

CENTRO UNIVERSITÁRIO DE MACEIÓ
CURSO DE CIÊNCIA DA COMPUTAÇÃO

Gabriel Canuto de Alencar

Marcus Cezar Moreira Ferraz

Otimização de Timing no Buy and Hold para Ações Perenes com Random Forest:
Análise Comparativa Contra Índices de Mercado

Gabriel Canuto de Alencar

Marcus Cezar Moreira Ferraz

**Otimização de Timing no Buy and Hold para Ações Perenes com Random Forest:
Análise Comparativa Contra Índices de Mercado**

Trabalho de Conclusão de Curso
apresentado ao Centro
Universitário de Maceió como um
dos pré-requisitos para a obtenção
de grau de Bacharel em Ciência da
Computação.

Orientador: Prof. Me. João Gabriel Gama Vila Nova

Maceió, AL
2025

Catálogo na fonte: Biblioteca do Centro Universitário de Maceió, Unima | Afya

A368o

Alencar, Gabriel Canuto de

Otimização de timing no Buy and hold para ações perenes com random forest : análise comparativa contra índices de mercado / Gabriel Canuto de Alencar, Marcus Cezar Moreira Ferraz ; orientação [de] João Gabriel Gama Vila Nova. – Maceió, 2025.
73 f. : il.

Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) -Afya Centro Universitário de Maceió | Afya, Maceió, 2025.

Inclui Bibliografias: p. 58-60.

1. Aprendizado de máquina. 2. Otimização de portfólio. 3. Random Forest. 4. Buy and hold. 5. Backtestin. I. Ferraz, Marcus Cezar Moreira. II. Vila Nova, João Gabriel Gama. (orient.). III Centro Universitário de Maceió. IV. Título.

CDU : 004

Bibliotecária responsável: Adriele da Silva Lima CRB-4/1898

Gabriel Canuto de Alencar

Marcus Cezar Moreira Ferraz

Otimização de Timing no Buy and Hold para Ações Perenes com Random Forest:

Análise Comparativa contra Índices de Mercado

Monografia apresentada ao Centro
Universitário de Maceió como um dos
pré-requisitos para a obtenção de grau de
bacharel em Ciência da Computação.

Aprovada em 10/11/2025.

Banca Examinadora

Me. João Gabriel Gama Vila Nova
Centro Universitário de Maceió

Prof. Me. Izaac Duarte de Alencar
Centro Universitário de Maceió

Prof. Esp. Victor Brunno Dantas de S. Rosas
Centro Universitário de Maceió

DEDICATÓRIA

Gostaríamos de dedicar este trabalho para nossas famílias, que sempre nos apoiaram e ajudaram em nosso caminho acadêmico.

AGRADECIMENTOS

Gostaríamos de agradecer conjuntamente em especial ao nosso Orientador, Professor João Gabriel, pela dedicação, paciência e valiosas contribuições que nortearam este trabalho. Agradecemos também aos membros da banca, Prof. Isaac Alencar e Prof. Victor Brunno, por disporem do seu tempo para avaliar e contribuir com nosso trabalho.

Gabriel Canuto de Alencar

Agradeço especialmente à minha família e à namorada, por ser meu porto seguro com seu apoio incondicional, paciência e palavras de incentivo que me deram força para continuar. Aos meus amigos, que tornaram a jornada mais leve. Por fim, ao meu colega e amigo Marcus, minha gratidão pela parceria exemplar, pelas longas horas de trabalho e pelo aprendizado mútuo.

Marcus Cezar Moreira Ferraz

Agradeço especialmente à minha família e à minha namorada, pelo apoio e tranquilidade transmitidos durante todo o período de produção deste trabalho, e por serem minha fortaleza nos momentos mais complicados. E ao meu amigo, Gabriel, agradeço pela colaboração exemplar. Sua dedicação e profissionalismo foram fundamentais para a conclusão do trabalho.

RESUMO

A otimização dos pontos de compra é um fator crítico para potencializar os retornos de longo prazo na estratégia *buy and hold*. Este trabalho aborda esse desafio ao desenvolver e avaliar um *framework* que utiliza modelos *random forest* para prever momentos oportunos de compra em uma carteira de ações perenes. A eficácia dos sinais gerados pelos modelos é validada via um sistema de backtesting, que compara o desempenho de estratégias de alocação dinâmica contra *benchmarks* do mercado financeiro. Os resultados indicam que o *timing* guiado por *machine learning* pode oferecer vantagens significativas, e contribui com a própria estrutura metodológica para o teste de estratégias quantitativas.

Palavras-chave: Aprendizado de Máquina, Otimização de Portfólio, Random Forest, Buy and Hold, Backtesting

ABSTRACT

The optimization of entry points is a critical factor in enhancing long-term returns within the buy and hold strategy. This project addresses this challenge by developing and evaluating a framework that utilizes random forest models to predict opportune buying moments in a portfolio of blue-chip stocks. The effectiveness of the signals generated by the models is validated through a backtesting system, which compares the performance of dynamic allocation strategies against financial market benchmarks. The results indicate that machine learning-guided timing can offer significant advantages, and the project also contributes its methodological structure for testing quantitative strategies.

Keywords: Machine Learning, Portfolio Optimization, Random Forest, Buy and Hold, Backtesting

LISTA DE FIGURAS

Figura 1 - Ilustração da árvore de decisão.....	21
Figura 2 - Processos da metodologia.....	32
Figura 3 - Diagrama do <i>pipeline</i>	39

LISTA DE GRÁFICOS

Gráfico 1 - Rentabilidade cumulativa - Estratégias vs. <i>benchmarks</i> (2020 - 2024).....	49
Gráfico 2 - Rentabilidade cumulativa - Estratégias vs. <i>benchmarks</i> (Simulação de aportes)...	51
Gráfico C.1 - Histórico de <i>drawdown</i> (IA-FS-R2).....	71
Gráfico C.2 - Histórico de <i>drawdown</i> (IVVB11).....	71
Gráfico C.3 - Histórico de <i>drawdown</i> (IA-FS-R4).....	72
Gráfico C.4 - Histórico de <i>drawdown</i> (DIVO11).....	72
Gráfico C.5 - Histórico de <i>drawdown</i> (BOVA11).....	73

LISTA DE TABELAS

Tabela 1 - Lista dos ativos perenes selecionados, ordenados pela participação perenes.....	33
Tabela 2 - Estatísticas Descritivas das Variáveis (PETR4).....	36
Tabela 3 - Dicionário de Hiperparâmetros para o <i>RandomizedSearchCV</i>	40
Tabela 4 - Lista de estratégias definidas para o processo de <i>backtesting</i>	43
Tabela 5 - Métricas do modelo treinado de cada ativo.....	45
Tabela 6 - Resultados das estratégias ordenadas pelo retorno.....	46
Tabela 7 - Distribuição e desempenho da carteira da estratégia IA-FS-R2.....	47
Tabela 8 - Distribuição e desempenho da carteira da estratégia IA-FS-R4.....	48
Tabela 9 - Resultados de simulação de aportes dos <i>benchmarks</i>	50
Tabela 10 - Métricas de risco consolidadas.....	52
Tabela A.1 - Estatísticas descritivas das variáveis (BBDC4).....	61
Tabela A.2 - Estatísticas descritivas das variáveis (VALE3).....	61
Tabela A.3 - Estatísticas descritivas das variáveis (BBAS3).....	62
Tabela A.4 - Estatísticas descritivas das variáveis (CMIG4).....	62
Tabela A.5 - Estatísticas descritivas das variáveis (ITUB4).....	63
Tabela A.6 - Estatísticas descritivas das variáveis (BBSE3).....	63
Tabela A.7 - Estatísticas descritivas das variáveis (ITSA4).....	64
Tabela A.8 - Estatísticas descritivas das variáveis (TIMS3).....	64
Tabela A.9 - Estatísticas descritivas das variáveis (KLBN11).....	65
Tabela A.10 - Estatísticas descritivas das variáveis (Índice BOVESPA).....	65
Tabela A.11 - Estatísticas descritivas das variáveis (Taxa de câmbio - USD/BRL).....	66
Tabela A.12 - Estatísticas descritivas das variáveis (Barril de petróleo).....	66
Tabela A.13 - Estatísticas descritivas das variáveis (Minério de ferro).....	67
Tabela B.1 - Distribuição e desempenho da carteira da estratégia IA-L3-R4.....	68
Tabela B.2 - Distribuição e desempenho da carteira da estratégia IA-L3-R2.....	68
Tabela B.3 - Distribuição e desempenho da carteira da estratégia IA-L3-R3.....	69
Tabela B.4 - Distribuição e desempenho da carteira da estratégia IA-FS-R3.....	69
Tabela B.5 - Distribuição e desempenho da carteira da estratégia IA-L3-R1.....	70
Tabela B.6 - Distribuição e desempenho da carteira da estratégia IA-FS-R1.....	70

LISTA DE EQUAÇÕES

Equação 1 - Cálculo do SMA.....	23
Equação 2 - Cálculo do EMA.....	23
Equação 3 - Cálculo do WMA.....	23
Equação 4 - Cálculo do MACD.....	24
Equação 5 - Cálculo do MFM.....	24
Equação 6 - Cálculo do MFV.....	25
Equação 7 - Cálculo do A/D.....	25
Equação 8 - Cálculo do PT.....	25
Equação 9 - Cálculo do DM.....	26
Equação 10 - Cálculo do CCI.....	26
Equação 11 - Cálculo do RS.....	27
Equação 12 - Cálculo do RSI.....	27
Equação 13 - Cálculo do MOM.....	27
Equação 14 - Cálculo do %K.....	28
Equação 15 - Cálculo do %D.....	28
Equação 16 - Cálculo do %R.....	28

LISTA DE SIGLAS

A/D	Oscilador de Acumulação/Distribuição
API	<i>Application Programming Interface</i> (Interface de Programação de Aplicações)
ARIMA	<i>Autoregressive Integrated Moving Average</i> (Média Móvel Integrada Autoregressiva)
B3	Brasil, Bolsa, Balcão
BOVA11	ETF (Fundo de Índice) que replica o IBOVESPA
CCI	<i>Commodity Channel Index</i> (Índice de Canal de Commodities)
CDI	Certificado de Depósito Interbancário
DIVO11	ETF (Fundo de Índice) que replica o IDIV
DM	Desvio Médio
EMA	<i>Exponential Moving Average</i> (Média Móvel Exponencial)
ETF	<i>Exchange Traded Fund</i> (Fundo de Índice)
IA	Inteligência Artificial
IBOV	Índice Bovespa
IDIV	Índice de Dividendos
IPO	<i>Initial Public Offering</i> (Oferta Pública Inicial)
IVVB11	ETF (Fundo de Índice) que replica o S&P 500
LSTM	<i>Long Short-Term Memory</i> (Memória de Longo Prazo)
MACD	<i>Moving Average Convergence Divergence</i> (Convergência e Divergência de Médias Móveis)
MFM	Multiplicador de Fluxo de Dinheiro

MFV	Volume de Fluxo de Dinheiro
ML	<i>Machine Learning</i> (Aprendizado de Máquina)
MOM	<i>Momentum</i> (Momento)
OHLVC	<i>Open, High, Low, Close, Volume</i> (Abertura, Máxima, Mínima, Fechamento e Volume)
ON	Ações Ordinárias
PN	Ações Preferenciais
PT	Preço Típico
Q3	Terceiro Quartil
RF	<i>Random Forest</i> (Floresta Aleatória)
RFECV	<i>Recursive Feature Elimination with Cross-Validation</i> (Eliminação Recursiva de <i>Features</i> com Validação Cruzada)
RNN	<i>Recurrent Neural Network</i> (Rede Neural Recorrente)
RS	<i>Relative Strength</i> (Força Relativa)
RSI	<i>Relative Strength Index</i> (Índice de Força Relativa)
S&P 500	<i>Standard & Poor's 500</i>
SMA	<i>Simple Moving Average</i> (Média Móvel Simples)
SMA PT	Média Móvel Simples do Preço Típico
USD/BRL	Dólar Americano / Real Brasileiro
WMA	<i>Weighted Moving Average</i> (Média Móvel Ponderada)
%D	Linha %D do Oscilador Estocástico
%K	Linha %K do Oscilador Estocástico
%R	William's %R

SUMÁRIO

INTRODUÇÃO.....	16
2 OBJETIVOS.....	18
2.1 Objetivo geral.....	18
2.2 Objetivos específicos.....	18
3 ASPECTOS TEÓRICOS.....	19
3.1 Mercado de ações.....	19
3.2 Ações.....	19
3.3 <i>Machine learning</i> aplicado ao mercado financeiro.....	20
3.4 <i>Random forest</i>	20
3.5 <i>Web scraping</i>	21
3.6 Indicadores de análise técnica.....	22
3.6.1 Indicadores de tendência e volatilidade.....	22
3.6.2 Osciladores de momento.....	26
3.7 Indicadores macroeconômicos.....	29
3.8 <i>Backtesting</i>	29
3.9 <i>Benchmarks</i> do mercado financeiro.....	30
3.10 Métricas de risco e retorno.....	30
4 METODOLOGIA.....	32
4.1 Coleta de dados.....	32
4.2 Análise de dados.....	33
4.2.1 Seleção e ranqueamento de empresas perenes.....	33
4.2.2 Enriquecimento dos dados com indicadores técnicos e macroeconômicos.....	35
4.3 Treinamento dos modelos.....	37
4.3.1 Engenharia de variáveis de entrada.....	37
4.3.2 Engenharia de variável-alvo.....	38
4.3.3 Treinamento dos modelos.....	38
4.3.4 Otimização e redução de dimensionalidade dos modelos.....	40
4.4 Arquitetura do <i>trade system</i>	41
4.4.1 Criação das estratégias.....	41
4.4.2 Criação do <i>trade system</i>	43
4.4.3 Validação das estratégias e análise dos resultados.....	44
5 RESULTADOS.....	45

6 DISCUSSÕES.....	53
7 CONCLUSÕES.....	55
8 SUGESTÕES PARA TRABALHOS FUTUROS.....	57
REFERÊNCIAS.....	58
APÊNDICE A - Estatísticas descritivas das variáveis dos ativos.....	61
APÊNDICE B - Distribuições de carteira das estratégias.....	68
APÊNDICE C - Gráficos do histórico de <i>drawdown</i> das estratégias e <i>benchmarks</i>.....	71

INTRODUÇÃO

A construção de um portfólio de investimentos eficiente está diretamente relacionada à seleção criteriosa de ativos e à adoção de estratégias que maximizem o retorno ajustado ao risco. No Brasil, apesar do crescente interesse pelo mercado financeiro, o número de investidores pessoa física ainda é relativamente baixo, representando apenas 11% da população (B3, 2024). Esse cenário é influenciado pela falta de conhecimento financeiro e pela incerteza econômica do país, fatores que dificultam a escolha de ativos sólidos e a definição de estratégias de investimento de longo prazo (PIXLEY, 2004).

A volatilidade e instabilidade intrínsecas aos mercados financeiros representam uma preocupação central, especialmente para estratégias focadas no curto prazo. Em contrapartida, a literatura frequentemente aponta o investimento com horizonte de longo prazo como uma abordagem para mitigar os riscos associados a essas flutuações diárias. Nesse sentido, as estratégias de longo prazo, como o *buy and hold*, são uma forma de o investidor se beneficiar do crescimento econômico estrutural, minimizando o impacto do ruído de mercado (SILVA SANTOS; MARQUES, 2021).

Diante desse cenário de incerteza, uma das estratégias mais consolidadas para o investidor individual é o *buy and hold* que significa “comprar” um ativo e “segurar”, que destaca-se por ser uma forma de se beneficiar com os rendimentos, resultados e valorização que as empresas porventura apresentarão no futuro, como aponta Damodaran (1997). Esta filosofia de investimento consiste na aquisição de ativos de alta qualidade e resilientes com a intenção de mantê-los na carteira por longos períodos, ignorando movimentos de curto prazo, cuja previsibilidade é frequentemente apontada como baixa em comparação com estratégias de longo prazo (HARYANTO, 2024). A sua premissa fundamental é que o valor de empresas sólidas tende a crescer consistentemente ao longo do tempo, e que tentar prever os movimentos de mais curto prazo, é muitas vezes uma tarefa ineficiente e custosa para o investidor não profissional.

