

LUAN THIAGO DE SOUZA RIBEIRO

**INTELIGÊNCIA ARTIFICIAL GENERATIVA NA
CONSTRUÇÃO DE CARTEIRAS DE
INVESTIMENTOS: UM ESTUDO
EXPERIMENTAL COM LLMS NA B3**

São Paulo
2025

LUAN THIAGO DE SOUZA RIBEIRO

**INTELIGÊNCIA ARTIFICIAL GENERATIVA NA
CONSTRUÇÃO DE CARTEIRAS DE
INVESTIMENTOS: UM ESTUDO
EXPERIMENTAL COM LLMS NA B3**

Dissertação apresentada à Escola Politécnica da Universidade de São Paulo para obtenção do Título de Bacharel em Engenharia de Produção.

São Paulo
2025

LUAN THIAGO DE SOUZA RIBEIRO

**INTELIGÊNCIA ARTIFICIAL GENERATIVA NA
CONSTRUÇÃO DE CARTEIRAS DE
INVESTIMENTOS: UM ESTUDO
EXPERIMENTAL COM LLMS NA B3**

Dissertação apresentada à Escola Politécnica da Universidade de São Paulo para obtenção do Título de Bacharel em Engenharia de Produção.

Orientador:

Pr. Dr. Renato de Oliveira Moraes

São Paulo
2025

Aos meus pais, Cleucio e Sandra

AGRADECIMENTOS

Agradeço primeiramente a Deus por toda essa jornada, em que não me faltou nada.

Aos meus pais, Cleucio e Sandra, expresso minha mais profunda gratidão. Esta conquista só foi possível graças à dedicação incansável, à educação, ao amor e ao apoio incondicional que sempre me ofereceram. Nada do que alcancei teria sido possível sem vocês.

Ao meu irmão, Leonardo, agradeço pelo companheirismo constante, pela amizade sincera e por todo apoio ao longo desses anos.

À minha namorada, Camilla, que esteve ao meu lado em cada etapa deste ciclo. Obrigado por todo amor, motivação, apoio constante e por tornar esse caminho mais leve, mesmo nos dias mais difíceis.

Aos meus amigos, que preencheram essa trajetória com risadas, leveza e acolhimento, tornando cada passo mais tranquilo e memorável.

Por fim, gostaria de agradecer ao meu orientador, o professor Renato Moraes, pelos ensinamentos, pelas correções criteriosas e pela dedicação com que acompanhou este trabalho.

*“Se eu vi mais longe, foi porque estava
sobre os ombros de gigantes”*

-- Sir Isaac Newton

RESUMO

O presente trabalho investiga o desempenho de carteiras de ações construídas por modelos de linguagem de larga escala (LLMs) no mercado acionário brasileiro, com foco em avaliar sua eficiência relativa quando comparadas a benchmarks tradicionais e a fundos de investimento. Para isso, foram geradas oito carteiras equiponderadas por quatro modelos de IA, em versões com e sem alimentação de dados, e analisadas ao longo do período de agosto a outubro de 2025. O estudo empregou métricas de desempenho amplamente consolidadas, como rentabilidade acumulada, Índice de Sharpe e Alfa de Jensen, além de testes estatísticos para comparação entre grupos.

Os resultados indicaram que todas as carteiras superaram o CDI no intervalo analisado e apresentaram retornos próximos aos observados em fundos de gestão ativa, ainda que influenciados por um ambiente de forte valorização do Ibovespa. Observou-se, porém, que o número limitado de carteiras e o curto horizonte temporal restringiram a robustez estatística das conclusões. Apesar de evidenciarem potencial em aplicações institucionais, os modelos não se mostraram suficientemente consistentes para recomendações autônomas a investidores de varejo. O estudo aponta caminhos promissores para trabalhos futuros, especialmente a ampliação da amostra, a inclusão de diferentes regimes de mercado e a construção de delineamentos experimentais mais abrangentes.

Palavras-Chave – Mercado de capitais; Inteligência artificial; Modelos de linguagem; Delineamento experimental; Construção de carteiras; Análise de desempenho; Índice de Sharpe; Alfa de Jensen; ANOVA

ABSTRACT

This study investigates the performance of equity portfolios constructed by large language models (LLMs) in the Brazilian stock market, with the objective of assessing their relative efficiency when compared to traditional benchmarks and professional investment funds. Eight equally weighted portfolios were generated by four AI models, in versions with and without structured data feeding, and evaluated over the period from August to October 2025. The analysis employed well-established performance metrics, including cumulative returns, Sharpe Ratio, and Jensen's Alpha, in addition to statistical tests for group comparison.

The results indicate that all portfolios outperformed the CDI benchmark during the analyzed period and achieved returns comparable to those of active equity funds, although their performance was influenced by a strong upward trend in the Ibovespa. However, the limited number of portfolios and the short time horizon reduced the statistical robustness of the findings. While the models demonstrate potential for institutional applications, they did not exhibit sufficient consistency to support autonomous investment recommendations for retail investors. The study highlights promising avenues for future research, particularly the expansion of the sample size, evaluation across different market regimes, and the development of more comprehensive experimental designs.

Keywords – Capital markets; Artificial intelligence; Language models; Experimental design; Portfolio construction; Performance analysis; Sharpe Ratio; Jensen's Alpha; ANOVA.

LISTA DE FIGURAS

1	Estrutura SFN	18
2	Evolução de Investidores em Renda Variável no Brasil)	21
3	Faixa Etária dos Investidores Brasileiros	21
4	Rentabilidade Acumulada do CDI	65
5	CDI vs Ibovespa	65
6	Tabela de Distribuição F de Snedecor a 5%	71
7	ChatGPT vs CDI vs Ibovespa	80
8	Copilot vs CDI vs Ibovespa	81
9	DeepSeek vs CDI vs Ibovespa	82
10	Gemini vs CDI vs Ibovespa	83

LISTA DE TABELAS

1	Decomposição da soma dos quadrados na ANOVA de dois fatores	42
2	Equações principais da Two-Way ANOVA	42
3	Tabela típica de resultados da ANOVA de dois fatores	43
4	Carteiras geradas pela IA Copilot	62
5	Carteiras geradas pela IA Gemini	62
6	Carteiras geradas pela IA ChatGPT	63
7	Carteiras geradas pela IA DeepSeek	63
8	Carteiras geradas pela IA DeepSeek	66
9	Carteiras geradas pela IA ChatGPT	66
10	Carteiras geradas pela IA Gemini	67
11	Carteiras geradas pela IA Copilot	67
12	Índice de Sharpe das Carteiras Geradas (Ago–Out/2025)	68
13	Covariância, Variância e Coeficiente Beta das Carteiras (Ago–Out/2025) .	69
14	Alfa de Jensen Trimestral das Carteiras (Ago–Out/2025)	69
15	Rentabilidade média (%) por IA e condição de alimentação de dados . . .	70
16	ANOVA de dois fatores (sem repetição) — até a estatística F	71
17	Comparação: F calculado vs. F crítico (Snedecor, 5%)	71
18	Ativos mais frequentes nas oito carteiras de IA	73
19	Distribuição setorial das ações selecionadas nas carteiras	75
20	Ranking comparativo das carteiras em relação ao CDI e ao Ibovespa	86
21	Índice de Sharpe e Alfa de Jensen (trimestral) das carteiras	87
22	Empresas listadas na B3	97

SUMÁRIO

1	Introdução	12
1.1	Motivação	13
1.2	Descrição do Problema	13
1.3	Descrição dos Objetivos	14
2	Revisão da Literatura	16
2.1	Mercado de Capitais e Ações	17
2.1.1	Estrutura do Mercado	17
2.1.2	Mercado Acionário	19
2.1.3	Tipos de Ações	19
2.2	Perfil do Investidor Brasileiro	20
2.3	Risco e Retorno em Finanças	22
2.3.1	Conceito de Risco Sistemático e Não Sistemático	23
2.4	Teoria Moderna do Portfólio	24
2.4.1	Formulação Matemática	25
2.4.2	Fronteira Eficiente e Diversificação	27
2.5	Benchmarks e Avaliação de Carteiras	28
2.5.1	CDI como Proxy de Taxa Livre de Risco	29
2.5.2	Ibovespa como Benchmark de Mercado	31
2.5.3	Outras Referências (IBrX, SMILL e Índices Setoriais)	32
2.5.3.1	IBrX 50 e IBrX 100	32
2.5.3.2	SMILL (Índice Small Caps)	33
2.5.3.3	Índices Setoriais (IFNC, IMAT, ICON e outros)	34
2.6	Métricas de Desempenho	34

2.6.1	Índice de Sharpe	35
2.6.2	Beta	36
2.6.3	Alpha de Jensen	36
2.6.4	Information Ratio	37
2.7	Estatística Aplicada à Análise de Carteiras	37
2.7.1	Conceitos Fundamentais de Estatística em Finanças	38
2.7.2	ANOVA de Dois Fatores	41
2.8	Inteligência Artificial e Finanças	44
2.8.1	Modelos de Linguagem de Larga Escala (LLMs)	45
2.8.2	Construção de Prompts e Engenharia de Prompting na IA	47
3	Materiais e Métodos	50
3.1	Delineamento da Pesquisa	50
3.2	Universo e Amostra	52
3.3	Construção da Carteira de Investimentos	54
3.4	Análise de Rentabilidade	56
3.5	Análise Estatística de Rentabilidade	58
4	Resultados	60
4.1	Composição das Carteiras Geradas pelas Inteligências Artificiais	60
4.2	Análises Implementadas	63
4.2.1	Rentabilidade do CDI e Ibovespa	64
4.2.2	Rentabilidade das Carteiras Geradas	66
4.2.3	Índice Sharpe	67
4.2.4	Alfa de Jensen	68
4.3	ANOVA de Dois Fatores	69
5	Análise dos Resultados	72

5.1	Análise das Carteiras de Investimento	72
5.2	Análise do CDI e Ibovespa	76
5.3	Análise das Rentabilidades	79
5.3.1	Chat GPT	80
5.3.2	Copilot	81
5.3.3	DeepSeek	82
5.3.4	Gemini	83
5.3.5	Comparação Consolidada vs CDI e Ibovespa	83
5.3.6	Indíce Sharpe e Alfa de Jensen	87
5.3.7	Análise Estatística	89
6	Conclusões	91
Referências		94
Apêndice A		97

1 INTRODUÇÃO

O mercado de capitais desempenha um papel central na estrutura financeira moderna ao facilitar o fluxo de recursos entre agentes superavitários e deficitários, permitindo que excedentes de poupança sejam direcionados para investimentos produtivos e sustentando o crescimento econômico de longo prazo (PRESENTE, 2019; BARTONCELLO, 2022). Como alternativa ao financiamento bancário tradicional, o mercado de capitais oferece instrumentos capazes de viabilizar captações diretas por meio de ações, debêntures e outros valores mobiliários, ampliando a capacidade de investimento das empresas e contribuindo para o fortalecimento institucional do sistema financeiro brasileiro (CVM, 2024; BCB, 2024).

No âmbito desse ecossistema, a atuação das bolsas de valores, como a B3, assegura a infraestrutura necessária para a negociação eficiente e transparente dos ativos financeiros, promovendo liquidez, alinhamento informacional e segurança jurídica aos participantes do mercado (B3, 2024). O mercado primário, responsável pela emissão inicial de títulos, e o mercado secundário, no qual se desenvolve a maior parte das negociações, funcionam de forma integrada e desempenham papéis complementares na formação de preços e na mobilidade do capital (PRESENTE, 2019; BARTONCELLO, 2022). A literatura demonstra que a existência de mercados secundários ativos reduz a incerteza dos investidores ao proporcionar mecanismos contínuos de liquidez e realocação de portfólios (GOMES, 2011).

A análise do mercado acionário brasileiro revela uma estrutura heterogênea, composta por empresas de grande porte e elevada liquidez, as chamadas blue chips, bem como por empresas de menor capitalização, que apresentam maiores níveis de volatilidade e risco (GOMES, 2011). Esses padrões de risco e retorno influenciam diretamente o processo de construção de carteiras, demandando abordagens analíticas cada vez mais sofisticadas (CASACCIA, 2011). Nesse cenário, os avanços recentes em inteligência artificial, especialmente com o surgimento de modelos de linguagem de larga escala (LLMs), têm ampliado as possibilidades de apoio à tomada de decisão financeira, oferecendo ferramentas capazes

de interpretar textos complexos, identificar padrões e gerar recomendações estruturadas (TIAN; GUO, 2024; NIE et al., 2025).

1.1 Motivação

A dinâmica do mercado financeiro contemporâneo caracteriza-se por elevada complexidade, marcada por assimetrias de informação, instabilidade macroeconômica e crescente volume de dados disponíveis aos investidores. Nesse contexto, a eficiência informacional e a capacidade de processamento se tornam elementos essenciais para a tomada de decisão. O mercado de capitais, por sua própria natureza, depende da interação contínua entre emissores, intermediários e investidores para formar preços e alocar recursos de maneira eficiente (PRESENTE, 2019; BARTONCELLO, 2022).

Dentro do mercado acionário brasileiro, essa complexidade se intensifica pela coexistência de empresas com diferentes graus de liquidez e risco, exigindo análises que integrem fundamentos econômicos, indicadores financeiros e métricas de risco ajustado (GOMES, 2011; CASACCIA, 2011). A distinção entre ações de alta liquidez, geralmente representadas por blue chips, e empresas de menor porte reforça a necessidade de metodologias robustas capazes de absorver e interpretar múltiplos fatores simultâneos (B3, 2024). Diante desse ambiente competitivo e dinâmico, surge a demanda por ferramentas analíticas avançadas que possam apoiar investidores em suas decisões de alocação de recursos.

Os avanços recentes em inteligência artificial, sobretudo com o desenvolvimento de modelos de linguagem de larga escala (LLMs), oferecem uma oportunidade singular para aprimorar processos de análise e seleção de ativos. Tais modelos demonstram capacidade de sintetizar grandes volumes de informação textual, identificar relações latentes em dados não estruturados e gerar sinais preditivos de retorno a partir de notícias, relatórios corporativos e outros documentos relevantes (TIAN; GUO, 2024). Abordagens multiagentes aprofundam ainda mais essa capacidade ao distribuir tarefas analíticas entre agentes especializados, produzindo avaliações mais consistentes e estruturadas (NIE et al., 2025).

1.2 Descrição do Problema

Nos últimos anos, o mercado de capitais brasileiro tem passado por um processo significativo de expansão. A B3 registrou, no primeiro trimestre de 2024, um total de

5,1 milhões de investidores pessoa física ativos, refletindo um crescimento expressivo no acesso ao mercado acionário nacional (B3, 2024). Embora esse número represente apenas cerca de 2,4% da população brasileira, considerando as estimativas do IBGE (2024), ele demonstra um aumento contínuo no interesse da população pela renda variável. Além disso, esses investidores já respondem por aproximadamente 18% do total investido em ações no país (B3, 2024).

Esse crescimento tem sido impulsionado por diversos fatores, como a digitalização de plataformas de investimento, a redução de custos operacionais e a maior democratização das informações financeiras, e ao mesmo tempo, o ambiente de investimentos permanece complexo, exigindo análises contínuas sobre riscos, fundamentos corporativos e contexto macroeconômico (CASACCIA, 2011; GOMES, 2011). Essa crescente demanda de investidores de varejo, ou seja, investidores pessoa física, evidencia um desafio central: investidores iniciantes e intermediários nem sempre possuem conhecimento técnico suficiente para interpretar dados, relatórios e eventos que influenciam decisões de alocação.

Isso amplia o espaço para ferramentas capazes de apoiar a análise financeira, sobretudo aquelas fundamentadas em inteligência artificial. Modelos LLMs, quando mobilizados de forma adequada, podem reduzir barreiras de entrada ao traduzir informações complexas em recomendações comprehensíveis, oferecendo suporte à tomada de decisão mesmo em cenários de elevada volatilidade (TIAN; GUO, 2024). O avanço de arquiteturas multi-agente reforça essa tendência ao permitir a especialização de agentes e a coordenação entre diferentes estratégias analíticas (NIE et al., 2025), sugerindo que novas tecnologias podem auxiliar na construção de carteiras mais consistentes e informadas.

1.3 Descrição dos Objetivos

O presente trabalho tem como objetivo principal investigar se diferentes modelos de inteligência artificial generativa (LLMs) são capazes de gerar carteiras de ações que superem benchmarks tradicionais do mercado brasileiro, como o CDI e o Ibovespa (B3, 2024; BCB, 2024). Ao analisar carteiras construídas com base em respostas de LLMs, busca-se determinar se esses modelos conseguem estruturar decisões de alocação comparáveis ou superiores às estratégias convencionais utilizadas por analistas humanos (TIAN; GUO, 2024).

Como objetivo secundário, o estudo busca avaliar se a inclusão de informações complementares, como relatórios corporativos, demonstrações financeiras e indicadores macroe-

conômicos, contribui para aumentar a consistência, precisão e desempenho das carteiras geradas. Evidências recentes indicam que LLMs podem enriquecer processos de análise ao integrar dados não estruturados e ao operar dentro de estruturas multiagentes, permitindo a construção de estratégias analíticas distribuídas e coordenadas (NIE et al., 2025).

Dessa forma, este trabalho pretende oferecer contribuições tanto ao campo da análise financeira quanto à discussão contemporânea sobre o uso prático de IAs generativas no mercado de capitais brasileiro, ampliando o entendimento sobre como essas tecnologias podem apoiar investidores na tomada de decisões em ambientes competitivos, dinâmicos e informacionalmente complexos.

2 REVISÃO DA LITERATURA

A revisão da literatura deste trabalho tem como objetivo consolidar o conjunto de fundamentos teóricos que sustenta a investigação sobre a aplicação de modelos de linguagem de larga escala (LLMs) na construção de carteiras de investimento. O capítulo está organizado em nove blocos temáticos, que percorrem desde as bases institucionais do funcionamento do mercado de capitais brasileiro até a fronteira tecnológica atual representada pelos modelos generativos e pelas técnicas de prompt engineering. A progressão dos temas reflete a lógica metodológica do estudo: iniciar pelos conceitos essenciais de finanças, avançar para os métodos quantitativos utilizados na avaliação de desempenho e culminar nas ferramentas tecnológicas que permitem a implementação prática dos experimentos conduzidos.

Os blocos iniciais (2.1 e 2.2) apresentam o contexto estrutural e comportamental do mercado de capitais brasileiro, discutindo o papel da bolsa de valores, o processo de intermediação financeira, os mecanismos de negociação de ações e a evolução recente do perfil do investidor nacional. Essa contextualização evidencia como transformações regulatórias, tecnológicas e macroeconômicas moldaram o ambiente de investimento no país e criaram condições para a adoção de estratégias mais quantitativas e orientadas a dados – cenário no qual o uso de inteligência artificial passa a ser não apenas possível, mas crescente.

Os blocos centrais (2.3 a 2.6) aprofundam-se nos princípios fundamentais de avaliação de ativos e gestão de carteiras. São discutidos os conceitos de risco e retorno, as métricas clássicas de performance, a Teoria Moderna do Portfólio, o modelo CAPM e os indicadores de desempenho ajustado ao risco, como o Índice Sharpe e o Alfa de Jensen. Em complemento, o capítulo apresenta os métodos estatísticos empregados em finanças, incluindo medidas descritivas, variância, covariância, correlação e técnicas inferenciais, com destaque para a utilização da ANOVA de dois fatores na comparação de desempenhos de portfólios. Esses tópicos constituem o eixo quantitativo que fundamenta a análise empírica desenvolvida posteriormente.

Por fim, os blocos finais (2.7 e 2.8) tratam do papel crescente da inteligência artificial no mercado financeiro, descrevendo a evolução das ferramentas computacionais, desde modelos tradicionais de aprendizado de máquina até a emergência dos modelos de linguagem de larga escala. São discutidas suas aplicações em análise textual, previsão, classificação de ativos, suporte à tomada de decisão e construção de carteiras. O capítulo também explora a lógica do prompt engineering e suas implicações metodológicas, destacando como a interação entre fundamentos financeiros, estatística aplicada e IA generativa estabelece o pano de fundo conceitual para os experimentos realizados neste estudo.

2.1 Mercado de Capitais e Ações

O mercado de capitais constitui um dos pilares fundamentais do sistema financeiro, desempenhando papel essencial na intermediação entre poupadore e tomadores de recursos, ao canalizar poupanças individuais e institucionais para o financiamento de atividades produtivas (BANCO CENTRAL DO BRASIL, s.d.). Por meio dele, empresas emitem títulos e valores mobiliários com o objetivo de captar recursos de longo prazo, enquanto investidores buscam oportunidades de retorno ajustado ao risco. Essa dinâmica contribui para o desenvolvimento econômico ao promover eficiência alocativa, liquidez e diversificação de investimentos dentro do Sistema Financeiro Nacional (CVM, s.d.).

Dentro desse contexto, o mercado acionário se destaca como principal segmento de capitalização das empresas, possibilitando tanto o financiamento de projetos quanto a ampliação de patrimônio líquido. A negociação de ações na B3, principal bolsa de valores do Brasil, reflete o comportamento agregado das expectativas de investidores, influenciado por variáveis macroeconômicas, políticas e corporativas (B3, s.d.). Assim, compreender a estrutura, o funcionamento e a regulação do mercado de capitais e de ações é fundamental para analisar o desempenho dos ativos financeiros e interpretar os mecanismos de formação de preço e de rentabilidade que embasam a análise de investimentos.

2.1.1 Estrutura do Mercado

O mercado de capitais constitui um dos segmentos do Sistema Financeiro Nacional (SFN), responsável por possibilitar a emissão e circulação de valores mobiliários como ações, debêntures e fundos. Sua principal função é transformar a poupança de investidores em capital para empresas e governo, contribuindo para o financiamento de projetos e a alocação eficiente de recursos (BANCO CENTRAL DO BRASIL, s.d.).

	Moeda, crédito, capitais e câmbio	Seguros privados	Previdência fechada	
Órgãos normativos	CMN Conselho Monetário Nacional	CNSP Conselho Nacional de Seguros Privados	CNPC Conselho Nacional de Previdência Complementar	
Supervisores	BCB Banco Central do Brasil	CVM Comissão de Valores Mobiliários	Susep Superintendência de Seguros Privados	
Operadores	 Bancos e caixas econômicas  Administradoras de consórcios	 Bolsa de valores  Seguradoras e Resseguradores	 Cooperativas de crédito  Corretoras e distribuidoras*	 Entidades abertas de previdência  Entidades fechadas de previdência complementar (fundos de pensão)
 Instituições de pagamento**  Demais instituições não bancárias	Bolsa de mercadorias e futuros		 Sociedades de capitalização	

Figura 1: Estrutura SFN
Fonte: Banco Central do Brasil, s.d.

No âmbito regulatório, o Conselho Monetário Nacional (CMN) exerce a função de instância máxima do SFN, enquanto a Comissão de Valores Mobiliários (CVM), instituída pela Lei nº 6.385/1976, atua na normatização, fiscalização e estímulo ao desenvolvimento do mercado de capitais, assegurando transparência e proteção ao investidor (BANCO CENTRAL DO BRASIL, s.d.; BRASIL, 1976).

Como infraestrutura de mercado, a B3 – Brasil, Bolsa, Balcão opera ambientes de bolsa e balcão e concentra a pós-negociação: atua como contraparte central e sistema de liquidação por meio da Câmara B3, além de manter funções de registro e central depositária. Seus regulamentos e páginas institucionais também apresentam o PUMA Trading System como plataforma de negociação (B3, s.d.).

Quanto aos participantes, o ecossistema envolve investidores pessoa física, institucionais (fundos, seguradoras, previdência), não residentes, intermediários (corretoras e distribuidoras), bancos de investimento e emissores (companhias abertas, securitizadoras, etc.). As páginas oficiais listam e definem essas categorias: a B3 mantém a seção de participantes (e seus tipos/cadastros), a CVM reúne regras e procedimentos para investidores não residentes, e o Banco Central sintetiza o papel de corretoras/distribuidoras na intermediação (CVM, 2020).

2.1.2 Mercado Acionário

As ações são valores mobiliários que representam participação societária em empresas abertas, conferindo ao acionista direitos econômicos (como dividendos) e, conforme a classe e o estatuto, direitos políticos (voto), além do potencial de valorização do capital. No ordenamento brasileiro, tais características decorrem da Lei nº 6.404/1976 (Lei das S.A.) e são explicadas em materiais oficiais da CVM/Portal do Investidor e da B3. Esses documentos também ressaltam que as empresas emitem ações para captar recursos próprios (equity) e que o investidor, ao adquiri-las, torna-se coproprietário no limite de sua participação. (BRASIL, 1976; B3, s.d.).

No contexto de alocação de portfólio, as ações cumprem três papéis recorrentes. Primeiro, são ativos historicamente associados a um prêmio de risco em relação à renda fixa, o comumente chamado equity premium, o que ajuda a explicar o potencial de retorno superior no longo prazo (MEHRA; PRESCOTT, 1985). Segundo, contribuem para diversificação: por não se moverem perfeitamente em conjunto com outros ativos, permitem reduzir risco para um dado nível de retorno (MARKOWITZ, 1952). Terceiro, quanto à proteção contra inflação, a evidência é mista e, em geral, imperfeita: estudos indicam que ações não são um hedge completo para inflação esperada no curto prazo, embora possam preservar poder de compra em horizontes extensos. (FAMA; SCHWERT, 1977).

