

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

FREDERICO TANNOUS SAGER

PREVISÃO DA TAXA DE CRESCIMENTO DO PIB DO BRASIL:
UMA COMPARAÇÃO ENTRE *MACHINE LEARNING* E MODELOS ARIMA

Porto Alegre

2023

FREDERICO TANNOUS SAGER

**PREVISÃO DA TAXA DE CRESCIMENTO DO PIB DO BRASIL:
UMA COMPARAÇÃO ENTRE *MACHINE LEARNING* E MODELOS ARIMA**

Trabalho de Conclusão de Curso apresentado
como requisito parcial para a obtenção do grau
de Bacharel em Ciências Econômicas pela
Faculdade de Ciências Econômicas da
Universidade Federal do Rio Grande do Sul.

Orientador(a): Prof. Dr. Hudson da Silva Torrent

Porto Alegre

2023

CIP - Catalogação na Publicação

Sager, Frederico Tannous
PREVISÃO DA TAXA DE CRESCIMENTO DO PIB DO BRASIL:
UMA COMPARAÇÃO ENTRE MACHINE LEARNING E MODELOS ARIMA
/ Frederico Tannous Sager. -- 2023.
50 f.
Orientador: Hudson Torrent.

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, 2023.

1. Taxa de Crescimento. 2. Machine Learning. 3.
PIB. 4. ARIMA. I. Torrent, Hudson, orient. II.
Título.

FREDERICO TANNOUS SAGER

PREVISÃO DA TAXA DE CRESCIMENTO DO PIB DO BRASIL:
UMA COMPARAÇÃO ENTRE *MACHINE LEARNING* E MODELOS ARIMA

Trabalho de Conclusão de Curso apresentado como requisito parcial para a obtenção do grau de Bacharel em Ciências Econômicas pela Faculdade de Ciências Econômicas da Universidade Federal do Rio Grande do Sul.

Aprovado em: ____ de _____ de _____.

BANCA EXAMINADORA:

Prof. Dr. Hudson da Silva Torrent – UFRGS

Prof. Dr. Fernando Augusto Boeira Sabino da Silva – UFRGS

Prof. Dr. Sabino da Silva Porto Júnior – UFRGS

Porto Alegre

2023

AGRADECIMENTOS

Primeiramente, agradeço ao meu pai, Prof Dr Fabio Sager, e à minha mãe, Dra Leila Tannous, aos quais sempre me proveram tudo. A vocês, muito obrigado por terem me acompanhado ao longo desses últimos anos com apoio incondicional e amor infinito.

Aos meus avós Adla Tannous e Aref Tannous, que vieram do outro lado do planeta para morar na grande Caxias do Sul para tentar uma vida melhor.

À minha irmã, Júlia Tannous, que sempre foi capaz de ouvir meus desabafos sem julgamentos.

Aos meus amigos Jorge Michael e Elaine, que trouxeram insights interessantes a respeito deste trabalho. Seus conhecimentos foram cruciais para a realização do mesmo.

À minha companheira Ana Clara, atualmente doutoranda em Ciências Criminais pela PUCRS. Aqui vem a ela meu inestimável carinho, admiração e amor.

Ao Prof Dr Hudson Torrent, ao qual tive a honra de ter sido orientado ao longo desse último ano. Sem ele, este trabalho não existiria.

Finalmente, gostaria de agradecer aos contribuintes brasileiros, que são os responsáveis pelo financiamento desta Universidade.

RESUMO

O PIB vem sendo utilizado comumente na análise econômica, sendo uma variável importante para *policy makers* nas tomadas de decisão. Este trabalho tem como objetivo a utilização de algoritmos de *machine learning*, tais como *Random Forest*, *Gradient Tree Boosting* e *LASSO* para predição da taxa de crescimento do PIB do Brasil. Para construir os algoritmos, foram utilizados 26 regressores de diferentes temas macroeconômicos, como Indústria, Energia, Serviços e Comércio, Índice Geral de Preços, Base Monetária e Meios de Pagamento. Para testar a validade dos modelos, foi utilizado o *benchmark* ARIMA. Os resultados indicaram uma superioridade do modelo *Random Forest*, que apresentou os menores erros em 67% dos resultados. Para testar a acurácia dos nossos modelos, foi utilizado o teste Diebold-Mariano. Dos doze testes realizados, cinco modelos concorrentes apresentaram performance preditiva superior ao *benchmark* para nível de significância de 5%. O trabalho concluiu que o modelo *Random Forest* pode ser útil em exercícios de predição, principalmente em horizontes de curto prazo, mas que o modelo *ARIMA* não deve ser descartado como ferramental em análises de *forecasting*, uma vez que também apresentou menores erros em horizontes temporais mais longos.

Palavras-chave: Aprendizado de Máquina. Predição. *Random Forest*. Taxa de Crescimento de PIB.

ABSTRACT

GDP has been commonly used in economic analysis, being an important variable for policy makers in decision-making. This work aims to use machine learning algorithms, such as Random Forest, Gradient Tree Boosting and LASSO to predict the GDP growth rate in Brazil. To build the algorithms, 26 regressors of different macroeconomic themes were used, such as Industry, Energy, Services and Commerce, General Price Index, Monetary Base and Means of Payment. To test the validity of the models, the ARIMA benchmark was used. The results indicated a superiority of the Random Forest model, which presented the smallest errors in 67% of the results. To test the accuracy of our models, the Diebold-Mariano test was used. Of the twelve tests performed, five competing models showed predictive performance superior to the benchmark at a significance level of 5%. The work concluded that the Random Forest model can be useful in a prediction exercise, mainly in short-term horizons, but that the ARIMA model should not be discarded as a tool in forecasting analysis, since it also presented smaller errors in longer time horizons.

Keywords: Machine Learning. Forecasting, Random Forest. GDP Growth Rate.

LISTA DE FIGURAS

Figura 1 - Inteligência Artificial e suas subáreas	13
Figura 2 - <i>K-Cross-Validation</i>	16
Figura 3 - Coeficientes Selecionados em Dez Regressões <i>LASSO</i>	19
Figura 4 - A criação de uma nova variável	20
Figura 5 - Janela Rolante com Janela Inicial	26
Figura 6 - <i>Random Forest</i>	27
Figura 7 - PIB Mensal a Preços Correntes	33
Figura 8 - Crescimento percentual do PIB mês a mês	35
Figura 9 - Valor real vs valores preditos para $h=1$	37
Figura 10 - Valor real vs valores preditos para $h=3$	38
Figura 11 - Valor real vs valores preditos para $h=6$	39
Figura 12 - Valor real vs valores preditos para $h=12$	40

LISTA DE TABELAS

Tabela 1 - Dados utilizados no exercício de previsão	34
Tabela 2 - Resultado do teste de Dickey-Fuller para crescimento do PIB	35
Tabela 4 - <i>MAPE</i> , <i>MSE</i> e <i>MAE</i> por horizonte temporal (h)	39
Tabela 5 - Resumo dos melhores modelos por métrica de avaliação	40
Tabela 6 - Teste Diebold-Mariano	41
Tabela 7 - Top 5 preditores selecionados ($h=1$)	43
Tabela 8 - Top 5 preditores selecionados ($h=3$)	44
Tabela 9 - Top 5 preditores selecionados ($h=6$)	44
Tabela 10 - Top 5 preditores selecionados ($h=12$)	45

LISTA DE ABREVIACOES E SIGLAS

<i>ANN</i>	<i>Multilayer Perceptron Artificial Neural Network</i>
AR	Modelo Autorregressivo puro
ARIMA	Modelo Autorregressivo Integrado de Medias Moveis
BCB	Banco Central do Brasil
<i>BOJ</i>	<i>Bank of Japan</i>
BRI	ndice de Relevncia Blomberg
<i>ENET</i>	<i>Elastic Net</i>
FMI	Fundo Monetrio Internacional
IBGE	Instituto Brasileiro de Geografia e Estatstica
<i>LARS</i>	<i>Least Angle Regression</i>
<i>LASSO</i>	<i>Least Absolute Shrinkage and Selection Operator</i>
<i>MAE</i>	<i>Mean Absolute Error</i>
<i>MAPE</i>	<i>Mean Absolute Percentage Error</i>
<i>ML</i>	<i>Machine Learning</i>
MSE	Mean Squared Error
<i>MSFE</i>	<i>Mean Square Forecast Error</i>
PIB	Produto Interno Bruto
<i>RF</i>	<i>Random Forest</i>
<i>RMSE</i>	<i>Root Mean Squared Error</i>
<i>SPCA</i>	<i>Sparse Principal Component Analysis</i>

SUMÁRIO

1 INTRODUÇÃO.....	11
2 MACHINE LEARNING.....	13
2.1 CONCEITUAÇÃO.....	13
2.2 MACHINE LEARNING VS ECONOMETRIA.....	16
2.3 ESTUDOS APLICADOS.....	21
3 METODOLOGIA.....	25
3.1 ESTRATÉGIAS DE PREVISÃO.....	25
3.2 MODELOS.....	26
3.2.1 <i>Random Forest (RF)</i>	26
3.2.2 <i>Least Absolute Shrinkage and Selection Operator (LASSO)</i>	28
3.2.3 <i>Gradient Tree Boosting</i>	28
3.2.4 <i>ARIMA</i>	29
3.3 MÉTRICAS DE AVALIAÇÃO.....	31
4 DADOS.....	33
4.1 TRATAMENTO DOS DADOS.....	35
5 RESULTADOS.....	37
5.1 PREDITORES SELECIONADOS.....	41
6 CONCLUSÃO.....	46
REFERÊNCIAS BIBLIOGRÁFICAS.....	47

1 INTRODUÇÃO

A literatura sobre *machine learning* (*ML*) na teoria econômica é relativamente recente, mas vem crescendo de maneira rápida, devido à crescente disponibilidade de dados econômicos. Essa área do conhecimento, que é estabelecida como uma subárea da inteligência artificial, tem aperfeiçoado soluções para diversos tipos de problema, como *forecasting*, previsão de cenários e segmentação, auxiliando analistas e *policy makers* nos processos de tomada de decisão.

O fato de termos hoje quantidades expressivas de dados disponíveis para análise corrobora com o uso de técnicas de maior força computacional, como propõem os métodos de *ML*. Esses dados têm sido utilizados por grandes empresas para resolver problemas de negócios relacionados à predição - como previsão da receita, lucro estimado e demanda estipulada. Adicionalmente, tais métodos computacionais também podem ser utilizados na previsão de variáveis econômicas. É importante ressaltar que métodos tradicionais de séries temporais têm apresentado limites na solução de problemas preditivos relacionados a fenômenos econômicos complexos. Medeiros *et al.* (2021) aponta que novas abordagens que se utilizam de inteligência artificial produzem performance superior quando comparadas com modelos lineares. Nota-se, também, que as pesquisas econômicas vêm apresentando dificuldades em prever situações de maior estresse, como a crise econômica de 2008 (NYMAN; ORMEROD, 2020), por exemplo, o que torna cada vez mais necessário o avanço no desenvolvimento de metodologias capazes de oferecer previsões com maior acurácia e em tempo real. Yoon (2021) se utiliza de algoritmos também utilizados neste trabalho como o *Random Forest* e o *Gradient Tree Boosting* para um *forecasting* do PIB ano-a-ano do Japão, onde conclui que modelos de *ML* tem maior poder preditivo quando comparados com o *benchmark*.

Para economias que passam por instabilidades macroeconômicas com certa recorrência, a existência de métodos alternativos de previsão também se faz necessária. Cepni, Güney e Swanson (2019) buscam fazer uma previsão do crescimento do PIB para mercados emergentes, que são: Brasil, Indonésia, México, África do Sul e Turquia. Neste estudo, os autores comparam métodos de *forecasting* tradicionais da literatura com modelos de *ML*, onde encontram resultados que apontam a superioridade dos modelos computacionais frente a literatura já bem estabelecida.

A questão a se investigar neste trabalho, portanto, é verificar a possibilidade dos algoritmos de *ML* realizarem um *forecasting out of sample* (fora da amostra) da taxa de

crescimento do PIB brasileiro com maior poder de predição e acurácia do que os métodos já tradicionais da literatura econométrica. Para isso, iremos comparar três modelos bem consolidados na literatura de *ML* – *Random Forest*, *Gradient Tree Boosting* e *LASSO* – contra um modelo *benchmark ARIMA*. Este estudo se justifica pela manifestação de novas tecnologias capazes de resolver problemas complexos como os problemas econômicos, e dos limites apresentados por métodos tradicionais de séries temporais na solução de problemas preditivos na área. Além disso, é pertinente uma vez que utiliza o estado da arte desses algoritmos para a realização de uma previsão do PIB, um dado chave para a avaliação de *policy makers*.

Este trabalho se divide da seguinte maneira: na seção 2 conceitua-se o que é *ML*, comparando este campo de estudo com a literatura econométrica e suas possíveis aplicações; na seção 3, discute-se a metodologia abordada, descrevendo os modelos utilizados e as métricas de avaliação para comparação entre modelos; na seção 4 são explicitados os dados e a estratégia de tratamento destes para a aplicação nos modelos. A seção 5 demonstra os resultados do exercício empírico proposto a partir das métricas de avaliação descritas na seção 3 e, finalmente, a seção 6 conclui o trabalho.

