

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
DEPARTAMENTO DE CIÊNCIAS ECONÔMICAS
GRADUAÇÃO EM CIÊNCIAS ECONÔMICAS

MARCUS PAULO DA SILVA ALVES

**APLICAÇÃO DE APRENDIZADO DE MÁQUINA À TEORIA MODERNA DO
PORTFÓLIO: UMA AVALIAÇÃO NO MERCADO ACIONÁRIO BRASILEIRO NO BIÊNIO
2018-2019**

JUIZ DE FORA
2025

MARCUS PAULO DA SILVA ALVES

**APLICAÇÃO DE APRENDIZADO DE MÁQUINA À TEORIA MODERNA DO
PORTFÓLIO: UMA AVALIAÇÃO NO MERCADO ACIONÁRIO BRASILEIRO NO BIÊNIO
2018-2019**

Monografia apresentada ao Departamento de Ciências Econômicas da Universidade Federal de Juiz de Fora, como requisito para obtenção do título de bacharel em Ciências Econômicas.

Orientador: Prof. Dr. Paulo César Coimbra
Lisbôa

JUIZ DE FORA

2025

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Alves, Marcus Paulo da Silva.

APLICAÇÃO DE APRENDIZADO DE MÁQUINA À TEORIA MODERNA DO PORTFÓLIO: UMA AVALIAÇÃO NO MERCADO ACIONÁRIO BRASILEIRO NO BIÊNIO 2018-2019 / Marcus Paulo da Silva Alves. -- 2025.

64 p. : il.

Orientador: Paulo Cesar Coimbra Lisbôa

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Juiz de Fora, Faculdade de Economia, 2025.

1. Aprendizado de máquina. 2. Random Forest. 3. Teoria Moderna do Portfólio. 4. Média-variância. 5. Mercado acionário brasileiro. I. Lisbôa, Paulo Cesar Coimbra, orient. II. Título.



UNIVERSIDADE FEDERAL DE JUIZ DE FORA
REITORIA - FACECON - Depto. de Economia

FACULDADE DE ECONOMIA / UFJF
ATA DE APROVAÇÃO DE MONOGRAFIA II

NA DATA DE 13/03/2025, A BANCA EXAMINADORA, COMPOSTA PELOS PROFESSORES:

- 1 – PAULO CÉSAR COIMBRA LISBÔA - ORIENTADOR; E
- 2 – SIDNEY MARTINS CAETANO,

REUNIU-SE PARA AVALIAR A MONOGRAFIA DO ACADÊMICO MARCUS PAULO DA SILVA ALVES, INTITULADA:

APLICAÇÃO DE APRENDIZADO DE MÁQUINA À TEORIA MODERNA DO PORTFÓLIO: UMA AVALIAÇÃO NO MERCADO ACIONÁRIO BRASILEIRO NO BIÊNIO 2018-2019.

APÓS PRIMEIRA AVALIAÇÃO, RESOLVEU A BANCA SUGERIR ALTERAÇÕES AO TEXTO APRESENTADO, CONFORME RELATÓRIO SINTETIZADO PELO ORIENTADOR. A BANCA, DELEGANDO AO ORIENTADOR A OBSERVÂNCIA DAS ALTERAÇÕES PROPOSTAS, RESOLVEU APROVAR A REFERIDA MONOGRAFIA.



Documento assinado eletronicamente por **Paulo César Coimbra Lisbôa, Professor(a)**, em 17/03/2025, às 16:50, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Sidney Martins Caetano, Professor(a)**, em 18/03/2025, às 20:36, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no Portal do SEI-Uffj (www2.uffj.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **2296894** e o código CRC **8A4898F4**.

RESUMO

O presente estudo avaliou a aplicação do algoritmo de aprendizado de máquina *Random Forest*, incorporado à Teoria Moderna do Portfólio, na seleção e otimização de portfólios financeiros no mercado acionário brasileiro durante o biênio 2018-2019. O modelo foi utilizado para identificar ativos com maior probabilidade de retorno positivo, integrando esses resultados a um otimizador baseado no critério de média-variância, e comparado à abordagem apresentada pelo *Capital Asset Pricing Model*. Os resultados indicam que o classificador de *Random Forest* supera o de *Capital Asset Pricing Model* em termos de retorno absoluto e ajustado ao risco, medidos por métricas como Índice Sharpe, Índice de Treynor e Alfa de Jensen. O algoritmo demonstrou melhor desempenho em carteiras com poucos ativos, enquanto, à medida que a diversificação aumenta, as estratégias tendem a convergir em termos de performance. Esses achados ressaltam o potencial do aprendizado de máquina como uma ferramenta robusta para otimizar a relação risco-retorno no mercado financeiro brasileiro.

Palavras-chave: Aprendizado de máquina. *Random Forest*. Teoria Moderna do Portfólio. Média-variância. Mercado acionário brasileiro.

ABSTRACT

This study evaluated the application of the Random Forest machine learning algorithm, integrated with Modern Portfolio Theory, in the selection and optimization of financial portfolios in the Brazilian stock market during the 2018-2019 biennium. The model was employed to identify assets with a higher probability of positive returns, incorporating these results into an optimizer based on the mean-variance criterion and comparing the approach to the Capital Asset Pricing Model.

The results indicate that the Random Forest classifier outperforms the Capital Asset Pricing Model in terms of absolute and risk-adjusted returns, as measured by metrics such as the Sharpe Ratio, Treynor Ratio, and Jensen's Alpha. The algorithm exhibited superior performance in portfolios with a limited number of assets, while, as diversification increased, the strategies tended to converge in terms of performance. These findings highlight the potential of machine learning as a robust tool for optimizing the risk-return trade-off in the Brazilian financial market.

Keywords: Machine Learning. *Random Forest*. Modern Portfolio Theory. Mean-variance. Brazilian stock market.

LISTA DE FIGURAS

| | |
|---|----|
| Figura 1 - Esquematização do processo de otimização de portfólios com pré-seleção de ativos | 29 |
| Figura 2 - Exemplificação da separação de períodos na base de validação | 39 |
| Figura 3 - Resultados da avaliação do modelo de Random Forest | 41 |
| Figura 4 - Importância das variáveis do modelo de Random Forest | 41 |
| Figura 5 - log-retorno acumulado, 2018-2019, por estratégia e número de ativos no portfólio | 47 |
| Figura 6 - log-retorno acumulado obtido pelo classificador de Random Forest | 48 |
| Figura 7 - log-retorno acumulado obtido pelo classificador de Capital Asset Pricing Model | 49 |
| Figura 8 – Índice Sharpe de acordo com o número de ativos disponibilizados ao otimizador | 57 |
| Figura 9 - Índice de Treynor de acordo com o número de ativos disponibilizados ao otimizador | 58 |
| Figura 10 - Alfa de Jensen de acordo com o número de ativos disponibilizados ao otimizador | 59 |
| Figura 11 - Log-retorno obtido de acordo com o número de ativos disponibilizados ao otimizador | 60 |
| Figura 12 - Comparação do Índice Sharpe das estratégias para carteiras com 1, 5, 15, 30, 50 e 71 ativos | 61 |
| Figura 13 - Comparação do Índice de Treynor das estratégias para carteiras com 1, 5, 15, 30, 50 e 71 ativos | 62 |
| Figura 14 - Comparação do Alfa de Jensen das estratégias para carteiras com 1, 5, 15, 30, 50 e 71 ativos | 63 |
| Figura 15 - Comparação do log-retorno das estratégias para carteiras com 1, 5, 15, 30, 50 e 71 ativos | 64 |

LISTA DE TABELAS

| | |
|---|----|
| Tabela 1 – Descrição das variáveis utilizadas no modelo de aprendizado de máquina | 32 |
| Tabela 2 - Bibliotecas utilizadas para coleta e manipulação dos dados..... | 33 |
| Tabela 3 - Divisão dos dados históricos e número de registros por etapa | 37 |
| Tabela 4 - Períodos de validação do otimizador | 40 |
| Tabela 5 - Dimensão das bases de treino e teste | 40 |
| Tabela 6 - indicadores médios por estratégia e número de ativos | 50 |

LISTA DE ABREVIATURAS E SIGLAS

| | |
|-----|---------------------------------|
| HME | Hipótese de Mercados Eficientes |
| TMP | Teoria Moderna do Portfólio |

SUMÁRIO

| | |
|--|-----------|
| 1. INTRODUÇÃO..... | 12 |
| 2. REFERENCIAL TEÓRICO | 16 |
| 2.1 Teoria da seleção de carteiras: o modelo de Markowitz | 16 |
| 2.2 Teorias sobre risco e retorno de ativos | 18 |
| 2.2.1 <i>Capital Asset Pricing Model</i> | 18 |
| 2.2.2 <i>Arbitrage Pricing Theory</i> | 19 |
| 2.2.3 Modelo de fatores | 20 |
| 2.3 Liquidez..... | 20 |
| 2.4 Métricas para mensuração do desempenho de um portfólio | 21 |
| 2.4.1 Índice Sharpe | 21 |
| 2.4.2 Índice de Treynor | 22 |
| 2.4.3 Alfa de Jensen | 23 |
| 2.5 Hipótese de Mercados Eficientes..... | 23 |
| 2.6 Momento | 24 |
| 2.7 Aprendizado de máquina | 25 |
| 2.7.1 Aplicações empíricas de aprendizado de máquina no mercado de ações | 25 |
| 2.7.2 <i>Random Forest</i> | 27 |
| 3. METODOLOGIA..... | 29 |
| 4. BASE DE DADOS | 32 |
| 4.1 Coleta dos dados | 33 |
| 4.1.1 Tratamento e preparação dos dados | 34 |
| 4.1.2 Setor | 35 |
| 4.2 Construção das variáveis para as etapas de treino e teste..... | 35 |
| 4.2.1 Variáveis explicativas..... | 37 |
| 4.2.2 Variável resposta | 38 |

| | | |
|-----------|---|-----------|
| 4.3 | Tratamento para a base de validação..... | 38 |
| 4.4 | Classificação das ações..... | 39 |
| 4.5 | Avaliação e comparação de portfólios | 42 |
| 4.6 | Estruturação da classificação de via <i>Capital Asset Pricing Model</i> | 43 |
| 5. | RESULTADOS | 45 |
| 5.1 | Avaliação e comparação dos classificadores..... | 45 |
| 5.2 | <i>Backtesting</i> de estratégias de investimento | 47 |
| 6. | CONCLUSÃO | 52 |
| | REFERÊNCIAS..... | 54 |
| | APÊNDICE A – RESULTADOS DOS CLASSIFICADORES POR PERÍODO DE VALIDAÇÃO | 57 |
| | APÊNDICE B – COMPARAÇÃO DAS ESTRATÉGIAS..... | 61 |

1. INTRODUÇÃO

O número de investidores brasileiros tem aumentado consideravelmente nos últimos anos. Segundo a 6ª edição do Raio X do Investidor Brasileiro (ANBIMA, 2023) a proporção de brasileiros investindo em produtos financeiros cresceu 5 pontos percentuais de 2021 para 2022, atingindo 36% da população. Na 7ª edição do estudo, divulgada em 2024, a ANBIMA trouxe novos dados: em 2023, mesmo com uma queda na disposição da população em poupar, o percentual de brasileiros investindo em produtos financeiros subiu para 37%, um acréscimo de 1 ponto percentual em relação ao ano anterior. O mesmo estudo projeta um aumento de 4 pontos percentuais no número de investidores para o ano de 2024.

Ainda de acordo com a 7ª edição do Raio X do Investidor Brasileiro, o comportamento da população em relação às aplicações financeiras segue o mesmo padrão dos últimos anos: os investidores brasileiros continuam a preferir produtos mais conservadores. A caderneta de poupança é o investimento mais utilizado, com destaque para dois grupos etários em particular: os *boomers* (31%), que correspondem a pessoas de 63 anos ou mais, e a geração X (29%), composta por indivíduos entre 43 e 62 anos.

Por outro lado, para a geração Z, composta principalmente por pessoas da classe C (48%), que representa aproximadamente 22% do total de investidores e engloba jovens entre 16 e 27 anos, essa realidade é um pouco diferente. Apenas 16% dos jovens preferem a poupança. Nessa geração, destacam-se as moedas digitais, criptomoedas e criptoativos (8%) e os fundos de investimento (6%), alternativas notoriamente caracterizadas pela expectativa de obtenção de um maior retorno frente a investimentos tradicionais mais conservadores. Apesar de ainda considerarem a segurança financeira a principal vantagem de investir em produtos financeiros, a geração Z foi a que mais se preocupou com o retorno dos investimentos em 2023, com 28% priorizando este aspecto.

Segundo Le Sourd (2007), mais do que única e exclusivamente o retorno, os investidores desejam dispor e avaliar todas as informações necessárias no processo de seleção de fundos de investimentos, isto é: se o seu retorno foi suficientemente elevado para recompensar os riscos assumidos, como a estratégia se compara aos seus pares e, por fim, se os resultados da gestão da carteira se deram devido à sorte ou porque a estratégia utilizada possui vantagens reais que se sustentarão no futuro.

Por essa razão, a avaliação da performance de carteiras deve ser mensurada não apenas em termos do retorno absoluto obtido, mas também tendo em mente os fatores acima expostos.

Dado o cenário de maior inserção da população brasileira no mercado financeiro e alinhadas às metas de preocupação com segurança, sobretudo da população mais velha, e de retorno do investimento, reivindicada pela população mais jovem, faz-se necessária a análise de risco-retorno proporcionadas pelas diferentes opções de investimentos para uma adequada tomada de decisão por parte dos investidores.

Em seu artigo seminal, *Portfolio Selection* (1952), Markowitz afirma que, graças à diversificação, o retorno esperado de um portfólio pode aumentar sem que necessariamente se incorra em um maior risco. Essa contribuição fundamental deu origem ao que hoje é conhecida como Teoria Moderna do Portfólio (TMP).

Outra área que ganhou bastante destaque após o surgimento da TMP foi a de finanças quantitativas, entendido como “a aplicação de matemática – especificamente a teoria da probabilidade – nos mercados financeiros” (BELL, 2016, p. 8). Nos últimos anos, com a implementação de modelos computacionais de aprendizado de máquina visando a projeção da direção dos movimentos dos preços das ações com base em dados históricos, esse campo passou a conquistar ainda maior notoriedade no estudo de finanças.

No Brasil, todavia, ainda há uma escassez significativa de estudos sobre a aplicação de modelos aprendizado de máquina na construção de portfólios. Trabalhos como as de Costa (2022) e Abreu (2021) avançaram nessa esfera, mas com algumas limitações. O estudo de Abreu (2021), por exemplo, cujo trabalho inspirou o vigente projeto de pesquisa, apesar de apresentar uma metodologia robusta na aplicação do algoritmo de *Random Forest* e dispor de uma ampla base histórica de cotações de ativos, utilizou o biênio 2020-2021 como período de validação das projeções realizadas pelo modelo, durante o auge da pandemia de COVID-19.

Durante períodos de crise ou volatilidade extrema, como o ocorrido durante a pandemia de COVID-19, as condições de mercado podem se desviar significativamente de padrões históricos, gerando movimentos atípicos nos preços e no comportamento dos investidores. Um exemplo claro dessa distorção foi a sequência recorde de *circuit breakers* entre os dias 9 e 18 de março de 2020, evidenciando níveis excepcionalmente elevados e persistentes de volatilidade. Esses

fatores podem fazer com que modelos de aprendizado de máquina capturem padrões específicos desse contexto anômalo, que podem não ser replicáveis em períodos de mercado mais estáveis.

Portanto, validar a utilização do modelo de aprendizado de máquina *Random Forest* em um período menos impactado por eventos excepcionais é fundamental para verificar se o algoritmo é capaz de manter um desempenho consistente e superior em condições de mercado mais regulares, como as observadas no biênio 2018-2019.

Mais do que apenas validar a metodologia de Abreu (2021), este estudo tem por objetivo investigar a eficácia do algoritmo *Random Forest* na classificação de ativos para a otimização da relação risco-retorno em portfólios financeiros. Busca-se avaliar se essa técnica pode contribuir significativamente na construção de carteiras de ações que maximizem o retorno esperado levando em consideração o nível de risco incorrido. Por fim, os resultados obtidos serão comparados com aqueles gerados a partir da classificação de ativos via *Capital Asset Pricing Model*.

O algoritmo de *Random Forest* foi selecionado com base em evidências empíricas que indicam seu bom desempenho na classificação de ativos financeiros. Estudos como os de Kumar e Thenmozhi (2006), Haughland (2023), Tan et al. (2019), Ma et al. (2021) e Costa (2022) apontam que o *Random Forest* consistentemente figura entre os algoritmos mais eficientes para essa tarefa, superando outras abordagens de aprendizado de máquina. Nesse sentido, sua escolha se justifica pela robustez, capacidade de processar grandes volumes de dados e resistência ao *overfitting*, características essenciais para modelos aplicados ao mercado financeiro.