Quando comparadas a outros instrumentos, ações se distinguem de:

- Títulos públicos e de renda fixa: possuem sua remuneração conhecida por fórmula no momento da aplicação (B3, 2024);
- Fundos de investimento: são veículos coletivos com políticas e riscos próprios (B3, 2024);
- Derivativos contratos cujo valor deriva de um ativo subjacente, usados para hedge, alavancagem ou exposição tática.(B3, 2024).

2.1.3 Tipos de Ações

Visando o entendimento maior sobre o mercado acionário, há a necessidade de descrever os tipos de subprodutos que existem nesse mercado, sendo eles:

- Ações ordinárias (ON) e preferenciais (PN): As primeiras garantem, como regra geral, o direito de voto em assembleias de acionistas, enquanto as preferenciais ofer-

recentes vantagens econômicas, como prioridade no recebimento de dividendos ou no reembolso de capital, ainda que, em alguns casos, restrinjam o direito de voto. Essas características derivam da Lei das S.A. e são detalhadas em materiais regulatórios da CVM e da B3 (BRASIL, 1976; BRASIL, 2022a; B3, 2024).

- Units: são pacotes negociados em conjunto que reúnem mais de uma classe de valores mobiliários, tipicamente combinações de ON e PN (ou ON e bônus), negociadas como uma única unidade. A própria B3 define units como “ativos compostos por mais de uma classe (...) comprados e vendidos como uma unidade”. (B3, s.d. a; B3, 2023).
- Blue chips: designa ações de empresas grandes, tradicionais e com alta liquidez, muitas vezes com peso relevante nos principais índices (ex.: Petrobras, Vale, Itaú). Materiais educativos da B3 descrevem “blue chip” como ação de grande liquidez e de empresas consolidadas; a ligação entre liquidez/relevância e presença em índices é reforçada pela metodologia do Ibovespa, que seleciona ativos de maior negociabilidade e representatividade do mercado. (B3, 2022; B3, 2020; B3, 2019).
- Penny stocks: designa ações de baixo preço unitário e baixa liquidez, associadas a maior volatilidade e risco de movimentos especulativos. Conteúdos da B3 destacam que ações abaixo de R\$ 1,00 podem ser classificadas como penny stocks, havendo casos em que companhias realizam grupamentos para evitar essa condição; a própria metodologia do Ibovespa exclui papéis negociados persistentemente abaixo de R\$ 1,00 de sua carteira teórica. (B3, 2024; B3, 2025; B3, 2023).

2.2 Perfil do Investidor Brasileiro

A literatura setorial recente e relatórios de mercado mostram um processo de massificação do acesso à renda variável no Brasil. Entre 2018 e 2023, os CPFs com posição em ações/BDRs/ETFs na B3 passaram de aproximadamente 700 mil para aproximadamente 5 milhões (Figura 2), fenômeno associado à digitalização das corretoras e ao ciclo de juros que favoreceu o risco (XP INC., 2023). Apesar do estudo ser referente a 2023, percebe-se que o número de manteve-se estável dado que em 2025, temos 5,1 milhões de investidores.

Do ponto de vista demográfico, ocorreu uma rejuvenescimento da base. Em 2013, a faixa 25–39 anos representava 33% dos investidores; em 2023, atinge 49%. A participação de 40–59 anos caiu de 40% para 26%, enquanto 19–24 anos avançou para 18% (Figura 3). A evidência converge para um perfil digital-first, jovem-adulto, com horizonte de



Figura 2: Evolução de Investidores em Renda Variável no Brasil)

Fonte: XP INC., 2023

investimento mais longo e alta propensão a experimentar ferramentas algorítmicas (XP INC., 2023).

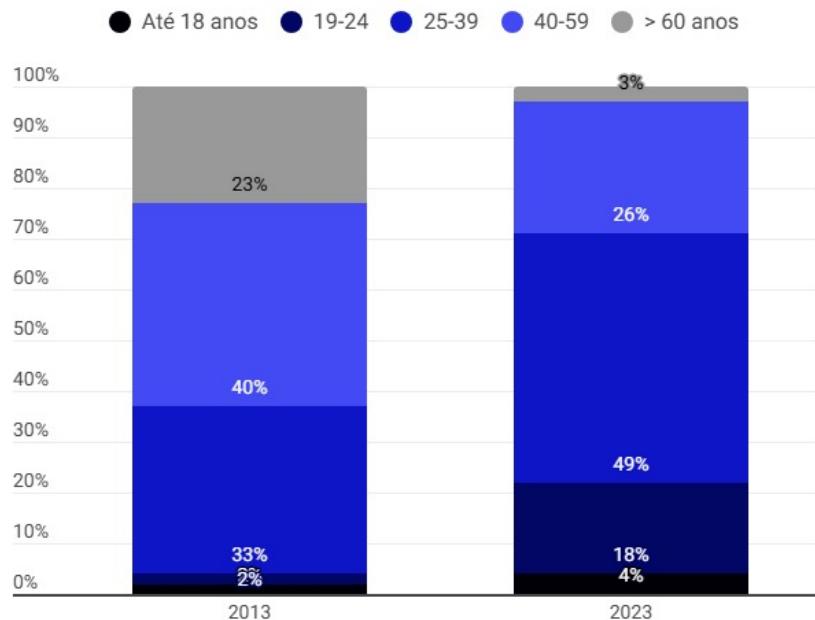


Figura 3: Faixa Etária dos Investidores Brasileiros

Fonte: XP INC., 2023

Quanto ao perfil de entrada e comportamento inicial, levantamento de mercado divulgado pela InfoMoney indica que o brasileiro ingressa na bolsa com tíquete médio de R\$ 352,50, que 50% dos novos entrantes estão na faixa 25–39 anos e que muitos já iniciam com mais de cinco ativos em carteira, sendo isso um sinal de busca de diversificação mínima desde o início (INFOMONEY, s.d., com base em dados da B3). A mesma tendência

aparece em séries institucionais: entre 2016 e 2T21, a B3 registra redução das carteiras com “apenas 1 ativo” e aumento daquelas com 4 ou mais, indicando amadurecimento progressivo da diversificação no varejo (B3, 2021; ver gráfico de diversificação).

2.3 Risco e Retorno em Finanças

A relação entre risco e retorno constitui o eixo central da teoria financeira moderna e baseia-se no princípio de que investidores racionais buscam maximizar retornos esperados em função do nível de risco que estão dispostos a assumir. O conceito de risco, nesse contexto, expressa a incerteza associada aos resultados futuros de um investimento, enquanto o retorno representa a compensação esperada pelo capital aplicado. Essa interdependência entre risco e retorno é o fundamento das decisões de alocação de recursos, da especificação de ativos e da construção de carteiras eficientes, sendo formalizada por modelos como a Teoria Moderna do Portfólio (MARKOWITZ, 1952) e o Capital Asset Pricing Model – CAPM (SHARPE, 1964).

Em termos práticos, o risco pode ser decomposto em duas dimensões: o risco sistemático, relacionado a fatores macroeconômicos que afetam o mercado como um todo, e o risco não sistemático, específico de empresas ou setores e passível de diversificação. A análise da volatilidade, da covariância e do coeficiente beta (β) permite quantificar essas exposições e compreender a sensibilidade de um ativo em relação ao mercado. Já o retorno é avaliado a partir de indicadores como retorno médio, retorno esperado e prêmio de risco, que refletem a remuneração exigida pelo investidor em troca da exposição à incerteza (FAMA; FRENCH, 1992).

A compreensão da relação risco-retorno é indispensável para a formulação de estratégias de investimento e para o desempenho de carteiras em ambientes de diferentes condições de mercado. Em economias emergentes, como a brasileira, a volatilidade macroeconômica, as oscilações de taxa de juros e a instabilidade política intensificam a relevância dessa relação, exigindo modelos de análise mais robustos e métodos quantitativos que capturem a dinâmica temporal do risco. Assim, o estudo do equilíbrio entre risco e retorno transcende o aspecto teórico, constituindo um pilar para a gestão de portfólios, a mensuração de performance e a tomada de decisão em finanças corporativas e de mercado (FAMA; FRENCH, 1992; MARKOWITZ, 1952; SHARPE, 1964).

2.3.1 Conceito de Risco Sistemático e Não Sistemático

A literatura financeira moderna sustenta que o risco total de um ativo não é homogêneo, mas resulta da combinação de diferentes fontes de incerteza. Esse risco pode ser decomposto em duas dimensões principais: o risco não sistemático, também denominado idiossincrático ou diversificável, e o risco sistemático, igualmente conhecido como risco de mercado ou não diversificável. Essa distinção é fundamental para compreender os mecanismos de precificação dos ativos, bem como as estratégias de diversificação utilizadas por investidores na busca de carteiras mais eficientes (SHARPE, 1964).

O risco não sistemático corresponde a fatores específicos de uma empresa ou setor, que afetam de forma isolada o desempenho de determinados ativos. Entre suas principais fontes destacam-se questões de governança corporativa, falhas na gestão, disputas trabalhistas, inovações tecnológicas, acidentes operacionais ou mudanças regulatórias setoriais. Como esses eventos tendem a impactar apenas uma fração restrita do mercado, o investidor pode mitigar esse risco por meio da diversificação. A lógica é que resultados negativos em uma empresa podem ser compensados por desempenhos positivos em outras, reduzindo a variabilidade agregada do portfólio. Nesse sentido, manter uma carteira concentrada em poucos ativos expõe o investidor a riscos não remunerados, pois tais incertezas podem ser neutralizadas com relativa facilidade (BRUSA et al., 2010).

Já o risco sistemático deriva de fatores macroeconômicos que afetam o mercado como um todo, de forma simultânea, ainda que com intensidades diferentes entre setores e empresas. Choques de política monetária, variações abruptas nas taxas de juros, episódios de inflação elevada, mudanças fiscais, crises financeiras globais e instabilidade política são exemplos típicos. Esse conjunto de variáveis atinge de forma generalizada todos os ativos, o que torna impossível eliminá-lo por meio da diversificação. Por essa razão, mesmo portfólios amplos e diversificados permanecem expostos a esse componente irreduzível de risco, que representa a parcela inevitável da incerteza de mercado (SHARPE, 1964).

Do ponto de vista teórico, a distinção entre os dois tipos de risco remonta ao trabalho de Harry Markowitz (1952), cuja Teoria Moderna do Portfólio demonstrou que a variância de uma carteira depende não apenas das variâncias individuais dos ativos, mas também das covariâncias entre eles. Essa formulação revelou que a diversificação é eficaz para reduzir a parcela idiossincrática do risco, mas não elimina a variabilidade relacionada a fatores macroeconômicos. Embora Markowitz não tenha utilizado explicitamente a terminologia de risco sistemático e não sistemático, sua modelagem já antecipava essa separação conceitual. Posteriormente, autores como Brusa et al. (2010) reforçaram que

essa distinção é essencial para a compreensão dos mecanismos de precificação de ativos.

A formalização clara dessa diferenciação surge com o Capital Asset Pricing Model (CAPM), proposto por William Sharpe em 1964. Em seu artigo seminal, Sharpe (1964) introduziu o conceito de que apenas o risco sistemático deve ser considerado na determinação do retorno esperado de um ativo. Nesse modelo, o risco sistemático é mensurado pelo coeficiente beta (β), que expressa a sensibilidade do retorno de um ativo em relação às variações do retorno do mercado. Um beta maior que 1 indica que o ativo tende a amplificar as oscilações do mercado, enquanto um beta menor que 1 indica uma resposta mais contida. O risco não sistemático, por sua vez, corresponde à parcela residual não correlacionada ao mercado e, portanto, não relevante para a precificação em equilíbrio (SHARPE, 1964).

Essa formulação leva a duas implicações práticas fundamentais. Em primeiro lugar, o mercado não remunera o risco diversificável. Isso significa que investidores que optam por manter carteiras pouco diversificadas estão assumindo riscos desnecessários, sem qualquer compensação adicional em termos de retorno esperado (BRUSA et al., 2010). Em segundo lugar, a remuneração do investidor decorre exclusivamente da exposição ao risco sistemático, capturado pelo beta. Assim, ativos com maior sensibilidade às oscilações do mercado devem oferecer retornos esperados mais elevados para compensar o risco inevitável que representam (SHARPE, 1964).

Assim, a distinção entre risco sistemático e não sistemático constitui uma das bases centrais da moderna teoria financeira. Além de justificar a prática da diversificação como mecanismo para reduzir riscos ineficientes, fornece a estrutura lógica que fundamenta a relação entre risco e retorno nos modelos de precificação de ativos (BRUSA et al., 2010; SHARPE, 1964).

2.4 Teoria Moderna do Portfólio

A Teoria Moderna do Portfólio (Modern Portfolio Theory – MPT), introduzida por Harry Markowitz em seu artigo seminal *Portfolio Selection*, publicado no *Journal of Finance* em 1952, constitui um dos marcos fundacionais da moderna teoria financeira. Seu objetivo principal é propor uma abordagem racional e quantitativa para a construção de carteiras de investimento, fundamentada na relação entre risco e retorno. Diferentemente das abordagens anteriores, que avaliavam ativos de forma isolada, a MPT inaugura a visão de que os investidores devem considerar os ativos no contexto de sua contribuição

ao risco e ao retorno da carteira como um todo (MARKOWITZ, 1952).

A formulação apresentada por Markowitz parte da premissa de que o investidor busca simultaneamente maximizar o retorno esperado e minimizar a variância dos retornos, estabelecendo um modelo matemático em que o risco de uma carteira não depende apenas da soma dos riscos individuais, mas também das covariâncias entre os ativos. Essa percepção introduziu formalmente o conceito de diversificação eficiente, mostrando que a inclusão de ativos com correlação imperfeita pode reduzir o risco total sem comprometer proporcionalmente o retorno esperado (MARKOWITZ, 1952).

2.4.1 Formulação Matemática

A formulação matemática da Teoria Moderna do Portfólio (TMP) consolidou-se como um dos avanços mais relevantes da teoria financeira moderna, ao estabelecer que investidores devem avaliar carteiras com base na relação entre retorno esperado e risco total (MARKOWITZ, 1952). A principal contribuição foi demonstrar que o risco de uma carteira não se resume à soma ponderada das variâncias individuais, mas depende também das covariâncias entre os ativos, introduzindo formalmente o conceito de diversificação (MARKOWITZ, 1952; ELTON; GRUBER, 1997).

O retorno esperado de uma carteira é representado pela média ponderada dos retornos esperados dos ativos que a compõem, conforme:

$$E(R_p) = \sum_{i=1}^n w_i E(R_i) \quad (2.1)$$

Em que $E(R_p)$ corresponde ao retorno esperado da carteira, w_i ao peso de cada ativo, e $E(R_i)$ ao retorno esperado do ativo i . Essa formulação demonstra que o retorno de uma carteira depende diretamente da contribuição proporcional de cada ativo, estabelecendo uma base linear para a análise de desempenho (MARKOWITZ, 1952; ELTON; GRUBER, 1997).

O risco da carteira, contudo, apresenta caráter quadrático e é definido pela variância dos retornos. A expressão geral é:

$$\sigma_p^2 = \sum_{i=1}^n \sum_{j=1}^n w_i w_j \sigma_{ij} \quad (2.2)$$

Em que σ_{ij} representa a covariância entre os ativos i e j . Essa equação evidencia que,

mesmo que os ativos possuam elevada variabilidade individual, a combinação de ativos com baixa correlação pode reduzir substancialmente o risco agregado (MARKOWITZ, 1952). Estudos posteriores confirmam que a diversificação eficaz resulta da inclusão de ativos com covariâncias imperfeitas, reforçando a lógica de que o risco específico pode ser mitigado em nível de portfólio (ELTON; GRUBER, 1997).

A construção da matriz de variância-covariância tornou-se, desde então, um elemento fundamental para a prática de otimização de carteiras, pois permite organizar as variâncias individuais na diagonal e as covariâncias nos elementos fora da diagonal (MARKOWITZ, 1952; ELTON; GRUBER, 1997). Entretanto, estimativas tradicionais dessa matriz apresentam problemas quando o número de ativos é grande em relação ao número de observações, tornando-a instável ou até não invertível. Para lidar com essa limitação, Ledoit e Wolf (2004) propuseram um estimador bem condicionado baseado em técnicas de shrinkage, no qual a matriz amostral é ajustada em direção a uma matriz alvo, geralmente a identidade, resultando em maior robustez estatística e aplicabilidade prática. Esse avanço é amplamente reconhecido como indispensável para a implementação da teoria em ambientes de alta dimensionalidade (LEDOIT; WOLF, 2004).

A formulação de média-variância também foi expandida para contextos intertemporais. Merton (1973), em seu modelo de precificação intertemporal de ativos (ICAPM), introduziu o uso de processos estocásticos em tempo contínuo para descrever a evolução dos preços. Nesse arcabouço, o retorno esperado de cada ativo é composto por um termo determinístico e por um termo aleatório associado à volatilidade, enquanto a covariância entre os ativos captura as interações instantâneas entre os processos de Wiener. Dessa forma, a teoria mantém sua lógica central, mas incorpora a dinâmica temporal, tornando-se aplicável em cenários de negociação contínua e revisão sucessiva de portfólios (MERTON, 1973).

Portanto, a formulação matemática da TMP estabelece que o retorno esperado de uma carteira é linear, enquanto o risco depende da estrutura quadrática das variâncias e covariâncias. Essa abordagem fundamenta não apenas o conceito de fronteira eficiente e a noção de diversificação, mas também sustenta modelos de extensão, como o ICAPM de Merton e os estimadores robustos de covariância de Ledoit e Wolf, que ampliam a aplicabilidade da teoria em cenários reais (MARKOWITZ, 1952; MERTON, 1973; ELTON; GRUBER, 1997; LEDOIT; WOLF, 2004).

2.4.2 Fronteira Eficiente e Diversificação

A noção de fronteira eficiente constitui um desdobramento direto da formulação de média-variância apresentada anteriormente. A partir do modelo de Markowitz (1952), em que o retorno esperado de uma carteira é uma média ponderada de retornos individuais e o risco depende das variâncias e covariâncias dos ativos, é possível identificar o conjunto de carteiras que oferece a melhor relação risco-retorno. Esse conjunto, denominado fronteira eficiente, representa as carteiras que maximizam o retorno para um dado nível de risco, ou, de forma equivalente, minimizam o risco para um dado retorno esperado (MARKOWITZ, 1952).

Graficamente, a fronteira eficiente é representada como a parte superior de uma curva côncava no espaço risco-retorno, onde a porção inferior é descartada por ser dominada em eficiência. O ponto de variância mínima global é particularmente relevante, pois indica a composição de ativos que apresenta o menor risco possível entre todas as combinações, independentemente dos retornos esperados (MARKOWITZ, 1952; MARTIN, 2020).

O princípio da diversificação é central nesse contexto. A inclusão de ativos com correlação imperfeita desloca a fronteira eficiente para fora, permitindo ao investidor reduzir o risco agregado sem sacrificar proporcionalmente o retorno. Esse resultado explica por que deter apenas o ativo de maior retorno esperado não é a solução ótima: a variabilidade conjunta entre ativos pode ser explorada para suavizar os resultados da carteira (MARKOWITZ, 1952; KASHEFI; MCKEE, 2001).

Aplicações empíricas confirmam que a diversificação setorial e internacional desloca a fronteira eficiente. Estudos sobre carteiras globais demonstram que a adição de ativos estrangeiros em carteiras domésticas melhora a relação risco-retorno, especialmente quando as correlações são baixas (DAVIS et al., 2013). Contudo, a crescente globalização e interdependência dos mercados têm aumentado as correlações médias entre países, reduzindo, embora não eliminando, os ganhos potenciais da diversificação internacional (KASHEFI; MCKEE, 2001).

A fronteira eficiente também se estende a outros mercados além das ações. Em carteiras de renda fixa, por exemplo, restrições relacionadas a duração e risco de crédito influenciam a forma da fronteira, exigindo abordagens complementares, como a programação linear, para calcular combinações ótimas (MARTIN, 2020). Esse resultado reforça a adaptabilidade da teoria para diferentes classes de ativos.

Além disso, a formalização da fronteira eficiente abriu espaço para métricas de ava-

liação do desempenho ajustado ao risco. O Índice de Sharpe, criado como uma medida que relaciona retorno e risco, é definido como:

$$S = \frac{E(R_p) - R_f}{\sigma_p}$$

Na qual $E(R_p)$ é o retorno esperado da carteira, R_f é a taxa livre de risco, e σ_p é o desvio-padrão dos retornos da carteira (NBER, 2014). Essa métrica permite comparar carteiras sob um mesmo critério de eficiência, indicando quais delas oferecem maior retorno em relação ao risco assumido. Apesar de sua ampla aceitação, o índice apresenta limitações em cenários nos quais os retornos não seguem a distribuição normal, já que ativos com assimetria ou caudas gordas podem distorcer sua interpretação. Por esse motivo, estudos mais recentes propõem medidas alternativas que incorporam momentos de ordem superior, como skewness e curtose, embora o Índice de Sharpe siga como a métrica mais utilizada pela simplicidade e praticidade de sua aplicação (NBER, 2014).

Em síntese, a fronteira eficiente demonstra que os investidores racionais devem escolher carteiras não pela maximização isolada do retorno, mas pela otimização da relação risco-retorno. A diversificação, seja setorial, internacional ou entre classes de ativos, é o mecanismo que sustenta esse resultado. Apesar das mudanças no mercado global e do aumento das correlações internacionais, a lógica proposta por Markowitz permanece válida, constituindo até hoje a base de modelos modernos de seleção de carteiras (MARKOWITZ, 1952; KASHEFI; MCKEE, 2001; DAVIS et al., 2013; MARTIN, 2020; NBER, 2014).

2.5 Benchmarks e Avaliação de Carteiras

A avaliação de desempenho de carteiras de investimento constitui uma etapa fundamental na análise empírica de estratégias financeiras, permitindo verificar se os retornos obtidos superam os padrões de mercado ou apenas os reproduzem. Para isso, torna-se necessário a definição de benchmarks adequados e de métricas de desempenho capazes de capturar a relação entre risco e retorno de forma comparável e objetiva (ELTON et al., 2014).

No contexto brasileiro, o CDI é amplamente utilizado como referência de taxa livre de risco, enquanto o Ibovespa representa o principal benchmark de mercado para carteiras de ações (B3, 2024). Além desses, índices como o IBrX, o SMLL e os índices setoriais da B3 ampliam as possibilidades de comparação, permitindo análises mais segmentadas

e coerentes com o perfil da carteira avaliada (B3, 2024; ANBIMA, 2023).

Por fim, a mensuração da eficiência das carteiras baseia-se em indicadores clássicos de desempenho ajustado ao risco, como os índices de Sharpe e Treynor, o Alpha de Jensen, o Sortino Ratio e o Information Ratio (SHARPE, 1966; TREYNOR, 1965; JENSEN, 1968). Esses parâmetros serão discutidos nos subtópicos seguintes, permitindo compreender os fundamentos teóricos e as aplicações práticas das métricas utilizadas para avaliar a performance das carteiras neste estudo.

2.5.1 CDI como Proxy de Taxa Livre de Risco

O Certificado de Depósito Interbancário (CDI) é amplamente empregado no mercado financeiro brasileiro como proxy da taxa livre de risco (R_f), constituindo o parâmetro padrão em pesquisas de desempenho de carteiras e fundos de investimento (CASACCIA et al., 2011; GOMES, 2011; OLIVEIRA, 2010). Essa escolha se justifica por sua natureza operacional, uma vez que o CDI reflete o custo médio das operações interbancárias de um dia, servindo como medida de remuneração mínima para aplicações financeiras de curtíssimo prazo e expressando o custo de oportunidade do capital no sistema bancário nacional (CASACCIA et al., 2011).

A adoção do CDI como ativo livre de risco está associada às suas características de liquidez e estabilidade, que o tornam adequado para representar um ativo sem incerteza de retorno. O ativo livre de risco é, por definição, aquele cujo desvio padrão dos retornos é nulo e cuja covariância com o portfólio de mercado é insignificante (CASACCIA et al., 2011). No contexto brasileiro, o CDI satisfaz essas condições de forma aproximada, apresentando baixa volatilidade e reduzida correlação com o mercado acionário (CASACCIA et al., 2011). Dessa forma, ele se consolida como o parâmetro mais apropriado para mensurar o retorno mínimo esperado pelos investidores, substituindo o papel que os Treasury Bills cumprem em economias desenvolvidas (OLIVEIRA, 2010).

Nas análises empíricas de desempenho, o CDI é essencial para o cálculo do excesso de retorno (excess return), definido como a diferença entre o retorno do portfólio e o retorno do ativo livre de risco ($R_p - R_f$). Esse componente é indispensável para os principais indicadores de desempenho ajustado ao risco, como os índices de Sharpe, Treynor e o Alpha de Jensen (CASACCIA et al., 2011; GOMES, 2011). No Índice de Sharpe, o CDI é utilizado como taxa livre de risco para padronizar o retorno da carteira pelo risco total medido pelo desvio-padrão. Já no Índice de Treynor, ele serve de base para a medição do excesso de retorno em relação ao risco sistemático (β). Por fim, no Alpha de Jensen, o

CDI representa a taxa livre de risco na equação derivada do modelo CAPM, que avalia o diferencial de desempenho em relação à linha do mercado de capitais (CASACCIA et al., 2011).