A previsão de mercados financeiros é amplamente reconhecida como um problema desafiador, dada a natureza de suas séries temporais, que são dinâmicas, não-lineares e caóticas (DEBOECK, 1994), frequentemente influenciadas por fatores macroeconômicos e pelo comportamento dos investidores. Diante dessa complexidade, um crescente corpo de pesquisa tem se dedicado a aplicar modelos de *Machine Learning* (ML), um subcampo da Inteligência Artificial (IA) focado no desenvolvimento de modelos capazes de aprender padrões a partir de dados e melhorar o desempenho em tarefas específicas (EL NAQA;

MURPHY, 2015). Essa abordagem tem se mostrado eficaz na análise e previsão de ativos, buscando capturar padrões que métodos tradicionais podem não identificar (MA; HAN; WANG, 2021).

Diversas arquiteturas têm demonstrado resultados promissores neste campo. Modelos de *ensemble* (conjunto) como o *Random Forest* (RF) (BALLINGS et al., 2015), por exemplo, são empregados para previsão de tendências de preços de ações. Outros modelos envolvendo técnicas de *deep learning* têm sido aplicados com sucesso na otimização de portfólios baseada na predição de retornos, como demonstram Ma et al. (2021).

No vasto universo de mercado de ações, um desafio fundamental para o investidor é a alocação de capital entre uma grande quantidade de ativos disponíveis. Uma abordagem comum para gerenciar essa complexidade é a pré-seleção de ativos, um processo no qual o universo de investimentos é deliberadamente restrito a um conjunto de empresas que atendem aos critérios estabelecidos (SILVA et al., 2024).

Isto posto, o presente estudo propõe o desenvolvimento de um *framework* de investimento sistemático que une a filosofia de *buy and hold* com a inteligência de técnicas de ML para otimizar a alocação em ativos perenes. A metodologia parte de uma pré-seleção criteriosa de empresas que pertencem simultaneamente às carteiras teóricas dos índices BOVESPA (IBOV) e Índice de Dividendos (IDIV), restringindo o universo de análise a companhias de alta liquidez, estabilidade e histórico de dividendos. Sobre essa base de ativos, em vez de uma abordagem puramente passiva, emprega-se um modelo de RF para qualificar a estratégia de acumulação, uma aplicação sustentada por trabalhos que demonstram o potencial do modelo para otimizar os momentos de entrada no mercado (ZHENG et al., 2024).

O modelo é treinado para identificar os momentos potencialmente mais favoráveis para a entrada no mercado, otimizando a alocação de capital ao capturar padrões não lineares que métodos tradicionais poderiam ignorar. Dessa forma, a contribuição deste trabalho é criar uma estratégia de *buy and hold* otimizada, que busca potencializar os retornos de longo prazo através de entradas mais estratégicas no mercado.

2 OBJETIVOS

2.1 Objetivo geral

Desenvolver e avaliar um *framework* de alocação dinâmica de capital para um portfólio de ativos perenes, pré-selecionados pela intersecção do índice IBOV e IDIV, utilizando RF como modelo preditivo com o objetivo de aprimorar a estratégia de *buy and hold*.

2.2 Objetivos específicos

- Estruturar a base de dados do estudo, selecionando um universo de ativos perenes a partir da intersecção dos índices IBOV e IDIV e calculando um conjunto abrangente de indicadores técnicos e informações históricas para cada ativo.
- Desenvolver e otimizar um modelo de classificação binária para cada ativo, utilizando RF, para identificar oportunidades de compra com alto potencial de valorização.
- Implementar um *framework* de *backtesting* de portfólio para simular o desempenho de múltiplas estratégias de alocação de capital e gestão de risco que são guiadas pelo modelo preditivo.
- Avaliar o valor agregado do modelo, comparando o desempenho das estratégias ativas contra *benchmarks* passivos (Índices do Mercado Financeiro) por meio de métricas quantitativas de retorno e risco.

3 ASPECTOS TEÓRICOS

Os presentes aspectos teóricos apresentam o mercado financeiro e os conceitos que envolvem as ações. Assim, como a sinergia entre técnicas computacionais avançadas e a análise de mercados financeiros. Abordam-se os princípios de ML, com foco no algoritmo RF, e os indicadores de análise técnica e macroeconômicas que servem como variáveis de entrada para o modelo. Adicionalmente, são detalhadas as metodologias de *web scraping* para a captura de dados e o *backtesting* para validação da estratégia de investimento proposta neste trabalho. Por fim, são apresentados os *benchmarks* do mercado financeiro que servem como comparativos para as estratégias testadas e explicados os conceitos das métricas de risco, importantes para a análise dos resultados.

3.1 Mercado de ações

O mercado de ações é um segmento do mercado de capitais onde são negociadas ações de companhias de capital aberto. Este mercado organiza-se fundamentalmente em duas etapas: o mercado primário e o secundário. Entende-se por mercado primário a etapa em que os valores mobiliários (como as ações) são emitidos pelas companhias e subscritos (comprados) pelos investidores pela primeira vez, geralmente através de uma Oferta Pública Inicial (IPO). Subsequentemente, esses valores passam a ser negociados entre os investidores no mercado secundário, proporcionando liquidez aos ativos (ASSAF NETO, 2012).

3.2 Ações

As ações, principal ativo negociado neste mercado, são títulos que representam uma fração do capital social de empresas de capital aberto que passaram pelo processo de IPO. Elas conferem aos seus detentores, conhecidos como acionistas, direitos e deveres societários (BTG PACTUAL, 2025).

No mercado de ações brasileiro, os ativos são classificados principalmente em dois tipos: Ações Ordinárias (ON) e Ações Preferenciais (PN). Essa classificação é refletida no final do código de negociação (*ticker*) do ativo (BTG PACTUAL, 2025).

- Ações ordinárias: Conhecidas como ações de governança, conferem ao acionista o direito a voto nas assembleias da empresa. Suas características beneficiam investidores

que buscam participação nas decisões corporativas. O *ticker* de uma ação ordinária geralmente termina com o número 3.

- Ações preferenciais: Possuem características mais voltadas ao investidor que busca primariamente o retorno financeiro (como dividendos), geralmente com prioridade no recebimento destes em relação às ações ON, mas sem direito a voto ou com voto restrito. O código de uma ação preferencial geralmente termina com o número 4. Existem também outras classes de ações preferenciais menos comuns, identificadas pelos números 5, 6, 7 ou 8 no final do *ticker*.

Além desses dois tipos básicos, existe também a *Unit*, um ativo composto por um "pacote" que inclui um número definido de ações ordinárias e preferenciais, negociado como um único título no mercado. O *ticker* de uma *Unit* geralmente termina com o número 11.

3.3 *Machine learning* aplicado ao mercado financeiro

O *machine leaning* é um subcampo da IA focado no desenvolvimento de modelos capazes de aprender padrões a partir de dados e melhorar seu desempenho em tarefas específicas (EL NAQA; MURPHY, 2015).

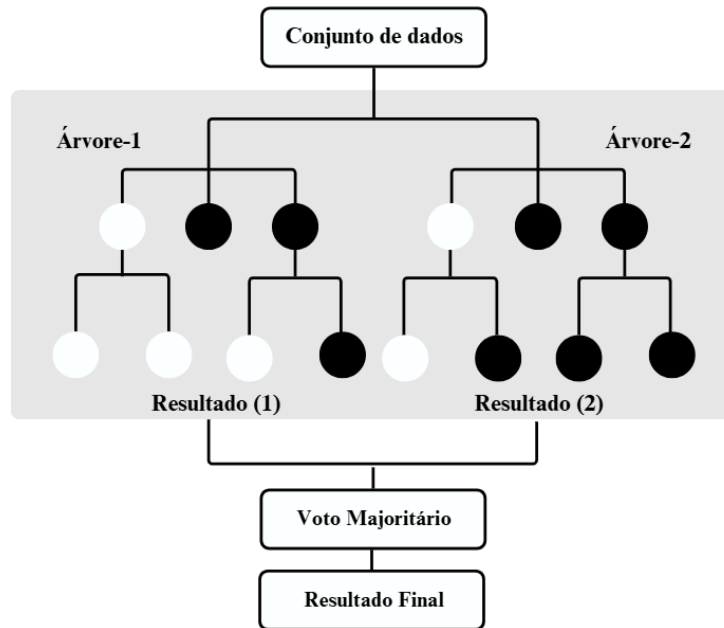
No contexto financeiro, caracterizado por seu grande volume de dados e complexidade, as técnicas de ML têm se tornado uma ferramenta poderosa para capturar padrões valiosos que métodos estatísticos tradicionais muitas vezes não conseguem. Os estudos de Ma et al. (2021), por exemplo, demonstram que modelos de aprendizado de máquina e *deep learning* possuem capacidade preditiva superior na previsão de retorno de ativos, pois, essas técnicas são aptas a capturar padrões não lineares complexos.

Suas aplicações são vastas e incluem a análise de séries temporais para previsão de tendências, otimização de portfólios, alocação dinâmica de ativos, gestão de risco e até análise de sentimento baseados em notícias e mídias sociais (AZIZ et al., 2022).

3.4 *Random forest*

Floresta aleatória ou RF, é um algoritmo de aprendizado de máquina do tipo *ensemble*, que constrói múltiplas árvores de decisão durante sua fase de treinamento. Para tarefas de classificação o modelo agrega as previsões de todas as árvores individuais e determina a classe final por meio de uma votação majoritária (SHAIK; SRINIVASAN, 2018).

Figura 1 - Ilustração da árvore de decisão.



Fonte: Autores (2025)

Sua principal vantagem reside na robustez contra o sobreajuste (*overfitting*), resultado da combinação de múltiplos modelos. Adicionalmente, o algoritmo é eficaz para lidar com dados de alta dimensionalidade e permite mensurar a importância relativa de cada variável (*feature*) para a construção do modelo preditivo, o que torna o processo de seleção de variáveis para um determinado modelo menos custoso (SHAIK; SRINIVASAN, 2018).

No contexto da previsão de tendências no mercado financeiro, ele é amplamente utilizado por sua capacidade de capturar relações não-lineares complexas e lidar com a alta dimensionalidade dos dados históricos. Sua eficácia é corroborada por estudos como o de Zheng et al. (2024), que destaca o potencial do modelo RF na previsão de tendências de preços de ações, superando os métodos convencionais de análise de tendências.

3.5 Web scraping

Uma das técnicas de aquisição de dados para este estudo foi realizada por meio de *web scraping*, uma técnica de extração automatizada que converte informações não estruturadas de páginas *web* em um formato estruturado e analisável. Este processo consiste, fundamentalmente, em utilizar *scripts* para requisitar e analisar o código-fonte de um *website*,

extraindo os dados de interesse para um formato tabular (NAVLANI; FANDANGO; IDRIS, 2021).

A aplicação desta técnica automatizada representou uma etapa metodológica fundamental para o presente trabalho, visto que a coleta manual de um vasto conjunto de informações financeiras, dispersas em múltiplos portais, seria impraticável em termos de escala e tempo, além de ser suscetível a erros operacionais. Portanto, o *web scraping* foi a técnica que viabilizou a construção do conjunto de dados primário, permitindo a obtenção sistemática dos indicadores fundamentalistas que subsidiaram a seleção do universo de empresas perenes aqui analisadas (NAVLANI; FANDANGO; IDRIS, 2021).

3.6 Indicadores de análise técnica

A importância dos indicadores de mercado é um pilar da análise técnica moderna. Conforme ressalta um dos textos do proeminente autor da área, Achelis e Achelis (2001):

Os indicadores de mercado adicionam uma profundidade significativa à análise técnica, pois contêm muito mais informação do que preço e volume... a analogia sendo ‘todos os barcos sobem na maré alta.’ (ACHELIS; ACHELIS, 2001, p. 38, tradução nossa).

Os indicadores de análise técnica são amplamente utilizados para interpretar os movimentos do mercado financeiro e auxiliar na tomada de decisões sobre ativos (OĞUZ; EYERCIOĞLU; AVCI, 2019). Eles são calculados com base nos dados *Open, High, Low, Close e Volume* (OHLVC) de um determinado período. Para este trabalho, foi selecionado um conjunto de indicadores de diferentes categorias baseados em estudos anteriores (KARA; BOYACIOĞLU; BAYKAN, 2011; PATEL et al., 2015), detalhados a seguir.

3.6.1 Indicadores de tendência e volatilidade

Estes indicadores buscam identificar a direção predominante do mercado, geralmente utilizam a mesma escala dos preços do ativo em análise.

- Média Móvel Simples (SMA): Determinada pela média aritmética dos preços de fechamento de um ativo em um determinado período de tempo. Usava para suavizar

grandes flutuações de preço em momento de euforia ou descrença do mercado. Sua fórmula é:

Equação 1 - Cálculo do SMA

$$SMA = \frac{1}{n} \sum_{i=1}^n P_i \quad (1)$$

Em que:

P_i = é o preço de fechamento no período i ; n = é o número total de períodos.

- Média Móvel Exponencial (EMA): É semelhante ao SMA, mas atribui maior peso aos preços mais recentes, o que torna mais responsiva às mudanças de mercado.

Equação 2 - Cálculo do EMA

$$EMA_t = (P_t \times k) + (EMA_{t-1} \times (1 - k)) \quad (2)$$

Em que:

P_t = é o preço de fechamento no tempo t ; EMA_{t-1} = é a média móvel exponencial do período anterior; k = é o fator de suavização, calculado como $k = \frac{2}{n+1}$, sendo n o número de períodos.

- Média Móvel Ponderada (WMA): Similar à SMA, ela também calcula um preço médio para um determinado período, porém atribui maior peso aos preços mais recentes. Essa característica a torna mais sensível às novas informações e mudanças de curto prazo no movimento do preço do que a SMA. Para um período n , o preço mais recente recebe peso n , o anterior $n-1$, e assim sucessivamente. A fórmula é:

Equação 3 - Cálculo do WMA

$$WMA = \frac{\sum_{i=1}^n P_i \cdot i}{\sum_{i=1}^n i} = \frac{n \cdot P_n + (n-1) \cdot P_{n-1} + \dots + 1 \cdot P_1}{\frac{n(n+1)}{2}} \quad (3)$$

Em que:

P_i = é o preço no período i (com P_n sendo o mais recente); n = é o número total de períodos

- Convergência e Divergência de Médias Móveis (MACD): É um indicador de momento que segue a tendência e mostra a relação entre duas médias móveis exponenciais de preços, uma de período mais curto e outra de período longo, mais comum sendo 12 e 26 dias respectivamente. A linha MACD é calculada pela seguinte fórmula:

Equação 4 - Cálculo do MACD

$$MACD = [(EMA_{12}) - (EMA_{26})] \quad (4)$$

Em que:

EMA_{12} = é a Média Móvel Exponencial de 12 períodos; EMA_{26} = é a Média Móvel Exponencial de 26 períodos.

- Oscilador de Acumulação/Distribuição (A/D): Este é um indicador de volume que mede o fluxo de dinheiro cumulativo para dentro ou para fora de um ativo. Embora baseado em volume, ele é frequentemente usado para confirmar a força de uma tendência de preço. Um A/D crescente sugere que o ativo está sendo acumulado (comprado), reforçando uma tendência de alta (PATEL et al., 2015). Ele é calculado em três etapas:

- Multiplicador de Fluxo de Dinheiro (MFM):

Equação 5 - Cálculo do MFM

$$MFM = \frac{(Fechamento - Mínima) - (Máxima - Fechamento)}{Máxima - Mínima} \quad (5)$$

- Volume de Fluxo de Dinheiro (MFV):

Equação 6 - Cálculo do MFV

$$MFV = MFM \times Volume \quad (6)$$

- Linha de Acumulação/Distribuição (A/D):

Equação 7 - Cálculo do A/D

$$A/D_t = A/D_{t-1} + MFV_t \quad (7)$$

Em que:

A/D_t é o valor atual da linha A/D e A/D_{t-1} é o valor do período anterior.

- Índice de Canal de Commodities (CCI): O CCI mede a diferença entre a variação de preço de um ativo e sua variação média histórica. Leituras positivas indicam que os preços estão acima da média, portanto, demonstrando força, enquanto leituras negativas mostram fraqueza. Ele é útil tanto para identificar tendência quanto para sinalizar níveis de sobrecompra e sobrevenda. O cálculo é realizado da seguinte forma:

- Preço Típico (PT):

Equação 8 - Cálculo do PT

$$PT = \frac{Máxima + Mínima + Fechamento}{3} \quad (8)$$

- Média Móvel Simples do Preço Típico (SMA PT): Calcula-se uma SMA de n períodos do PT.

