

Stacked Generalization na Previsão de Retornos na B3

José Márcio Arcanjo dos Santos

Universidade Federal de Pernambuco – UFPE
marcio.arcanjo@ufpe.br

Joséte Florêncio dos Santos

Universidade Federal de Pernambuco – UFPE
josete.santos@ufpe.br

Resumo

Este estudo investiga se a técnica de *stacked generalization* em *machine learning* potencializa as previsões dos retornos das ações na B3. Para isso, combina-se um conjunto de cinco modelos de precificação de ativos — CAPM, *Learning CAPM*, *Downside CAPM*, *Conditional CAPM* e a regressão *cross-section* de Fama e French (1996) — com quatro meta-modelos: *Random Forest*, *Gradient Boosted Regression Trees*, *Neural Networks* e *Linear Regression*. A amostra é composta por 202 empresas, com dados mensais de janeiro de 2003 a dezembro de 2022, e a avaliação empírica é conduzida por meio de janelas de expansão, com foco no R^2 fora da amostra e em medidas de previsão de sinal. Os resultados mostram que o *stacked generalization* melhora o desempenho preditivo em diferentes contextos, com destaque para o meta-modelo linear, mas não supera de forma significativa a média histórica durante choques exógenos, como a pandemia de Covid-19.

Palavras-chave: apreçamento de ativos, previsão de retornos, *stacked generalization*, *machine learning*, B3.

Classificação JEL: C13; C22; C69; C81; G12; G17

1 Introdução

Há um longo debate sobre se os preços das ações são previsíveis. O tema é importante porque desempenha um papel fundamental em diversas áreas de finanças, como precificação de ativos, alocação de portfólio e avaliação de desempenho de gestores de investimentos. No entanto, a previsibilidade do retorno das ações ainda é controversa e a especificação do melhor modelo permanece uma questão em aberto (Timmermann, 2008). Nos últimos 50 anos, uma quantidade significativa de pesquisas foi realizada sobre o tema, com destaque para duas categorias: (i) estudos que buscam encontrar o melhor modelo de previsão individual, e (ii) estudos que se concentram em combinar previsões (Wang *et al.*, 2022). Esse debate permanece atual porque a instabilidade dos mercados e a heterogeneidade dos ativos tornam improvável que um único modelo seja suficientemente robusto em diferentes contextos.

O argumento principal para combinar previsões na precificação de ativos de capital é a crença de que a previsão de retorno das ações não pode ser descrita eficientemente por um único modelo, e que modelos de *machine learning*, como o *stacked generalization*, são capazes de melhorar significativamente a explicação do retorno das ações por meio da combinação de diferentes modelos que utilizam algoritmos sofisticados (Zhao e Cheng, 2022). Nessa perspectiva, a combinação de previsões não representa apenas uma alternativa metodológica, mas uma resposta à possibilidade de erro de especificação associada ao uso isolado de um único modelo. Isso é particularmente relevante em finanças, onde relações lineares e não lineares podem coexistir e variar ao longo do tempo.

Zhao e Cheng (2022) descobriram que realizar o *stacked generalization* com um meta-modelo um pouco mais complexo pode superar o *benchmark* da média histórica tradicional em termos de desempenho dentro e fora da amostra. Além disso, o *stacked generalization* geralmente pode ter um desempenho melhor do que vários outros modelos em termos de previsão. Os autores também descobriram que os ganhos de previsão fora da amostra aumentam em mercados de desaceleração extrema. Para os autores, a principal diferença entre o *stacked generalization* e os métodos tradicionais de combinação de previsões está em dois aspectos: (i) o meta-modelo pode ser uma função complicada, enquanto os métodos tradicionais usam uma função simples, como média ou mediana; e (ii) os modelos base podem ser modelos lineares ou não lineares, ao passo que os métodos tradicionais de combinação normalmente utilizam apenas modelos lineares (Zhao e Cheng, 2022). Essas evidências reforçam a pertinência de investigar se tais ganhos também podem ser observados

no mercado acionário brasileiro.

Diante deste cenário de inovações e avanços tecnológicos, surge a seguinte pergunta de pesquisa: as previsões dos retornos das ações dos modelos de precificação de ativos são potencializadas quando a técnica de *stacked generalization* em *machine learning* é aplicada nos ativos da B3? A relevância dessa questão decorre não apenas de seu interesse acadêmico, mas também de suas implicações práticas para investidores, gestores e pesquisadores interessados em aprimorar a previsão de retornos em ambientes marcados por incerteza e mudanças de regime.

Para responder a essa questão, a amostra foi composta por 202 empresas com informações disponíveis no período. Para aplicação da técnica de *stacked generalization* foram considerados cinco modelos de precificação de ativos recorrentes na literatura: *Capital Asset Pricing Model* (CAPM), *Learning CAPM* (L-CAPM), *Downside CAPM* (D-CAPM), *Conditional CAPM* (C-CAPM) e a Regressão *Cross-section* de Fama e French (1996), e quatro algoritmos de *machine learning*: *Random Forest* (RF), *Gradient Boosted Regression Trees* (GBRT), *Neural Networks* (NN) e *Linear Regression* (LR). Para atingir os objetivos da pesquisa, a amostra foi dividida em três conjuntos de dados em um esquema de janelas de expansão, em que os dados são segmentados em partes, seguindo um formato de estimativas recursivas. A eficácia das previsões foi avaliada por meio dos ganhos estatísticos, mensurados pelo R^2 fora da amostra (R_{OOS}^2) de Campbell e Thompson (2008), e por medidas de previsão de sinal, como acurácia, sensibilidade e especificidade.

Os resultados evidenciam que a aplicação da técnica de *stacked generalization* para combinar modelos de precificação de ativos pode melhorar significativamente os resultados das previsões em diferentes contextos e condições de mercado. Além disso, treinar os meta-modelos de *stacked generalization* com a seleção dos melhores modelos base pode melhorar a previsão de acurácia negativa, especialmente em períodos de queda. Todavia, as evidências indicam que os modelos estudados foram incapazes de superar a média histórica dos retornos das ações de forma significativa durante um choque exógeno. Em conjunto, esses resultados sugerem que os ganhos associados à combinação de previsões existem, mas não são uniformes, estando condicionados ao ambiente de mercado e à natureza do choque enfrentado.

Assim, este estudo contribui para a literatura de precificação de ativos ao investigar, no contexto da B3, se a combinação de previsões via *stacked generalization* pode ampliar o desempenho preditivo de modelos tradicionais e recentes de apreçamento de ativos, especialmente em termos de capacidade preditiva fora da amostra. Além desta introdução, o

artigo está estruturado da seguinte forma. A próxima seção apresenta o referencial teórico e o posicionamento do estudo na literatura. Em seguida, são descritos os dados, os modelos e os procedimentos metodológicos adotados. Posteriormente, são apresentados e discutidos os resultados empíricos. Por fim, a última seção apresenta as conclusões do estudo.

2 Revisão da Literatura

Os modelos de precificação de ativos ocupam posição central na teoria financeira por buscarem explicar a relação entre risco e retorno no mercado de capitais. Entre os principais modelos utilizados nessa tradição, destacam-se o Capital Asset Pricing Model (CAPM), desenvolvido por Sharpe (1964), Lintner (1965), Mossin (1966) e Black (1972), o Arbitrage Pricing Theory (APT) de Ross (1976), e os modelos multifatoriais de Fama e French (1993; 2015) e Carhart (1997). Em conjunto, essas abordagens representam avanços importantes na tentativa de explicar a variação dos retornos das ações, ao relacionar o comportamento dos ativos a fatores sistemáticos de risco e a características observáveis das firmas. Apesar de suas diferenças, esses modelos compartilham a ambição de oferecer uma estrutura analítica capaz de melhorar a compreensão e a previsão dos retornos no mercado de capitais.

Todavia, os resultados empíricos desses modelos não são convergentes. Estudos como os de Friend e Blume (1970), Jensen, Black e Scholes (1972) e Fama e French (2004) apontam limitações importantes do CAPM na explicação do comportamento observado dos retornos. De forma semelhante, a literatura posterior mostrou que mesmo modelos multifatoriais mais sofisticados não eliminam integralmente os problemas de instabilidade, erro de especificação e sensibilidade a diferentes contextos de mercado. Nesse sentido, autores como Wei (1988), Chen, Roll e Ross (1986), Rapach, Strauss e Zhou (2010), Rapach e Zhou (2020) e Zhao e Cheng (2022) argumentam que a combinação de diferentes previsões pode oferecer uma explicação mais robusta para os retornos médios das ações, o prêmio de risco de capital e o retorno de títulos corporativos.

A literatura de combinação de previsões parte da ideia de que previsões individuais podem conter informações complementares. Bates e Granger (1969) são considerados os primeiros a sistematizar um método para combinar os resultados dos modelos de previsão mais eficientes, com a finalidade de agregar as características mais relevantes de cada modelo utilizado. O argumento central é que uma previsão descartada pode conter informações independentes e potencialmente úteis, seja porque utiliza variáveis distintas, seja porque assume formas diferentes de relacionamento entre variáveis. Ao longo das décadas seguintes,

a literatura financeira passou a registrar diferentes estratégias de combinação de previsões, incluindo abordagens probabilísticas, adaptativas, simples, iteradas e, mais recentemente, métodos baseados em *machine learning* (Von Holstein, 1972; Timmermann, 2008; Lin, Wu e Zhou, 2018; Rapach e Zhou, 2020; Wang et al., 2023).

O método de combinação por *stacked generalization* foi desenvolvido por Wolpert (1992) e pertence à família de técnicas de aprendizado conjunto (*ensemble learning*) no âmbito do *machine learning*. O autor apresentou o método como uma técnica para melhorar o desempenho preditivo por meio da combinação das previsões de vários modelos em um nível mais alto, permitindo a criação de um novo modelo que aprende a combinar as previsões dos modelos base e a aproveitar suas forças individuais. Para Wolpert (1992), a estrutura ideal do *stacked generalization* possui dois níveis: no primeiro, vários modelos base são treinados com os dados de treinamento; no segundo, um meta-modelo é treinado utilizando as previsões dos modelos base como entrada, com o objetivo de aprender a ponderá-las de modo a otimizar a performance geral. Em linha com essa formulação, Zhao e Cheng (2022) mostram que o *stacked generalization* pode superar o *benchmark* da média histórica tradicional em termos de desempenho dentro e fora da amostra, além de apresentar melhor desempenho preditivo do que diversos modelos alternativos. Os autores também destacam que a técnica é particularmente adequada em problemas de previsão marcados por ruído, não estacionariedade e comutação de regime, características frequentemente observadas nos mercados de capitais.

No caso brasileiro, a relevância dessa discussão é ampliada por dois fatores. O primeiro é a carência de estudos empíricos que utilizam métodos de combinação de previsões no mercado de ações brasileiro, com exceção das pesquisas realizadas por Cavaleri e Ribeiro (2011) e Rosa (2016), voltadas à combinação de previsões aplicadas à volatilidade. O segundo é o crescimento recente das publicações que empregam abordagens de *machine learning* na previsão do comportamento das ações no cenário nacional, como Santos (2020), Abreu (2021), Torres (2021), Bluvol (2022) e Silva (2023). É nesse ponto que o presente estudo se posiciona. Ao invés de propor um novo modelo teórico de apreçamento, o artigo investiga se a combinação de previsões via *stacked generalization* pode potencializar o desempenho preditivo de modelos de precificação de ativos já consolidados na literatura quando aplicados aos ativos da B3. Dessa forma, o estudo se insere na interface entre a literatura tradicional de precificação de ativos e a literatura contemporânea de combinação de previsões com *machine learning*, buscando avaliar se a combinação de diferentes estruturas preditivas produz ganhos estatisticamente relevantes fora da amostra no mercado

acionário brasileiro.

3 Dados, Modelos e Procedimentos Metodológicos

Esta pesquisa utiliza uma amostra composta por 202 empresas com informações disponíveis no período analisado, totalizando 240 observações mensais. A variável dependente é o retorno da carteira, calculado a partir dos retornos das ações das empresas que compõem cada um dos grupos. Já as variáveis independentes são os fatores de risco pertencentes aos modelos de precificação de ativos observados na literatura considerada neste estudo. Para formar as carteiras, os ativos foram classificados em ordem crescente com base na variável tamanho (ME), mensurada pelo valor de mercado do ativo. Em seguida, os ativos foram distribuídos em quintis, dando origem a cinco carteiras distintas, em que o menor quintil abrange os 20% de ativos com os valores de mercado mais baixos e o maior quintil inclui os 20% com os valores de mercado mais altos. Para fins de comparação empírica, o estudo concentra a análise no desempenho preditivo dos modelos e meta-modelos para as diferentes carteiras formadas com base em tamanho.