Em modelos de regressão baseados no Capital Asset Pricing Model (CAPM), o CDI é incorporado para calcular os retornos excedentes do portfólio e do mercado. A estrutura clássica dessas regressões pode ser expressa como:

$$R_{it} - R_{ft} = \alpha_i + \beta_i(R_{mt} - R_{ft}) + \varepsilon_{it},$$

Em que R_{ft} é o retorno do CDI, representando a taxa livre de risco do período. Assim, o CDI é utilizado tanto na variável dependente (excesso de retorno do portfólio) quanto na variável explicativa (excesso de retorno do mercado), permitindo a estimativa do alfa (α_i) e do beta (β_i), que medem, respectivamente, a habilidade do gestor e a exposição da carteira ao risco de mercado (OLIVEIRA, 2010; CASACCIA et al., 2011).

A aplicação do CDI também se estende a modelos de avaliação de market timing, como os de Treynor-Mazuy (1966) e Henriksson-Merton (1981). Nesses modelos, a taxa livre de risco é utilizada como limite comparativo para identificar se os gestores conseguem prever corretamente o comportamento do mercado. O modelo de Henriksson e Merton, por exemplo, introduz o termo $\gamma_i \max[0, R_{ft} - R_{mt}]$, no qual o CDI (R_{ft}) é o parâmetro que define os momentos de mercado em alta ou baixa. Dessa forma, quando o retorno do mercado (R_{mt}) é inferior ao retorno do CDI, o gestor deveria reduzir a exposição ao risco, ajustando o beta da carteira (CASACCIA et al., 2011). Esse tipo de análise reforça a importância empírica do CDI na decomposição do desempenho entre habilidades de seletividade e de market timing, especialmente em estudos que investigam a eficiência de gestores no Brasil (GOMES, 2011).

Embora o CDI não seja absolutamente livre de risco, por refletir as condições do sistema bancário e da política monetária, sua volatilidade mínima, liquidez diária e ampla aceitação no mercado o consolidam como o referencial empírico mais adequado à realidade brasileira. Como destacam Casaccia et al. (2011), Oliveira (2010) e Gomes (2011), o CDI é o padrão metodológico dominante na literatura nacional, servindo como base para mensurações de desempenho, estimativa de parâmetros de risco e validação de modelos de precificação de ativos em estudos aplicados ao mercado de capitais brasileiro.

2.5.2 Ibovespa como Benchmark de Mercado

O Índice Bovespa (Ibovespa) é reconhecido como o principal indicador do desempenho médio das ações negociadas no mercado brasileiro, refletindo a evolução dos preços e volumes transacionados das companhias de maior representatividade na B3 (OLIVEIRA; SOUSA, 2015). Sua metodologia de composição busca representar de forma abrangente o comportamento do mercado acionário, ao selecionar as ações mais líquidas e negociadas em pregão, ponderadas por seu volume financeiro e presença em mercado (REZENDE, 2013). Dessa forma, o Ibovespa sintetiza a performance das empresas que respondem por aproximadamente 80% do volume financeiro negociado à vista na bolsa, sendo recalculado e reponderado a cada quatro meses para assegurar sua atualidade e representatividade (REZENDE, 2013).

A estrutura do índice confere a cada ativo um peso proporcional à sua liquidez, o que o torna uma aproximação empírica do portfólio de mercado (M) na formulação do Capital Asset Pricing Model (CAPM). Essa correspondência metodológica justifica a utilização do Ibovespa como proxy do retorno de mercado em modelos de precificação de ativos aplicados ao Brasil (MACHADO, 2013). Assim, o índice serve como base para a estimativa do beta (β) das carteiras e fundos, sendo este calculado pela razão entre a covariância do retorno do ativo e o retorno do mercado e a variância dos retornos do próprio Ibovespa, ou seja:

$$\beta_j = \frac{\text{Cov}(R_j, R_m)}{\sigma_m^2}$$

Esse coeficiente expressa a sensibilidade do retorno de um ativo em relação às oscilações do mercado, sendo a variável central no CAPM clássico (MACHADO, 2013).

O papel do Ibovespa como benchmark não se limita à precificação de ativos, mas se estende à avaliação de desempenho de portfólios e fundos de investimento em ações. O Ibovespa é utilizado como índice de mercado (R_m) na aplicação dos principais indicadores de performance ajustada ao risco, como o Alfa de Jensen, o Índice de Treynor, o Information Ratio e o Índice de Modigliani (OLIVEIRA; SOUSA, 2015). Esses autores observam que o uso de benchmarks de mercado é uma prática consolidada na mensuração da habilidade dos gestores, sendo o Ibovespa o padrão mais adotado por refletir com maior fidedignidade o comportamento da carteira média do investidor institucional brasileiro (OLIVEIRA; SOUSA, 2015).

Contudo, a escolha inadequada de benchmark pode distorcer a análise de desempenho,

pois índices genéricos como o Ibovespa podem não reproduzir fielmente o perfil de risco de determinados fundos (OLIVEIRA; SOUSA, 2015). Apesar disso, por representar uma carteira diversificada de ativos de alta capitalização e liquidez, o índice é amplamente reconhecido como o parâmetro de comparação mais consistente para fundos de ações no Brasil. Esse uso é consistente com a tradição internacional de comparar o retorno dos fundos com índices amplos de mercado, como o S&P 500 nos Estados Unidos e o FTSE 100 no Reino Unido (MACHADO, 2013).

Empiricamente, o Ibovespa apresenta correlação positiva com os retornos médios das carteiras representativas do mercado brasileiro, o que reforça sua adequação como variável de referência nos modelos de precificação (REZENDE, 2013). As empresas de maior valor de mercado, como Petrobras, Vale, Itaú e Bradesco, estão consistentemente incluídas em sua composição, o que garante que os movimentos do índice expressem o comportamento agregado das companhias com maior influência sobre o mercado. Além disso, o índice tem sido utilizado em análises comparativas com outros indicadores, como o Índice de Sustentabilidade Empresarial (ISE), demonstrando que, em períodos como dezembro de 2005 a março de 2007, a diferença entre os retornos médios do ISE e do Ibovespa não foi estatisticamente significativa, evidenciando a capacidade do Ibovespa de representar o desempenho geral do mercado (REZENDE, 2013).

2.5.3 Outras Referências (IBrX, SMLL e Índices Setoriais)

Além do Ibovespa, outros índices da B3 podem ser utilizados como benchmarks complementares, oferecendo uma perspectiva mais ampla e menos concentrada da performance do mercado acionário brasileiro (ROGERS et al., 2005). Entre os mais relevantes estão o Índice Brasil 50 (IBrX 50), o Índice Brasil 100 (IBrX 100), o Índice Small Caps (SMLL) e os índices setoriais, como IFNC, IMAT e ICON, que permitem análises segmentadas por setor econômico (SIMPLÍCIO, 2024). A utilização combinada desses índices em análises de desempenho reduz o viés decorrente da concentração excessiva de determinados ativos, característica observada no Ibovespa, e melhora a comparabilidade entre carteiras de perfis distintos (VALENTE, 2020; ROQUETE, 2019).

2.5.3.1 IBrX 50 e IBrX 100

Os índices IBrX 50 e IBrX 100 são indicadores de retorno total que mensuram o desempenho médio das ações mais negociadas e representativas do mercado acionário brasileiro (B3, 2024). Ambos refletem a variação dos preços de 50 e 100 ativos, respecti-

vamente, ponderados pelo valor de mercado do free float, isto é, o conjunto de ações em livre circulação. Os índices incluem apenas ações e units listadas na B3 e excluem BDRs ou papéis de empresas em recuperação judicial ou sob intervenção (B3, 2024).

Para integrar o IBrX, os ativos devem atender simultaneamente a critérios de negociabilidade, frequência de presença em pregão (mínimo de 95%) e exclusão de penny stocks, o que assegura a representatividade dos papéis mais líquidos do mercado (B3, 2024). A metodologia dos índices IBrX busca corrigir as distorções observadas em benchmarks excessivamente concentrados, como o Ibovespa, oferecendo maior diversificação setorial e melhor distribuição de pesos entre empresas de diferentes tamanhos (ROGERS Et Al, 2005).

Comparações entre índices amplos como o IBrX 100, o Ibovespa e o IGC, as métricas de desempenho ajustado ao risco, como o índice de Sharpe, tendem a ser semelhantes, ainda que o IBrX 50 e o IBrX 100 apresentem volatilidade marginalmente inferior, o que reforça sua utilidade como benchmarks alternativos em estudos de carteiras (VALENTE, 2020). Além disso, o IBrX 100 tem sido amplamente utilizado em aplicações empíricas de indexação fundamental, servindo de base para comparar estratégias de ponderação de carteiras por valor de mercado ou por fundamentos contábeis (ROQUETE, 2019).

2.5.3.2 SMLL (Índice Small Caps)

O Índice Small Caps (SMML) foi criado pela B3 em 2005 com o objetivo de medir o desempenho médio de empresas de menor capitalização de mercado, que geralmente apresentam maior potencial de crescimento, mas também maior risco e volatilidade (CARLUCCI, 2021). Para compor o índice, são selecionadas ações que, em ordem decrescente de valor de mercado, estão fora dos 85% superiores da capitalização total das empresas listadas, o que garante a representatividade das companhias de menor porte (B3, 2024).

As empresas integrantes do SMML devem apresentar presença mínima em 95% dos pregões e não serem classificadas como penny stocks, além de atender aos critérios de liquidez e elegibilidade semelhantes aos aplicados aos índices IBrX (MAGALHÃES; TIMÓTEO, 2020). O índice é ponderado pelo valor de mercado em free float e revisado quadrimensalmente, assegurando que sua composição reflete o comportamento das small caps ao longo do tempo (B3, 2024).

O SMML apresenta volatilidade assimétrica, sendo mais sensível a choques negativos do que positivos, comportamento compatível com o maior risco intrínseco dessas empresas (CARLUCCI, 2021). Modelagens econômicas baseadas em GARCH e ARCH

confirmam que períodos de queda do SMLL são seguidos por aumentos significativos na variância condicional dos retornos, o que reforça o papel do índice como um indicador de risco e sensibilidade macroeconômica (CARLUCCI, 2021). Além disso, Magalhães-Timóteo (2020) observou que, em contextos de baixa taxa Selic, o SMLL tende a superar benchmarks tradicionais, refletindo a migração de investidores para ativos de maior risco e potencial de retorno.

2.5.3.3 Índices Setoriais (IFNC, IMAT, ICON e outros)

Os índices setoriais da B3, como IFNC (Financeiro), IMAT (Materiais Básicos) e ICON (Consumo), funcionam como benchmarks específicos para o desempenho de setores econômicos, permitindo análises comparativas mais detalhadas entre carteiras temáticas (SIMPLÍCIO, 2024). Cada índice é composto por ações de empresas que atuam predominantemente em determinado segmento, e suas ponderações seguem a mesma lógica dos índices amplos, baseada no valor de mercado ajustado pelo free float (B3, 2024).

A utilização de índices setoriais é especialmente relevante em períodos de forte variação macroeconômica, pois diferentes setores respondem de forma heterogênea às mudanças na taxa de juros e nas condições de crédito (SIMPLÍCIO, 2024). Os resultados empíricos obtidos pelo autor indicam que, em períodos de juros baixos, os índices IMAT, UTIL e INDX apresentaram retornos médios superiores, enquanto, em cenários de juros elevados, índices como IFNC e IEEEX se destacaram pelo desempenho positivo, dada sua correlação direta com a Selic.

A relação inversa entre taxa de juros e desempenho setorial é estatisticamente confirmada em parte das regressões realizadas, com significância de 5% para os índices IEEEX, IFNC, INDX e UTIL (SIMPLÍCIO, 2024). Esses resultados reforçam o potencial dos índices setoriais como instrumentos analíticos para gestores de portfólios com foco em alocação estratégica por setor, oferecendo subsídios para a construção de carteiras mais eficientes e alinhadas às condições macroeconômicas predominantes.

2.6 Métricas de Desempenho

As métricas de desempenho ajustadas ao risco são instrumentos fundamentais para a análise de carteiras, permitindo avaliar a eficiência de uma estratégia de investimento considerando simultaneamente retorno e volatilidade. No contexto contemporâneo, essas medidas continuam sendo amplamente utilizadas por gestores e analistas para mensurar

a performance de fundos de investimento e portfólios em relação à taxa livre de risco e a benchmarks de mercado (THOMYAMONGKOL, 2020; KUMAR, 2015).

Essas métricas, sendo elas o Sharpe Ratio, Treynor Ratio, Jensen's Alpha, Sortino Ratio e Information Ratio, possibilitam avaliar o retorno excedente de um portfólio ajustado por diferentes definições de risco. Cada uma delas adota uma abordagem distinta para a mensuração do desempenho, refletindo diferentes dimensões do comportamento dos retornos e da sensibilidade do portfólio ao mercado. De forma geral, quanto maior o valor desses índices, melhor o desempenho ajustado ao risco do fundo ou carteira (THOMYAMONGKOL, 2020).

2.6.1 Índice de Sharpe

O Índice de Sharpe foi proposto por William F. Sharpe como uma métrica de avaliação da eficiência de um portfólio em gerar retorno adicional por unidade de risco total assumido (SHARPE, 1966). Originalmente denominado Reward-to-Variability Ratio, o índice mede a compensação de risco em relação à volatilidade total da carteira. Ele é definido pela seguinte equação:

$$SHR = \frac{R_p - R_f}{\sigma_p} \quad (2.3)$$

Em que R_p é o retorno do portfólio, R_f é a taxa livre de risco e σ_p é o desvio-padrão dos retornos.

O índice de Sharpe interpreta-se como o retorno excedente por unidade de risco total, permitindo comparações diretas entre carteiras com diferentes níveis de volatilidade (SHARPE, 1966). O Sharpe Ratio apresentou desempenho superior na análise de fundos de ações e mistas, por capturar de forma adequada a relação entre retorno e risco total (THOMYAMONGKOL, 2020).

Apesar de sua ampla utilização, o índice de Sharpe apresenta limitações, pois penaliza igualmente volatilidades positivas e negativas. Esse aspecto torna-o menos eficaz em carteiras com retornos assimétricos, nas quais o investidor se preocupa mais com perdas do que com ganhos (KUMAR, 2015).

2.6.2 Beta

O Beta (β) é uma métrica de risco sistemático que representa a sensibilidade do retorno de um ativo ou portfólio em relação às variações do mercado (KUMAR, 2015). É calculado pela covariância entre o retorno da carteira (R_p) e o retorno do mercado (R_m), dividida pela variância do mercado:

$$\beta = \frac{\text{Cov}(R_p - R_f, R_m - R_f)}{\text{Var}(R_m - R_f)} \quad (2.4)$$

Um $\beta_p = 1$ indica que o ativo se move em linha com o mercado; $\beta_p > 1$ indica maior volatilidade (ativo agressivo); e $\beta_p < 1$ caracteriza ativos defensivos (SHARPE, 1966). O Beta é o componente central do CAPM e serve como base para métricas como o Treynor Ratio e o Alpha de Jensen. Já Thomyamongkol (2020) destaca que o uso do Beta permite compreender a relação de risco-retorno em ambientes de mercado diferentes, sendo essencial para avaliar a exposição das carteiras a oscilações sistêmicas (KUMAR, 2015). Assim, o Beta não apenas quantifica o risco de mercado, mas também fornece o parâmetro de referência para a avaliação de desempenho ajustado ao risco sistemático (KUMAR, 2015).

2.6.3 Alpha de Jensen

O Alpha de Jensen foi desenvolvido por Michael C. Jensen (1968) como um desdobramento do modelo CAPM, sendo utilizado para medir o desempenho de fundos de investimento em relação ao retorno esperado para seu nível de risco sistemático (KUMAR, 2015). Ele é calculado conforme a expressão:

$$\alpha_p = R_p - [R_f + \beta_p(R_m - R_f)] \quad (2.5)$$

Em que R_m é o retorno de mercado e β_p é o coeficiente de sensibilidade ao risco sistemático.

O α_p representa o retorno excedente gerado pelo gestor após ajustar o portfólio pelo risco de mercado. Um α positivo indica que o fundo superou o desempenho previsto pelo CAPM, enquanto um α negativo revela subdesempenho (SHARPE, 1966; THOMYAMONGKOL, 2020).

O Jensen's Alpha apresenta forte capacidade discriminatória em setores de menor vo-

latilidade, como fundos de títulos e de renda fixa, evidenciando sua eficiência em contextos onde o retorno está mais correlacionado com o risco sistemático do mercado (THOMYAMONGKOL, 2020).

2.6.4 Information Ratio

O Information Ratio mede o excesso de retorno da carteira em relação ao seu benchmark, ajustado pela volatilidade do tracking error, ou seja, a variação entre o desempenho do portfólio e o índice de referência (KUMAR, 2015). A fórmula é expressa por:

$$IR = \frac{R_p - R_b}{\sigma(R_p - R_b)} \quad (2.6)$$

Em que R_b é o retorno do benchmark. O Information Ratio indica a consistência do gestor na geração de alfa em relação ao índice de mercado, sendo amplamente utilizado para avaliar estratégias de gestão ativa (KUMAR, 2015). Essa métrica é particularmente eficaz em fundos de ações com gestão ativa, nos quais a consistência do retorno é um fator crítico de performance (THOMYAMONGKOL, 2020).

2.7 Estatística Aplicada à Análise de Carteiras

A estatística desempenha um papel central na moderna teoria financeira, servindo como base quantitativa para a mensuração, comparação e inferência de resultados obtidos em estudos de investimento (MONTGOMERY, 2017). Desde a formulação da Teoria Moderna do Portfólio por Markowitz (1952), a análise estatística consolidou-se como instrumento para a avaliação da relação entre risco e retorno, a validação de hipóteses de eficiência de mercado e a comparação de desempenhos entre carteiras. No contexto deste trabalho, a estatística constitui o elo entre o delineamento experimental e a interpretação empírica dos resultados, permitindo avaliar se as diferenças observadas entre carteiras geradas por inteligências artificiais com e sem alimentação de dados estruturados são estatisticamente significativas e não decorrentes do acaso.

Em finanças, as técnicas estatísticas são aplicadas tanto na etapa descritiva, que envolve a síntese de indicadores como média, variância, desvio-padrão e correlação, quanto na etapa inferencial, voltada à testagem de hipóteses e comparação de grupos amostrais (MONTGOMERY; RUNGER, 2014). Essas ferramentas possibilitam transformar observações empíricas em conclusões generalizáveis, fundamentando a análise de desempe-

nho e eficiência de diferentes estratégias de investimento. Ao longo das últimas décadas, o avanço da econometria e da estatística aplicada ampliou a capacidade de mensurar e interpretar fenômenos complexos do mercado de capitais, como volatilidade, covariância entre ativos e efeitos de interação entre variáveis explicativas (HAIR et al., 2019).

No âmbito deste estudo, a estatística é utilizada como instrumento de verificação empírica das hipóteses centrais do experimento, particularmente no que se refere ao impacto do uso de dados estruturados sobre a performance das carteiras geradas por modelos de linguagem de larga escala (LLMs). Entre as diversas técnicas disponíveis, destaca-se o uso da Análise de Variância (ANOVA), que permite comparar médias entre grupos e identificar diferenças significativas no desempenho das carteiras.

2.7.1 Conceitos Fundamentais de Estatística em Finanças

A estatística constitui um dos pilares da análise financeira moderna, sendo responsável por fornecer os instrumentos necessários para a mensuração, a interpretação e a previsão de fenômenos econômicos e de mercado. De modo geral, ela permite que o pesquisador transforme dados empíricos em inferências objetivas e quantificáveis, viabilizando a tomada de decisão em ambientes de incerteza (MONTGOMERY; RUNGER, 2014). A estatística é definida como a ciência que trata da coleta, organização, análise e interpretação de dados, sendo aplicável em qualquer área na qual existam variações e incertezas observáveis (MONTGOMERY; RUNGER, 2014). Em finanças, essas variações estão presentes nos preços de ativos, nas taxas de retorno e nos indicadores de risco, o que torna os métodos estatísticos indispensáveis à formulação e validação de modelos quantitativos de investimento.

A estatística divide-se tradicionalmente em dois grandes domínios: estatística descritiva e estatística inferencial. A estatística descritiva tem como objetivo organizar e resumir os dados observados, permitindo compreender o comportamento de uma variável por meio de representações gráficas, tabelas ou medidas numéricas de síntese (COOKSEY, 2020). Essa abordagem é amplamente utilizada em finanças para descrever séries temporais de retornos, analisar a volatilidade de ativos e visualizar tendências de mercado. As principais medidas descritivas são a média, a mediana, a variância e o desvio-padrão, que possibilitam quantificar a tendência central e a dispersão dos dados (MONTGOMERY; RUNGER, 2014). A representação visual por histogramas e gráficos de dispersão também facilita a identificação de padrões e anomalias, como caudas longas ou assimetrias em distribuições de retornos (COOKSEY, 2020).

A estatística inferencial permite extrapolar conclusões obtidas de uma amostra para toda uma população, baseando-se na teoria da probabilidade e reconhecendo que as observações empíricas estão sujeitas a variações aleatórias (MONTGOMERY; RUNGER, 2014). Em estudos financeiros, essa abordagem é utilizada para testar hipóteses, por exemplo, se dois portfólios possuem retornos médios significativamente diferente, e para construir intervalos de confiança em torno de estimativas, quantificando o grau de incerteza associado aos resultados. O conceito de probabilidade fornece a estrutura matemática que permite mensurar a chance de que os resultados observados representem padrões reais e não apenas flutuações ocasionais (COOKSEY, 2020).

A aplicação desses conceitos torna-se evidente ao analisar o risco e o retorno dos ativos financeiros. A relação risco-retorno constitui o eixo central da tomada de decisão em finanças, pois investidores avessos ao risco exigem uma compensação adicional, denominada prêmio de risco, para deter ativos mais voláteis (SCHOENMAKER; SCHRAMADE, 2023). O retorno esperado $E[r]$ é calculado como uma média ponderada dos retornos possíveis, em que as probabilidades associadas a cada cenário servem como pesos:

$$E[r] = \sum_{i=1}^n p_i \cdot r_i \quad (2.7)$$

Essa métrica sintetiza o ganho médio esperado em um investimento e serve de referência para a avaliação de carteiras e modelos de precificação (RISK-RETURN ANALYSIS, 2023).

O risco é tradicionalmente mensurado por meio da variância e do desvio-padrão, medidas que expressam a dispersão dos retornos em torno da média. A variância representa o desvio quadrático médio, enquanto o desvio-padrão, sua raiz quadrada, fornece uma medida mais intuitiva de variabilidade, mantendo as mesmas unidades do dado original (MONTGOMERY; RUNGER, 2014). Em finanças, o desvio-padrão é conhecido como volatilidade e constitui uma das métricas mais utilizadas para avaliar o risco total de um ativo ou portfólio. Uma volatilidade elevada reflete maior incerteza sobre o retorno futuro, demandando prêmios de risco mais altos (SCHOENMAKER; SCHRAMADE, 2023).

Além disso, a variância σ_i^2 representa o risco total associado aos retornos de um ativo individual, constituindo a medida fundamental de incerteza na Teoria Moderna do Portfólio (MARKOWITZ, 1952). O desvio-padrão σ_i , definido como a raiz quadrada da variância, expressa a volatilidade dos retornos e é amplamente utilizado como medida prática de risco.

A partir dessas medidas individuais, define-se a covariância entre dois ativos $\text{Cov}(R_1, R_2)$, que quantifica como os retornos variam conjuntamente. A covariância é dada por:

$$\text{Cov}(R_1, R_2) = \sigma_{12}$$

A correlação, por sua vez, normaliza a covariância pelo produto dos desvios-padrão de cada ativo, resultando no coeficiente:

$$\rho_{12} = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$$

Esse coeficiente varia entre -1 e 1 e expressa a intensidade e a direção da relação linear entre os retornos dos ativos. A correlação é central para o princípio da diversificação, pois combina tanto o risco individual quanto o risco compartilhado entre ativos, permitindo a formação de carteiras mais eficientes (RISK-RETURN ANALYSIS, 2023).

Em termos práticos, as medidas estatísticas mais utilizadas na análise financeira incluem: média, mediana, variância, desvio-padrão, coeficiente de variação e correlação (MONTGOMERY; RUNGER, 2014). A média representa o retorno médio histórico ou esperado; a mediana é particularmente útil em distribuições assimétricas, pois reduz o impacto de valores extremos; a variância e o desvio-padrão quantificam a volatilidade dos retornos; o coeficiente de variação expressa o risco relativo de ativos com diferentes magnitudes de retorno; e a correlação de Pearson permite avaliar o grau de associação linear entre séries financeiras (COOKSEY, 2020, p. 72).

Evidências empíricas demonstram o uso dessas medidas em estudos de longo prazo. Entre 1870 e 2015, ações renderam em média 10,7% ao ano, com desvio-padrão de 22,6%, enquanto títulos governamentais tiveram retorno médio de 6,1% e volatilidade de 8,9%, confirmado a relação positiva entre risco e retorno (SCHOENMAKER; SCHRAMADE, 2023). Essa relação é a base do Capital Asset Pricing Model (CAPM), no qual o beta (β_i) quantifica a sensibilidade de um ativo em relação ao portfólio de mercado, medindo o componente sistemático de risco não diversificável (RISK-RETURN ANALYSIS, 2023).

Em síntese, a estatística fornece o arcabouço matemático e conceitual necessário para descrever, analisar e inferir o comportamento dos ativos financeiros. As medidas de tendência central e dispersão permitem compreender a rentabilidade e o risco; a correlação e a covariância fundamentam a teoria da diversificação; e os testes de inferência garantem rigor científico às comparações de desempenho entre carteiras. Assim, a estatística atua como elo entre o delineamento experimental e a interpretação empírica dos resultados

neste estudo, sendo indispensável à análise comparativa entre as carteiras geradas por modelos de linguagem artificial com e sem alimentação de dados estruturados.

