–Phillips–Schmidt–Shin (KPSS) com nível de significância de 5% para garantir a estacionariedade das séries.

Algumas séries não estavam disponíveis para todo o período, como é o caso do desemprego e da massa real de rendimentos, ambos publicados pela Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua) do IBGE, que só começa a ser publicada em 2012. Para o caso do desemprego, aplicou-se aos dados da PNAD Contínua anteriores a 2012 a variação mensal na desocupação segundo a Pesquisa

Mensal de Emprego (PME), também do IBGE e extinta em fevereiro de 2016. Para a massa de rendimentos, utilizou-se o mesmo procedimento, porém tendo por base a variação mensal do rendimento médio real efetivo das pessoas ocupadas segundo o BCB.

Para a Selic, calculou-se a média no mês da taxa anual. Para as séries de consumo energético, que estão disponíveis apenas a partir de janeiro de 2004 (e, deste modo, não seria possível tirar a diferença para garantir a estacionariedade nem aplicar atrasos posteriormente), imputaram-se os dados no período setembro a dezembro de 2003 por meio de regressão linear. Acreditamos que isso não deve ter um impacto relevante nos dados por se tratar de um período curto.

No caso do *LASSO* e da *RR*, é necessário antes normalizarmos os dados, isto é, transformá-los de modo que fiquem com média igual a zero e desvio-padrão igual a um. Isso se deve ao fato deles serem medidos em unidades diferentes. Dado que os parâmetros estão restritos a um determinado  $\lambda$ , as variáveis em escalas grandes (digamos, milhões, como em variáveis de base monetária) serão pouco regularizadas quando comparadas a variáveis medidas em escalas menores (digamos, variação mensal de desemprego). De fato, no artigo em que introduz o modelo, Tibshirani presume que os dados estão normalizados (Tibshirani, 1996). Por fim, as previsões resultantes devem ser desnormalizadas posteriormente.

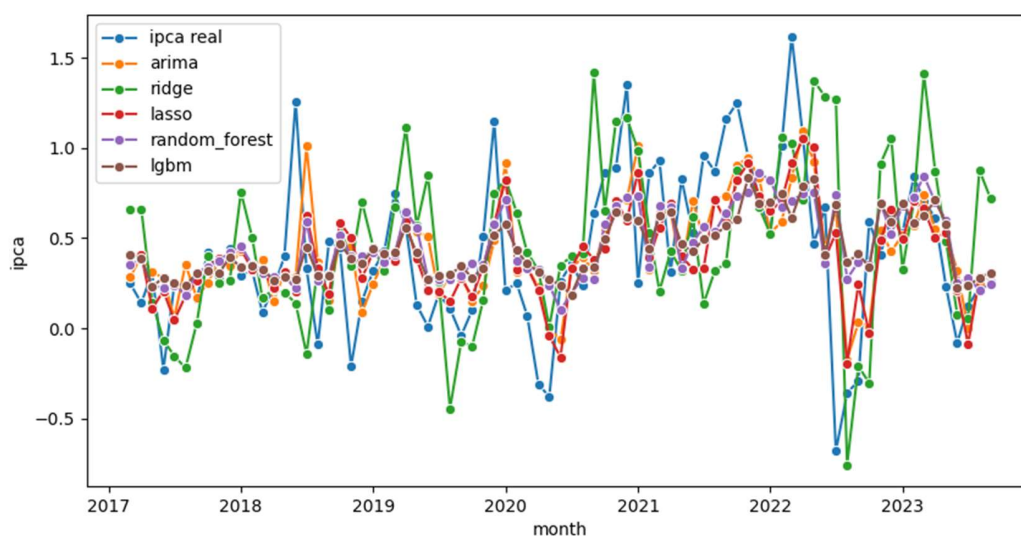
## 5 RESULTADOS

Nesta seção, apresentamos os resultados dos modelos para os cinco horizontes de previsão estudados ( $h = 1, 2, 3, 6, 12$ ). Inspiramos nossa apresentação em Sager (2023).

### 5.1 RESULTADOS DOS MODELOS

A Tabela 1 expõe as métricas de erro para cada um dos modelos, sombreando-se os resultados que superam o modelo de referência (*ARIMA*). A Tabela 2 apresenta os resultados do teste de Diebold-Mariano para a comparação entre cada um dos modelos e o *ARIMA*, sombreando-se os resultados dentro de um intervalo de confiança de 5%. Além disso, o melhor resultado para cada métrica e em cada horizonte foi destacado em amarelo. Os valores em ambas as tabelas foram arredondados para quatro casas decimais. No Apêndice B estão as séries de *CSFE* para todos os modelos por horizonte.