- Desvio Médio (DM):

Equação 9 - Cálculo do DM

$$DM = \frac{1}{n} \sum_{i=1}^n |PT_i - SMA PT| \quad (9)$$

- Por fim o cálculo do CCI:

Equação 10 - Cálculo do CCI

$$CCI = \frac{PT - SMA PT}{0,015 \times DM} \quad (10)$$

Em que:

O fator constante de 0,015 é usado para normalizar o indicador, fazendo com que a maioria dos valores oscile dentro do intervalo de -100 a +100 (PATEL et al., 2015).

3.6.2 Osciladores de momento

Diferentemente dos indicadores de tendências, os osciladores de momento são projetados para medir a velocidade e a força dos movimentos de preço, geralmente operando dentro de uma faixa definida. Sua principal utilidade reside na capacidade de sinalizar quando um ativo pode estar em condições de "sobrecompra" ou "sobrevenda". Esses termos descrevem situações de mercado em que o preço de um ativo se moveu de forma acentuada em uma direção. Assim, ao identificar esses pontos, os osciladores indicam uma probabilidade crescente de uma futura correção de preço ou de uma reversão da tendência vigente.

- Índice de Força Relativa (RSI): Indica a velocidade e a variação dos movimentos dos preços. Ajuda a identificar condições de sobrecompra e sobrevenda. O cálculo é feito em duas etapas. Primeiro, calcula-se a Força Relativa (RS), que é a razão entre a média dos ganhos e a média das perdas em um determinado período:

Equação 11 - Cálculo do RS

$$RS = \frac{\text{Média de Ganhos}}{\text{Média de Perdas}} \quad (11)$$

Em seguida, com o valor de RS, o RSI é calculado e normalizado para uma escala de 0 a 100, através da seguinte fórmula:

Equação 12 - Cálculo do RSI

$$RSI = 100 - \frac{100}{1 + RS} \quad (12)$$

Em que:

A média de ganhos e a média de perdas são calculadas com base nos preços de fechamento do período de tempo analisado.

- **Momento (MOM):** Este é o indicador que mede a taxa de variação do preço de um ativo. Ele compara o preço de fechamento atual com o preço de fechamento de n períodos atrás. Um valor positivo indica que o preço está subindo, já um valor negativo indica que está caindo. Sua fórmula é:

Equação 13 - Cálculo do MOM

$$MOM = P_t - P_{t-n} \quad (13)$$

Em que:

$p_t =$ é o preço de fechamento atual e $P_{t-n} =$ é o preço de fechamento n períodos atrás.

- **Oscilador Estocástico (%K e %D):** Ele compara o preço de fechamento de um ativo com o intervalo de seus preços máximo e mínimo ao longo de um determinado período. Ele baseia-se na premissa de que em uma tendência de alta, os preços de fechamento de um ativo tendem a se aproximar das máximas do período e em uma tendência de baixa, tendem a se aproximar das mínimas. O indicador é composto por duas linhas:

- Linha %K: Representa o valor atual do osculado.

Equação 14 - Cálculo do %K

$$\%K = \frac{Fechamento_{Atual} - Mínima_n}{Máxima_n - Mínima_n} \times 100 \quad (14)$$

- Linha %D: Corresponde a uma média móvel de n períodos da linha %K, atuando como um sinal suavizado para reduzir a volatilidade e gerar sinais de negociação.

Equação 15 - Cálculo do %D

$$\%D = SMA_m(\%K) \quad (15)$$

Em que:

$Mínima_n$ e $Máxima_n$ são, respectivamente, o menor e o maior preço registrado nos últimos n períodos.

- William's %R: Esse indicador funciona de forma similar ao oscilador estocástico, contudo, sua escala é invertida (de -100 a 0) e não possui uma linha de sinal. O %R mede o nível do preço de fechamento em relação ao intervalo entre a máxima e mínima de um determinado período, sendo altamente eficaz para identificar condições de sobrecompra e sobrevenda. Sua fórmula é:

Equação 16 - Cálculo do %R

$$\%R = \frac{Máxima_n - Fechamento_{Atual}}{Máxima_n - Mínima_n} \times (-100) \quad (16)$$

Em que:

$Máxima_n$ e $Mínima_n$ são o maior e o menor preço registrados nos últimos n períodos.

A combinação de indicadores técnicos com algoritmos de aprendizado de máquina pode fornecer a previsão mais bem-sucedida sobre a direção do movimento dos preços no

mercado financeiro (OĞUZ; EYERCIOĞLU; AVCI, 2019). Esses indicadores desempenharam um papel crucial ao fornecer *inputs* para o modelo de ML, permitindo a análise de padrões históricos e otimização do desempenho da estratégia.

3.7 Indicadores macroeconômicos

Além dos indicadores técnicos do ativo, os modelos de ML podem ter sua capacidade preditiva enriquecida ao receberem informações sobre o contexto macroeconômico (DARAPANENI et al., 2022). Fatores exógenos, como o sentimento geral do mercado ou o preço de commodities, influenciam diretamente o valor das empresas. Por essa razão, indicadores macroeconômicos relevantes foram incluídos no treinamento dos modelos.

Nesse contexto, certos indicadores se destacam pela sua capacidade de capturar a dinâmica da economia e seu impacto setorial. O IBOV, por exemplo, funciona como um termômetro da confiança dos investidores na economia brasileira. A taxa de câmbio (USD/BRL), por sua vez, é um vetor de transmissão crucial, sua variação afeta a competitividade de empresas exportadoras. Adicionalmente, os preços de *commodities*, como o barril de petróleo e o minério de ferro, são de vital importância para empresas baseadas em economia setorial.

3.8 Backtesting

Backtesting é uma técnica amplamente utilizada no mercado financeiro para avaliar a eficácia de uma estratégia de investimento com base em dados históricos (CAMPBELL, 2005). Essa prática consiste em simular como a estratégia teria se comportado no passado, utilizando dados reais de mercado para validar seu desempenho.

No contexto deste projeto, o *backtesting* foi essencial para mensurar a performance do modelo preditivo e da carteira teórica, permitindo ajustes na parametrização dos indicadores e nas decisões de alocação de ativos.

3.9 Benchmarks do mercado financeiro

Para avaliar a eficácia das estratégias ativas utilizando RF, é essencial compará-las com *benchmarks* de mercado capazes de representar o sentido e retorno base do mercado ou de um setor.

- Índice Bovespa (IBOV): É o principal indicador de desempenho do mercado de ações brasileiro. A comparação com o IBOV (neste trabalho, representado pelo *Exchange Trade Fund* (ETF) BOVA11) responde se a estratégia de RF foi capaz de superar o retorno médio do mercado nacional.
- Índice de Dividendos (IDIV): É um índice da B3 que mede o desempenho das ações de empresas que se destacam por pagar dividendos. É um *benchmark* crucial, pois o portfólio de ativos perenes foi selecionado justamente com base nesse critério (interseção IBOV e IDIV), sendo o DIVO11 seu ETF correspondente.
- S&P 500: É o principal indicador do mercado de ações dos Estados Unidos. Usado como *benchmark* de performance global, sua variação (representada pelo ETF IVVB11) reflete tanto a performance do mercado externo quanto o impacto da variação cambial do dólar (USD/BRL).
- Certificado de Depósito Interbancário (CDI): É a taxa de juros que serve como principal *benchmark* de renda fixa no Brasil. Ela representa o "custo de oportunidade" do capital, ou seja, o retorno que se obteria sem assumir os riscos do mercado brasileiro de ações.

3.10 Métricas de risco e retorno

A avaliação de uma estratégia de investimento transcende a simples análise do retorno total, exige uma análise aprofundada dos riscos incorridos para alcançar tal resultado. Enquanto o retorno quantifica a lucratividade, as métricas de risco mensuram a incerteza e a magnitude das perdas potenciais ao longo do período analisado. Em nossa simulação de *trading*, usamos o *drawdown* máximo e a volatilidade anualizada como métricas de risco.

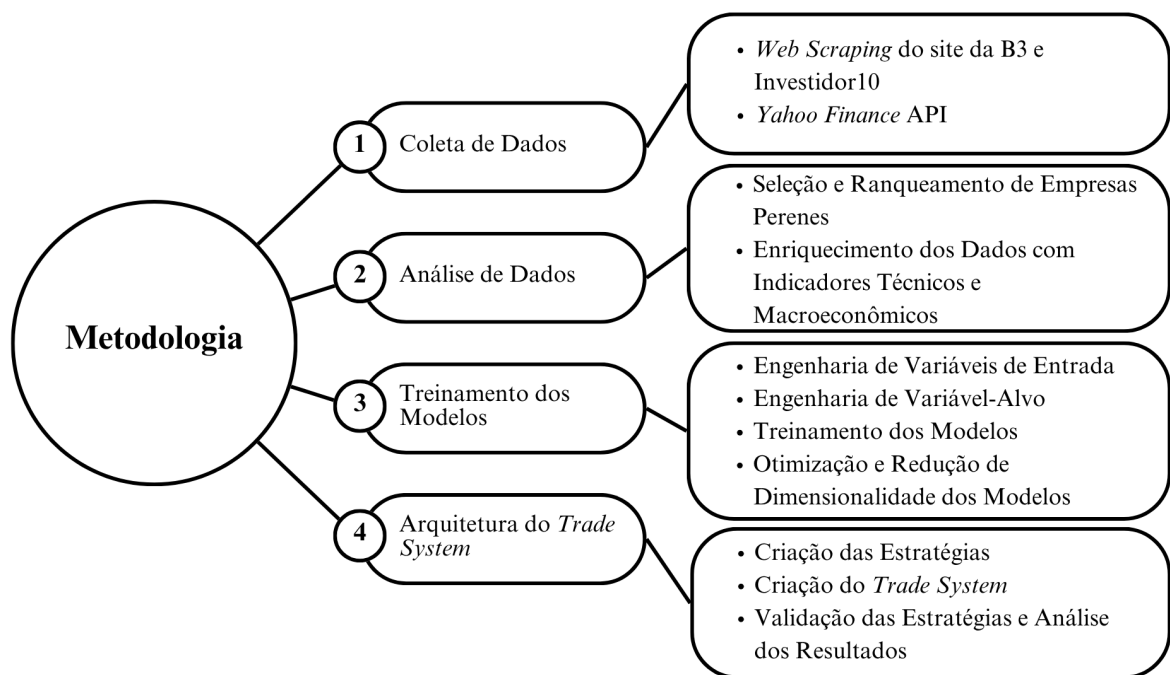
Drawdown é a medida da queda de um portfólio a partir de um pico de valor até seu ponto mais baixo durante um período específico, antes que um novo pico seja atingido. Ele quantifica a maior perda percentual acumulada que um investidor teria experimentado se tivesse entrado no pico e saído no fundo. O *drawdown* máximo é o maior valor de *drawdown* observado durante todo o período de análise (GOLDBERG; MAHMOUD, 2017).

Já a volatilidade é a medida estatística de dispersão dos retornos de um portfólio. Ele quantifica a intensidade das oscilações de preço. A volatilidade anualizada projeta essa métrica diária para o período de um ano (multiplicando-a pela raiz quadrada de 252 pregões), permitindo que o risco de diferentes ativos seja comparado na mesma base.

4 METODOLOGIA

A metodologia adotada neste projeto seguiu uma abordagem estruturada em quatro etapas principais: coleta de dados, análise de dados, treinamento dos modelos e arquitetura do *trade system*. O objetivo foi garantir uma análise consistente e alinhada à estratégia de *buy and hold* aplicada a ativos perenes. Segue abaixo, na Figura 2, o esquema dos processos realizados na fase de metodologia da pesquisa.

Figura 2 - Processos da metodologia.



Fonte: Autores (2025)

4.1 Coleta de dados

Inicialmente, utilizou-se a *Application Programming Interface* (API) do *Yahoo Finance*, um recurso que permite a comunicação automatizada entre sistemas, para obter informações como preços de abertura, fechamento, máximas e mínimas, além do volume de negociação. Entretanto, devido a limitações dessa API no fornecimento de históricos extensos, foram incorporados dados complementares obtidos via *Web Scraping* de plataformas como B3, *Status Invest* e Investidor10 (B3, 2025; STATUS INVEST, 2025; INVESTIDOR10, 2025).

4.2 Análise de dados

A segunda etapa consistiu no processamento e preparação dos dados, utilizando a biblioteca Pandas, amplamente utilizada para limpeza e formatação de dados (NAVLANI; FANDANGO; IDRIS, 2021). A partir disso, a metodologia seguiu as seguintes fases:

4.2.1 Seleção e ranqueamento de empresas perenes

Esta etapa consistiu na definição e no ranqueamento de um universo de “ações perenes”. Para isso, foram identificadas as empresas cujas ações compunham simultaneamente as carteiras teóricas do IBOV e do IDIV. A partir dessa interseção, e com o objetivo de concentrar a análise nos ativos de maior relevância e liquidez, o universo de estudo foi deliberadamente restrito aos 10 ativos com a maior pontuação no ranqueamento, aplicando-se ainda critérios adicionais: não selecionar múltiplos códigos da mesma empresa e excluir ativos que não estivessem mais sendo negociados no mercado de ações brasileiro até o presente momento.

Para quantificar a relevância de cada ativo, foi criado um *score* customizado. Inicialmente, os ativos foram ranqueados separadamente segundo sua participação percentual no IBOV e no IDIV. A média aritmética desses dois *rankings* gerou um *rank* médio, e o inverso desse valor foi usado para criar um *score* bruto, onde um número maior indica mais relevância. Por fim, o *score* foi normalizado para se tornar um peso relativo percentual dentro do grupo de ativos selecionados.

O resultado completo deste processo de ranqueamento, que fundamentou a seleção e a ponderação dos ativos para as estratégias subsequentes, está consolidado na Tabela 1. Ela detalha os *rankings* individuais, o *score* bruto e o peso relativo final para cada um dos ativos perenes selecionados.

Tabela 1 - Lista dos ativos perenes selecionados, ordenados pela participação perenes.

(continua)

Ativo	Part. IBOV (%)	Part. IDIV (%)	Posição IBOV	Posição IDIV	Score	Part. Perenes (%)
PETR4	5,84	4,79	2	5	0,286	9,93
BBDC4	4,13	4,67	3	7	0,200	6,95

Ativo	Part. IBOV (%)	Part. IDIV (%)	Posição IBOV	Posição IDIV	Score	Part. Perenes (%)
VALE3	12,40	5,00	7	3	0,200	6,95
BBAS3	2,77	4,86	6	4	0,200	6,95
CMIG4	0,96	5,86	9	1	0,200	6,95
ITUB4	8,10	3,93	1	10	0,182	6,32
PETR3	3,71	4,45	4	8	0,167	5,79
BBSE3	0,90	5,50	10	2	0,167	5,79
BBDC3	1,01	4,75	8	6	0,143	4,97
ITSA4	3,04	4,26	5	9	0,143	4,97
TIMS3	0,87	3,77	11	11	0,091	3,16
KLBN11	0,63	3,03	12	12	0,083	2,90
EGIE3	0,46	2,77	14	13	0,074	2,58
CMIN3	0,45	2,71	15	14	0,069	2,40
ISAE4	0,44	2,70	16	15	0,065	2,24
CXSE3	0,40	2,44	17	16	0,061	2,11
TAE11	0,37	2,26	18	17	0,057	1,99
CPFE3	0,35	2,12	19	18	0,054	1,88
GOAU4	0,31	1,88	20	19	0,051	1,78
MBRF3	0,56	0,92	13	28	0,049	1,70
CSNA3	0,31	1,87	21	20	0,049	1,70
FLRY3	0,29	1,77	22	21	0,047	1,62
DIRR3	0,25	1,50	23	22	0,044	1,55
POMO4	0,23	1,42	24	23	0,043	1,48
CURY3	0,21	1,28	25	24	0,041	1,42
BRAP4	0,21	1,26	26	25	0,039	1,36
AURE3	0,16	0,95	27	26	0,038	1,31
RECV3	0,15	0,93	28	27	0,036	1,26

Fonte: Autores (2025)

Aplicando as diretrizes e os critérios definidos, os 10 ativos selecionados para compor a carteira foram: PETR4, BBDC4, VALE3, BBAS3, CMIG4, ITUB4, BBSE3, ITSA4, TIMS3 e KLBN11. Ativos como PETR3 e BBDC3 não fizeram parte do conjunto final para evitar redundância com PETR4 e BBDC4, priorizando-se estes últimos por sua característica mais voltada ao investidor pessoa física individual.