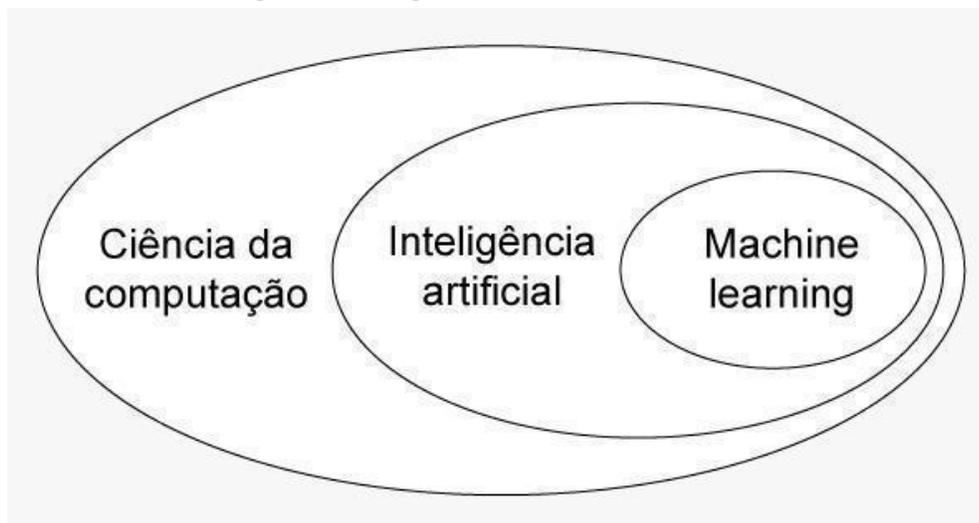
2 MACHINE LEARNING

Neste capítulo iremos conceituar o que é *ML* e suas funcionalidades para os problemas do mundo real, além de apontar quais são suas diferenças e semelhanças frente à econometria. Além disso, visitaremos trabalhos relacionados ao exercício de predição não apenas referente a variável PIB, mas também a outras variáveis econômicas como inflação e índices de preços.

2.1 CONCEITUAÇÃO

O *ML* pode ser definido como uma subárea da ciência da computação, fazendo parte do escopo da inteligência artificial, conforme ilustrado no diagrama abaixo (Figura 1). Lantz (2015) oferece uma definição funcional do campo, salientando que o *ML* destina-se ao desenvolvimento de algoritmos computacionais capazes de transformar dados em ação, razão pela qual vem sendo amplamente utilizado por diversos setores da indústria para oferecer *insights* a partir da grande quantidade de dados de que dispõem.

Figura 1 - Inteligência Artificial e suas subáreas



Fonte: Elaborado pelo autor (2022).

Para Athey (2017), o *ML* é um campo que desenvolve algoritmos desenhados para aplicação de *datasets*, cujos principais objetivos são (a) predição (regressão) – que é o enfoque

deste trabalho –, (b) classificação, (c) *clustering* e (d) tarefas de agrupamento. Dentro desse escopo, Géron (2019) divide ainda os métodos de *ML* em duas categorias, quais sejam, os métodos de aprendizado supervisionado e os métodos de aprendizado não-supervisionado.

O aprendizado supervisionado é utilizado na análise de problemas que já incluem as soluções desejadas, isto é, diz-se à máquina antecipadamente aquilo o que se quer prever. Dado um *dataset* com *target* y (discreto ou contínuo) e preditores X , o objetivo dos métodos de aprendizado supervisionado é encontrar y num *subset*. Assim, o aprendizado supervisionado abrange os objetivos de predição e de classificação de dados (GÉRON, 2019). Segundo Jordan e Mitchell (2015), os algoritmos mais populares são *support vector machine* (SVM), redes neurais, regressão linear, árvores de decisão e *random forests* (RF).

Já nos métodos de aprendizado não-supervisionado, o *target* a ser encontrado não é indicado à máquina. Dessa forma, o sistema aprende de maneira autônoma a encontrar padrões nos dados de entrada (GÉRON, 2019). Exemplos de problemas que podem ser resolvidos por métodos de aprendizado não-supervisionado são agrupamento de imagens em *clusters* e categorização de documentos por tópicos. Athey (2017) propõe a utilização destes métodos para criação de covariáveis, atuando como um passo intermediário para a análise econômica.

Em regra, algoritmos de *ML* costumam apresentar bons resultados em termos de predição. Contudo, não estão isentos de apresentar eventuais problemas como o *overfitting* - também chamado sobre-ajuste - e o *underfitting*. De acordo com o IBM Cloud Education (2021) o *overfitting* ocorre quando um modelo estatístico se ajusta exatamente ao *dataset* de treino, tornando-se impossível aplicar o modelo aos dados fora da amostra. Muitas vezes, o que pode acontecer é o algoritmo treinar por tempo demais, começando a aprender também os ruídos dos dados e informações irrelevantes. Esse fenômeno pode ser verificado quando o modelo apresenta alta variância e baixo erro. Por outro lado, o *underfitting* ocorre quando o algoritmo não foi treinado o suficiente ou quando não consegue encontrar relações significativas entre os dados de entrada e os de saída. Diferentemente do *overfitting*, o *underfitting* é caracterizado por apresentar alto viés e menor variância (IBM CLOUD EDUCATION, 2021).

De acordo com Varian (2014), tais fenômenos podem ser evitados com o emprego de métodos que acrescentem aleatoriedade aos modelos. Esse é o fundamento principal do chamado *ensemble learning*, que consiste em combinar múltiplos modelos a fim de reduzir o erro a variância final, mantendo seu poder de generalização. Os métodos mais conhecidos na literatura e

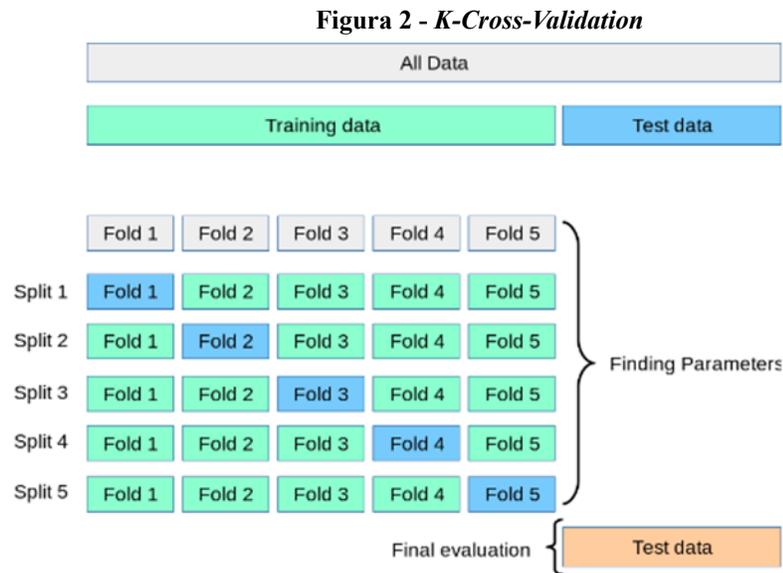
na indústria para adicionar aleatoriedade aos modelos são os chamados (i) *bootstrap*; (ii) *bagging*; e (iii) *boosting*.

De acordo com Friedman e Hall (2007), o método de *bootstrap* fornece inferências estatísticas (como estimativas de desvio padrão, viés, intervalos de confiança e testes de hipótese), sem supor a existência de distribuição normal na amostra ou mesma variância. Esse método pode ser mais preciso que inferências clássicas baseadas em distribuições normais ou *t* de Student. Além disso, o método de *bootstrap* utiliza os mesmos procedimentos básicos independentemente da estatística a ser calculada, sem necessidade de uma fórmula específica (FRIEDMAN; HALL, 2007).

Já o método de *bagging*, que consiste em uma junção das palavras “*bootstrap*” e “agregação”, costuma ser utilizado para reduzir o erro de previsão dos algoritmos. O *bagging* extrai amostras de *bootstrap* do *dataset* de treinamento e aplica o algoritmo para cada amostra deste *bootstrap* para, depois, fazer uma média das previsões estimadas (FRIEDMAN; HALL, 2007).

Por fim, o método de *boosting* é um procedimento que combina diversos classificadores fracos com o objetivo de melhorar a acurácia geral de um determinado modelo. Em cada iteração, o algoritmo atualiza os pesos dos exemplos e constrói um classificador adicional (FRIEDMAN; HALL, 2007).

Conforme apontado por Yoon (2021), além das técnicas que visam evitar o *overfitting* dos modelos, existem também técnicas que se propõem a otimizar os seus hiperparâmetros, visando melhorar sua acurácia final. O processo de *cross-validation* foi desenvolvido com este fim, uma vez que seleciona hiperparâmetros ótimos que produzem o menor erro-quadrado-médio (MSE, sigla em inglês para “*mean-squared-error*”) possível (YOON, 2021). Para realizar a tarefa com sucesso, o método separa os dados de treino em K partes e testa separadamente cada uma dessas partes, analisando qual delas se ajusta melhor ao modelo (FRIEDMAN; HALL, 2007), conforme ilustrado na Figura 2:



O *ML*, portanto, traz um quesito computacional importante na solução de variados problemas. Em problemas de predição, por exemplo, a literatura demonstra ótimos resultados em sua utilização. Problemas como *overfitting* e *underfitting* são facilmente contornáveis quando adicionamos aleatoriedade aos modelos, e também são facilmente otimizados quando usadas técnicas de validação cruzada.

2.2 MACHINE LEARNING VS ECONOMETRIA

Como se sabe, apesar de o *ML* estar se apresentando útil na solução de problemas de predição, a teoria econômica tradicional já trabalhava com essas questões sem esse instrumental, a partir da utilização de métodos econométricos. De acordo com Pinto (2011), a econometria pode ser definida como o resultado da incorporação de conhecimentos matemáticos, econômicos e estatísticos para estimar parâmetros entre variáveis dependentes e independentes, estabelecendo, assim, relações entre cofatores.

Varian (2014) busca estabelecer uma distinção entre as duas áreas. Segundo o autor, a análise dos dados a partir da econometria utiliza-se majoritariamente da regressão linear, e pode ser dividida em quatro categorias, quais sejam: (i) predição; (ii) sumarização, (iii) estimação e (iv) testes de hipóteses. O *ML*, por outro lado, está preocupado principalmente com a predição e

com o desenvolvimento de sistemas integrados de alta performance, que possam utilizar essas previsões frente às restrições computacionais.

Athey (2017) aprofunda a discussão sobre o impacto do *ML* na análise econômica. Segundo a autora, o *ML* não acrescenta questões relacionadas à identificação e estimação de parâmetros - objetivo principal da ciência econômica moderna. Por outro lado, a autora argumenta que o *ML* apresenta vantagens frente à econometria, salientando que esse trata a análise empírica de forma algorítmica, estimando e comparando diversos modelos alternativos. Na análise econômica tradicional, o pesquisador escolhe um modelo baseado em princípios da teoria econômica e o estima. Desta forma, Athey (2017) evidencia as vantagens dos métodos de *ML* frente aos métodos econométricos, destacando a performance desses e permitindo aos pesquisadores serem sistemáticos ao descrever o processo de escolha de seus modelos.

De acordo com Mullainathan e Spiess (2017), o sucesso dos métodos de *ML* se deve à sua capacidade de descobrir padrões generalizáveis. A vantagem de algoritmos de *ML* sobre modelos de regressão tradicionais é que eles encontram estruturas complexas nos dados sem que essas estruturas tenham sido previamente especificadas. Desse modo, os algoritmos de *ML* são capazes de ajustar formas funcionais complexas e flexíveis aos dados. Ao mesmo tempo em que algoritmos são capazes de se ajustar a formas mais complexas, eles apresentam sua limitação na estimação de parâmetros únicos para o modelo.

Para ilustrar as questões comentadas, os autores desenvolveram um estudo comparativo entre métodos de *ML* e métodos econométricos, realizando uma aplicação costumeira no universo dos dados – a predição de preços de imóveis. Para tanto, utilizam-se de um *dataset* contendo 10.000 observações selecionadas aleatoriamente do *American Housing Survey*¹, contando com 150 variáveis explanatórias que contêm informações como localização, número de quartos, metragem, dentre outras, e aplicam a ele métodos de *ML* – *RF*, *ensemble* e árvores de decisão (*tuned by depth*) – e métodos econométricos (OLS). Os autores concluem que, tanto para as previsões dentro da amostra (*training set*) quanto para previsões fora da amostra (*test set*), os métodos econométricos conseguiram superar apenas as árvores de decisão, a partir da utilização do R^2 como métrica de comparação.