3.1 Dados e formação das carteiras

Os dados para calcular os modelos de precificação de ativos foram coletados por meio de fontes secundárias, com informações provenientes de fontes públicas. Os dados relacionados ao retorno de mercado e ao retorno das ações usados na formação das carteiras foram obtidos em consulta à B3. Os dados relacionados à taxa livre de risco, utilizada para calcular o prêmio de risco de mercado, e o Certificado de Depósito Interbancário (CDI), utilizado para calcular o *spread* de juros, foram coletados no site do Banco Central do Brasil (BACEN). Os dados relacionados às variáveis tamanho e *book-to-market* foram obtidos por meio da base de dados da plataforma ComDinheiro, e o fator de risco HML foi obtido na base de dados do Núcleo de Pesquisa em Economia Financeira da Universidade de São Paulo (NEFIN). Os dados coletados foram organizados e processados por meio do *software* de planilha eletrônica Microsoft Excel e as estimativas das regressões, os testes econométricos e a implementação do algoritmo de *machine learning* da técnica de *stacked generalization* foram realizados por meio da linguagem Python, com o desenvolvimento de um código de programação específico para essa finalidade, elaborado pelo próprio autor.

3.2 Modelos de precificação de ativos e técnica de *stacked generalization*

Diferentemente de Zhao e Cheng (2022), não foram utilizados algoritmos de *machine learning* como modelos base. Foram considerados cinco modelos de precificação de ativos citados por diversas pesquisas que envolvem a previsão de retorno das ações nos últimos anos: CAPM, L-CAPM, D-CAPM, C-CAPM e F&F (1996). A escolha desses modelos decorre de sua recorrência na literatura e de sua capacidade de representar diferentes formas de modelagem do risco e do retorno no mercado acionário.

O CAPM tradicional é utilizado como uma especificação de referência para a previsão dos retornos esperados das carteiras:

$$E(R_i) = R_f + \beta_i[E(R_m) - R_f] \quad (1)$$

Onde $E(R_i)$ = retorno esperado do ativo i ; R_f é o retorno do ativo livre de risco; β_i é a sensibilidade do ativo i em relação ao mercado; $E(R_m)$ é o retorno esperado do mercado e $E(R_m) - R_f$ é o prêmio do risco de mercado.

O L-CAPM mantém a estrutura do CAPM tradicional, diferenciando-se pela forma de obtenção da sensibilidade do ativo em relação ao mercado, uma vez que o beta é construído a partir do filtro de Kalman. Em razão dessa diferença recair sobre o procedimento de estimação, e não sobre a estrutura formal da equação, o modelo é mantido nesta seção em nível descritivo.

Para capturar assimetrias associadas ao risco de mercado em cenários de queda, o estudo também utiliza o D-CAPM:

$$R_i = R_f + (R_m - R_f) \beta^D \quad (2)$$

$$\beta^D = \frac{E(\text{Min}((r_i - \bar{r}_i), 0)) \times E(\text{Min}((r_M - \bar{r}_M), 0))}{E(\text{Min}((r_i - \bar{r}_i), 0)^2)}$$

Onde: R_i é o retorno esperado do ativo; R_f é o retorno do ativo livre de risco; R_m é o retorno do mercado; $R_m - R_f$ é prêmio do risco de mercado e β^D é o beta *downside* do ativo.

Além disso, emprega-se o Conditional CAPM, que amplia a estrutura tradicional ao incorporar variáveis condicionantes de natureza macroeconômica e financeira:

$$R_{t+1}^i = \alpha^i + \beta^{1i} R_{t+1}^M + \beta^{2i} R_{t+1}^M R_m - R_{f_t} + \beta^{3i} R_{t+1}^M HML_t + \beta^{4i} R_{t+1}^M Spread_t + \epsilon_{t+1}$$

(3)

Onde R_{t+1}^i = retorno esperado do ativo i ; β^i é o beta condicional do ativo i ; R_{t+1}^M é o retorno do mercado; $Rm - Rf_t$ é o prêmio de risco do mercado; HML_t é o *retorno da carteira high minus low*; $Spread_t$ é o *spread* de juros e ϵ_{t+1} = erro aleatório

Por fim, utiliza-se a regressão *cross-section* de Fama e French (1996), que permite relacionar os retornos esperados às características das firmas e à sensibilidade em relação ao mercado:

$$R_{i,t+\tau} = \lambda_{t+\tau}^0 + \lambda_{t+\tau}^M \beta_{i,t-1} + \lambda_{t+\tau}^{ME} \ln(ME_{i,t-1}) + \lambda_{t+\tau}^{BM} \ln(BM_{i,t-1}) + v_{i,t+\tau}, \forall i, \text{ para cada } (t + \tau) \quad (4)$$

Onde $R_{i,t+\tau}$ = retorno esperado do ativo; $t, t - 1$ são os meses de julho e junho, respectivamente; τ é o período 0,1, ..., 11; $t + \tau$ é o período de julho t a junho $t - 1$ do ano seguinte; $\lambda_{t+\tau}^0$ é a taxa livre de risco; $\beta_{i,t-1}$ é a sensibilidade do ativo em relação ao mercado; $ME_{i,t-1}$ é o valor de mercado; $BM_{i,t-1} = (BE_{i,t-1} / ME_{i,t-1})$ é o valor do patrimônio líquido ao valor de mercado e $\bar{\lambda}^M = (T^{-1} \sum_{t=1}^T \lambda_t^M) > 0$.

A partir das previsões geradas por esses modelos base, aplica-se a técnica de *stacked generalization* com o objetivo de combinar as informações preditivas em um meta-modelo. Em termos gerais, a técnica pode ser representada como uma função em que a previsão final do retorno resulta da combinação das previsões produzidas pelos modelos base selecionados no conjunto de validação:

$$\hat{R}_{t+1}^{ST} = g(\hat{R}_{t+1}^{CAPM}, \hat{R}_{t+1}^{L-CAPM}, \hat{R}_{t+1}^{D-CAPM}, \hat{R}_{t+1}^{C-CAPM}, \hat{R}_{t+1}^{F\&F}) \quad (5)$$

onde $g(\cdot)$ representa o meta-modelo estimado por meio dos algoritmos RF, GBRT, NN ou LR. No procedimento empírico adotado neste estudo, seguindo Zhao e Cheng (2022), apenas os modelos base com melhor desempenho preditivo no conjunto de validação são utilizados no treinamento do meta-modelo.

3.3 Estratégia de estimação e métricas de avaliação

Para atingir os objetivos da pesquisa, a amostra completa de janeiro de 2003 a dezembro de 2022 foi dividida em três conjuntos de dados: (i) treinamento, (ii) validação e (iii) teste. Optou-se pela abordagem de janela de expansão proposta por Timmermann (2008), na qual os dados

são segmentados em partes, seguindo um formato de estimativas recursivas. As janelas de estimativas recursivas utilizadas foram determinadas com base nos estudos de Rapach e Zhou (2020) e Zhao e Cheng (2022). A amostra inicial de treinamento abrange o período de janeiro de 2003 a dezembro de 2005, totalizando 36 meses (15%). O conjunto de validação abrange o intervalo de janeiro de 2006 a dezembro de 2012, totalizando 84 meses (35%), enquanto o período de teste se estende de janeiro de 2013 a dezembro de 2022, com duração de 120 meses (50%).

Em seguida, utilizou-se o conjunto de dados de treinamento para empregar a janela de expansão de Timmermann (2008), visando realizar previsões um passo à frente para os modelos base no conjunto de dados de validação. Desta forma, a primeira previsão foi gerada para janeiro de 2006, utilizando dados de janeiro de 2003 a dezembro de 2005. No mês seguinte, a janela de dados de treinamento é expandida para incluir também janeiro de 2006, e os parâmetros dos modelos de previsão são novamente estimados, sendo depois utilizados para prever os retornos das ações para fevereiro de 2006, e assim sucessivamente até o final da amostra de validação.

Este mesmo procedimento foi aplicado ao conjunto de dados de teste com relação ao conjunto de dados de validação, e os resultados das previsões dos modelos de precificação de ativos foram armazenados em uma matriz de previsões. Esse procedimento permite aproximar o exercício empírico das condições efetivamente enfrentadas pelos agentes de mercado, uma vez que as previsões são sempre geradas com base apenas nas informações disponíveis até cada ponto do tempo.

Na sequência, seguindo Zhao e Cheng (2022), avaliou-se a eficácia do desempenho das previsões no conjunto de dados de validação por meio do R_{OOS}^2 de Campbell e Thompson (2008). Nesta etapa, os três modelos de precificação de ativos com os melhores R_{OOS}^2 foram selecionados para treinar o meta-modelo de *stacked generalization*, enquanto as duas piores previsões foram descartadas. O R_{OOS}^2 de Campbell e Thompson (2008) é dado por:

$$R_{OOS,\mathbb{A}}^{2,(j)} = 1 - \frac{\sum_{\tau \in \mathbb{A}} (y_{\tau} - \hat{y}_{\tau|\tau-1}^{(j)})^2}{\sum_{\tau \in \mathbb{A}} (y_{\tau} - \bar{y}_{\tau|\tau-1})^2}, \quad (6)$$

Onde j indica cada um dos modelos em consideração, e \mathbb{A} é a janela de tempo para comparação. y_{τ} é o prêmio de risco das ações observado na data τ , e $\hat{y}_{\tau|\tau-1}^{(j)}$ é a previsão na data $\tau - 1$ do modelo j . $\bar{y}_{\tau|\tau-1} = \frac{1}{\tau-1} \sum_{i=1}^{\tau-1} y_i$ é a previsão feita pela média histórica (MH).

Neste sentido, $R_{OOS,\mathbb{A}}^{2,(j)}$ também pode ser escrito como: $R_{OOS,\mathbb{A}}^{2,(j)} = 1 - \frac{MFSE_{\mathbb{A}}^{(j)}}{MFSE_{\mathbb{A}}^{(MH)}}$, onde $MFSE_{\mathbb{A}}^{(j)} = \frac{1}{\#\mathbb{A}} \sum_{\tau \in \mathbb{A}} (y_{\tau} - \hat{y}_{\tau|\tau-1}^{(j)})^2$ e $MFSE_{\mathbb{A}}^{(MH)} = \frac{1}{\#\mathbb{A}} \sum_{\tau \in \mathbb{A}} (y_{\tau} - \bar{y}_{\tau|\tau-1})^2$ representam os erros de previsão quadráticos médios (*Mean Squared Forecast Error* - MSFE) no conjunto \mathbb{A} de um determinado modelo j e MH, respectivamente.

$\#\mathbb{A}$ é a cardinalidade do conjunto \mathbb{A} . Portanto, $R_{OOS,\mathbb{A}}^{2,(j)}$ é uma estatística conveniente para comparar MSFE de um modelo concorrente em relação ao MH. Quando $R_{OOS,\mathbb{A}}^{2,(j)}$ for maior que zero, o modelo concorrente supera a MH em termos de MSFE.

A significância estatística foi avaliada pela estatística ajustada ao MSFE proposta por Clark e West (2007). A hipótese nula é que a MH possui um MSFE menor ou igual em relação ao modelo de previsão concorrente, e a hipótese alternativa é que a MH tem um MSFE maior que o modelo concorrente. Se a hipótese nula for rejeitada, o modelo concorrente tem um desempenho melhor que a MH (Zhao e Cheng, 2022).

Além do R^2 fora da amostra (R_{OOS}^2), a eficácia preditiva também é avaliada por medidas de previsão de sinal, como acurácia, sensibilidade e especificidade. A utilização conjunta dessas métricas permite examinar não apenas a capacidade dos modelos em reduzir o erro de previsão em relação ao benchmark, mas também sua habilidade de antecipar corretamente a direção dos retornos.

A seção seguinte apresenta os resultados empíricos do estudo, contemplando tanto o desempenho preditivo no conjunto geral de teste quanto a análise em condições extremas de mercado.