2.7.2 ANOVA de Dois Fatores

A Análise de Variância de dois fatores (Two-Way ANOVA) é uma extensão da ANOVA tradicional que permite avaliar, de forma simultânea, o efeito de duas variáveis independentes (fatores) sobre uma variável dependente, além da interação entre elas (WILLIAM, 2004). Essa metodologia é particularmente relevante em estudos financeiros que buscam identificar como variáveis de natureza distinta, como o sentimento de mercado e o tamanho das empresas, influenciam o comportamento dos retornos (YOSHINAGA; CASTRO JR., 2012). A Two-Way ANOVA amplia a capacidade explicativa dos modelos ao decompor a variabilidade total dos dados em parcelas atribuíveis aos efeitos principais de cada fator e à sua interação combinada (WILLIAM, 2004).

O modelo estatístico geral é representado pela equação:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \quad (2.8)$$

Onde:

- Y_{ijk} representa a k -ésima observação associada ao nível i do fator A e ao nível j do fator B;
- μ é a média geral de todas as observações;
- α_i e β_j são os efeitos principais dos fatores A e B, respectivamente;
- $(\alpha\beta)_{ij}$ é o efeito de interação entre os dois fatores;
- ε_{ijk} é o erro aleatório, assumido como normalmente distribuído, com média zero e variância constante σ^2 (WILLIAM, 2004; YOSHINAGA; CASTRO JR., 2012).

A decomposição da variância total é expressa por:

$$SS_{Total} = SS_A + SS_B + SS_{AB} + SS_{Error} \quad (2.9)$$

Onde SS_A e SS_B são as somas de quadrados dos efeitos principais dos fatores A e B, SS_{AB} é a soma de quadrados da interação entre os fatores, e SS_{Error} representa a

variabilidade residual (WILLIAM, 2004). Essa estrutura permite analisar a contribuição individual de cada fator e a interação combinada sobre a variância total dos resultados (YOSHINAGA; CASTRO JR., 2012).

Decomposição da Variância

Tabela 1: Decomposição da soma dos quadrados na ANOVA de dois fatores

Componente	Soma de Quadrados (SS)	Graus de Liberdade (df)
Fator A	$SS_A = \sum_{i=1}^a bn(\bar{Y}_{i..} - \bar{Y}_{...})^2$	$a - 1$
Fator B	$SS_B = \sum_{j=1}^b an(\bar{Y}_{.j.} - \bar{Y}_{...})^2$	$b - 1$
Interação A×B	$SS_{AB} = \sum_{i=1}^a \sum_{j=1}^b n(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$	$(a - 1)(b - 1)$
Erro	$SS_{Error} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (Y_{ijk} - \bar{Y}_{ij.})^2$	$ab(n - 1)$

Cada componente da soma dos quadrados está associado a um número específico de graus de liberdade. A partir dessas quantidades, são calculados os Quadrados Médios (MS) e as estatísticas F , que avaliam a significância estatística dos efeitos (WILLIAM, 2004; YOSHINAGA; CASTRO JR., 2012).

O Cálculo dos Quadrados Médios e Estatísticas F pode ser sintetizado da seguinte forma:

Tabela 2: Equações principais da Two-Way ANOVA

Parâmetro	Cálculo do Quadrado Médio (MS)	Estatística F
Fator A	$MS_A = \frac{SS_A}{a - 1}$	$F_A = \frac{MS_A}{MS_{Error}}$
Fator B	$MS_B = \frac{SS_B}{b - 1}$	$F_B = \frac{MS_B}{MS_{Error}}$
Interação A×B	$MS_{AB} = \frac{SS_{AB}}{(a - 1)(b - 1)}$	$F_{AB} = \frac{MS_{AB}}{MS_{Error}}$
Erro	$MS_{Error} = \frac{SS_{Error}}{ab(n - 1)}$	—

As estatísticas F_A , F_B e F_{AB} seguem distribuições F sob a hipótese nula de igualdade das médias (H_0), e são utilizadas para testar os efeitos principais e de interação. Quando

$F_{calc} > F_{crit}$, rejeita-se a hipótese nula, indicando diferenças estatisticamente significativas entre os níveis dos fatores (WILLIAM, 2004).

Estrutura de Resultados da ANOVA de Dois Fatores

Tabela 3: Tabela típica de resultados da ANOVA de dois fatores

Fonte de Variação	SS	df	MS	F / p-value
Fator A	SS_A	$a - 1$	MS_A	$F_A = MS_A/MS_{Error}$
Fator B	SS_B	$b - 1$	MS_B	$F_B = MS_B/MS_{Error}$
Interação A×B	SS_{AB}	$(a - 1)(b - 1)$	MS_{AB}	$F_{AB} = MS_{AB}/MS_{Error}$
Erro	SS_{Error}	$ab(n - 1)$	MS_{Error}	—
Total	SS_{Total}	$abn - 1$	—	—

Em aplicações financeiras, a Two-Way ANOVA é uma ferramenta utilizada para explorar como múltiplos fatores influenciam retornos. No estudo de YOSHINAGA e CASTRO JR. (2012), o modelo foi empregado para analisar o efeito combinado do sentimento de mercado e das características das empresas sobre as taxas de retorno trimestrais. As empresas foram agrupadas em quintis de valor de mercado, risco e idade, e seus retornos médios foram avaliados sob diferentes condições de sentimento (positivo e negativo). O modelo revelou uma interação estatisticamente significativa entre o sentimento de mercado e o tamanho das empresas, demonstrando que o impacto do sentimento é condicionado pelas características corporativas (YOSHINAGA; CASTRO JR., 2012).

Após períodos de sentimento positivo, empresas de menor valor de mercado apresentaram retornos mais baixos, enquanto as de maior valor mantiveram desempenho superior, evidência de que o comportamento do mercado não é homogêneo, mas dependente de fatores interativos. Essa relação cruzada entre variáveis de natureza econômica e psicológica exemplifica o poder da ANOVA de dois fatores para detectar efeitos combinados (WILLIAM, 2004; YOSHINAGA; CASTRO JR., 2012).

A formulação matricial da Two-Way ANOVA pode ser expressa como:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.10)$$

Em que \mathbf{Y} é o vetor das observações, \mathbf{X} é a matriz de delineamento que representa os níveis dos fatores e suas interações, $\boldsymbol{\beta}$ é o vetor de parâmetros estimados, e $\boldsymbol{\varepsilon}$ é o

vetor dos erros aleatórios. A estimação é realizada por Mínimos Quadrados Ordinários (OLS), minimizando $S = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$, sob os pressupostos de normalidade, homocedasticidade e independência dos resíduos (WILLIAM, 2004).

Em síntese, a ANOVA de dois fatores é uma técnica estatística robusta que permite identificar não apenas os efeitos principais de variáveis explicativas, mas também as interações entre elas. No contexto de finanças, sua aplicação possibilita avaliar simultaneamente fatores de natureza econômica e comportamental, oferecendo um arcabouço teórico e empírico para estudos sobre previsibilidade de retornos, formação de carteiras e comportamento do investidor (YOSHINAGA; CASTRO JR., 2012; WILLIAM, 2004).

2.8 Inteligência Artificial e Finanças

A Inteligência Artificial (IA) tem se consolidado como uma das tecnologias mais transformadoras do setor financeiro, redefinindo a forma como dados são processados, decisões são tomadas e estratégias de investimento são formuladas (HE et al., 2023). Desde os primeiros sistemas automatizados de negociação, baseados em regras fixas, até as redes neurais profundas e os modelos de linguagem de larga escala (Large Language Models – LLMs), a IA percorreu uma trajetória de contínua sofisticação e integração com as finanças quantitativas (HEATON; POLSON; WITTE, 2017). Essa evolução é impulsionada pela combinação entre o aumento exponencial do poder computacional, a disponibilidade de big data e o desenvolvimento de novas arquiteturas de aprendizado, como Deep Learning e Ensemble Models (HE et al., 2023).

A utilização de algoritmos inteligentes em finanças ampliou o alcance analítico de investidores e instituições, permitindo a modelagem de fenômenos complexos como a dinâmica de preços, o risco sistêmico e o comportamento de mercado (CHEN et al., 2023). O aprendizado de máquina e suas derivações possibilitam a identificação de padrões não lineares, a previsão de séries temporais e a extração de informações relevantes de bases massivas de dados estruturados e não estruturados (CHEN et al., 2023).

Sendo assim, a integração entre IA e finanças representa não apenas um avanço tecnológico, mas uma reconfiguração epistemológica da análise financeira. A transição da modelagem estatística tradicional para o aprendizado hierárquico e adaptativo redefine a natureza do investimento contemporâneo, consolidando a IA como eixo fundamental da engenharia financeira moderna (HEATON; POLSON; WITTE, 2017; HE et al., 2023; CHEN et al., 2023).

2.8.1 Modelos de Linguagem de Larga Escala (LLMs)

Os Modelos de Linguagem de Larga Escala (LLMs) representam um dos avanços mais disruptivos da Inteligência Artificial (IA) aplicada ao setor financeiro, marcando a transição de modelos estatísticos tradicionais para sistemas capazes de compreender e gerar linguagem natural em contextos complexos (LIU et al., 2024). Fundamentados na arquitetura Transformer, introduzida em 2017, os LLMs são treinados em extensos conjuntos de dados textuais, permitindo-lhes capturar relações semânticas de longo alcance e compreender nuances linguísticas. Essa arquitetura é caracterizada pelo mecanismo de self-attention, que possibilita a ponderação contextual de cada token em relação a todos os outros, resultando em um aprendizado hierárquico e contextualizado da linguagem (KONG, 2024).

Os LLMs têm se mostrado ferramentas excepcionais na análise de textos financeiros, com capacidade para processar volumes massivos de dados não estruturados, como relatórios corporativos, comunicados de imprensa e notícias de mercado (RAM-AI, 2024). Essa habilidade é crucial em um cenário em que a informação textual influencia fortemente os preços de ativos e as decisões de investimento. A aplicação de LLMs a essas tarefas permite a extração automatizada de informações relevantes e a geração de insights sobre o sentimento do mercado.

Entre as tarefas linguísticas em que os LLMs se destacam estão: (i) a sumarização e extração de dados, em que os modelos sintetizam relatórios financeiros extensos em resumos estruturados, reduzindo o tempo de análise de investidores e analistas (KONG, 2024); (ii) a análise de sentimento, que quantifica o tom positivo, negativo ou neutro de textos financeiros, capturando percepções implícitas que influenciam o comportamento de mercado (RAM-AI, 2024); e (iii) o raciocínio contextual, pelo qual os modelos inferem relações entre eventos financeiros, narrativas macroeconômicas e reações de mercado (LIU et al., 2024). Essas aplicações redefinem o papel dos analistas, pois os LLMs conseguem integrar informações de fontes heterogêneas, identificar padrões de correlação semântica e oferecer percepções interpretativas que seriam inviáveis por meio de modelos lineares ou algoritmos baseados em regras (RAM-AI, 2024).

Estudos recentes destacam o potencial dos LLMs em compreender e prever fenômenos financeiros a partir da interpretação textual. Modelos como o FinBERT e o FinDeBERTa foram aplicados à análise de notícias financeiras e demonstraram desempenho superior a métodos tradicionais baseados em dicionários, apresentando correlação positiva entre o sentimento agregado e o retorno futuro das ações (RAM-AI, 2024). Pesquisas subsequentes

ampliaram essa abordagem, testando a eficácia de modelos como o GPT-4 na previsão de retornos com base em manchetes e fluxos de notícias (LIU et al., 2024). Os resultados indicam que o GPT-4 supera significativamente versões anteriores e modelos clássicos de aprendizado de máquina, especialmente em contextos de alta volatilidade e em ações de menor capitalização. Observa-se, ainda, que o impacto de notícias negativas é capturado de forma mais precisa por LLMs de última geração, refletindo uma compreensão semântica mais refinada do contexto (KONG, 2024).

Além da previsão de preços, os LLMs vêm sendo aplicados à análise de risco e à estimativa de volatilidade. Liu et al. (2024) destacam que modelos baseados em transformers conseguem extrair variáveis latentes de sentimento e incorporar essas informações em modelos de covariância para otimização de portfólios. Esses resultados corroboram o potencial dos LLMs em unir dados textuais e numéricos, criando representações integradas que elevam o desempenho preditivo de modelos financeiros. Essa sinergia entre o texto e os retornos quantitativos representa uma das fronteiras mais promissoras da pesquisa em finanças comportamentais computacionais (RAM-AI, 2024; LIU et al., 2024).

O papel emergente dos LLMs como “co-pilotos” de analistas humanos destaca o potencial colaborativo entre IA generativa e julgamento humano. Segundo Liu et al. (2024), os modelos de linguagem não substituem os analistas, mas ampliam suas capacidades cognitivas ao fornecer resumos, simulações de cenários e interpretações rápidas de dados complexos. Esse paradigma de interação humano-IA reforça a ideia de sistemas híbridos, nos quais o analista atua como supervisor e interpretador dos resultados gerados pelo modelo.

Entre as principais contribuições dessa integração, destacam-se: (i) a análise qualitativa aprimorada, em que LLMs como o GPT-4 conseguem identificar nuances discursivas em relatórios e notícias, oferecendo interpretações semânticas que apoiam decisões estratégicas (RAM-AI, 2024); (ii) a integração entre análise qualitativa e quantitativa, na qual modelos combinam dados textuais com métricas de mercado, construindo pontuações compostas de risco e retorno (KONG, 2024); e (iii) a interpretação transparente, em que modelos explicam a lógica subjacente às suas previsões, favorecendo a auditoria e a rastreabilidade das decisões (LIU et al., 2024). Esses avanços impulsionam o surgimento dos chamados Human-AI Interactive Systems, que unem as capacidades analíticas das máquinas à intuição humana. Exemplos incluem plataformas experimentais como o Alpha-GPT, que permite a interação direta entre analistas e modelos, traduzindo ideias de trading em estratégias automatizadas de investimento (LIU et al., 2024).

A incorporação de LLMs em finanças inaugura uma nova fase de automação cognitiva, na qual a análise textual e o raciocínio quantitativo convergem em um mesmo sistema inteligente. Os estudos analisados demonstram que esses modelos não apenas interpretam informações de forma contextual, mas também aprendem padrões de comportamento de mercado a partir de dados linguísticos, revelando correlações latentes entre sentimento, preço e risco (RAM-AI, 2024; KONG, 2024). Essa capacidade de generalização faz dos LLMs ferramentas valiosas para a geração de alpha, otimização de portfólios e mitigação de riscos comportamentais. A evolução contínua das arquiteturas, aliada à ampliação de conjuntos de dados multimodais e ao avanço dos LLMs voltados para o domínio financeiro, tende a consolidar o papel desses modelos como núcleo da próxima geração de finanças baseadas em IA. No contexto da engenharia financeira contemporânea, os LLMs representam a convergência entre linguagem, decisão e aprendizado, um ponto de inflexão que redefine a relação entre análise humana e inteligência artificial (LIU et al., 2024; RAM-AI, 2024; KONG, 2024).

2.8.2 Construção de Prompts e Engenharia de Prompting na IA

Os modelos de linguagem de larga escala (LLMs) dependem diretamente da forma como são instruídos para gerar suas respostas. Nesse contexto, o conceito de prompt, isto é, o texto ou comando fornecido ao modelo como entrada, assume papel central no desempenho e na coerência das respostas geradas. Um prompt funciona como o ponto de partida da inferência linguística, definindo o contexto, o tom e os limites da geração textual (BROWN et al., 2020). Em termos práticos, é a instrução que direciona o raciocínio do modelo e influencia diretamente sua capacidade de fornecer respostas precisas, consistentes e alinhadas à intenção do usuário.

A engenharia de prompts (prompt engineering) emerge como um campo especializado da interação com modelos de linguagem, voltado à formulação sistemática de instruções que otimizam o desempenho dos LLMs. Enquanto o prompting simples envolve apenas a redação de uma pergunta ou instrução direta, a engenharia de prompts requer o uso de técnicas estruturadas para induzir o modelo a raciocinar de forma ordenada e contextualizada (LIU et al., 2023). Trata-se, portanto, de uma abordagem metodológica que combina linguística, lógica e análise de comportamento de modelos para guiar o fluxo de raciocínio e mitigar ambiguidades.

Segundo Brown et al. (2020), os LLMs como o GPT-3 e suas versões subsequentes foram projetados para compreender instruções em linguagem natural, ajustando suas res-

postas com base nos exemplos e padrões apresentados. Essa característica permitiu o surgimento de estratégias avançadas de prompting, que transformaram a interação com modelos em um processo iterativo e refinado, aproximando-se de uma engenharia cognitiva. Liu et al. (2023) reforçam que a engenharia de prompts representa um novo paradigma de controle de modelos de linguagem, substituindo o ajuste de pesos (tradicional no machine learning) pelo ajuste semântico da entrada textual.

No campo financeiro, essa capacidade é particularmente relevante, pois a formulação de prompts bem definidos pode alterar substancialmente o comportamento analítico do modelo, conduzindo-o a interpretações mais coerentes de relatórios corporativos, demonstrações financeiras e tendências de mercado. O domínio da engenharia de prompts torna-se, portanto, uma competência essencial na aplicação de LLMs para análise quantitativa e qualitativa, especialmente em tarefas que exigem precisão semântica e inferência contextual (LIU et al., 2023).

As estratégias de prompting podem ser classificadas em quatro grandes categorias, cada uma com diferentes níveis de complexidade e aplicação:

- Zero-shot prompting: nesta abordagem, o modelo é solicitado a realizar uma tarefa sem exemplos prévios. O LLM deve inferir diretamente o padrão de resposta a partir do contexto da instrução. Essa técnica avalia a capacidade de generalização do modelo e é útil em análises financeiras exploratórias, quando se busca respostas rápidas sem fornecer dados de referência (BROWN et al., 2020).
- Few-shot prompting: consiste em incluir no prompt alguns exemplos de entrada e saída para demonstrar ao modelo o formato desejado de resposta. Essa estratégia melhora significativamente a consistência e reduz a variabilidade das respostas, sendo amplamente usada em tarefas de classificação de sentimento financeiro e summarização de relatórios (BROWN et al., 2020; LIU et al., 2023).
- Chain-of-thought prompting: introduzido por Wang et al. (2022), esse método incentiva o modelo a explicitar seu raciocínio passo a passo, antes de fornecer a resposta final. Essa forma de prompting melhora a capacidade de raciocínio lógico e matemático do modelo, tornando-o mais transparente e confiável. Em aplicações financeiras, o chain-of-thought pode ser utilizado para decompor etapas de valuation, explicar decisões de alocação de portfólio ou justificar previsões de retorno de ativos (WANG et al., 2022).
- Role prompting: esta técnica designa ao modelo um papel específico, como “analista

de equity research”, “gestor de portfólio” ou “especialista em risco de crédito”. Essa contextualização orienta o estilo de escrita, o grau de formalidade e o nível técnico das respostas, aproximando a produção textual das práticas profissionais do setor financeiro (LIU et al., 2023). Por exemplo, ao instruir o modelo como um “analista de mercado”, ele tende a estruturar respostas com argumentos, premissas e linguagem técnica compatíveis com relatórios de investimento.

No contexto das aplicações financeiras, a precisão do prompt está diretamente ligada à coerência e à qualidade das respostas. Estudos recentes indicam que instruções bem estruturadas reduzem inconsistências, aumentam a estabilidade do raciocínio e permitem que modelos como o GPT-4 ou o DeepSeek atinjam níveis mais elevados de precisão na interpretação de dados de mercado (LIU et al., 2023). Assim, a engenharia de prompting torna-se uma forma de controle semântico sobre os LLMs, permitindo que o usuário atue como projetista da interação cognitiva.

Além disso, a engenharia de prompts oferece um novo caminho para a integração entre análise qualitativa e quantitativa em finanças. Por meio de instruções específicas, o modelo pode ser direcionado a realizar sínteses de relatórios, interpretar tendências econômicas e correlacionar variáveis de desempenho com informações textuais. Em experimentos recentes, prompts cuidadosamente elaborados foram capazes de induzir LLMs a interpretar dados financeiros com maior consistência, reduzindo vieses e maximizando a relevância das respostas (BROWN et al., 2020; WANG et al., 2022).

3 MATERIAIS E MÉTODOS

O presente capítulo descreve de forma estruturada todo o arcabouço metodológico utilizado para desenvolver, organizar e analisar o experimento central desta pesquisa. Inicialmente, é apresentado o delineamento adotado, explicitando como o estudo foi planejado para comparar carteiras formadas por diferentes modelos de linguagem, com e sem alimentação adicional de dados financeiros. Em seguida, detalham-se os procedimentos de coleta, limpeza e preparação das bases de dados, bem como os critérios aplicados na construção das oito carteiras analisadas.

O capítulo também descreve os métodos utilizados para mensuração de desempenho, contemplando o cálculo dos retornos, métricas financeiras derivadas e indicadores comparativos. Por fim, são apresentados os procedimentos estatísticos empregados para testar diferenças entre grupos, com destaque para a ANOVA de dois fatores, e as ferramentas utilizadas para garantir precisão, replicabilidade e rastreabilidade dos resultados obtidos ao longo do estudo.

3.1 Delineamento da Pesquisa

O presente estudo adota um delineamento metodológico de natureza quantitativa, experimental e comparativa. A abordagem quantitativa se justifica pelo foco na mensuração de retornos financeiros, volatilidade e demais indicadores estatísticos de desempenho das carteiras de investimento, permitindo análises objetivas e replicáveis. Diferentemente de métodos qualitativos, voltados à interpretação subjetiva de percepções ou discursos, a ênfase deste trabalho recai sobre a mensuração numérica de resultados, em consonância com a natureza financeira do problema investigado.

No que tange ao caráter experimental, este se manifesta na criação e avaliação de diferentes carteiras de investimento geradas a partir de modelos de linguagem de larga escala (LLMs). O experimento consiste em estruturar dois grupos principais de cenários: carteiras formuladas apenas com instruções gerais, sem suporte informacional adicional, e

carteiras elaboradas mediante a incorporação de um alimentação de dados de informações estruturadas. Esse alimentação de dados é composto por planilhas contendo indicadores fundamentais das companhias listadas, extraídos diretamente do terminal Bloomberg, o que assegura confiabilidade e padronização dos dados.

Os dados utilizados abrangem tanto métricas anuais quanto mensais. No nível anual, incluem-se variáveis como o valor de mercado (market cap, em milhões de reais), o lucro líquido consolidado ao final de cada ano (net income, em milhões de reais), os dividendos pagos e o dividend yield, todos coletados de forma consistente desde 2015. Adicionalmente, uma planilha independente fornece especificamente o lucro líquido anual (net income) das companhias ao final de cada exercício, também extraído da Bloomberg, garantindo maior granularidade e precisão na mensuração dessa variável. Em complemento, foram considerados dados mensais de estrutura de capital, notadamente a razão entre o total investido somado ao caixa (em milhares de reais) e o patrimônio líquido (em milhares de reais), igualmente disponíveis desde 2015.

Esse conjunto de informações, ao ser incorporado no processo de geração de carteiras, representa um acréscimo substancial de fundamentação analítica, permitindo às inteligências artificiais basearem suas recomendações em variáveis que refletem tanto a performance histórico-financeira das empresas quanto aspectos de retorno ao acionista. Dessa forma, cria-se uma distinção clara entre os dois grupos experimentais: de um lado, carteiras orientadas apenas por comandos gerais aos LLMs; de outro, carteiras que se apoiam em dados quantitativos consistentes e padronizados, o que possibilita investigar em que medida a introdução desse alimentação de dados de informações financeiras influencia a qualidade e a robustez das decisões de investimento sugeridas.

O delineamento também possui caráter comparativo, uma vez que os resultados obtidos pelas carteiras geradas serão confrontados com os benchmarks de referência do mercado brasileiro, a saber: o Certificado de Depósito Interbancário (CDI), considerado a taxa livre de risco local, e o Índice Bovespa (Ibovespa), que sintetiza o desempenho médio das ações mais líquidas negociadas na B3. Essa estratégia metodológica viabiliza avaliar não apenas o retorno absoluto das carteiras, mas também sua capacidade de gerar valor em relação a parâmetros consagrados pela literatura e pela prática do mercado de capitais.

3.2 Universo e Amostra

Este estudo toma como universo de referência o conjunto de companhias com ações negociadas na B3 – Brasil, Bolsa, Balcão. Para garantir investibilidade, liquidez mínima e comparabilidade entre alternativas, a seleção efetiva de ativos é realizada a partir da carteira teórica do índice Ibovespa vigente nas datas de formação das carteiras experimentais. Essa escolha estabelece um marco amostral coerente com a prática de avaliação de desempenho no mercado brasileiro, reduz o risco de inclusão de papéis com baixa negociabilidade e mitiga ruídos operacionais que poderiam contaminar a inferência estatística.