Gráfico 2 – IPCA real e previsões para  $h = 1$

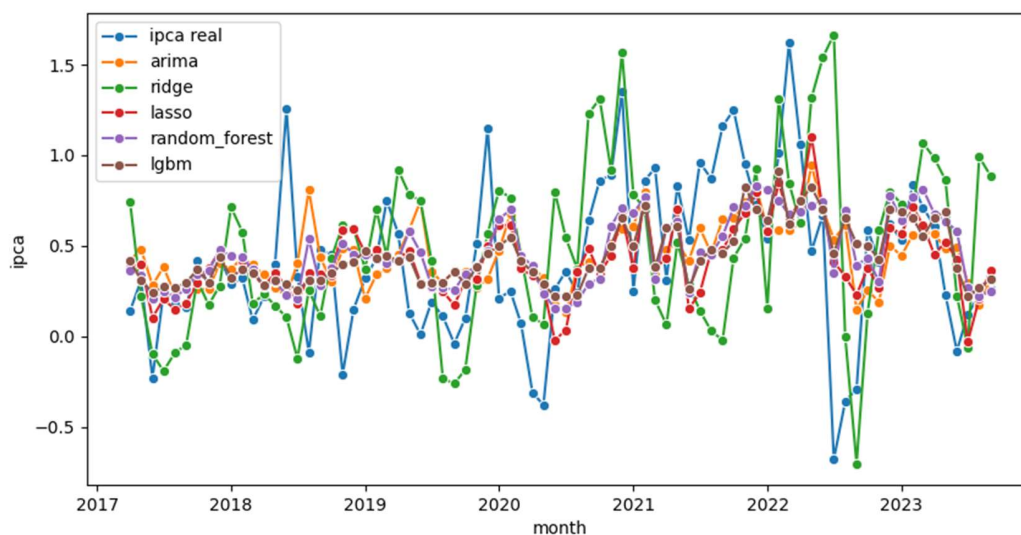


Fonte: Elaboração própria

É interessante observar que a RR apresentou resultados inferiores ao *ARIMA* para todos os horizontes e todas as métricas, excetuando-se o EPAM para os dois horizontes mais longos ( $h = 6$  e  $h = 12$ ). Em contrapartida, o *LASSO* obteve resultados superiores em todas as métricas e para todos os horizontes, exceto o EQM e o EMA para  $h = 6$ . É possível que essa diferença se deva à capacidade do *LASSO* de penalizar certas

variáveis até o zero, isto é, de selecionar variáveis, o que indicaria que algumas das séries selecionadas não são úteis para a previsão.

Gráfico 3 - IPCA real e previsões para  $h = 2$



Fonte: Elaboração própria

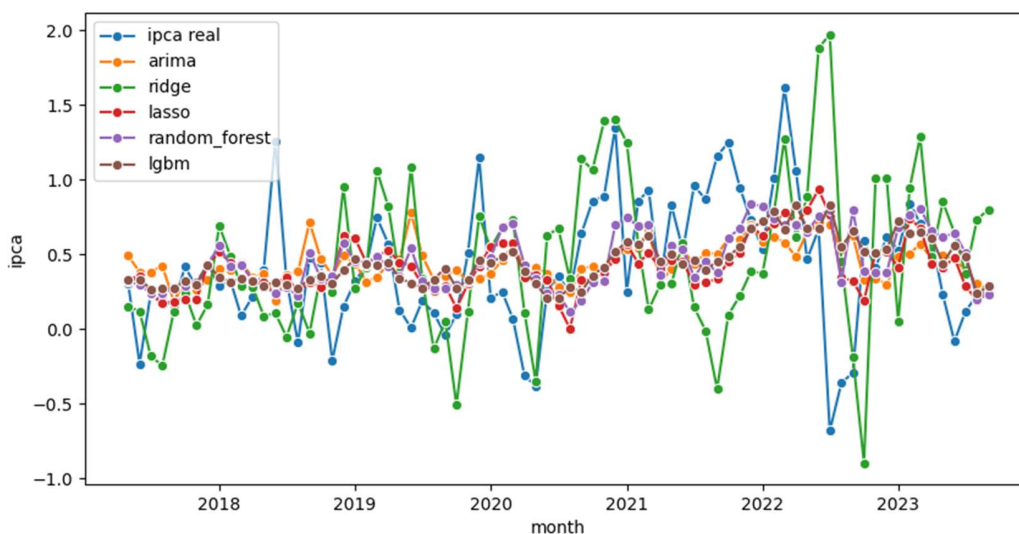
Os outros modelos também conseguiram resultados positivos: de modo semelhante ao *LASSO*, o *Random Forest* retornou erros menores para todas as métricas nos horizontes  $h = 2$ ,  $h = 3$  e  $h = 12$ . Para  $h = 1$  e  $h = 6$ , o modelo superou o referencial apenas em EPAM, obtendo erros maiores nas outras métricas.

Tabela 1 - Métricas de erro para os cinco horizontes de previsão

Métrica	Modelo	h = 1	h = 2	h = 3	h = 6	h = 12
EMA	ARIMA	0,2817	0,3243	0,3297	0,3429	0,3658
	RR	0,3754	0,4243	0,4616	0,4470	0,4866
	LASSO	0,2718	0,2927	0,3201	0,3638	0,3474
	Random Forest	0,2892	0,3069	0,3173	0,3474	0,3625
	LightGBM	0,2824	0,2850	0,2985	0,3268	0,3576
EQM	ARIMA	0,1438	0,1798	0,1944	0,1987	0,2192
	RR	0,2392	0,2930	0,3888	0,4283	0,3823
	LASSO	0,1332	0,1501	0,1798	0,2357	0,2001
	Random Forest	0,1442	0,1620	0,1865	0,2150	0,2191
	LightGBM	0,1446	0,1534	0,1746	0,2014	0,2078
EPAM	ARIMA	1,5908	2,1089	2,1358	2,1630	2,3804
	RR	2,1893	2,2689	2,7965	2,1511	2,0519
	LASSO	1,1619	1,3373	1,5845	1,9103	1,4730
	Random Forest	1,3054	1,6774	1,8180	1,6891	2,0886
	LightGBM	1,2704	1,3502	1,4434	1,5544	1,5791