4 Resultados Empíricos

Esta seção apresenta os resultados empíricos do estudo, com foco no desempenho preditivo dos modelos base e dos meta-modelos de *stacked generalization* no conjunto de teste, bem como em condições extremas de mercado. Inicialmente, são discutidos os resultados gerais fora da amostra para as carteiras analisadas. Em seguida, examina-se o comportamento dos modelos em episódios de maior instabilidade, com o objetivo de verificar se os ganhos preditivos observados se mantêm em contextos marcados por choques exógenos.

4.1 Desempenho preditivo no conjunto de teste

A análise dos resultados no conjunto de teste foi realizada com base no período de janeiro de 2013 a dezembro de 2022, considerando três carteiras de ativos: empresas de pequeno porte (Small-ME), empresas de médio porte (Median-ME) e grandes empresas (Large-ME). Esse recorte permite avaliar se os ganhos associados à técnica de *stacked generalization* são observados de forma consistente em diferentes segmentos do mercado acionário brasileiro.

A seleção dos modelos base utilizados no treinamento dos meta-modelos foi realizada com base no desempenho preditivo observado no conjunto de validação, conforme os resultados apresentados anteriormente. De forma geral, esse procedimento permitiu restringir o treinamento do *stacked generalization* aos modelos com melhor desempenho relativo, preservando a lógica metodológica adotada no estudo.

Os resultados centrais do artigo para o conjunto de teste são apresentados na Tabela 1. De maneira geral, as evidências sugerem que nenhum modelo de precificação de ativos, considerado isoladamente, é significativamente superior à média histórica (MH) em termos de R^2_{OOS} nas três carteiras simultaneamente, embora possam ser observados alguns destaques específicos, como o D-CAPM na carteira de pequeno e médio porte, o C-CAPM na carteira das médias empresas e o modelo F&F (1996) na carteira das médias e grandes empresas. Esses achados indicam que o desempenho dos modelos base é heterogêneo e dependente do segmento de mercado analisado.

Tabela 1 – Resultados de previsão no conjunto de teste

janeiro de 2013 a dezembro de 2022					
Painel A: Carteira <i>Small-ME</i>	R^2_{OOS}	MSFE	Acu. (%)	Sen. (%)	Esp. (%)
Modelos base					
CAPM	0,45	0,0052	74,17	80,00	68,33
D-CAPM	0,04**	0,0091	58,33	98,33	18,33
L-CAPM	-0,04***	0,0099	50,00	100,00	0,00
C-CAPM	-58,31***	0,5604	55,00	81,67	28,33
F&F (1996)	-11,03	0,1136	41,67	35,00	48,33
MediaM	-1,74***	0,0259	50,83	78,33	23,33
Stacked com todos os modelos base					
St-LR	0,34***	0,0062	69,17	93,33	45,00
St-RF	0,24***	0,0072	62,50	95,00	30,00
St-GBRT	-0,02***	0,0096	62,50	85,00	40,00
St-NN	0,01***	0,0094	66,67	88,33	45,00
Stacked com os modelos CAPM, D-CAPM e L-CAPM					
St-LR	0,36***	0,006	69,17	96,67	41,67
St-RF	0,20***	0,0076	65,00	96,67	33,33
St-GBRT	0,01***	0,0093	65,83	91,67	40,00
St-NN	-0,20***	0,0114	50,00	100,00	0,00
Painel B: Carteira <i>Small-Median-ME</i>					
Modelos base	R^2_{OOS}	MSFE	Acu. (%)	Sen. (%)	Esp. (%)
CAPM	-201,24	1,5489	19,17	18,18	20,00
D-CAPM	0,01***	0,0076	48,33	100,00	4,62

L-CAPM	-0,03***	0,0079	45,83	100,00	0,00
C-CAPM	-819,68***	6,2856	29,17	40,00	20,00
F&F (1996)	-25,12***	0,2	45,83	29,09	60,00
Stacked com todos os modelos base					
MediaM	-68,08***	0,5291	21,67	25,45	18,46
St-LR	-8886,77	68,0716	37,50	29,09	44,62
St-RF	-15996,49*	122,525	61,67	98,18	30,77
St-GBRT	-39339,76*	301,3114	62,50	98,18	32,31
St-NN	-1028,48**	7,8848	39,17	25,45	50,77
Stacked com os modelos CAPM, C-CAPM e F&F (1996)					
St-LR	-7166,32	54,8946	23,33	20,00	26,15
St-RF	-13267,86*	101,6264	64,17	100,00	33,85
St-GBRT	-19106,36*	146,3435	63,33	100,00	32,31
St-NN	-595,71**	4,5702	43,33	38,18	47,69
Painel C: Carteira Median-ME					
	R^2_{00s}	MSFE	Acu. (%)	Sem. (%)	Esp. (%)
Modelos base					
CAPM	0,71	0,0016	80,83	83,58	77,36
D-CAPM	0,28***	0,0039	80,00	94,03	62,26
L-CAPM	-0,03***	0,0056	55,83	100,00	0,00
C-CAPM	0,69**	0,0017	80,83	91,04	67,92
F&F (1996)	0,07***	0,005	65,83	76,12	52,83
MediaM	0,51***	0,0027	85,00	95,52	71,70
Stacked com todos os modelos base					
St-LR	0,71**	0,0016	84,17	92,54	73,58
St-RF	0,60**	0,0021	78,33	88,06	66,04
St-GBRT	0,52***	0,0026	70,00	86,57	49,06
St-NN	0,58***	0,0023	79,17	98,51	54,72
Stacked com os modelos CAPM, C-CAPM e F&F (1996)					
St-LR	0,71**	0,0016	84,17	91,04	75,47
St-RF	0,59***	0,0022	78,33	95,52	56,60
St-GBRT	0,52***	0,0026	72,50	89,55	50,94
St-NN	0,50***	0,0027	75,00	62,69	90,57
janeiro de 2013 a dezembro de 2022					
Painel D: Carteira Median-Large-ME					
	R^2_{00s}	MSFE	Acu. (%)	Sen. (%)	Esp. (%)
Modelos base					
CAPM	0,72	0,0013	82,50	86,15	78,18
D-CAPM	0,27	0,0034	77,50	96,92	54,55
L-CAPM	-0,05	0,0048	54,17	100,00	0,00
C-CAPM	-0,88***	0,0087	63,33	98,46	21,82
F&F (1996)	-0,63***	0,0075	39,17	52,31	23,64
MediaM	0,43***	0,0026	67,50	98,46	30,91
Stacked com todos os modelos base					
St-LR	-8,03*	0,0416	43,33	41,54	45,45
St-RF	0,60	0,0018	77,50	95,38	56,36
St-GBRT	0,55	0,0021	77,50	92,31	60,00
St-NN	0,45***	0,0025	72,50	89,23	52,73
Stacked com os modelos CAPM, C-CAPM e F&F (1996)					
St-LR	-4,86	0,027	55,00	70,77	36,36
St-RF	0,59	0,0019	80,00	93,85	63,64
St-GBRT	0,58	0,0019	80,00	95,38	61,82
St-NN	0,34***	0,003	75,83	96,92	50,91
Painel E: Large-ME					
	R^2_{00s}	MSFE	Acu. (%)	Sen. (%)	Esp. (%)
Modelos base					
CAPM	0,78	0,0007	84,17	83,10	85,71
D-CAPM	0,33	0,0021	80,83	95,77	59,18
L-CAPM	-0,04	0,0034	59,17	100,00	0,00

C-CAPM	0,73	0,0009	82,50	83,10	81,63
F&F (1996)	0,04**	0,0031	60,83	73,24	42,86
MediaM	0,67	0,0011	85,83	90,14	79,59
Stacked com todos os modelos base					
St-LR	-4,85***	0,0189	60,83	42,25	87,76
St-RF	-1,97	0,0096	75,00	69,01	83,67
St-GBRT	-2,44	0,0111	78,33	71,83	87,76
St-NN	0,81*	0,0006	86,67	91,55	79,59
Stacked com os modelos CAPM, C-CAPM e F&F (1996)					
St-LR	0,19***	0,0026	80,00	67,61	97,96
St-RF	-1,16**	0,007	74,17	64,79	87,76
St-GBRT	-0,88***	0,006	77,50	73,24	83,67
St-NN	0,64***	0,0012	75,83	59,15	100,00

Notas: CAPM: Capital Asset Pricing Model (modelo de precificação de ativos de capital); D-CAPM: é o modelo Downside CAPM; L-CAPM: é o modelo Learning CAPM; C-CAPM: é o modelo Conditional CAPM; F&F (1996): é o modelo de precificação de ativos de Fama e French (1996). Abreviaturas do nome do modelo: St significa stacked. St-LR: é o stacked com o modelo de Regressão Linear; St-RF: é o stacked com o modelo de Random Forest; St-GBRT: é o stacked com o modelo de Gradient Boosted Regression Trees; St-NN: é o stacked com o modelo de Neural Networks; MediaM: é a combinação de média simples dos modelos financeiros. MSFE é o erro quadrático médio de previsão; Acu. (%) é a medida de acurácia; Sen. (%) é a medida de sensibilidade; e Esp. (%) é a medida de especificidade. *, **, e *** indicam significância estatística nos níveis de 10%, 5% e 1%, respectivamente.

Fonte: Elaborada pelos autores (2026)

Esse resultado é importante porque reforça a ideia de que a previsão dos retornos das ações dificilmente pode ser descrita de forma eficiente por um único modelo. Em vez disso, a evidência empírica sugere que diferentes modelos capturam aspectos distintos da dinâmica dos retornos, o que torna economicamente plausível a utilização de uma técnica de combinação de previsões.

Na carteira Small-ME, a Tabela 1 mostra que o modelo D-CAPM foi o único modelo base com R_{00S}^2 significativo e positivo, ainda que com valor reduzido. Em termos de previsão de sinal, o modelo apresentou bom desempenho na previsão das instâncias positivas, mas desempenho limitado na previsão das instâncias negativas, o que restringiu sua acurácia geral. Em contraste, os meta-modelos de *stacked generalization* apresentaram resultados mais consistentes, com destaque para o St-LR e o St-RF. Na especificação com seleção dos melhores modelos base, o St-LR apresentou R_{00S}^2 de 0,36, MSFE de 0,0060, acurácia de 69,17%, sensibilidade de 96,67% e especificidade de 41,67%, enquanto o St-RF apresentou R_{00S}^2 de 0,20 e acurácia de 65,00%.

Esses resultados sugerem que, na carteira de empresas de pequeno porte, a combinação de previsões via *stacked generalization* produz desempenho superior ao da maior parte dos modelos base, especialmente quando se considera o meta-modelo linear. Além disso, o melhor desempenho do St-LR em relação a meta-modelos mais complexos indica que maior sofisticação algorítmica não implica, necessariamente, melhor desempenho preditivo fora da amostra.

Nas demais carteiras, os resultados seguem padrão menos uniforme. Ainda que determinados modelos base e meta-modelos apresentem desempenhos relativamente melhores em contextos específicos, a evidência geral da Tabela 1 sugere que os ganhos associados ao *stacked generalization* não se distribuem de forma homogênea entre todos os segmentos do mercado. Em outras palavras, a combinação de previsões parece explorar complementaridades entre modelos de forma mais eficiente em alguns contextos do que em outros.

Em conjunto, os resultados do conjunto de teste indicam que a técnica de *stacked generalization* pode melhorar significativamente o desempenho das previsões em diferentes contextos e condições de mercado, com destaque para o meta-modelo St-LR. Ao mesmo tempo, os achados mostram que esses ganhos não são uniformes entre carteiras, dependendo da interação entre os modelos base selecionados, do tipo de meta-modelo utilizado e das características específicas do ambiente de mercado.

Em síntese, os resultados do conjunto de teste indicam que o melhor desempenho relativo se concentrou no meta-modelo St-LR, especialmente na carteira Small-ME.

4.2 Desempenho preditivo em condições extremas de mercado

Para complementar a análise do conjunto geral de teste, esta subseção examina o comportamento dos modelos e meta-modelos em condições extremas de mercado. O objetivo é verificar se os ganhos preditivos observados em contextos usuais permanecem quando o ambiente econômico é marcado por choques exógenos e maior instabilidade.