4.2.2 Enriquecimento dos dados com indicadores técnicos e macroeconômicos

Com o universo de ações perenes definido pela etapa de ranqueamento, o passo seguinte consistiu no enriquecimento dos dados para cada empresa selecionada, foram incorporadas as séries históricas de ativos macroeconômicos correlacionados. O Quadro 1 detalha os ativos macroeconômicos que foram selecionados para enriquecer o modelo preditivo de cada ação perene.

Quadro 1 - Lista de ativos macroeconômicos.

Ativo	Ativos Macroeconômicos
PETR4	Índice BOVESPA, Taxa de câmbio (USD/BRL), Preço do Barril de Petróleo
BBDC4	Índice BOVESPA e Taxa de câmbio (USD/BRL)
VALE3	Índice BOVESPA, Taxa de câmbio (USD/BRL), Preço do Minério de Ferro
BBAS3	Índice BOVESPA e Taxa de câmbio (USD/BRL)
CMIG4	Índice BOVESPA e Taxa de câmbio (USD/BRL)
ITUB4	Índice BOVESPA e Taxa de câmbio (USD/BRL)
BBSE3	Índice BOVESPA e Taxa de câmbio (USD/BRL)
ITSA4	Índice BOVESPA e Taxa de câmbio (USD/BRL)
TIMS3	Índice BOVESPA e Taxa de câmbio (USD/BRL)
KLBN11	Índice BOVESPA e Taxa de câmbio (USD/BRL)

Fonte: Autores (2025)

Pode-se notar no Quadro 1 que a seleção dos indicadores macro foi direcionada setorialmente. Para empresas de commodities ou com forte relação com elas, como PETR4 e VALE3, foram adicionados os preços dos seus respectivos produtos. De forma geral, o Índice

Bovespa e a Taxa de câmbio (USD/BRL) foram adicionados a todos os ativos. Isso se justifica pois o IBOV funciona como um termômetro para o sentimento do mercado de ações brasileiro, enquanto a taxa de câmbio impacta diretamente empresas com receitas ou dívidas dolarizadas.

A incorporação das variáveis macroeconômicas foi fundamentada na premissa de que a combinação de diferentes fontes de informação enriquece o espaço de características e melhora a precisão dos modelos preditivos, permitindo uma compreensão mais completa da dinâmica do mercado (BAMPOUNTIS et al., 2024).

Foram assim, calculados indicadores técnicos de múltiplas categorias, como tendência e volatilidade (SMA, WMA, MACD, A/D, CCI), osciladores de momento (RSI, MOM, Estocástico, William's %R) do ativo principal e dos ativos macroeconômicos correlacionados a cada um deles. A seleção desses indicadores seguiu práticas consolidadas na análise técnica financeira (ACHELIS; ACHELIS, 2001; OĞUZ; EYERCIOĞLU; AVCI, 2019).

A Tabela 2 apresenta, como exemplo, as estatísticas descritivas para o conjunto de dados de treinamento e otimização do ativo PETR4, abrangendo o período de 2010 a 2019.

Tabela 2 - Estatísticas Descritivas das Variáveis (PETR4).

Variável	Média	Desvio Padrão	Mínimo	Mediana	Máximo
SMA_10	4,84	1,61	1,23	4,76	8,94
WMA_10	4,84	1,62	1,22	4,76	8,95
MACD	0,007	0,138	-0,570	0,010	0,492
A/D (em Milhões)	-1.838,60	1.101,00	-3.893,60	-1.777,54	-354,05
CCI	-2.705,34	1.599,97	-9.720,47	-2326,30	-315,16
RSI	50,86	12,82	14,05	50,57	84,70
MOM	0,016	0,463	-2,960	0,020	1,880
STOCHk	50,43	29,43	0,00	50,76	99,03
STOCHd	50,42	28,64	1,82	50,74	98,17
WILL	-49,55	31,50	-100	-49,23	0,00

Fonte: Autores (2025)

De forma análoga, as estatísticas descritivas para os conjuntos de dados de treinamento dos demais nove ativos do portfólio encontram-se no Apêndice A.

Ademais, foram criadas variáveis categóricas direcionais, para cada uma das variáveis calculadas anteriormente, aplicando a técnica de *Trend Deterministic Data* (TDA). Para o conjunto selecionado de indicadores, foram geradas novas variáveis que discretizaram a informação contínua em um sinal binário, representando a “opinião” daquele indicador sobre a tendência de mercado (PATEL et al. 2015).

A fundamentação para a aplicação desta técnica reside no fato de que, para algoritmos de classificação, a discretização de variáveis contínuas em representações binárias pode aprimorar a performance preditiva (PATEL et al. 2015). Essa transformação tende a tornar os padrões mais simples e compreensíveis para o classificador, além de aumentar a eficácia e torná-lo menos suscetível a ruído.

A consolidação dessas variáveis resultou em um conjunto de dados abrangente, permitindo que os modelos fossem treinados não apenas com base em padrões técnicos, mas também com informações sobre o ciclo econômico e seus impactos setoriais.

4.3 Treinamento dos modelos

Esta seção descreve o processo metodológico para a construção e otimização dos modelos preditivos de *machine learning*. A aplicação de técnicas de aprendizado de máquina para a previsão de mercados financeiros é uma tarefa reconhecidamente desafiadora, dada a natureza não estacionária e complexa das séries temporais de preços (PATEL et al., 2015).

4.3.1 Engenharia de variáveis de entrada

O conjunto de variáveis de entrada usado pelos modelos preditivos consistiu nos valores numéricos contínuos dos indicadores técnicos e macroeconômicos calculados na etapa anterior, juntamente com as suas versões discretizadas pelo TDA.

As variáveis contínuas, por suas diferentes características de escala, foram posteriormente submetidas a processos de padronização ou normalização, conforme a variável. Variáveis com distribuição bem definida, como os osciladores (RSI, Estocástico e William's %R), foram submetidas ao processo de normalização, enquanto as demais variáveis contínuas foram padronizadas.

Já as variáveis discretizadas foram inseridas nos modelos sem transformação, por sua característica binária. Todo esse processo foi encapsulado em um pré-processador da biblioteca *Scikit-Learn* (SCIKIT-LEARN DEVELOPERS, 2024).

4.3.2 Engenharia de variável-alvo

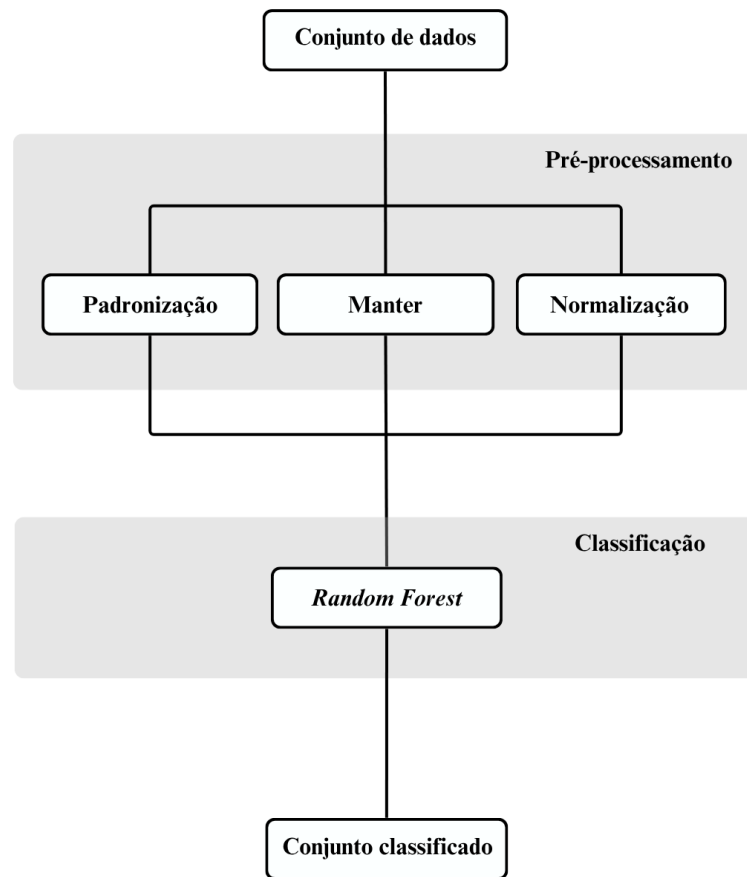
A variável-alvo foi desenhada para refletir diretamente a estratégia definida de *buy and hold* otimizada, buscando otimizar os aportes mensalmente. Primeiramente, para cada ponto de dado histórico, foi calculado o retorno percentual em um horizonte de tempo futuro de 30 dias. A escolha de um horizonte de 30 dias está diretamente alinhada à cadência de aportes buscada nas estratégias.

Dessa forma, para que o modelo desse foco apenas em oportunidades de valorização estatisticamente relevantes, foi definido um limiar de performance, que corresponde ao quartil superior (Q3) da distribuição dos retornos calculados anteriormente, representando os 25% melhores resultados históricos do ativo. Assim, com o limiar definido, a variável-alvo foi discretizada em duas classes, Classe 1 (comprar) e Classe 0 (manter). Uma consequência desta formulação é a criação de um conjunto de dados desbalanceados.

4.3.3 Treinamento dos modelos

Como primeiro passo, e por se tratar de um problema de série temporal, o conjunto de dados foi dividido cronologicamente na proporção de 90% para treinamento e 10% para teste. Esta é uma abordagem mandatória para prevenir o viés de *lookahead*, no qual o modelo seria treinado com informações do futuro para prever o passado (BOTACHE; PÉREZ; ARBELÁEZ, 2023). Adicionalmente, para garantir reprodutibilidade e assegurar que não haja vazamento de dados entre o processo de treinamento e teste, todo fluxo de modelagem foi encapsulado em um objeto *pipeline* da biblioteca *Scikit-Learn*. A utilização do *pipeline* é uma prática que assegura que as etapas de pré-processamento sejam ajustadas apenas para o conjunto de dados de treinamento e posteriormente para os conjuntos de teste e validação.

O *pipeline* desenvolvido para esse processo é composto por etapas sequenciais de pré-processamento e classificação, conforme ilustrado na Figura 3.

Figura 3 - Diagrama do *pipeline*.

Fonte: Autores (2025)

Após a definição do *pipeline*, realizou-se um processo iterativo para selecionar as variáveis de entrada e ajustar os hiperparâmetros, usando a validação cruzada para séries temporais para garantir a robustez. A otimização foi configurada para maximizar o *recall* da classe 1 (compra), uma escolha estratégica deliberada e alinhada à filosofia de *buy and hold*.

Para este portfólio, o custo de um falso negativo (perder uma compra) é considerado maior que o de um falso positivo (comprar equivocadamente). Essa assimetria de custos se justifica pela própria natureza do portfólio, que é formado por ativos perenes. Presume-se que o impacto de um falso positivo (compra não ideal) seja mitigado pela valorização de longo prazo da empresa. Em contrapartida, um falso negativo (perda da oportunidade de compra) representa um custo de oportunidade que compromete significativamente o retorno acumulado da estratégia.

Para tratar o forte desbalanceamento de classes, foi utilizado um recurso de ajuste de pesos das classes no próprio *RandomForestClassifier* da biblioteca *Scikit-Learn*.

4.3.4 Otimização e redução de dimensionalidade dos modelos

Como primeira etapa do processo de otimização, optou-se por uma fase de redução de dimensionalidade. Devido à etapa de engenharia de variáveis, o modelo passou a ter um número muito grande de variáveis de entrada, o que poderia causar sobreajuste (*overfitting*) e inferências equivocadas.

Para mitigar isso, realizou-se uma fase de eliminação recursiva de *features* com validação cruzada, a fim de identificar o subconjunto de variáveis com melhor performance. Foi utilizado nesse processo o *Recursive Feature Elimination with Cross-Validation* (RFECV) da biblioteca *Scikit-Learn* (SCIKIT-LEARN DEVELOPERS, 2024), que realiza a seleção baseando-se na importância de cada variável ao longo de sucessivos treinamentos e validações.

Em seguida, já com o conjunto otimizado de *features* selecionado, o modelo passou por um processo de otimização de hiperparâmetros. Esta otimização foi dividida em duas fases: uma busca ampla seguida de um ajuste fino.

Na primeira fase, foi utilizado o *RandomizedSearchCV* da biblioteca *Scikit-Learn* para explorar aleatoriamente um espaço de busca amplo e encontrar regiões de hiperparâmetros promissoras de forma computacionalmente eficiente. Para isso, foi definido um dicionário de parâmetros e distribuições que passariam por esse processo. Na Tabela 3 a seguir, os detalhes desse espaço de busca estarão dispostos.

Tabela 3 - Dicionário de Hiperparâmetros para o *RandomizedSearchCV*.

Hiperparâmetro	Opções
<i>class_weight</i>	<i>balanced</i>
<i>n_estimators</i>	300, 500, 700, 900, 1000
<i>max_depth</i>	3, 4, 5
<i>min_samples_leaf</i>	30, 50, 70
<i>criterion</i>	<i>entropy, gini</i>

Fonte: Autores (2025)

Por fim, para a fase de ajuste fino, foi utilizado o *GridSearchCV* da biblioteca *Scikit-Learn*. Os melhores hiperparâmetros *n_estimators*, *max_depth* e *min_samples_leaf* encontrados pelo processo anterior serviram como centro para uma nova grade de busca mais restrita.

Esta nova grade foi definida adicionando valores próximos para cada hiperparâmetro: ± 50 unidades para *n_estimators*, ± 1 unidade para *max_depth* e ± 5 unidades para *min_samples_leaf*. No caso do *class_weight*, foi mantido o valor '*balanced*', e o *criterion* foi fixado com o melhor valor encontrado na etapa anterior.

Ao final desse processo de otimização, um *pipeline* final e otimizado foi estabelecido para cada um dos ativos. O resultado é um conjunto de 10 modelos, que então servem como motores preditivos para o *trade system*.

Vale ressaltar que para gerar os sinais de previsão para o período de validação (1° de janeiro de 2020 a 31 de dezembro de 2024) utilizado no *trade system*, foi adotado uma estratégia de retreino anual. No início de cada ano, os modelos de RF foram treinados novamente utilizando uma janela deslizando dos 10 anos de histórico precedentes. Para fins de simplificação metodológica, este treino utilizou os mesmos hiperparâmetros e variáveis finais encontrados no processo de otimização. As previsões de cada ano foram então consolidadas em um conjunto de dados que alimenta as próximas etapas.

4.4 Arquitetura do *trade system*

Para a validação empírica das hipóteses, foi desenvolvido um sistema de negociação algorítmica, ou *trade system*, projetado para simular a aplicação de diversas estratégias em *backtesting*. A arquitetura do sistema é orientada a eventos, processando os dados de mercado de forma sequencial, dia a dia, a fim de simular a gestão de um portfólio que recebe aportes periódicos e executa ordens de compra com base em regras predefinidas.

4.4.1 Criação das estratégias

Implementou-se um desenho experimental para conduzir uma análise comparativa e robusta entre diferentes estratégias de investimento. Estas estratégias foram definidas em três eixos principais: estratégias de timing, métodos de alocação de capital e parâmetros de restrição operacional, conforme detalhado a seguir.

- Estratégias de *timing* (Gatilho de Compra): Define a lógica que dispara uma ordem de compra:
 - Estratégias baseadas nos modelos: As estratégias baseadas nos modelos foram desenhadas para testar duas lógicas distintas de atuação sobre os sinais. A primeira, *first_buy_signal*, adota uma abordagem ágil, executando a ordem de compra para um ativo assim que o primeiro sinal positivo é gerado no mês, limitando-se a um aporte mensal por ativo. A segunda, *last_day_top3_abs*, utiliza uma lógica de acumulação de evidências, aguardando até o último dia útil do mês para alocar o capital nos três ativos que receberam a maior contagem de sinais de compra ao longo do período.
- Métodos de alocação de capital: Determina como o capital disponível é distribuído entre os ativos selecionados para compra.
 - Divisão baseada nos sinais (*signal-weighted*): Este método foi desenhado para distribuir o capital de forma não-igualitária, priorizando os ativos com maior "força" de sinal. A lógica de ponderação, no entanto, se adapta ao gatilho de compra utilizado: para a estratégia *last_day_top3_abs*, a força é medida pela contagem de sinais de compra acumulados no mês, com o capital sendo distribuído proporcionalmente a essa contagem. Já para a estratégia *first_buy_signal*, como a compra é imediata, a força é determinada pelo *score* bruto do ativo, alocando mais capital para as empresas com maior *ranking* dentre as empresas perenes avaliadas.
- Parâmetros de restrição operacional:
 - Lote mínimo (*min_shares*): Impõe um lote mínimo de ações para evitar operações de valor irrisório. Os valores de lote mínimo foram determinados empiricamente para as estratégias.
 - Alocação máxima (*max_allocation*): Limita a concentração de capital em um único ativo a uma porcentagem definida do portfólio, atuando como uma

ferramenta de gestão de risco. Os valores percentuais da alocação foram definidos empiricamente.