Apesar das técnicas de inteligência artificial superarem modelos econométricos em termos de acurácia, Mullainathan e Spiess (2017) nos trazem avisos sobre o que não é possível

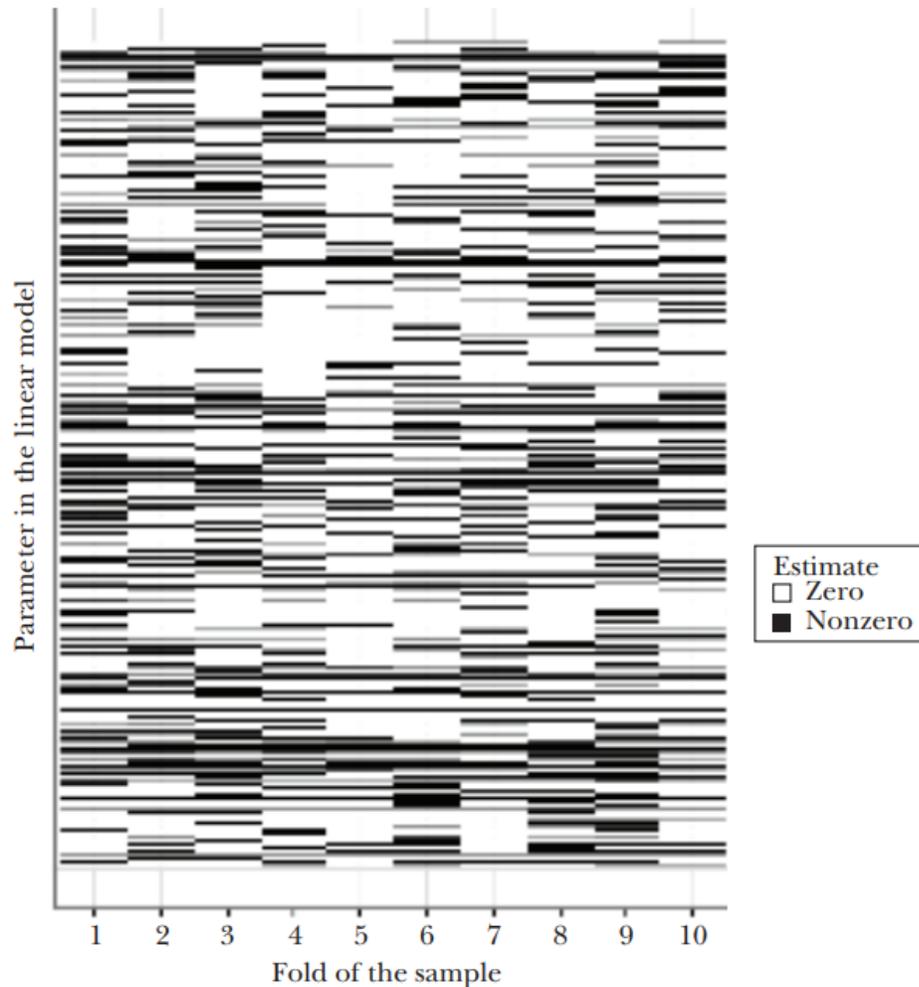
¹ A *American Housing Survey* é uma pesquisa estatística financiada pelo Departamento de Habitação e Desenvolvimento Urbano dos Estados Unidos e conduzida pelo US Census Bureau (HUD USER, 2015).

aprender com esse *output* gerado. O primeiro, é a impossibilidade de contabilizar erros-padrão dos coeficientes. O segundo, e o que talvez seja o Calcanhar de Aquiles do *ML*, é a inabilidade de selecionar os mesmos cofatores para a mesma regressão. Para ilustrar esta questão, os autores selecionaram aleatoriamente dez partições da amostra com 5000 observações cada, a fim analisar como os algoritmos selecionam os regressores. A figura abaixo mostra como a seleção de variáveis varia de partição para partição, impossibilitando, portanto, uma análise de inferência causal. Essa inabilidade, no entanto, não se reflete numa instabilidade na predição, já que o R^2 se mantém constante ao longo das partições.

O grande apelo desses algoritmos é que eles podem se encaixar em muitas funções diferentes. Mas isso cria um calcanhar de Aquiles: mais funções significam uma chance maior de que duas funções com coeficientes muito diferentes possam produzir uma qualidade de previsão semelhante. Como resultado, como um algoritmo escolhe entre duas funções muito diferentes pode se resumir ao lançamento de uma moeda (MULLAINATHAN; SPIESS, 2017, tradução livre)².

² “The very appeal of these algorithms is that they can fit many different functions. But this creates an Achilles’ heel: more functions mean a greater chance that two functions with very different coefficients can produce similar prediction quality. As a result, how an algorithm chooses between two very different functions can effectively come down to the flip of a coin.” (MULLAINATHAN; SPIESS, 2017).

Figura 3 - Coeficientes Seleccionados em Dez Regressões *LASSO*



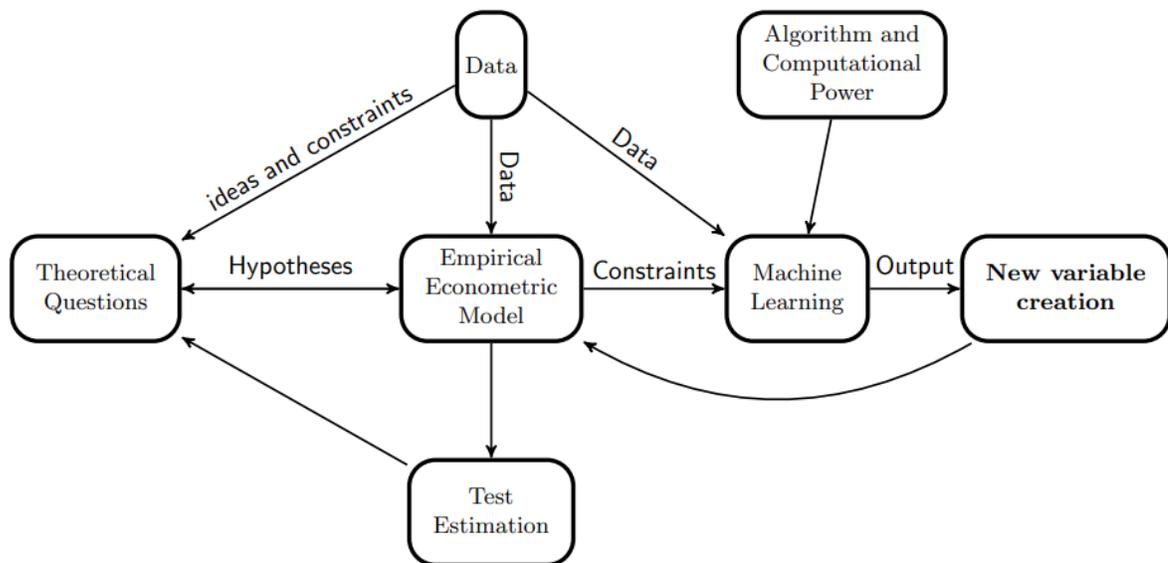
Fonte: Mullainathan e Spiess, 2017.

Apesar de a econometria e o *ML* apresentarem diferenças metodológicas e fins distintos, existem trabalhos que conseguem condensar as duas técnicas, a fim de extrair o que há de melhor dos dois mundos. Nesse sentido, Liu e Xie (2018) conduziram exercícios empíricos com o objetivo de contrastar os métodos num contexto de predição de bilheteria de cinemas, utilizando dados da indústria cinematográfica em conjunção com dados das mídias sociais. Os resultados deste exercício mostraram que métodos de *ML* superaram os econométricos no curto prazo, e vice-versa. Isso porque métodos de *ML* conseguem lidar com estruturas heterogêneas de dados, sendo, assim, capazes de capturar ruídos de curto prazo como não-linearidade e irregularidades. Em contraste, métodos econométricos performam melhor no longo prazo ou quando a heterogeneidade dos dados reduz, ganhando assim competitividade frente aos modelos de

inteligência artificial. Os autores concluem que o modelo híbrido que combina ambas abordagens performa melhor, em média, que modelos puramente econométricos ou puramente baseados em métodos de *ML* (LIU; XIE, 2018).

Com uma proposta semelhante a Liu e Xie (2018), Guerzoni, Nava e Nuccio (2019) adotam uma estrutura econométrica-computacional a fim de mensurar a taxa de sobrevivência de *startups* inovativas. Para isso, os autores propõem uma metodologia em que se utiliza um algoritmo de *ML* supervisionado para criar um indicador de inovação, já que em trabalhos anteriores da área se utilizavam variáveis *proxies*, o que afeta o *output*. Tendo esse indicador calculado pelo algoritmo, este é “imputado” num modelo econométrico clássico. Este método vai ao encontro do estudo realizado por Athey (2017), em que a autora sugere a utilização de *ML* para a criação de covariáveis. Figura 4 ilustra a criação da variável e um *roadmap* do método aplicado:

Figura 4 - A criação de uma nova variável



Fonte: Guerzoni, Nava e Nuccio (2019)

A conclusão dos autores é de que são necessários *feedbacks* entre a teoria adotada para a interpretação do fenômeno (*assumptions*), a econometria e o *ML*, a fim de que sejam criadas propriedades robustas entre as três áreas para que, assim, seja possível uma integração sólida entre os campos (GUERZONI; NAVA; NUCCIO, 2019).

De maneira geral, podemos afirmar que o *ML* é útil em tarefas de predição devido sua capacidade em encontrar padrões em estruturas de dados heterogêneas sem especificação prévia, os tornando capazes de ajustar a formas funcionais complexas e flexíveis aos dados. No entanto, esse método apresenta suas limitações na estimação de parâmetros únicos para os modelos. A econometria, por outro lado, pode ser usada não apenas na predição, mas também na estimação de parâmetros, na sumarização e também no teste de hipóteses. Esse trabalho se propõe a comparar a capacidade de predição de modelos econométricos com modelos de *ML*. A seção a seguir irá demonstrar estudos que se propuseram a fazer o mesmo e que demonstraram ótimos resultados na utilização dos novos métodos computacionais comparados com o *benchmark*.

2.3 ESTUDOS APLICADOS

Nesta seção, iremos analisar trabalhos de variados autores que utilizam algoritmos de *ML* para realizar predições de variáveis econômicas, como esta tese também se propõe. Os estudos mostram certa superioridade dos métodos computacionais frente à literatura clássica de séries temporais; isso porque, como já explicitado anteriormente, os métodos de *ML* conseguem lidar com estruturas heterogêneas de dados.

Flanerry (2020) utiliza métodos de *ML* no *forecast* do PIB da Irlanda e de outros países da Europa com o objetivo de comprovar a eficácia preditiva dos métodos de *ML* para esta finalidade. Ao invés de comparar métodos, o autor compara dois algoritmos: (i) o *Multilayer Perceptron Artificial Neural Network (ANN)*, que é bastante utilizado quando temos variáveis independentes não-lineares; e (ii) a *Random Forest (RF)*, que é um algoritmo baseado em árvores de decisão. A métrica de avaliação do modelo selecionada pelo autor foi o R^2 e o *Mean Squared Error (MSE)* – erro quadrado médio. O modelo de *Random Forest* apresentou uma acurácia de 0.98, enquanto que a *ANN* apresentou 0.94. Apesar do modelo ser altamente explicativo, a *ANN* apresentou diferenças significativas de *output* quando comparado o PIB da Irlanda com o de outros países. Segundo o autor, isso acontece pela presença de multinacionais no país, que utilizam o sistema tributável favorável para declarar lucros e outros resultados operacionais, alterando assim resultados finais.

Em Maccarrone, Morelli e Spadaccini (2021), os autores buscam estimar o PIB dos Estados Unidos utilizando modelos tradicionais de séries temporais, modelos de regressão e algoritmos de *ML*. O objetivo de seu trabalho é comparar empiricamente os três métodos, analisando qual tem melhor poder preditivo. A métrica de comparação é o MSE. Para realizar o objetivo proposto, os autores utilizam dados de curvas de rendimento e de variáveis macroeconômicas tradicionais, como índice de preços do consumidor, taxa de desemprego etc., aplicando-os em cinco modelos de *forecast*: (i) modelo autorregressivo puro (AR), (ii) regressão linear, (iii) algoritmo *k-nearest-neighbors* (KNN), (iv) modelo autorregressivo integrado de médias móveis (ARIMA), e (v) modelo autorregressivo integrado de médias móveis ajustado à sazonalidade (SARIMA).

Os autores concluem que o algoritmo KNN performa melhor no curto prazo (*one-step-ahead*), enquanto a regressão linear performa melhor no longo prazo (*multi-step-ahead*). Isso vai diretamente ao encontro dos resultados encontrados por Yoon (2021), que utiliza métodos tanto de *boosting* quanto RF para prever o crescimento do PIB do Japão entre os períodos de 2001-2018. Para isso, o autor utiliza indicadores tradicionais relacionados a contas nacionais, emprego, moeda e inflação como preditores. Apesar de utilizar regressores altamente correlacionados, o autor ressalta a eficiência dos métodos de *ML* em lidar com problemas de multicolinearidade. Além disso, um processo de *k fold cross-validation* é utilizado para otimizar hiperparâmetros do modelo. Além disso, o autor se utiliza de *rolling windows* a fim de evitar que o modelo sofra sobre-ajuste. Ele conclui a partir do RMSE e do MAPE que os modelos de *ML* superam os *benchmarks* propostos pelo FMI (Fundo Monetário Internacional) e pelo BOJ (*Bank of Japan*) (MACCARRONE; MORELLI; SPADACCINI, 2021).

Não apenas o PIB é objeto de estudo para comparação de *forecastings* no campo econômico. A literatura também nos traz aplicações para tratar de uma variável chave no cálculo econômico: a inflação. Na tentativa de fazer previsões com maior acurácia, Gaglianone e Araújo (2020) propõem comparar dezesseis métodos de *forecasting*, variando seu horizonte de predição de um a doze meses. Os métodos escolhidos são: *random walk*; *random walk* (Atkeson-Ohanian); ARMA; VAR; *Phillips curve (backward-looking or hybrid)*; *factor model (direct or iterated forecast; with or without targeted predictors)*; *ridge regression*; LASSO; *elastic net*; *random forest*; *quantile regression forest* e *Focus survey* (incluído como referência) (GAGLIANONE; ARAÚJO, 2020).