Para identificar períodos de movimentos extremos de mercado, foi calculado o *drawdown* (DD) do retorno mensal do Ibovespa, seguindo Geronazzo (2019). Ao considerar um horizonte de tempo fixo $T \in [0, \infty]$, o processo de *drawdown* $D^X = \{D_t^{(X)}\}_{t \in [0, T]}$ correspondente a um processo estocástico $X \in R^\infty$ é definido pelas seguintes equações:

$$D_t^{(X)} = M_t^{(X)} - X_t \quad (7)$$

Onde:

$$M_t^{(X)} = \sup_{\mu \in [0, t]} X_\mu \quad (8)$$

é o máximo local de X até o tempo t .

Os períodos de perdas mais significativas identificados pelo DD coincidem com episódios de forte instabilidade no mercado brasileiro, com destaque para a Crise Financeira Global de 2008 e 2009, a crise político-econômica de 2015 e 2016 e a pandemia de Covid-19 em 2020.

No contexto deste artigo, a análise empírica concentra-se nos dois períodos contidos no conjunto de teste: a crise político-econômica de 2015–2016 e a pandemia de Covid-19.

As Tabelas 2 e 3 apresentam os resultados de previsão de três carteiras de ativos — empresas de pequeno porte (Small-ME), empresas de médio porte (Median-ME) e grandes empresas (Large-ME) — para esses dois episódios de crise. Para treinamento e previsão dos meta-modelos de *stacked generalization*, foram considerados os modelos de precificação com os três melhores R_{00s}^2 no conjunto de validação, em linha com a estratégia metodológica adotada no artigo.

Tabela 2 – Resultado de previsão durante a crise político-econômica de 2015–2016

janeiro de 2015 a dezembro de 2016					
Painel A: Carteira <i>Small-ME</i>	R_{00s}^2	MSFE	Acu. (%)	Sen. (%)	Esp. (%)
Modelos base					
CAPM	0,21**	0,0078	70,83	72,73	69,23
D-CAPM	-0,04***	0,0102	45,83	100,00	0,00
L-CAPM	-0,08***	0,0106	45,83	100,00	0,00
C-CAPM	-53,18***	0,5324	54,17	90,91	23,08
F&F (1996)	-20,45***	0,2108	29,17	27,27	30,77
MediaM	-9,56***	0,1038	54,17	90,91	23,08
Stacked com os modelos CAPM, D-CAPM e L-CAPM					
St-LR	0,31***	0,0068	79,17	90,91	69,23
St-RF	0,21***	0,0078	54,17	72,73	38,46
St-GBRT	0,06***	0,0093	54,17	63,64	46,15
St-NN	0,02***	0,0096	75,00	90,91	61,54
Painel B: Carteira <i>Median-ME</i>					
	R_{00s}^2	MSFE	Acu. (%)	Sen. (%)	Esp. (%)
Modelos base					
CAPM	0,71*	0,0018	87,50	84,62	90,91
D-CAPM	0,22***	0,0049	87,50	100,00	72,73
L-CAPM	-0,12***	0,0069	54,17	100,00	0,00
C-CAPM	0,67***	0,002	87,50	92,31	81,82
F&F (1996)	0,00***	0,0062	75,00	76,92	72,73
MediaM	0,62***	0,0024	91,67	92,31	90,91
Stacked com os modelos CAPM, C-CAPM e F&F (1996)					
St-LR	0,68***	0,0020	91,67	92,31	90,91
St-RF	0,65**	0,0022	95,83	92,31	100,00
St-GBRT	0,64**	0,0023	91,67	92,31	90,91
St-NN	0,42***	0,0036	83,33	100,00	63,64
Painel C: <i>Large-ME</i>					
	R_{00s}^2	MSFE	Acu. (%)	Sen. (%)	Esp. (%)
Modelos base					
CAPM	0,73	0,001	83,33	83,33	83,33
D-CAPM	0,24	0,0027	75,00	100,00	50,00
L-CAPM	-0,17	0,0042	50,00	100,00	0,00
C-CAPM	0,68	0,0011	79,17	75,00	83,33
F&F (1996)	-0,04	0,0038	70,83	75,00	66,67
MediaM	0,76	0,0008	83,33	83,33	83,33

<i>Stacked com os modelos CAPM, C-CAPM e F&F (1996)</i>					
St-LR	0,00***	0,0036	75,00	50,00	100,00
St-RF	-1,22	0,008	75,00	50,00	100,00
St-GBRT	-2,55*	0,0128	83,33	75,00	91,67
St-NN	0,81	0,0007	83,33	83,33	83,33

Notas: CAPM: Capital Asset Pricing Model (modelo de precificação de ativos de capital); D-CAPM: é o modelo Downside CAPM; L-CAPM: é o modelo Learning CAPM; C-CAPM: é o modelo Conditional CAPM; F&F (1996): é o modelo de precificação de ativos de Fama e French (1996). Abreviaturas do nome do modelo: St significa stacked. St-LR: é o stacked com o modelo de Regressão Linear; St-RF: é o stacked com o modelo de Random Forest; St-GBRT: é o stacked com o modelo de Gradient Boosted Regression Trees; St-NN: é o stacked com o modelo de Neural Networks; MediaM: é a combinação de média simples dos modelos financeiros. MSFE é o erro quadrático médio de previsão; Acu. (%) é a medida de acurácia; Sen. (%) é a medida de sensibilidade; e Esp. (%) é a medida de especificidade. *, **, e *** indicam significância estatística nos níveis de 10%, 5% e 1%, respectivamente.

Fonte: Elaborada pelos autores (2026)

Na crise político-econômica de 2015–2016, os resultados da Tabela 2 mostram que, na carteira Small-ME, o modelo CAPM apresentou R^2_{OOS} positivo e significativo de 0,21, enquanto o meta-modelo St-LR alcançou R^2_{OOS} ainda superior, de 0,31, com MSFE de 0,0068, acurácia de 79,17%, sensibilidade de 90,91% e especificidade de 69,23%. O St-RF apresentou R^2_{OOS} de 0,21, e os meta-modelos St-GBRT e St-NN também registraram resultados positivos, embora menores. Esses achados indicam que, ao menos para empresas de pequeno porte no episódio de 2015–2016, a técnica de *stacked generalization* preservou parte de sua capacidade de gerar previsões superiores às dos modelos base isolados.

Essa evidência é relevante porque mostra que a combinação de previsões não perde completamente sua utilidade em ambientes adversos. Ainda assim, os resultados não autorizam concluir que a técnica apresenta superioridade generalizada em contextos de crise. O que se observa, mais precisamente, é a existência de ganhos pontuais e condicionais, mais perceptíveis em segmentos específicos da amostra.

No caso da pandemia de Covid-19, os resultados da Tabela 3 devem ser interpretados como um teste mais severo para a abordagem. Embora alguns modelos e meta-modelos possam apresentar acertos pontuais ou desempenho relativamente melhor em determinadas carteiras, a evidência consolidada do estudo indica que o choque exógeno associado à pandemia impôs restrições severas à capacidade preditiva de todos os modelos analisados. Nesse sentido, a pandemia funciona como o principal teste de limite da abordagem, ao mostrar que os ganhos da técnica não devem ser interpretados como universais nem invariantes ao regime de mercado.

As evidências são claras ao demonstrarem que nenhum modelo ou meta-modelo foi capaz de superar a média histórica dos retornos das ações de forma significativa durante um choque exógeno. Esse ponto é central para a contribuição do artigo, pois qualifica a interpretação dos resultados: a técnica de *stacked generalization* melhora o desempenho preditivo em diversos

contextos, mas encontra limites relevantes quando o mercado é atingido por choques exógenos de grande magnitude.

Tabela 3 – Resultado de Previsão durante a Pandemia de Covid-19 em 2020

janeiro de 2020 a junho de 2021					
Painel A: Carteira <i>Small</i>-ME	R^2_{00s}	MSFE	Acu. (%)	Sen. (%)	Esp. (%)
Modelos base					
CAPM	0,47	0,0071	88,89	83,33	100,00
D-CAPM	-0,23	0,0167	88,89	91,67	83,33
L-CAPM	-0,43	0,0193	66,67	100,00	0,00
C-CAPM	-67,56	0,9282	61,11	75,00	33,33
F&F (1996)	-1,61***	0,0354	33,33	16,67	66,67
MediaM	-6,10	0,0961	61,11	66,67	50,00
Stacked com os modelos CAPM, D-CAPM e L-CAPM					
St-LR	0,51	0,0066	88,89	91,67	83,33
St-RF	0,07***	0,0126	77,78	91,67	50,00
St-GBRT	0,19**	0,011	77,78	91,67	50,00
St-NN	-0,09**	0,0148	72,22	100,00	16,67
Painel B: Carteira <i>Median</i>-ME					
R^2_{00s}	MSFE	Acu. (%)	Sen. (%)	Esp. (%)	
Modelos base					
CAPM	0,80	0,0021	88,89	83,33	100,00
D-CAPM	0,08	0,0096	88,89	83,33	100,00
L-CAPM	-0,39	0,0145	66,67	100,00	0,00
C-CAPM	0,85	0,0015	83,33	83,33	83,33
F&F (1996)	-0,36	0,0142	55,56	66,67	33,33
MediaM	0,66	0,0035	77,78	75,00	83,33
Stacked com os modelos CAPM, C-CAPM e F&F (1996)					
St-LR	0,79	0,0022	77,78	75,00	83,33
St-RF	0,57	0,0044	83,33	83,33	83,33
St-GBRT	0,61	0,004	83,33	83,33	83,33
St-NN	0,50***	0,0053	72,22	100,00	16,67
Painel C: <i>Large</i>-ME					
R^2_{00s}	MSFE	Acu. (%)	Sen. (%)	Esp. (%)	
Modelos base					
CAPM	0,85	0,0008	94,44	90,91	100,00
D-CAPM	0,15	0,0047	94,44	90,91	100,00
L-CAPM	-0,38	0,0076	61,11	100,00	0,00
C-CAPM	0,70	0,0017	88,89	90,91	85,71
F&F (1996)	-0,30	0,0072	50,00	63,64	28,57
MediaM	0,80	0,0011	88,89	90,91	85,71
Stacked com os modelos CAPM, C-CAPM e F&F (1996)					
St-LR	0,55	0,0025	83,33	81,82	85,71
St-RF	0,62	0,0021	94,44	100,00	85,71
St-GBRT	0,86	0,0008	94,44	100,00	85,71
St-NN	0,72	0,0016	88,89	81,82	100,00

Notas: CAPM: Capital Asset Pricing Model (modelo de precificação de ativos de capital); D-CAPM: é o modelo Downside CAPM; L-CAPM: é o modelo Learning CAPM; C-CAPM: é o modelo Conditional CAPM; F&F (1996): é o modelo de precificação de ativos de Fama e French (1996). Abreviaturas do nome do modelo: St significa stacked. St-LR: é o stacked com o modelo de Regressão Linear; St-RF: é o stacked com o modelo de Random Forest; St-GBRT: é o stacked com o modelo de Gradient Boosted Regression Trees; St-NN: é o stacked com o modelo de Neural Networks; MediaM: é a combinação de média simples dos modelos financeiros. MSFE é o erro quadrático médio de previsão; Acu. (%) é a medida de acurácia; Sen. (%) é a medida de sensibilidade; e Esp. (%) é a medida de especificidade. *, **, e *** indicam significância estatística nos níveis de 10%, 5% e 1%, respectivamente.

Fonte: Elaborada pelos autores (2026)

4.3 Síntese dos resultados

Em conjunto, os achados desta seção mostram que a principal contribuição empírica do estudo não está em demonstrar superioridade irrestrita da técnica de *stacked generalization*, mas em evidenciar que a combinação de previsões pode ampliar o desempenho preditivo fora da amostra em diversos contextos do mercado acionário brasileiro. Os resultados também mostram, contudo, que esses ganhos são condicionais e não absolutos, tornando-se menos expressivos ou mesmo insuficientes em contextos marcados por choques exógenos severos, como a pandemia de Covid-19.

Assim, a evidência empírica do artigo sustenta uma conclusão equilibrada: a técnica de *stacked generalization* representa uma alternativa promissora para potencializar previsões de retornos na B3, mas sua efetividade depende do ambiente de mercado e da natureza do choque enfrentado. A próxima seção apresenta as conclusões do estudo.