Este desenho experimental multifacetado permite isolar o impacto de cada componente da estratégia e aferir, de forma quantitativa, se a inteligência gerada pelos modelos de ML oferece uma vantagem real sobre abordagens passivas ou mais simples de investimento. As configurações detalhadas de cada estratégia testada estão resumidas na Tabela 4, a seguir.

Tabela 4 - Lista de estratégias definidas para o processo de *backtesting*.

Estratégia	Gatilho de Compra	Alocação de Capital	Lote Min.	Alocação Máxima (%)
IA-FS-R1	<i>first_buy_signal</i>	<i>signal_weighted</i>	1	100
IA-FS-R2	<i>first_buy_signal</i>	<i>signal_weighted</i>	10	100
IA-FS-R3	<i>first_buy_signal</i>	<i>signal_weighted</i>	1	25
IA-FS-R4	<i>first_buy_signal</i>	<i>signal_weighted</i>	10	25
IA-L3-R1	<i>last_day_top3_abs</i>	<i>signal_weighted</i>	1	100
IA-L3-R2	<i>last_day_top3_abs</i>	<i>signal_weighted</i>	10	100
IA-L3-R3	<i>last_day_top3_abs</i>	<i>signal_weighted</i>	1	25
IA-L3-R4	<i>last_day_top3_abs</i>	<i>signal_weighted</i>	10	25

Fonte: Autores (2025)

4.4.2 Criação do *trade system*

A arquitetura do *trade system* é orientada a eventos, processando os dados de mercado de forma sequencial. A simulação opera sobre um conjunto de dados de validação, abrangendo o período de 1º de janeiro de 2020 a 31 de dezembro de 2024, que é temporalmente distinto e subsequente ao utilizado para o treinamento dos modelos (01/01/2010 a 31/12/2019), garantindo uma avaliação puramente *out-of-sample* (fora da amostra) e aferindo a real capacidade de generalização do sistema.

Para a execução do *backtesting*, foi definida uma política de aportes que simula um cenário realista e acessível para o investidor pessoa física individual no Brasil. Esta escolha

foi fundamentada nos dados da pesquisa “Raio X do Investidor Brasileiro”, que entrevistou 5.846 pessoas, de norte a sul do país (ANBIMA, 2024).

A pesquisa indica que as faixas de renda mais comuns para investidores de ações no país concentram-se entre R\$ 2.825,00 e R\$ 14.125,00 (ANBIMA, 2024). Dessa forma, foi definido um valor de aporte mensal de R\$ 300,00, que representa uma taxa de poupança conservadora (aproximadamente 10%) da faixa mais baixa deste grupo. O capital inicial, por sua vez, foi definido em R\$ 1.000,00, valor que representa um montante acessível para este mesmo perfil.

No início da simulação, o conjunto de dados de previsões (gerado pelo processo de retreino anual) é carregado. O motor de *backtesting* então executa um laço temporal que itera dia a dia sobre um índice de datas unificado, garantindo que todas as decisões sejam baseadas apenas em informações disponíveis até o momento da transação e eliminando, assim, o viés de *lookahead*.

Para a gestão de estado do portfólio, o sistema mantém dinamicamente duas estruturas centrais: uma que armazena o estado de cada posição (quantidade, preço médio, etc.), e uma variável, que representa a liquidez disponível, ajustada por aportes periódicos. Um ponto crucial para a imparcialidade da simulação é a aleatorização da ordem de avaliação dos ativos diariamente, mitigando o viés que poderia favorecer sistematicamente certos ativos na alocação de capital. Ao final de cada pregão, o sistema realiza a reconciliação do portfólio, calculando seu valor total de mercado para registrar a evolução patrimonial.

4.4.3 Validação das estratégias e análise dos resultados

Após a definição das estratégias e a criação do *trade system*, a etapa final da metodologia consistiu na validação empírica. Esta fase é crucial para assegurar a robustez dos resultados, mitigar riscos de sobreajuste e, fundamentalmente, avaliar o valor prático dos sinais gerados pelos modelos dos ativos em um ambiente que simula condições reais de mercado.

5 RESULTADOS

O treinamento inicial dos modelos foi conduzido utilizando um conjunto de dados históricos das empresas citadas acima que abrange o período de 1 de janeiro de 2010 a 31 de dezembro de 2019. Para a fase de testes e consolidação do processo de treinamento, foi utilizada a validação cruzada específica para séries temporais (*TimeSeriesSplit*), a fim de preservar a integridade temporal durante a busca pelos melhores parâmetros.

A validação seguiu um processo iterativo e multifásico, que partiu de um modelo de linha base para, subsequentemente, conduzir uma análise de técnicas de balanceamento. Foi realizada uma seleção recursiva de *features* (RFECV), guiada pela importância de cada variável, a fim de diminuir a dimensionalidade do modelo e buscar melhores métricas. Por fim, foi feita a otimização de hiperparâmetros de cada um dos modelos. Todo esse processo foi focado no recall da Classe 1 (compra).

A configuração final foi então encontrada por meio da própria validação cruzada para garantir uma estimativa de performance confiável. Este processo de validação e otimização culminou na seleção de um conjunto final e otimizado de variáveis preditoras e de parâmetros específicos para cada ativo. As métricas de performance consolidadas no primeiro ano subsequente ao treino dos modelos são apresentadas na Tabela 5.

Tabela 5 - Métricas do modelo treinado de cada ativo.

(continua)

Ativo	<i>Recall</i> (Classe 1) (%)	Acurácia Geral (%)	Precisão (Classe 1) (%)	F1-Score (Classe 1) (%)	ROC AUC
ITUB4	100,00	53,44	40,20	57,34	0,73
BBDC4	100,00	45,42	37,28	54,31	0,59
TIMS3	100,00	25,95	25,95	41,21	0,36
BBSE3	96,55	43,13	27,59	42,91	0,82
ITSA4	95,24	58,02	43,01	59,26	0,75
CMIG4	92,77	54,96	40,74	56,62	0,67
KLBN11	89,76	74,05	67,46	77,03	0,71
PETR4	82,35	88,55	87,50	84,35	0,9

Ativo	Recall (Classe 1) (%)	Acurácia Geral (%)	Precisão (Classe 1) (%)	F1-Score (Classe 1) (%)	ROC AUC
BBAS3	45,71	75,95	56,14	50,39	0,71
VALE3	42,96	53,82	60,40	50,21	0,57

Fonte: Autores (2025)

Concluída a fase de treinamento, a performance de cada estratégia foi validada empiricamente por meio de um *backtesting*. A simulação foi conduzida em um *trading system* customizado, abrangendo o período de 1 de janeiro de 2020 a 31 de dezembro de 2024. Conforme a política de aportes definida, cada estratégia simulada iniciou com um capital de R\$ 1.000,00 e foi alimentada com aportes mensais de R\$ 300,00. Os resultados comparativos, que demonstram a evolução patrimonial de cada abordagem, são exibidos na Tabela 6.

Tabela 6 - Resultados das estratégias ordenadas pelo retorno.

Estratégia	Retorno Total (%)	Retorno Anual (%)	Custo Final (R\$)	Saldo Final (R\$)	Caixa Final (R\$)
IA-FS-R2	57,90	11,58	18.527,44	29.254,31	177,35
IA-FS-R4	55,98	11,20	18.689,67	29.151,87	6,59
IA-L3-R4	46,79	9,36	18.395,39	26.958,17	335,88
IA-L3-R2	39,57	7,91	18.364,32	25.630,89	335,57
IA-L3-R3	38,34	7,67	18.375,45	25.420,98	327,85
IA-FS-R3	37,66	7,53	18.698,55	25.740,17	4,71
IA-L3-R1	36,15	7,23	18.387,26	25.034,99	319,27
IA-FS-R1	35,25	7,05	18.705,12	25.299,51	3,12

Fonte: Autores (2025)

A Tabela 6 resume a performance das 8 estratégias testadas, destacando a IA-FS-R2 (57,90%) e a IA-FS-R4 (55,98%) como as de melhor resultado, ambas superando os 50 pontos percentuais. Para dissecar os fatores que levaram a estes desempenhos, a análise a seguir foca em uma comparação direta entre estas estratégias de melhor resultado e os *benchmarks* do mercado financeiro. As tabelas a seguir detalham a distribuição final de ativos nas carteiras destas estratégias, permitindo a análise de suas composições.

Tabela 7 - Distribuição e desempenho da carteira da estratégia IA-FS-R2.

Ativo	Quantidade	Preço Médio (R\$)	Preço Final (R\$)	Retorno Final (%)	Alocação na Carteira (%)
PETR4	261	15,75	32,78	108,13	29,25
TIMS3	271	9,11	14,87	63,23	13,77
KLBN11	181	15,99	19,82	23,95	12,26
CMIG4	290	2,88	10,11	251,04	10,02
VALE3	47	54,59	53,04	-2,84	8,52
BBDC4	179	13,17	12,49	-5,16	7,64
BBSE3	67	18,14	29,84	64,50	6,83
ITSA4	223	5,51	8,85	60,62	6,75
ITUB4	42	18,21	28,76	57,94	4,13
BBAS3	10	8,64	24,03	178,12	0,82

Fonte: Autores (2025)

A Tabela 7 detalha a composição final da carteira da estratégia IA-FS-R2, o que permite analisar os fatores de seu desempenho. A primeira observação é que a estratégia foi capaz de alocar os 10 ativos perenes disponíveis.

Apesar da sua diversificação, a alocação de PETR4 foi significativamente maior que as demais, concentrando 29,25% da carteira apenas nesse ativo. Essa exposição maior à PETR4, que obteve um retorno individual de 108,13%, foi um dos motores para o resultado superior da estratégia.

A estratégia também foi capaz de alocar CMIG4, que, embora com uma participação menor (10,02%), teve a performance individual mais explosiva da carteira (251,04%). A diversificação no restante da carteira também ajudou a mitigar o impacto negativo da participação de VALE3 (-3,12%) e BBDC4 (-5,16%), que obtiveram uma alocação controlada de 8,52% e 7,64% respectivamente. Estes resultados negativos podem ser atribuídos a movimentos de mercado no período, como o ciclo de commodities para a VALE3 e problemas de gestão e inadimplência reportados para o BBDC4.

Tabela 8 - Distribuição e desempenho da carteira da estratégia IA-FS-R4.

Ativo	Quantidade	Preço Médio (R\$)	Preço Final (R\$)	Retorno Final (%)	Alocação na Carteira (%)
PETR4	250	16,25	32,78	101,72	28,11
TIMS3	258	9,10	14,87	63,41	13,16
KLBN11	170	15,85	19,82	25,05	11,56
CMIG4	290	2,95	10,11	242,71	10,06
BBSE3	88	17,85	29,84	67,17	9,01
VALE3	47	54,59	53,04	-2,84	8,55
BBDC4	196	12,92	12,49	-3,33	8,40
ITSA4	231	5,54	8,85	59,75	7,01
ITUB4	42	18,59	28,76	54,71	4,14

Fonte: Autores (2025)

A Tabela 8, por sua vez, detalha a composição final da carteira da estratégia IA-FS-R4. Primeiramente, pode-se observar que esta estratégia foi capaz de alocar capital em 9 dos 10 ativos disponíveis.

A estratégia IA-FS-R4 também apresentou sua maior concentração em PETR4, com 28,11% da carteira. Embora esta estratégia possua um limite de alocação de 25% por ativo, este valor final foi ultrapassado. Tal fato se explica pela mecânica do *backtesting*, o limite de 25% é verificado apenas no momento da compra. Após as últimas ordens de aquisição de PETR4 terem sido executadas, a subsequente oscilação de preço do ativo fez com que sua participação percentual no portfólio crescesse, sem que houvesse um período hábil para o rebalanceamento das posições antes do final do período de validação.

Além disso, o retorno individual de PETR4 na carteira foi de 101,72%. Este desempenho, inferior ao da mesma ação na carteira IA-FS-R2 (108,13%), é explicado por um preço médio de aquisição mais baixo (R\$ 16,25 contra R\$ 15,75).

A estratégia também foi capaz de alocar em CMIG4, com participação de 10,06% e a performance mais explosiva da carteira também (242,71%). A diversificação restante da carteira, assim como IA-FS-R2, foi capaz de mitigar os impactos dos retornos negativos de VALE3 (-2,84%) e BBDC4 (-3,33%).

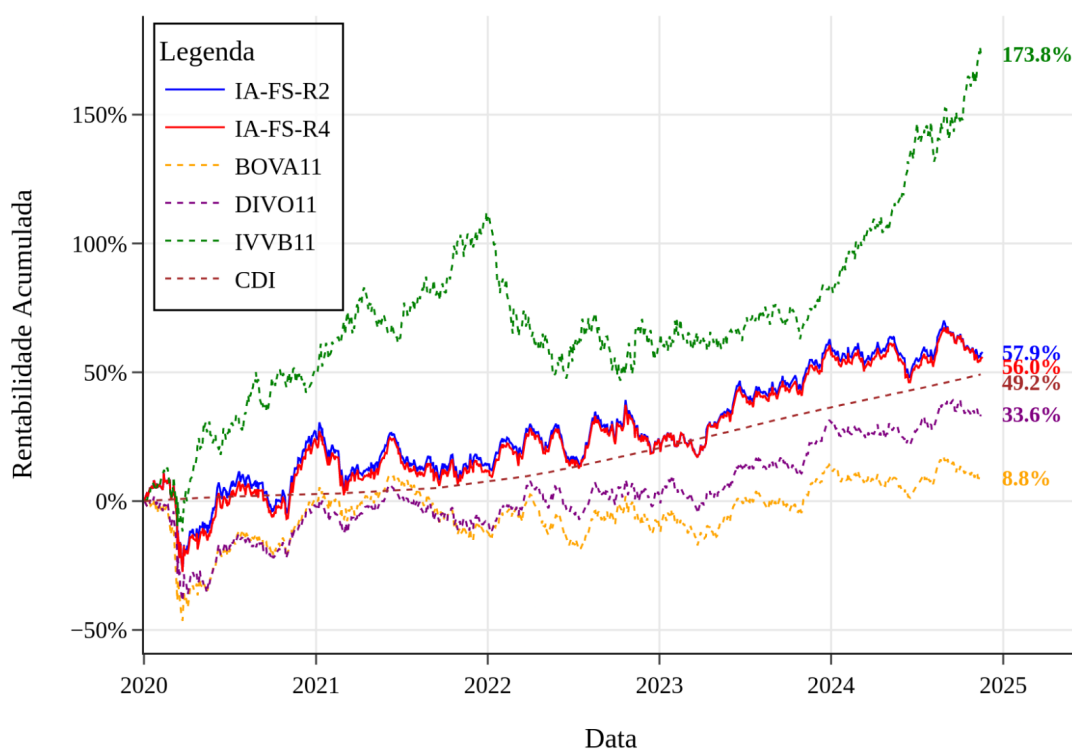
A análise da composição das duas estratégias de melhor performance (IA-FS-R2 e IA-FS-R4) ilustra os principais fatores que influenciaram seus respectivos retornos. Para fins

de completude e consulta, a distribuição final da carteira para as outras estratégias restantes encontra-se detalhada no Apêndice B.

Compreendida a dinâmica das carteiras ativas de melhor resultado, a análise final compara a performance das estratégias IA-FS-R2 e IA-FS-R4 com os principais *benchmarks* passivos do mercado. Para viabilizar a simulação e por questões de acessibilidade e liquidez, os índices de ações (IBOVESPA, IDIV e S&P 500) foram representados por seus respectivos ETFs (BOVA11, DIVO11 e IVVB11). O CDI foi incluído como o *benchmark* de renda fixa, com sua rentabilidade diária obtida diretamente do Banco Central.

Para uma comparação justa, foram feitas duas análises: (1) uma comparação da rentabilidade cumulativa entre as estratégias e os índices; e (2) uma simulação de investimentos utilizando a mesma política de aportes das estratégias (R\$ 1.000,00 iniciais e R\$ 300,00 mensais). O gráfico a seguir apresenta as rentabilidades acumuladas das estratégias ativas e os índices no período de validação.

Gráfico 1 - Rentabilidade cumulativa - Estratégias vs. *benchmarks* (2020 - 2024).