Do ponto de vista operacional, a amostra é composta por oito carteiras de ações, formadas a cada data de formação definida no desenho experimental. Quatro carteiras são geradas por modelos de linguagem de larga escala (LLMs) sem alimentação de dados estruturados (grupo “sem alimentação de dados”), baseando-se exclusivamente em instruções gerais para seleção de ativos; e quatro carteiras são geradas por LLMs expostos a um alimentação de dados padronizado de informações financeiras (grupo “com alimentação de dados”). O alimentação de dados consiste em planilhas extraídas do terminal Bloomberg contendo, em base anual desde 2015, valor de mercado (market cap, em milhões de reais), lucro líquido (net income, em milhões de reais), dividendos pagos e dividend yield; e, em base mensal desde 2015, a razão entre total investido somado ao caixa (em milhares de reais) e patrimônio líquido (em milhares de reais). Essas variáveis foram escolhidas por sintetizarem dimensões centrais de tamanho, rentabilidade, distribuição de resultados e estrutura de capital, fornecendo insumos objetivos para o processo decisório dos modelos com alimentação de dados. Para assegurar consistência interna, as unidades monetárias e periodicidades são mantidas conforme as planilhas originais e, quando necessário, harmonizadas para a escala da análise (por exemplo, conversões para bases mensais ou anuais compatíveis com a janela de formação).

A definição do Ibovespa como filtro inicial implica que, em cada data de formação, o conjunto elegível de ativos corresponde aos constituintes do índice naquele período. Esse procedimento com composição contemporânea evita viés de sobrevivência (survivorship bias), pois não se utiliza uma lista ex post de vencedores; também reduz risco de viés de antecipação (look-ahead bias), já que a elegibilidade é determinada pela informação disponível na própria data de formação. Para os grupos com alimentação de dados, somente permanecem elegíveis os ativos cujas observações atendam a requisitos mínimos de completude: (i) presença de histórico anual desde 2015 para market cap, net income, dividendos e dividend yield, admitindo-se lacunas pontuais a serem tratadas por regras de

imputação conservadoras; e (ii) disponibilidade do indicador mensal de estrutura de capital para o período de estimativa imediatamente anterior à formação. Quando a ausência de dados for material (por exemplo, mais de 20% de faltas na janela de referência) o ativo é excluído do universo elegível do grupo com alimentação de dados naquela formação, preservando a comparabilidade entre carteiras e evitando imputações que distorçam os sinais informacionais.

O recorte amostral limita-se a ações e units elegíveis no Ibovespa nas respectivas datas de formação. São excluídos, por definição, BDRs, ETFs, FIIs, notas estruturadas e quaisquer instrumentos que não representem diretamente participação acionária da companhia emissora. A presença de diferentes espécies ou classes acionárias (ON/PN/Units) segue a elegibilidade do próprio índice; eventuais sobreposições decorrentes de ações de mesma companhia com liquidez relevante são tratadas como ativos distintos, respeitando-se os limites e regras de pesos definidos na subseção de construção de carteiras.

Eventos corporativos são tratados por séries de retorno total (total return) ajustadas para proventos e desdobramentos, de modo a refletir reinvestimento de dividendos e neutralização de efeitos puramente mecânicos de splits ou grupamentos. Em casos de ofertas subsequentes, reorganizações societárias, cisões ou aquisições com troca de ações, adota-se o histórico ajustado do provedor de dados (Bloomberg) e, quando aplicável, o mapeamento para o ativo sucessor. Para deslistagens durante a janela de avaliação, interrompe-se a contribuição do papel a partir da data de saída, substituindo-se conforme regra de rebalanceamento estabelecida ou mantendo-se o caixa residual, conforme o protocolo definido na subseção de procedimentos (a escolha é documentada para transparência e replicabilidade).

A amostra observacional, portanto, possui duas camadas: (i) a camada de ativos elegíveis em cada data de formação (derivada da carteira teórica do Ibovespa vigente, com os filtros de dados aplicáveis ao grupo com alimentação de dados); e (ii) a camada de carteiras efetivamente formadas pelos LLMs (quatro com alimentação de dados e quatro sem alimentação de dados), que constituem as unidades experimentais para comparação de desempenho. A dimensão efetiva da amostra (número de observações) resulta do produto entre o número de carteiras e o número de janelas/intervalos de avaliação utilizados no estudo (por exemplo, janelas de 1, 3 e 6 meses), totalizando um painel carteira-tempo cuja estrutura é explicitada na subseção de avaliação de rentabilidade. Esse arranjo permite o uso de testes estatísticos entre grupos (com alimentação de dados versus sem alimentação de dados) controlando a variação ao longo do tempo.

Para assegurar independência entre grupos, proíbe-se a reutilização direta de sinais informacionais do alimentação de dados nas carteiras sem alimentação de dados. Diferenças entre as quatro carteiras de cada grupo derivam de variações controladas (por exemplo, prompts, amostragens e regras de desempate) previamente documentadas, garantindo replicabilidade e evitando enviesamentos por “melhor de N” (data-snooping). Adicionalmente, o protocolo de formação utiliza apenas informações disponíveis até a data de formação (por exemplo, dados anuais publicados até o final do ano anterior e dados mensais até o mês imediatamente precedente), evitando vazamentos temporais de informação.

Por fim, reconhece-se que a escolha do Ibovespa como referência confere à amostra um perfil de companhias tipicamente maiores e mais líquidas, o que favorece a investibilidade do resultado, mas restringe a generalização para o conjunto total de listadas na B3, em especial small caps de baixa liquidez. Para lidar com essa limitação, o estudo documenta a composição elegível em cada formação e, quando pertinente, realiza análises de sensibilidade (por exemplo, estratificando por setores ou por faixas de capitalização de mercado dentro do próprio marco amostral). Esse desenho de universo e amostra preserva o equilíbrio entre rigor acadêmico, viabilidade prática e pertinência ao objetivo central do trabalho: comparar, de forma causal e estatisticamente informada, o desempenho de carteiras geradas por LLMs com e sem exposição a um alimentação de dados estruturado de informações financeiras.

3.3 Construção da Carteira de Investimentos

A construção das carteiras de investimento neste estudo segue um protocolo experimental que visa avaliar o impacto do uso de informações estruturadas sobre a capacidade de modelos de linguagem de larga escala (LLMs) em selecionar ativos. Para tanto, foram utilizadas quatro inteligências artificiais distintas, sendo elas: ChatGPT, Gemini, Copilot e DeepSeek, sendo cada uma responsável por gerar carteiras em dois contextos: sem acesso a dados financeiros estruturados (grupo sem alimentação de dados) e com acesso a um conjunto padronizado de informações extraídas do terminal Bloomberg (grupo com alimentação de dados). Dessa forma, ao todo, foram geradas oito carteiras de ações, quatro sem alimentação de dados e quatro com alimentação de dados, assegurando comparabilidade entre os grupos.

No grupo sem alimentação de dados, os modelos foram instruídos apenas por prompts contendo orientações gerais para seleção de ativos pertencentes ao Ibovespa. O objetivo

dessas instruções foi induzir a construção de carteiras diversificadas, compostas por 10 ativos distintos, respeitando limites de concentração por ativo e por setor. Como não havia acesso a dados quantitativos, as escolhas se basearam exclusivamente em critérios qualitativos inferidos pelos modelos a partir de suas bases de conhecimento. Para garantir replicabilidade, as variações entre as carteiras decorreram de ajustes controlados nos parâmetros de geração (como seeds e temperatura), evitando múltiplas tentativas ad hoc e reduzindo risco de viés de seleção.

No grupo com alimentação de dados, as mesmas inteligências artificiais receberam, em complemento, planilhas contendo dados fundamentalistas e de estrutura de capital das empresas elegíveis no Ibovespa. Esse material incluiu variáveis anuais desde 2015, como valor de mercado (market cap, em milhões de reais), lucro líquido (net income, em milhões de reais), dividendos pagos e dividend yield, além de um indicador mensal referente à razão entre o total investido somado ao caixa e o patrimônio líquido (em milhares de reais). As planilhas foram organizadas de modo a apresentar apenas informações disponíveis até a data de formação das carteiras, prevenindo vazamento de dados futuros. Cada IA recebeu instruções padronizadas para gerar carteiras de 10 ativos, mas pôde interpretar e ponderar as variáveis do alimentação de dados segundo sua própria “lógica” interna. As diferenças entre as quatro carteiras com alimentação de dados refletem a diversidade das inteligências artificiais empregadas, o que permite observar como distintas arquiteturas e mecanismos de raciocínio respondem ao mesmo conjunto de dados.

A alocação dos ativos em todas as carteiras adotou o critério de equiponderação ($1/N$), de modo que cada ação recebeu peso inicial idêntico de 10% sobre o capital total. Essa escolha elimina distorções provenientes da estimação de retornos esperados ou covariâncias em horizontes curtos e assegura neutralidade na comparação entre carteiras. Como análise complementar, em etapa posterior, pode-se avaliar se alocações otimizadas, como por exemplo, de mínima variância com restrições realistas, alterariam os resultados obtidos com a alocação igualitária, mas tal abordagem não constitui o núcleo metodológico do experimento.

O horizonte de análise definido para o estudo é de três meses. Em cada data de formação, as carteiras foram constituídas e mantidas até o final do período, sem rebalanceamento intermediário. Os retornos foram apurados em reais, considerando preços de fechamento para entrada e saída, bem como séries de retorno total ajustadas para dividendos e eventos corporativos. Custos de transação foram incluídos de forma simplificada, à razão de 0,10% por operação, a fim de aproximar o experimento da prática de mercado. Em caso de deslistagens ou eventos extraordinários durante o horizonte, os ativos foram

liquidados na data do evento e os recursos mantidos em caixa até o encerramento do período.

Esse desenho experimental garante que as diferenças observadas nos resultados entre os grupos possam ser atribuídas, de forma causal, ao acesso ou não às informações estruturadas. O uso de múltiplas inteligências artificiais adiciona robustez ao estudo, permitindo verificar se os padrões de desempenho se repetem entre arquiteturas distintas. A opção por equiponderação e horizonte de três meses reforça a comparabilidade entre carteiras, reduzindo fontes de ruído e assegurando clareza na avaliação de rentabilidade e risco. Assim, a etapa de construção das carteiras estabelece a base necessária para a análise estatística posterior, na qual se busca identificar se o acesso ao alimentação de dados financeiro efetivamente contribui para decisões de investimento mais consistentes e alinhadas às práticas profissionais de mercado.

3.4 Análise de Rentabilidade

A avaliação da rentabilidade das carteiras constitui etapa central deste trabalho, pois é a partir dela que se estabelece a base empírica para comparar o desempenho das carteiras geradas por inteligências artificiais com e sem acesso ao alimentação de dados de informações financeiras.

O cálculo dos retornos de cada carteira foi realizado em reais (BRL), utilizando séries de preços ajustados para dividendos, desdobramentos e grupamentos, de modo a refletir o conceito de retorno total. Cada carteira, composta por 10 ativos selecionados, recebeu pesos iniciais iguais (10% por ativo), e não houve rebalanceamento ao longo do horizonte de três meses. Dessa forma, a evolução de cada carteira reflete a valorização acumulada dos papéis escolhidos, acrescida do efeito dos proventos distribuídos durante o período, e descontada de custos de transação fixados em 0,10% por operação de compra e venda.

A rentabilidade de cada carteira foi comparada com dois benchmarks de referência no mercado brasileiro. O primeiro é o Certificado de Depósito Interbancário (CDI), considerado a taxa livre de risco local e parâmetro básico para avaliar se estratégias ativas oferecem retorno superior ao mínimo esperado por um investimento conservador. O segundo é o Índice Bovespa (Ibovespa), que sintetiza o desempenho médio das ações mais líquidas e relevantes da B3. A inclusão desses dois benchmarks possibilita analisar, em paralelo, se as carteiras geradas superam tanto uma referência de renda fixa segura quanto o índice acionário mais representativo do país.

A sequência metodológica para o cálculo da rentabilidade e das métricas de desempenho foi estruturada conforme os seguintes passos:

1. **Cálculo do retorno total de cada ativo**, considerando o preço ajustado e os proventos distribuídos:

$$r_{i,t} = \frac{P_{i,t} + D_{i,t}}{P_{i,t-1}} - 1 \quad (3.1)$$

2. **Definição dos pesos equiponderados** para todos os ativos da carteira:

$$w_i = \frac{1}{N} \quad \text{para } i = 1, 2, \dots, N \quad (3.2)$$

3. **Cálculo do retorno diário da carteira**:

$$r_{p,t} = \sum_{i=1}^N w_i r_{i,t} \quad (3.3)$$

4. **Cálculo do retorno acumulado no horizonte de três meses**:

$$R_p^{(0 \rightarrow T)} = \prod_{t=1}^T (1 + r_{p,t}) - 1 \quad (3.4)$$

5. **Cálculo do Índice de Sharpe**, utilizando o excesso de retorno em relação à taxa livre de risco (CDI):

$$S = \frac{\bar{r}_p - \bar{r}_f}{\sigma_p} \quad (3.5)$$

O cálculo será realizado a partir dos retornos obtidos pelas oito carteiras geradas por inteligências artificiais, considerando o período de agosto a outubro de 2025. O retorno livre de risco (R_f) será extraído a partir da série histórica de fatores diários do CDI, enquanto o retorno acumulado de cada carteira (R_p) corresponde à variação percentual no período. Para tornar as magnitudes comparáveis, o desvio-padrão dos retornos diários será ajustado para a base trimestral por meio da multiplicação por $\sqrt{63}$, que corresponde ao número médio de dias úteis em três meses.

As volatilidades (σ_p) das carteiras serão obtidas diretamente a partir dos dados de mercado divulgados pela B3 (Brasil, Bolsa, Balcão), de modo a garantir maior aderência empírica e precisão estatística nos valores de risco.

6. **Estimativa do Alfa de Jensen**, obtido a partir da regressão do excesso de retorno da carteira em relação ao excesso de retorno do mercado (Ibovespa):

$$r_{p,t} - r_{f,t} = \alpha + \beta(r_{m,t} - r_{f,t}) + \varepsilon_t \quad (3.6)$$

$$\alpha = \bar{r}_p - [\bar{r}_f + \beta(\bar{r}_m - \bar{r}_f)] \quad (3.7)$$

Essas etapas compõem o procedimento de mensuração de desempenho utilizado neste trabalho, permitindo a análise comparativa entre as carteiras geradas por inteligências artificiais com diferentes níveis de informação, de forma quantitativa, objetiva e replicável.

3.5 Análise Estatística de Rentabilidade

A etapa de análise estatística tem como objetivo verificar se as diferenças observadas entre as carteiras formadas pelas inteligências artificiais com e sem alimentação de dados estruturados são estatisticamente significativas. Para isso, foi empregada a Análise de Variância (ANOVA) de dois fatores, sem repetição, adequada para comparar médias de grupos considerando dois critérios de classificação independentes.

O primeiro fator corresponde ao tipo de modelo de linguagem utilizado, e o segundo fator representa a condição de alimentação de dados (com ou sem alimentação de dados). A variável dependente é o retorno acumulado de cada carteira, calculado conforme os procedimentos descritos na Seção 3.4.

A aplicação da ANOVA segue as etapas formais de cálculo das somas de quadrados, médias quadráticas e estatísticas de teste, conforme descrito a seguir.

1. Cálculo da Soma Total dos Quadrados (SQT):

$$SQ_T = \sum_{i=1}^a \sum_{j=1}^b (Y_{ij} - \bar{Y}_{..})^2 \quad (3.8)$$

2. Soma dos Quadrados entre os Níveis do Fator A (Modelos de IA):

$$SQ_A = b \sum_{i=1}^a (\bar{Y}_{i.} - \bar{Y}_{..})^2 \quad (3.9)$$

3. Soma dos Quadrados entre os Níveis do Fator B (Condição de Alimentação):

$$SQ_B = a \sum_{j=1}^b (\bar{Y}_{.j} - \bar{Y}_{..})^2 \quad (3.10)$$

4. Soma dos Quadrados do Erro (Resíduo):

$$SQ_E = \sum_{i=1}^a \sum_{j=1}^b (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2 \quad (3.11)$$

5. Graus de Liberdade:

$$gl_A = a - 1 \quad (3.12)$$

$$gl_B = b - 1 \quad (3.13)$$

$$gl_E = (a - 1)(b - 1) \quad (3.14)$$

6. Cálculo das Médias Quadráticas:

$$MQ_A = \frac{SQ_A}{gl_A} \quad (3.15)$$

$$MQ_B = \frac{SQ_B}{gl_B} \quad (3.16)$$

$$MQ_E = \frac{SQ_E}{gl_E} \quad (3.17)$$

7. Estatísticas F calculadas:

$$F_A = \frac{MQ_A}{MQ_E} \quad (3.18)$$

$$F_B = \frac{MQ_B}{MQ_E} \quad (3.19)$$

8. Decisão estatística: Os valores de F_A e F_B são comparados aos valores críticos F_{α, gl_A, gl_E} e F_{α, gl_B, gl_E} , com nível de significância de 5%. Caso F calculado $> F$ crítico, rejeita-se a hipótese nula de igualdade das médias entre os grupos.

O procedimento foi executado em ambiente computacional, utilizando-se python. O resultado da ANOVA permite identificar se há diferença estatisticamente significativa no desempenho das carteiras em função do tipo de modelo de linguagem e da presença de alimentação de dados estruturados.

4 RESULTADOS

Este capítulo apresenta os resultados empíricos obtidos a partir da aplicação dos procedimentos metodológicos descritos no Capítulo 3. As análises foram conduzidas de forma a avaliar o desempenho das carteiras de investimento geradas por diferentes inteligências artificiais generativas, comparando seus resultados com benchmarks tradicionais do mercado financeiro brasileiro, como o CDI e o Ibovespa. São discutidas tanto as características estruturais das carteiras, em termos de composição e pesos atribuídos aos ativos, quanto suas respectivas rentabilidades acumuladas, índices de desempenho e medidas estatísticas. Por fim, o capítulo encerra com a aplicação da ANOVA de dois fatores, que permite testar a significância estatística das diferenças de rentabilidade observadas entre as IAs e entre as condições com e sem alimentação de dados.

4.1 Composição das Carteiras Geradas pelas Inteligências Artificiais

A construção das carteiras de investimento constitui a etapa inicial da análise empírica deste trabalho, permitindo compreender como diferentes modelos de linguagem de larga escala (LLMs) se comportam diante da tarefa de selecionar ações no mercado acionário brasileiro. Ao todo, foram formadas oito carteiras, quatro sem alimentação de dados estruturados (“sem alimentação de dados”) e quatro com alimentação quantitativa (“com alimentação de dados”), todas compostas por dez ativos equiponderados (10% cada), integrantes do índice Ibovespa. Essa estrutura padronizada assegura comparabilidade entre as estratégias e reduz o impacto de vieses de ponderação.

A geração das carteiras pelos modelos de linguagem de larga escala baseou-se na formulação de um prompt estruturado e padronizado, garantindo consistência metodológica entre as diferentes IAs avaliadas. O comando utilizado foi:

“Dado as empresas que compõem o Índice Ibovespa, monte uma carteira teórica de investimentos com horizonte de 3 meses com 10 ações, adicionando também os pesos de

cada ação na carteira.”

Esse prompt foi definido de forma a refletir o procedimento descrito na metodologia, permitindo que os modelos operassem sob as mesmas condições experimentais e utilizando exclusivamente as informações internas disponíveis em sua base de treinamento, exceto nos cenários em que receberam alimentação de dados adicional. Ao restringir o escopo do comando às empresas integrantes do Ibovespa, buscou-se aproximar o experimento de práticas reais de seleção de ativos baseadas em carteiras representativas e amplamente utilizadas por investidores e analistas.

A formulação do prompt também segue as recomendações da literatura sobre interação com modelos de linguagem, segundo as quais comandos diretos, objetivos e com restrição contextual tendem a gerar respostas mais consistentes e menos ambíguas. Essa estratégia favorece não apenas a comparabilidade entre as carteiras, mas também a análise estatística subsequente, pois minimiza variações geradas por interpretações divergentes do modelo em relação ao objetivo do usuário. Em linha com abordagens recentes de sistemas multi-agentes aplicados a finanças, a definição de instruções padronizadas possibilita avaliar de forma mais clara a capacidade dos modelos em estruturar recomendações de investimento baseadas em informações gerais do mercado.

Sendo assim, as carteiras do grupo sem alimentação de dados refletem a capacidade das IAs em selecionar empresas apenas com base em conhecimento contextual e sem acesso direto a indicadores financeiros. Já as carteiras com alimentação de dados incorporaram informações de natureza quantitativa, como indicadores de rentabilidade, alavancagem, valor de mercado e dividendos, permitindo decisões potencialmente mais fundamentadas. A seguir, apresentam-se as composições de cada uma das oito carteiras:

Tabela 4: Carteiras geradas pela IA Copilot

(a) Sem alimentação de dados		(b) Com alimentação de dados	
Ativo	Peso (%)	Ativo	Peso (%)
ITUB4	10	VALE3	10
EQTL3	10	PETR4	10
SUZB3	10	ITUB4	10
VALE3	10	WEGE3	10
PETR4	10	EQTL3	10
BBSE3	10	PRI03	10
EMBR3	10	RDOR3	10
TIMS3	10	LREN3	10
CPFE3	10	B3SA3	10
CYRE3	10	SUZB3	10

Tabela 5: Carteiras geradas pela IA Gemini

(a) Sem alimentação de dados		(b) Com alimentação de dados	
Ativo	Peso (%)	Ativo	Peso (%)
VALE3	10	ITUB4	10
PETR4	10	PETR4	10
ITUB4	10	VALE3	10
BBAS3	10	BBAS3	10
WEGE3	10	B3SA3	10
PRI03	10	BPAC11	10
RDOR3	10	SUZB3	10
RAIL3	10	EQTL3	10
ELET3	10	RADL3	10
RENT3	10	WEGE3	10

Tabela 6: Carteiras geradas pela IA ChatGPT

(a) Sem alimentação de dados		(b) Com alimentação de dados	
Ativo	Peso (%)	Ativo	Peso (%)
VALE3	10	SANB11	10
PETR4	10	BBDC3	10
ITUB4	10	ELET6	10
WEGE3	10	CPFE3	10
VIVT3	10	MRFG3	10
PRIO3	10	CMIN3	10
RENT3	10	CXSE3	10
SUZB3	10	EGIE3	10
BBAS3	10	BBAS3	10
LREN3	10	PETR3	10

Tabela 7: Carteiras geradas pela IA DeepSeek

(a) Sem alimentação de dados		(b) Com alimentação de dados	
Ativo	Peso (%)	Ativo	Peso (%)
VALE3	10	VALE3	10
PETR4	10	ITUB4	10
ITUB4	10	PETR4	10
WEGE3	10	WEGE3	10
RENT3	10	RADL3	10
SUZB3	10	BBSE3	10
BBAS3	10	SUZB3	10
PRIO3	10	EQTL3	10
EQTL3	10	PRIO3	10
MGLU3	10	BRFS3	10

4.2 Análises Implementadas

A presente seção apresenta a análise das rentabilidades acumuladas das carteiras geradas, bem como a comparação com os benchmarks de referência. O objetivo é verificar se as carteiras produzidas pelas diferentes inteligências artificiais apresentaram desempe-

nho superior ao CDI e ao Ibovespa no período considerado. Para tanto, são inicialmente demonstradas as rentabilidades dos índices de mercado e, em seguida, as rentabilidades das carteiras formadas, tanto na condição com alimentação de dados quanto sem alimentação. Essa estrutura permite observar o comportamento das IAs sob diferentes níveis de informação e fornece a base empírica para as etapas subsequentes de comparação estatística. Para tais análises, a ferramenta utilizado foi o Python.

4.2.1 Rentabilidade do CDI e Ibovespa

Durante o período de análise, compreendido entre agosto e outubro de 2025, o Certificado de Depósito Interbancário (CDI) apresentou uma rentabilidade acumulada de 3,70%, refletindo o comportamento de um ativo de baixo risco e de natureza essencialmente estável. A trajetória do CDI manteve-se próxima a uma linha de crescimento contínuo, caracterizando uma evolução previsível e compatível com a dinâmica de capitalização diária da taxa interbancária brasileira.

Em contrapartida, o Índice Bovespa (Ibovespa) registrou uma rentabilidade acumulada de 12,91% no mesmo intervalo. O desempenho significativamente superior ao CDI demonstra a existência de um prêmio de risco positivo no mercado acionário, associado à maior exposição a fatores macroeconômicos e expectativas corporativas. A variação do índice apresentou oscilações mais pronunciadas ao longo do período, compatíveis com o comportamento típico de um portfólio composto por ativos de renda variável, mas mantendo tendência ascendente na maior parte do horizonte observado.

A diferença entre os dois indicadores evidencia o papel do Ibovespa como benchmark representativo do retorno médio do mercado de capitais brasileiro e confirma o CDI como proxy adequado de taxa livre de risco. O intervalo analisado caracterizou-se por um ambiente de valorização expressiva do mercado acionário, impulsionado por expectativas de flexibilização monetária e melhora no fluxo de capitais estrangeiros. Assim, o desempenho de 12,91% do Ibovespa em comparação aos 3,70% do CDI representa um excesso de retorno de aproximadamente 9,2 pontos percentuais, valor que baliza a avaliação das carteiras experimentais desenvolvidas neste estudo.