Fonte: Elaboração própria

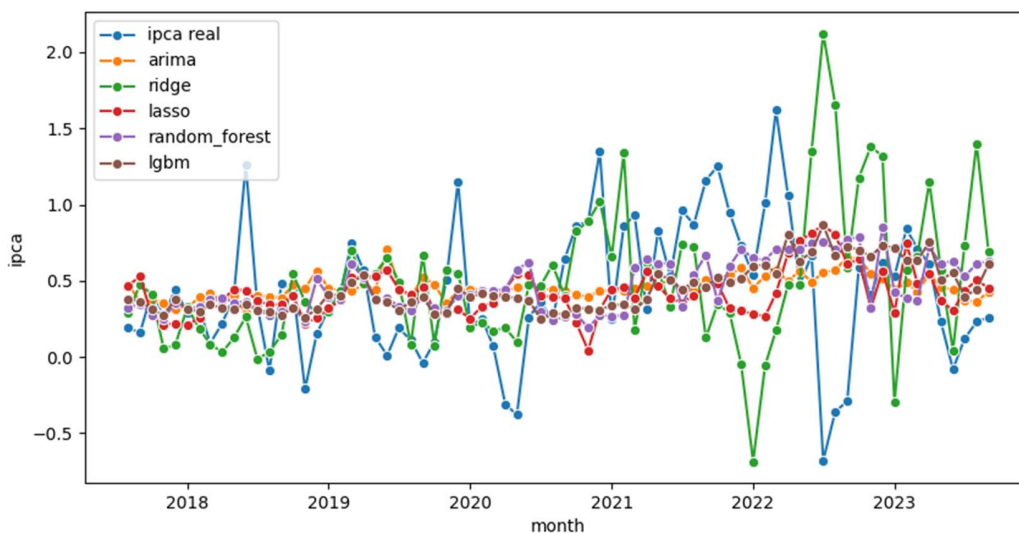
*LightGBM*, o outro modelo baseado em árvores de decisão, obteve resultados semelhantes aos do *Random Forest*: superioridade em todas as métricas de erro para  $h = 2$ ,  $h = 3$  e  $h = 12$ ; apenas EPAM superior para  $h = 1$  e EPAM e EMA superior para  $h = 6$ .

Gráfico 4 - IPCA real e previsões para  $h = 3$ 



Fonte: Elaboração própria

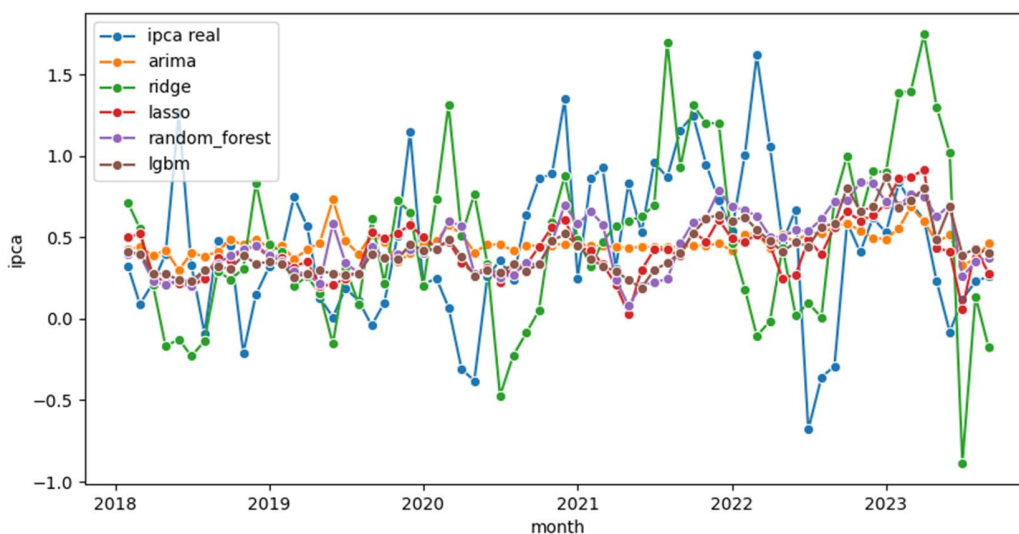
Gráfico 5 - IPCA real e previsões para  $h = 6$



Fonte: Elaboração própria

Ao analisarmos os melhores resultados por métrica e em cada horizonte, chama a atenção o fato de que nem a RR nem o *Random Forest* obtiveram algum deles. O *LASSO* e o *LightGBM* monopolizaram essas graças, com a exceção do melhor EQM para  $h = 6$ , conquistado pelo *ARIMA*.

Gráfico 6 - IPCA real e previsões para  $h = 12$



Fonte: Elaboração própria

Tabela 2 - Valores-p para o teste de Diebold-Mariano

Modelo	h = 1	h = 2	h = 3	h = 6	h = 12
RR	0,9995	0,9987	0,9987	0,9882	0,9965
LASSO	0,1865	0,0147	0,2569	0,9191	0,1356
Random Forest	0,6785	0,0700	0,1833	0,6208	0,4170
LightGBM	0,5152	0,0025	0,0066	0,1091	0,2952

Fonte: Elaboração própria

No que tange ao teste de Diebold-Mariano, apenas o *LASSO* e o *LightGBM* apresentaram algum valor-p inferior a 5%: para aquele, apenas em  $h = 2$ ; para este, em  $h = 2$  e  $h = 3$ . Também é interessante observar que as previsões do *Random Forest* foram significativamente superiores às do *ARIMA* no horizonte  $h = 2$  se considerarmos um nível de significância de 10%.

Em suma, os modelos de aprendizado de máquina apresentam resultados superiores aos do *ARIMA* para a maior parte dos horizontes temporais (com a exceção da RR, que se mostrou inferior em todos os horizontes). Contudo, a vitória não foi absoluta: apenas dois modelos apresentaram resultados significativamente superiores segundo o teste de Diebold-Mariano para um nível de significância de 5%, e mesmo estes para poucos horizontes (um no caso do *LASSO* e dois no caso do *LGBM*).