A comparação com o *Capital Asset Pricing Model*, por sua vez, é explicada pelo fato de este ser um modelo consolidado e amplamente utilizado na precificação de ativos, servindo como um *benchmark* adequado para avaliar o desempenho do *Random Forest*. Dessa forma, é possível verificar se a abordagem baseada em aprendizado de máquina oferece ganhos significativos em relação a um modelo tradicional.

Com esse projeto de pesquisa, portanto, pretende-se responder à seguinte questão: a implementação do algoritmo *Random Forest* como classificador de ativos, integrado à Teoria Moderna do Portfólio, com o propósito de prever o retorno esperado futuro dos preços de ativos financeiros e construir um portfólio de investimentos no mercado acionário brasileiro, apresentou retornos superiores para os investidores, tanto em termos absolutos quanto ponderados pelo risco, conforme as métricas Índice

Sharpe, Índice de Treynor e Alfa de Jensen, quando comparados aos resultados obtidos pela classificação de ativos via *Capital Asset Pricing Model* no biênio 2018-2019?

Para isso, foi desenvolvido, com base no algoritmo de *Random Forest*, um modelo capaz de estimar, com base num conjunto de variáveis explicativas, a probabilidade de cada ativo apresentar um retorno acumulado positivo nos 20 dias úteis seguintes. Em seguida, os ativos foram ordenados de acordo com essa probabilidade e submetidos a um otimizador de portfólios baseado no critério de média-variância.

O impacto da pré-seleção de ativos sobre a performance das carteiras otimizadas foi examinado, verificando a capacidade do modelo de identificar consistentemente os ativos mais promissores. Por fim, foi comparado o desempenho das carteiras otimizadas com base na seleção de ativos pelo *Capital Asset Pricing Model*.

Os resultados encontrados mostram que o algoritmo de *Random Forest* superou consistentemente o *Capital Asset Pricing Model* em termos de log-retorno e das demais métricas avaliadas. As carteiras mais concentradas foram as que melhor performaram, evidenciando a excelente capacidade do algoritmo em identificar ativos com alto potencial de retorno. O log-retorno acumulado da carteira de 1 ativo classificada pelo algoritmo, por exemplo, atingiu 130,4% no biênio 2018-2019, demonstrando que a estratégia de aprendizado de máquina foi capaz de selecionar consistentemente os ativos mais promissores ao longo do período analisado.

Este projeto está dividido em seis seções, sendo a primeira a presente unidade introdutória. Na Seção 2, é apresentado o referencial teórico. Na Seção 3, é apresentada a metodologia implementada no projeto de pesquisa, seguida pela base de dados, na Seção 4. A Seção 5 discute-se os resultados obtidos. Por fim, na Seção 6 é realizada a conclusão e as implicações práticas do estudo.

2. REFERENCIAL TEÓRICO

Esta seção apresenta uma revisão dos principais estudos teóricos que embasam a construção do classificador de aprendizado de máquina para seleção de carteiras. São abordados conceitos fundamentais, como o modelo de média-variância, as principais teorias sobre risco e retorno de ativos, a relevância da liquidez e as métricas utilizadas na avaliação de desempenho de portfólios. Além disso, discute-se a Hipótese de Mercados Eficientes (HME) e o conceito de momento, que desafia os pressupostos da HME ao sugerir a persistência de tendências nos preços dos ativos.

Por fim, são analisados estudos empíricos que aplicam técnicas de aprendizado de máquina na classificação de ativos, com ênfase na implementação do algoritmo *Random Forest* e em sua eficácia na otimização de portfólios.

2.1 Teoria da seleção de carteiras: o modelo de Markowitz

O Modelo de Seleção de Carteiras, concebido em 1952 pelo economista Harry Markowitz, representa um marco pioneiro na análise de risco nos investimentos. Para fundamentar esse novo paradigma de seleção de ativos, Markowitz estabeleceu duas premissas centrais visando determinar as características essenciais de uma carteira:

A) o investidor busca maximizar o retorno para um dado nível de risco; e

B) o risco pode ser minimizado por meio da diversificação do portfólio entre ativos não correlacionados.

Markowitz assumiu que o investidor procura maximizar o retorno esperado e reduzir a variância, entendida como o risco do portfólio. Carteiras eficientes, portanto, buscam a menor variância por meio da diversificação dos ativos.

A variância da carteira é influenciada pela covariância entre os pares de ativos, e essa covariância está vinculada à correlação entre os ativos. Nesse sentido, o modelo indica a importância de selecionar ativos com baixa correlação entre si, para que um possa minimizar as perdas do outro. Para os investidores, é vantajoso construir uma carteira com ativos negativamente correlacionados, pois a baixa correlação não apenas reduz o risco, mas também proporciona uma combinação de ativos em que o risco é menor, sem uma diminuição proporcional do retorno.

Segundo Silva et al. (2020, p. 32), “quando se diversifica uma carteira, o risco pode ser mais bem distribuído e assim, o retorno pode ser maior, não importando qual

seja a combinação de ativos”. Conforme destacado por Markowitz (1952), a redução do risco no investimento não se limita exclusivamente à estratégia de diversificação; é essencial considerar também a covariância entre os ativos ao tomar decisões de investimento.

[...] ao tentar tornar a variância pequena, não basta investir em muitos títulos. É necessário evitar investir em títulos com altas covariâncias entre si. Devemos diversificar em todos os setores porque empresas em diferentes indústrias, especialmente indústrias com características econômicas diferentes, têm covariâncias menores do que as empresas dentro de uma mesma indústria (MARKOWITZ, 1952, p. 89, tradução nossa).¹

Assim, Markowitz (1952) introduziu a variância do retorno da carteira como um critério para sua seleção. A apresentação desse estudo marcou o início do que hoje é conhecida como a TMP.

O retorno esperado do portfólio $E(R_{port})$ é definido conforme a Equação (1).

$$E(R_{port}) = \sum_{i=1}^n w_i E(R_i) \quad (1)$$

Nessa formulação, w_i e $E(R_i)$ são, respectivamente, o peso e o retorno esperado para o ativo i .

Enquanto isso, o risco pode ser medido tanto pela variância quanto pelo desvio-padrão do retorno do ativo. O risco de um ativo individual é diferente de seu risco no contexto de um portfólio de investimentos (SILVA et al., 2020), diluindo-se à medida que mais ativos são adicionados, o que reduz a variância da carteira. Sendo assim, esse efeito corrobora com o princípio da diversificação, que quanto mais um portfólio for diversificado, menor será seu risco, como demonstrado na Equação (2).

$$\theta^2(c) = W_a^2 \theta_a^2 + W_b^2 \theta_b^2 + 2W_a^2 W_b^2 Cov_{a,b} \quad (2)$$

Na equação supracitada, $\theta^2(c)$ corresponde à variância da carteira W_a ; W_b ao percentual (peso) investido em cada ativo; θ_a^2 e θ_b^2 à variância de cada ativo; e $Cov_{a,b}$ é a covariância entre dois ativos.

Por sua vez, a covariância entre dois ativos é apresentada na Equação (3).

¹ No original: [...] in trying to make variance small it is not enough to invest in many securities. It is necessary to avoid investing in securities with high covariances among themselves. We should diversify across industries because firms in different industries, especially industries with different economic characteristics, have lower covariances than firms within an industry.

$$Cov_{x,y} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n-1} \quad (3)$$

Nela, $Cov_{x,y}$ corresponde à covariância entre dois ativos, X_i e Y_i são os retornos dos ativos da carteira, \bar{X} e \bar{Y} , o retorno médio de cada ativo e n é o número de ativos.

Assim, por meio dessas equações, é possível mensurar o retorno e o risco de uma carteira de ativos, possibilitando sua otimização, ou seja, obter uma carteira que proporcione a maximização do retorno para um dado um nível de risco.

Ao examinar a Equação 1, é possível inferir que o retorno dos ativos não está condicionado à relação entre eles. Dessa forma, a busca por ativos na construção do portfólio visa preservar o nível de retorno, ao mesmo tempo em que se minimiza o risco da carteira.

2.2 Teorias sobre risco e retorno de ativos

“As questões de como o risco é medido, como é recompensado e quanto risco assumir são fundamentais em cada decisão de investimento, desde a alocação de ativos até sua avaliação” (DAMODARAN, 2005, p. 25). Segundo o autor, existem três principais modelos que relacionam o *trade-off* incorrido entre risco e retorno: o *Capital Asset Pricing Model*, o *Arbitrage Pricing Model* e os modelos multifatoriais.

2.2.1 *Capital Asset Pricing Model*

O *Capital Asset Pricing Model* é um modelo financeiro utilizado para estimar o retorno esperado de um ativo com base no risco associado a esse ativo em relação ao mercado como um todo. Desenvolvido independentemente por William Sharpe (1964), John Lintner (1965) e Jan Mossin (1966), o *Capital Asset Pricing Model* é uma extensão da teoria de portfólios de Harry Markowitz e fornece uma estrutura para precificação de ativos em um ambiente de mercado eficiente. Foi concebido para abordar a questão do risco sistemático, não mitigado pela diversificação proposta pela TMP.

Partindo da premissa de uma carteira perfeitamente diversificada, na qual o risco dos ativos individuais não é recompensado, exceto pelo risco de mercado sistemático, o *Capital Asset Pricing Model* estabelece a taxa teórica de retorno de um determinado ativo em relação à carteira de mercado diversificada. Essa taxa é

derivada do equilíbrio dos ativos de risco em relação às suas covariâncias com a carteira de mercado.

O *Capital Asset Pricing Model* assume que o retorno esperado de um ativo é determinado por uma função linear de três variáveis: o beta (β), a taxa de retorno do ativo livre de risco (R_f) e o retorno da carteira de mercado (R_m), conforme demonstrado na Equação (4).

$$R(i) = R_f + \beta(R_m - R_f) \quad (4)$$

Na equação acima, o β capta parte do risco não-diversificável em relação ao mercado como um todo.

2.2.2 *Arbitrage Pricing Theory*

A *Arbitrage Pricing Theory* parte da mesma premissa que o *Capital Asset Pricing Model*, mas considera os riscos não-diversificáveis de forma mais ampla. Embora os ativos sejam tratados de forma individualizada, eles são suscetíveis a influências conjuntas de fatores macroeconômicos. Nesse contexto, Ross (1976) propõe a *Arbitrage Pricing Theory*, estabelecendo a relação entre a sensibilidade de um ativo a um componente específico de risco e o prêmio de risco associado a esse mesmo fator.

Uma das principais críticas ao modelo clássico de média-variância é a suposição de que os retornos dos ativos seguem uma distribuição normal. Além disso, os modelos de média-variância assumem que a variância da média dos retornos é uma medida adequada de risco, desconsiderando as preferências individuais dos investidores.

Buscando contornar essas limitações, Rom e Ferguson (1993) apresentam a *Post-Modern Portfolio Theory*. Nesse modelo, os autores redefinem o conceito de risco, considerando apenas os eventos adversos que estejam abaixo do limite estabelecido pelo investidor como risco relevante. Esse tipo de risco é denominado *Downside Risk*, e o retorno associado é definido como o retorno mínimo aceitável. Diferentemente da TMP, a *Post-Modern Portfolio Theory* aceita distribuições assimétricas, o que se aproxima mais da realidade.

2.2.3 Modelo de fatores

Contemporâneo à *Arbitrage Pricing Theory*, o modelo de fatores é uma abordagem analítica que busca explicar os retornos dos ativos financeiros por meio da identificação e quantificação de fatores específicos que influenciam esses retornos. Esse modelo parte do pressuposto de que os preços dos ativos são influenciados por múltiplos fatores, além do simples risco de mercado. Em vez de considerar os preços como reflexos diretos de todas as informações disponíveis, o modelo foca na relação entre os ativos e os fatores que podem explicar suas variações de retorno.

O mais difundido é o modelo de 3 fatores, proposto por Fama e French (1992). Esse modelo incorpora três fatores principais: (i) o prêmio de risco, como definido pelo *Capital Asset Pricing Model*; (ii) a diferença entre os retornos de obtidos por um portfólio composto por ações de baixa capitalização de mercado em relação a outro, com empresas de alta capitalização de mercado; (iii) a diferença entre os retornos de um portfólio com alta razão *book-to-market*² e um com baixa razão. A ideia por trás desse modelo é que empresas de diferentes tamanhos e com diferentes perfis de investimento podem ter desempenhos distintos devido a esses fatores específicos.

O modelo de três fatores de Fama e French sugere que os retornos dos ativos podem ser mais bem explicados pela combinação desses três fatores do que apenas pelo risco de mercado considerado pelo *Capital Asset Pricing Model*. Ele tem sido amplamente utilizado na pesquisa financeira e na prática de gestão de investimentos para avaliar e explicar os retornos de diferentes classes de ativos e estratégias de investimento.

2.3 Liquidez

Embora não exista um conceito dominante, a literatura descreve liquidez como a facilidade de negociar rapidamente um ativo em grande volume, convertendo-o em papel-moeda sem afetar significativamente seu preço ou o preço de mercado (HODRICK e MOULTON, 2009). Mercados líquidos permitem negociações com mínimo impacto nos preços (O'HARA, 1999), implicando em baixos custos de transação.

² *Book-to-market ratio* corresponde à divisão do valor contábil do patrimônio líquido pelo valor de mercado de uma empresa.

Desse modo, a liquidez de mercado - representada pelo volume médio de negociação - pode ser entendida como o custo de troca de um ativo em relação ao seu valor justo. Portanto, quanto maior a liquidez menor as fricções no mercado, tornando-o mais eficiente e atrativo para investidores.

Ao se definir um referencial para balizar a liquidez de um ativo, a escolha do critério deve ser baseada em princípios que garantam a liquidez suficiente. Isso assegura que os ativos selecionados tenham um nível adequado de negociação, reduzindo a volatilidade anômala e a incerteza nos preços decorrentes da baixa liquidez, sem limitar excessivamente a diversidade de ativos.

2.4 Métricas para mensuração do desempenho de um portfólio

Mensurar o desempenho de um portfólio é uma prática crucial para investidores e gestores por diversas razões. Primeiramente, essa análise permite uma avaliação detalhada das estratégias de investimento adotadas, possibilitando entender se as decisões tomadas estão gerando os retornos esperados e se estão alinhadas com os objetivos estabelecidos para o portfólio. Além disso, ao comparar o desempenho do portfólio com estratégias concorrentes, os investidores podem determinar se o portfólio está superando ou ficando aquém de determinado *benchmark*.

Segundo Oliveira Filho e Sousa (2015, p. 64), “a avaliação de performance é constituída por um conjunto de técnicas, muitas das quais originadas na Moderna Teoria de Carteiras, e está relacionada à área de riscos.” A mensuração de desempenho não se limita apenas aos retornos obtidos, mas também engloba a análise da volatilidade, a exposição a diferentes tipos de risco e a eficiência na gestão desses riscos ao longo do tempo.

Com base nessa análise mais abrangente, os investidores podem tomar decisões informadas sobre a diversificação da carteira, realocação de ativos e adoção de estratégias de proteção em cenários adversos. Em suma, mensurar o desempenho do portfólio é essencial para uma gestão financeira sólida, adaptável às mudanças do mercado e alinhada com as metas e tolerâncias ao risco dos investidores.

2.4.1 Índice Sharpe

O Índice Sharpe, desenvolvido em 1966 por William Sharpe, surgiu da necessidade de se criar uma métrica para avaliar o desempenho de portfólios de ativos. Sharpe, fundamentando sua tese nos conceitos de Markowitz, trouxe uma abordagem inovadora ao considerar não apenas a rentabilidade do portfólio, mas também a taxa livre de risco, levando em conta o risco associado à própria carteira, medido pelo desvio-padrão conforme definido por Markowitz.

Essa métrica oferece uma medida robusta do sucesso de um investimento, considerando que um rendimento que envolve riscos deve proporcionar um retorno superior à taxa livre de risco. Ele mensura o retorno extra para cada unidade de risco global assumido pelo investidor, conforme a Equação (5):

$$\text{Índice Sharpe} = \frac{(R_p - R_f)}{\sigma_p} \quad (5)$$

sendo que R_p corresponde ao retorno do portfólio, R_f à taxa livre de risco e σ_p é o desvio padrão (ou volatilidade) do portfólio.

Essencialmente, o Índice Sharpe possibilita a avaliação comparativa entre carteiras, permitindo determinar se uma carteira supera outra com base no seu valor. Valores entre 0 e 1 são considerados aceitáveis, pois indicam que o retorno supera a taxa livre de risco, embora a relação entre risco e retorno ainda não seja plenamente eficiente. No entanto, o ideal é que o índice seja superior a 1, pois esse patamar reflete uma alocação de risco mais eficiente em relação ao retorno obtido. Esta métrica torna-se, assim, uma ferramenta valiosa para os investidores na tomada de decisões informadas sobre alocação de ativos, identificando portfólios que equilibram efetivamente o retorno desejado com o risco assumido.