Fonte: Autores (2025)

O Gráfico 1 demonstra a trajetória de rentabilidade acumulada das estratégias e dos índices de mercado. Nele, observa-se o quanto o IVVB11 (ETF atrelado ao S&P 500) se

destaca em relação aos outros, acumulando o maior retorno no período (173,8%), beneficiando-se tanto da performance do mercado americano quanto da valorização do dólar no período de validação.

As duas estratégias ativas, IA-FS-R2 e IA-FS-R4, tiveram um desempenho muito similar ao longo de todo o período, com suas linhas quase sobrepostas. Ambas superaram com vantagem todos os *benchmarks* nacionais: o CDI, o BOVA11 (ETF atrelado ao IBOVESPA) e o DIVO11 (ETF atrelado ao IDIV).

Além da comparação da rentabilidade cumulativa pura, a segunda análise proposta, a simulação de aportes, apresenta uma perspectiva mais realista da performance das estratégias em relação aos *benchmarks*.

Tabela 9 - Resultados de simulação de aportes dos *benchmarks*.

Estratégia	Retorno Total (%)	Retorno Anual (%)	Custo Final (R\$)	Saldo Final (R\$)	Caixa Final (R\$)
IVVB11	72,79	14,56	18.491,70	31.951,68	208,30
DIVO11	32,93	6,59	18.608,67	24.736,80	91,33
CDI	30,38	6,08	18.700,00	24.381,91	0,00
BOVA11	15,05	3,01	18.685,76	21.498,71	14,24

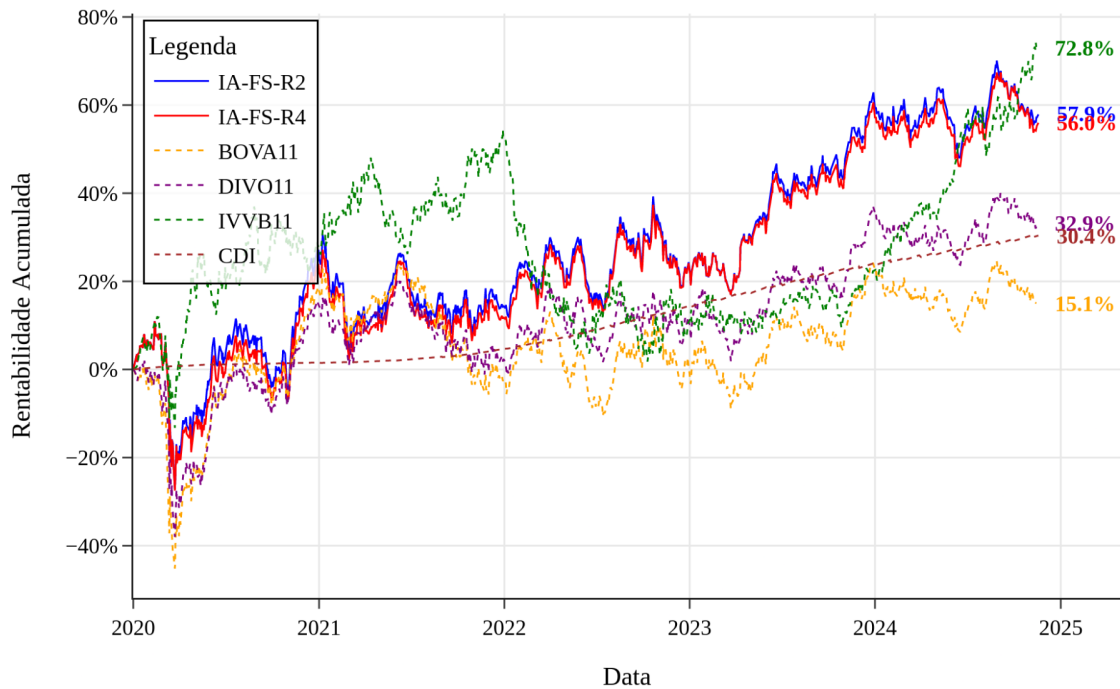
Fonte: Autores (2025)

A Tabela 9 demonstra a performance dos *benchmarks* passivos sob a mesma política de aportes das estratégias guiadas pelo modelo de RF. O IVVB11 (ETF atrelado ao S&P 500) se destacou com um retorno total de 72,79% (14,56% anual).

Ao comparar esses resultados com os da Tabela 6, fica claro o desempenho das estratégias ativas IA-FS-R2 (com 57,90% de retorno total) e IA-FS-R4 (com 55,98% de retorno total). Tais estratégias tiveram um resultado levemente inferior à do IVVB11, mas superaram com ampla margem o DIVO11 (32,93%), o CDI (30,38%) e o BOVA11 (15,05%).

Isso indica que a seleção ativa de ativos perenes, guiada pelos modelos de RF, foi capaz de gerar um alfa significativo (retorno acima do *benchmark*) em relação ao principal índice do mercado brasileiro (BOVA11) e aos ativos de dividendos (DIVO11).

Gráfico 2 - Rentabilidade cumulativa - Estratégias vs. *benchmarks* (Simulação de aportes).



Fonte: Autores (2025)

O Gráfico 2 ilustra visualmente a evolução da rentabilidade diária desta simulação de aportes nos *benchmarks* passivos em relação às estratégias ativas (IA-FS-R2 e IA-FS-R4). A trajetória do IVVB11 demonstra sua superioridade em relação ao retorno, porém, as estratégias IA-FS-R2 e IA-FS-R4 navegam muito mais próximas dele e estabelecem uma clara vantagem sobre os *benchmarks* nacionais (BOVA11, DIVO11 e CDI), terminando a simulação com um patrimônio líquido significativamente maior.

Completadas as duas análises de retorno (cumulativa e por simulação de aportes), é fundamental verificar se o retorno das estratégias ativas em relação aos *benchmarks* de mercado foi alcançado com um nível de risco aceitável. A Tabela 10 compara as volatilidade e as quedas máximas de cada uma delas.

Tabela 10 - Métricas de risco consolidadas.

Estratégia/Benchmark	Volatilidade Anualizada (%)	Drawdown Máximo (%)
CDI	0,26	0,00
IA-FS-R2	22,66	-20,65
IVVB11	22,34	-21,49
IA-FS-R4	22,23	-22,25
DIVO11	22,09	-36,45
BOVA11	26,13	-40,70

Fonte: Autores (2025)

As métricas de risco revelam que o *drawdown* máximo das estratégias ativas (IA-FS-R2 com -20,65% e IA-FS-R4 com -22,25%) foi significativamente inferior ao dos *benchmarks* do mercado brasileiro, como o BOVA11 (-40,70%) e o DIVO11 (-36,45%).

Em termos de volatilidade, o BOVA11 (26,13%) foi o que apresentou maior oscilação. A estratégia IA-FS-R2 (22,66%) apresentou uma volatilidade mais controlada, ficando alinhada com a IA-FS-R4 (22,23%) e os *benchmarks* DIVO11 (22,09%) e IVVB11 (22,34%).

A estratégia IA-FS-R2 apresentou o perfil de risco mais conservador entre as estratégias de renda variável, com o menor *drawdown* (-20,65%). O IVVB11 apresentou o perfil mais próximo, com a menor volatilidade (22,34%) e um *drawdown* ligeiramente superior (-21,49%).

Ressalta-se que o CDI não é incluído diretamente nesta comparação de risco por ser um ativo de classe distinta. Seus valores de volatilidade (0,26%) e *drawdown* (0,00%) são inerentes à renda fixa e servem, no contexto deste trabalho, primariamente como um *benchmark* de custo de oportunidade de capital, e não como um parâmetro de risco comparável aos portfólios de ações.

Para uma análise visual da profundidade, frequência e duração dessas quedas, os gráficos com o histórico de *drawdown* diário das estratégias e dos *benchmarks* encontram-se no Apêndice C.

6 DISCUSSÕES

A análise das métricas finais dos modelos permite extrair conclusões importantes sobre o processo de modelagem. A principal delas é a estratégia de generalizar a otimização de *features* e hiperparâmetros para o processo de retreino anual dos modelos dos ativos. Isso evidencia que, mesmo em uma carteira de ativos perenes, cada empresa possui dinâmicas de preço e comportamentos distintos ao longo do tempo, fazendo com que uma configuração otimizada para um período do tempo não seja transferível para outros.

Observam-se dois padrões de falha contrastantes. Por um lado, modelos como os de ITUB4, BBDC4 e TIMS3 apresentaram um *recall* de 100%, combinado a uma precisão extremamente baixa (40,20%, 37,28% e 25,95%, respectivamente). Tal comportamento indica que a arquitetura desenvolveu um viés de previsão extremo para esses ativos, classificando quase todas as instâncias como "comprar" e gerando uma quantidade massiva de falsos positivos.

Em contrapartida, para VALE3, o modelo tornou-se mais conservador, prevendo a classe "comprar" com menos frequência e resultando em um *recall* de 42,96%, embora com uma precisão maior de 60,40%. O modelo para PETR4 (*recall* 82,35%, precisão 87,50%) manteve-se equilibrado, validando o sucesso da utilização de indicadores macroeconômicos.

As diferenças no desempenho dos modelos individuais, discutidas anteriormente, impactam diretamente a performance das estratégias de investimento. A análise dos resultados do *backtesting* permite quantificar o quanto os comportamentos dos modelos se traduziram em retorno financeiro. Dessa forma, observa-se que as estratégias IA-FS-R2 e IA-FS-R4 emergiram como as de melhor desempenho, com um retorno total de 57,90% (11,58% anual) e 55,98% (11,20% anual) respectivamente, conforme Tabela 6. Ambas as configurações combinam a execução de compra no primeiro sinal do modelo, diferindo primariamente nas regras de alocação máxima por ativo e lote mínimo. Este resultado sugere que a agilidade na alocação de capital, característica central deste grupo IA-FS, parece ter sido fundamental para capturar o início dos movimentos de alta.

Para contextualizar a magnitude deste resultado, é fundamental compará-lo com os *benchmarks* de mercado, apresentados na Tabela 9. As estratégias de melhor performance, IA-FS-R2 (57,90%) e IA-FS-R4 (55,98%), geraram um retorno significativo sobre os principais índices nacionais, como o BOVA11 (15,05%) e o DIVO11 (32,93%). Isso valida que a seleção de ativos perenes combinada ao timing de compra guiado por ML foi capaz de agregar valor substancial sobre uma estratégia puramente passiva no mercado brasileiro.

Contudo, ambas as estratégias foram superadas pelo *benchmark* internacional IVVB11 (72,79%), um ETF que replica o maior índice do mercado americano e que se beneficiou tanto da performance dele quanto da forte valorização do dólar no período de validação.

Contudo, a principal contribuição do *framework* não está apenas no retorno, mas na gestão de risco. A análise do *drawdown* máximo (Tabela 10) revela que a IA-FS-R2 (-20,65%) e a IA-FS-R4 (-22,25%) ofereceram uma proteção de capital muito superior à de um investimento passivo no BOVA11 (-40,70%) ou DIVO11 (-36,45%) e resultados próximos ao do IVVB11 (-21,49%). Isso sugere que o modelo foi eficaz em evitar compras em picos de euforia ou em momentos de risco elevado, mitigando as piores quedas do mercado e protegendo o capital do investidor em períodos de tensão, o que justifica a abordagem de otimização de *timing* no *buy and hold*.

Em contraste com o sucesso das outras estratégias, o grupo de estratégias IA-L3 apresentou um desempenho notadamente inferior, com retornos totais entre 36,15% e 46,79%. Embora utilizassem os sinais gerados pelos modelos, a regra de aguardar até o final do mês para alocar o capital mostrou-se contraproducente. Este fraco desempenho reforça uma hipótese central, já sugerida pelo sucesso da melhor estratégia: a latência na execução é prejudicial. A espera fez com que o sistema perdesse o *timing* ideal dos movimentos de alta, que aparentemente ocorrem bem antes do fechamento do período mensal.

7 CONCLUSÕES

Este trabalho propôs e avaliou um *framework* para aprimorar a estratégia de *buy and hold* por meio de um sistema de negociação algorítmica, que utiliza RF como modelo preditivo para otimizar os momentos de compra em um portfólio de ativos perenes do mercado brasileiro. Os estudos conduzidos através do processo de *backtesting*, geraram conclusões fundamentais sobre a modelagem preditiva e a implementação das estratégias de investimento propostas.

Embora o escopo deste trabalho não tenha incluído um estudo comparativo direto com outras arquiteturas, como modelos de *deep learning*, a escolha pelo RF foi deliberadamente fundamentada em evidências da literatura. Trabalhos como o de Ma et al. (2021) já demonstraram a sua superioridade em tarefas de previsão de retornos em comparação com modelos mais complexos, validando-o como um ponto de partida robusto e eficaz para a otimização de estratégias financeiras.

Os resultados demonstraram, primeiramente, que a generalização de variáveis de entrada e hiperparâmetros de um modelo, especialmente durante o processo de retreino anual, é uma abordagem ineficaz. As dinâmicas únicas de cada ativo exigem uma otimização individualizada para cada período de tempo, a fim de ter modelos preditivos confiáveis.

Em segundo lugar, a superioridade das estratégias IA-FS sobre as IA-L3 confirmou que o *timing* na execução dos aportes é um fator crítico para maiores retornos. Por fim, a análise do *backtesting* revelou que a maior contribuição do *framework* foi a gestão de risco, as estratégias ativas protegeram o capital com um *drawdown* máximo substancialmente inferior ao dos índices brasileiros.

É fundamental, contudo, reconhecer as limitações do presente estudo. A principal delas foi a generalização das características do modelo no processo de retreino, que se mostrou uma simplificação metodológica. Além disso, a simulação não considerou alguns elementos do mercado real, como custos transacionais, *spreads* de compra, dividendos e bonificações.

Tais limitações, no entanto, abrem caminhos claros para evolução e seguimento deste trabalho. As perspectivas de continuidade incluem, como prioridade, a otimização individualizada do modelo preditivo de cada ativo durante os retreinos, exploração de arquiteturas de modelos mais complexos, entre outras.

Em suma, o *framework* modular e escalável desenvolvido serve como base para futuras pesquisas quantitativas, permitindo o teste e a implementação de novas ideias no dinâmico cenário do mercado financeiro brasileiro.

8 SUGESTÕES PARA TRABALHOS FUTUROS

O *framework* metodológico estabelecido neste estudo oferece um fundamento robusto para a avaliação de estratégias, mas também revela diversas oportunidades para pesquisas futuras. A primeira e mais fundamental linha de aprimoramento seria a otimização individualizada do modelo preditivo de cada ativo durante os retreinos. Embora a abordagem de generalizar um template único garanta consistência, reconhece-se que a realização de um processo de otimização específico para cada retreino resultaria em modelos mais calibrados e, potencialmente, em uma performance preditiva superior para o portfólio como um todo.

Uma segunda vertente de pesquisa consiste na exploração de arquiteturas de modelos alternativos. Propõe-se a implementação de modelos estatísticos clássicos, como o *Auto-Regressive Integrated Moving Average* (ARIMA), para servir como um rigoroso *benchmark*, quantificando o valor agregado dos modelos de aprendizado de máquina. Adicionalmente, pretende-se investigar Redes Neurais Recorrentes (RNN), especificamente as arquiteturas *Long Short-Term Memory* (LSTM), que são projetadas para capturar dependências temporais de longo prazo e padrões não-lineares complexos, frequentemente presentes em dados financeiros.

Por fim, o desdobramento mais avançado desta pesquisa seria a criação de um sistema híbrido e especialista. Após uma análise de performance entre modelos para comparar os resultados do *random forest* com os modelos ARIMA e LSTM, esta abordagem permitiria alocar o modelo de melhor performance para cada ativo específico. Com isso, a lógica de alocação de capital evoluiria para um sistema de pontuação ponderada, onde um sinal de compra gerado por um modelo de alta acurácia histórica para um ativo receberia um maior score de confiança, resultando em um mecanismo de alocação de risco mais robusto e consistente.