## 5.2 VARIÁVEIS SELECIONADAS

Nesta seção, apresentamos as cinco principais variáveis selecionadas por cada modelo para cada horizonte. O objetivo deste exercício é tentar evidenciar as correlações encontradas pelos modelos entre a inflação e as variáveis exógenas. Isso é particularmente interessante para modelos de aprendizado conjunto que utilizam árvores, pois são considerados de difícil interpretação (James *et al.*, 2023e).

Para tanto, atribuímos pesos percentuais para cada variável (ignorando-se as diferentes defasagens). No caso da RR e do *LASSO*, somaram-se os valores absolutos de todos os coeficientes, e o peso de cada variável é sua porcentagem deste total. É possível fazer isso pois as variáveis foram todas normalizadas; do contrário, os diversos pesos seriam meramente reflexos das escalas diferentes das variáveis. Para os métodos baseados em árvores (*RF* e *LGBM*), utiliza-se a redução total da métrica de erro

selecionada por variável (conhecida como importância de Gini) (Sci-kit learn, [c2024]). Apresentamos esses resultados em cinco tabelas, uma para cada horizonte.

Para o primeiro horizonte, vê-se que todos os modelos selecionaram índices de receita nominal do varejo e do varejo ampliado. Outras variáveis comuns (mas de uso não tão amplo) foram o dólar e o próprio IPCA. Além disso, todos os modelos (com exceção da RR) deram grande peso a alguma métrica de incerteza da política econômica, seja para o Brasil (como foi o caso dos métodos baseados em árvore), seja para os EUA (como no caso do *LASSO*), e o *LASSO* foi o único a ter séries de consumo energético entre suas cinco variáveis mais relevantes. Por fim, vê-se o poder da capacidade de seleção de variáveis do *LASSO*: comparado à RR, os pesos das cinco variáveis mais relevantes estão bem maiores. Esse fenômeno se repete para os demais horizontes.

Tabela 3 - Cinco variáveis de maior peso para  $h = 1$  por modelo

Variável	Peso
<b>Regressão Ridge</b>	
Índice da receita nominal do varejo	5,99%
IPCA	5,20%
Taxa de câmbio BRL/USD	4,57%
Índice da receita nominal do varejo ampliado	4,10%
M1	2,91%
<b>LASSO</b>	
Índice da receita nominal do varejo	20,70%
Índice da receita nominal do varejo ampliado	18,99%
<i>Economic Policy Uncertainty Index USA</i>	8,31%
Consumo de energia total	7,16%
Consumo de energia comercial	6,21%
<b>Random Forest</b>	
Índice da receita nominal do varejo	16,67%
Índice da receita nominal do varejo ampliado	15,35%
IPCA	3,89%
<i>Economic Policy Uncertainty Index Brazil</i>	3,43%
Taxa de câmbio BRL/USD	3,21%
<b>LGBM</b>	
Índice da receita nominal do varejo	8,29%
Índice da receita nominal do varejo ampliado	7,01%
<i>Economic Policy Uncertainty Index Brazil</i>	4,91%
IPCA	4,32%
Taxa de câmbio BRL/USD	4,10%

Fonte: Elaboração própria

Para  $h = 2$ , percebe-se que as variáveis se mantêm estáveis de modo geral, algo que se repetirá nos próximos horizontes. Para dois dos modelos (*LASSO* e *RF*), a ordem das variáveis se manteve idêntica. Para a *RR*, o M1 foi substituído pelo M2. Para o *LGBM*, o dólar foi substituído por uma métrica de incerteza: o índice de confiança do consumidor (ICC).

Tabela 4 - Cinco variáveis de maior peso para  $h = 2$  por modelo

Variável	Peso
<b>Regressão Ridge</b>	
Índice da receita nominal do varejo	6,10%
IPCA	5,23%
Taxa de câmbio BRL/USD	4,70%
Índice da receita nominal do varejo ampliado	4,15%
M2	2,88%
<b>LASSO</b>	
Índice da receita nominal do varejo	19,97%
Índice da receita nominal do varejo ampliado	18,13%
<i>Economic Policy Uncertainty Index USA</i>	8,66%
Consumo de energia total	6,81%
Consumo de energia comercial	6,04%
<b>Random Forest</b>	
Índice da receita nominal do varejo	16,36%
Índice da receita nominal do varejo ampliado	14,99%
IPCA	4,01%
<i>Economic Policy Uncertainty Index Brazil</i>	3,68%
Taxa de câmbio BRL/USD	3,05%
<b>LGBM</b>	
Índice da receita nominal do varejo	8,61%
Índice da receita nominal do varejo ampliado	7,26%
<i>Economic Policy Uncertainty Index Brazil</i>	5,06%
IPCA	3,78%
Índice de confiança do consumidor	3,71%

Fonte: Elaboração própria

Para  $h = 3$ , novamente a situação se mantém basicamente a mesma. A principal diferença é que, no *LGBM*, o IPCA perdeu sua posição para o ICC, e o quinto lugar foi tomado pelo preço do barril de petróleo Brent.

Tabela 5 - Cinco variáveis de maior peso para  $h = 3$  por modelo

Variável	Peso
<b>Regressão Ridge</b>	
Índice da receita nominal do varejo	6,07%
IPCA	5,24%
Taxa de câmbio BRL/USD	4,69%
Índice da receita nominal do varejo ampliado	4,14%
M2	2,87%
<b>LASSO</b>	
Índice da receita nominal do varejo	20,31%
Índice da receita nominal do varejo ampliado	18,75%
<i>Economic Policy Uncertainty Index USA</i>	9,51%
Consumo de energia total	6,95%
Consumo de energia comercial	6,21%
<b>Random Forest</b>	
Índice da receita nominal do varejo	16,29%
Índice da receita nominal do varejo ampliado	15,04%
IPCA	4,08%
<i>Economic Policy Uncertainty Index Brazil</i>	3,49%
Taxa de câmbio BRL/USD	3,15%
<b>LGBM</b>	
Índice da receita nominal do varejo	9,40%
Índice da receita nominal do varejo ampliado	7,76%
<i>Economic Policy Uncertainty Index Brazil</i>	5,02%
Índice de confiança do consumidor	3,65%
Preço do barril de petróleo Brent	3,64%

Fonte: Elaboração própria

Para  $h = 6$ , o M1 retoma sua posição do M2 na RR. Ademais, o ICC perde sua posição para o dólar no *LGBM*.