5 Conclusões

Este estudo investigou se as previsões dos retornos das ações dos modelos de precificação de ativos são potencializadas quando a técnica de *stacked generalization* em *machine learning* é aplicada aos ativos da B3. Para isso, foram utilizados cinco modelos de precificação de ativos recorrentes na literatura — CAPM, *Learning CAPM*, *Downside CAPM*, *Conditional CAPM* e a regressão *cross-section* de Fama e French (1996) — combinados por meio de quatro algoritmos de *machine learning* empregados como meta-modelos: *Random Forest*, *Gradient Boosted Regression Trees*, *Neural Networks* e *Linear Regression*. A avaliação empírica foi realizada em esquema de janelas de expansão, com foco no desempenho fora da amostra, por meio do R_{OOS}^2 de Campbell e Thompson (2008) e de medidas de previsão de sinal, como acurácia, sensibilidade e especificidade.

Os resultados evidenciam que a técnica de *stacked generalization* pode melhorar significativamente o desempenho das previsões em diferentes contextos e condições de mercado. No conjunto geral de teste, os meta-modelos apresentaram, em diversos casos, desempenho superior ao dos modelos base avaliados isoladamente, com destaque para o St-LR, sobretudo na carteira de empresas de pequeno porte. Esses achados reforçam a ideia de que a previsão dos retornos das ações dificilmente pode ser descrita de forma eficiente por um único modelo, e sugerem que a combinação de previsões pode explorar complementaridades informacionais relevantes entre diferentes especificações de apreçamento de ativos.

Ao mesmo tempo, os resultados mostram que esses ganhos não são uniformes. Sua magnitude varia entre carteiras e depende tanto do conjunto de modelos base selecionados quanto do tipo de meta-modelo empregado. Além disso, embora a seleção dos melhores modelos base não produza diferenças substantivas generalizadas em termos de R^2_{OOS} , as evidências sugerem que esse procedimento pode contribuir para melhorar aspectos específicos da previsão, especialmente a acurácia negativa em períodos de queda.

A análise em condições extremas de mercado permitiu qualificar de forma importante a interpretação dos resultados. Durante a crise político-econômica de 2015–2016, a técnica ainda apresentou ganhos preditivos pontuais em determinados segmentos da amostra, especialmente na carteira de empresas de pequeno porte. Todavia, no contexto da pandemia de Covid-19, a evidência consolidada do estudo indica que nenhum modelo ou meta-modelo foi capaz de superar a média histórica dos retornos das ações de forma significativa durante o choque exógeno. Esse resultado sugere que os benefícios associados ao *stacked generalization* são condicionais, e não absolutos, encontrando limites relevantes em ambientes marcados por rupturas severas e mudanças abruptas de regime.

Dessa forma, a principal contribuição do artigo para a literatura de precificação de ativos está em mostrar, no contexto da B3, que a combinação de previsões via *stacked generalization* pode ampliar o desempenho preditivo fora da amostra de modelos tradicionais e recentes de apreçamento de ativos, sem, contudo, eliminar as limitações impostas por episódios de elevada instabilidade e choques exógenos. Assim, o estudo se posiciona na interface entre a literatura tradicional de precificação de ativos e a literatura contemporânea de combinação de previsões com *machine learning*, oferecendo evidências empíricas sobre o potencial e os limites dessa integração no mercado acionário brasileiro.

Como implicação prática, os resultados sugerem que investidores, gestores e pesquisadores podem se beneficiar do uso de técnicas de combinação de previsões na modelagem de retornos de ações, especialmente em contextos de mercado sem rupturas extremas. Como agenda de pesquisa futura, recomenda-se ampliar a investigação para outros conjuntos de fatores de risco, diferentes algoritmos de *machine learning*, novas estruturas de combinação de previsões e distintos recortes do mercado brasileiro, de modo a aprofundar a compreensão sobre as condições sob as quais esses ganhos preditivos podem ser sustentados.

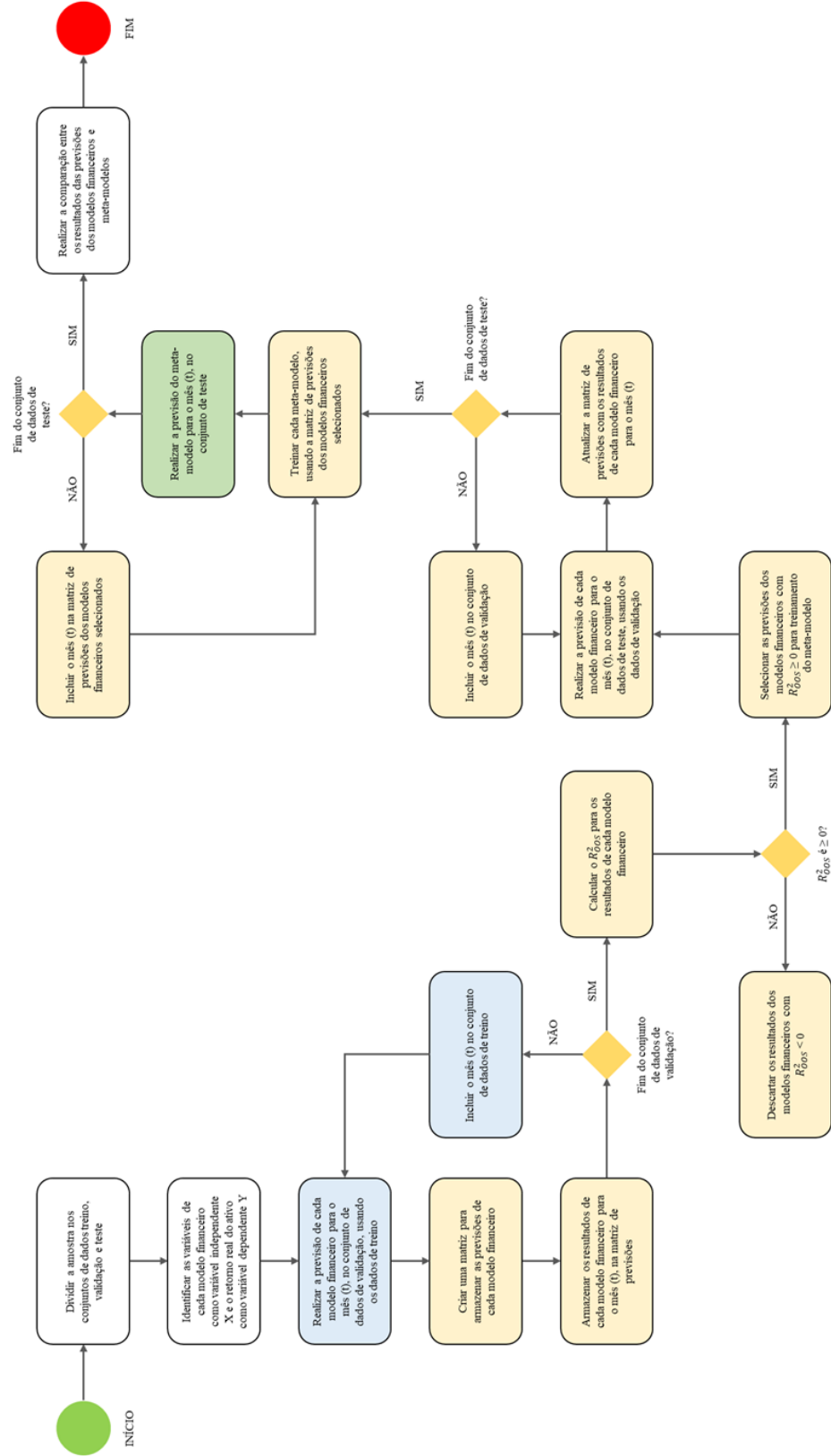
Referências

- Abreu, A. F. F. D. (2021). Aplicação de machine learning na pré-seleção de ativos para portfólios de investimento (*Doctoral dissertation*).
- Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20(4), 451-468.
- Black, F. (1972). Capital market equilibrium with restricted borrowing. *The Journal of Business*, 45(3), 444-455.
- Bluvol, L. (2022). Análise de algoritmos de *machine learning* para previsão de preços de IBOVESPA. (*Doctoral Dissertation*).
- Campbell, J. Y., & Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average?. *The Review of Financial Studies*, 21(4), 1509-1531.
- Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1), 57-82.
- Cavaleri, R., & Ribeiro, E. P. (2011). Combinação de previsões de volatilidade: um estudo. *Revista Economia*, 12(2), 239-261.
- Clark, T. E., & West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1), 291-311.
- Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), 3-56.
- Fama, E. F., & French, K. R. (1996). Multifactor explanations of asset pricing anomalies. *The Journal of Finance*, 51(1), 55-84.
- Fama, E. F., & French, K. R. (2004). The capital asset pricing model: Theory and evidence. *Journal of Economic Perspectives*, 18(3), 25-46.
- Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1-22.
- Friend, I., & Blume, M. (1970). Measurement of portfolio performance under uncertainty. *The American Economic Review*, 60(4), 561-575.
- Geronazzo, Arthur. Técnicas de estresse teste de mercado usando maximum drawdown / Arthur Geronazzo. - 2019. 137 f.
- Granger, C. W., & Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, 3(2), 197-204.
- Jensen, M. C., Black, F., & Scholes, M. S. (1972). The capital asset pricing model: Some empirical tests. *STUDIES IN THE THEORY OF CAPITAL MARKETS*, Praeger Publishers Inc.
- Lin, H., Wu, C., & Zhou, G. (2018). Forecasting corporate bond returns with a large set of predictors: An iterated combination approach. *Management Science*, 64(9), 4218-4238.
- Lintner, J. (1975). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. In *Stochastic Optimization Models in Finance* (pp. 131-155). Academic Press.
- Markowitz, H. M. (1952). Portfolio selection. *Journal of finance*, 7(1), 71-91.
- Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica: Journal of The Econometric Society*, 768-783.

- Rapach, D. E., Strauss, J. K., & Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies*, 23(2), 821-862.
- Rapach, D. E., & Zhou, G. (2020). Time-series and cross-sectional stock return forecasting: New machine learning methods. *Machine learning for asset management: New developments and financial applications*, 1-33.
- Rosa, F. G. D. D. (2016). Combinação de previsão de matrizes de covariância avaliada na aplicação em gestão de portfólios.
- Ross, S. A. (2013). The arbitrage theory of capital asset pricing. In *Handbook of The Fundamentals of Financial Decision Making: Part I* (pp. 11-30).
- Santos, Gustavo Carvalho *et al* (2020). Algoritmos de machine learning para previsão de ações da B3. 2020. *Dissertação de Mestrado*.
- Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3), 425-442.
- Silva, M. O. D. (2023). Precificação empírica das ações brasileiras com o uso de machine learning e gradient boosted trees: aplicação no período 2007-2022.
- Timmermann, A. (2008). Elusive return predictability. *International Journal of Forecasting*, 24(1), 1-18.
- Torres, S. (2021). Previsão do preço de ações brasileiras utilizando redes neurais artificiais. 2021. 51 f. *Dissertação (Mestrado em Administração de Empresas) - Universidade Presbiteriana Mackenzie, São Paulo*.
- Von Holstein, C. A. S. S. (1972). Probabilistic forecasting: An experiment related to the stock market. *Organizational Behavior and Human Performance*, 8(1), 139-158.
- Wang, X. *et al*. (2022). Forecast combinations: an over 50-year review. *International Journal of Forecasting*.
- Wei, K. C. John (1988). An asset-pricing theory unifying the capm and apt. *The Journal of Finance*, v. 43, n. 4, p. 881-892.
- Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, 5(2), 241-259.
- Zhao, A. B., & Cheng, T. (2022). Stock return prediction: Stacking a variety of models. *Journal of Empirical Finance*, 67, 288-317.