Para tal, segue o gráfico de rentabilidade do CDI:

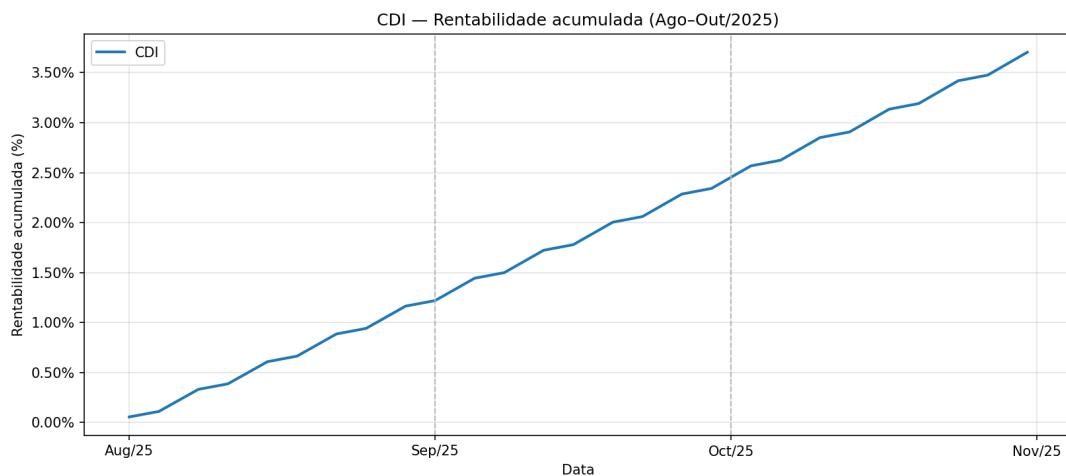


Figura 4: Rentabilidade Acumulada do CDI

Fonte: Autor

Segue a comparação de rentabilidade do CDI vs Ibovespa:

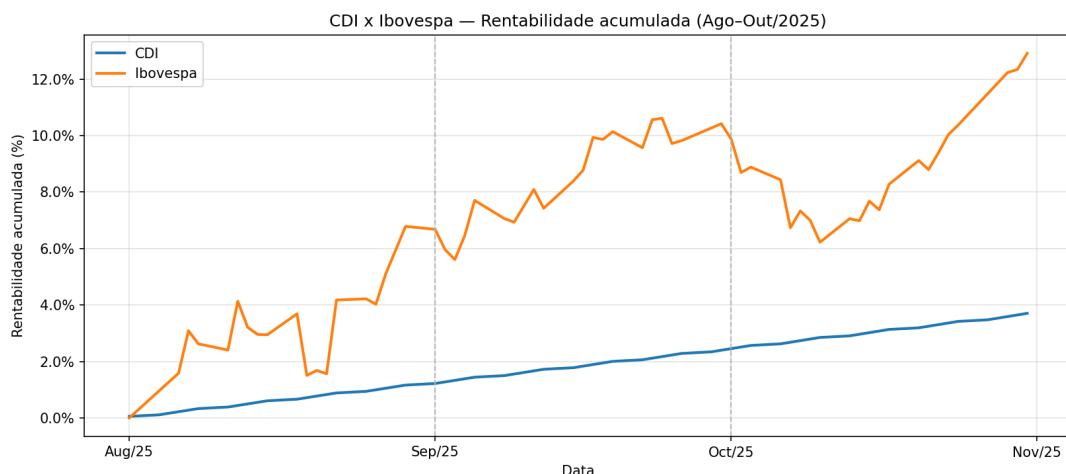


Figura 5: CDI vs Ibovespa

Fonte: Autor

Observa-se, portanto, que o cenário de referência apresenta um contraste marcante entre estabilidade e volatilidade, refletindo a diferença estrutural entre aplicações de renda fixa e renda variável. Esse comportamento serve como base para a análise subsequente das carteiras formadas por inteligências artificiais, que serão avaliadas em relação a ambos os benchmarks.

4.2.2 Rentabilidade das Carteiras Geradas

Os resultados a seguir apresentam, para cada carteira gerada, a rentabilidade acumulada por ativo no período analisado e o retorno acumulado do portfólio equiponderado. As quatro tabelas estão organizadas por modelo de IA, separando as condições sem alimentação de dados e com alimentação de dados:

Tabela 8: Carteiras geradas pela IA DeepSeek

(a) Sem alimentação de dados			(b) Com alimentação de dados		
Ativo	Peso (%)	Ret. acum. (%)	Ativo	Peso (%)	Ret. acum. (%)
BBAS3	10	19.35	BBSE3	10	-1.76
EQTL3	10	6.76	BRFS3	10	-11.66
ITUB4	10	12.91	EQTL3	10	6.76
MGLU3	10	8.27	ITUB4	10	12.91
PETR4	10	-7.64	PETR4	10	-7.64
PRI03	10	-12.10	PRI03	10	-12.10
RENT3	10	18.74	RADL3	10	45.99
SUZB3	10	-5.51	SUZB3	10	-5.51
VALE3	10	21.41	VALE3	10	21.41
WEGE3	10	13.23	WEGE3	10	13.23
Carteira (equiponderada)		8.05	Carteira (equiponderada)		6.16

Tabela 9: Carteiras geradas pela IA ChatGPT

(a) Sem alimentação de dados			(b) Com alimentação de dados		
Ativo	Peso (%)	Ret. acum. (%)	Ativo	Peso (%)	Ret. acum. (%)
ABEV3	10	1.47	ABEV3	10	1.47
B3SA3	10	0.56	B3SA3	10	0.56
BBDC4	10	16.19	BBDC3	10	14.62
BPAC11	10	16.43	BPAC11	10	16.43
HAPV3	10	12.35	EQTL3	10	6.76
ITUB4	10	12.91	ITUB4	10	12.91
PETR4	10	-7.64	PETR4	10	-7.64
PRI03	10	-12.10	RADL3	10	45.99
SUZB3	10	-5.51	SUZB3	10	-5.51
VALE3	10	21.41	VALE3	10	21.41
Carteira (equiponderada)		4.79	Carteira (equiponderada)		12.76

Tabela 10: Carteiras geradas pela IA Gemini

(a) Sem alimentação de dados			(b) Com alimentação de dados		
Ativo	Peso (%)	Ret. acum. (%)	Ativo	Peso (%)	Ret. acum. (%)
BBAS3	10	19.35	B3SA3	10	0.56
ELET3	10	47.10	BBAS3	10	19.35
ITUB4	10	12.91	BPAC11	10	16.43
PETR4	10	-7.64	EQTL3	10	6.76
PRIO3	10	-12.10	ITUB4	10	12.91
RAIL3	10	18.18	PETR4	10	-7.64
RDOR3	10	33.77	RADL3	10	45.99
RENT3	10	18.74	SUZB3	10	-5.51
VALE3	10	21.41	VALE3	10	21.41
WEGE3	10	13.23	WEGE3	10	13.23
Carteira (equiponderada)		15.86	Carteira (equiponderada)		13.20

Tabela 11: Carteiras geradas pela IA Copilot

(a) Sem alimentação de dados			(b) Com alimentação de dados		
Ativo	Peso (%)	Ret. acum. (%)	Ativo	Peso (%)	Ret. acum. (%)
BBSE3	10	-1.76	B3SA3	10	0.56
CPFE3	10	9.68	EQTL3	10	6.76
CYRE3	10	24.32	ITUB4	10	12.91
EMBR3	10	10.04	LREN3	10	-10.01
EQTL3	10	6.76	PETR4	10	-7.64
ITUB4	10	12.91	PRIO3	10	-12.10
PETR4	10	-7.64	RDOR3	10	33.77
SUZB3	10	-5.51	SUZB3	10	-5.51
TIMS3	10	17.23	VALE3	10	21.41
VALE3	10	21.41	WEGE3	10	13.23
Carteira (equiponderada)		8.74	Carteira (equiponderada)		5.34

4.2.3 Índice Sharpe

Usando os dados de volatilidade e de evolução dos ativos obtidos a partir das informações de mercado da B3, bem como a série histórica do CDI divulgada pelo Banco Central do Brasil, foram calculados os Índices de Sharpe das oito carteiras analisadas. O cálculo considerou o retorno acumulado no período de agosto a outubro de 2025 (R_p), o retorno livre de risco (R_f) correspondente ao CDI acumulado no mesmo intervalo, e o desvio-padrão trimestral (σ_p) de cada ativo obtido diretamente a partir das estatísticas de volatilidade da B3.

A Tabela 12 apresenta os resultados obtidos para cada carteira, dispostos em ordem alfabética.

Tabela 12: Índice de Sharpe das Carteiras Geradas (Ago–Out/2025)

Carteira	$R_p - R_f$	σ_p (trimestral)	Índice de Sharpe (S)
ChatGPT - Com Alimentação	0,0906	0,0628	1,44
ChatGPT - Sem Alimentação	0,0108	0,0622	0,17
Copilot - Com Alimentação	0,0163	0,0616	0,27
Copilot - Sem Alimentação	0,0504	0,0521	0,97
DeepSeek - Com Alimentação	0,0246	0,0558	0,44
DeepSeek - Sem Alimentação	0,0434	0,0711	0,61
Gemini - Com Alimentação	0,0950	0,0702	1,35
Gemini - Sem Alimentação	0,1216	0,0576	2,11

Fonte: Autor, com base em dados da B3 e do Banco Central do Brasil.

4.2.4 Alfa de Jensen

Usando as séries históricas diárias de retornos das carteiras e do Ibovespa, juntamente com o CDI diário como taxa livre de risco, foi estimado o Alfa de Jensen (α) de cada portfólio. O cálculo baseou-se na regressão dos retornos em excesso das carteiras sobre os retornos em excesso do mercado, conforme o modelo de Jensen, considerando o período de agosto a outubro de 2025. O α foi posteriormente trimestralizado pela multiplicação pelo número médio de dias úteis no trimestre ($\times 63$).

As medidas de covariância e variância foram calculadas diretamente a partir das séries históricas diárias de preços e volatilidade divulgadas pela B3, e serviram de base para o cálculo do coeficiente de sensibilidade (β) de cada carteira, conforme apresentado na Tabela 13.

Tabela 13: Covariância, Variância e Coeficiente Beta das Carteiras (Ago–Out/2025)

Carteira	Covariância	Variância	β
ChatGPT - Com Alimentação	0,000029	0,000033	0,867
ChatGPT - Sem Alimentação	0,000031	0,000035	0,899
Copilot - Com Alimentação	0,000032	0,000035	0,924
Copilot - Sem Alimentação	0,000028	0,000038	0,743
DeepSeek - Com Alimentação	0,000026	0,000036	0,720
DeepSeek - Sem Alimentação	0,000030	0,000029	1,026
Gemini - Com Alimentação	0,000033	0,000031	1,067
Gemini - Sem Alimentação	0,000034	0,000029	0,854

Fonte: Autor, com base em dados da B3 e do Banco Central do Brasil.

A partir dos valores de β , foi estimado o Alfa de Jensen (α), conforme a Equação 4.1. O resultado trimestralizado para cada carteira é apresentado na Tabela 14.

$$\alpha = R_p - [R_f + \beta(R_m - R_f)] \quad (4.1)$$

Tabela 14: Alfa de Jensen Trimestral das Carteiras (Ago–Out/2025)

Carteira	α (trimestral)
ChatGPT - Com Alimentação	0,0088
ChatGPT - Sem Alimentação	-0,0655
Copilot - Com Alimentação	-0,0650
Copilot - Sem Alimentação	-0,0158
DeepSeek - Com Alimentação	-0,0407
DeepSeek - Sem Alimentação	-0,0432
Gemini - Com Alimentação	-0,0043
Gemini - Sem Alimentação	0,0176

Fonte: Autor, com base em dados da B3 e do Banco Central do Brasil.

4.3 ANOVA de Dois Fatores

A Tabela 15 apresenta as rentabilidades percentuais obtidas por cada Inteligência Artificial (IA), ChatGPT, Gemini, Copilot e Deepseek, sob duas condições experimen-

tais: com e sem alimentação de dados financeiros. As linhas representam as IAs analisadas, enquanto as colunas correspondem às rentabilidades médias observadas em cada condição (Com Alimentação e Sem Alimentação). Essas observações servem de base para a aplicação da ANOVA de dois fatores sem repetição.

Tabela 15: Rentabilidade média (%) por IA e condição de alimentação de dados

IA	Com Alimentação (%)	Sem Alimentação (%)
ChatGPT	12,76	4,79
Gemini	13,20	15,86
Copilot	5,34	8,74
Deepseek	6,16	8,05

A ANOVA de dois fatores sem repetição foi conduzida considerando os fatores:

- Linhas (IA): representa as diferentes inteligências artificiais;
- Colunas (Alimentação): representa a condição com ou sem alimentação de dados.

Para a aplicação da ANOVA de dois fatores sem repetição, é necessário formular os conjuntos de hipóteses associados a cada fator avaliado. Assim, as hipóteses estatísticas testadas são:

- **Fator 1 (Modelos de IA):**

- H_0 : As médias dos retornos entre os diferentes modelos de IA são iguais.
- H_1 : Pelo menos um dos modelos de IA apresenta média de retornos diferente dos demais.

- **Fator 2 (Condição de Alimentação de Dados):**

- H_0 : As médias dos retornos entre os grupos com alimentação de dados e sem alimentação de dados são iguais.
- H_1 : As médias dos retornos dos grupos diferem entre si.

Tabela 16: ANOVA de dois fatores (sem repetição) — até a estatística F

Fonte de Variação	Soma de Quadrados	GL	Quadrado Médio	Estatística F
Linhas	75,08	3	25,03	1,75
Colunas	0,00005	1	0,00005	0,000003
Residual	42,86	3	14,29	—
Total	117,94	—	—	—

E para dar sequência na análise, utilizou-se da tabela de Distribuição F de Snedecor a 5%, para a qual encontrou-se os seguintes resultados:

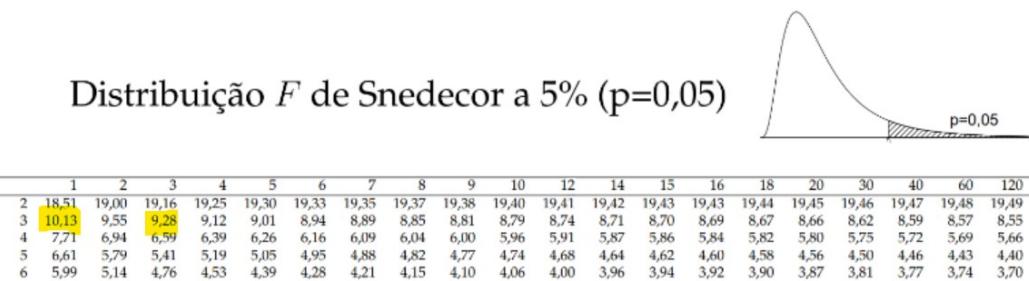


Figura 6: Tabela de Distribuição F de Snedecor a 5%

E a partir desses dados, foi possível concluir a análise:

Tabela 17: Comparaçāo: F calculado vs. F crítico (Snedecor, 5%)

Fator	GL	F_{calc}	$F_{\text{crit}} (5\%)$	Decisão (5%)
IA (linhas)	(3, 3)	1,75	9,28	Não rejeitar H_0
Alimentação (colunas)	(1, 3)	0,000003	10,13	Não rejeitar H_0

5 ANÁLISE DOS RESULTADOS

A presente seção reúne e discute os resultados obtidos ao longo do estudo, relacionando o desempenho das carteiras geradas pelos diferentes modelos de inteligência artificial com os benchmarks estabelecidos para o mercado brasileiro. A análise contempla tanto a comparação direta dos retornos e indicadores de performance quanto a interpretação dos efeitos associados aos fatores avaliados, permitindo uma visão integrada sobre o comportamento das carteiras nos distintos cenários testados. Além disso, são examinados os achados estatísticos provenientes da ANOVA, de forma a identificar diferenças significativas entre modelos e condições de alimentação de dados, consolidando as evidências empíricas que sustentam as conclusões da pesquisa.

5.1 Análise das Carteiras de Investimento

A análise da composição das carteiras permite compreender como as diferentes inteligências artificiais internalizaram suas visões de mercado ao selecionar ativos, bem como avaliar em que medida a alimentação de dados estruturados alterou essas escolhas. No total, as oito carteiras geradas (quatro IAs, com e sem alimentação de dados) resultaram em 31 ações distintas, distribuídas em 79 posições individuais, com cada carteira contendo dez ativos. Apesar dessa variedade, a concentração em um subconjunto de blue chips é bastante marcada, o que aponta para um comportamento relativamente convergente entre as IAs.

Quando se observa a frequência com que cada ativo aparece nas oito carteiras, emerge um núcleo bastante concentrado de ações recorrentes. Três papéis, sendo eles ITUB4, PETR4 e VALE3, aparecem em 7 das 8 carteiras (87,5% dos portfólios). Em seguida, SUZB3 e WEGE3 surgem em 6 carteiras (75%), enquanto BBAS3, EQTL3 e PRIO3 aparecem em 5 carteiras (62,5%). No agregado, apenas 8 ações respondem por 48 das 80 posições, ou seja, cerca de 60% de todas as alocações concentram-se nesse conjunto de nomes recorrentes.

Tabela 18: Ativos mais frequentes nas oito carteiras de IA

Ativo	Frequência	% das carteiras
ITUB4	7	87,5
PETR4	7	87,5
VALE3	7	87,5
SUZB3	6	75,0
WEGE3	6	75,0
BBAS3	5	62,5
EQTL3	5	62,5
PRIOR3	5	62,5

Do ponto de vista interpretativo, esse padrão sugere que, independentemente da IA ou da presença de alimentação de dados, os modelos convergem para um conjunto de large caps brasileiras de alta liquidez e relevância no Ibovespa, combinando bancos, commodities, óleo e gás e nomes industriais consolidados. Em outras palavras, as IAs tendem a replicar o núcleo do mercado acionário brasileiro, o que é coerente com uma seleção baseada em empresas de maior visibilidade, liquidez e cobertura por analistas.

Seguindo a análise, a comparação entre cada par de carteiras permite avaliar se o alimentação de dados mudou de fato a composição das carteiras ou apenas ajustou a periferia das posições.

No Copilot, metade da carteira permanece idêntica entre os dois cenários: EQTL3, ITUB4, PETR4, SUZB3 e VALE3 aparecem tanto na versão sem alimentação de dados quanto na versão com alimentação de dados. O que muda é a borda da carteira: sem alimentação de dados surgem BBSE3, CPFE3, CYRE3, EMBR3 e TIMS3, enquanto com alimentação de dados esses nomes são substituídos por B3SA3, LREN3, PRIOR3, RDOR3 e WEGE3. A alimentação de dados desloca a carteira de uma composição mais espalhada em elétricas, seguros e construção civil para uma combinação que inclui infraestrutura de mercado, varejo consolidado, saúde e exploração de petróleo, mantendo o núcleo de blue chips.

No caso do Gemini, novamente há um núcleo comum entre versões com e sem alimentação de dados: BBAS3, ITUB4, PETR4, VALE3 e WEGE3 aparecem em ambas. Sem alimentação de dados, a carteira inclui ELET3, PRIOR3, RDOR3 e RENT3; com alimentação de dados, esses nomes são substituídos por B3SA3, BPAC11, EQTL3, RADL3 e SUZB3. O alimentação de dados reforça a presença de financeiros, adiciona energia

elétrica e varejo farmacêutico, removendo parte da exposição a consumo cíclico.

O efeito do alimentação de dados no ChatGPT é o mais acentuado: apenas BBAS3 aparece nos dois cenários. Sem alimentação de dados, a carteira reúne ITUB4, LREN3, PETR4, PRIO3, RENT3, SUZB3, VALE3, VIVT3 e WEGE3. Com alimentação de dados, surge uma composição quase inteiramente nova, com BBDC3, CMIN3, CPFE3, CXSE3, EGIE3, ELET6, MRFG3, PETR3 e SANB11. Aqui, o alimentação de dados substitui uma carteira centrada em bancos tradicionais, varejo, commodities e telecom por uma mistura de bancos múltiplos adicionais, elétricas, seguros, mineração e proteínas, sugerindo uma preferência por setores mais defensivos e estáveis.

No DeepSeek, a sobreposição é alta: EQTL3, ITUB4, PETR4, PRIO3, SUZB3, VALE3 e WEGE3 aparecem em ambas as carteiras, com apenas três substituições. Sem alimentação de dados surgem BBAS3, MGLU3 e RENT3; com alimentação de dados, aparecem BBSE3, BRFS3 e RADL3. Assim como nas demais IAs, o alimentação de dados reduz a presença de consumo cíclico e aumenta a exposição a setores defensivos.

Em síntese, a alimentação de dados tende a preservar o núcleo das carteiras (as blue chips que aparecem de forma recorrente) e a recalibrar a borda do portfólio, substituindo empresas mais sensíveis ao ciclo ou à volatilidade por nomes de setores de maior resiliência setorial e previsibilidade de caixa.

Passando da análise de ativos para a análise setorial, o padrão de convergência torna-se ainda mais claro. Considerando todas as carteiras, os setores mais frequentes são:

Tabela 19: Distribuição setorial das ações selecionadas nas carteiras

Setor	Sem feed	Com feed	Total
Financeiro	8	9	17
Óleo e gás	7	6	13
Energia elétrica	4	6	10
Materiais básicos / Mineração	4	4	8
Indústria / Bens de capital	4	3	7
Materiais básicos / Papel e celulose	3	3	6
Consumo cíclico / Serviços	3	0	3
Consumo cíclico / Varejo	2	1	3
Consumo não cíclico / Alimentos	0	2	2
Financeiro / Infraestrutura de mercado	0	2	2
Saúde	1	1	2
Saúde / Varejo farmacêutico	0	2	2
Telecomunicações	2	0	2
Consumo cíclico / Construção	1	0	1
Financeiro / Seguros	0	1	1

O conjunto de carteiras é claramente ancorado em setores que representam o centro de gravidade estrutural da bolsa brasileira, como bancos, petróleo, mineração e energia elétrica. Esses setores são intensivos em capital, altamente regulamentados ou exportadores de commodities, alinhando-se com a composição média do Ibovespa.

Comparando versões com e sem alimentação de dados:

- Nas carteiras com alimentação de dados, aparecem setores que não surgem nas versões sem alimentação de dados, como:
 - Alimentos;
 - Infraestrutura de mercado financeiro;
 - Seguradoras;
 - Varejo farmacêutico.
- Nas carteiras sem alimentação de dados, surgem setores tipicamente mais cílicos e sensíveis à atividade doméstica, tais como:

- Consumo sensível;
- Varejo eletrônico;
- Construção civil;
- Telecomunicações.

Apenas nas carteiras com alimentação de dados aparecem setores como alimentos, infraestrutura de mercado financeiro, seguradoras e varejo farmacêutico.

Apenas nas carteiras sem alimentação de dados aparecem setores tipicamente mais cíclicos: consumo sensível, varejo eletrônico, construção civil e telecomunicações.

Essa diferença indica que o alimentação de dados incentiva as IAs a privilegiar setores mais defensivos e estáveis, enquanto sem alimentação de dados há maior propensão a escolher nomes com maior volatilidade e maior dependência do ciclo doméstico.

Todas as carteiras são equiponderadas, e em termos absolutos, tratam-se de portfólios concentrados. Contudo, a concentração efetiva é ainda maior: 31 ações diferentes foram selecionadas, mas 8 delas respondem por mais de 60% de todas as posições.

Do ponto de vista fundamental, a grande maioria dos nomes selecionados são empresas de grande porte, elevada liquidez e ampla cobertura por analistas. Não há microcaps, small caps extremas ou papéis de liquidez irrigária. Quando surgem empresas mais cíclicas, elas ainda pertencem ao universo de ativos acompanhados pelo mercado institucional.

O alimentação de dados reforça essa característica ao introduzir ativos de setores como infraestrutura financeira, seguros, saúde e alimentos, enquanto as versões sem alimentação de dados selecionam mais varejo, consumo cíclico e serviços dependentes do ciclo econômico interno.

5.2 Análise do CDI e Ibovespa

O período analisado foi marcado por uma valorização expressiva do Ibovespa, sustentada por vetores macroeconômicos internos e externos que ampliaram o apetite por risco e favoreceram setores específicos da economia brasileira. Essa dinâmica não apenas explica a alta do índice como também permite compreender, de maneira estrutural, o comportamento das carteiras geradas pelas quatro IAs avaliadas. A performance dos portfólios está diretamente associada à interação entre: (i) o ciclo de commodities, (ii) o fluxo estrangeiro para emergentes, (iii) a estabilidade inflacionária, (iv) a manutenção da

Selic em 15% ao ano e (v) a composição setorial das carteiras, especialmente sua exposição relativa a setores vencedores ou penalizados no trimestre.

No ambiente global, a expectativa de flexibilização monetária nos Estados Unidos, ainda que marcada por incertezas e divergências internas entre dirigentes do Federal Reserve, reduziu a aversão ao risco e impulsionou os mercados emergentes. Commodities metálicas e energéticas tiveram recuperação relevante, e o dólar perdeu força frente às moedas emergentes. Esse contexto favoreceu diretamente ações como Vale, Petrobras, Suzano, Prio e Gerdau, todas com forte peso relativo no Ibovespa e presentes, em maior ou menor grau, nas carteiras construídas pelas IAs. Paralelamente, tensões geopolíticas continuaram a introduzir volatilidade nos preços de energia, mas sem impedir a trajetória de alta desses ativos no agregado.

No plano doméstico, a combinação entre inflação em desaceleração, câmbio relativamente estável e entrada líquida de capital estrangeiro contribuiu para sustentar a valorização das ações. Mesmo com a Selic mantida em 15% ao ano, nível elevado que encarece crédito, reprime consumo e aumenta o custo de oportunidade da renda variável, setores menos sensíveis ao ciclo econômico, como mineração, petróleo, bancos e energia elétrica, apresentaram desempenho robusto. Esses setores possuem características estruturais que reduzem sua dependência de condições financeiras expansionistas, como elevada geração de caixa, receitas dolarizadas, contratos regulados ou margens estáveis. Assim, o Ibovespa, cuja composição é majoritariamente concentrada nesses setores, respondeu positivamente ao cenário, atingindo máximas históricas e superando amplamente o retorno do CDI no período.