Tabela 6 - Cinco variáveis de maior peso para  $h = 6$  por modelo

Variável	Peso
<b>Regressão Ridge</b>	
Índice da receita nominal do varejo	6,08%
IPCA	5,11%
Taxa de câmbio BRL/USD	4,70%
Índice da receita nominal do varejo ampliado	4,03%
M1	2,96%
<b>LASSO</b>	
Índice da receita nominal do varejo	21,90%
Índice da receita nominal do varejo ampliado	20,05%
<i>Economic Policy Uncertainty Index USA</i>	10,24%
Consumo de energia total	7,23%
Consumo de energia comercial	6,37%
<b>Random Forest</b>	
Índice da receita nominal do varejo	16,04%
Índice da receita nominal do varejo ampliado	14,85%
IPCA	4,16%
<i>Economic Policy Uncertainty Index Brazil</i>	3,86%
Taxa de câmbio BRL/USD	3,02%
<b>LGBM</b>	
Índice da receita nominal do varejo	9,10%
Índice da receita nominal do varejo ampliado	7,05%
<i>Economic Policy Uncertainty Index Brazil</i>	5,92%
Taxa de câmbio BRL/USD	4,17%
Preço do barril de petróleo Brent	3,70%

Fonte: Elaboração própria

Para  $h = 12$ , há um pouco mais de movimento: na RR, a receita nominal do varejo ampliado e o dólar trocam de posição. No LGBM, o preço do barril de petróleo Brent é substituído pelo IPCA. Por fim, vemos a primeira mudança no LASSO: o consumo de energia total e comercial trocam de lugar, e aquele é substituído pelo M1. Apenas o RF mantém sua ordem intacta para todos os horizontes.

Tabela 7 - Cinco variáveis de maior peso para h = 12 por modelo

Variável	Peso
<b>Regressão Ridge</b>	
Índice da receita nominal do varejo	6,44%
IPCA	5,34%
Índice da receita nominal do varejo ampliado	4,09%
Taxa de câmbio BRL/USD	3,93%
M1	3,04%
<b>LASSO</b>	
Índice da receita nominal do varejo	19,80%
Índice da receita nominal do varejo ampliado	18,10%
<i>Economic Policy Uncertainty Index USA</i>	10,14%
Consumo de energia comercial	6,02%
M1	5,00%
<b>Random Forest</b>	
Índice da receita nominal do varejo	16,46%
Índice da receita nominal do varejo ampliado	15,17%
IPCA	4,01%
<i>Economic Policy Uncertainty Index Brazil</i>	3,35%
Taxa de câmbio BRL/USD	2,91%
<b>LGBM</b>	
Índice da receita nominal do varejo	8,70%
Índice da receita nominal do varejo ampliado	7,27%
<i>Economic Policy Uncertainty Index Brazil</i>	4,87%
Taxa de câmbio BRL/USD	4,58%
IPCA	3,54%

Fonte: Elaboração própria

Apesar das (poucas) variações, é possível perceber um certo padrão entre os modelos: dão-se maiores pesos a variáveis de atividade econômica (índice de receita nominal do varejo e do varejo ampliado), ao próprio IPCA, à taxa de câmbio (apenas com o dólar) e a índices de confiança ou incerteza.

## 6 CONCLUSÃO

O objetivo deste trabalho é comparar a capacidade de previsão da inflação de quatro modelos de aprendizado de máquina (*RR*, *LASSO*, *RF* e *LGBM*) ao *ARIMA*, numa tentativa de incorporar os avanços na área da inteligência artificial e o fenômeno de *big data* à previsão de inflação. O tema é relevante simplesmente do ponto de vista teórico, mas também é pertinente devido às aplicações práticas para agentes econômicos (firmas e indivíduos), pois expectativas de inflação balizam a indexação de contratos de emprego e de aluguel, para investidores e instituições financeiras, pois precisam de previsões precisas para precificar corretamente ativos, e para banqueiros centrais, pois é valioso ter alguma previsão para a inflação ao se conduzir política monetária.

A pesquisa atinge seu objetivo de comparar os resultados das previsões de métodos de aprendizado de máquina às do *ARIMA*. Nossos resultados demonstram que os métodos de aprendizado de máquina obtêm resultados superiores aos do referencial (*ARIMA*), mas não significativamente superiores para a maior parte dos horizontes e para nenhum horizonte em alguns dos modelos. Assim, apesar desses novos modelos serem adições interessantes ao ferramental de econométricos, o *ARIMA* não pode ser esquecido.

Ademais, dentre as cinco variáveis com maior peso, os modelos de aprendizado de máquina tenderam a selecionar variáveis relacionadas à atividade econômica (representada pela receita do varejo e do varejo ampliado), a moeda (taxa de câmbio real/dólar e indicadores de base monetária), indicadores de incerteza e de confiança (como EPU e ICC) e o próprio IPCA. Esses dados podem ser úteis na construção de bases de dados que busquem replicar os resultados no futuro.