2.4.2 Índice de Treynor

Desenvolvida por Treynor (1965), é uma das medidas de desempenho de investimentos financeiros mais conhecidas e utilizadas. Conforme Guimarães Júnior et al. (2015, p. 88), o Índice de Treynor

utiliza premissas do modelo *Capital Asset Pricing Model* para decompor o risco em duas parcelas: risco associado à variação do mercado (risco sistemático ou não diversificável) e risco decorrente de aspectos únicos relacionados a cada ativo individualmente (risco não sistemático ou diversificável) (GUIMARÃES JÚNIOR et al., 2015, p.88).

O Índice de Treynor, portanto, é definido de acordo com a Equação (6).

$$\text{Índice de Treynor} = \frac{(R_p - R_f)}{\beta_p} \quad (6)$$

De forma similar ao Índice Sharpe, R_p representa o retorno esperado do portfólio e R_f , o retorno do ativo livre de risco. Já o β , corresponde ao risco sistemático do portfólio.

De acordo com Le Sourd (2007), devido ao fato de considerar apenas o risco não-diversificável do portfólio, este indicador é particularmente apropriado para avaliar carteiras bem diversificadas e que, conseqüentemente, já eliminaram todo o risco não sistemático.

2.4.3 Alfa de Jensen

Elaborado por Jensen (1968), o Alfa de Jensen corresponde à diferença entre o retorno observado do portfólio em excesso ao ativo livre de risco e o retorno explicado pelo *Capital Asset Pricing Model*. Conforme Le Sourd (2007, p. 15). Essa relação é expressa conforme a Equação (7):

$$E(R_p) - R_f = \alpha_p + B_p(E[R_m] - R_f) + \varepsilon_{pt} \quad (7)$$

sendo que $B_p(E[R_m] - R_f)$ corresponde ao retorno previsto pelo *Capital Asset Pricing Model*, enquanto o α_p representa a parcela de retorno adicional que pode ser creditada à estratégia de investimento.

Este método, ao contrário dos Índices Sharpe e de Treynor, não permite comparar portfólios com diferentes níveis de risco. O valor do α_p é proporcional ao nível de risco assumido, medido pelo B_p .

2.5 Hipótese de Mercados Eficientes

A HME é uma teoria fundamental na área de finanças que postula que os preços dos ativos financeiros refletem todas as informações disponíveis de forma eficiente e instantânea. Segundo Damodaran (2005), seria impossível para

investidores obterem retornos consistentemente acima da média no mercado, uma vez que todas as informações relevantes já estão incorporadas aos preços dos ativos.

Fama (1970) define três formas de eficiência de mercado, conforme Damodaran (2005):

- Forma fraca: os preços correntes refletem todas as informações contidas em preços passados, o que implica que gráficos e análises técnicas que utilizam preços passados são ineficazes para descobrir ações subvalorizadas.
- Forma semi-forte: o preço corrente reflete não apenas as informações contidas nos preços passados, mas todas as informações publicamente disponíveis. Desse modo, nenhuma abordagem baseada em utilizar e tratar estas informações seria útil para descobrir ações subvalorizadas.
- Forma forte: os preços correntes refletem todas as informações, tanto públicas quanto privadas, tornando impossível para os investidores obterem consistentemente lucros extraordinários com base nessas informações.

2.6 Momento

“Momento é um conjunto de estratégias de investimentos nas quais retornos passados podem prever retornos futuros” (JEGADEESH e TITMAN, 1993 *apud* ABREU, 2021, p. 12). Entende-se por momento a taxa de variação de um ativo por unidade de tempo. Matematicamente, pode ser descrito conforme a Equação (8), apresentada a seguir.

$$\text{Momento} = \left(\frac{P_{i+n} - P_i}{n} \right) \quad (8)$$

Nela, P_i é o preço de um ativo financeiro no i -ésimo instante e n o número de unidades de tempo escolhido pelo usuário.

Essa abordagem se baseia na tendência de que ativos com boa performance continuem a apresentar bons resultados no futuro, enquanto ativos com desempenho ruim tendem a se manter dessa forma. A estratégia de momento capitaliza sobre a inércia das tendências de preços de ativos, explorando a persistência de retornos positivos ou negativos ao longo de um determinado período.

Utilizando essa estratégia, os estudos de Rouwenhorst (1999), realizado no mercado Europeu e países emergentes, e de Chan et. al (2006), no mercado norte-

americano, chegaram a resultados similares aos de Jegadeesh e Titman (1993). Rejeitando a proposta da HME em sua forma fraca, ambos os trabalhos encontraram evidências de momentos nestes mercados.

2.7 Aprendizado de máquina

O aprendizado de máquina consiste em programar computadores de modo a otimizar um determinado parâmetro, com base em amostras de dados e informações passadas (ALPAYDIN, 2014). A partir do treinamento desses parâmetros, é possível utilizar o modelo para realizar inferências e prever comportamentos futuros de determinada variável.

Os sistemas de aprendizado de máquina podem ser classificados de acordo com o nível e o tipo de supervisão que eles recebem durante a etapa de treinamento. Existem quatro categorias principais: aprendizado supervisionado, aprendizado não supervisionado, aprendizado semi-supervisionado e aprendizado por reforço (GÉRON, 2019).

2.7.1 Aplicações empíricas de aprendizado de máquina no mercado de ações

De acordo com Kumar e Thenmozhi (2006), a previsão do mercado financeiro apresenta grande complexidade devido à natureza das séries temporais financeiras, que são ruidosas, não estacionárias e caóticas em termos determinísticos. Fatores inesperados, como eventos econômicos, políticos, desastres e guerras, assim como as expectativas dos investidores, podem influenciar significativamente uma série financeira, como índices de mercado de ações e taxas de câmbio.

Ademais, a relação dessas séries com outros dados financeiros pode se alterar ao longo do tempo, o que torna a previsão dos movimentos de mercado extremamente desafiadora. Nesse contexto, métodos de aprendizado de máquina mostram-se eficazes ao evitar a modelagem detalhada da estrutura dos dados, concentrando-se em identificar padrões e relações que ajudam a fazer previsões mais precisas.

Apesar de ainda muito incipiente no Brasil, estudos realizados com o intuito de se comparar a performance de modelos de aprendizado de máquina na predição do comportamento dos mercados financeiros ganharam bastante notoriedade em âmbito internacional nas últimas décadas. Nesse interim, foram conduzidas pesquisas não só

em mercados desenvolvidos como o norte-americano e o europeu, mas também em países em desenvolvimento, como China e Índia.

Kumar e Thenmozhi (2006) realizaram um estudo cujo objetivo era prever a direção do índice de ações do mercado indiano, o S&P CNX NIFTY. Para isso, os autores testaram os 4 diferentes algoritmos de aprendizado de máquina e um modelo de rede neural. Os modelos que melhor performaram foram os de *Support Vector Machine* e *Random Forest*, ambos superando o algoritmo de rede neural.

Haugland (2023), em sua tese de mestrado, realizou uma análise comparativa de diversos modelos de aprendizado de máquina para prever o preço de 4 ações negociadas no mercado norueguês. Como resultado, concluiu que o modelo de *Random Forest* foi o que apresentou a melhor performance ao prever o preço de fechamento ajustado dentre os modelos de aprendizado de máquina testados. Seu desempenho, todavia, foi superado pelo modelo LSTM, que utiliza redes neurais, mecanismo mais sofisticado se comparado ao aprendizado de máquina.

Determinados trabalhos se aprofundaram além da predição de um número restrito de ativos ou índices de mercado. Esses estudos se propuseram a aplicar mecanismos de aprendizado de máquina visando não só prever o comportamento de determinados ativos financeiros, mas também para a seleção de carteiras de investimentos.

Tan et al. (2019) avaliaram a robustez do modelo de *Random Forest* no contexto da seleção de ações negociadas no índice chinês CSI 500. Combinando indicadores fundamentalistas e técnicos, os autores treinaram um modelo capaz de atribuir probabilidades a cada ativo, de acordo sua capacidade de alcançar retornos excedentes positivos. Com essa base, as 20 ações com as maiores probabilidades eram selecionadas e mantidas por um determinado período, havendo a ocorrência de rebalanceamentos periódicos. Como resultado, o portfólio construído através do algoritmo apresentou um retorno médio diário 3 vezes superior, se comparado ao retorno médio do índice de referência do mercado chinês.

Ma et al. (2021), analisaram o desempenho de modelos de otimização de portfólios cuja previsão dos retornos provinham de modelos de aprendizado de máquina. Para tal, os autores utilizaram dados entre 2007 e 2015 das ações pertencentes ao Índice *China Securities* 100. Para projeção de retornos futuros, os otimizadores utilizaram apenas os dados de retorno históricos dos ativos em questão. Entre os modelos avaliados no estudo, a combinação do algoritmo de *Random Forest*

junto ao otimizador de média-variância foi a que apresentou o melhor resultado entre as estratégias comparadas.

Por sua vez, com base num universo de ativos presentes no índice S&P 500 entre janeiro de 1999 e março de 2021, Wolff e Echterling (2023) treinaram uma série de modelos de aprendizado de máquina com base em indicadores fundamentalistas e técnicos dos papéis. Os resultados encontrados mostraram um desempenho ajustado ao risco substancial e significativo dos modelos de seleção de ações baseados em aprendizado de máquina em comparação ao retorno do índice norte-americano no período. Resultados análogos foram encontrados nos testes de robustez com ativos pertencentes ao índice STOXX Europe 600.

No Brasil, Costa (2022) realizou a otimização de portfólios compostos por 10 ETFs que reproduziam os retornos diários dos índices dos mercados acionários de países emergentes. Foram testados os algoritmos de *Random Forest*, *Support Vector Regression* e *K-Nearest Neighbor* durante uma janela de 8 anos. Como resultado, o autor encontrou que as técnicas de pré-seleção aliadas à otimização de portfólios via média-variância foram a que melhor desempenharam no que diz respeito ao retorno e Índice Sharpe obtidos.

2.7.2 *Random Forest*

Apresentado por Breiman (2001), o modelo de *Random Forest* é uma técnica de aprendizado de máquina que pertence à categoria de métodos de *ensemble*, que combinam múltiplos modelos para melhorar a precisão das previsões. A ideia por trás do algoritmo é criar várias árvores de decisão durante o processo de treinamento, onde cada árvore é construída com uma amostra aleatória dos dados de treinamento e características (*features*) selecionadas aleatoriamente em cada nó. Durante a fase de previsão, as estimativas individuais de cada árvore no *Random Forest* são combinadas de forma ponderada para gerar um resultado mais preciso e robusto.

Esse algoritmo se enquadra na categoria de aprendizado supervisionado. Segundo Haugland (2023, p. 10),

quando se aprende sob supervisão, o modelo recebe dados de entrada que podem incluir uma variedade de métricas financeiras, como preços de ações, volumes de negociação, variáveis econômicas subjacentes e indicadores

técnicos, um dado de saída correspondente ou valor-alvo descrevendo a previsão desejada ou resultado está presente para cada ponto de dados de entrada. No caso de previsões de ações, por exemplo, os dados de entrada podem ser dados de preços passados, e o de saída poderia ser o movimento de preço futuro antecipado (HAUGLAND, 2023, p. 10, tradução nossa).³

Conforme Hastie et al. (2009, p. 588), o algoritmo de *Random Forest* é definido pela Equação (9), disposta a seguir:

$$\hat{f}_{rf}^B(x) = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (9)$$

Nessa formulação, $\hat{f}_{rf}^B(x)$ corresponde ao valor previsto para um determinado dado de entrada x . Cada árvore de decisão individual T_b no *Random Forest* contribui com a previsão $T_b(x)$. As previsões de cada árvore de decisão são combinadas e, então, ponderadas pelo número de árvores B .

O algoritmo de *Random Forest* é conhecido por sua capacidade de lidar bem com dados de alta dimensionalidade, variáveis correlacionadas e presença de *outliers*. Ele também é resistente a *overfitting*, devido à aleatoriedade introduzida na construção das árvores. Essas características fazem do *Random Forest* uma escolha popular para muitas aplicações de aprendizado de máquina, especialmente em problemas complexos onde a precisão é crucial.

³ No original: When learning under supervision, the model is given input data that can include a variety of financial metrics, such as stock prices, trade volumes, underlying economic variables, and technical indicators. A corresponding output label or target value describing the desired prediction or result is present for each input data point. In the case of stock predictions, the input could for example be past price data, and the output label could be the anticipated future price movement.

3. METODOLOGIA

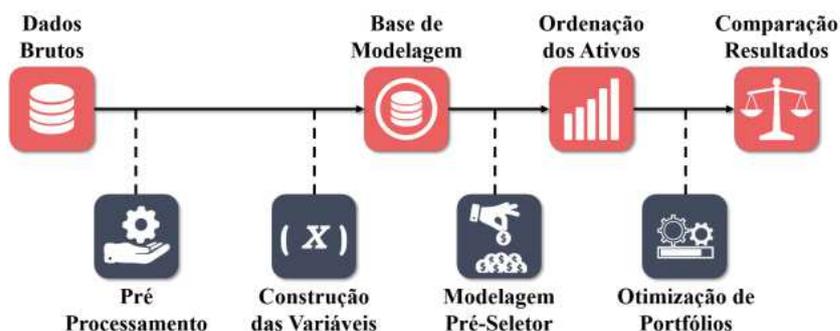
Este trabalho tem como objetivo realizar um estudo sobre a utilização de modelos computacionais e matemáticos para a identificação de portfólios de ativos. A natureza da pesquisa é caracterizada como aplicada, pois objetiva gerar conhecimentos para aplicação prática, direcionados à solução de problemas específicos.

No que diz respeito à abordagem adotada, a pesquisa é quantitativa. De acordo com Kauark et al. (2010), esse tipo de pesquisa concentra-se no uso de métodos e técnicas estatísticas para analisar e classificar informações, utilizando a grande quantidade de dados disponível para identificar e relacionar causas e efeitos.

Conforme mencionado na Seção 1, o vigente trabalho consiste na utilização da metodologia aplicada por Abreu (2021), com uma modificação do período de análise dos dados, bem como a comparação da performance desse portfólio com carteiras classificadas através do *Capital Asset Pricing Model*, realizada através das métricas de performance Índice Sharpe, Índice de Treynor e Alfa de Jensen.

Para alcançar esse resultado, as etapas necessárias são dispostas conforme apresentado na Figura 1:

Figura 1 - Esquemática do processo de otimização de portfólios com pré-seletor de ativos



Fonte: Abreu (2021).

Inicialmente, deve-se obter uma série temporal de preços de ativos financeiros – definidos a partir de um critério de liquidez – abrangendo um período suficientemente longo para permitir todas as avaliações necessárias.

Posteriormente, essas séries devem ser transformadas em log-retornos. Computar os retornos logarítmicos dos ativos financeiros ao realizar otimizações de portfólio é uma prática comum por diversas razões técnicas e teóricas. Entre suas

principais vantagens, cabem citar que os retornos são aditivos ao longo do tempo, o que simplifica a análise de desempenho e facilita o cálculo de retornos acumulados, além de sua tendência a ser mais próximos de uma distribuição normal do que os retornos simples. Isso é especialmente útil porque muitos métodos estatísticos e modelos financeiros, incluindo o modelo de média-variância, assumem que os retornos dos ativos são normalmente distribuídos.

Depois, deve-se criar uma variável de resposta binária com base no log-retorno acumulado futuro das séries temporais dos ativos financeiros (1 para positivo e 0 para negativo), bem como variáveis históricas com potencial explicativo da variável resposta elaborada, com ênfase em variáveis de momento.

Com a base de dados pronta e tratada, ela deve ser dividida em três conjuntos: treino, teste e validação. Segundo Abreu (2021), a proporção de cada conjunto pode ser flexível, mas é recomendado destinar pelo menos 20% dos dados para a validação. Utilizando as variáveis explicativas e os conjuntos de dados de treino e teste, deve-se treinar o modelo de aprendizado de máquina *Random Forest* para prever a variável resposta.

Em seguida, é necessário estruturar a base de validação em intervalos segmentados, levando em conta o tempo requerido para criar as variáveis explicativas e de resposta. O término de cada intervalo será considerado o momento de recalibração do portfólio. Dessa forma, parte do período anterior será incluído na construção das variáveis explicativas para cada intervalo subsequente. Cada segmento de tempo terá um registro para cada ativo financeiro, contendo todas as variáveis necessárias para a aplicação do modelo de aprendizado de máquina.

Feito isso, chega-se ao momento da validação dos resultados. Nesta fase, o modelo desenvolvido é aplicado aos intervalos da base de validação. As probabilidades extraídas do modelo são usadas para classificar os ativos financeiros em ordem decrescente, do melhor para o pior. Com essa classificação, serão construídas carteiras, expandindo gradualmente o número de ativos elegíveis para o otimizador média-variância, de acordo com o tamanho de cada portfólio, seguindo a ordem obtida na etapa anterior.