É justamente nesse ponto que o desempenho das carteiras geradas pelas IAs se torna plenamente comprensível. De forma geral, as carteiras com alimentação de dados (feed) apresentaram maior concentração em setores resilientes ao juro elevado e beneficiados pelo cenário externo, como commodities, utilities e bancos. Já as carteiras sem alimentação exibiram maior dispersão e maior presença de setores cíclicos, varejo, tecnologia local, construção civil e consumo discricionário, segmentos fortemente penalizados pelo custo de capital elevado e pela restrição de crédito imposta pela Selic. Essa diferença estrutural é consistente com o comportamento observado: carteiras com feed performaram de maneira mais estável e menos volátil, enquanto carteiras sem feed foram mais sensíveis ao ciclo doméstico e à volatilidade de curto prazo, apresentando retornos inferiores ou mais erráticos.

A performance particularmente elevada de Gemini e ChatGPT com alimentação, por

exemplo, está em linha com sua forte exposição a empresas como Vale, Petrobras, bancos privados e companhias elétricas. Ao serem beneficiadas simultaneamente pela alta das commodities, pelo fluxo estrangeiro e por resultados corporativos resilientes em ambiente de juro alto, essas carteiras capturaram de forma eficiente o movimento de valorização predominante no período. A carteira Gemini sem alimentação, embora tenha apresentado o maior retorno absoluto entre todas as oito carteiras, alcançou esse desempenho por motivos distintos: maior concentração em ativos de beta alto (alto risco sistemático) e maior sensibilidade a movimentos amplos de mercado. Assim, seu resultado elevado está associado a uma exposição mais agressiva, mas não necessariamente mais eficiente em termos de risco.

Em direção oposta, as carteiras sem alimentação de dados que alocaram recursos em small caps, varejo, construção e tecnologia local, segmentos que enfrentaram desaceleração do consumo, endividamento elevado, impacto no custo financeiro e compressão de margens, tiveram desempenho inferior. Esse fenômeno reflete não apenas o efeito restritivo da Selic em 15% ao ano, mas também a valorização concentrada do Ibovespa em poucos setores, de modo que carteiras com alocação mais pulverizada sofreram com a ausência de motores de retorno equivalentes aos grandes pesos do índice.

A relação entre CDI e Ibovespa reforça ainda mais o diagnóstico. Com o CDI acumulando cerca de 3,7% no trimestre, qualquer performance inferior a esse patamar representa incapacidade da carteira de superar o ativo livre de risco, e apenas parte das carteiras analisadas atingiu esse nível. O fato de o Ibovespa ter registrado retorno próximo de 12,9% no mesmo período demonstra que a captura dos setores líderes do índice foi determinante para superar o CDI, e apenas as carteiras mais alinhadas à estrutura setorial vencedora conseguiram atingir esse patamar. Assim, o contexto macroeconômico não apenas explica a alta do Ibovespa, mas também ilumina com clareza a lógica por trás do desempenho relativo entre as carteiras avaliadas.

Em síntese, o período foi caracterizado por um mercado acionário impulsionado por vetores externos e internos que favoreceram setores específicos de grande peso no Ibovespa. As carteiras geradas pelas IAs refletiram, com maior ou menor precisão, esses vetores: aquelas com maior exposição a commodities, energia e bancos foram naturalmente favorecidas, enquanto aquelas com maior presença de setores penalizados pelo juro elevado apresentaram resultados mais modestos. A análise conjunta do cenário macroeconômico e da composição das carteiras evidencia, portanto, que a performance obtida não é obra do acaso, mas o resultado da interação entre o ambiente econômico, a estrutura setorial do mercado brasileiro e as escolhas de alocação realizadas por cada IA.

5.3 Análise das Rentabilidades

A análise comparativa do desempenho das oito carteiras geradas pelas IAs (ChatGPT, Copilot, Gemini e DeepSeek), em suas versões com alimentação de dados e sem alimentação de dados, constitui uma etapa fundamental para avaliar a efetividade das diferentes arquiteturas de modelos generativos na construção de portfólios de ações no contexto brasileiro. Além de mensurar o retorno absoluto dos portfólios, essa seção busca interpretar, à luz de fatores macroeconômicos e setoriais, a capacidade dessas carteiras de superar benchmarks tradicionais, notadamente o CDI e o Ibovespa, entendendo se há evidências de geração de alfa, consistência decisória e sensibilidade às dinâmicas do mercado acionário nacional.

Ao longo do período analisado (agosto a outubro de 2025), o CDI acumulou aproximadamente 3,7%, refletindo a manutenção de uma taxa Selic em patamar extremamente elevado (15% a.a.), o que estabelece um custo de oportunidade alto para estratégias de renda variável. Em contraste, o Ibovespa registrou forte valorização, acumulando cerca de 12,91%, impulsionado principalmente pela performance expressiva de empresas ligadas a commodities metálicas, petróleo, energia elétrica e serviços financeiros. Essa divergência estrutural entre os benchmarks reforça que superar o CDI é condição mínima para justificar o risco da renda variável, enquanto superar o Ibovespa implica a capacidade de capturar, de forma eficiente, as principais tendências setoriais que impulsionaram o índice.

A partir dessa referência, passa-se à análise da trajetória das carteiras ao longo do tempo. As curvas de rentabilidade acumulada revelam não apenas o retorno final, mas também a estabilidade relativa, a consistência direcional e o comportamento das IAs frente a momentos de volatilidade de mercado.

Os quatro gráficos individuais (um para cada IA) permitem observar como cada arquitetura reagiu às condições de mercado em tempo real, evidenciando padrões específicos de volatilidade, divergência entre as versões com e sem alimentação de dados, e diferentes sensibilidades às inflexões do Ibovespa e do CDI.

5.3.1 Chat GPT

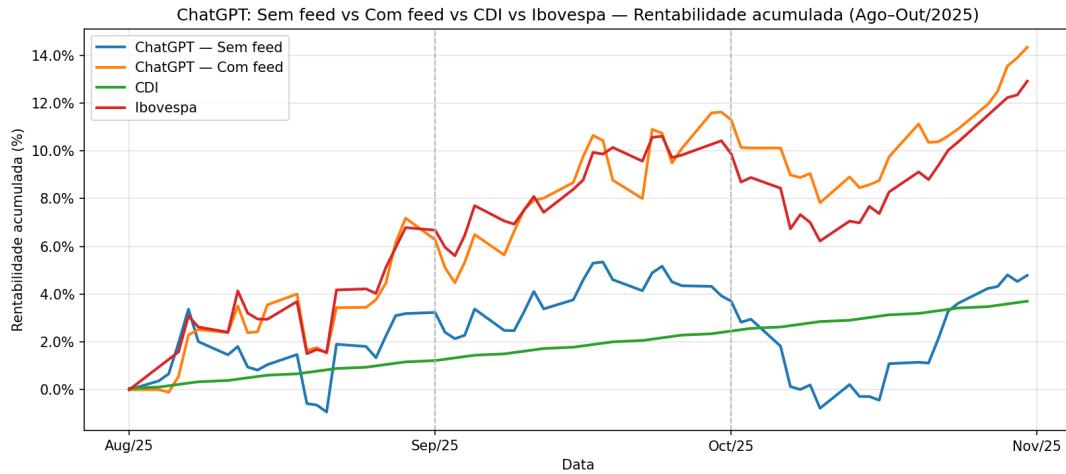


Figura 7: ChatGPT vs CDI vs Ibovespa

Fonte: Autor

As carteiras geradas pelo ChatGPT apresentam comportamento distinto dos demais modelos e evidenciam um interessante contraste entre as versões com e sem alimentação de dados. A versão com alimentação de dados foi consistentemente superior, com retorno acumulado próximo de 14%, acompanhando de forma muito próxima a trajetória do Ibovespa e superando-o em momentos específicos. Esse desempenho sugere que o ChatGPT se beneficia fortemente da adição de dados estruturados, utilizando-os para ajustar sua inferência em direção a setores que, no período analisado, apresentaram crescimento notável, especialmente petróleo e bancos.

Em contraposição, a versão sem alimentação de dados apresenta comportamento mais errático, com volatilidade elevada e retorno acumulado inferior ao do índice, embora ainda acima do CDI na maior parte do período. Essa diferença revela que o ChatGPT, ao operar sem dados externos, tende a diversificar com maior intensidade, o que reduz sua capacidade de capturar movimentos direcionais fortes do mercado.

5.3.2 Copilot

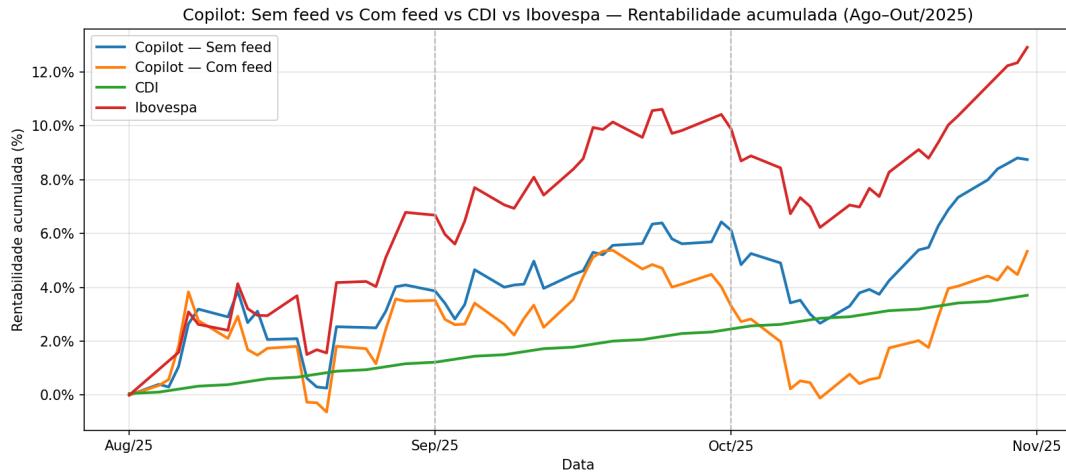


Figura 8: Copilot vs CDI vs Ibovespa

Fonte: Autor

O Copilot apresenta um comportamento intermediário entre os dois grupos anteriores. A versão sem alimentação de dados demonstra trajetória positiva relativamente estável, superando o CDI com facilidade e oscilando próxima à curva do Ibovespa. Contudo, sua capacidade de capturar os movimentos ascendentes mais intensos do mercado é limitada, o que sugere que o Copilot adota, por padrão, uma postura mais conservadora, possivelmente priorizando ativos de menor volatilidade ou maior diversificação em sua inferência.

A versão com alimentação de dados apresenta desempenho inferior ao da versão sem alimentação de dados e ao do próprio Ibovespa. Há momentos de queda acentuada, especialmente no final de outubro, quando o índice apresentava trajetória de recuperação. Essa defasagem indica que, para o Copilot, o fornecimento de dados estruturados não apenas não aprimorou sua capacidade de seleção, como pode ter perturbado o equilíbrio inferencial, levando-o a priorizar setores ou ativos que tiveram performance mais fraca no período.

5.3.3 DeepSeek

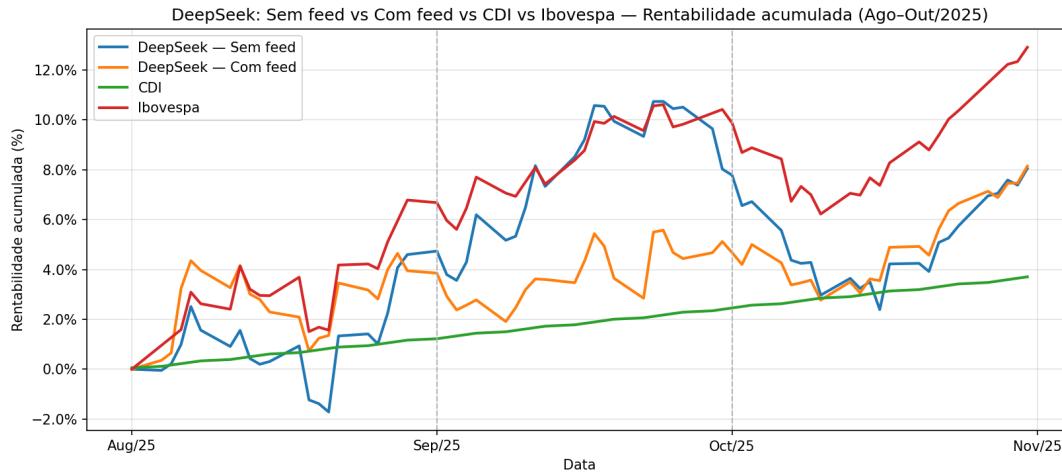


Figura 9: DeepSeek vs CDI vs Ibovespa

Fonte: Autor

A análise da evolução temporal do DeepSeek evidencia um comportamento caracterizado por elevada oscilação no retorno acumulado, sobretudo na versão sem alimentação de dados. Observa-se que a carteira apresenta movimentos mais amplos, com quedas e recuperações sucessivas, sugerindo que a IA, ao operar sem dados estruturados, apresenta maior exposição a ativos cíclicos e a setores voláteis. Apesar disso, a versão sem alimentação de dados demonstra uma capacidade não trivial de recuperação, aproximando-se do Ibovespa em momentos específicos, especialmente na transição de setembro para outubro. Essa dinâmica sugere que a IA capturou parte das tendências setoriais dominantes, mas com menor precisão e maior volatilidade.

Por outro lado, a versão com alimentação de dados apresenta comportamento marcadamente mais conservador, com curvas de retorno mais próximas ao CDI durante boa parte do período. A ausência de inclinação acentuada na curva sugere que, ao receber dados estruturados, o modelo desloca sua inferência para um portfólio com maior estabilidade e menor exposição a riscos, sacrificando potencial de ganho nos momentos em que o mercado apresentou tendência claramente positiva. Ambas as versões, entretanto, não superam o Ibovespa, evidenciando dificuldade em antecipar corretamente os vetores de valorização mais relevantes do trimestre.

5.3.4 Gemini

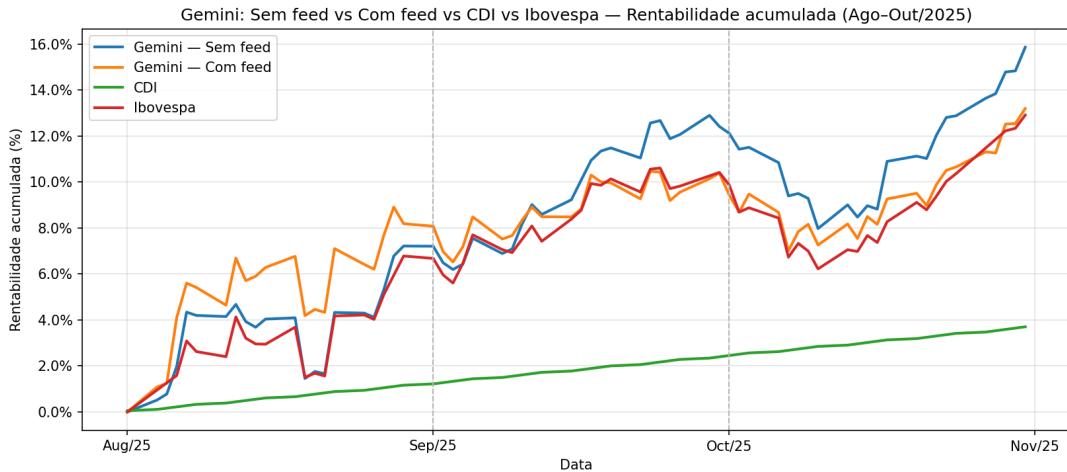


Figura 10: Gemini vs CDI vs Ibovespa

Fonte: Autor

O comportamento das carteiras derivadas do Gemini contrasta significativamente com as do DeepSeek. A versão sem alimentação de dados apresenta um desempenho notavelmente superior, com trajetória consistente de valorização e patamar final de retorno acima de 16%, superando inclusive o Ibovespa. Esse resultado indica que a IA foi capaz de, espontaneamente, selecionar uma combinação de ativos fortemente alinhada aos setores vencedores do período, sobretudo mineração, petróleo e grandes bancos, capturando tanto o movimento estrutural quanto as inflexões de curto prazo.

A versão com alimentação de dados, ainda que também apresente retorno elevado, fica ligeiramente abaixo do retorno da versão sem alimentação de dados. Isso sugere que o fornecimento de dados estruturados pode, paradoxalmente, restringir a capacidade exploratória do modelo, levando-o a escolhas moderadas que diminuem a agressividade ou a exposição aos setores de maior impulso. Ainda assim, ambas as versões superaram amplamente o CDI e tiveram desempenho superior à maioria das demais IAs, posicionando o Gemini como o modelo mais eficiente na tarefa de seleção de ativos ao longo do período estudado.

5.3.5 Comparação Consolidada vs CDI e Ibovespa

A comparação consolidada das oito carteiras frente aos benchmarks tradicionais, CDI e Ibovespa, constitui a etapa central da avaliação de desempenho deste estudo, pois permite identificar se as IAs foram capazes de gerar retornos ajustados à realidade do mercado

brasileiro e se tais retornos foram superiores ao que seria alcançado por um investidor passivo. Esta análise requer não apenas a observação do retorno final das carteiras, mas também a compreensão do comportamento relativo ao longo do período, a sensibilidade às oscilações do mercado e a consistência das escolhas dos modelos em contextos distintos de volatilidade e tendência.

A primeira comparação relevante é entre as carteiras e o CDI, que acumulou aproximadamente 3,7% no período. Como o CDI funciona como uma proxy para o ativo livre de risco no Brasil, qualquer portfólio que não supere esse benchmark torna-se financeiramente pouco atraente, uma vez que toma risco sem oferecer compensação adequada. Nesse sentido, observa-se que a maioria das carteiras das IAs superou o CDI de forma consistente, indicando que suas estratégias foram capazes de capturar movimentos positivos do mercado mesmo com metodologias não otimizadas e totalmente automatizadas. A superação do CDI é, portanto, um indicativo inicial de que as IAs possuem potencial para auxiliar investidores no processo de seleção de ativos, ainda que não substituam por completo processos tradicionais de análise fundamentalista ou técnica.

Entretanto, a análise se torna substancialmente mais exigente quando se adota o Ibovespa como parâmetro. Ao registrar aproximadamente 13% de retorno no período observado, o índice estabelece um benchmark de alta performance, refletindo a forte valorização de setores que historicamente possuem peso significativo em sua composição, como mineração, petróleo, bancos e energia elétrica. Em um trimestre em que os setores líderes apresentaram desempenhos acima da média histórica, superar o Ibovespa exige que a carteira possua uma concentração adequada nesses vetores, ou que o modelo de IA tenha sido capaz de capturar empresas que, mesmo fora da composição principal do índice, exibiram comportamento ascendente excepcional.

Quando se compararam as curvas de rentabilidade acumulada das carteiras com a do Ibovespa, torna-se evidente que a superação do benchmark é um fenômeno restrito a poucos modelos. As duas principais estratégias que se destacam são: a carteira Gemini sem alimentação de dados, que apresenta retorno acumulado superior ao Ibovespa e trajetória temporal consistente, e a carteira ChatGPT com alimentação de dados, que acompanha de forma próxima o desempenho do índice e o supera em momentos específicos. Esse resultado é particularmente relevante porque demonstra que não há uma relação monotônica entre acesso a dados estruturados e desempenho. Em outras palavras, a alimentação adicional de informações não garante, por si só, uma alocação mais eficiente; a eficácia da IA depende de sua própria arquitetura interna, da lógica de inferência estatística utilizada e da forma como interpreta informações complementares.

O comportamento das demais carteiras reforça esse ponto. No caso do Copilot, a versão sem alimentação de dados apresentou desempenho razoavelmente alinhado ao índice, enquanto a versão com alimentação de dados ficou significativamente abaixo, demonstrando sensibilidade excessiva a momentos de correção do mercado e limitada capacidade de capturar os movimentos de alta. No caso do DeepSeek, ambas as versões não superaram o Ibovespa, embora tenham tido momentos de recuperação expressiva. Esse desempenho sinaliza que a lógica inferencial do DeepSeek tende a construir portfólios mais expostos à volatilidade idiossincrática e menos alinhados ao beta do mercado, o que reduz sua capacidade de acompanhar tendências setoriais dominantes.

A análise do ranking comparativo das carteiras reforça os padrões identificados nas curvas de rentabilidade acumulada e permite quantificar, de forma objetiva, o grau de superação ou defasagem de cada estratégia em relação aos benchmarks. A carteira Gemini sem alimentação de dados apresenta a maior superioridade relativa, com diferença de 12,30 pontos percentuais acima do CDI e 3,09 pontos percentuais acima do Ibovespa, consolidando-se como a única carteira que não apenas supera de maneira expressiva o retorno livre de risco, mas também excede, com margem confortável, o desempenho do principal índice do mercado acionário brasileiro. Em seguida, a carteira ChatGPT com alimentação de dados ocupa posição de destaque ao apresentar retorno 10,60 pontos percentuais superior ao CDI e 1,39 ponto percentual acima do Ibovespa, revelando que a alimentação de dados estruturados contribuiu de forma significativa para o aprimoramento da performance do modelo.

De forma intermediária, mas ainda positiva, observa-se que a carteira Gemini com alimentação de dado apresenta pequena vantagem em relação ao Ibovespa (0,19 ponto percentual), além de retorno sólido frente ao CDI. Essa proximidade com o benchmark sugere que, embora o alimentação de dados limite marginalmente o nível de retorno em relação à versão sem alimentação de dados, ele mantém a carteira alinhada aos movimentos estruturais do mercado. Em contraste, as carteiras Copilot sem alimentação de dados e DeepSeek sem alimentação de dados, apesar de apresentarem retornos superiores ao CDI, ficam entre 4 e 5 pontos percentuais abaixo do Ibovespa, evidenciando dificuldade em acompanhar o ritmo ascendente do índice, especialmente em momentos de forte valorização setorial.

Por fim, os desempenhos mais fracos concentram-se nas versões sem e com alimentação de dados do Copilot e do DeepSeek, que exibem diferenças negativas substanciais em relação ao Ibovespa, variando entre -7 e -9 pontos percentuais. Esses resultados mostram que, além de retornos absolutos inferiores, tais modelos não foram capazes de capturar

apropriadamente os vetores de valorização que dominaram o mercado no período analisado. A dispersão de resultados entre as oito carteiras revela a influência determinante da arquitetura e da lógica interna de cada IA, reforçando que o desempenho não depende apenas da inserção de dados estruturados, mas da forma como cada modelo processa, generaliza e prioriza informações para construir seus portfólios.

Assim, o ranking comparativo sintetiza, de maneira numérica e objetiva, a eficiência relativa das estratégias, fortalecendo a evidência de que o sucesso das carteiras está intrinsecamente relacionado à sua capacidade de alinhar-se aos setores de maior relevância no período e à habilidade do modelo de IA em identificar padrões de mercado dominantes, com ou sem alimentação adicional de dados.

Tabela 20: Ranking comparativo das carteiras em relação ao CDI e ao Ibovespa

Carteira	Retorno (%)	Δ CDI (%)	Δ Ibovespa (%)
Gemini Sem alimentação de dados	15,86	12,16	2,95
Gemini Com alimentação de dados	13,20	9,50	0,29
ChatGPT Com alimentação de dados	12,79	9,09	-0,12
Copilot Sem alimentação de dados	8,74	5,04	-4,17
DeepSeek Sem alimentação de dados	8,05	4,35	-4,86
Copilot Com alimentação de dados	5,34	1,64	-7,57
ChatGPT Sem alimentação de dados	4,79	1,09	-8,12
DeepSeek Com alimentação de dados	6,16	2,46	-6,75
CDI	3,70	—	—
Ibovespa	12,91	—	—

A comparação consolidada também revela um aspecto estrutural: as carteiras que performaram melhor são aquelas que, de forma intencional ou não, concentraram-se nos setores vencedores descritos no tópico anterior. Embora cada IA tenha construído portfólios distintos, observa-se que a carteira Gemini sem alimentação de dados selecionou ativos que refletem o comportamento do núcleo do Ibovespa, enquanto a carteira ChatGPT com alimentação de dados incorporou empresas de alta capitalização e setores resilientes. Por outro lado, Copilot e DeepSeek dispersaram suas escolhas ou selecionaram setores com comportamento inferior ao índice.

Essa constatação reforça um princípio amplamente discutido na literatura de finanças: a composição setorial explica parcela substancial do retorno de uma carteira, frequente-

mente superando o impacto da seleção individual de ações. Assim, o desempenho superior das duas carteiras vencedoras pode estar mais relacionado à captura da tendência setorial dominante do período do que propriamente à escolha intrínseca de cada papel.

Ao observar o conjunto das oito carteiras, torna-se claro que a superação do Ibovespa permanece como um desafio, refletindo a própria dificuldade enfrentada por gestores profissionais. Mesmo assim, o fato de uma das carteiras ter superado o índice de forma consistente e outra ter se aproximado fortemente sugere que o uso de IAs pode ter potencial de complementaridade na construção de portfólios, sobretudo em horizontes curtos e contextos de forte tendência macroeconômica.

5.3.6 Indice Sharpe e Alfa de Jensen

A análise conjunta do Sharpe e do Alfa de Jensen permite avaliar não apenas o retorno observado, mas a eficiência risco-retorno e a capacidade de geração de valor acima do retorno esperado pelo risco sistemático. A tabela evidencia diferenças marcantes entre as estratégias das quatro IAs, revelando tanto padrões consistentes quanto impactos assimétricos da alimentação de dados.