APÊNDICE A. Fluxograma detalhado de *Stacked Generalization* aplicado aos modelos de precificação de ativos

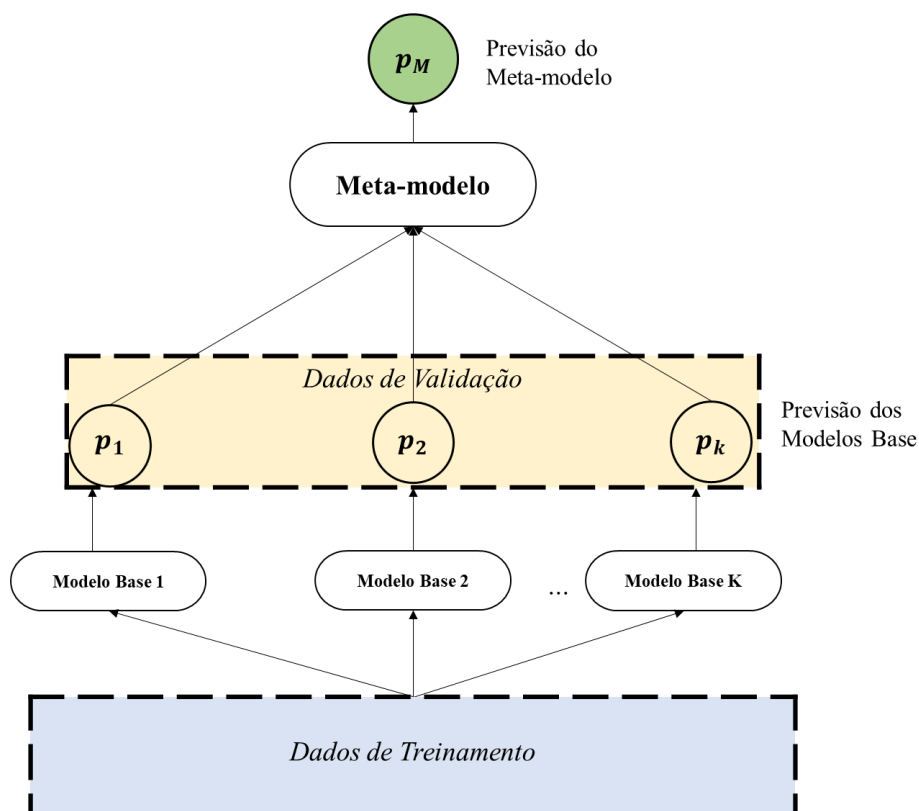
Figura A1 - Fluxograma detalhado de *Stacked Generalization* aplicado aos modelos de precificação de ativos
 Este apêndice apresenta, de forma esquemática, o fluxo completo de aplicação da técnica de *stacked generalization* no contexto do artigo. O objetivo é tornar mais transparente a sequência de procedimentos adotados, desde a divisão da amostra em conjuntos de treinamento, validação e teste, passando pela geração e seleção das previsões dos modelos base, até o treinamento dos meta-modelos e a avaliação final de seu desempenho preditivo. Dessa forma, o fluxograma complementa a descrição metodológica apresentada no corpo do artigo e facilita a compreensão da estratégia empírica utilizada no estudo.



APÊNDICE B. Estrutura sintética de *Stacked Generalization*

Este apêndice apresenta, de forma sintética, a estrutura da estratégia empírica adotada no artigo. Inicialmente, são estimados os modelos base de precificação de ativos, a saber: CAPM, Learning CAPM, Downside CAPM, Conditional CAPM e a regressão *cross-section* de Fama e French (1996). Em seguida, suas previsões são geradas de forma recursiva, com base no esquema de janelas de expansão utilizado ao longo do estudo. As previsões produzidas pelos modelos base no conjunto de validação são então avaliadas por meio do R_{00s}^2 de Campbell e Thompson (2008), permitindo a seleção dos três modelos com melhor desempenho preditivo. Na etapa seguinte, as previsões dos modelos selecionados são utilizadas como insumo para o treinamento dos meta-modelos de *stacked generalization*. Por fim, o desempenho dos meta-modelos é avaliado no conjunto de teste por meio de métricas estatísticas e medidas de previsão de sinal. A Figura B1 resume visualmente esse fluxo de estimação, seleção, combinação e avaliação preditiva.

Figura B1 - Estrutura sintética de *Stacked Generalization*



Fonte: Adaptado de Zhao e Cheng (2022)

APÊNDICE C. Dados, carteiras e janelas de estimação

A base empírica do estudo é composta por 202 empresas com informações disponíveis no período de janeiro de 2003 a dezembro de 2022, totalizando 240 observações mensais. Para a formação das carteiras, os ativos foram classificados em ordem crescente com base no tamanho da empresa, mensurado pelo valor de mercado do ativo, e posteriormente distribuídos em quintis. Esse procedimento deu origem a cinco carteiras distintas: *Small-ME*, *Small-Median-ME*, *Median-ME*, *Median-Large-ME* e *Large-ME*. A opção pela formação de carteiras por tamanho busca garantir comparabilidade entre os ativos ao longo do tempo e captar diferenças relevantes no comportamento dos retornos entre empresas de portes distintos.

A Tabela C1 apresenta as estatísticas descritivas das carteiras baseadas no valor de mercado. De forma geral, os resultados indicam diferenças relevantes entre os grupos formados, tanto em termos de retorno médio quanto de dispersão dos retornos e magnitude do valor de mercado médio. As evidências também sugerem que as carteiras de empresas maiores não apresentam, necessariamente, melhor relação entre retorno e risco, o que reforça a pertinência de analisar separadamente o desempenho preditivo dos modelos e meta-modelos em diferentes segmentos do mercado acionário brasileiro.

No que se refere à estimação, a amostra foi dividida em três conjuntos de dados. O conjunto de treinamento abrange o período de janeiro de 2003 a dezembro de 2005, totalizando 36 meses. O conjunto de validação compreende o intervalo de janeiro de 2006 a dezembro de 2012, com 84 meses. Por fim, o conjunto de teste cobre o período de janeiro de 2013 a dezembro de 2022, totalizando 120 meses. A escolha dessa estrutura segue a lógica de janelas de expansão e permite avaliar o desempenho dos modelos em um ambiente mais próximo daquele enfrentado pelos agentes de mercado, uma vez que as previsões são sempre geradas com base apenas nas informações disponíveis até cada ponto do tempo.

Tabela C1 – Estatística Descritiva das Carteiras baseadas no Valor de Mercado

Carteiras	Média do Retorno	Desvio Padrão dos Retornos	Coefficiente de Variação dos Retornos	Mínimo Retorno	Máximo Retorno	Valor de Mercado Médio (Milhões R\$)
<i>Small-ME</i>	0,0197	0,0934	4,7416	-0,3870	0,3064	242,84
<i>Small-Median-ME</i>	0,0199	0,0869	4,3723	-0,3476	0,3348	978,29
<i>Median-ME</i>	0,0162	0,0803	4,9663	-0,3615	0,3327	2.554,77
<i>Median-Large-ME</i>	0,0147	0,0735	5,0067	-0,3418	0,3040	5.941,74
<i>Large-ME</i>	0,0080	0,0715	8,8862	-0,4968	0,1891	39.399,89

Nota: Esta tabela apresenta as estatísticas descritivas (média dos retornos, desvio padrão dos retornos, coeficiente de variação dos retornos, mínimo retorno, máximo retorno e valor de mercado médio em milhões) para as cinco carteiras de ativos baseadas no tamanho, mensuradas pelo valor de mercado dos ativos. *Small-ME* são os ativos do primeiro quintil; *Small-Median-ME* são os ativos do segundo quintil; *Median-ME* são os ativos do terceiro quintil; *Median-Large-ME* são os ativos do quarto quintil; e *Large-ME* são os ativos do maior quintil.

APÊNDICE D. Especificação dos modelos base

Este apêndice sintetiza os modelos base utilizados no artigo. O CAPM constitui a especificação de referência e relaciona o retorno esperado do ativo ao prêmio de risco de mercado. O Learning CAPM mantém a estrutura do CAPM tradicional, diferenciando-se pela forma de obtenção do coeficiente beta, que é construído a partir do filtro de Kalman. O Downside CAPM busca captar assimetrias associadas ao risco de mercado em cenários de queda, tornando-se particularmente útil em contextos de perdas relevantes. O Conditional CAPM amplia a estrutura tradicional ao incorporar variáveis condicionantes de natureza macroeconômica e financeira. Por fim, a regressão *cross-section* de Fama e French (1996) permite relacionar os retornos esperados às características das firmas e à sensibilidade dos ativos em relação ao mercado. Em conjunto, esses modelos representam diferentes formas de modelagem do risco e do retorno, fornecendo a base informacional que posteriormente será combinada pelos meta-modelos de *stacked generalization*.

D1 - CAPM e Learning CAPM:

$$E(R_i) = R_f + \beta_i[E(R_m) - R_f]$$

D2 - Downside CAPM (D-CAPM):

$$R_i = R_f + (R_m - R_f) \beta^D$$

$$\beta^D = \frac{E(\text{Min}((r_i - \bar{r}_i), 0)) \times E(\text{Min}((r_M - \bar{r}_M), 0))}{E(\text{Min}((r_i - \bar{r}_i), 0)^2)}$$

D3 - Conditional CAPM (C-CAPM):

$$R_{i,t+1}^i = \alpha^i + \beta^{1i} R_{t+1}^M + \beta^{2i} R_{t+1}^M R_m - R_{f,t} + \beta^{3i} R_{t+1}^M HML_t + \beta^{4i} R_{t+1}^M Spread_t + \epsilon_{t+1}$$

D4 - Regressão *cross-section* de Fama & French (1996):

$$R_{i,t+\tau} = \lambda_{t+\tau}^0 + \lambda_{t+\tau}^M \beta_{i,t-1} + \lambda_{t+\tau}^{ME} \ln(ME_{i,t-1}) + \lambda_{t+\tau}^{BM} \ln(BM_{i,t-1}) \\ + v_{i,t+\tau}, \forall i, \text{ para cada } (t + \tau)$$

APÊNDICE E. Meta-modelos e métricas de avaliação

Os meta-modelos empregados na técnica de *stacked generalization* foram Linear Regression (LR), Random Forest (RF), Gradient Boosted Regression Trees (GBRT) e Neural Networks (NN). Cada um desses algoritmos recebe como entrada as previsões geradas pelos modelos base selecionados no conjunto de validação, com o objetivo de produzir uma previsão refinada para os retornos das carteiras. A utilização de meta-modelos com diferentes níveis de complexidade permite avaliar se estruturas mais flexíveis são capazes de ampliar o desempenho preditivo das combinações de previsões no mercado acionário brasileiro. A avaliação empírica do desempenho preditivo foi realizada por meio de duas famílias de métricas. A primeira é o R_{005}^2 de Campbell e Thompson (2008), que permite comparar o desempenho das previsões com o benchmark da média histórica. A segunda é composta por medidas de previsão de sinal, como acurácia, sensibilidade e especificidade. Essas medidas são particularmente úteis porque captam a capacidade dos modelos em prever corretamente a direção dos retornos, aspecto relevante em contextos de decisão de investimento.

Figura E1 – Matriz de confusão da previsão de sinais

		PREVISÃO	
		Evento Positivo	Evento Negativo
REAL	Evento Positivo	Verdadeiro Positivo (TP)	Falso Negativo (FN)
	Evento Negativo	Falso Positivo (FP)	Verdadeiro Negativo (TN)