Como *target*, os autores escolheram o Índice de Preços do Consumidor Amplo (IPCA), utilizando quinze anos de dados, de janeiro de 2004 a dezembro de 2018, totalizando 180 observações. Destas, 85 são utilizadas para estimar o modelo (*training set*) enquanto que 95 são empregadas para comparar *forecasts* (*test set*). Como preditores, são utilizadas tanto variáveis macroeconômicas como financeiras, totalizando 120 regressores mensalmente observados. Estes estão inseridos em categorias como índice de preços, taxas de juros, atividade econômica, mercado financeiro, mercado de trabalho, dívidas do governo, importação e exportação de bens e serviços e variáveis internacionais potencialmente relacionadas à economia brasileira (GAGLIANONE; ARAÚJO, 2020).

Para analisar as variáveis que os métodos de *ML* selecionam para fazer as previsões, os autores elaboraram uma nuvem de palavras a fim de destacar quais regressores são mais utilizados pelos algoritmos. Estes incluem inflação passada, produção industrial (principalmente bens de consumo e bens intermediários), mercado de trabalho, setor externo e setor público. Como métrica de avaliação, os autores optaram por utilizar o *Mean Squared Error* (*MSE*) e, a partir dele, concluíram que os algoritmos de *ML* superam os modelos econométricos, principalmente o algoritmo *RF* e o *Quantile Regression Forest*, oferecendo uma contribuição para o campo da previsão econômica (GAGLIANONE; ARAÚJO, 2020).

Cepni, Güney e Swanson (2019) buscam fazer uma previsão do crescimento do PIB para mercados emergentes, sendo estes: Brasil, Indonésia, México, África do Sul e Turquia. A abordagem escolhida envolve utilizar o *dynamic factor modeling*, técnica introduzida por Giannone, Reichlin e Small (2008), a fim de construir previsões mensais do crescimento do PIB trimestral. As previsões são construídas utilizando todo o *dataset* selecionado ou apenas com preditores indicados pelos seguintes algoritmos: *Sparse Principal Component Analysis* (*SPCA*), *Elastic Net* (*ENET*), o *Least Angle Regression* (*LARS*) e o *LASSO*. Além destas técnicas, os autores optaram por também utilizar o Índice de Relevância Blomberg (*BRI*), que atribui uma medida de importância para cada variável baseada na sua utilização por agentes do mercado (CEPNI; GÜNEY; SWANSON, 2019).

Os dados selecionados pelos autores abrangem um período de dois anos compreendido entre setembro de 2005 e setembro de 2017, sendo dividido em seis categorias: (i) habitação (preços); (ii) mercado de trabalho (emprego e desemprego); (iii) preços (preços do consumidor e preços de produção); (iv) variáveis financeiras (taxas de juros, taxas de câmbio e preços de

ações); (v) moeda, crédito e agregados monetários (oferta de moeda, depósitos à vista); e (vi) atividade real (produção industrial, capacidade utilizada, vendas de varejo etc.). As séries são feitas estacionárias e as defasagens também aplicadas. No caso do Brasil, os algoritmos selecionaram variáveis relacionadas à economia internacional – nesse caso, exportação – e também regressores ligados ao consumo privado e ao setor de construção civil. Além disso, os resultados encontrados a partir do *Mean Square Forecast Error (MSFE)* sugerem que modelos que utilizaram redução de dimensionalidade produziram previsões superiores àqueles que não utilizam essa técnica – nesse caso, modelos autorregressivos e *dynamic factor modeling*.

Os trabalhos analisados utilizaram diversos modelos para construir seus *forecastings* dos diferentes dados econômicos e também escolheram suas métricas de acurácia. No próximo capítulo, iremos explicitar o *framework* utilizado para a construção dos nossos modelos, como também iremos descrever os algoritmos utilizados neste estudo e as métricas de avaliação consideradas.

3 METODOLOGIA

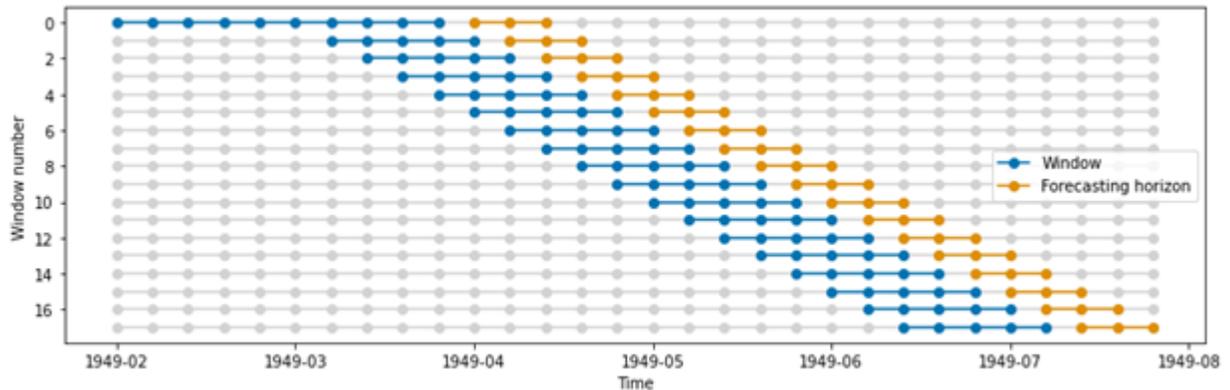
Na seção de Metodologia, iremos analisar a estratégia de previsão dos modelos, além de detalhar o funcionamento de cada modelo de *ML* proposto neste trabalho como também o *benchmark ARIMA*; finalmente, iremos evidenciar as métricas de avaliação utilizadas para medir a acurácia dos modelos propostos.

3.1 ESTRATÉGIAS DE PREVISÃO

Existem duas estratégias para previsão de séries temporais: a recursiva e a direta. Na estratégia de previsão recursiva, o modelo é primeiro treinado nos dados históricos até um determinado ponto no tempo. Em seguida, o modelo é usado para prever o próximo valor na série temporal, que se torna o primeiro valor previsto. Este valor previsto é então usado como entrada para prever o próximo valor da série, e este processo é repetido para o número desejado de passos à frente. Na estratégia de previsão direta, por outro lado, o modelo é treinado com os regressores até $t-h$, enquanto que a variável dependente correspondente está no ponto t . O modelo, portanto, é avaliado no instante t e, assim, obtém-se a previsão para o ponto $t+h$. Neste trabalho, foi utilizada a estratégia recursiva.

Para estimar os modelos, foi utilizada a técnica de janelas rolantes com uma janela inicial. Nesta abordagem, uma janela de dados é escolhida primeiro e os parâmetros do modelo são estimados utilizando esta janela inicial. A janela inicial é usada para definir o ponto de partida para as estimativas e serve também para capturar as propriedades da série temporal. Em seguida, uma janela rolante de tamanho fixo é rolada ao longo dos pontos e os valores preditos são atualizados, utilizando-se os novos dados, sem atualização de estimação dos parâmetros. Neste trabalho, utilizamos como janela inicial 75% da amostra, com uma janela rolante de tamanho 12 – ou seja, nossa janela tem tamanho de 1 ano. Além disso, o tamanho do passo para a movimentação da janela é igual a 1 – ou seja, um mês. O gráfico da Figura 5 exemplifica a abordagem.

Figura 5 - Janela Rolante com Janela Inicial



Fonte: SKTIME (2021).

3.2 MODELOS

Para a aplicação dos modelos (*ARIMA*, *Gradient Tree Boosting*, *Random Forest* e *LASSO*), utilizaremos a biblioteca *sktime*, a qual oferece uma estrutura unificada para tarefas de predição em séries temporais (LÖNING; KIRÁLY, 2020). A vantagem dessa biblioteca é que ela oferece integração com os pacotes convencionais de *ML*, como o *sci-kit learn*, oferecendo uma interface útil para tarefas de regressão, classificação, etc.

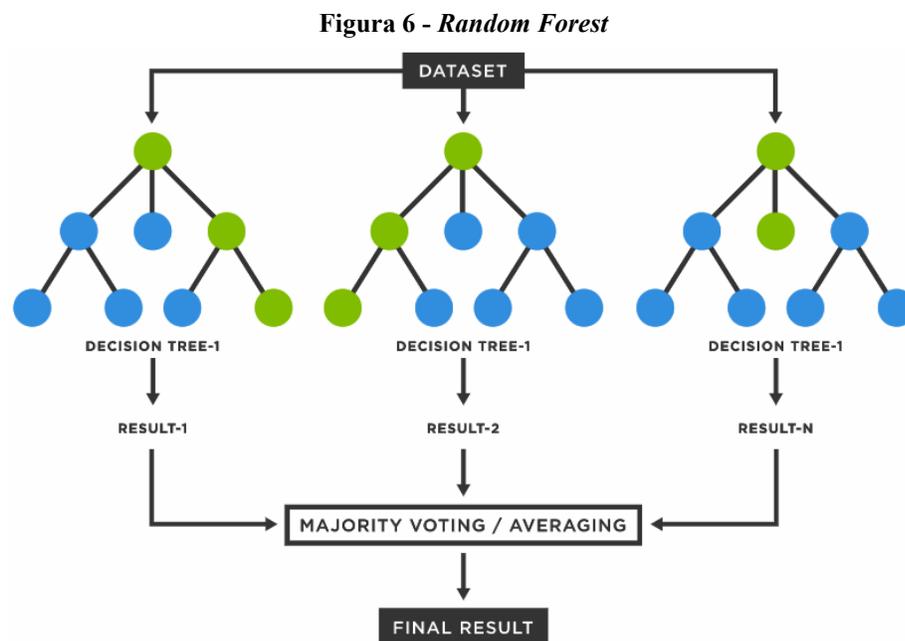
3.2.1 *Random Forest (RF)*

Segundo Breiman (2001), um algoritmo *RF* é uma combinação de preditores de árvores de decisão feita a partir de uma seleção aleatória de amostras dos dados de treino, conforme ilustrado no diagrama abaixo (Figura 6). A predição é feita a partir da média do resultado de cada árvore de decisão. Assim, o algoritmo *RF* mantém os benefícios classificatórios das árvores de decisão, mas a partir do uso do *bagging* nas amostras, assim evitando ajustes excessivos. Segundo Horning (2010), as vantagens dos algoritmos *RF* são que eles solucionam problemas de *overfitting* com facilidade; são pouco sensíveis a *outliers*; não requerem a poda das árvores; geram de maneira automática a importância das variáveis e a acurácia do modelo. Matematicamente, podemos definir a *RF* sendo:

$$\hat{y}_{t+h} = \frac{1}{B} \sum_{b=1}^B f_b(\mathbf{x}_t),$$

(1)

onde y_{t+h} é variável que queremos prever, B é o número de árvores de regressão e $f_b(x_t)$ é o resultado de cada árvore de regressão construída considerando uma subamostra dos dados de treino e um subconjunto aleatório de variáveis que levam a melhor divisão em cada nó.



Fonte: TIBICO CLOUD (2022).

Ali, Khan e Ahmad (2012) comentam que algoritmos *RF* são apropriados para *datasets* com alta dimensionalidade, pois são capazes de lidar com valores faltantes, além de serem capazes de solucionar problemas com dados contínuos, categóricos e binários.

3.2.2 *Least Absolute Shrinkage and Selection Operator (LASSO)*

A regressão *LASSO* é um modelo linear de regularização do tipo *L1* proposto por Tibshirani (1996), que adiciona uma penalidade igual ao valor absoluto da magnitude dos coeficientes da regressão (BEYER, 2002):

$$p(\beta, \lambda) = \lambda \|\beta\|_1 = \lambda \sum_{i=1}^n |\beta_i|. \quad (2)$$

É utilizado em áreas como genética e economia, em que se observam dados em que o número de preditores é geralmente maior que o número de observações, como é o caso deste trabalho, e muitas vezes há pouco ou nenhum conhecimento a priori para a seleção de variáveis (RANSTAM; COOK, 2018).

Segundo Ranstam e Cook (2018), a regressão *LASSO* visa identificar as variáveis e seus respectivos coeficientes de regressão que levam a um modelo que minimiza o erro da previsão. Isso é possível impondo uma restrição nos parâmetros do modelo, que “encolhe” da regressão para zero, forçando que a soma do valor absoluto dos coeficientes de regressão seja inferior a um valor fixo (λ). De forma prática, isso restringe a complexidade do modelo. Variáveis com coeficiente de regressão zero após o encolhimento são excluídas do modelo. A escolha de λ é feita usando uma abordagem de validação cruzada (*k-cross-validation*).

De acordo com Cheng, Huang e Shi (2019), o método *LASSO* tem sido o mais estudado nos últimos quinze anos, uma vez que mostra sucesso em muitas situações práticas onde se tem mais preditores que observações.