Tabela 21: Índice de Sharpe e Alfa de Jensen (trimestral) das carteiras

IA	Sharpe	Alfa (trimestral)
ChatGPT - Com Alimentação	1,44	0,0088
ChatGPT - Sem Alimentação	0,17	-0,0655
Copilot - Com Alimentação	0,27	-0,0650
Copilot - Sem Alimentação	0,97	-0,0158
DeepSeek - Com Alimentação	0,44	-0,0407
DeepSeek - Sem Alimentação	0,61	-0,0432
Gemini - Com Alimentação	1,35	-0,0043
Gemini - Sem Alimentação	2,11	0,0176

O primeiro ponto de destaque é o desempenho da carteira Gemini sem alimentação, que apresenta o maior Sharpe (2,11) e o maior Alfa do conjunto (0,0176). Trata-se, portanto, da estratégia mais eficiente dentre todas as avaliadas: combina retorno absoluto elevado, volatilidade relativamente moderada e capacidade efetiva de superar o retorno esperado pelo CAPM. Seu desempenho sugere que, para este modelo específico, a ausência de alimentação de dados potencializou o caráter exploratório da IA, levando-a a identificar

oportunidades de mercado superiores às da versão com alimentação.

A versão Gemini com alimentação também apresenta desempenho expressivo, com Sharpe de 1,35, embora o alfa seja próximo de zero. Isso indica que a carteira se alinhou ao mercado e transformou risco em retorno de forma eficiente, mas sem gerar valor adicional além do esperado pelo seu nível de risco sistemático. Em outras palavras, trata-se de uma carteira sólida, estável e bem calibrada, porém sem componente significativo de geração de alfa.

O ChatGPT com alimentação ocupa posição semelhante: Sharpe elevado (1,44) e alfa positivo (0,0088). A combinação sugere que o fornecimento de dados estruturados aprimorou a qualidade da seleção de ativos, reduzindo ruído de decisão e permitindo que o modelo se aproxime da fronteira eficiente. A performance contrasta de forma marcante com a versão sem alimentação, cujo Sharpe é muito inferior (0,17) e alfa fortemente negativo (-0,0655). Essa assimetria revela que o alimentação de dados desempenha papel central no funcionamento do ChatGPT, sendo determinante para sua capacidade de alocação eficiente.

As estratégias do Copilot e do DeepSeek mostram padrões mais modestos. Ambas apresentam alfa negativo em suas duas versões, indicando que o retorno obtido foi inferior ao retorno esperado para o nível de risco sistemático assumido. Ainda assim, observam-se nuances relevantes: o Copilot sem alimentação apresenta Sharpe de 0,97, muito acima do Sharpe de sua versão com alimentação (0,27), o que sugere que o alimentação de dados, para esse modelo, introduz excesso de cautela ou escolhas menos alinhadas aos vetores de valorização do mercado brasileiro. O comportamento do DeepSeek é semelhante: sua versão sem alimentação supera a com alimentação (Sharpe 0,61 versus 0,44), ainda que ambas permaneçam abaixo das estratégias mais eficientes.

A tabela evidencia um padrão central: o impacto do alimentação de dados é profundamente dependente da arquitetura da IA. Enquanto ChatGPT e Gemini se beneficiam da alimentação de dados estruturados (ainda que em magnitudes distintas), Copilot e DeepSeek apresentam degradação na eficiência risco-retorno quando recebem alimentação de dados. Isso reforça que o desempenho das IAs não depende apenas da disponibilização de informação, mas da forma como cada modelo interpreta, prioriza e transforma esses dados em decisões de portfólio.

No conjunto, a análise dos indicadores Sharpe e Alfa confirma que o retorno absoluto por si só não é suficiente para avaliar o desempenho das estratégias. As melhores carteiras, especialmente Gemini sem alimentação e ChatGPT com alimentação, são também aquelas

com maior eficiência risco-retorno e maior capacidade de geração de valor não explicado pelo mercado. Já as estratégias com maiores fragilidades apresentam simultaneamente baixos Sharpe e alfa negativo, revelando limitações estruturais de processo decisório.

5.3.7 Análise Estatística

Os resultados mostraram que o fator IA apresentou soma de quadrados de 75,08, com três graus de liberdade e quadrado médio de 25,03, resultando em estatística F de 1,75. Esse valor, inferior ao F crítico para o nível de significância usual, revela que não foi possível rejeitar a hipótese nula de igualdade de médias entre ChatGPT, Gemini, Copilot e DeepSeek. Embora as análises financeiras anteriores tenham demonstrado diferenças expressivas, como o superior desempenho absoluto e ajustado ao risco da Gemini Sem Alimentação, a ANOVA sugere que, estatisticamente, essas diferenças não podem ser consideradas significativas dentro do delineamento disponível. O resultado deriva do fato de que, com apenas uma observação por combinação de fatores, grande parte da variabilidade é absorvida no termo residual, reduzindo a sensibilidade do teste.

O fator Alimentação de dados também não apresentou significância estatística. Sua soma de quadrados foi praticamente nula (0,00005), gerando quadrado médio igualmente ínfimo e estatística F igual a zero. Isso indica que, considerando apenas os valores médios, as versões com e sem feed não diferem estatisticamente em termos de retorno. No entanto, essa interpretação exige cautela: as análises qualitativas e financeiras apontaram efeitos pronunciados e assimétricos da alimentação entre os modelos. O feed elevou significativamente a performance do ChatGPT (como evidenciado pelos seus indicadores de retorno, Sharpe e alfa), mas deteriorou a eficiência do Copilot e do DeepSeek. A ANOVA, ao agrupar tais efeitos contrastantes em um único fator, dilui suas diferenças, resultando em ausência de significância estatística.

Outro ponto metodológico relevante é a impossibilidade de estimar a interação IA × Alimentação. Como o delineamento não possui replicações (apenas uma carteira por IA e por condição), o termo de interação não pode ser isolado da variabilidade residual. Essa limitação impede avaliar estatisticamente se o efeito do feed depende da IA, justamente um dos achados mais importantes deste estudo. As evidências financeiras mostraram que a alimentação melhora de forma substancial algumas IAs (como o ChatGPT) e piora outras (como Copilot e DeepSeek). Contudo, do ponto de vista estatístico, essa assimetria não pode ser testada no presente delineamento. Essa restrição representa uma oportunidade clara para trabalhos futuros: a geração de múltiplas carteiras por IA e por condição

permitiria a estimação rigorosa da interação, ampliaria o poder estatístico e possibilitaria conclusões mais precisas sobre o papel da alimentação de dados.

No conjunto, a ANOVA indica que nenhum dos fatores testados apresentou significância estatística. Isso não contradiz as análises financeiras, que identificaram diferenças expressivas entre as carteiras, mas revela que o número reduzido de observações limita a capacidade do método estatístico de capturar tais diferenças. Em outras palavras, há significância econômica clara nas diferenças, refletida nos retornos, no Índice Sharpe e no Alfa de Jensen, mas não há significância estatística no delineamento atual.

Assim, a análise estatística cumpre papel complementar: fornece evidências de que as conclusões financeiras encontradas neste trabalho derivam de padrões reais observados nas carteiras, mas que não podem ser confirmados formalmente em testes estatísticos devido à limitação estrutural da amostra. A expansão do número de portfólios analisados é, portanto, o passo natural para pesquisas futuras, permitindo avaliação mais robusta e modelagem apropriada da interação entre arquitetura das IAs e alimentação de dados estruturados.

6 CONCLUSÕES

A partir das análises realizadas, constatou-se que, embora as IAs tenham apresentado rentabilidades competitivas e, em alguns casos, superiores ao CDI no intervalo considerado, os resultados ainda não são suficientes para caracterizar o uso de LLMs como substitutos robustos de processos tradicionais de análise para investidores de varejo. A conclusão decorre tanto do comportamento das carteiras quanto das limitações metodológicas inerentes ao experimento.

Em primeiro lugar, a performance observada deve ser interpretada à luz do contexto macroeconômico favorável. A expressiva valorização do Ibovespa no período, impulsuada por fatores como entrada de capital estrangeiro, estabilidade inflacionária e apreciação de commodities-chave (minério de ferro e petróleo), exerceu influência direta sobre o resultado das carteiras. Como grande parte das IAs, especialmente nas versões com alimentação de dados, convergiu para escolhas concentradas em setores resilientes, como mineração, petróleo, energia elétrica e bancos, a alta do índice atuou como vetor determinante de performance. Em outras palavras, a capacidade das carteiras de superar o CDI está fortemente atrelada ao fato de que o período analisado beneficiou justamente os setores mais selecionados pelas IAs, o que limita generalizações sobre sua eficiência em diferentes regimes de mercado, especialmente em períodos de queda, alta volatilidade ou ciclo de juros distinto.

Outro ponto central refere-se à própria construção do experimento. A elaboração de apenas oito carteiras reduz substancialmente a variabilidade da amostra e compromete o poder estatístico dos testes realizados, inclusive da ANOVA de dois fatores. Sem múltiplas repetições por IA e por condição de alimentação de dados, torna-se impossível estimar adequadamente o termo de interação e capturar nuances de comportamento que poderiam emergir em universos maiores. Uma pesquisa mais robusta requereria dezenas de carteiras por IA, variações sistemáticas de prompts, múltiplas janelas temporais e replicações sob diferentes condições de mercado. Assim, ainda que os resultados obtidos sejam indicativos, eles não possuem amplitude suficiente para conclusões definitivas.

Adicionalmente, a própria formulação dos prompts representa um componente crítico da análise. Embora o objetivo deste estudo tenha sido avaliar o uso das IAs sob a ótica do investidor de varejo, com prompts simples e informações acessíveis ao público geral, permanece a questão sobre se prompts mais sofisticados teriam produzido carteiras mais consistentes. Entretanto, exigir do investidor comum o mesmo nível de detalhamento empregado por profissionais institucionais contraria o propósito destas ferramentas para o público geral. Nesse sentido, os resultados obtidos refletem com maior fidelidade o que investidores comuns provavelmente alcançariam ao utilizar essas tecnologias no seu estágio atual de maturidade.

No intuito de contextualizar o desempenho das IAs, comparou-se as rentabilidades das carteiras com fundos profissionais de referência. Entre agosto e outubro de 2025, fundos como XP Investor Ibovespa Ativo (14,35%), Kinea Gama (15,08%), Absolute Endurance FIA (12,82%) e SPX Long Bias Prev (11,23%) apresentaram resultados próximos às melhores carteiras de IA. Ainda que nenhum dos modelos tenha reproduzido desempenhos excepcionais como o do Pátira Pioneiro Long Only (34,97%), observa-se que IAs como Gemini e ChatGPT com alimentação de dados atingiram patamares compatíveis com gestores tradicionais de alto nível. Esse resultado é significativo, pois sugere que, mesmo com limitações de metodologia, amostragem e janela temporal, as IAs já conseguem gerar portfólios com qualidade comparável à de fundos consolidados do mercado brasileiro no curto prazo.

Por outro lado, o fato de algumas carteiras terem ficado próximas do desempenho de gestores profissionais não implica que as IAs estejam prontas para substituir processos de análise humana ou modelos quantitativos estruturados. A ausência de consistência estatística, a sensibilidade a pequenas mudanças nos prompts, a dependência do regime macroeconômico e a dificuldade de interpretar risco de forma sistemática são limitações que restringem o uso dessas ferramentas por investidores de varejo. O desempenho observado pode ser explicado mais pela conjuntura favorável do mercado do que por uma habilidade explícita dos modelos em identificar relações risco-retorno superiores.

Assim, os resultados sugerem que as IAs têm potencial como instrumentos complementares, sobretudo em ambientes institucionais. Investidores profissionais, que dispõem de expertise para interpretar, validar e ajustar as sugestões dos modelos, podem se beneficiar significativamente do uso de LLMs como suporte analítico, fonte de geração de hipóteses ou ferramenta de otimização. Já para o investidor de varejo, que não dispõe da mesma capacidade técnica, o uso das IAs ainda deve ser visto com cautela, uma vez que decisões de investimento demandam profundidade interpretativa que os modelos não são

capazes de garantir de forma autônoma.

Por fim, este trabalho abre espaço para pesquisas futuras mais amplas e metodologicamente rigorosas. Recomenda-se a construção de um número maior de carteiras por IA, a realização de testes em diferentes regimes de mercado (alta, baixa e lateralização), a experimentação com prompts de complexidade variável e a comparação mais extensa com benchmarks institucionais. Estudos complementares poderiam ainda investigar métricas adicionais de risco, como drawdown, tracking error e Information Ratio, ou explorar abordagens híbridas que combinem IA generativa com modelos quantitativos clássicos.

Em suma, conclui-se que, embora as IAs tenham apresentado desempenho competitivo e, em algumas condições, superior ao CDI, ainda não estão maduras para serem utilizadas como assessoras de investimento para o público de varejo. Seu melhor uso, no momento, reside no suporte a investidores institucionais e equipes de research que possuam conhecimento técnico suficiente para integrar as recomendações das IAs dentro de arcabouços de decisão mais robustos. O caminho para uma adoção ampla e segura ainda requer maior maturidade tecnológica, evolução metodológica e pesquisas adicionais que consolidem evidências em múltiplos cenários de mercado.

REFERÊNCIAS

- [1] BANCO CENTRAL DO BRASIL (BCB). Sistema Financeiro Nacional: composição e regulação. Brasília, 2024. Disponível em: <https://www.bcb.gov.br/estabilidedefinanceira/composicaosfn>.
- [2] BARTONCELLO, M. G. Mercado de capitais: notas sobre a importância do mercado de capitais brasileiro. Foz do Iguaçu: UNILA, 2022.
- [3] B3 – Brasil, Bolsa, Balcão. Metodologia do Índice Brasil 50 (IBRX 50). São Paulo: B3, 2024.
- [4] B3 – Brasil, Bolsa, Balcão. Metodologia do Índice Brasil 100 (IBRX 100). São Paulo: B3, 2024.
- [5] BROWN, T. et al. Language Models Are Few-Shot Learners. *Advances in Neural Information Processing Systems*, v. 33, p. 1877–1901, 2020.
- [6] CARLUCCI, A. G. Análise Temporal do Índice Small Cap (SMALL). São Paulo: Universidade Presbiteriana Mackenzie, 2021.
- [7] CASACCIA, M. C.; GALLI, O. C.; MACÊDO, G. R.; LEITÃO, C. Análise do desempenho de fundos de investimentos: um estudo em ações brasileiras no período de janeiro de 2004 a agosto de 2009. *Revista Organizações em Contexto*, v. 7, n. 13, p. 1–30, 2011.
- [8] CASACCIA, M. C. Análise do desempenho de fundos de investimento em ações no Brasil. Universidade de São Paulo, 2011.
- [9] COMISSÃO DE VALORES MOBILIÁRIOS (CVM). Mercado de valores mobiliários. Brasília, 2024. Disponível em: <https://www.gov.br/cvm/pt-br/assuntos/mercado-de-valores-mobiliarios>.
- [10] COOKSEY, R. Descriptive Statistics for Summarising Data. In: *Practical Data Analysis: Case Studies in Business Statistics*. Singapore: Springer, 2020.
- [11] DAVIS, J.; HAGELSTEIN, P.; LACKNER, I.; PIZIAK, R. The Efficient Frontier and International Portfolio Diversification. *Journal of Financial and Investment Analysis*, v. 2, n. 3, p. 25–45, 2013.
- [12] EL KAROUI, N. High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: Risk underestimation. arXiv preprint, 2012.
- [13] ELTON, E. J.; GRUBER, M. J. Modern Portfolio Theory, 1950 to Date. *Journal of Banking Finance*, Amsterdam, v. 21, n. 11, p. 1743–1759, 1997.
- [14] GOMES, F. A. R. Avaliação do desempenho dos fundos long-short no Brasil. *Revista de Administração Contemporânea*, v. 15, n. 4, p. 580–602, 2011.

- [15] GOMES, M. P. Testes empíricos do modelo intertemporal de precificação de ativos para dados brasileiros. *Revista Brasileira de Finanças*, v. 9, n. 4, p. 469–503, 2011.
- [16] INFOMONEY. Brasileiro ingressa na bolsa com investimento médio de R\$ 352,50, tem de 25 a 39 anos e mais de 5 ativos na carteira. 2024. Disponível em: <https://www.infomoney.com.br/onde-investir/brasileiro-ingressa-na-bolsa-com-investimento-medio-de-r-352-50-tem-de-25-a-39-anos-e-mais-de-5-ativos-na-carteira/>. Acesso em: 22 fev. 2025.
- [17] INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). Estimativas populacionais. Brasília, 2024. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao.html>.
- [18] KONDOR, I.; PAFKA, S.; NAGY, G. Noise Sensitivity of Portfolio Selection under Various Risk Measures. arXiv preprint, 2006.
- [19] KASHEFI, J.; MCKEE, G. Efficient Frontier: Global Equity Markets and International Diversification. *Clute Institute Journals*, v. 15, n. 2, p. 61–78, 2001.
- [20] KONG, Y. Large Language Models for Financial and Investment Applications. MIT Media Lab, 2024.
- [21] LEDOIT, O.; WOLF, M. Honey, I Shrunk the Sample Covariance Matrix. Zurich: University of Zurich Working Paper, 2003.
- [22] LIU, J. et al. A Comprehensive Overview of Prompt Engineering Techniques for Large Language Models. 2023.
- [23] MACHADO, M. A. V.; SANTOS, A. C.; FAMÁ, R. Inter-temporal CAPM: An Empirical Test with Brazilian Equity Market. *Revista Contabilidade Finanças*, v. 24, n. 62, p. 44–56, 2013.
- [24] MARTIN, R. Fixed income portfolio optimisation: Interest rates, credit, and the efficient frontier. arXiv preprint, 2020.
- [25] MEHRA, R.; PRESCOTT, E. C. The Equity Premium in Retrospect. NBER Working Paper, 2003.
- [26] MERTON, R. C. An Intertemporal Capital Asset Pricing Model. *Econometrica*, v. 41, n. 5, p. 867–887, 1973.
- [27] MONTGOMERY, D. C.; RUNGER, G. C. *Estatística Aplicada e Probabilidade para Engenheiros*. 4. ed. Rio de Janeiro: LTC, 2014.
- [28] MONTGOMERY, D. C. *Introdução ao Controle Estatístico da Qualidade*. Rio de Janeiro: LTC, 2000.
- [29] NBER. A Sharper Ratio: A General Measure for Correctly Ranking Portfolios. NBER Working Paper Series, n. 19500, 2014.
- [30] NIE, W. et al. A multi-agent framework for quantitative finance. EMNLP Industry Track, 2025.

- [31] OLIVEIRA, G. G. Carteiras de Variância Mínima no Mercado de Ações. Encontro Nacional da ANPAD, 2010.
- [32] OLIVEIRA, B. G. de; SOUSA, A. F. de. Fundos de Investimento em Ações no Brasil: Métricas para Avaliação de Desempenho. REGE/USP, v. 22, n. 1, p. 61–76, 2015.
- [33] PRESENTE, R. Mercados financeiros. Fundação Cecierj, 2019.
- [34] RAM ACTIVE INVESTMENTS. Financial Sentiment Analysis with Large Language Models. Lausanne, 2024.
- [35] REZENDE, F. A. Um Estudo sobre o Desempenho Financeiro do Índice Bovespa. Revista Gestão Tecnologia, v. 13, n. 1, p. 45–59, 2013.
- [36] ROGERS, P. et al. Comparações múltiplas das carteiras de bolsa no Brasil. REGE/USP, v. 12, n. 3, p. 77–94, 2005.
- [37] ROQUETE, R. M. Indexação fundamental no Brasil: aplicação e implicações. RBGN, v. 21, n. 2, p. 260–278, 2019.
- [38] SCHOENMAKER, D.; SCHRAMADE, W. Risk-Return Analysis. In: Corporate Finance for Long-Term Value. Cham: Springer, 2023.
- [39] SIMPLÍCIO, E. E. de O. Análise do desempenho dos índices setoriais da B3 e sua relação com a taxa básica de juros no Brasil. UFPB, 2024.
- [40] TIAN, S.; GUO, Y. Fine-tuning large language models for stock return prediction using newsflow. arXiv, 2024.
- [41] VALENTE, M. P. N. Modelo Greenblatt: Aplicação e Comparação. Anais do SEGeT, v. 17, p. 1–15, 2020.
- [42] WANG, X. et al. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. NeurIPS, v. 35, 2022.
- [43] WILLIAM, R. Two-Way Analysis of Variance. University of Notre Dame, 2004.
- [44] XP INVESTIMENTOS. B3: Número de investidores de renda fixa, variável e FIIs – dezembro de 2023. 2023. Disponível em: <https://conteudos.xpi.com.br/conteudos-gerais/b3-investidores-renda-fixa-variavel-fiis-dez-2023/>. Acesso em: 22 fev. 2025.
- [45] YANG, L.; COUILLET, R.; MCKAY, M. A Robust Statistics Approach to Minimum Variance Portfolio Optimization. arXiv preprint, 2015.
- [46] YOSHINAGA, C. E.; CASTRO JR., F. H. F. The Relationship between Market Sentiment Index and Stock Rates of Return: a Panel Data Analysis. BAR, v. 9, n. 2, p. 189–210, 2012.
- [47] ZHANG, H. Limitations and Critique of Modern Portfolio Theory: A Comprehensive Literature Review. SSRN/ResearchGate Preprint, 2023.

APÊNDICE A

Tabela 22: Empresas listadas na B3

Empresa	BBG Ticker
Vale S.A.	VALE3 BZ Equity
Petrobras PN	PETR4 BZ Equity
Petrobras ON	PETR3 BZ Equity
Itaú Unibanco PN	ITUB4 BZ Equity
Banco do Brasil ON	BBAS3 BZ Equity
WEG ON	WEGE3 BZ Equity
Eletrobras ON	ELET3 BZ Equity
Sabesp ON	SBSP3 BZ Equity
Bradesco PN	BBDC4 BZ Equity
B3 ON	B3SA3 BZ Equity
Ambev ON	ABEV3 BZ Equity
Itaúsa PN	ITSA4 BZ Equity
JBS ON	JBSS3 BZ Equity
Embraer ON	EMBR3 BZ Equity
Suzano ON	SUZB3 BZ Equity
BTG Pactual Unit	BPAC1 BZ Equity
Equatorial Energia ON	EQTL3 BZ Equity
PetroRio ON	PRI03 BZ Equity
Localiza ON	RENT3 BZ Equity
Rede D'Or ON	RDOR3 BZ Equity
Raia Drogasil ON	RADL3 BZ Equity
BB Seguridade ON	BBSE3 BZ Equity
Gerdau PN	GGBR4 BZ Equity
Rumo ON	RAIL3 BZ Equity
BRF ON	BRFS3 BZ Equity
Cemig PN	CMIG4 BZ Equity
Vivo ON	VIVT3 BZ Equity

Empresa	BBG Ticker
Eneva ON	ENEV3 BZ Equity
Vibra Energia ON	VBBR3 BZ Equity
Klabin Unit	KLBN1 BZ Equity
Ultrapar ON	UGPA3 BZ Equity
Bradesco ON	BBDC3 BZ Equity
Copel PNB	CPLE6 BZ Equity
Totvs ON	TOTS3 BZ Equity
Lojas Renner ON	LREN3 BZ Equity
Energisa Unit	ENGI1 BZ Equity
TIM ON	TIMS3 BZ Equity
CCR ON	CCRO3 BZ Equity
Cosan ON	CSAN3 BZ Equity
Hapvida ON	HAPV3 BZ Equity
Eletrobras PNB	ELET6 BZ Equity
CSN ON	CSNA3 BZ Equity
Fleury ON	FLRY3 BZ Equity
Marfrig ON	MRFG3 BZ Equity
Cyrela ON	CYRE3 BZ Equity
Minerva ON	BEEF3 BZ Equity
Magazine Luiza ON	MGLU3 BZ Equity
Cogna ON	COGN3 BZ Equity
MRV ON	MRVE3 BZ Equity
LWSA ON	LWSA3 BZ Equity
Petz ON	PETZ3 BZ Equity
Azul PN	AZUL4 BZ Equity
Alpargatas PN	ALPA4 BZ Equity
EZTEC ON	EZTC3 BZ Equity
Pão de Açúcar ON	PCAR3 BZ Equity
CVC ON	CVCB3 BZ Equity
Americanas ON	AMER3 BZ Equity
BR Malls ON	BRML3 BZ Equity
BR Properties ON	BRPR3 BZ Equity
Ecorodovias ON	ECOR3 BZ Equity
EDP Brasil ON	ENBR3 BZ Equity

Empresa	BBG Ticker
Gol PN	GOLL4 BZ Equity
Grupo Soma ON	SOMA3 BZ Equity
Iguatemi ON	IGTA3 BZ Equity
NotreDame Intermédica ON	GNDI3 BZ Equity
Locamérica ON	LCAM3 BZ Equity
Lojas Americanas PN	LAME4 BZ Equity
Positivo Tecnologia ON	POSI3 BZ Equity
Qualicorp ON	QUAL3 BZ Equity
Rumo ON	RUMO3 BZ Equity
Sanepar Unit	SAPR1 BZ Equity
Suzano PNA	SUZB5 BZ Equity
Vivo PN	VIVT4 BZ Equity
Vale PNA	VALE5 BZ Equity