Por fim, para cada período, devem ser calculados o log-retorno, o Índice Sharpe, o Índice de Treynor e o Alfa de Jensen dos portfólios otimizados conforme a classificação realizada pelo algoritmo. Além disso, os resultados devem ser

comparados com a classificação obtida por meio do *Capital Asset Pricing Model*, a fim de avaliar a efetividade do modelo.

4. BASE DE DADOS

Essa Seção se dedica à exposição do processo de coleta das bases de dados a serem utilizadas no experimento, bem como a explicação dos tratamentos realizados de modo a torná-la propícia a uso. Ao final desta etapa, o intuito é gerar três bases de dados como resultado (treino, teste e validação).

Os conjuntos de treino e teste serão utilizados para construir um modelo capaz de prever quais ativos terão um retorno acumulado positivo nos 20 dias úteis seguintes. Por sua vez, o conjunto de validação será usado para realização do *backtesting*, comparando as carteiras de investimentos construídas com e sem a influência do algoritmo pré-seletor de ativos.

Em suma, cada base criada contará com as seguintes variáveis, conforme disposto na Tabela 1.

Tabela 1 – Descrição das variáveis utilizadas no modelo de aprendizado de máquina

| Nome da variável | Descrição |
|-----------------------|---|
| lag_0 | Log-retorno correspondente ao dia de referência (d0). |
| lag_1 | Log-Retorno em d-1. |
| lag_2 | Log-Retorno em d-2. |
| lag_3 | Log-Retorno em d-3. |
| lag_4 | Log-Retorno em d-4. |
| lag_5 | Log-Retorno em d-5. |
| lag_6 | Log-Retorno em d-6. |
| lag_7 | Log-Retorno em d-7. |
| lag_8 | Log-Retorno em d-8. |
| lag_9 | Log-Retorno em d-9. |
| lag_10 | Log-Retorno em d-10. |
| liquidez | Total do Volume comercializado entre d0 e d-10. |
| media_logretorno | Média dos Log-Retornos entre d0 e d-10. |
| std_logretorno | Desvio Padrão dos Log-Retornos entre d0 e d-10. |
| ID | Valor do ID (Momentum Absoluto Quantitativo) dos Log-Retornos entre d0 e d-10. |
| setor_comunicacoes | Variável categórica com valor 1 caso o ativo pertença ao setor de Comunicações. Caso contrário, seu valor é 0. |
| setor_consumo_ciclico | Variável categórica com valor 1 caso o ativo pertença ao setor de consumo cíclico. Caso contrário, seu valor é 0. |

| | |
|------------------------------------|---|
| setor_consumo_nao_ciclico | Variável categórica com valor 1 caso o ativo pertença ao setor de consumo não cíclico. Caso contrário, seu valor é 0. |
| setor_financeiro | Variável categórica com valor 1 caso o ativo pertença ao setor financeiro. Caso contrário, seu valor é 0. |
| setor_materiais_basicos | Variável categórica com valor 1 caso o ativo pertença ao setor de materiais básicos. Caso contrário, seu valor é 0. |
| setor_petróleo_gás_biocombustíveis | Variável categórica com valor 1 caso o ativo pertença ao setor de petróleo, gás e biocombustíveis. Caso contrário, seu valor é 0. |
| setor_saude | Variável categórica com valor 1 caso o ativo pertença ao setor de saúde. Caso contrário, seu valor é 0. |
| setor_ti | Variável categórica com valor 1 caso o ativo pertença ao setor de tecnologia da informação. Caso contrário, seu valor é 0. |
| setor_utilidade_publica | Variável categórica com valor 1 caso o ativo pertença ao setor de utilidade pública. Caso contrário, seu valor é 0. |
| target | Variável resposta. É uma variável categórica com valor 1 caso o Log-Retorno acumulado dos próximos 20 dias for positivo. Caso contrário, seu valor é 0. |

Fonte: elaborado pelo autor (2025).

Todo o processo de coleta e manipulação de dados, desde a fase de pré-seleção dos ativos até a otimização dos portfólios, foram feitas por meio da utilização da linguagem de programação *python 3.10.12* junto à plataforma *Google Colab*. As bibliotecas empregadas nesse processo estão listadas na Tabela 2.

Tabela 2 - Bibliotecas utilizadas para coleta e manipulação dos dados

| Biblioteca | Versão |
|-------------------|---------------|
| pandas | 2.2.2 |
| numpy | 1.26.4 |
| yfinance | 0.2.51 |
| scikit-learn | 1.6 |
| seaborn | 0.13.2 |

Fonte: elaborado pelo autor (2025).

4.1 Coleta dos dados

Através da biblioteca *yfinance*, realizou-se a obtenção dos dados históricos do preço de fechamento ajustado e do volume diário de negociação dos papéis que eram

transacionados no mercado brasileiro entre janeiro de 2010 e dezembro de 2019, totalizando 10 anos e 2.480 observações para cada um dos 326 ativos negociados.

A escolha dessa janela temporal para a coleta de dados é fundamentada em duas razões principais. Primeiro, iniciar a análise em 2010 permite capturar um momento em que o mercado brasileiro já apresentava um número significativo de empresas listadas, garantindo uma amostra representativa. Um período anterior a 2010 resultaria em uma limitação nas possibilidades de diversificação entre os ativos disponíveis, o que poderia comprometer a eficácia do otimizador.

Em segundo lugar, optou-se por excluir os dados de 2020, um ano caracterizado por intensa volatilidade nos mercados devido à pandemia de COVID-19, para garantir que a etapa de validação não fosse afetada por condições excepcionais. A escolha de um período mais estável, como o biênio 2018-2019, permite avaliar a eficácia do modelo em um contexto de mercado menos atípico, contribuindo para a validade externa da estratégia proposta por Abreu (2021). Dessa forma, é possível validar o modelo em um cenário com menor influência de eventos extraordinários, proporcionando uma análise mais precisa e confiável de sua aplicação à otimização de portfólios no mercado brasileiro.

4.1.1 Tratamento e preparação dos dados

Antes de prosseguir para a segmentação dos dados, duas questões devem ser consideradas: (i) conforme Wolff e Echterling (2023), alguns papéis possuem liquidez insuficiente, de modo que seus retornos podem apresentar considerável volatilidade de um período para o outro, decorrentes dessa ausência de negociação e não de sua performance em si; e (ii) determinadas empresas negociam mais de um tipo papel - como no caso das empresas que possuem ações ordinárias e preferenciais -, de modo que desconsiderar esse fato poderia levar o algoritmo a adicionar mais de um papel de uma única companhia, gerando superconcentração em um único ativo cujo modelo não seria capaz de diferenciar, eliminando parte considerável dos benefícios da diversificação e comprometendo a performance do modelo.

Para resolver a primeira questão, foram removidos da base de dados todos os papéis que apresentaram um volume médio de negociação diário ao longo de 2019 inferior a R\$1.000.000,00 (um milhão de reais). Por fim, para admitir um único papel por empresa, foram removidos *tickers* duplicados, de modo que apenas o papel com

maior liquidez (todos correspondentes a ações das classes PN ou PNB) de cada companhia fosse considerado. Feitos esses ajustes, restaram 76 papéis para análise.

Outra questão que emerge nesse cenário é que nem todos os ativos presentes na base de dados de 2019 estavam listados no ano 2010. Portanto, ativos que não eram listados em janeiro de 2010 também foram desconsiderados da análise, totalizando 71 ativos, base suficiente o bastante para que o otimizador consiga construir carteiras otimizadas e eficientes.

Por razões desconhecidas, o papel B3SA3 não possuía cotação nos dias 24/02/2011 e 09/03/2011. Para corrigir o problema de dados faltantes, esses dois dias foram preenchidos com a preço de fechamento ajustado do dia de negociação imediatamente posterior às datas, ou seja, 25/02/2011 e 10/03/2011, respectivamente, sem nenhum prejuízo significativo ao otimizador.

Por fim, como o objetivo do projeto é utilizar o retorno acumulado mensal (utilizado como 20 dias úteis de negociação como *proxy*), a transformação do preço de fechamento diário para retornos logaritmos se faz necessária. O log-retorno entre os dias t e $t + 1$ pode ser definido conforme a Equação (10).

$$R_i = \ln \left(\frac{P_{t+1}}{P_t} \right) \quad (10)$$

Sendo que R_i representa a variação logarítmica do preço do ativo i entre t e $t+1$.

4.1.2 Setor

Com base na classificação setorial da B3, listagem que consta o setor econômico correspondente a cada uma das empresas negociadas no mercado brasileiro, criou-se um *dataframe*, denominação dada a tabelas criadas através da biblioteca *pandas*, contendo o *ticker* de cada uma das 71 empresas e seu respectivo setor de atuação.

4.2 Construção das variáveis para as etapas de treino e teste

Para garantir o sucesso do modelo, é fundamental estruturar corretamente os dados antes de dividir os conjuntos de treino, teste e validação. Este processo envolve

a criação e organização das variáveis necessárias ao modelo, assegurando que cada uma delas seja consistente em todas as etapas da análise.

Com isso em mente, as variáveis foram inicialmente criadas para toda a base de dados, a fim de garantir essa padronização entre as bases utilizadas pelo algoritmo. Somente após a criação completa das variáveis, os dados foram separados nos conjuntos acima mencionados, permitindo uma divisão consistente e representativa para o modelo.

A partir da base de dados tratada, contendo o log-retorno diário para cada um dos 71 ativos em colunas distintas, foi realizada uma reorganização, unindo essas informações com dados adicionais sobre o setor de atuação e o volume diário de negociação de cada ativo, mencionados na Subseção 4.1.

Dessa forma, em vez de uma coluna para cada ativo, passa-se a ter 71 entradas para cada dia de negociação, cada uma representando um ativo específico. Com essa reestruturação, o *dataframe* passa a ter cinco colunas: "Data", "Ticker", "Log-retorno", "Volume de negociação" e "Setor". Essa modificação facilita o cálculo das demais variáveis necessárias para o modelo, além de proporcionar uma estrutura mais adequada para as análises subsequentes.

Com essa reformulação, foi possível adicionar cada variável criada como uma nova coluna ao *dataframe*, permitindo uma organização mais clara dos dados. Assim, ao final do processo, restou apenas realizar a separação final dos dados em conjuntos de treino, validação e teste.

Em projetos de aprendizado de máquina, é muito comum a utilização da função *train_test_split*, presente no pacote *scikit-learn* para a separação dos dados em conjuntos de treino e teste. Esse método permite selecionar aleatoriamente uma proporção pré-estabelecida dos dados para cada conjunto.

Acontece que, no caso de dados de séries temporais, a ordem dos dados é muito importante. Cada ponto de dados depende dos anteriores, e a sequência temporal precisa ser respeitada para preservar a autocorrelação e o padrão da série ao longo do tempo. Dividir aleatoriamente uma série temporal pode levar a um "vazamento de dados", onde informações futuras, contidas no conjunto de teste, "vazam" para o treinamento, comprometendo a avaliação e a capacidade do modelo de generalizar. Em séries temporais, o modelo deve ser treinado em dados passados para prever dados futuros, então é essencial que o conjunto de treinamento contenha apenas os dados que ocorreram antes dos dados no conjunto de teste.

De modo a manter a ordenação original do *dataframe*, a separação dos dados nos três conjuntos necessários foi feita com base na delimitação das datas de corte, mantendo uma proporção de 60%, 20% e 20% para treino, teste e validação, respectivamente. A distribuição dos períodos delimitados, assim como o número total de registros de cada, pode ser observada na Tabela 3.

Tabela 3 - Divisão dos dados históricos e número de registros por etapa

| Etapa | Início | Fim | Número de Registros |
|--------------|---------------|------------|----------------------------|
| Treino | 04/01/2010 | 30/12/2015 | 1483 |
| Teste | 04/01/2016 | 29/12/2017 | 503 |
| Validação | 02/01/2018 | 30/12/2019 | 494 |

Fonte: elaborado pelo autor (2025).

4.2.1 Variáveis explicativas

No modelo de *Random Forest*, a inclusão de variáveis explicativas é crucial para melhorar a previsão do log-retorno acumulado, pois essas variáveis capturam diferentes aspectos da qualidade dos ativos financeiros. Com elas, o modelo pode avaliar múltiplos fatores que influenciam o desempenho dos ativos, permitindo não apenas prever, mas também classificar os ativos conforme a probabilidade de obterem um log-retorno acumulado positivo nos próximos 20 dias.

Com base nos dados já disponíveis no *dataframe*, primeiramente foi criada uma variável que representa o volume total negociado entre o dia atual e os últimos 10 dias, servindo como uma *proxy* de liquidez dos ativos.

Para que o modelo pudesse incorporar informações sobre o setor de atuação de cada empresa, foi necessário converter esses dados categóricos, representados em texto, em variáveis numéricas. Desse modo, utilizou-se de variáveis *dummy*, permitindo que as diferenças setoriais fossem incorporadas ao processo de decisão do modelo.

Seguindo a abordagem econométrica tradicional, a inclusão dos *lags* dos log-retornos das séries temporais foi uma das primeiras variáveis explicativas adicionadas ao modelo. Nessa implementação, foram utilizados 10 *lags* para cada ativo, possibilitando capturar informações sobre a evolução recente de seus retornos.

Para complementar essa análise temporal, foram calculadas a média e o desvio padrão a partir dos log-retornos defasados, capazes de fornecer uma visão geral da tendência e volatilidade histórica dos ativos.

Por fim, foi construída uma variável de momento que avalia a direção e a consistência dos retornos acumulados de um ativo nos últimos 10 dias. Essa variável calcula o retorno acumulado do período, identifica seu sinal (positivo ou negativo) e mede a proporção de dias com retornos positivos e negativos. O sinal do retorno acumulado é então combinado com a diferença entre essas proporções, o que oferece uma indicação da direção predominante e da consistência da tendência recente do ativo.

4.2.2 Variável resposta

Neste projeto, a seleção dos melhores ativos é realizada por meio de uma classificação que utiliza como variável resposta o log-retorno acumulado dos ativos dentro de um período previamente estabelecido. Esse intervalo de análise pode variar, mas, para o propósito dessa pesquisa, foi definido como 20 dias, representando aproximadamente um mês de dias úteis.

Como o objetivo do estudo é desenvolver portfólios que sejam reavaliados e rebalanceados periodicamente, optou-se, assim, por prever o log-retorno acumulado ao longo dos 20 dias, em vez de fazer previsões diárias de preços, o que reduz potenciais erros de modelagem.

Para simplificar ainda mais a tarefa do otimizador, foi aplicado um tratamento adicional à variável de resposta, R_{acum} , inicialmente numérica. Esta foi convertida em uma variável *dummy* com base em um limite mínimo estabelecido: se R_{acum} for positivo, a classificação atribuída é 1; caso contrário, 0.

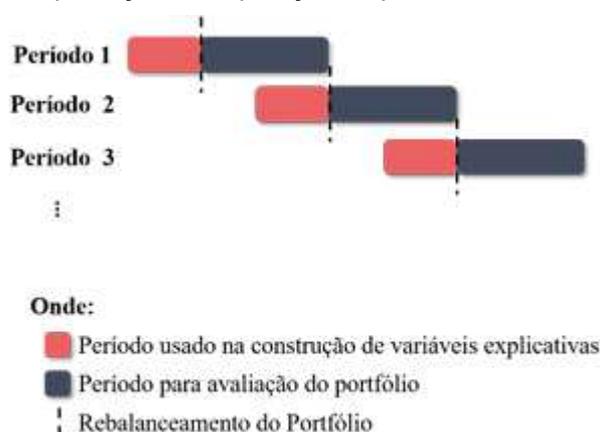
O tema central para a compreensão deste estudo centra-se na seguinte etapa: para cada janela de 20 dias, o algoritmo calcula a probabilidade de cada ativo financeiro apresentar um log-retorno acumulado positivo nos próximos 20 dias. Com essas probabilidades, os ativos são ordenados, posicionando aqueles com maior probabilidade de retorno positivo no topo da lista.

4.3 Tratamento para a base de validação

A base de validação recebeu um tratamento específico, diferente das bases de treino e teste. Com o intuito de avaliar a performance dos portfólios de investimentos criados pelo pré-seletor de ativos, o período reservado para validação foi segmentado em blocos de 31 dias. Esse intervalo contempla o tempo necessário para incluir os 10 *lags*, a data de referência e mais 20 dias, destinados à variável resposta.

Nessa configuração, considera-se que o portfólio será rebalanceado na data de referência com base nas previsões do pré-seletor e, em seguida, mantido ao longo dos 20 dias subsequentes, conforme ilustrado pela Figura 2.

Figura 2 - Exemplificação da separação de períodos na base de validação



Fonte: Abreu (2021).

Para a modelagem da base de validação, foi criada uma função específica para essa etapa, que é aplicada de forma independente a cada um dos 24 intervalos de validação. Os períodos considerados em cada intervalo estão apresentados na Tabela 4.