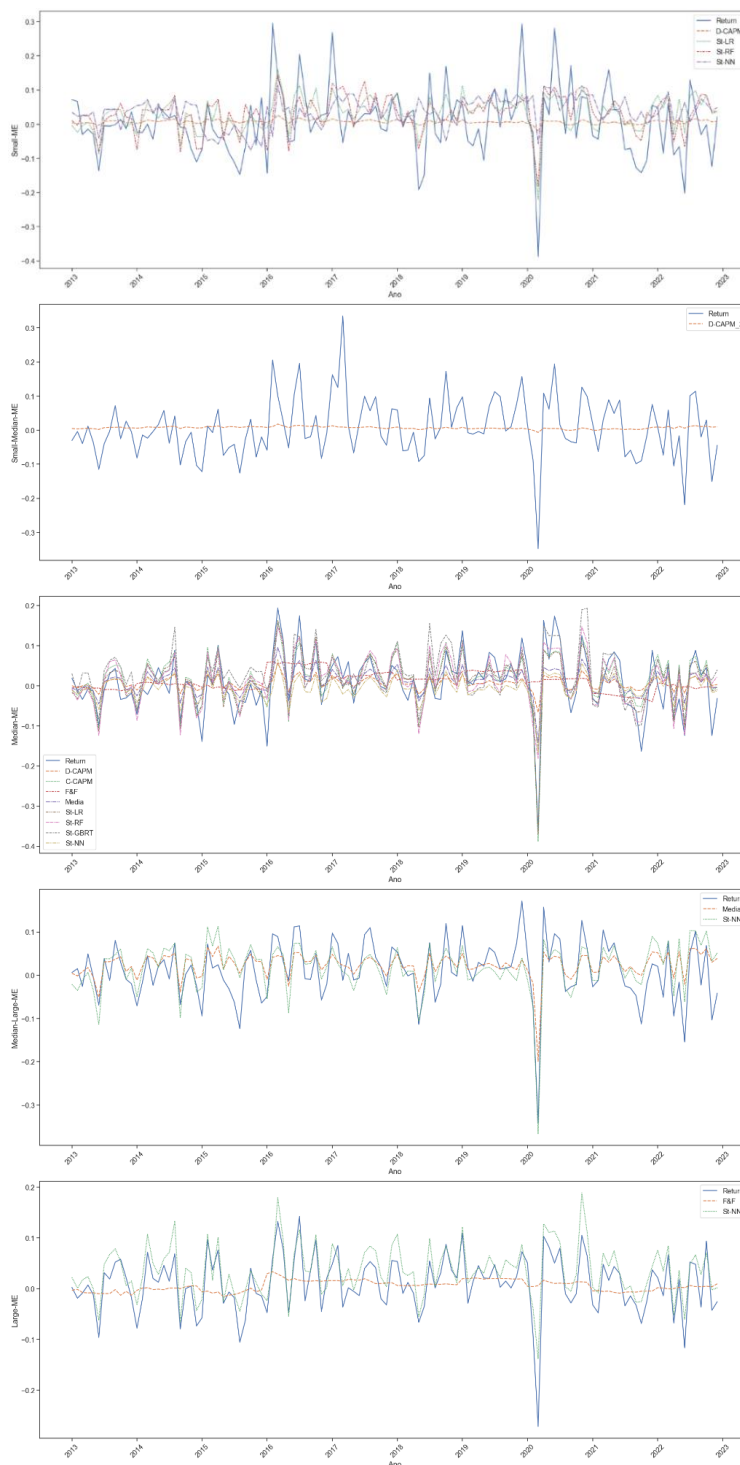
A Figura E1 apresenta a matriz de confusão da previsão de sinais, que serve de base para o cálculo dessas métricas. Em termos práticos, a matriz permite organizar os acertos e erros de classificação entre movimentos positivos e negativos dos retornos, facilitando a interpretação da acurácia geral e do desempenho específico dos modelos em antecipar altas e quedas do mercado. Essa representação é particularmente importante no contexto do artigo, pois parte relevante da contribuição empírica está associada à capacidade dos meta-modelos em melhorar a previsão de sinais, especialmente em períodos de declínio.

APÊNDICE F. Evidências complementares do conjunto de teste

F1 - Retorno Previsto das Carteiras No conjunto de Teste

Este apêndice reúne evidências complementares para a análise do conjunto geral de teste. A Figura F1 apresenta os retornos previstos dos principais modelos e meta-modelos para as carteiras estudadas, permitindo uma avaliação visual do alinhamento entre as previsões e os movimentos observados dos retornos. Em particular, a figura reforça o destaque do meta-modelo St-LR na carteira Small-ME, bem como o bom desempenho relativo do St-RF nesse mesmo segmento. As evidências visuais também mostram que os ganhos associados à técnica não são uniformes entre todas as carteiras, o que está em linha com a discussão apresentada no corpo principal do artigo.

Figura F1 – Retorno previsto das carteiras no conjunto de teste



F2 - Resultados do Teste de Wilcoxon na Previsão dos Meta-modelos

A Tabela F2 apresenta os resultados do teste de Wilcoxon aplicado às previsões dos meta-modelos com e sem seleção dos melhores modelos base. O objetivo desse teste é verificar se há diferenças estatisticamente significativas entre as previsões geradas sob essas duas estratégias. Os resultados indicam que, em algumas carteiras e para determinados meta-modelos, há diferenças significativas entre as previsões. Contudo, a interpretação substantiva dessas diferenças deve ser feita com cautela, pois o teste funciona aqui como evidência complementar e não altera a conclusão central do artigo. Assim, a principal utilidade deste apêndice está em documentar que o processo de seleção dos modelos base pode afetar aspectos específicos das previsões, ainda que não modifique de forma generalizada o núcleo dos resultados centrais.

Tabela F2 – Resultados do Teste de Wilcoxon na Previsão dos Meta-modelos

Modelos	<i>Small</i>-ME	<i>Small-Median</i>-ME	<i>Median</i>-ME	<i>Median-Large</i>-ME	<i>Large</i>-ME
St-LR	0,1067	0,0031***	0,2106	0,0034***	0,0000***
St-RF	0,7139	0,4050	0,5297	0,0297**	0,0008***
St-GBRT	0,0000	0,3947	0,0053***	0,1149	0,033**
St-NN	0,0601*	0,0000***	0,0000***	0,0000***	0,0000***

Notas: St significa *stacked*. St-LR: é o *stacked* com o modelo de Regressão Linear; St-RF: é o *stacked* com modelo de *Random Forest*; St-GBRT: é o *stacked* com o modelo de *Gradient Boosted Regression Trees*; St-NN: é o *stacked* com o modelo de *Neural Networks*. *, **, e *** indicam significância estatística nos níveis de 10%, 5% e 1%, respectivamente.

APÊNDICE G. Identificação dos períodos extremos de mercado

A identificação dos períodos extremos de mercado considerados no artigo foi realizada com base no comportamento do retorno mensal do Ibovespa e na análise do *drawdown* mensal do índice. As Figuras G1 e G2 apresentam, respectivamente, a distribuição anual do retorno logarítmico mensal do Ibovespa e do *drawdown* mensal ao longo do período analisado. Essas representações permitem visualizar de forma mais clara os episódios em que o mercado brasileiro experimentou perdas mais intensas e persistentes. A leitura conjunta dessas figuras indica que os episódios de maior deterioração do mercado coincidem com momentos de forte instabilidade econômica e financeira, com destaque para a crise financeira global de 2008–2009, a crise político-econômica de 2015–2016 e a pandemia de Covid-19 em 2020. No contexto do artigo, a análise empírica concentrou-se nos dois últimos episódios, por estarem contidos no conjunto de teste. Esse procedimento fortalece a interpretação dos resultados em condições extremas de mercado, pois mostra que a escolha dessas janelas não decorreu de arbitrariedade, mas de evidências observáveis no próprio comportamento do Ibovespa.

Figura G1 – Boxplot Anual do Retorno Logarítmico Mensal do Ibovespa

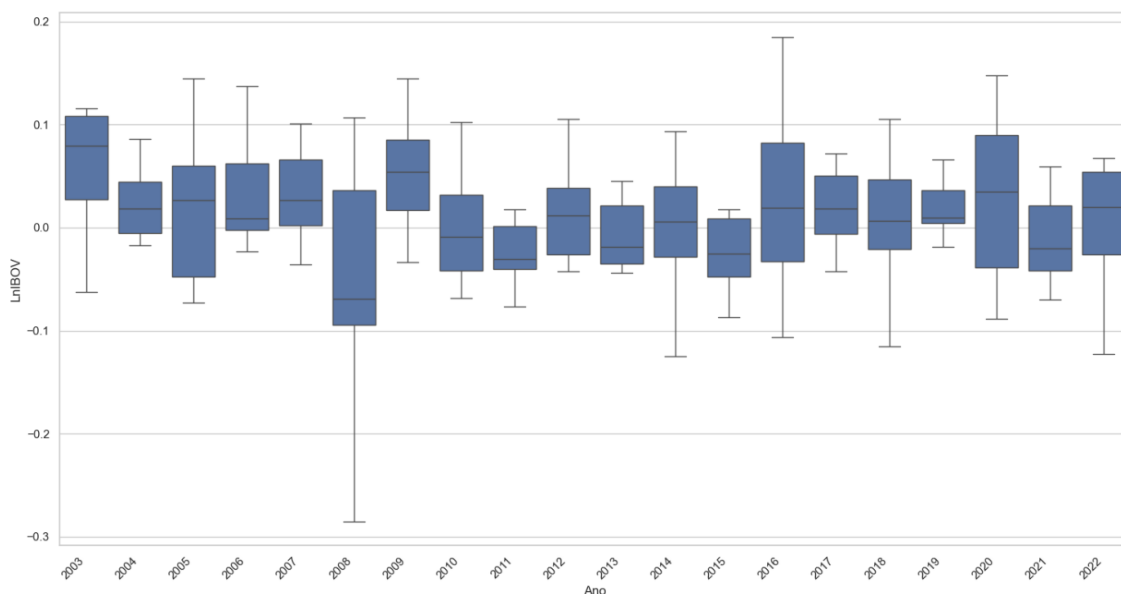
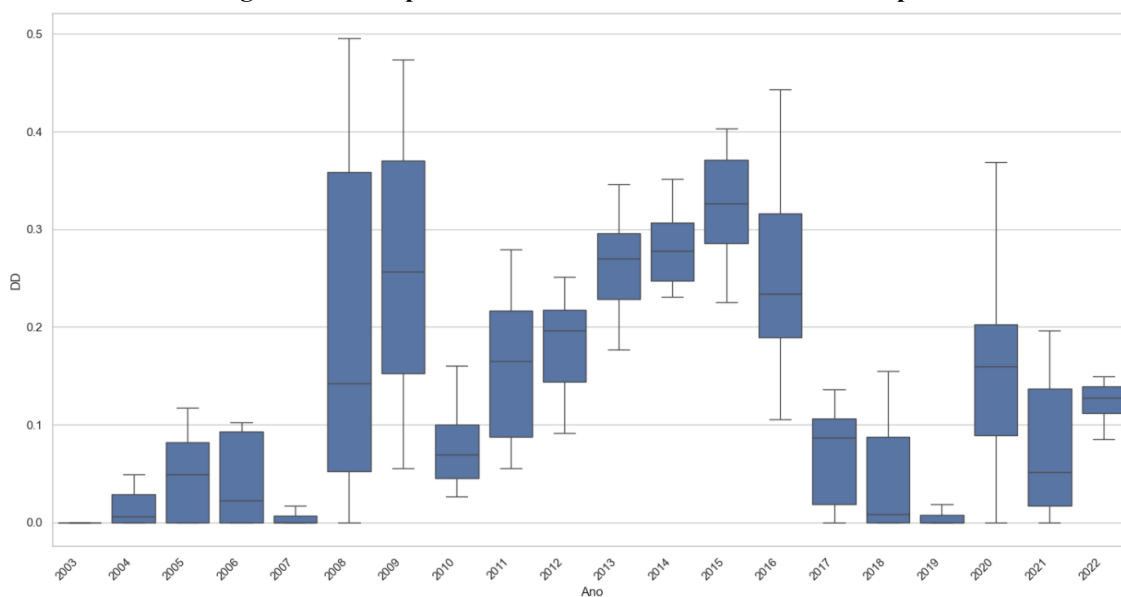