REFERÊNCIAS

- ACHELIS, S. B.; ACHELIS, S. **Technical Analysis from A to Z**. New York: McGraw Hill, 2001.
- ASSAF NETO, Alexandre. **Mercado Financeiro**. 11. ed. São Paulo: Atlas, 2012.
- ASSOCIAÇÃO BRASILEIRA DAS ENTIDADES DOS MERCADOS FINANCEIRO E DE CAPITAIS (ANBIMA). **Raio X do Investidor Brasileiro**. 2024. Disponível em: https://www.anbima.com.br/pt_br/especial/raio-x-do-investidor-brasileiro.htm. Acesso em: 4 nov. 2025.
- AZIZ, S.; DOWLING, M.; GHANDOUR, A.; TRIGONOMETRICIO, P. Machine learning in finance: A topic modeling approach. **European Financial Management**, v. 28, n. 3, p. 744–770, 2022.
- B3. Pessoas Físicas na B3. 2024. Disponível em: https://www.b3.com.br/pt_br/noticias/pessoas-fisicas-na-b3.htm. Acesso em: 13 ago. 2024.
- B3. B3 - Brasil, Bolsa, Balcão. 2025. Disponível em: <https://www.b3.com.br/>. Acesso em: 13 ago. 2024.
- BALLINGS, M.; VAN DEN POEL, D.; HESPEELS, N.; GRYP, R. Evaluating multiple classifiers for stock price direction prediction. **Expert systems with Applications**, v. 42, n. 20, p. 7046–7056, 2015.
- BAMPOUNTIS, S.; NIFAKOS, S.; GOGAS, P.; PAPADIMITRIOU, T. Using Machine Learning on Macroeconomic, Technical, and Sentiment Indicators for Stock Market Forecasting. **Algorithms**, v. 17, n. 7, p. 584, 2024.
- BOTACHE, D.; PÉREZ, D.; ARBELÁEZ, P. Unraveling the complexity of splitting sequential data: tackling challenges in video and time series analysis. **arXiv preprint arXiv:2307.14294**, 2023.
- BTG PACTUAL. **O que são ações**. [s.l.]: BTG Pactual Blog, [s.d.]. Disponível em: <https://content.btgpactual.com/blog/acoes/o-que-sao-acoes>. Acesso em: 11 nov. 2025.
- CAMPBELL, S. D. A review of backtesting and backtesting procedures. **Finance and Economics Discussion Series**, 2005-21, Board of Governors of the Federal Reserve System (U.S.), 2005.
- DAMODARAN, A. **Avaliação de investimentos**: ferramentas e técnicas para determinação do valor. Rio de Janeiro: Qualitymark Editora Ltda., 1997.
- DARAPANENI, Narayana et al. Stock price prediction using sentiment analysis and deep learning for Indian markets. **arXiv preprint arXiv:2204.05783**, 2022.
- DEBOECK, G. J. **Trading on the edge**: neural, genetic, and fuzzy systems for chaotic financial markets. New York: John Wiley & Sons, 1994.

- EL NAQA, I.; MURPHY, M. J. What is machine learning?. *In: Machine learning in radiation oncology: theory and applications*. Cham: Springer International Publishing, 2015. p. 3–11.
- GOLDBERG, Lisa R.; MAHMOUD, Ola. Drawdown: from practice to theory and back again. **Mathematics and Financial Economics**, v. 11, n. 3, p. 275-297, 2017.
- HARYANTO, J. T. Short-Term Versus Long-Term Portfolio Management Strategies and the Selection of Securities. **Advances in Management & Financial Reporting**, v. 2, n. 1, p. 11–23, 2024.
- INVESTIDOR10. Investidor10 - Ajudando você a investir melhor. 2025. Disponível em: <https://investidor10.com.br/>. Acesso em: 13 ago. 2024.
- KARA, Yakup; BOYACIOGLU, Melek Acar; BAYKAN, Ömer Kaan. Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. **Expert systems with Applications**, v. 38, n. 5, p. 5311-5319, 2011.
- MA, Y.; HAN, R.; WANG, W. Portfolio optimization with return prediction using deep learning and machine learning. **Expert Systems with Applications**, v. 165, p. 113973, 2021.
- NAVLANI, A.; FANDANGO, A.; IDRIS, I. **Python Data Analysis**: Perform data collection, data processing, wrangling, visualization, and model building using Python. Birmingham: Packt Publishing Ltd., 2021.
- OĞUZ, R. F.; EYERCIOĞLU, Ö. K.; AVCI, D. On the use of technical analysis indicators for stock market price movement direction prediction. *In: 2019 27th SIGNAL PROCESSING AND COMMUNICATIONS APPLICATIONS CONFERENCE (SIU)*, 2019. **Anais... IEEE**, 2019. p. 1–4.
- PATEL, Jigar et al. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. **Expert systems with applications**, v. 42, n. 1, p. 259-268, 2015.
- PIXLEY, J. **Emotions in finance**: Distrust and uncertainty in global markets. Cambridge: Cambridge University Press, 2004.
- SCIKIT-LEARN DEVELOPERS. `sklearn.pipeline.Pipeline`. 2024. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.pipeline.Pipeline.html>. Acesso em: 29 jul. 2025.
- SHAIK, A. B.; SRINIVASAN, S. A brief survey on random forest ensembles in classification model. *In: INTERNATIONAL CONFERENCE ON INNOVATIVE COMPUTING AND COMMUNICATIONS: PROCEEDINGS OF ICICC 2018, VOLUME 2*, 2018. **Anais...** Singapore: Springer, 2018. p. 253–260.
- SILVA, N. F.; de ANDRADE, L. P.; da SILVA, W. S.; de MELO, M. K.; TONELLI, A. O. Portfolio optimization based on the pre-selection of stocks by the Support Vector Machine model. **Finance Research Letters**, v. 61, p. 105014, 2024.
- SILVA SANTOS, G. F.; da SILVA SANTOS, C. A.; MARQUES, N. L. A estratégia de buy and hold na composição da carteira de jovens investidores da baixada fluminense. **Caderno de Administração**, v. 29, n. 2, p. 6–31, 2021.

STATUS INVEST. Status Invest - Análise de Ações, Fundos Imobiliários e muito mais. 2025. Disponível em: <https://statusinvest.com.br/>. Acesso em: 13 ago. 2024.

ZHENG, J.; XIN, D.; CHENG, Q.; TIAN, M.; YANG, L. The random forest model for analyzing and forecasting the us stock market in the context of smart finance. **arXiv preprint arXiv:2402.17194**, 2024.

APÊNDICE A - Estatísticas descritivas das variáveis dos ativos

Tabela A.1 - Estatísticas descritivas das variáveis (BBDC4).

Variável	Média	Desvio Padrão	Mínimo	Mediana	Máximo
SMA_10	9,02	3,99	4,76	7,49	19,44
WMA_10	9,03	4,00	4,73	7,49	19,54
MACD	0,037	0,193	-0,720	0,015	0,651
A/D (em Milhões)	1.223,19	784,89	-86,18	1.226,68	2.864,11
CCI	-3.427,93	1.751,91	-32.903,50	-3.064,08	-584,06
RSI	52,15	12,24	16,61	51,58	90,75
MOM	0,075	0,637	-2,430	0,060	2,920
STOCHk	53,59	28,85	0,32	55,03	99,91
STOCHd	53,59	28,01	2,87	55,47	98,36
WILL	-46,41	30,94	-100,00	-44,80	0,00

Fonte: Autores (2025)

Tabela A.2 - Estatísticas descritivas das variáveis (VALE3).

Variável	Média	Desvio Padrão	Mínimo	Mediana	Máximo
SMA_10	17,34	6,88	4,76	16,27	34,33
WMA_10	17,34	6,89	4,73	16,29	34,36
MACD	0,027	0,397	-1,516	-0,011	1,392
A/D (em Milhões)	-207,10	258,09	-571,69	-275,60	163,06
CCI	-2.636,61	1.372,59	-13.233,96	-2.450,43	-463,17
RSI	50,40	11,84	12,39	49,96	87,21
MOM	0,059	1,460	-7,490	0,020	5,720
STOCHk	49,47	29,02	0,54	49,56	100,00
STOCHd	49,46	28,14	1,03	49,43	99,18
WILL	-50,54	31,24	-100,00	-50,58	0,00

Fonte: Autores (2025)

Tabela A.3 - Estatísticas descritivas das variáveis (BBAS3).

Variável	Média	Desvio Padrão	Mínimo	Mediana	Máximo
SMA_10	7,44	3,50	3,49	5,82	17,14
WMA_10	7,44	3,51	3,48	5,83	17,21
MACD	0,031	0,209	-0,687	0,012	0,796
A/D (em Milhões)	454,92	236,84	-127,66	419,09	990,25
CCI	-2.649,01	1.387,65	-9.487,26	-2.358,60	-338,53
RSI	51,72	13,25	15,52	51,29	90,22
MOM	0,064	0,678	-2,550	0,050	3,420
STOCHk	51,84	29,82	0,00	52,27	100,00
STOCHd	51,85	28,99	0,00	52,27	100,00
WILL	-48,16	31,79	-100,00	-47,87	0,00

Fonte: Autores (2025)

Tabela A.4 - Estatísticas descritivas das variáveis (CMIG4).

Variável	Média	Desvio Padrão	Mínimo	Mediana	Máximo
SMA_10	2,28	0,66	0,90	2,10	3,85
WMA_10	2,28	0,67	0,89	2,11	3,86
MACD	0,005	0,066	-0,278	0,003	0,224
A/D (em Milhões)	-1.270,88	883,41	-2.322,29	-1.570,01	394,80
CCI	-2.909,59	1.536,03	-9.737,19	-2.610,93	-412,48
RSI	51,59	12,82	14,15	51,26	87,99
MOM	0,011	0,200	-1,110	0,020	0,870
STOCHk	50,71	28,48	-13,69	52,02	100,00
STOCHd	50,70	27,64	-6,94	52,02	100,00
WILL	-49,29	30,65	-132,18	-48,26	0,00

Fonte: Autores (2025)

Tabela A.5 - Estatísticas descritivas das variáveis (ITUB4).

Variável	Média	Desvio Padrão	Mínimo	Mediana	Máximo
SMA_10	11,89	5,59	5,58	9,60	24,76
WMA_10	11,90	5,60	5,53	9,61	24,85
MACD	0,048	0,228	-0,933	0,032	0,907
A/D (em Milhões)	1.707,60	601,19	471,70	1.867,23	2.650,48
CCI	-3.552,43	1.752,59	-11.832,08	-3.174,41	-592,04
RSI	52,26	11,96	16,66	52,22	89,08
MOM	0,097	0,759	-2,770	0,100	3,090
STOCHk	53,02	28,37	0,00	54,99	100,00
STOCHd	53,03	27,50	0,46	54,97	97,54
WILL	-46,99	30,58	-100,00	-45,04	0,00

Fonte: Autores (2025)

Tabela A.6 - Estatísticas descritivas das variáveis (BBSE3).

Variável	Média	Desvio Padrão	Mínimo	Mediana	Máximo
SMA_10	14,15	2,33	9,17	13,88	22,32
WMA_10	14,16	2,34	9,08	13,89	22,31
MACD	0,054	0,237	-0,546	0,053	0,959
A/D (em Milhões)	631,45	71,62	395,89	660,97	709,67
CCI	-3.632,98	1.646,54	-10.029,45	-3.273,21	-689,28
RSI	52,60	10,99	10,99	52,17	89,72
MOM	0,116	0,822	-2,730	0,090	3,910
STOCHk	52,34	27,32	2,23	53,01	99,67
STOCHd	52,31	26,41	3,78	52,55	98,84
WILL	-47,62	29,59	-100,00	-46,31	0,00

Fonte: Autores (2025)

Tabela A.7 - Estatísticas descritivas das variáveis (ITSA4).

Variável	Média	Desvio Padrão	Mínimo	Mediana	Máximo
SMA_10	3,31	1,54	1,66	2,66	7,50
WMA_10	3,32	1,54	1,64	2,67	7,51
MACD	0,015	0,066	-0,295	0,008	0,253
A/D (em Milhões)	2.603,57	1.027,99	2.995,13	3.055,89	4.029,68
CCI	-3.630,97	1.781,10	-12.414,75	-3.297,61	-700,85
RSI	52,47	12,15	13,03	52,55	88,51
MOM	0,030	0,219	-0,900	0,020	0,940
STOCHk	53,65	28,18	0,00	54,41	100,00
STOCHd	53,66	27,31	0,78	55,06	98,23
WILL	-46,35	30,33	-100,00	-45,08	0,00

Fonte: Autores (2025)

Tabela A.8 - Estatísticas descritivas das variáveis (TIMS3).

Variável	Média	Desvio Padrão	Mínimo	Mediana	Máximo
SMA_10	6,13	1,52	3,67	6,06	10,29
WMA_10	6,14	1,52	3,66	6,07	10,39
MACD	0,016	0,122	-0,298	0,022	0,406
A/D (em Milhões)	220,25	135,48	-22,28	268,76	563,04
CCI	-3.291,53	1.673,73	-11.928,39	-2.910,71	-721,87
RSI	51,69	11,32	18,15	51,63	80,25
MOM	0,036	0,430	-1,830	0,050	1,430
STOCHk	53,12	27,25	0,27	54,26	100,00
STOCHd	53,09	26,42	2,59	54,42	99,72
WILL	-46,85	29,68	-100,00	-45,46	0,00

Fonte: Autores (2025)

Tabela A.9 - Estatísticas descritivas das variáveis (KLBN11).

Variável	Média	Desvio Padrão	Mínimo	Mediana	Máximo
SMA_10	11,31	1,51	6,71	11,23	14,68
WMA_10	11,31	1,51	6,71	11,22	14,72
MACD	0,031	0,208	-0,784	0,033	0,541
A/D (em Milhões)	275,59	48,48	111,04	287,33	338,89
CCI	-3.692,68	1.870,73	-19.052,27	-3.311,06	-614,79
RSI	52,15	12,24	15,52	51,74	89,70
MOM	0,067	0,698	-3,640	0,095	1,940
STOCHk	52,10	27,81	0,00	53,81	98,88
STOCHd	52,10	27,03	0,00	54,49	98,53
WILL	-47,91	29,83	-100,00	-46,67	0,00

Fonte: Autores (2025)

Tabela A.10 - Estatísticas descritivas das variáveis (Índice BOVESPA).

Variável	Média	Desvio Padrão	Mínimo	Mediana	Máximo
SMA_10	64.571,29	16.269,97	38.049,00	59.127,30	115.782,10
WMA_10	64.599,76	16.325,76	37.946,73	59.185,47	116.070,80
MACD	127,01	911,18	-3.110,45	53,40	2.353,85
A/D (em Milhões)	126,33	91,38	4,01	93,91	424,52
CCI	60.118,65	16.058,05	36.101,54	55.048,01	113.320,04
RSI	51,70	12,33	13,45	50,88	84,44
MOM	261,689	2,998,832	-10.431,000	158,000	9.321,000
STOCHk	53,34	29,89	1,50	55,56	100,00
STOCHd	53,35	29,04	1,74	54,40	99,89
WILL	-46,66	31,97	-100,00	-44,23	0,00

Fonte: Autores (2025)

Tabela A.11 - Estatísticas descritivas das variáveis (Taxa de câmbio - USD/BRL).

Variável	Média	Desvio Padrão	Mínimo	Mediana	Máximo
SMA_10	2,83	0,82	1,55	3,10	4,22
WMA_10	2,83	0,82	1,55	3,10	4,22
MACD	0,01	0,03	-0,09	0,01	0,14
A/D (em Milhões)	0,00	0,00	0,00	0,00	0,00
CCI	-7.890,69	4.738,55	-37.219,65	-6.836,45	-1.445,06
RSI	52,35	11,21	23,23	51,69	87,78
MOM	0,014	0,101	-0,416	0,009	0,454
STOCHk	53,65	24,58	-51,57	55,87	99,80
STOCHd	53,65	23,71	-40,19	56,16	97,50
WILL	-46,37	27,54	-225,34	-44,50	22,51

Fonte: Autores (2025)

Tabela A.12 - Estatísticas descritivas das variáveis (Barril de petróleo).

Variável	Média	Desvio Padrão	Mínimo	Mediana	Máximo
SMA_10	80,22	26,69	30,28	73,74	125,13
WMA_10	80,21	26,70	29,86	73,70	125,22
MACD	-0,04	1,54	-5,61	0,15	4,46
A/D (em Milhões)	0,22	0,43	-0,64	0,07	1,37
CCI	-3.921,40	2.466,33	-23.546,38	-3.302,33	-715,58
RSI	50,17	13,26	11,44	51,45	81,96
MOM	-0,084	4,851	-17,950	0,170	15,480
STOCHk	52,81	29,60	-0,49	54,98	99,82
STOCHd	52,82	28,85	0,66	54,72	99,50
WILL	-47,20	31,62	-104,81	-45,37	2,15

Fonte: Autores (2025)

Tabela A.13 - Estatísticas descritivas das variáveis (Minério de ferro).