Possíveis extensões deste trabalho incluem rodar o mesmo experimento para um índice de inflação que exclua itens voláteis como alimentos e energia (semelhante ao *core CPI* dos EUA) e o uso de outros modelos de aprendizado de máquina. Dentre estes, recomenda-se utilizar modelos de aprendizado profundo (*deep learning*), aplicações mais modernas de RNAs que prometem capturar melhor não-linearidades e lidar melhor com grandes bases de dados (alinhando-se perfeitamente com o fenômeno de *big data*). Além disso, poder-se-ia simular uma carteira de investimentos que siga uma determinada



estratégia de compra ou venda de títulos públicos indexados à inflação com base em previsões de inflação geradas por modelos de aprendizado de máquina.

## REFERÊNCIAS

- 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**, Amsterdã, v. 4, n. 2, p. 100087, 2023.
- ARCE, R. de; MAHÍA, R. Modelos ARIMA. **Programa CITUS: Técnicas de Variables Financieras**, Madri, p. 31, 2003.
- ASOKAN, A. P. Nvidia's AI Revolution. Story behind How it Became a Trillion-Dollar Company. *In*: MEDIUM. 30 maio 2023. Disponível em: <https://medium.com/@ArunPrakashAsokan/nvidias-ai-revolution-story-behind-how-it-became-a-trillion-dollar-company-620e7235fdc>. Acesso em: 26 nov. 2023.
- ATHEY, S. The Impact of Machine Learning on Economics. *In*: AGRAWAL, A.; GANS, J.; GOLDFARB, A. (ed.). **The Economics of Artificial Intelligence: An Agenda**. Chicago: University of Chicago Press, 2019. p. 507–547.
- BAKER, S. R.; BLOOM, N.; DAVIS, S. J. Measuring Economic Policy Uncertainty. **The Quarterly Journal of Economics**, Estados Unidos, v. 131, n. 4, p. 1593–1636, 2016.
- BAYER, F. M.; SOUZA, A. M. Wavelets e Modelos Tradicionais de Previsão: um Estudo Comparativo. **Revista Brasileira de Biometria**, São Paulo, v. 28, n. 5, p. 40–61, 2010.
- BREIMAN, L. (org.). **Classification and regression trees**. 1. CRC Press reprod. Boca Raton, Fla.: Chapman & Hall/CRC, 1998. Disponível em: <https://doi.org/10.1201/9781315139470>.
- BREIMAN, L. Random Forests. **Machine Learning**, Estados Unidos, v. 45, n. 1, p. 5–32, 2001.
- CALDEIRA, J.; TORRENT, H. Forecasting the US Term Structure of Interest Rates Using Nonparametric Functional Data Analysis. **Journal of Forecasting**, [s. l.], v. 36, n. 1, p. 56–73, 2017.
- CARBONELL, J. G. Introduction: Paradigms for machine learning. **Artificial Intelligence**, Países Baixos, v. 40, n. 1–3, p. 1–9, 1989.
- CARNEIRO, L. **Desemprego retoma padrão sazonal, mas também pode ter efeito de desaceleração, diz IBGE**. Rio de Janeiro, 2023. Disponível em: <https://valor.globo.com/brasil/noticia/2023/04/28/desemprego-retoma-padro-sazonal-mas-tambm-pode-ter-efeito-de-desacelerao-diz-ibge.ghtml>. Acesso em: 27 dez. 2023.
- CLARK, B. D.; GOLDBERG, E. M.; RUDY, B. Electrogenic Tuning of the Axon Initial Segment. **The Neuroscientist**, Estados Unidos, v. 15, n. 6, p. 651–668, 2009.

DIEBOLD, F. X. Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests. **Journal of Business & Economic Statistics**, Estados Unidos, v. 33, n. 1, p. 1–9, 2015.

DIEBOLD, F. X.; MARIANO, R. S. Comparing Predictive Accuracy. **Journal of Business & Economic Statistics**, Estados Unidos, v. 13, n. 3, p. 253–263, 1995.

DING, P. **Linear Model and Extensions**. 1 jan. 2024. Manuscrito submetido para publicação. Disponível em: <https://arxiv.org/abs/2401.00649>. Acesso em: 5 jan. 2024.

FORECASTING WITH SKTIME — SKTIME DOCUMENTATION. [S. l.], [c2024]. Disponível em: [https://www.sktime.net/en/latest/examples/01\\_forecasting.html](https://www.sktime.net/en/latest/examples/01_forecasting.html). Acesso em: 25 dez. 2023.

FRIEDMAN, J. H. Stochastic gradient boosting. **Computational Statistics & Data Analysis**, Países Baixos, v. 38, n. 4, p. 367–378, 2002.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. H. **The elements of statistical learning: data mining, inference, and prediction**. Second edition, corrected at 12th printing 2017ed. New York, NY: Springer, 2017. (Springer series in statistics).

HOERL, A. E.; KENNARD, R. W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. **Technometrics**, Estados Unidos, v. 42, n. 1, p. 80–86, 2000.

JAMES, G. *et al.* Deep Learning. *In*: JAMES, G. *et al.* **An Introduction to Statistical Learning**. Cham: Springer International Publishing, 2023a. (Springer Texts in Statistics). p. 399–467. Disponível em: [https://link.springer.com/10.1007/978-3-031-38747-0\\_10](https://link.springer.com/10.1007/978-3-031-38747-0_10). Acesso em: 26 nov. 2023.

JAMES, G. *et al.* Linear Model Selection and Regularization. *In*: JAMES, G. *et al.* **An Introduction to Statistical Learning**. Cham: Springer International Publishing, 2023b. (Springer Texts in Statistics). p. 229–288. Disponível em: [https://link.springer.com/10.1007/978-3-031-38747-0\\_6](https://link.springer.com/10.1007/978-3-031-38747-0_6). Acesso em: 20 nov. 2023.