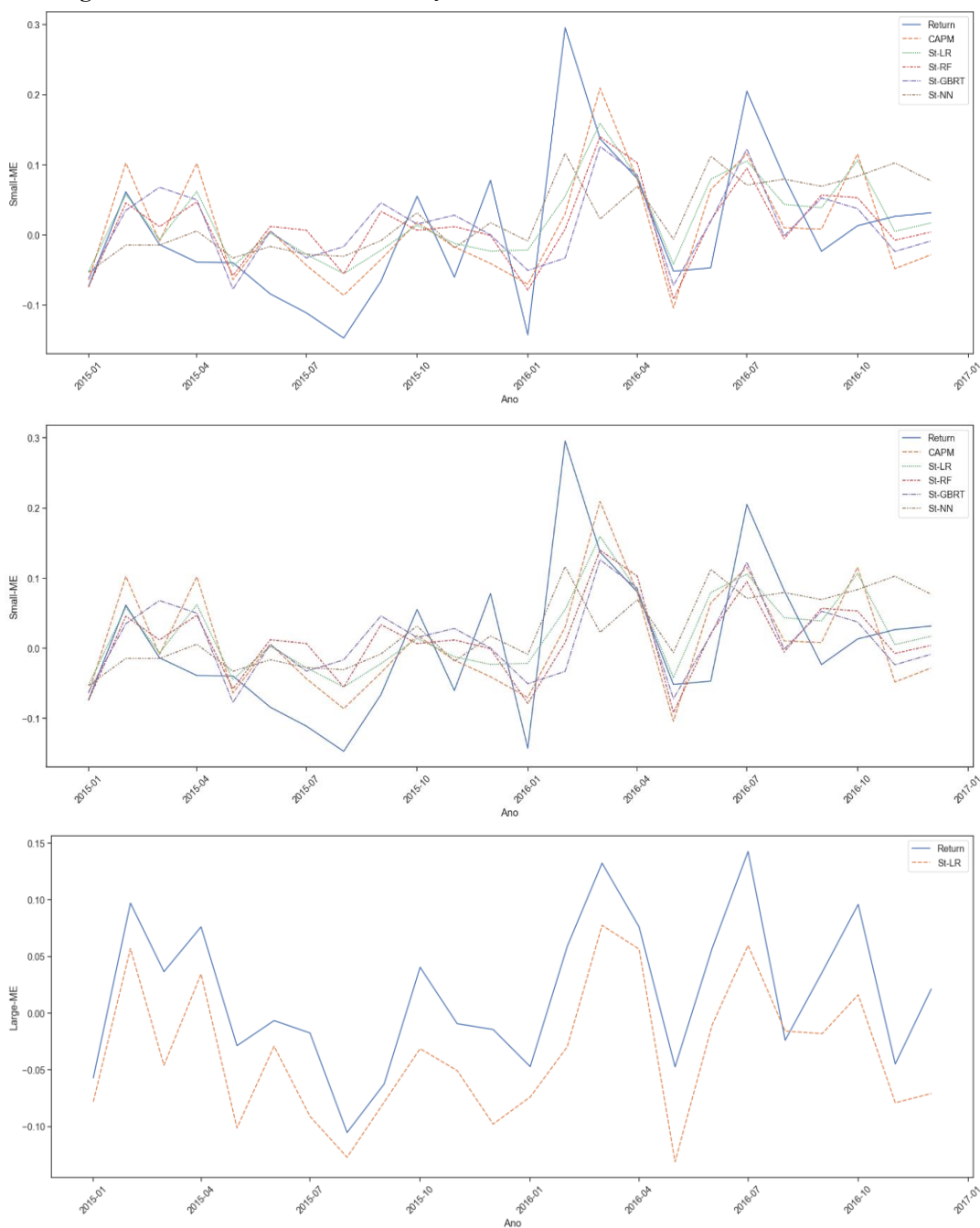
Figura G2 – Boxplot anual do DrawDown Mensal do Ibovespa



APÊNDICE H. Evidências visuais das crises

A Figura H1 apresenta os retornos previstos das ações durante a crise político-econômica de 2015–2016. Seu objetivo é complementar a evidência da tabela principal de resultados para esse período, permitindo visualizar de forma mais intuitiva como os modelos base e os meta-modelos acompanharam os movimentos dos retornos em um ambiente de forte turbulência. A figura é especialmente útil para mostrar que, em determinados segmentos da amostra, a técnica de *stacked generalization* preserva parte de sua capacidade preditiva mesmo em contextos adversos. Ao mesmo tempo, a evidência visual também reforça a interpretação de que esses ganhos são condicionais. Ainda que alguns meta-modelos acompanhem relativamente bem os movimentos dos retornos em certos momentos, o desempenho não é uniforme ao longo de toda a crise nem entre todas as carteiras. Por isso, este apêndice deve ser lido como complemento visual da análise principal, e não como substituto da evidência quantitativa apresentada no corpo do artigo.

Figura H1 – Retorno Previsto das Ações durante a Crise Político-Econômica de 2016

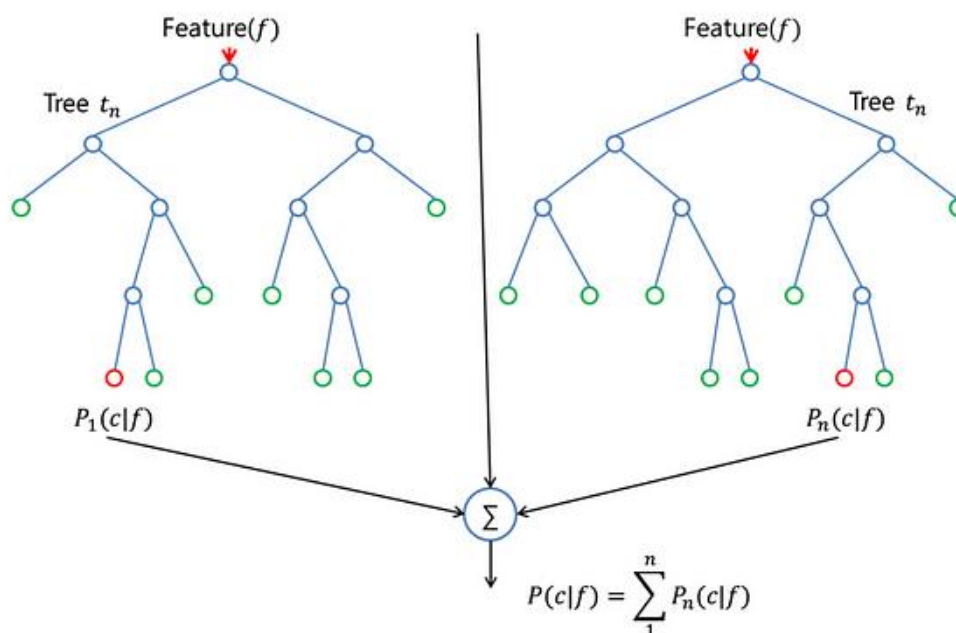


APÊNDICE I. Introdução aos meta-modelos

I1 - *Random Forest* (RF)

A Floresta Aleatória, *Random Forest* (RF), foi proposta formalmente por Breiman (2001). Como o nome sugere, é uma técnica usada para combinar previsões ou classificadores de um certo número de modelos conhecidos como Árvores de Decisão (*Decision Trees* – DT). A DT é um modelo intuitivo que faz perguntas simples sequencialmente. Uma pergunta geralmente é chamada de “nó”, e quando um nó não possui nós filhos, é chamado de nó folha, que encerra o processo de perguntar. A previsão é realizada pela média dos valores das amostras que caem no mesmo nó. As perguntas e as ordens dessas perguntas feitas nas DT’s são essenciais para o desempenho dos modelos. Uma desvantagem da DT é que, embora a cada passo ele encontre o melhor par para dividir os dados, ele não necessariamente encontra a sequência ótima geral. Outra desvantagem desse modelo é que ele é instável. Uma pequena alteração nos dados geralmente resulta em uma alteração significativa do modelo, tornando o modelo menos propenso a generalizar bem para novos dados.

Figura I1 – Exemplo de *Random Forest* (RF)



A RF é frequentemente utilizada para superar a instabilidade da DT, pois calcula a média das previsões sobre muitas árvores. Por exemplo, para uma amostra X_i ser previsto, se a primeira árvore de decisão prever o valor 0,01, a segunda árvore prever o valor 0,02 e a terceira árvore de decisão prever o valor 0,025, então a decisão final da RF com essas três árvores seria 0,0183, a média dessas três previsões. A aleatoriedade na RF geralmente se refere à reamostragem das amostras de treinamento para cada árvore (ZHAO e CHENG, 2022). Uma grande vantagem da técnica de RF é que ela pode ser utilizada tanto para tarefas de classificação quanto para regressão. A Figura I1, acima, apresenta uma RF com duas árvores.

I2 - *Gradient Boosted Regression Trees* (GBRT)

Árvores de Regressão Impulsionadas por Gradiente, *Gradient Boosted Regression Trees* (GBRT), é um algoritmo de *ensemble learning* em que ocorre a combinação de diversas árvores de decisão. Resumidamente, o GBRT funciona organizando as DT’s de maneira sequencial, fazendo com que as próximas árvores corrijam os erros residuais das anteriores. Esse método é diferente da *Random Forest*, pois as árvores não são paralelas, mas uma após a outra. Um pseudocódigo simples, elaborado por Zhao e Cheng (2022), é apresentado na Figura I2. Para uma introdução mais detalhada, ver Gu, Kelly e Xiu (2020).

Figura I2 – Exemplo de *Gradient Boosted Regression Trees (GBRT)*

GBRT example

Tree 1:

```
tree1 = DecisionTreeRegressor(max_depth = 3)
tree1.fit(X, y)
resid1 = y - tree1.predict(X)
```

Tree 2:

```
tree2 = DecisionTreeRegressor(max_depth = 2)
tree1.fit(X, resid1)
resid2 = resid1 - tree2.predict(X)
```

Tree 3:

```
tree3 = DecisionTreeRegressor(max_depth = 2)
tree3.fit(X, resid2)
```

Combine:

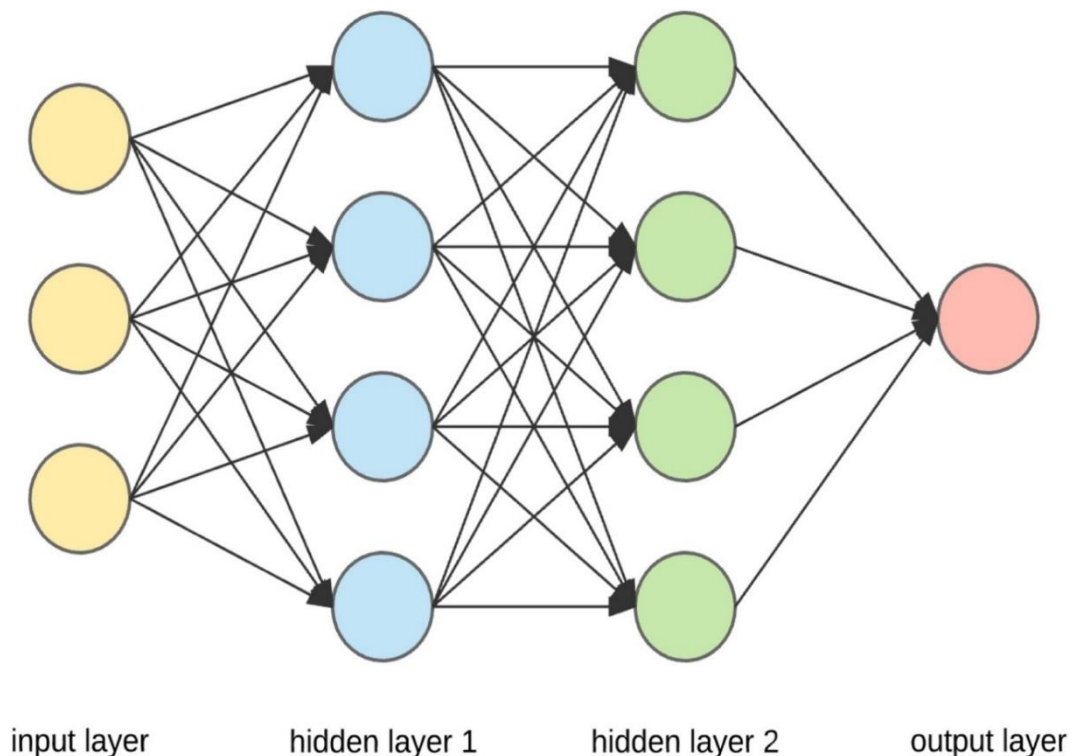
```
y_pred = tree1.predict(X_new) + tree2.predict(X_new) + tree3.predict(X_new)
```

Fonte: Zhao e Cheng (2020)

I3 - *Neural Networks (NN)*

As redes neurais, *Neural Networks (NN)*, como o nome sugere, tem suas origens nas tentativas de encontrar representações matemáticas do processamento de informações em sistemas biológicos (SANTOS, 2020). As NN's são modelos inspirados na estrutura do cérebro humano, e consistem em várias camadas de neurônios interconectados, cada neurônio recebendo entradas, realizando operações e passando os resultados destas operações para os neurônios na camada seguinte. A ideia central por trás de um algoritmo de NN é que a rede neural aprende uma função que mapeia as entradas para as saídas desejadas. Durante o treinamento, os pesos das conexões entre os neurônios são ajustados iterativamente com base nos dados de treinamento, de modo a minimizar uma função de perda que quantifica o quão distantes as saídas previstas da rede estão dos valores reais. Uma NN pode consistir de múltiplas camadas de neurônios, incluindo uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Cada neurônio em uma camada está conectado a todos os neurônios na camada subsequente. A Figura I3 apresenta um exemplo de NN denso com uma camada de entrada, duas camadas ocultas e uma camada de saída.

Figura I3 – Exemplo de *Neural Networks (NN)*



Fonte: Zhao e Cheng (2020)