Variável	Média	Desvio Padrão	Mínimo	Mediana	Máximo
SMA_10	92,19	34,51	38,79	82,16	178,90
WMA_10	92,14	34,46	38,76	82,24	178,84
MACD	-0,25	2,31	-10,59	-0,08	7,86
A/D (em Milhões)	0,00	0,00	0,00	0,00	0,00
CCI	-5.183,28	5.161,00	-42.719,94	-3.663,15	-555,08
RSI	48,70	23,00	2,74	49,45	95,11
MOM	-0,493	6,852	-37,560	0,220	25,310
STOCHk	51,24	38,88	-42,47	57,19	104,24
STOCHd	51,28	38,42	-38,28	57,19	104,24
WILL	-48,80	40,02	-152,98	-42,89	12,71

Fonte: Autores (2025)

APÊNDICE B - Distribuições de carteira das estratégias

Tabela B.1 - Distribuição e desempenho da carteira da estratégia IA-L3-R4.

Ativo	Quantidade	Preço Médio (R\$)	Preço Final (R\$)	Retorno Final (%)	Alocação na Carteira (%)
PETR4	173	22,78	32,78	43,90	21,04
KLBN11	259	15,55	19,82	27,46	19,04
TIMS3	277	9,00	14,87	65,22	15,28
BBSE3	93	18,39	29,84	62,26	10,29
ITSA4	263	5,42	8,85	63,28	8,63
VALE3	42	45,15	53,04	17,50	8,26
CMIG4	211	2,91	10,11	247,42	7,91
BBDC4	146	12,40	12,49	0,73	6,76
ITUB4	26	17,23	28,76	66,92	2,77

Fonte: Autores (2025)

Tabela B.2 - Distribuição e desempenho da carteira da estratégia IA-L3-R2.

Ativo	Quantidade	Preço Médio (R\$)	Preço Final (R\$)	Retorno Final (%)	Alocação na Carteira (%)
TIMS3	349	9,22	14,87	61,28	20,25
KLBN11	254	15,38	19,82	28,87	19,64
PETR4	150	25,33	32,78	29,41	19,18
BBSE3	97	18,65	29,84	60,00	11,29
BBDC4	188	12,73	12,49	-1,89	9,16
CMIG4	180	3,10	10,11	226,13	7,10
VALE3	29	52,25	53,04	1,51	6,00
ITSA4	142	5,53	8,85	60,04	4,90
ITUB4	22	17,26	28,76	66,63	2,47

Fonte: Autores (2025)

Tabela B.3 - Distribuição e desempenho da carteira da estratégia IA-L3-R3.

Ativo	Quantidade	Preço Médio (R\$)	Preço Final (R\$)	Retorno Final (%)	Alocação na Carteira (%)
VALE3	85	49,27	53,04	7,65	17,73
PETR4	135	23,96	32,78	36,81	17,41
KLBN11	207	15,71	19,82	26,16	16,14
TIMS3	222	9,20	14,87	61,63	12,99
BBSE3	94	18,53	29,84	61,04	11,03
CMIG4	202	3,02	10,11	231,48	8,03
BBDC4	135	13,28	12,49	-5,95	6,63
ITSA4	157	5,37	8,85	64,80	5,47
ITUB4	32	17,61	28,76	63,32	3,62
BBAS3	10	10,12	24,03	137,45	0,95

Fonte: Autores (2025)

Tabela B.4 - Distribuição e desempenho da carteira da estratégia IA-FS-R3.

Ativo	Quantidade	Preço Médio (R\$)	Preço Final (R\$)	Retorno Final (%)	Alocação na Carteira (%)
PETR4	171	21,61	32,78	51,69	21,78
VALE3	84	50,12	53,04	5,83	17,31
TIMS3	225	9,27	14,87	60,41	13,00
KLBN11	168	16,29	19,82	21,67	12,94
BBSE3	105	19,64	29,84	51,93	12,17
ITUB4	56	18,34	28,76	56,82	6,26
BBDC4	124	13,65	12,49	-8,50	6,02
CMIG4	139	2,98	10,11	239,26	5,46
ITSA4	115	5,61	8,85	57,75	3,95
BBAS3	12	10,79	24,03	122,71	1,12

Fonte: Autores (2025)

Tabela B.5 - Distribuição e desempenho da carteira da estratégia IA-L3-R1.

Ativo	Quantidade	Preço Médio (R\$)	Preço Final (R\$)	Retorno Final (%)	Alocação na Carteira (%)
PETR4	132	24,50	32,78	33,80	17,28
KLBN11	214	15,75	19,82	25,84	16,94
VALE3	72	52,85	53,04	0,36	15,25
TIMS3	231	9,27	14,87	60,41	13,72
BBSE3	97	17,80	29,84	67,64	11,56
ITUB4	63	17,44	28,76	64,91	7,24
BBDC4	127	12,88	12,49	-3,03	6,34
ITSA4	156	5,58	8,85	58,60	5,51
CMIG4	119	3,05	10,11	231,48	4,81
BBAS3	14	10,12	24,03	137,45	1,34

Fonte: Autores (2025)

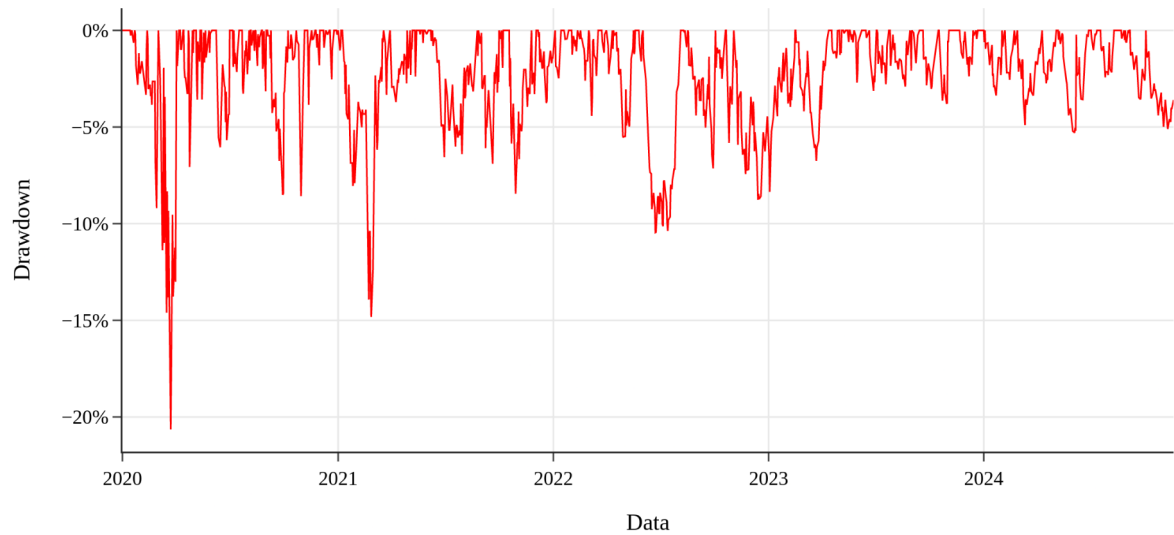
Tabela B.6 - Distribuição e desempenho da carteira da estratégia IA-FS-R1.

Ativo	Quantidade	Preço Médio (R\$)	Preço Final (R\$)	Retorno Final (%)	Alocação na Carteira (%)
PETR4	177	21,75	32,78	50,71	22,93
VALE3	90	51,70	53,04	2,59	18,87
KLBN11	191	16	19,82	23,87	14,96
TIMS3	186	9,35	14,87	59,04	10,93
BBSE3	88	19,05	29,84	56,64	10,38
BBDC4	120	13,79	12,49	-9,43	5,92
ITUB4	48	18,29	28,76	57,24	5,46
CMIG4	127	2,92	10,11	246,23	5,08
ITSA4	121	5,62	8,85	57,47	4,23
BBAS3	13	11,33	24,03	112,09	1,23

Fonte: Autores (2025)

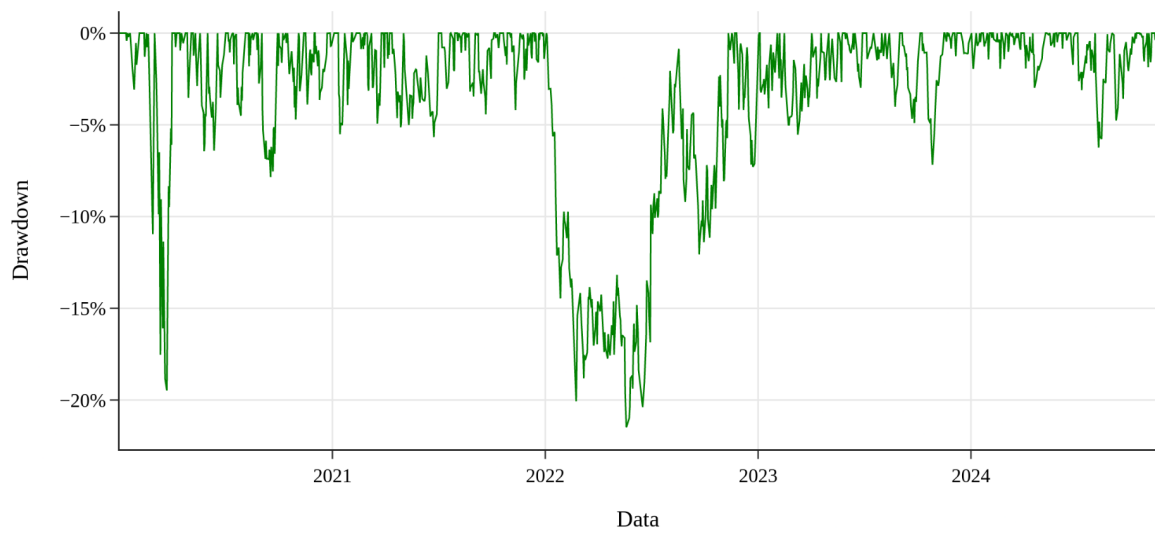
APÊNDICE C - Gráficos do histórico de *drawdown* das estratégias e *benchmarks*

Gráfico C.1 - Histórico de *drawdown* (IA-FS-R2).



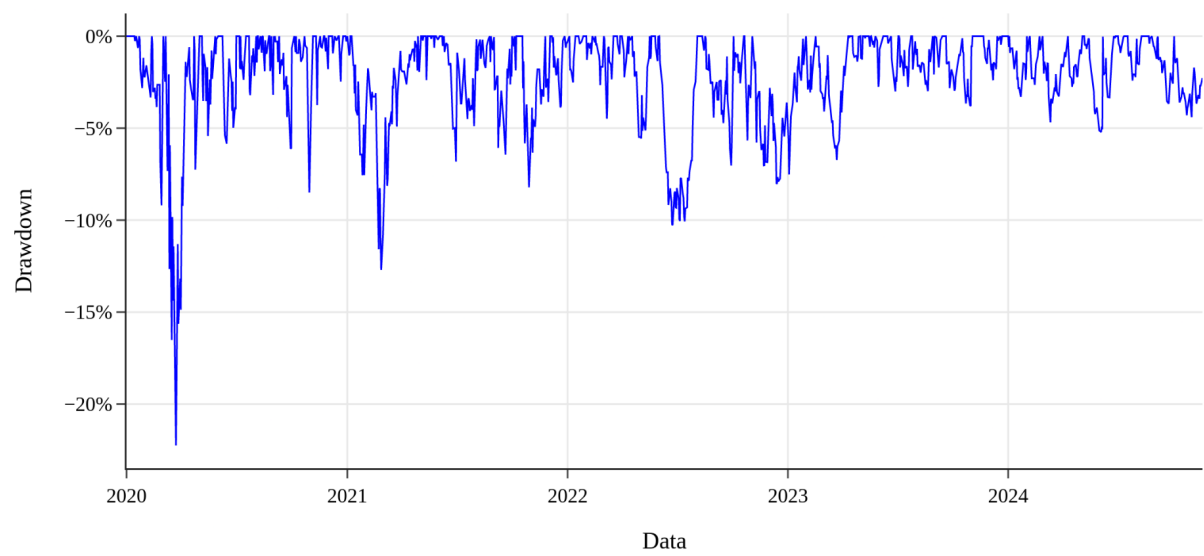
Fonte: Autores (2025)

Gráfico C.2 - Histórico de *drawdown* (IVVB11).



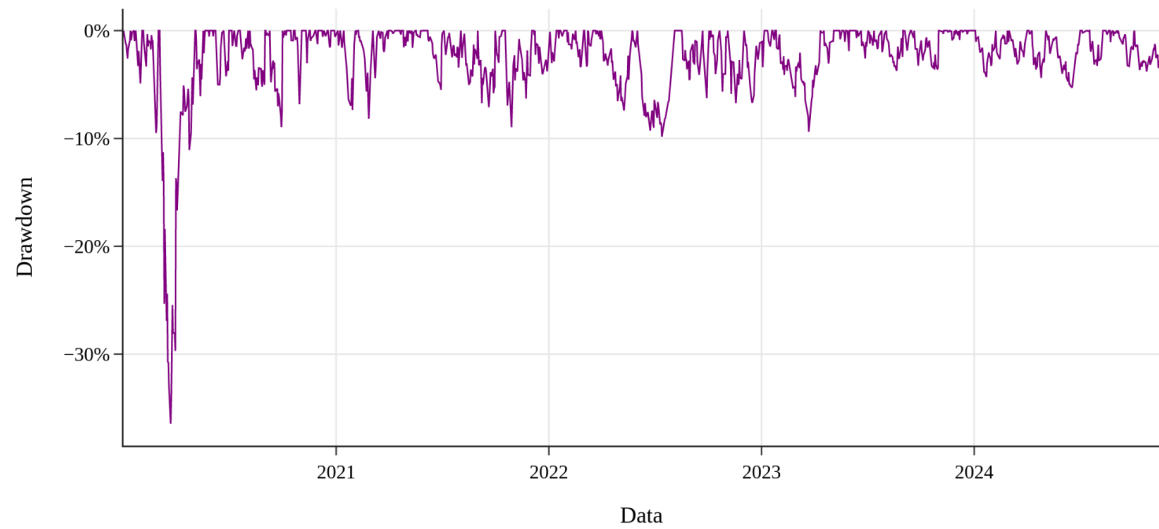
Fonte: Autores (2025)

Gráfico C.3 - Histórico de *drawdown* (IA-FS-R4).



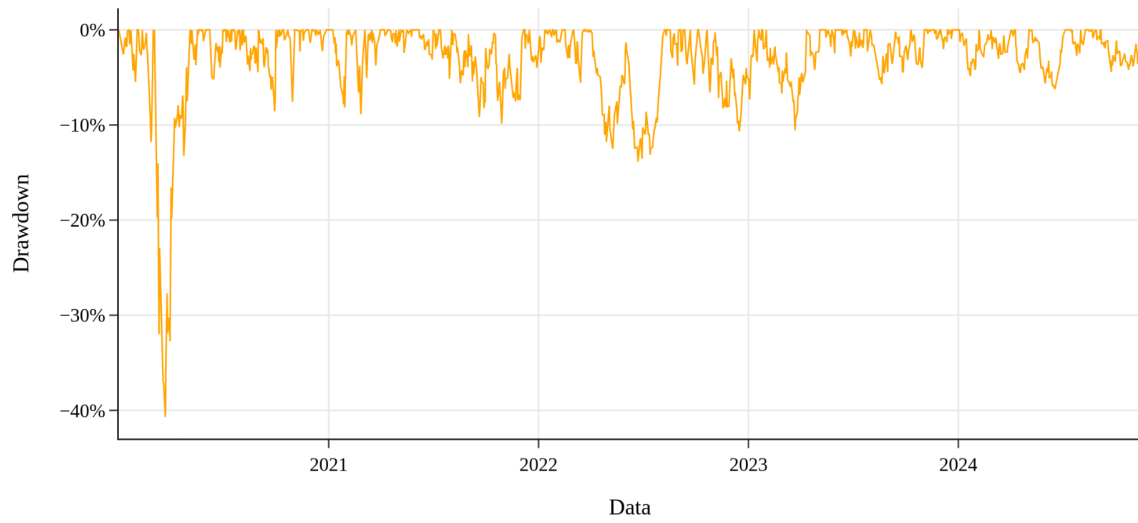
Fonte: Autores (2025)

Gráfico C.4 - Histórico de *drawdown* (DIVO11).



Fonte: Autores (2025)

Gráfico C.5 - Histórico de *drawdown* (BOVA11).



Fonte: Autores (2025)