4.4 Classificação das ações

Com as bases de modelagem já definidas, o próximo estágio envolveu a estruturação de um fluxo de trabalho para construir o pré-seletor de ativos. O processo começou pela divisão das bases de treino e teste, separando as variáveis explicativas (X) – composta pela série de variáveis mencionadas na Subseção 4.2.1 – da variável de resposta (y) – variável *dummy* com valor igual a 1 quando o retorno mensal acumulado é positivo. As dimensões dessas bases encontram-se detalhadas na Tabela 5.

Tabela 4 - Períodos de validação do otimizador

| Período | Início | Fim |
|----------------|---------------|------------|
| 1 | 02/01/2018 | 14/02/2018 |
| 2 | 30/01/2018 | 14/03/2018 |
| 3 | 27/02/2018 | 11/04/2018 |
| 4 | 27/03/2018 | 10/05/2018 |
| 5 | 24/04/2018 | 07/06/2018 |
| 6 | 23/05/2018 | 05/07/2018 |
| 7 | 20/06/2018 | 02/08/2018 |
| 8 | 18/07/2018 | 30/08/2018 |
| 9 | 15/08/2018 | 28/09/2018 |
| 10 | 13/09/2018 | 29/10/2018 |
| 11 | 11/10/2018 | 28/11/2018 |
| 12 | 12/11/2018 | 27/12/2018 |
| 13 | 11/12/2018 | 24/01/2019 |
| 14 | 10/01/2019 | 22/02/2019 |
| 15 | 07/02/2019 | 22/03/2019 |
| 16 | 07/03/2019 | 18/04/2019 |
| 17 | 04/04/2019 | 20/05/2019 |
| 18 | 03/05/2019 | 17/06/2019 |
| 19 | 31/05/2019 | 15/07/2019 |
| 20 | 28/06/2019 | 12/08/2019 |
| 21 | 26/07/2019 | 09/09/2019 |
| 22 | 23/08/2019 | 07/10/2019 |
| 23 | 20/09/2019 | 04/11/2019 |
| 24 | 18/10/2019 | 03/12/2019 |

Fonte: elaborado pelo autor (2025).

Tabela 5 - Dimensão das bases de treino e teste

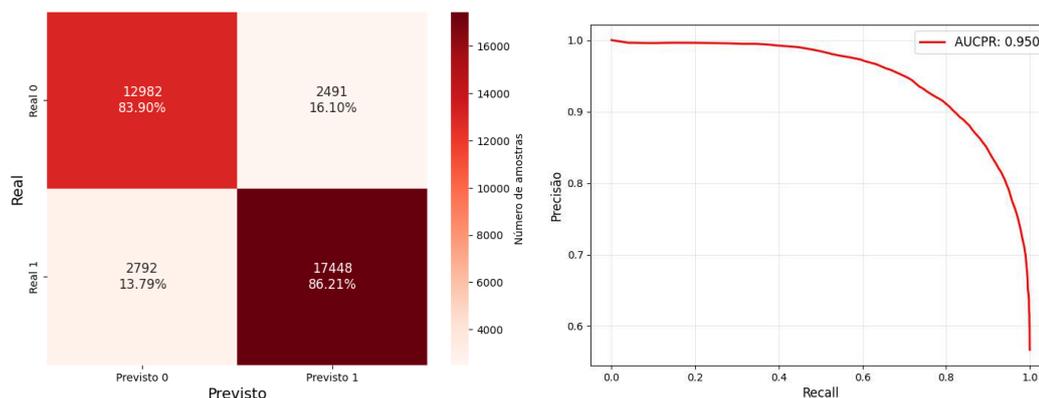
| Dataset | Linhas | Colunas |
|----------------|---------------|----------------|
| X_treino | 103.873 | 24 |
| y_treino | 103.873 | 1 |
| X_teste | 35.713 | 24 |
| y_teste | 35.713 | 1 |

Fonte: elaborado pelo autor (2025).

O algoritmo de *Random Forest* foi aplicado através da versão 1.6 da biblioteca *scikit-learn*. Para a avaliação do modelo, foi utilizada a matriz de confusão, que mede o desempenho do algoritmo ao contabilizar verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos. No entanto, por limitar-se a um único limiar de decisão, essa abordagem foi complementada pela análise das curvas de precisão e *recall*. Essas curvas permitem avaliar como o modelo se comporta em

diferentes limiares, com o *recall* destacando a capacidade do modelo de identificar corretamente os exemplos positivos, essencial em cenários onde minimizar falsos negativos é prioritário. Os resultados da matriz de confusão e da análise das curvas de Precisão e Recall são dispostos na Figura 2.

Figura 3 - Resultados da avaliação do modelo de *Random Forest*



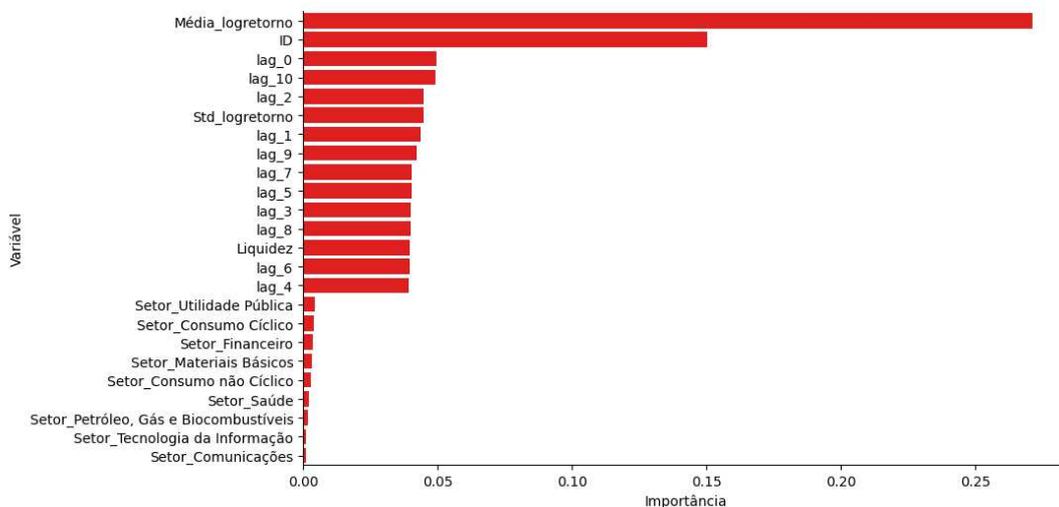
(a) Matriz de Confusão

(b) Curva de Precisão-Recall

Fonte: elaborado pelo autor (2025).

Por fim, realizou-se a análise de importância. A análise de importância de variáveis é uma técnica utilizada em modelos de aprendizado de máquina para identificar quais variáveis possuem maior influência na previsão do modelo. Os resultados obtidos são podem ser observados na Figura 4.

Figura 4 - Importância das variáveis do modelo de *Random Forest*



Fonte: elaborado pelo autor (2025).

Esses resultados indicam que as características individuais dos ativos, especialmente o log-retorno médio e as variáveis defasadas do log-retorno, exercem a maior influência sobre as previsões do modelo. A alta relevância dessas variáveis sugere que o desempenho passado dos ativos contém informações significativas para a classificação. Além disso, o desvio padrão e a liquidez também apresentam impacto moderado, evidenciando que a oscilação dos retornos e a facilidade de negociação dos ativos desempenham um papel relevante.

Por outro lado, as variáveis setoriais possuem importância significativamente menor, indicando que a classificação por setor econômico não contribui de forma substancial para as previsões do modelo. Esses resultados reforçam a ideia de que a dinâmica de precificação dos ativos analisados é mais influenciada por características individuais e padrões históricos de retorno do que por fatores setoriais.

4.5 Avaliação e comparação de portfólios

Para validar a pré-seleção de ativos criado na Subseção 4.4, foi utilizado o modelo de média-variância. Esse modelo utiliza o retorno e a volatilidade esperados anualizados, calculados com base nos 10 dias anteriores ao rebalanceamento de cada período, para definir os pesos de cada ativo no portfólio, com o objetivo de otimizar o índice Sharpe. Cabe destacar que a soma dos pesos dos ativos é sempre igual a 1 (ou 100%), o que implica que o investidor direciona todo o capital disponível para ativos de risco, sem recorrer a alavancagem.

Para evitar a concentração excessiva em poucos ativos e preservar os benefícios da diversificação, foram incorporadas ao otimizador de média-variância duas restrições adicionais em relação ao modelo convencional: (i) a alocação máxima por ativo é ajustada dinamicamente de acordo com o número de ativos disponíveis, limitando a exposição individual e promovendo uma distribuição mais equilibrada; (ii) cada ativo na carteira deve possuir um peso mínimo de 0,5%, garantindo que todos os ativos tenham representatividade nas carteiras geradas.

O primeiro passo para validar os resultados foi aplicar o pré-seleção aos diferentes períodos definidos para *backtest*, conforme listados na Tabela 4. Em cada período de validação, os ativos foram classificados em ordem decrescente conforme os resultados do pré-seleção, que estima a probabilidade de o ativo apresentar um log-retorno positivo nos próximos 20 dias.

Para avaliar o impacto do pré-seletor, adotou-se a seguinte abordagem: em cada um dos 24 períodos, os ativos foram progressivamente adicionados ao universo de escolhas do otimizador de portfólio, com as métricas de desempenho sendo calculadas para cada portfólio gerado. O primeiro portfólio foi composto exclusivamente pelo ativo com a maior pontuação atribuída pelo modelo; o segundo incluiu os dois ativos mais bem classificados, e assim sucessivamente.

4.6 Estruturação da classificação de via *Capital Asset Pricing Model*

Conforme o guia técnico do Ministério da Fazenda (2018), os principais parâmetros do *Capital Asset Pricing Model* podem ser obtidos através de fontes específicas de mercado.

A taxa livre de risco, comumente baseada nos títulos de dívida pública de longo prazo dos Estados Unidos (Tesouro de 10 anos), foi coletada através da API do *Federal Reserve*, utilizando-se uma média móvel de 12 meses retroativa ao período de análise em questão.

Por sua vez, o prêmio de risco de mercado, é calculado pela diferença entre o retorno esperado do mercado (comumente representado pelo índice S&P 500) e a taxa livre de risco. Para o vigente projeto, foram utilizados os dados disponibilizados por Damodaran (2024), cujo trabalho dispõe o prêmio de risco de mercado histórico para diversos países do mundo, incluindo o Brasil.

Para o cálculo do beta de 5 anos para cada ativo em cada período, aplicou-se uma regressão linear simples para estimar a relação entre o retorno do ativo e o de mercado. Utilizou-se a mesma base de dados de preços históricos empregada na criação das variáveis para o modelo de aprendizado de máquina. Além disso, foram incluídos os dados de cotações históricas do Ibovespa, obtidos através da biblioteca *yfinance*.

Por fim, para a mensuração do risco país, foi considerado a taxa vigente do *Credit Default Swap* de 5 anos em cada período analisado. Ele reflete as percepções do mercado sobre a solvência do país, por conta de sua alta liquidez e de atualizações constantes. Foram utilizados os dados do site *Investing* (INVESTING, 2025) para a construção das bases de dados históricas.

Em posse dessas variáveis, computou-se o retorno esperado de todos os ativos para cada período que, de forma idêntica à realizada para o classificador de *Random*

Forest, foram ordenados e disponibilizados ao otimizador de média-variância, possuindo as mesmas restrições apresentadas na Subseção 4.5.

5. RESULTADOS

Esta Seção apresenta e discute os principais achados da aplicação do algoritmo na construção de portfólios de investimento.

Inicialmente, é realizada uma análise comparativa do desempenho dos classificadores ao longo dos períodos, considerando diferentes quantidades de ativos disponíveis ao otimizador de média-variância. Em seguida, são exploradas em maior profundidade as carteiras formadas por 1, 5, 15, 30, 50 e 71 ativos, avaliando sua performance em termos de retorno e risco.

5.1 Avaliação e comparação dos classificadores

Os resultados obtidos através da otimização de portfólios com o algoritmo de *Random Forest* estão dispostos no Apêndice A, onde são apresentados gráficos que ilustram o desempenho dos portfólios otimizados em diferentes períodos e para variados números de ativos.

Os gráficos exibem o Índice Sharpe, o Índice de Treynor, o Alfa de Jensen e o log-retorno observado, com o eixo das abscissas representando o número de ativos no portfólio e o eixo das ordenadas refletindo o respectivo valor do indicador. As linhas azuis representam os resultados obtidos pelos portfólios classificados pelo algoritmo de *Random Forest*, as linhas verdes, pelos portfólios classificados via *Capital Asset Pricing Model*, e as linhas pretas tracejadas indicam os valores correspondentes para uma carteira de 71 ativos igualmente ponderados, utilizada como um *benchmark* adicional para análise comparativa das estratégias.

Para o Índice Sharpe, apresentado na Figura 8, os portfólios de *Random Forest* demonstram, de forma geral, um desempenho superior na maioria dos períodos analisados. Eles não apenas superam sistematicamente os classificados através do *Capital Asset Pricing Model*, como também apresentam picos de Índice Sharpe significativamente mais altos, evidenciando sua maior capacidade para gerar retornos expressivos. Conforme esperado, a estratégia de *Random Forest* também apresentou um desempenho superior à carteira igualmente ponderada na maioria dos períodos analisados, resultado acompanhado de uma volatilidade significativamente maior.

No caso do Índice de Treynor, apresentado na Figura 9, observa-se que, de modo geral, os portfólios construídos com *Random Forest* caminham em linha com os

classificados via *Capital Asset Pricing Model* e à carteira igualmente ponderada. No entanto, os portfólios de *Random Forest* tendem a exibir maior variação nesse indicador, especialmente em carteiras mais concentradas. Por utilizar padrões não lineares, os portfólios de *Random Forest* tendem a ser mais sensíveis às mudanças de mercado, resultando em um Índice de Treynor mais volátil, mas também indicando um potencial de retornos excedentes. Já o *Capital Asset Pricing Model*, ao assumir que o mercado precifica corretamente os ativos, tende a apresentar um Índice de Treynor mais estável.

Os resultados do Alfa de Jensen são exibidos na Figura 10. A análise desse indicador sugere uma possível correlação negativa entre as estratégias *Random Forest* e *Capital Asset Pricing Model* no que diz à obtenção de retorno excedentes ao longo do tempo, indicando que, em diversos momentos, uma estratégia captura ganhos enquanto a outra apresenta desempenho inferior. Vale notar que as carteiras mais concentradas do *Capital Asset Pricing Model* frequentemente exibem um Alfa de Jensen negativo. Já o Alfa de Jensen da carteira igualmente ponderada tende a zero, indicando que ela tende a acompanhar o retorno de mercado.

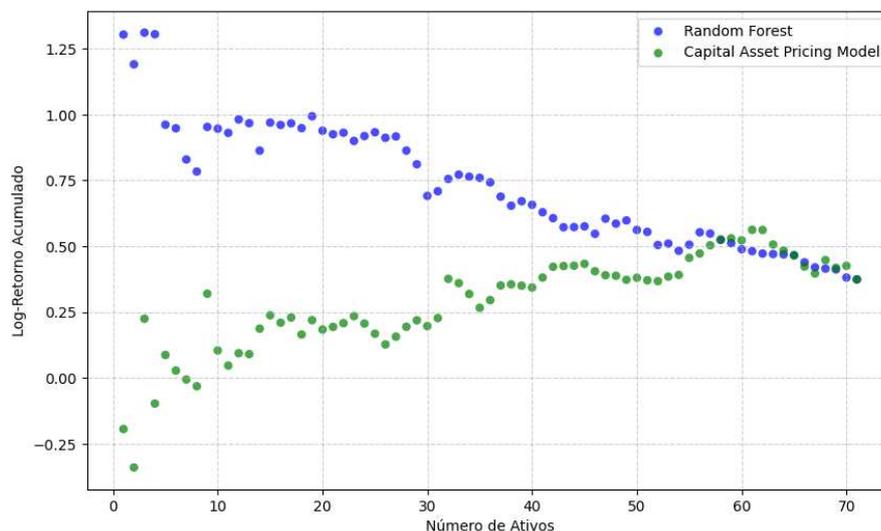
Por fim, conforme exposto na Figura 11, a análise do log-retorno entre o modelo baseado em *Random Forest* e o *Capital Asset Pricing Model* evidencia uma vantagem significativa do algoritmo de aprendizado de máquina. Nos períodos analisados, observa-se que as carteiras classificadas pelo *Random Forest* apresentam sistematicamente retornos superiores às classificadas pelo *Capital Asset Pricing Model*.

Também cabe destacar que, à medida que o número de ativos na carteira aumenta, os retornos das estratégias tendem a convergir. Esse comportamento é esperado, pois a diversificação reduz a influência dos ativos individuais, aproximando os portfólios de um desempenho médio do mercado. Essa relação pode ser mais bem visualizada na Figura 5.