JAMES, G. *et al.* Resampling Methods. *In*: JAMES, G. *et al.* **An Introduction to Statistical Learning**. Cham: Springer International Publishing, 2023c. (Springer Texts in Statistics). p. 201–228. Disponível em: [https://link.springer.com/10.1007/978-3-031-38747-0\\_5](https://link.springer.com/10.1007/978-3-031-38747-0_5). Acesso em: 26 dez. 2023.

JAMES, G. *et al.* Statistical Learning. *In*: JAMES, G. *et al.* **An Introduction to Statistical Learning**. Cham: Springer International Publishing, 2023d. (Springer Texts in Statistics). p. 15–67. Disponível em: [https://link.springer.com/10.1007/978-3-031-38747-0\\_2](https://link.springer.com/10.1007/978-3-031-38747-0_2). Acesso em: 25 nov. 2023.

JAMES, G. *et al.* Tree-Based Methods. *In*: JAMES, G. *et al.* **An Introduction to Statistical Learning**. Cham: Springer International Publishing, 2023e. (Springer Texts in Statistics). p. 331–366. Disponível em: [https://link.springer.com/10.1007/978-3-031-38747-0\\_8](https://link.springer.com/10.1007/978-3-031-38747-0_8). Acesso em: 20 nov. 2023.

JUNG, J.-K.; PATNAM, M.; TER-MARTIROSYAN, A. **An Algorithmic Crystal Ball: Forecasts-Based on Machine Learning**. Rochester, NY, 1 nov. 2018. SSRN Scholarly Paper. Disponível em: <https://papers.ssrn.com/abstract=3297651>. Acesso em: 25 nov. 2023.

LAGO, J. *et al.* Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. **Applied Energy**, Reino Unido, v. 293, p. 116983, 2021.

LENCZUK, J. **Why start using sktime for forecasting?**. Toronto, 2021. Disponível em: <https://towardsdatascience.com/why-start-using-sktime-for-forecasting-8d6881c0a518>. Acesso em: 25 dez. 2023.

LONDON, S. *et al.* **Ask the AI experts: What's driving today's progress in AI? | McKinsey**. Londres, 2017. Disponível em: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/ask-the-ai-experts-whats-driving-todays-progress-in-ai>. Acesso em: 26 nov. 2023.

LÖNING, M.; KIRÁLY, F. **Forecasting with sktime: Designing sktime's New Forecasting API and Applying It to Replicate and Extend the M4 Study**. 9 jun. 2020. Manuscrito submetido para publicação. Disponível em: <http://arxiv.org/abs/2005.08067>. Acesso em: 25 dez. 2023.

MACHADO, V. D. G.; PORTUGAL, M. S. Measuring inflation persistence in Brazil using a multivariate model. **Revista Brasileira de Economia**, Rio de Janeiro, v. 68, n. 2, p. 225–241, 2014.

MATOS, O. C. de. **Econometria básica: teoria e aplicações**. 3 ed. rev. e amplad. São Paulo: Atlas, 2000.

MEDEIROS, M. C. *et al.* Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods. **Journal of Business & Economic Statistics**, Estados Unidos, v. 39, n. 1, p. 98–119, 2021.

MEDEIROS, M. C.; VASCONCELOS, G.; FREITAS, E. Forecasting Brazilian Inflation with High-Dimensional Models. **Brazilian Review of Econometrics**, Rio de Janeiro, v. 36, n. 2, p. 223–254, 2016.

PADILHA, J. B. J. *et al.* Modelagem Box-Jenkins (ARIMA) para Gestão de Risco de Preço na Pecuária de Corte do Estado do Paraná. *In*: CONVIBRA – CONGRESSO VIRTUAL BRASILEIRO DE ADMINISTRAÇÃO, X., 2013, São Paulo. **Anais [...]**. São Paulo: [s. n.], 2013. p. 1–12. Disponível em: [https://convibra.org/congresso/res/uploads/pdf/2013\\_30\\_8283.pdf](https://convibra.org/congresso/res/uploads/pdf/2013_30_8283.pdf). Acesso em: 23 nov. 2023.

PORTUGAL, M. S. Neural networks versus time series methods: a forecasting exercise. **Revista Brasileira de Economia**, Rio de Janeiro, v. 49, n. 4, p. 611–629, 1995.

PORTUGAL, M. S.; FERNANDES, L. G. L. Redes Neurais Artificiais e Previsão de Séries Econômicas: Uma Introdução. **Nova Economia**, Belo Horizonte, v. 6, n. 1, p. 51–74, 1996.

ROUGAS, S. What is an Expert System and What is Used For? *In*: EXPERTISE FINDER. 22 jul. 2020. Disponível em: <https://expertisefinder.com/expert-systems/>. Acesso em: 25 nov. 2023.

SAGER, F. T. **Previsão da Taxa de Crescimento do PIB do Brasil: Uma Comparação entre Machine Learning e Modelos ARIMA**. 2023. Trabalho de conclusão de curso (Graduação em Ciências Econômicas) - Faculdade de Ciências Econômicas, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2023. Disponível em: [https://www.lume.ufrgs.br/handle/10183/258451?locale-attribute=en&trk=public\\_post\\_comment-text](https://www.lume.ufrgs.br/handle/10183/258451?locale-attribute=en&trk=public_post_comment-text). Acesso em: 9 jun. 2023.

SARGENT, T. J. Sources of Artificial Intelligence. [s. l.], 2023. Disponível em: [http://www.tomsargent.com/research/AI\\_Sargent.pdf](http://www.tomsargent.com/research/AI_Sargent.pdf). Acesso em: 15 nov. 2023.

SCI-KIT LEARN. **Sklearn.ensemble.RandomForestRegressor**. [S. l.], [c2024]. Disponível em: <https://scikit-learn/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>. Acesso em: 11 jan. 2024.

