



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA
COMPUTAÇÃO

KÉTHLYN CAMPOS SILVA

**Otimização de Portfólio de Ativos do
Mercado Financeiro Brasileiro:
Integrando Notícias e Indicadores
Fundamentalistas com Aprendizado por
Reforço Profundo**

Goiânia
2025



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE INFORMÁTICA

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a [Lei 9.610/98](#), o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

Dissertação Tese Outro*: _____

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

Kéthlyn Campos Silva

3. Título do trabalho

Otimização de Portfólio de Ativos do Mercado Financeiro Brasileiro: Integrando Notícias e Indicadores Fundamentalistas com Aprendizado por Reforço Profundo

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento SIM NÃO¹

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:

- a) consulta ao(à) autor(a) e ao(à) orientador(a);
 - b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.
- O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.



Documento assinado eletronicamente por **Deborah Silva Alves Fernandes, Vice-Diretora**, em 27/01/2026, às 18:25, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Kéthlyn Campos Silva, Discente**, em 28/01/2026, às 13:15, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5916938** e o código CRC **3F775050**.

KÉTHLYN CAMPOS SILVA

Otimização de Portfólio de Ativos do Mercado Financeiro Brasileiro: Integrando Notícias e Indicadores Fundamentalistas com Aprendizado por Reforço Profundo

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Informática da Universidade Federal de Goiás, como requisito parcial para obtenção do título de Mestra em Ciência da Computação.

Área de concentração: Ciência da Computação.

Orientadora: Profa. Deborah Silva Alves Fernandes

Goiânia
2025

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Silva, Kéthlyn Campos
Otimização de Portfólio de Ativos do Mercado Financeiro Brasileiro:
Integrando Notícias e Indicadores Fundamentalistas com Aprendizado por Reforço Profundo [Manuscrito] / Kéthlyn Campos Silva. - 2025.
98 f. : 2025

Orientadora: Prof(a). Dra. Deborah Silva Alves Fernandes
Dissertação (Mestrado) - Universidade Federal de Goiás, Instituto de Informática (INF), Programa de Pós-Graduação em Ciência da Computação, Goiânia, 2025.

Bibliografia.
Inclui: siglas, lista de figuras, lista de tabelas.

1. Aprendizado por Reforço Profundo. 2. Mercado Financeiro Brasileiro. 3. Processamento de Linguagem Natural. 4. Análise Fundamentalista. 5. Grandes Modelos de Linguagem.

I. Fernandes, Deborah Silva Alves, orient. II. Título.

CDU 004



UNIVERSIDADE FEDERAL DE GOIÁS

INSTITUTO DE INFORMÁTICA

ATA DE DEFESA DE DISSERTAÇÃO

Ata nº 18 da sessão de Defesa de Dissertação de **Kéthlyn Campos Silva**, que confere o título de Mestra em Ciência da Computação, na área de concentração em Ciência da Computação.

Aos dezesseis dias do mês de setembro de dois mil e vinte e cinco, a partir das catorze horas, via sistema de webconferência, realizou-se a sessão pública de Defesa de Dissertação intitulada “**Otimização de Portfólio de Ativos do Mercado Financeiro Brasileiro: Integrando Notícias e Indicadores Fundamentalistas com Aprendizado por Reforço Profundo**”. Os trabalhos foram instalados pela Orientadora, Professora Doutora Deborah Silva Alves Fernandes (INF/UFG) com a participação dos demais membros da Banca Examinadora: Professor Doutor Fabrizzio Alphonsus Alves de Melo Nunes Soares (INF/UFG), membro titular interno; Professor Doutor Sandrerley Ramos Pires (EMC/UFG), membro titular externo. A realização da banca ocorreu por meio de videoconferência. Durante a arguição os membros da banca não fizeram sugestão de alteração do título do trabalho. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Dissertação, tendo sido a candidata **aprovada** pelos seus membros. Proclamados os resultados pela Professora Doutora Deborah Silva Alves Fernandes, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, aos dezesseis dias do mês de setembro de dois mil e vinte e cinco.

TÍTULO SUGERIDO PELA BANCA



Documento assinado eletronicamente por **Deborah Silva Alves Fernandes, Professor do Magistério Superior**, em 16/09/2025, às 15:50, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Fabrizzio Alphonsus Alves De Melo Nunes Soares, Professor do Magistério Superior**, em 16/09/2025, às 15:51, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Sandrerley Ramos Pires, Professor do Magistério Superior**, em 16/09/2025, às 15:53, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Kéthlyn Campos Silva, Discente**, em 22/09/2025, às 19:45, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5607221** e o código CRC **4D834670**.