A análise do gráfico revela que a estratégia baseada em *Random Forest*, representada pelos pontos azuis, apresenta um desempenho superior em termos de log-retorno acumulado, especialmente para carteiras com um número reduzido de ativos. Isso sugere que o modelo de aprendizado de máquina possui uma maior capacidade de seleção de ativos com retornos mais elevados. Em contrapartida, a abordagem baseada no *Capital Asset Pricing Model*, representada pelos pontos

verdes, demonstra retornos acumulados significativamente menores para carteiras de poucos ativos à disposição do otimizador de média-variância.

Figura 5 - log-retorno acumulado, 2018-2019, por estratégia e número de ativos no portfólio



Fonte: elaborado pelo autor (2025).

À medida que o número de ativos nas carteiras aumenta, observa-se uma convergência gradual entre as duas abordagens, indicando que a diversificação reduz o impacto da seleção individual de ativos. Isso sugere que, embora o *Random Forest* se destaque na escolha de ativos mais rentáveis quando há poucas opções, sua vantagem relativa diminui conforme a diversificação aumenta, tornando os retornos observados pelas duas estratégias mais semelhantes.

5.2 *Backtesting* de estratégias de investimento

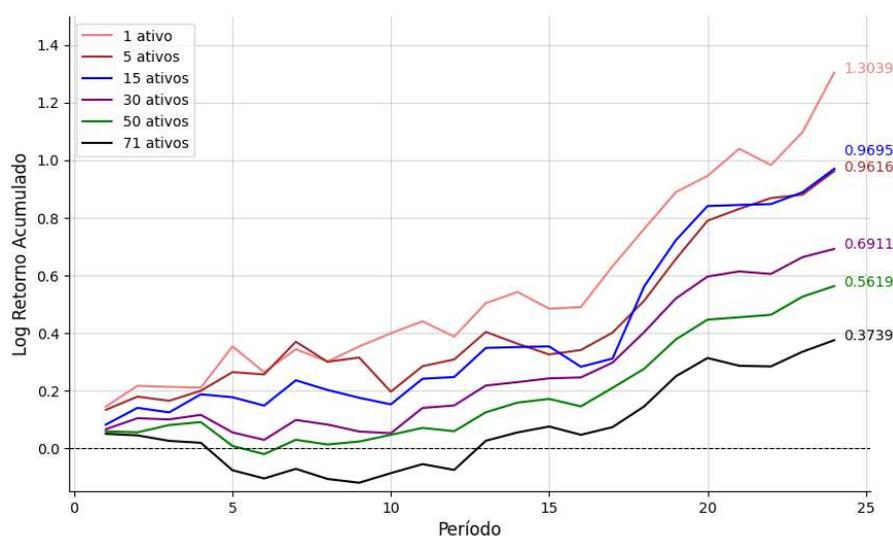
Para a realização do *backtesting* do retorno acumulado e das métricas de risco-retorno obtidas pelas diferentes estratégias de investimento avaliadas neste estudo, selecionou-se os portfólios com 1, 5, 15, 30, 50 e 71 ativos, representando diferentes graus de concentração e diversificação.

O portfólio com 1 ativo maximiza o potencial de retorno, apostando sempre no ativo mais promissor, mas assume o maior risco idiossincrático. O portfólio com 5 ativos ainda mantém alta concentração, mas é capaz de reduzir parcialmente o risco. Os portfólios com 15, 30, 50 e 71 ativos representam estratégias de diversificação crescente, onde o risco idiossincrático é progressivamente reduzido, resultando em

maior estabilidade nos retornos e indicadores mais consistentes ao longo dos períodos.

Para simular a aplicação prática dos métodos de otimização como estratégia de investimento, inicialmente foi analisado o desempenho do log-retorno acumulado ao longo dos 24 períodos de estudo. Os resultados obtidos pelo algoritmo de *Random Forest* são expostos na Figura 6.

Figura 6 - log-retorno acumulado obtido pelo classificador de *Random Forest*



Fonte: elaborado pelo autor (2025).

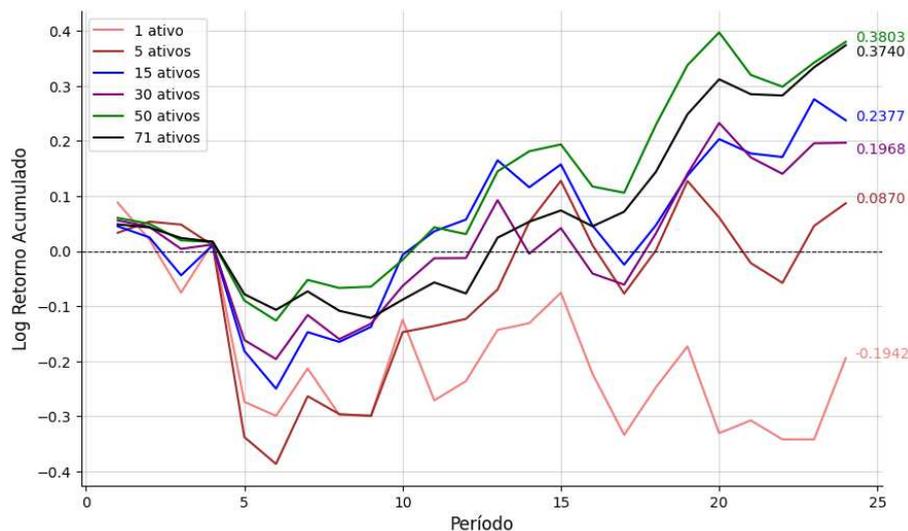
É possível concluir que o algoritmo demonstrou eficácia ao selecionar os ativos com melhores potenciais de retorno, conforme evidenciado pelo desempenho superior das carteiras mais concentradas, como as de 1 e 5 ativos. Essas carteiras apresentaram um crescimento expressivo no log-retorno acumulado, indicando que o modelo foi capaz de priorizar os ativos mais promissores em cada período.

As carteiras que ofereceram um leque maior de ativos ao otimizador, como as formadas por 30, 50 e 71 ativos, apresentaram retornos acumulados menores em comparação às carteiras mais concentradas. No entanto, ainda registraram rentabilidades elevadas, destacando que a diversificação preservou resultados positivos.

Já os resultados do *Capital Asset Pricing Model*, expostos na Figura 7, evidenciam que o modelo prioriza a diversificação e a redução do risco sistemático, alinhado ao seu pressuposto teórico de que carteiras mais diversificadas são menos suscetíveis a flutuações idiossincráticas de ativos individuais. Essa característica se

reflete no desempenho positivo das carteiras com maior número de ativos, como as de 50 e 71 ativos.

Figura 7 - log-retorno acumulado obtido pelo classificador de *Capital Asset Pricing Model*



Fonte: elaborado pelo autor (2025).

Em contrapartida, as carteiras mais concentradas, compostas por 1 e 5 ativos, apresentam os menores log-retornos acumulados, com a carteira de 1 ativo registrando até mesmo um resultado negativo (-0,1942). Esse achado ressalta a vulnerabilidade da estratégia para a construção de carteiras não diversificadas, que sofrem com a exposição direta ao desempenho dos poucos ativos selecionados.

É importante destacar que, pelo fato de ambas as estratégias terem utilizado o critério de média-variância, as carteiras de 71 ativos obrigatoriamente convergem em termos de retorno e indicadores, pois possuem todos os ativos disponíveis e as mesmas restrições aplicadas ao otimizador.

Em última análise, cabe notar que todas as carteiras otimizadas com o classificador baseado no *Capital Asset Pricing Model* apresentaram prejuízos acumulados durante uma parte significativa do período de análise, iniciando sua recuperação apenas por volta do décimo período. Em contraste, as carteiras construídas com *Random Forest* exibiram retornos positivos de forma consistente desde o início. Esse padrão sugere que a recuperação do resultado das carteiras do *Capital Asset Pricing Model* pode estar mais relacionada a um movimento favorável do mercado do que à real eficácia do classificador em identificar boas oportunidades de investimento.

Ademais, no Apêndice B é apresentada uma comparação da performance das estratégias para cada carteira ao longo do período de análise para o Índice Sharpe (Figura 12), o Índice de Treynor (Figura 13), o Alfa de Jensen (Figura 14) e o log-retorno observado (Figura 15).

A análise dos gráficos revela importantes diferenças no desempenho das estratégias de aprendizado de máquina e do *Capital Asset Pricing Model*. Os resultados mais relevantes são sumarizados na Tabela 6, que exhibe o Índice Sharpe, o Índice de Treynor e o Alfa de Jensen médios por período de cada carteira avaliada.

Tabela 6 - indicadores médios por estratégia e número de ativos

| Métrica | Random Forest | Capital Asset Pricing Model |
|-----------------------------------|---------------|-----------------------------|
| 1. Índice Sharpe médio | | |
| 1.1 Carteira com 1 ativo | 0,178 | -0,096 |
| 1.2 Carteira com 5 ativos | 0,149 | -0,046 |
| 1.3 Carteira com 15 ativos | 0,414 | -0,043 |
| 1.4 Carteira com 30 ativos | 0,122 | -0,07 |
| 1.5 Carteira com 50 ativos | 0,149 | -0,048 |
| 1.6 Carteira com 71 ativos | 0,029 | 0,029 |
| 2. Índice de Treynor médio | | |
| 2.1 Carteira com 1 ativo | 0,179 | -0,01 |
| 2.2 Carteira com 5 ativos | 0,032 | -0,007 |
| 2.3 Carteira com 15 ativos | 0,024 | -0,007 |
| 2.4 Carteira com 30 ativos | 0,003 | -0,011 |
| 2.5 Carteira com 50 ativos | -0,001 | -0,008 |
| 2.6 Carteira com 71 ativos | -0,015 | -0,015 |
| 3. Alfa de Jensen médio | | |
| 3.1 Carteira com 1 ativo | 0,036 | -0,018 |
| 3.2 Carteira com 5 ativos | 0,021 | -0,008 |
| 3.3 Carteira com 15 ativos | 0,021 | -0,005 |
| 3.4 Carteira com 30 ativos | 0,008 | -0,009 |
| 3.5 Carteira com 50 ativos | 0,002 | -0,003 |
| 3.6 Carteira com 71 ativos | -0,006 | -0,006 |

Fonte: elaborado pelo autor (2025).

No que se refere ao Índice Sharpe médio, o *Random Forest* apresentou valores positivos em todas as carteiras, com destaque para a carteira de 15 ativos (0,414), evidenciando uma relação risco-retorno bastante favorável. Por outro lado, as carteiras de *Capital Asset Pricing Model* registraram resultados negativos em todas as composições, com exceção à carteira de 71 ativos. Esses resultados sugerem que, de forma geral, a abordagem baseada em aprendizado de máquina consegue capturar melhor a relação entre risco e retorno, superando consistentemente a estratégia tradicional.

No caso do Índice de Treynor médio, o *Random Forest* também se destacou, apresentando resultados positivos na maioria das configurações. O fato de o melhor

resultado para esse indicador ter sido apresentado na carteira com apenas 1 ativo demonstra que o modelo foi capaz de identificar consistentemente ativos com um beta relativamente baixo e um retorno significativo, maximizando a relação entre retorno excedente e risco sistemático. O *Capital Asset Pricing Model*, por sua vez, exibiu valores negativos em todas as carteiras, sugerindo uma relação pouco eficiente entre risco sistemático e retorno. Dessa forma, o *Random Forest* se demonstrou mais eficaz na seleção de ativos que capazes de remunerar adequadamente o risco de mercado.

De forma similar ao Índice de Treynor, o Alfa de Jensen médio da carteira com 1 ativo do algoritmo de *Random Forest* foi o mais elevado. O valor encontrado indica que, em média, a cada período o portfólio de *Random Forest* obtém um log-retorno excedente de 3,60% além do que seria previsto pelo mercado. Esse resultado merece destaque, pois demonstra que, mesmo considerando o risco sistemático do mercado, a estratégia baseada em aprendizado de máquina conseguiu agregar valor extra ao investidor, superando o retorno que seria justificado apenas pelo nível de risco assumido. Em contraste, o *Capital Asset Pricing Model* apresentou valores negativos em todas as configurações, demonstrando menor habilidade para gerar retornos acima do previsto pelo próprio modelo.

6. CONCLUSÃO

O presente estudo analisou a aplicação do algoritmo de aprendizado de máquina *Random Forest*, integrado à TMP, no contexto de seleção e otimização de portfólios financeiros no mercado acionário brasileiro. A pesquisa explorou as diferenças de desempenho entre as carteiras otimizadas por esse método e as estruturadas segundo a abordagem tradicional do *Capital Asset Pricing Model*, avaliando as métricas Índice Sharpe, Índice de Treynor e Alfa de Jensen, durante o biênio 2018-2019.

Os resultados evidenciam que o algoritmo se destacou ao explorar padrões não-lineares e relações complexas entre ativos, gerando carteiras mais eficazes em termos de retorno ajustado ao risco. As carteiras compostas por um número menor de ativos apresentaram os melhores resultados, enquanto portfólios excessivamente diversificados demonstraram redução significativa de rentabilidade. Essa observação reforça que, embora a diversificação seja importante para mitigar riscos, sua aplicação excessiva pode comprometer o potencial de retorno.

Nesse sentido, a performance superior do *Random Forest* sugere que abordagens baseadas em aprendizado de máquina podem agregar valor na seleção de ativos, especialmente em carteiras menores. Em contraste, o *Capital Asset Pricing Model*, por ser um modelo mais simples, parece depender mais da diversificação para mitigar seus problemas de seleção de ativos, mas ainda assim não consegue superar a performance do modelo de *Random Forest*.

Os resultados comparativos entre as duas abordagens indicam uma clara vantagem do *Random Forest* na construção de portfólios, evidenciada pelos valores superiores nos três indicadores de desempenho analisados. Enquanto o *Capital Asset Pricing Model* apresentou resultados predominantemente negativos ou próximos de zero, sugerindo baixa eficiência na otimização do retorno ajustado ao risco, o algoritmo de *Random Forest* demonstrou maior capacidade de identificar carteiras com melhor relação risco-retorno, especialmente em portfólios menores e intermediários.

Além disso, o estudo reforçou a importância de validar a metodologia para a seleção de ativos, proposta por Abreu (2021), em contextos de mercado estáveis, de modo a garantir a robustez da estratégia proposta. A escolha desse período permitiu

a análise do modelo em condições menos influenciadas por eventos extraordinários, assegurando maior confiabilidade na comparação de desempenho.

Como teorizado, os resultados obtidos nesse projeto em termos de log-retorno superaram significativamente os encontrados por Abreu (2021). Apesar do tratamento dos períodos de análise realizado pelo autor, isso não impediu que o otimizador fosse afetado, minando parte relevante da rentabilidade potencialmente capaz de ser gerada pelo algoritmo em um cenário menos caótico.

Ainda assim, um ponto em comum de ambos os projetos é que carteiras mais concentradas foram as que apresentaram desempenhos mais satisfatórios em termos de rentabilidade. Isso indica que, independentemente do nível de volatilidade enfrentado, o classificador foi capaz de ordenar os ativos com maior potencial de retorno.

Os achados deste trabalho têm implicações práticas significativas para investidores e gestores, evidenciando o potencial do aprendizado de máquina como uma ferramenta complementar na gestão de portfólios. A integração de ferramentas como o *Random Forest* à gestão de investimentos permite atender tanto investidores com perfis mais conservadores, interessados em uma melhor relação risco-retorno, quanto aqueles que buscam a maximização de retornos em ambientes mais voláteis. Essa flexibilidade amplia a atratividade da estratégia, mostrando-se como uma solução robusta para diferentes objetivos financeiros.

Finalmente, estudos futuros podem ampliar essa abordagem testando outras técnicas de aprendizado de máquina, como redes neurais, e incorporando variáveis explicativas mais abrangentes. Além disso, a utilização de outros mecanismos de otimização de carteiras, como regularização Lasso e Ridge, pode contribuir para a seleção mais eficiente de ativos, reduzindo o impacto de ruídos nos dados e melhorando a estabilidade dos portfólios. Explorar novos cenários de mercado e períodos de análise mais longos também pode fortalecer a compreensão da aplicação do aprendizado de máquina na gestão de portfólios, ampliando as possibilidades de tomada de decisão fundamentada em dados.