STOCK, J.; WATSON, M. **Modeling Inflation After the Crisis**. Cambridge, MA: National Bureau of Economic Research, 2010. Disponível em: <http://www.nber.org/papers/w16488.pdf>. Acesso em: 23 nov. 2023.

THEOHARIDIS, A. F.; GUILLÉN, D. A.; LOPES, H. Deep learning models for inflation forecasting. **Applied Stochastic Models in Business and Industry**, Reino Unido, v. 39, n. 3, p. 447–470, 2023.

TIBSHIRANI, R. Regression Shrinkage and Selection Via the Lasso. **Journal of the Royal Statistical Society: Series B (Methodological)**, Reino Unido, v. 58, n. 1, p. 267–288, 1996.

UZUNIAN, A.; BIRNER, E. **Biologia**. 4. ed. São Paulo: HARBRA, 2013.

VON BARTHELD, C. S.; BAHNEY, J.; HERCULANO-HOUZEL, S. The Search for True Numbers of Neurons and Glial Cells in the Human Brain: A Review of 150 Years of Cell Counting. **The Journal of comparative neurology**, Estados Unidos, v. 524, n. 18, p. 3865–3895, 2016.

WHAT IS MACHINE LEARNING? | IBM. [S. l.], [c2023]. Disponível em: <https://www.ibm.com/topics/machine-learning>. Acesso em: 25 nov. 2023.

WHITE, H. **Artificial neural networks: approximation and learning theory**. Oxford, UK; Cambridge, USA: Blackwell, 1992.

## APÊNDICE A – LISTA DE VARIÁVEIS UTILIZADAS NAS PREVISÕES

	Categoria	Nome	Fonte	Unidade original	Nome de variável	Transf. <sup>6</sup>
1	Inflação	Índice Nacional de Preços ao Consumidor Amplo	IBGE	Var % mensal	ipca	1
2	Inflação	Índice Nacional de Preços ao Consumidor Amplo 15	IBGE	Var % mensal	ipca_15	1
3	Moeda	Selic (taxa nominal de juros)	BCB	% a.a.	selic	2
4	Moeda	M1	BCB	u.m.c. (milhões)	m1	2
5	Moeda	M2	BCB	u.m.c. (milhões)	m2	3
6	Moeda	M3	BCB	u.m.c. (milhões)	m3	3
7	Moeda	M4	BCB	u.m.c. (milhões)	m4	3
8	Mercado financeiro	Índice Bovespa	B3	Índice	ibovespa	2
9	Câmbio	Taxa de câmbio	BCB	BRL/USD	cambio	2
10	Câmbio	Taxa de câmbio efetiva real	FGV	Índice	cambio_real	2
11	Economia real	Receita nominal do varejo	IBGE	Índice	ind_varejo	2
12	Economia real	Receita nominal do varejo ampliado	IBGE	Índice	ind_varejo_ampl	2
13	Economia real	Produção física industrial	IBGE	Índice	ind_industria	2
14	Economia real	Índice de atividade econômica	BCB	Índice	ibc_br	2
15	Economia real	Nível de utilização da capacidade instalada	FGV	%	capacidade_ind	2
16	Economia real	Massa real de rendimentos	IBGE e BCB	R\$ milhões (reais)	rendimento	2
17	Economia real	PIB	BCB	u.m.c. (milhões)	pib	3
18	Trabalho	Salário mínimo federal	IpeaData	R\$	salario_minimo	2
19	Trabalho	Desemprego	IBGE	%	desemprego	1
20	Energia	Consumo total	EPE	GWh	energia_total	2
21	Energia	Consumo residencial	EPE	GWh	energia_residencial	2
22	Energia	Consumo industrial	EPE	GWh	energia_industrial	2
23	Energia	Consumo comercial	EPE	GWh	energia_comercial	2
24	Energia	Consumo outros	EPE	GWh	energia_outros	2
25	Produção industrial	Produção de bens de capital	IBGE	Índice	bens_capital	1
26	Produção industrial	Produção de bens intermediários	IBGE	Índice	bens_int	2
27	Produção industrial	Produção de bens de consumo duráveis	IBGE	Índice	bens_cons_d	2
28	Produção industrial	Produção de bens de consumo não-duráveis	IBGE	Índice	bens_cons_nd	2
29	Produção industrial	Produção de veículos	Anfavea	Unidades	prod_veiculos	2
30	Produção industrial	Produção de caminhões	Anfavea	Unidades	prod_caminhoes	1
31	Produção industrial	Produção de ônibus	Anfavea	Unidades	prod_onibus	2
32	Governo	Result. prim. setor público consolidado acum. 12 meses	BCB	u.m.c. (milhões)	result_primario	2
33	Clima	Oceanic Niño Index	NOAA	Índice	el_nino	1
34	Incerteza	Economic Policy Uncertainty Index for Brazil	EPU	Índice	epu_br	2
35	Incerteza	Economic Policy Uncertainty Index for USA	EPU	Índice	epu_usa	2
36	Incerteza	Índice de Confiança do Consumidor	Fecomércio	Índice	ind_conf_cons	2
37	Commodities	CRB all commodities index	Reuters	Índice	crb_cmdt	2
38	Commodities	CRB foodstuffs index	Reuters	Índice	crb_food	2
39	Commodities	CRB metals index	Reuters	Índice	crb_metal	2
40	Commodities	Preço do petróleo WTI	FRED	USD/barril	oil_usa	1
41	Commodities	Preço do petróleo Brent	FRED	USD/barril	oil_eu	1

<sup>6</sup> Código das transformações: (1) sem transformação, (2)  $\Delta x_t$  (3)  $\Delta^2 x_t$ .

## APÊNDICE B – GRÁFICOS DE CSFE POR MODELO E POR HORIZONTE