3.2.3 *Gradient Tree Boosting*

Gradient Tree Boosting é um método “*greedy*” proposto por Friedman (2001) baseado na construção de modelos aditivos, ajustados sequencialmente a uma *base learner* que se ajusta à pseudo-resíduos em cada iteração. O pseudo-resíduo é o gradiente da função perda sendo minimizado, com relação aos valores em cada ponto dos dados de treinamento avaliados na etapa

atual. Em outras palavras, se quisermos estimar uma função $f(x)$, nós iterativamente usamos árvores de regressão para calcular:

$$f_m(\mathbf{x}) = f_{m-1}(\mathbf{x}) - \rho_m \frac{\partial \ell(f(\mathbf{x}))}{\partial f(\mathbf{x})} \Big|_{f(\mathbf{x})=f_{m-1}(\mathbf{x})}, \quad (3)$$

onde $\ell(f(x))$ é a função perda e ρ_m é a taxa de aprendizado. O método repete o processo ao número de iterações especificado (IZENMAN, 2008).

O *Gradient Tree Boosting* é um caso especial do *gradient boosting*, onde os “*base learners*” são pequenas árvores de regressão. Assim, adiciona-se iterativamente árvores de regressão ao modelo que se ajustam ao gradiente da função perda, reduzindo o erro.

3.2.4 ARIMA

Para testarmos nossos modelos e entender se estes foram úteis para previsões de séries temporais, precisamos comparar suas performances com um *benchmark*. Um *benchmark* serve para verificar a validade de um modelo proposto. Como modelos *ARIMA* já são conhecidos na literatura e amplamente utilizados, optou-se por utilizar este para verificação neste trabalho.

O *ARIMA* é um modelo autorregressivo integrado de médias móveis. Diferente dos outros métodos propostos neste trabalho, o *ARIMA* é um método estatístico univariado, ou seja, utiliza a própria série para fins de predição. Diferente dos modelos de *ML*, onde os parâmetros dos modelos nos dizem pouco em relação à sua predição, os métodos econométricos – neste caso, o modelo *ARIMA* – necessitam de avaliação de seus principais parâmetros: o número de defasagens AR (auto regressivo), a ordem da *MA* (média móvel) e também o número de vezes que os dados foram diferenciados (*I*).

Segundo Padilha *et al.* (2013), os modelos Autorregressivos (AR) foram criados com a ideia de que a presente observação de série Z_t pode ser explicada como uma função das p observações passadas, $Z_{t-1}, Z_{t-2}, \dots, Z_{t-p}$, onde p determina o número de passos entre as

observações passadas e a previsão da próxima observação. A estrutura autorregressiva geral é expressa por:

$$Z_t = \varphi_1 Z_{t-1} + \varphi_2 Z_{t-2} + \dots + \varphi_p Z_{t-p} + a_t \quad (4)$$

em que φ são os parâmetros da estrutura autorregressiva, $i = 1, \dots, p$ (ordem da estrutura); a_t é o ruído branco com média zero e variância σa^2 .

Já os modelos de médias móveis (MA) são formados por combinações lineares de ruído branco, a_t , ocorridos no período corrente e nos períodos passados (SOUZA *et al.*, 2006). A estrutura de médias móveis geral é expressa por:

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (5)$$

De acordo com Bayer (2010), o modelo *ARIMA* consiste em ajustar modelos autorregressivos integrados de média móvel a um conjunto específico de dados. Se Z_t segue um processo autorregressivo e de média móvel de ordens p, q , *ARMA(p, q)*, então:

$$\varphi(B)Z_t = c + \theta(B)\varepsilon_t \quad (6)$$

em que ε_t é ruído branco com média zero e variância σ^2 , c é um parâmetro que permite ao processo ter média diferente de zero, o polinômio $B^k Z_t = Z_{t-k}$ é o operador retroativo, $\varphi(B)$ e $\theta(B)$ são o polinômio autorregressivo e o polinômio de médias móveis, respectivamente, apresentados nas equações abaixo:

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p \quad (7)$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (8)$$

Uma série que não apresenta característica de estacionariedade pode torna-se estacionária pela utilização de um processo de diferenciação. Sendo assim, se uma série se torna estacionária

após d diferenças, a série é dita integrada (I) de ordem d . Com isso, o modelo *ARMA* integrado passa a ser denominado de modelo *ARIMA*.

3.3 MÉTRICAS DE AVALIAÇÃO

Para comparar nossas previsões, utilizaremos métricas de avaliação comumente utilizadas na literatura. Essas métricas visam comparar os diferentes modelos propostos nos diferentes horizontes temporais. Seu cálculo geral é feito a partir da diferença entre os valores reais observados e os valores preditos pelos modelos. Neste exercício, utilizaremos Erro Médio Absoluto (*MAE*), Erro Percentual Absoluto Médio (*MAPE*) e Erro Quadrado Médio (*MSE*).

O *MAE* utiliza a média aritmética na Distância Euclidiana entre os valores preditos pelo modelo e os valores reais observados.

Se \hat{y}_i é o valor previsto da i -ésima amostra, e y_i é o valor real observado, então o *MAE* estimado sobre o tamanho da amostra (n_{samples}) é:

$$\text{MAE}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} |y_i - \hat{y}_i|. \quad (9)$$

O *MAPE* visa compreender a proporção do erro com as observações reais. A ideia dessa métrica é ser sensível a erros relativos. Ela é definida matematicamente por:

$$\text{MAPE}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} \frac{|y_i - \hat{y}_i|}{\max(\epsilon, |y_i|)} \quad (10)$$

onde ϵ é um número arbitrário mas estritamente positivo para evitar resultados indefinidos quando y é zero.

Finalmente, o *MSE* calcula o erro quadrático médio, que corresponde ao valor esperado do erro ao quadrado. É definido por

$$\text{MSE}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} (y_i - \hat{y}_i)^2. \quad (11)$$

Apesar do bom desempenho das medidas de previsão amplamente utilizadas, elas não revelam por si só se a previsão de um modelo é estatisticamente melhor do que a de outro. Portanto, utilizaremos também o Teste Diebold-Mariano (DIEBOLD; MARIANO, 1995) para avaliar a hipótese nula de igual precisão preditiva entre os métodos de previsão concorrentes. O teste é feito a partir da função perda L associado ao erro do de previsão e_t . A perda diferencial entre previsões 1 e 2 em t pode ser definida como:

$$d_{12t} = L(e_{1t}) - L(e_{2t}) \quad (12)$$

Para satisfazer a hipótese nula, é necessário que $E(d_{12t}) = 0$. Segundo SABA (2015), se a hipótese de covariância estacionária da perda diferencial se sustentar, conseguimos testar a perda diferencial para todo o horizonte amostral T e avaliar a estatística DM_{12} resultante. Sendo D_{12} a média amostral das perdas diferenciais e um estimador consistente do desvio padrão de D_{12} :

$$D_{12} = \frac{1}{T} \sum_{t=1}^T d_{12t} \quad (13)$$

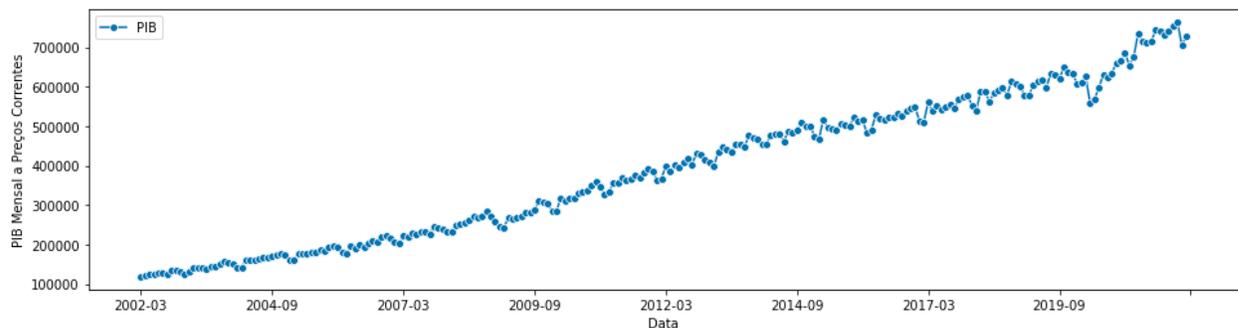
$$DM_{12} = \frac{D_{12}}{\hat{\sigma}_{D_{12}}} \xrightarrow{d} N(0,1) \quad (14)$$

4 DADOS

Este trabalho consiste em fazer uma previsão do crescimento do Produto Interno Bruto (PIB) mensal do Brasil utilizando algoritmos de *ML*. Apesar de o PIB oficial do Brasil ser divulgado trimestralmente pelo Instituto Brasileiro de Geografia e Estatística (IBGE), adota-se neste trabalho uma previsão mensal, tendo em vista que há maior abundância de dados divulgados e maior número de observações disponíveis nessa frequência, trazendo maior robustez e validade aos resultados.

O PIB mensal é um indicador de frequência mensal produzido pelo Banco Central do Brasil (BCB) para utilização no cálculo da relação entre agregados econômicos mensais (como dívida pública, saldo em transações correntes e saldo de crédito) e o PIB (IPEA DATA). A estimativa é feita via interpolação dos valores trimestrais já divulgados pelo IBGE ou dos projetados, não se tratando de cálculo do PIB a partir de informações primárias. A interpolação é feita uma vez que o PIB oficial é divulgado trimestralmente, ao passo que várias informações econômicas compiladas pelo Banco Central são mensais.

Figura 7 - PIB Mensal a Preços Correntes



Fonte: Elaborado pelo autor (2023).

Para a realização da presente estimativa, foram selecionadas 27 (vinte e sete) séries macroeconômicas (incluindo o PIB Mensal), de janeiro de 2002 a fevereiro de 2022, totalizando 242 observações. A tabela abaixo sumariza os dados selecionados na construção do preditor (Tabela 1). Indo de encontro com Gaglianone e Araújo (2020), foram selecionadas variáveis macroeconômicas de diferentes temas: Contas Nacionais, Indústria, Energia, Serviços e Comércio, Índice Geral de Preços e seus componentes, Base Monetária e Meios de Pagamento. Esses temas foram selecionados com o objetivo de oferecer uma visão ampla da economia

brasileira, tangendo temas variados que vão do lado real da economia até temas que abrangem fatores monetários.

Tabela 1 - Dados utilizados no exercício de previsão

Tema	Título da Série	Unidades	Fonte	Nº de Séries
Contas Nacionais	PIB Mensal	R\$ Milhões	BCB-Depec	1
Indústria	Indicador de Produção – Indústria de Transformação	Índice (base 100)	IBGE	11
	Indicador de Produção – Indústria Extrativa-Mineral	Índice (base 100)	IBGE	
	Indicador de Produção – Bens de Capital	Índice (base 100)	IBGE	
	Indicador de Produção – Bens intermediários	Índice (base 100)	IBGE	
	Indicador de Produção – Bens de Consumo Duráveis	Índice (base 100)	IBGE	
	Indicador de Produção – Bens de Consumo Semiduráveis e não-duráveis	Índice (base 100)	IBGE	
	Produção de Motociclos	Unidades	Anfavea	
	Produção de Ônibus	Unidades	Anfavea	
	Produção de Caminhões	Unidades	Anfavea	
	Produção de Automóveis e Comerciais Leves	Unidades	Anfavea	
	Energia	Produção de Derivados do Petróleo	Barris/dia (mil)	
Consumo de Energia Elétrica Comercial		GWH	Eletrobras	
Consumo de Energia Elétrica Residencial		GWH	Eletrobras	
Consumo de Energia Elétrica Industrial		GWH	Eletrobras	
Serviços e Comércio	Índice de Confiança do Consumidor	Índice (base 100)	Fecomércio	2
	Índice de Expectativas Futuras	Índice (base 100)	Fecomércio	
Índice Geral de Preços e seus componentes	Índice Geral de Preços de Mercado e seus Componentes (IGP-M)	var % mensal	FGV	2
	Índice nacional de Preços ao Consumidor Amplo (IPCA)	var % mensal	IBGE	
Base Monetária	Papel moeda emitido (saldo em final de período)	u.m.c (mil)	BCB-DSTAT	7
	Reservas bancárias	u.m.c (mil)	BCB-DSTAT	
	Papel Moeda em Poder do Público	u.m.c (mil)	BCB-DSTAT	
Meios de Pagamento	M1	u.m.c (mil)	BCB-DSTAT	
	M2	u.m.c (mil)	BCB-DSTAT	

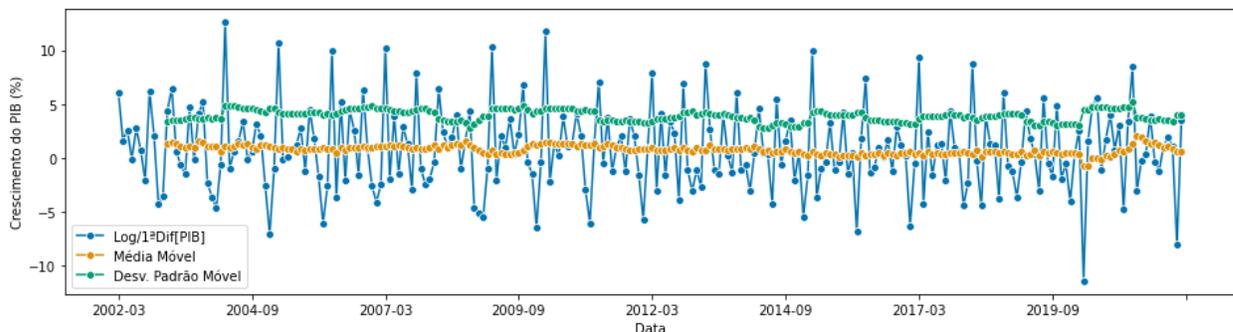
M3	u.m.c (mil)	BCB-DSTAT
M4	u.m.c (mil)	BCB-DSTAT

Fonte: Elaborado pelo autor (2023).