REFERÊNCIAS

- ABREU, André. **Aplicação de *machine learning* na pré-seleção de ativos para portfólios de investimento**. Tese (Mestrado em Economia) – Escola de Economia de São Paulo, Fundação Getúlio Vargas. São Paulo, p. 44. 2021.
- ALPAYDIN, Ethem. **Introduction to Machine Learning**. 3^a ed. The MIT Press, 2014.
- ANBIMA. **Raio X do Investidor Brasileiro**: 6^a edição. 2023. Disponível em: https://www.anbima.com.br/pt_br/especial/raio-x-do-investidor-2023.htm. Acesso em: 27 de abril de 2024.
- ANBIMA. **Raio X do Investidor Brasileiro**: 7^a edição. 2024. Disponível em: https://www.anbima.com.br/pt_br/especial/raio-x-do-investidor-brasileiro.htm. Acesso em: 22 de maio de 2024.
- BELL, Steve. **Quantitative Finance for dummies**. 1^a ed. Chichester: John Wiley & Sons, 2016.
- BRASIL. Ministério da Fazenda. **Metodologia de Cálculo do WACC**. Brasília, 2018.
- BREIMAN, Leo. **Random Forests**. Machine Learning, v. 45, n. 1, p. 5-32, 2001.
- CHAN, K. et al. Earnings quality and stock returns. The Journal of Business, v. 79, n. 3, p. 1041-1082, 2006.
- COSTA, Thiago. **Otimização de portfólio com pré-seleção de ativos usando machine learning**: uma aplicação no contexto dos mercados emergentes. Revista Debates em Economia Aplicada – REDEA, [S. l.], v. 2, n. 1, 2022. Disponível em: <https://www.portaldeperiodicos.idp.edu.br/redea/article/view/6637>. Acesso em: 10/08/2024.
- DAMODARAN, Aswath. **Avaliação de investimentos**: ferramentas e técnicas para a determinação do valor de qualquer ativo. 1^a ed., 7^a reimpressão. Rio de Janeiro: Qualitymark Ed., 2005.
- DAMODARAN, Aswath. **Equity Risk Premiums (ERP)**: Determinants, Estimation, and Implications – The 2024 Edition. Stern School of Business, 2024. Disponível em: <https://ssrn.com/abstract=4751941>. Acesso em: 07/11/2024.
- FAMA, E. F. **Efficient capital markets**: A review of theory and empirical work. The Journal of Finance, v.25, n. 2, p. 383-417,1970.
- FAMA, E. F.; FRENCH, K. R. **The cross-section of expected stock returns**. The Journal of Finance, v. 47, n. 2, p. 427-465, 1992.
- GÉRON, Aurelien. **Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow**: Concepts, Tools, and Techniques to Build Intelligent Systems. 2^a ed. Estados Unidos: O'Reilly Media, Inc., 2019.

GUIMARÃES JÚNIOR, F. R. F.; CARMONA, C. U. de M.; GUIMARÃES, L. G. de A. **Carteiras Formadas por Meio de Variáveis Fundamentalistas Apresentam Bom Desempenho de Mercado?** *Gestão & Regionalidade*. São Paulo, v. 31, n. 91, p. 87-104, 2015.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction** (2ª ed.). Stanford, CA: Stanford University, 2009.

HAUGLAND, Johannes. **A comparative study of machine learning models in stock price prediction**. Tese (Mestrado em Business Analytics) – BI Norwegian Business School. Oslo, p. 10. 2023.

HODRICK, L. S.; MOULTON, P. C. **Liquidity: Considerations of a portfolio manager**. *Financial Management*, v. 38, n. 1, p.59-74, 2009.

INVESTING. **Credit Default Swap 5 anos – Brasil**. Disponível em: <https://www.investing.com>. Acesso em: 15 fev. 2025.

JEGADEESH, N.; TITMAN. S. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, V. 48, n. 1, p. 65-91, 1993.

JENSEN, M. C. **The performance of mutual funds in the period 1945 – 1964**. *Journal of Finance*, v. 23 (2), p. 389-416, 1968.

KAUARK, F. da S.; MANHÃES, F. C.; MEDEIROS, C. H. **Metodologia da pesquisa: um guia prático**. Via Litterarum, Bahia, 2010.

KUMAR, M.; THENMOZHI, M. **Forecasting Stock Index Movement: A Comparison of Support Vector Machines and *Random Forest***. Indian Institute of Capital Markets, 2006. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=876544. Acesso em: 10/10/2024.

LE SOURD, Véronique. **Performance Measurement for Traditional Investment – Literature Survey**. Lille - Nice: EDHEC, 2007.

LINTNER, John. **The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets**. *The Review of Economics and Statistics*, Vol. 47, n. 1., p. 13-37, 1965.

MA, Y.; HAN, R.; WANG, W. **Portfolio optimization with return prediction using deep learning and machine learning**. *Expert Systems With Applications*, v. 165, n. 1, 2021.

MARKOWITZ, H. M. **Portfolio Selection**. *The Journal of Finance*, v. 7, n. 1, p. 77-91, 1952.

MOSSIN, Jan. **Equilibrium in a capital asset market**. *Econometrica*, Vol. 34, n. 4, p. 768-783, 1966.

O'HARA, Maureen. **Market microstructure theory**. Malden: Blackwell Publishers Ltd., 1999.

OLIVEIRA FILHO, B. G.; SOUSA, A. F. **Fundos de investimento em ações no Brasil: métricas para avaliação de desempenho**. REGE, São Paulo – SP, Brasil, v. 22, n. 1, p. 61-76, 2015.

ROM, B. M.; FERGUSON, K. W. **Post-Modern Portfolio Theory comes of age**. *The Journal of Investing*, v. 2, n. 4, p.27-33, 1993.

ROSS, S. A. **The arbitrage theory of capital asset pricing**. *Journal of Economic Theory*, v. 13, n. 3, p. 341-360, 1976.

ROUWENHORST, K. G. **International momentum strategies**. *The Journal of finance*, v. 53, n. 1, p. 267-284, 1999.

SHARPE, W. F. **Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk**. *The Journal of Finance*, v. 19, n. 3, p. 425-442, 1964.

SHARPE, W. F. **Mutual fund performance**. *The Journal of business*, v. 39, n. 1, p. 119-138, 1966.

SILVA, M. C.; MATTAR, B. M.; ROSA, M.; OLIVEIRA, E. F. **Diversificação do risco de um portfólio de ativos modelo de markowitz**. *Revista eletrônica do departamento de ciências contábeis e departamento de atuária e métodos quantitativos da FEA (Redeca)*. São Paulo, v.7, n. 2., p. 19-34, jul./dez. 2020.

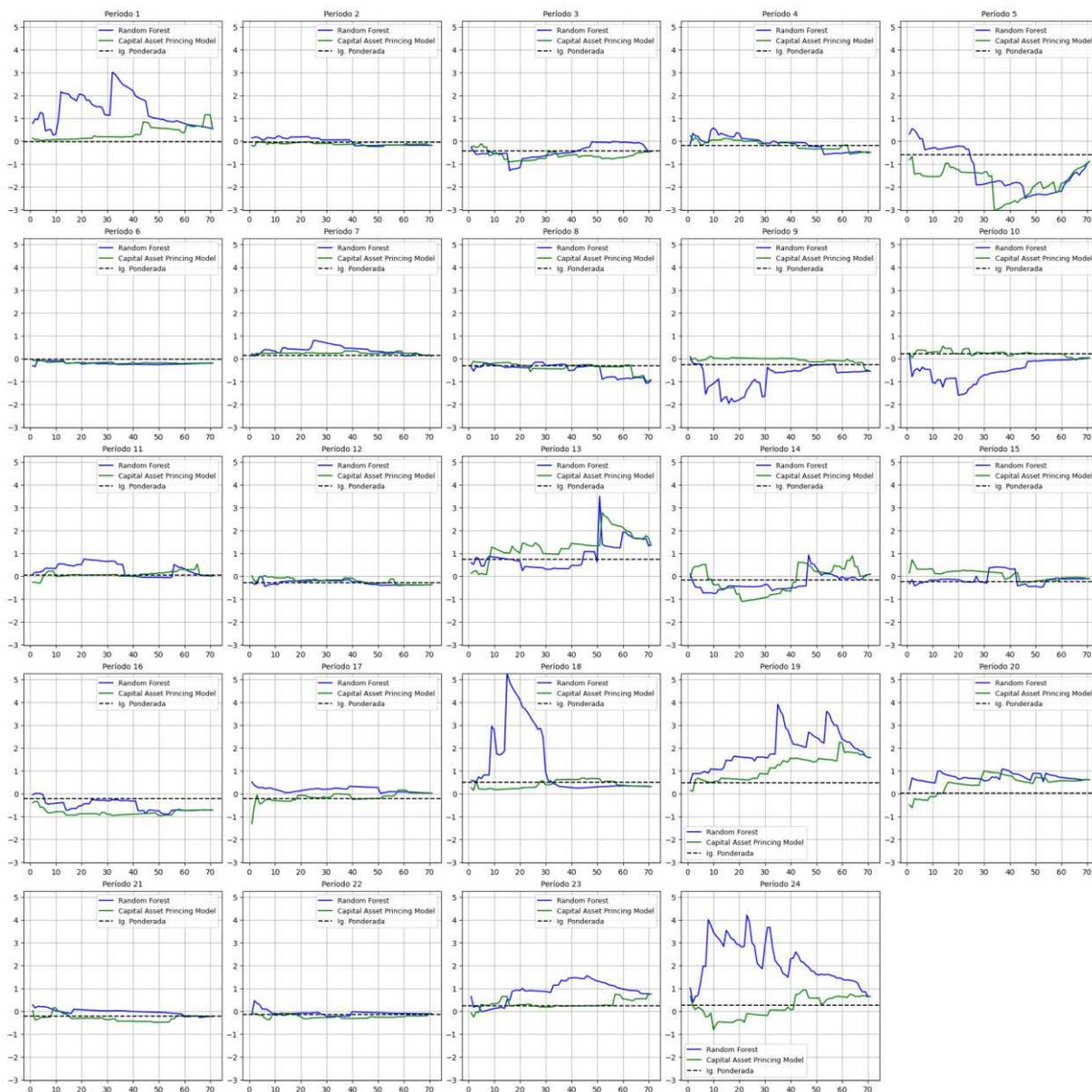
TAN, Z.; YAN, Z.; ZHU, G. **Stock selection with *Random Forest*: An exploitation of excess return in the Chinese stock market**. *Heliyon*, v. 5, n. 8, 2019.

TREYNOR, J. L. **How to Rate Management of Investment Funds**. *Harvard Business Review*, v. 43, n. 1 p. 63-75, 1965.

WOLFF, D.; ECHTERLING, F. **Stock Picing with Machine Learning**. *Journal of Forecasting*, v. 43, 2023.

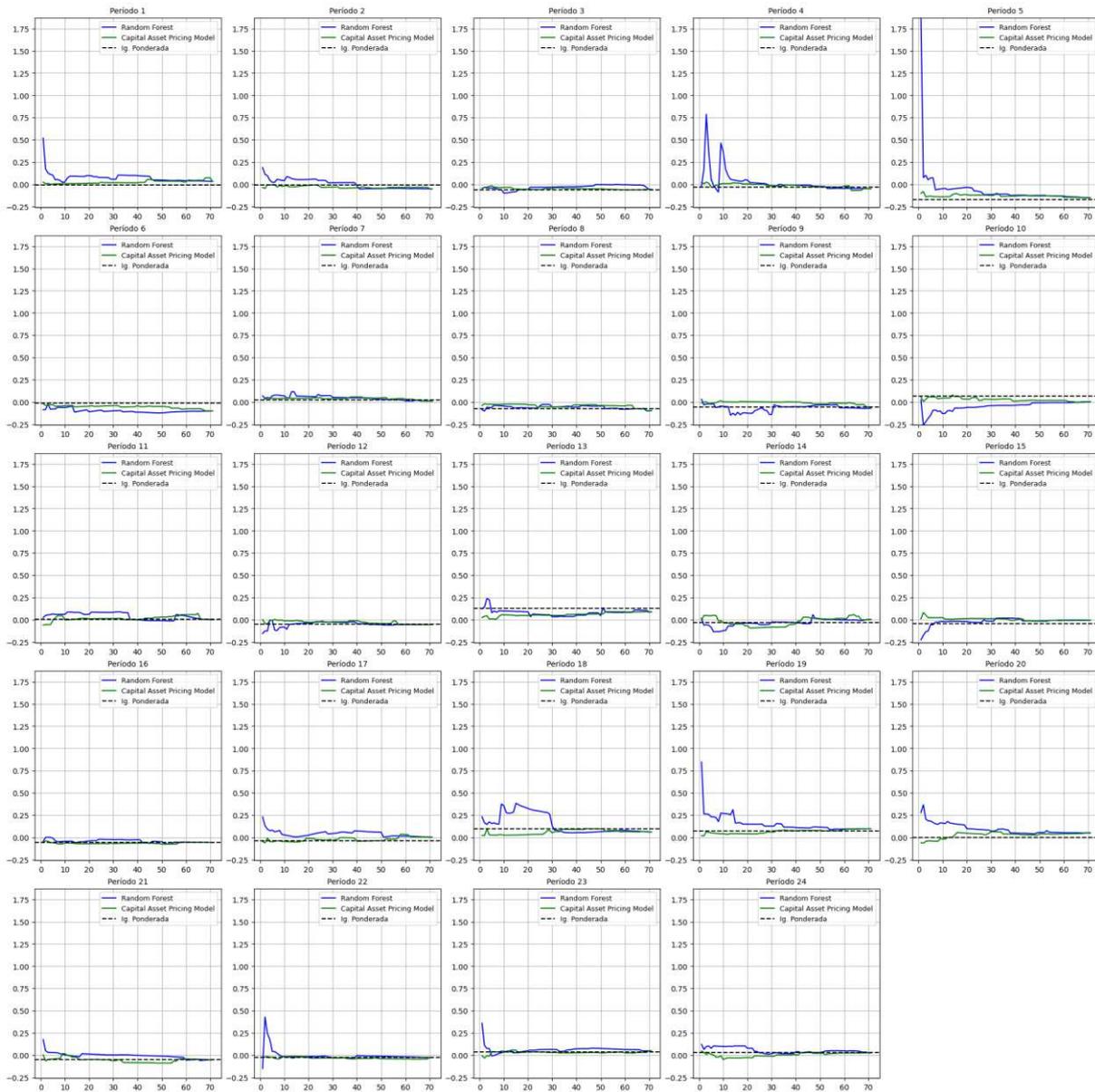
APÊNDICE A – RESULTADOS DOS CLASSIFICADORES POR PERÍODO DE VALIDAÇÃO

Figura 8 – Índice Sharpe de acordo com o número de ativos disponibilizados ao otimizador



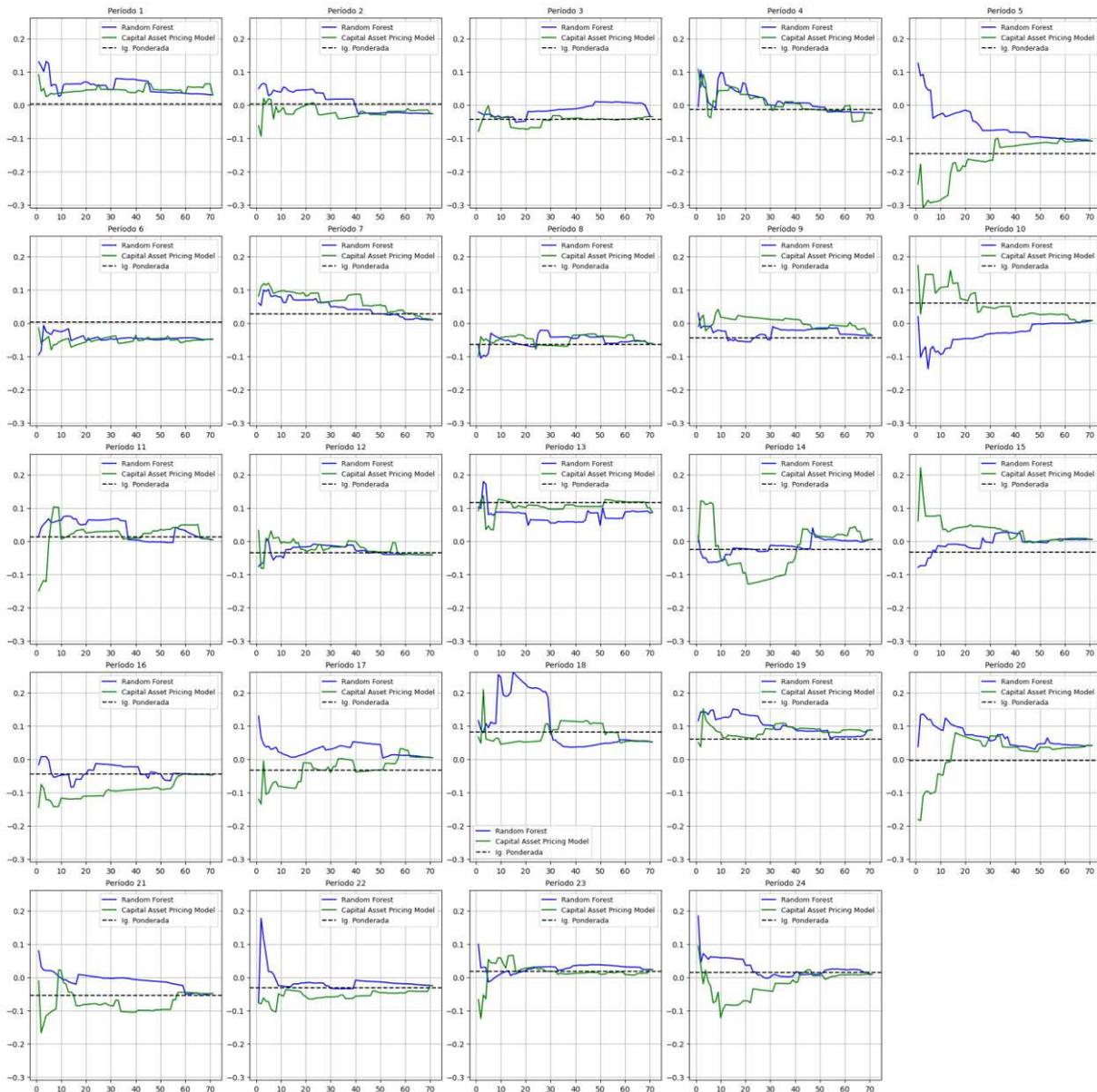
Fonte: elaborado pelo autor (2025).