4.1 TRATAMENTO DOS DADOS

Assim como em Skorin (2022), optou-se por tornar estacionárias todas as séries originais não-estacionárias. É desejável realizar esse tipo de transformação pois séries temporais estacionárias apresentam propriedades úteis – como média e variância constantes ao longo da série temporal – que as tornam mais fáceis de modelar e analisar. Aplicou-se, portanto, o logaritmo aos dados, a fim de estabilizar a variância das séries. A partir do teste Dickey-Fuller, analisou-se quais séries eram estacionárias ou não. Caso a série apresentasse raiz unitária, aplicou-se a primeira diferença no regressor; caso já estacionária, optou-se por deixar a variável com seus valores originais. Desse processo, foi possível verificar que 19 (dezenove) das 27 (vinte e sete) séries passaram por uma transformação de estacionariedade. A Figura 8 mostra o exemplo da série PIB Mensal após a transformação; a Tabela 2 mostra os resultados do teste de Dickey-Fuller da série para o Crescimento do PIB.

Figura 8 - Crescimento percentual do PIB mês a mês



Fonte: Elaborado pelo autor (2023).

Tabela 2 - Resultado do teste de Dickey-Fuller para crescimento do PIB

Estatística do Teste	-3.81395
p-valor	0.00237
Valor Crítico (1%)	-3.11530
Valor crítico (5%)	-2.73642
Valor crítico (10%)	-2.39746

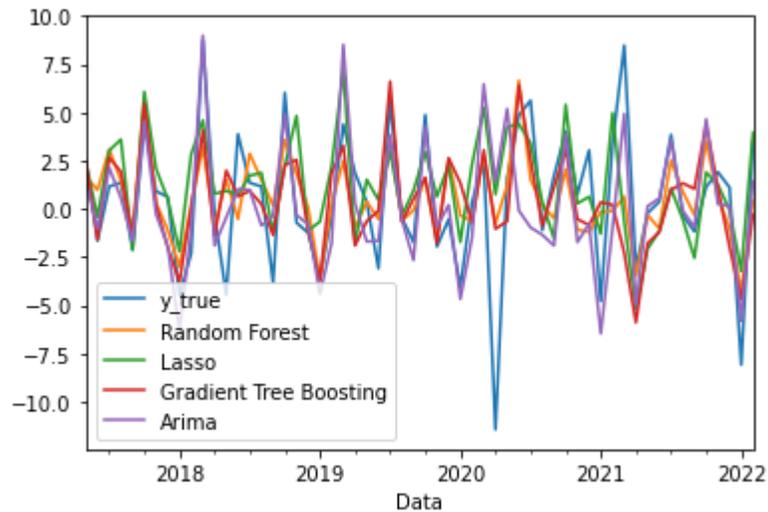
Fonte: Elaborado pelo autor (2023).

Optou-se também por criar defasagens de todas as séries, a fim de se obter uma base de dados maior. Para isso, foi realizada a defasagem de cada uma das variáveis selecionadas para até 12 (doze) pontos. Ou seja, considerando o tempo presente t para um regressor X , realizou-se defasagem onde $X = x_{(t)}, x_{(t-1)}, x_{(t-2)}, \dots, x_{(t-12)}$. Assim, perde-se 12 (doze) observações para a série ao mesmo tempo que se ganha novos preditores, totalizando, portanto, 230 observações e 340 regressores.

5 RESULTADOS

Neste capítulo, iremos demonstrar os resultados do exercício empírico de previsão da taxa de crescimento do PIB mensal do Brasil, com base nas métricas de avaliação discutidas na seção três. A Tabela 3 reporta os resultados para quatro horizontes temporais distintos: $h=1$, $h=3$, $h=6$ e $h=12$. Os resultados sombreados mostram os valores onde se teve o menor erro, ou seja, onde o modelo teve a melhor performance de previsão. A Tabela 4 mostra um resumo de quais foram os melhores modelos para cada horizonte temporal, enquanto que a Tabela 5 apresenta os valores do teste Diebold-Mariano.

Figura 9 - Valor real vs valores preditos para $h=1$



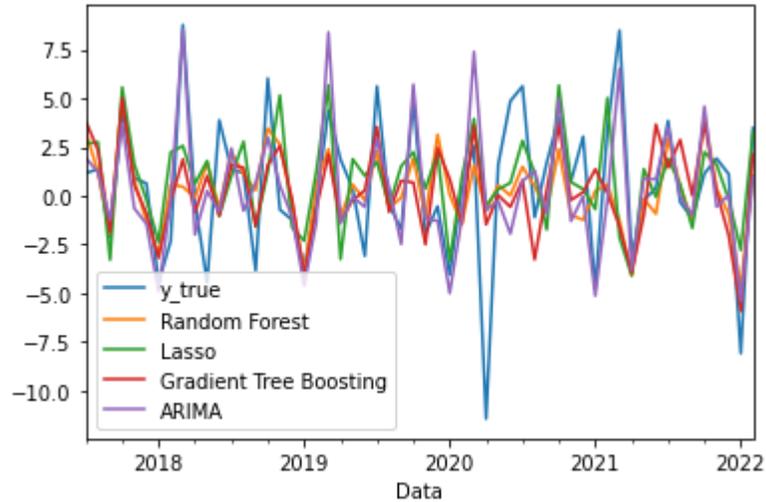
Fonte: Elaborado pelo autor (2023).

De maneira geral e analisando cada métrica de avaliação, notamos certa superioridade dos modelos de *ML* frente ao *benchmark ARIMA*³, além dos resultados se mostrarem consistentes ao longo de cada horizonte preditivo. Em termos do *MAPE*, nota-se que, para todos os horizontes temporais, o *benchmark* performou pior que os modelos baseados em árvores de decisão, enquanto que performou melhor que o modelo *LASSO* em horizontes temporais mais curtos ($h=1$ e $h=3$). O modelo *Gradient Tree Boosting* pode ser ranqueado em primeiro lugar para o horizonte temporal $h=1$, ficando o modelo *Random Forest* como segundo colocado. Para os horizontes

³ Para modelarmos nosso *benchmark ARIMA*, utilizamos o pacote *auto.arima* da linguagem de programação R. A ordem encontrada a partir deste foi $(p, d, q)(P, D, Q)$ foi $(2, 0, 1)(0, 1, 1)[12]$.

temporais restantes, considera-se o modelo *Random Forest* de maior poder preditivo, ficando o modelo *Gradient Tree Boosting* em segundo colocado.

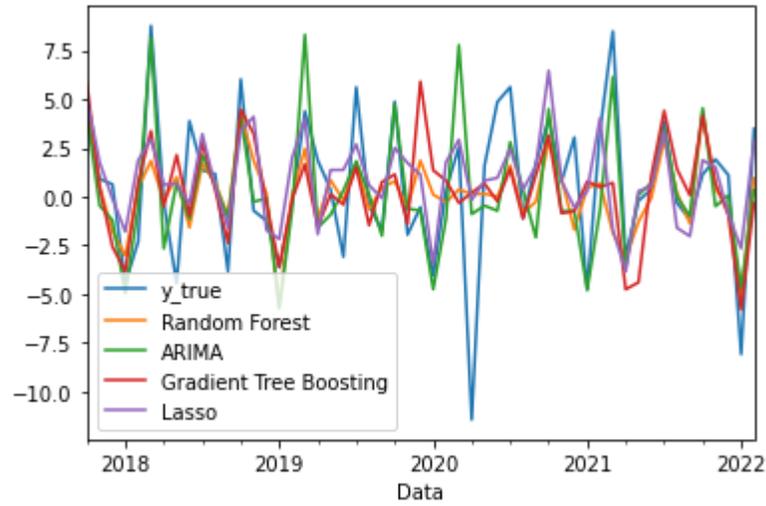
Figura 10 - Valor real vs valores preditos para $h=3$



Fonte: Elaborado pelo autor (2023).

Avaliando a métrica MSE, observa-se novamente superioridade preditiva dos algoritmos baseados em árvores de decisão frente aos demais modelos; além disso, nota-se que o modelo univariado performa melhor que o modelo linear *LASSO* em horizontes temporais $h=1$, $h=3$. Para $h=1$ e $h=3$, nota-se uma inversão dos primeiros colocados observados na métrica *MAPE*: o modelo *Random Forest* é ranqueado em primeiro lugar para o horizonte $h=1$, enquanto que o *Gradient Tree Boosting* é o modelo com menor erro para $h=3$. Nos horizontes temporais $h=6$ e $h=12$, relata-se os mesmos resultados encontrados no *MAPE*, com o modelo *Random Forest* sendo o primeiro colocado.

Figura 11 - Valor real vs valores preditos para $h=6$



Fonte: Elaborado pelo autor (2023).

Em relação aos valores apresentados pelo *MAE*, para horizontes de curto prazo, nota-se, novamente, o modelo *Random Forest* como melhor modelo, denotando robustez ao longo da avaliação dos modelos. No horizonte temporal $h=6$, o *benchmark ARIMA* se mostrou superior frente aos modelos concorrentes; para $h=12$, o modelo *Gradient Tree Boosting* pode ser considerado o melhor, ficando o modelo *ARIMA* em segundo colocado.

Tabela 3 - *MAPE*, *MSE* e *MAE* por horizonte temporal (h)

Métrica	Modelo	$h=1$	$h=3$	$h=6$	$h=12$
<i>MAPE</i>	<i>ARIMA</i>	1.23	1.34	1.72	1.95
	<i>LASSO</i>	1.31	1.39	1.41	1.63
	<i>Random Forest</i>	1.08	1.11	1.13	1.24
	<i>Gradient Tree Boosting</i>	1.02	1.16	1.21	1.27
<i>MSE</i>	<i>ARIMA</i>	10.88	11.43	13.27	14.89
	<i>LASSO</i>	11.48	11.91	13.14	14.46
	<i>Random Forest</i>	10.66	11.06	11.13	13.58
	<i>Gradient Tree Boosting</i>	10.76	10.88	12.36	13.59
<i>MAE</i>	<i>ARIMA</i>	2.16	2.38	2.61	2.92
	<i>LASSO</i>	2.37	2.49	2.71	3.02
	<i>Random Forest</i>	2.11	2.36	2.64	3.17
	<i>Gradient Tree Boosting</i>	2.25	2.41	2.63	2.84

Fonte: Elaborado pelo autor (2023).

Somando os horizontes temporais de cada modelo, encontramos os resultados da Tabela 4. Analisando a métrica *MAPE* e *MSE*, por exemplo, nota-se que os métodos de *ML* performaram

consideravelmente melhor, enquanto que para a métrica de avaliação *MAE* mostrou o modelo *ARIMA* com melhor capacidade preditiva para o horizonte temporal $h=6$. Dos modelos alternativos propostos, o modelo linear *LASSO* apresentou performance preditiva superior ao *benchmark* em $h=6$ e $h=12$ para *MAPE* e para o *MSE*, enquanto que o *ARIMA* se mostrou superior a esse em horizontes de mais curto prazo para as mesmas métricas de avaliação.

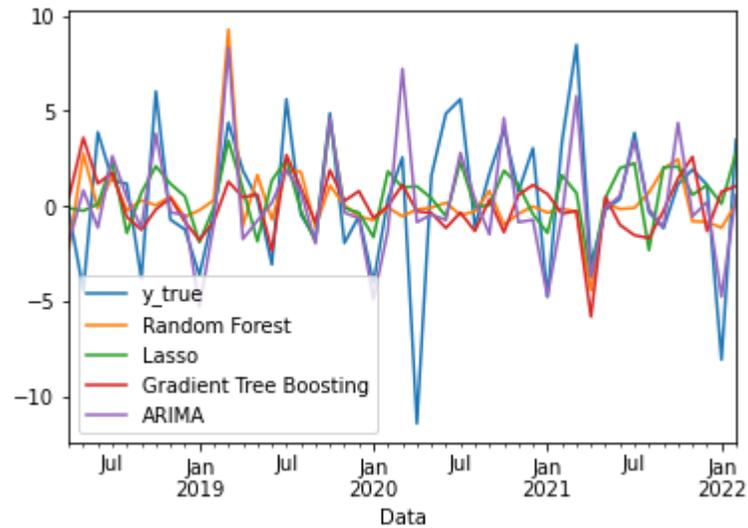
Tabela 4 - Resumo dos melhores modelos por métrica de avaliação

	<i>ARIMA</i>	<i>LASS</i>	<i>Random Forest</i>	<i>Gradient Tree Boosting</i>
	O			
<i>MAPE</i>	0	0	3	1
<i>MSE</i>	0	0	3	1
<i>MAE</i>	1	0	2	1

Fonte: Elaborado pelo autor (2023).

Até agora, usamos apenas os resultados encontrados pelas métricas de desempenho, sem entrar em contato com testes de significância estatística. A próxima análise aplica o teste Diebold-Mariano para comparar a precisão da previsão dos modelos apresentados contra o *benchmark ARIMA*. As propriedades do teste Diebold-Mariano provavelmente diferem sob a hipótese nula com modelos híbridos, pois os erros de previsão são assintoticamente iguais e, portanto, perfeitamente correlacionados (CLARK; McCRACKEN, 2001). No entanto, Giacomini e White (2006) provam que o teste é válido se estimarmos os modelos usando uma janela rolante, como é utilizado neste trabalho. A métrica de cálculo que utilizaremos para o cálculo do teste Diebold-Mariano é a *L2 (MSE)*.