Figura 9 - Índice de Treynor de acordo com o número de ativos disponibilizados ao otimizador



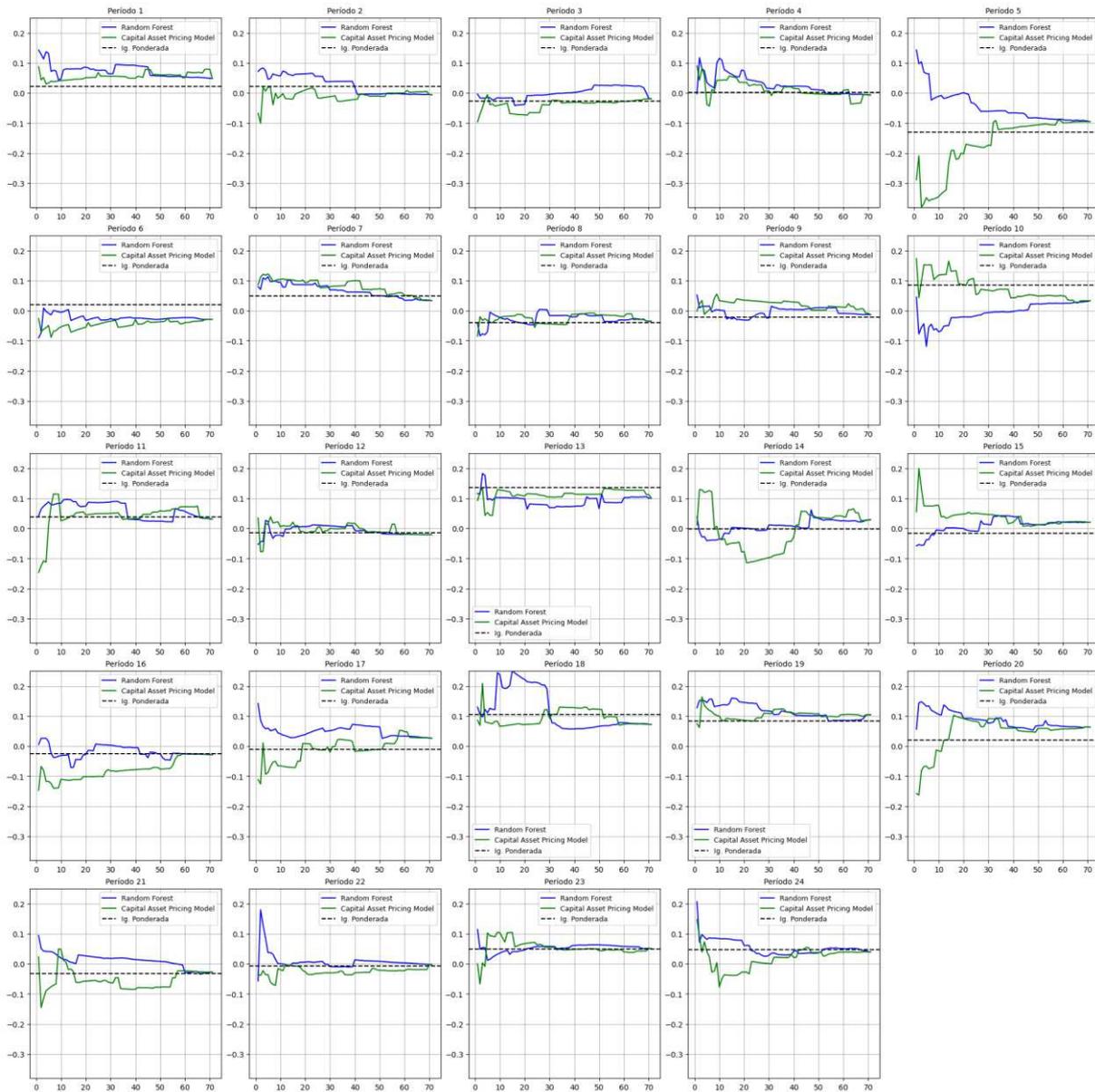
Fonte: elaborado pelo autor (2025).

Figura 10 - Alfa de Jensen de acordo com o número de ativos disponibilizados ao otimizador



Fonte: elaborado pelo autor (2025).

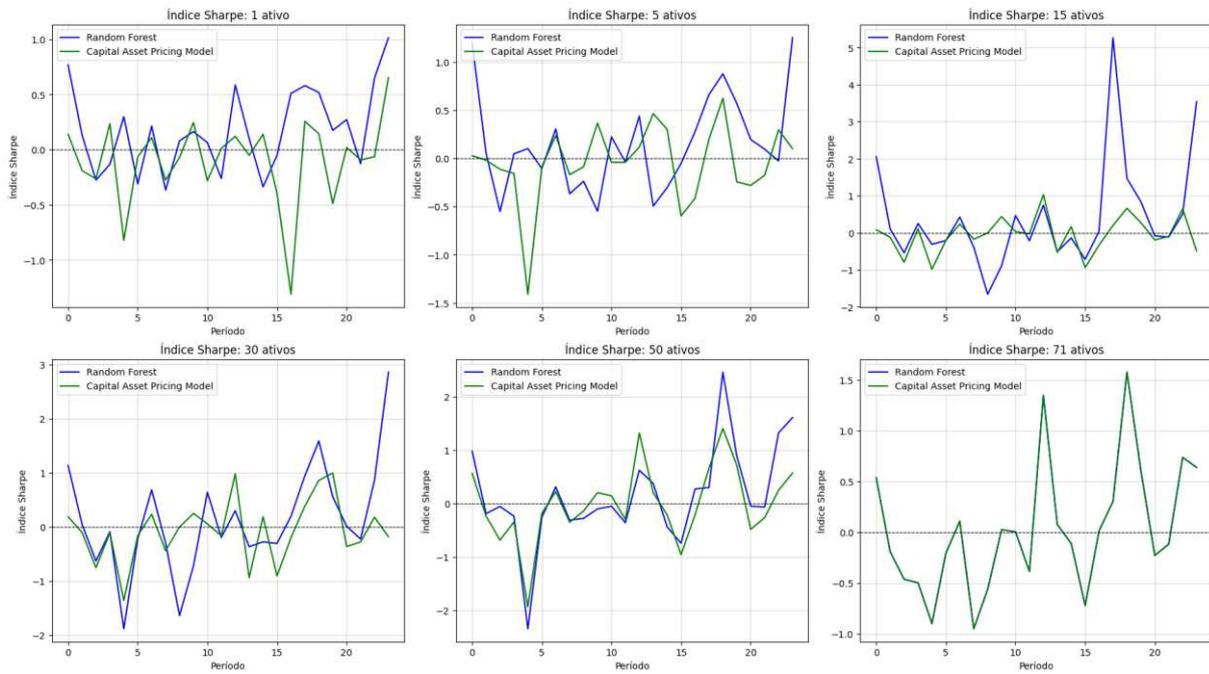
Figura 11 - Log-retorno obtido de acordo com o número de ativos disponibilizados ao otimizador



Fonte: elaborado pelo autor (2025).

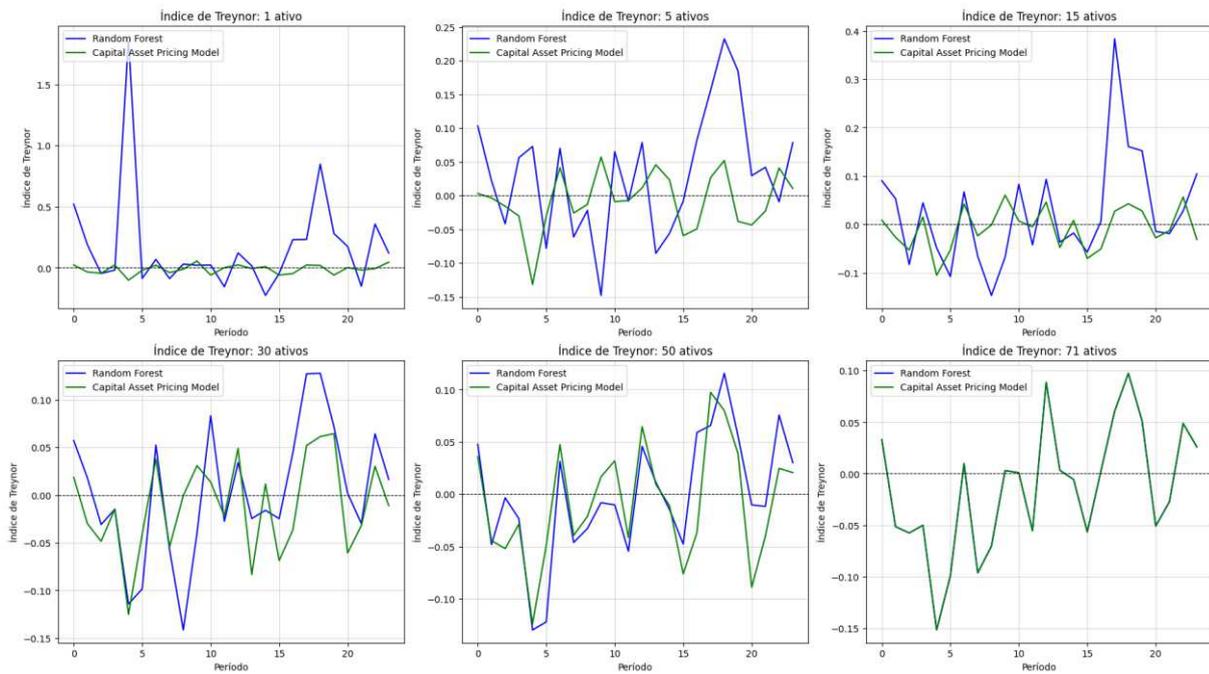
APÊNDICE B – COMPARAÇÃO DAS ESTRATÉGIAS

Figura 12 - Comparação do Índice Sharpe das estratégias para carteiras com 1, 5, 15, 30, 50 e 71 ativos



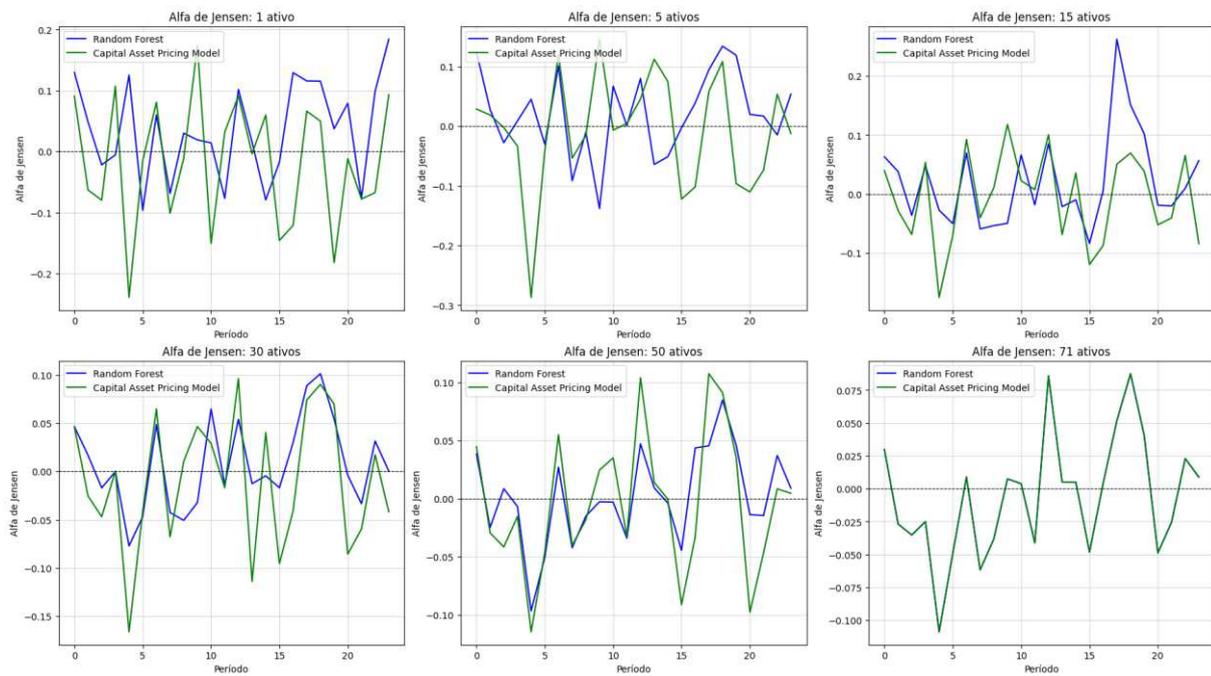
Fonte: elaborado pelo autor (2025).

Figura 13 - Comparação do Índice de Treynor das estratégias para carteiras com 1, 5, 15, 30, 50 e 71 ativos



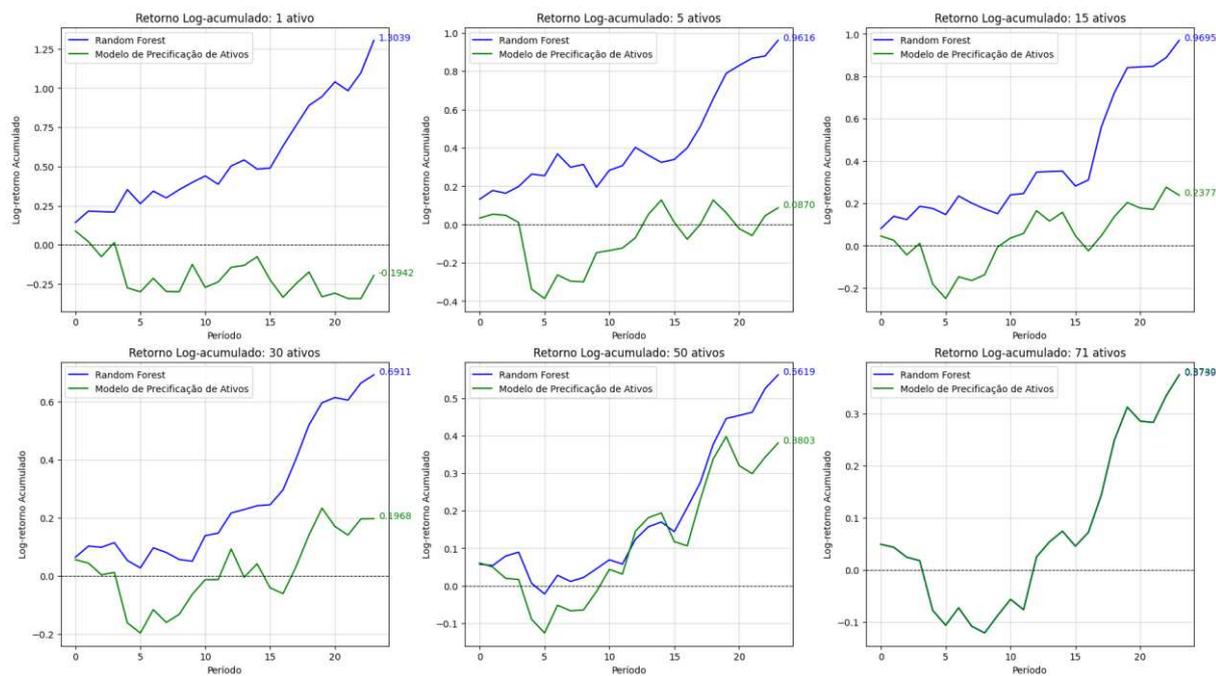
Fonte: elaborado pelo autor (2025).

Figura 14 - Comparação do Alfa de Jensen das estratégias para carteiras com 1, 5, 15, 30, 50 e 71 ativos



Fonte: elaborado pelo autor (2025).

Figura 15 - Comparação do log-retorno das estratégias para carteiras com 1, 5, 15, 30, 50 e 71 ativos



Fonte: elaborado pelo autor (2025).