Figura 12 - Valor real vs valores preditos para $h=12$



Fonte: Elaborado pelo autor (2023).

A tabela 5 apresenta os resultados do teste Diebold-Mariano e o seu valor-p para cada modelo contra o *benchmark ARIMA*. Caso a estatística do teste apresente sinal negativo, podemos interpretar como melhor performance preditiva do modelo concorrente ao *benchmark*; caso o resultado do teste apresente sinal positivo, assume-se que o *benchmark* é mais preciso que o modelo alternativo. Além disso, caso o valor-p seja menor que 0,05, a diferença de performance preditiva é significativa.

Tabela 5 - Teste Diebold-Mariano

	h=1		h=3		h=6		h=12	
	DM	valor-p	DM	valor-p	DM	valor-p	DM	valor-p
<i>RF</i>	-0.1942	0.0117	-0.2684	0.0356	-0.7142	0.0496	-0.8406	0.9643
<i>LASSO</i>	0.7891	0.2971	0.2381	0.9491	0.9341	0.6249	0.3581	0.0712
<i>Gradient Tree Boosting</i>	-0.2041	0.0387	-0.5195	0.2895	-0.8153	0.0311	-0.6423	0.5982

Fonte: Elaborado pelo autor (2023).

Para o horizonte $h=1$, conclui-se que tanto o modelo *Random Forest* quanto o modelo *Gradient Tree Boosting* apresentaram performance preditiva superior ao *benchmark* para o nível de significância de 5%; para $h=3$, apenas o modelo *RF* apresentou performance preditiva superior ao *benchmark* para o nível de significância de 5%; para $h=6$, conclui-se o mesmo resultado que tivemos para $h=1$; para $h=12$, nenhum dos modelos alternativos propostos apresentaram performance preditiva significativamente superior ao *benchmark ARIMA*.

De maneira geral, os modelos concorrentes ao modelo *ARIMA* performaram bem para todos horizontes temporais, mostrando não apenas erros menores para previsões fora da amostra mas também maior acurácia quando testados contra o *benchmark*.

5.1 PREDITORES SELECIONADOS

Após investigação dos resultados de capacidade preditiva dos modelos de *ML* propostos, analisaremos agora o funcionamento de escolha dos preditores de cada algoritmo e também a importância dada a cada regressor para cada modelo. Cabe aqui ressaltar que a importância verificada não significa em si uma inferência causal, mas sim a lógica de pensamento do algoritmo para extrair insights sobre seu funcionamento e, eventualmente, encontrar correlações entre a variável dependente – nesse caso, a taxa de crescimento do PIB brasileiro – e seus regressores (SKORIN, 2022). A análise também é válida para comparar as escolhas de diferentes modelos em diferentes horizontes de previsão.

O modelo *LASSO* adiciona um termo de penalidade à função custo do OLS. Esse termo de penalidade, como já comentado na seção 3 deste trabalho, é baseado na norma L1 dos coeficientes, que incentiva os coeficientes de alguns regressores serem reduzidos a zero ou totalmente eliminados. Isso pode ser visto como uma forma de seleção de variáveis, pois apenas as características mais importantes são mantidas no modelo (TIBSHIRANI, 1996).

No algoritmo *Random Forest*, a importância dos regressores é calculada medindo a diminuição da impureza das árvores de decisão quando um regressor em específico é utilizado para particionar os dados (STROBL, *et al*, 2007). Essa medida é conhecida como importância de Gini; regressores que são usados com maior frequência no particionamento dos dados e resultam em menor impureza são considerados os mais importantes. Já em um algoritmo *Gradient Tree Boosting*, a importância da variável é calculada medindo a redução total na função perda, que é alcançada por cada regressor no conjunto (*ensemble*) de árvores de decisão (FRIEDMAN, 2001). Os regressores que mais contribuem para a reduzir a função perda são considerados os mais importantes.

A fim de realizarmos uma análise de importância de preditores mais clara e simples, optou-se por escolher os cinco preditores mais importantes para cada modelo para todos os

horizontes temporais propostos. Cabe aqui ressaltar que se considera o regressor de maneira geral, ou seja, considerando também suas defasagens – lembrando: 12 defasagens para cada regressor.

Analisando para o horizonte $h=1$, nota-se a importância dos Indicadores de Produção para todos os modelos propostos. Para os modelos baseados em árvores de decisão, é possível observar que a Indústria de Transformação é considerada um elemento chave para o resultado do modelo, sendo a variável mais importante tanto no modelo *Random Forest* quanto no modelo *Gradient Tree Boosting*. Além disso, outros indicadores de produção também apresentam relevante importância para esses algoritmos, como a produção dos Bens de Consumo e a produção de Bens de Capital. As variáveis monetárias – M3 e Papel moeda emitido – não apresentaram grande relevância na explicação desses modelos. No modelo linear *LASSO*, cabe notar a grande importância da produção de Bens Intermediários e também Bens de Consumo, e também a baixa relevância de variáveis monetárias para explicar o crescimento do PIB.

Tabela 6 - Top 5 preditores selecionados ($h=1$)

Preditores selecionados pelo modelo <i>LASSO</i>	Importância
Indicador de Produção – Bens Intermediários (base 100)	31,42%
Indicador de Produção – Bens de consumo (base 100)	10,07%
Consumo de Energia Elétrica Industrial – GWH	3,18%
Índice nacional de Preços ao Consumidor Amplo (IPCA) - var % mensal	1,19%
Produção de Caminhões – unidades	0,93%
Preditores selecionados pelo modelo <i>Random Forest</i>	Importância
Indicador de Produção – Indústria de Transformação (base 100)	23,17%
Indicador de Produção – Bens de consumo (base 100)	8,41%
Indicador de Produção – Semiduráveis e não duráveis (base 100)	4,56%
Produção de Automóveis e Comerciais Leves – unidades	2,17%
M3 – u.m.c	1,06%
Preditores selecionados pelo modelo <i>Gradient Tree Boosting</i>	Importância
Indicador de Produção – Indústria de Transformação (base 100)	26,89%
Indicador de Produção – Bens de capital (base 100)	10,76%
Papel moeda emitido (saldo em final de período)	4,55%
Produção de Derivados do Petróleo – barris/dia	2,08%
Indicador de Produção – Semiduráveis e não-duráveis	2,21%

Fonte: Elaborado pelo autor (2023).

Para horizontes intermediários ($h=3$ e $h=6$), observa-se certa estabilidade do que foi observado para $h=1$, mas com a inclusão e exclusão de algumas variáveis antes observadas no curto prazo. Para o modelo *Random Forest*, observou-se uma mudança no ranqueamento do regressor mais importante; a variável Bens de Consumo é considerada a mais importante tanto

para $h=3$ quanto para $h=6$, enquanto que para $h=1$ a Indústria de Transformação foi considerada a mais relevante. Além disso, o Indicador de Produção de Semiduráveis e Não Duráveis ganhou maior importância nos horizontes de médio prazo para este algoritmo. Para o modelo *Gradient Tree Boosting*, a importância das variáveis pouco mudou, mantendo-se tanto a Indústria de Transformação quanto os Bens de Capital como os regressores de maior relevância. As variáveis monetárias – Papel moeda emitido e M3 – ganharam maior importância para esses horizontes temporais de médio prazo. Para o modelo linear *LASSO*, houve um aumento substancial da importância para o regressor Indicador de Produção de Semiduráveis e não duráveis, que para $h=1$ não se fez presente.

Tabela 7 - Top 5 preditores selecionados ($h=3$)

Preditores selecionados pelo modelo <i>LASSO</i>	Importância
Indicador de Produção – Bens Intermediários (base 100) -	21,71%
Indicador de Produção – Semiduráveis e não duráveis (base 100)	19,64%
Indicador de Produção – Bens de consumo (base 100) -	6,92%
Consumo de Energia Elétrica Comercial – GWH -	2,84%
Indicador de Produção – Bens de capital (base 100)	1,75%
Preditores selecionados pelo modelo <i>Random Forest</i>	Importância
Indicador de Produção – Bens de consumo (base 100)	16,18%
Indicador de Produção – Semiduráveis e não duráveis (base 100)	11,74%
Produção de Automóveis e Comerciais Leves – unidades	3,07%
M3 – u.m.c	1,65%
Indicador de Produção – Bens de capital (base 100)	0,38%
Preditores selecionados pelo modelo <i>Gradient Tree Boosting</i>	Importância
Indicador de Produção – Indústria de Transformação (base 100)	13,45%
Indicador de Produção – Bens de capital (base 100)	10,02%
Papel moeda emitido (saldo em final de período)	9,63%
Indicador de Produção – Semiduráveis e não-duráveis	4,01%
Produção de Derivados do Petróleo – barris/dia	1,31%

Fonte: Elaborado pelo autor (2023).

Tabela 8 - Top 5 preditores selecionados ($h=6$)

Preditores selecionados pelo modelo <i>LASSO</i>	Importância
Indicador de Produção – Bens Intermediários (base 100) -	22,81%
Indicador de Produção – Semiduráveis e não duráveis (base 100)	13,97%
Indicador de Produção – Bens de consumo (base 100) -	6,72%
Consumo de Energia Elétrica Comercial – GWH -	5,39%
Produção de Automóveis e Comerciais Leves – unidades	2,46%
Preditores selecionados pelo modelo <i>Random Forest</i>	Importância
Indicador de Produção – Bens de consumo (base 100)	18,09%
Indicador de Produção – Semiduráveis e não duráveis (base 100)	14,91%
Consumo de Energia Elétrica Residencial – GWH	12,05%
M3 – u.m.c	7,83%
Reservas Bancárias – u.m.c	4,09%
Preditores selecionados pelo modelo <i>Gradient Tree Boosting</i>	Importância
Indicador de Produção – Indústria de Transformação (base 100)	14,07%

Indicador de Produção – Bens de capital (base 100)	9,82%
Papel moeda emitido (saldo em final de período)	4,81%
Indicador de Produção – Semiduráveis e não-duráveis	2,09%
Produção de Derivados do Petróleo – barris/dia	0,93%

Fonte: Elaborado pelo autor (2023).

Para o horizonte temporal de mais longo prazo $h=12$, observa-se uma diluição de importância de cada variável, mas ainda mantendo certo nível de ranqueamento de importância ao longo dos horizontes temporais. Para o modelo linear *LASSO*, nota-se que o Indicador de Produção de Bens Intermediários se manteve como o regressor mais importante durante toda a série, mas tendo sua importância diminuída com o passar do tempo. O modelo *Random Forest* e *Gradient Tree Boosting* mantiveram suas duas variáveis mais importantes no médio-longo prazo, tendo essas sido modificadas apenas de $h=1$ para $h=3$. Importante ressaltar que os regressores mais importantes durante toda a série foram os ligados ao lado real da economia, enquanto que preditores ligados ao lado monetário foram pouco valorizados pelo modelo. Podemos concluir dessa análise que as variáveis monetárias não se apresentam como variável relevante para a previsão da taxa de crescimento do PIB brasileiro.

Tabela 9 - Top 5 preditores selecionados ($h=12$)

Preditores selecionados pelo modelo <i>LASSO</i>	Importância
Indicador de Produção – Bens Intermediários (base 100) -	16,41%
Indicador de Produção – Semiduráveis e não duráveis (base 100)	8,53%
Indicador de Produção – Bens de consumo (base 100) -	5,12%
Consumo de Energia Elétrica Comercial – GWH -	3,96%
Produção de Automóveis e Comerciais Leves – unidades	3,13%
Preditores selecionados pelo modelo <i>Random Forest</i>	Importância
Indicador de Produção – Bens de consumo (base 100)	10,92%
Indicador de Produção – Semiduráveis e não duráveis (base 100)	6,94%
Consumo de Energia Elétrica Residencial – GWH	6,01%
M3 – u.m.c	4,85%
Reservas Bancárias – u.m.c	4,01%
Preditores selecionados pelo modelo <i>Gradient Tree Boosting</i>	Importância
Indicador de Produção – Indústria de Transformação (base 100)	12,94%
Indicador de Produção – Bens de capital (base 100)	6,41%
Papel moeda emitido (saldo em final de período)	5,93%
Indicador de Produção – Semiduráveis e não-duráveis	5,71%
Produção de Derivados do Petróleo – barris/dia	5,02%

Fonte: Elaborado pelo autor (2023).

6 CONCLUSÃO

Este trabalho tinha como objetivo verificar a possibilidade dos algoritmos de *ML* realizarem um *forecasting* da taxa de crescimento do PIB brasileiro com maior poder de predição e acurácia do que métodos já tradicionais da literatura econométrica. Comparou-se três algoritmos bem consolidados no campo da inteligência artificial – *LASSO*, *Gradient Tree Boosting* e *Random Forest* – com o *benchmark ARIMA*. A partir das métricas de avaliação, evidencia-se certa superioridade dos métodos de *ML* frente ao *benchmark*, principalmente quando falamos do algoritmo de *Random Forest*, que apresentou os menores erros em 67% dos resultados, podendo ser considerado, portanto, uma ferramenta útil em exercícios de predição, principalmente em horizontes de curto prazo. O modelo *ARIMA*, no entanto, não deve ser descartado como ferramental em análises de *forecasting*, uma vez que também apresentou os menores erros em horizontes temporais mais longos quando utilizada a métrica *MAE*.

Para testar a acurácia dos nossos modelos, foi utilizado o teste Diebold-Mariano. Dos doze testes realizados, cinco modelos concorrentes apresentaram performance preditiva superior ao *benchmark* para nível de significância de 5%.

Uma possível extensão deste trabalho seria a utilização do *Model Confidence Set (MCS)*, introduzido por Hansen, Lunde e Nason (2011). A ideia por trás do *MCS* é que, em vez de escolher um único modelo e tratá-lo como o "melhor", pode ser mais apropriado considerar um conjunto de modelos, cada um dos quais pode ser o melhor em diferentes situações. A abordagem *MCS* pode ser vista como uma forma de combinar os pontos fortes de vários modelos e mitigar seus pontos fracos. Ao considerar um conjunto de modelos, o *MCS* fornece uma abordagem mais robusta e flexível para a previsão do que depender de um único modelo. Além do *MCS*, uma possível extensão do trabalho é a utilização de combinação de modelos, uma vez que estes podem incorporar pontos fortes de ambas abordagens e, assim, aumentar a precisão das previsões.

REFERÊNCIAS BIBLIOGRÁFICAS

ALI, Jahed; KHAN, Rehanullah; AHMAD, Nasir; MAQSOOD, Irman. Random Forests and Decision Trees. **International Journal of Computer Science Issues**, v. 9, n. 3, p. 271-278, 2012. Disponível em: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.402.3863&rep=rep1&type=pdf>. Acesso em: 04 set. 2022.

ARAUJO, Gustavo Silva; GAGLIANONE, Wagner Piazza. **Machine learning methods for inflation forecasting in Brazil: new contenders versus classical models**. 2020. Disponível em: https://www.cemla.org/actividades/2020-final/2020-10-xxv-meeting-cbrn/Session%202/3.%20Machine_Learning...%20Wagner%20Piazza.pdf. Acesso em: 04 mai. 2022.

ATHEY, S. The Impact of Machine Learning on Economics. In: AGRAWAL, A.; GANS, J.;

BAYER, Fábio. Mariano; SOUZA, Adriano Mendonça. Wavelets e modelos tradicionais de previsão: Um estudo comparativo. **Revista Brasileira Biometria**, v. 28, n. 2, p. 40–61, 2010. Disponível em: http://jaguar.fcav.unesp.br/RME/fasciculos/v28/v28_n2/A3_Fabio_Adriano.pdf. Acesso em: 10 mar. 2023.

BEYER, William Huntley. **CRC Standard Mathematical Tables**, 31st ed. Boca Raton: CRC Press, pp. 536-571, 2002.

BREIMAN, Leo. Random Forests. **Machine Learning**, v. 45, p. 5–32, 2001.

Breiman, Leo. Random Forests. **Machine Learning**, v. 45, p. 5–32, 2001. Disponível em: <https://doi.org/10.1023/A:1010933404324>. Acesso em 10 mar. 2023.

CEPNI, Oguzhan, GÜNEY, I. Ethem, SWANSON, Norman R. Nowcasting and forecasting GDP in emerging markets using global financial and macroeconomic diffusion indexes. **International Journal of Forecasting**, v. 35, n. 2, p. 555-572, 2019.

CHENG, Kayan; HUANG, Naijing; SHI, Zhentao. Survey-Based Forecasting: To Average or Not To Average. In: SRIBOONCHITTA, Songsak; KREINOVICH, Vladik; YAMAKA, Woraphon. (eds). **Behavioral Predictive Modeling in Economics**. Studies in Computational Intelligence, v. 897, p. 87-104, 2019. Disponível em: https://doi.org/10.1007/978-3-030-49728-6_5. Acesso em: 07 set. 2022.

Clark, Todd E.; MCCracken, Michael. W. Tests of equal forecast accuracy and encompassing for nested models. **Journal of Econometrics**, n. 105, p. 85–110, 2001. Disponível em: <https://EconPapers.repec.org/RePEc:eee:econom:v:105:y:2001:i:1:p:85-110>. Acesso em: 10 mar. 2023.

DIEBOL, Francis; MARIANO, Roberto. Comparing Predictive Accuracy. **Journal of Business & Economic Statistics**, v. 13, n. 3, p. 253-63, 1995. Disponível em: <https://EconPapers.repec.org/RePEc:bes:jnlbes:v:13:y:1995:i:3:p:253-63>. Acesso em: 10 mar. 2023.

FLANNERY, Ronan. A Machine Learning Approach to Predicting Gross Domestic Product. **MSc Research Project in School of Computing National College of Ireland**, 2020. Disponível em: <http://norma.ncirl.ie/4441/>. Acesso em: 04 mai. 2022.

FRIEDMAN, Jerome H. Greedy function approximation: A gradient boosting machine. **Ann. Statist.**, v. 29, n. 5, p. 1189 - 1232, 2001. Disponível em: <https://doi.org/10.1214/aos/1013203451>. Acesso em: 10 mar. 2023.

FRIEDMAN, Jerome; HALL, Peter. On Bagging and Nonlinear Estimation. **Journal of Statistical Planning and Inference**, v. 137, n. 3, p. 669–83, 2007. Disponível em: <https://doi.org/10.1016/j.jspi.2006.06.002>. Acesso em: 09 set. 2022.

GÉRON, Aurélien. **Hands-on Machine Learning with Scikit-Learn, Keras & TensorsFlow**. 2. ed. Sebastopol: O'Reilly Media, 2019.

GIACOMINI, Raffaella; WHITE, Halbert. Tests of conditional predictive ability. **Econometrica**, v. 74, n. 6, p. 1545–1578, 2006. Disponível em: <https://www.jstor.org/stable/4123083>. Acesso em 10 mar. 2023.

GIANNONE, Domenico; REICHLIN, Lucrezia; SMALL, David. Nowcasting: The real-time informational content of macroeconomic data. **Journal of Monetary Economics**, v. 55, p. 665–676, 2008.

GUERZONI, Marco; NAVA, Consuelo R.; NUCCIO, Massimiliano. The survival of start-ups in time of crisis. A machine learning approach to measure innovation. **General Economics**, v. 1, 2019. Disponível em: <https://doi.org/10.48550/arXiv.1911.01073>. Acesso em: 04 set. 2022.

HANSEN, Peter; R., ASGER, Lunde; JAMES M. Nason. The model confidence set. **Econometrica**, v. 79, n. 2, p. 453–97, 2011. Disponível em: <http://www.jstor.org/stable/41057463>. Acesso em: 10 mar. 2023.

Horning, Ned. Random Forests: An algorithm for image classification and generation of continuous fields data sets. **International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences**, 2010. Disponível em: <https://www.semanticscholar.org/paper/Random-Forests-%3A-An-algorithm-for-image-and-of-data-Horning/9484240b4dfe10a40b80037d6869c4b035615be9>. Acesso em: 10 mar. 2023.

HUD USER. **American Housin Survey**, 2015. AHS Updates. Disponível em: <https://www.huduser.gov/portal/datasets/ahs.html>. Acesso em: 04 set. 2022.

IBM Cloud Education. **Overfitting**. What is overfitting? 2021. Disponível em: <https://www.ibm.com/cloud/learn/overfitting>. Acesso em: 03 set. 2022.

JORDAN, Michael Irwin; MITCHELL, Tom Michael. Machine learning: Trends, perspectives and prospects. **Science**, v. 349, n. 6245, p. 255-60, jul. 2015. Disponível em: <https://www.science.org/doi/10.1126/science.aaa8415>. Acesso em: 03 set. 2022.

LANTZ, Brett. **Machine Learning with R**. 2. ed. Birmingham: Pack Publishing, 2015. Disponível em: <https://archive.org/details/Machine.Learning.with.R.2nd.Edition.2015.Lantz>. Acesso em: 03 set. 2022.

LÖNING, Markus; KIRÁLY, Franz. Forecasting with sktime: Designing sktime's New Forecasting API and Applying It to Replicate and Extend the M4 Study. **Machine Learning**, v. 1, 2020. Disponível em: <https://doi.org/10.48550/arXiv.2005.08067>. Acesso em: 07 set. 2022.

MACCARRONE, Giovanni, MORELLI, Giacomo; SPADACCINI, Sara. GDP Forecasting: Machine Learning, Linear or Autoregression? **Frontiers in artificial intelligence**, v. 4, n. 757864, 2021. Disponível em: <https://doi.org/10.3389/frai.2021.757864>. Acesso em: 03 set. 2022.

MEDEIROS, Marcelo C.; VASCONCELOS, Gabriel F. R.; VEIGA, Álvaro e ZILBERMAN, Eduardo. Forecasting inflation in a data-rich environment: the benefits of machine learning methods. **Journal of Business & Economic Statistics**, v. 39, n. 1, p. 98–119, 2021. Disponível em: <https://doi.org/10.1080/07350015.2019.1637745>. Acesso em 10 mar. 2023.

MULLAINATHAN, Sendhil; SPIESS, Jann. Machine Learning: An Applied Econometric Approach. **Journal of Economic Perspectives**, v. 31, n. 2, p. 87–106, 2017. Disponível em: <https://doi.org/10.1257/jep.31.2.87>. Acesso em: 03 set. 2022.

NYMAN, Rickard; ORMEROS, Paul. Understanding the Great Recession Using Machine Learning Algorithms. Bank of England/Federal Reserve conference on “Modelling with Big Data and Machine Learning”. **Bank of England**. November 2018. 2020. Disponível em: https://www.researchgate.net/publication/338447333_Understanding_the_Great_Recession_Using_Machine_Learning_Algorithms. Acesso em 10 mar. 2023.

PADILHA, J. B. J. *et al.* Modelagem box-jenkins (arima) para gestão de risco de preço na pecuária de corte do estado do paran . **X Congresso Virtual Brasileiro de Administra o–CONVIBRA**. [S.l.: s.n.]. p. 12, 2013.

PINTO, Hugo. The role of econometrics in economic science: an essay about the monopolization of economic methodology by econometric methods. **Journal of Socio-Economics**, v. 40, n. 4, p. 436- 443, 2011. Disponível em: <https://ideas.repec.org/a/eee/soceco/v40y2011i4p436-443.html>. Acesso em 10 mar. 2023.

RANSTAM Jonas; COOK J A. LASSO regression. **British Journal of Surgery**, v. 105, i. 10, Page 1348, 2018. Disponível em: <https://doi.org/10.1002/bjs.10895>. Acesso em: 09 set. 2022.

SABA, Nicole de Mendonça. **Avaliando o desempenho preditivo de modelos de taxa de c mbio real efetiva: an lise do caso brasileiro**. Disserta o (Mestrado em Economia) – Escola de Economia de S o Paulo, Funda o Get lio Vargas. S o Paulo, p. 70, 2015. Disponível em: <http://hdl.handle.net/10438/14091>. Acesso em: 10 mar. 2023.

SKTIME. Window Splitters in Sktime. 2021. Disponível em: https://www.sktime.net/en/latest/examples/forecasting/window_splitters.html?highlight=sliding#Sliding-windows-using-SlidingWindowSplitter. Acesso em: 10 mar. 2023.

SOUZA, V. *et al.* **Utiliza o da metodologia de box & jenkins na previs o do pre o futuro pago as exporta es paranaenses de madeira serrada**. Universidade Federal do Paran , 2006.

STROBL, Carolin; BOULESTEIX, Anne-Laure, ZEILEIS, Achim; HOTHORN Torsten. Bias in random forest variable importance measures: illustrations, sources and a solution. **BMC Bioinformatics**, v. 25, n. 8, 2007. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/17254353/>. Acesso em: 10 mar. 2023.

TIBICO CLOUD. **O que é uma floresta aleatória?**, 2022. Disponível em: <https://www.tibco.com/pt-br/reference-center/what-is-a-random-forest>. Acesso em: 04 set. 2022.

TIBSHIRANI, Robert. Regression Shrinkage and Selection via the Lasso. **Journal of the Royal Statistical Society. Series B (Methodological)**, v. 58, n. 1, p. 267–88, 1996. Disponível em: <https://www.jstor.org/stable/2346178>. Acesso em: 10 mar. 2023.

URANGA, Pedro Pablo Skorin. **Forecasting with high-dimensional data using linear and P-splines L2-boosting : an exercise for the unemployment rate in Brazil**. Trabalho de Conclusão de Curso (Graduação em Economia) – Faculdade de Ciências Econômicas, Universidade Federal do Rio Grande do Sul. Porto Alegre, p. 57, 2022. Disponível em: <http://hdl.handle.net/10183/241209>. Acesso em: 10 mar. 2023.

VARIAN, Hal Ronald. Big Data: New Tricks for Econometrics. **Journal of Economic Perspectives**, v. 28, n. 2, p. 3-28, 2014. Disponível em: <http://dx.doi.org/10.1257/jep.28.2.3>. Acesso em: 03 set. 2022.

XIE, Tian; LIU, Yan. Machine learning versus econometrics: prediction of box office. **Applied Economics Letters**, 2018. Disponível em: <https://doi.org/10.1080/13504851.2018.1441499>. Acesso em: 03 set. 2022.

YOON, Jaehyun. Forecasting of Real GDP Growth Using Machine Learning Models: Gradient Boosting and Random Forest Approach. **Comput Econ**, v. 57, p. 247–265, 2021. Disponível em: <https://doi.org/10.1007/s10614-020-10054-w>. Acesso em: 03 set. 2022.