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador(a).

Kéthlyn Campos Silva

Graduou-se em Engenharia de Computação na UFG - Universidade Federal de Goiás. Durante o Mestrado, na UFG, foi tutora da especialização em Processamento de Linguagem Natural do AKCIT (*Advanced Knowledge Center for Immersive Technologies*), associado ao Centro de Excelência em Inteligência Artificial (CEIA). Atualmente é Engenheira de Inteligência Artificial, desenvolvendo soluções baseadas em IA Generativa.

Dedico este trabalho aos meus pais, Eliane e José, ao meu amor, Felipe, e ao meu irmão, Kristhyan, pelo amor incondicional, apoio constante e incentivo em cada etapa da minha jornada acadêmica. Sem vocês, nada disso teria sido possível.

Agradecimentos

Agradeço primeiramente a Deus, que sempre me abençoou, permitindo-me chegar até este momento e me dando força para superar os desafios desta jornada. À minha orientadora, Deborah, que tem me acompanhado desde a graduação, devo grande parte das minhas conquistas pessoais e profissionais ao momento em que me escolheu como sua orientanda na iniciação científica. Refletindo sobre os pontos de inflexão da minha vida, considero este um dos mais importantes.

Ao professor Márcio Giovane, por sua coorientação, ainda que indireta, que contribuiu para o desenvolvimento deste trabalho. Aos professores Sandrerley e Fabrízio, pelo apoio e pelas valiosas contribuições ao avaliar meu trabalho. À Universidade Federal de Goiás, que me proporcionou uma formação acadêmica sólida desde a graduação.

Aos meus pais, Eliane e José, por sempre me darem o exemplo, incentivarem meus estudos e minha independência; sua inteligência, dedicação e persistência são minha inspiração. Ao meu noivo, Felipe, que foi meu porto seguro e companheiro durante toda essa trajetória. À minha sogra, Elisangela, pelo apoio constante, seja preparando uma refeição, cuidando de mim ou simplesmente me ouvindo nos momentos de desabafo.

Aos outros é dado o direito de desistir; a você, não!

José Pereira da Silva,
meu pai.

Resumo

Silva, K ethlyn Campos. **Otimiza o de Portf lio de Ativos do Mercado Financeiro Brasileiro: Integrando Not cias e Indicadores Fundamentalistas com Aprendizado por Refor o Profundo**. Goi nia, 2025. 98p. Disserta o de Mestrado. Instituto de Inform tica, Universidade Federal de Goi s.

Este estudo investiga o impacto da incorpora o de indicadores fundamentalistas e de sentimento extra dos de not cias em l ngua portuguesa na aplica o de algoritmos de aprendizagem por refor o profundo (ARP) para a otimiza o de portf lios no mercado financeiro brasileiro. A pesquisa abrange a coleta de not cias, dados hist ricos e indicadores financeiros de ativos, com extra o de sentimento e entidades-chave a partir de artigos jornal sticos por meio do *Large Language Model Gemini Pro*. Para refinar os indicadores de sentimento, s o aplicadas t cnicas de filtragem baseadas em entidades e t picos, com o objetivo de reduzir o ru do informacional. An lises estat sticas com coeficientes de correla o evidenciam que a filtragem por entidades contribui para o aumento da correla o entre o sentimento e os retornos di rios dos ativos. Em seguida, os indicadores de sentimento e os fundamentalistas s o integrados a cinco algoritmos de ARP, testados em quatro cen rios distintos: (1) apenas pre os, (2) pre os e sentimento, (3) pre os e indicadores fundamentalistas, e (4) todos os elementos combinados. Foram geradas 44 amostras no total. O teste de Kruskal-Wallis n o apontou diferen as estatisticamente significativas entre os cen rios, no entanto observou-se que todos os modelos baseados em ARP superaram as estrat gias *Buy and Hold* e o  ndice Ibovespa em termos de Sharpe Ratio, indicando maior retorno ajustado ao risco. Esses resultados sugerem que a utiliza o de ARP pode contribuir para um controle de risco mais eficaz em portf lios compostos por ativos do mercado financeiro brasileiro.

Palavras-chave

Aprendizado por Refor o Profundo, Mercado Financeiro Brasileiro, Processamento de Linguagem Natural, An lise Fundamentalista e Grandes Modelos de Linguagem.

Abstract

Silva, Kéthlyn Campos. **Portfolio Optimization of Brazilian Financial Market Assets: Integrating News and Fundamental Indicators with Deep Reinforcement Learning**. Goiânia, 2025. 98p. MSc. Dissertation. Insituto de Informática, Universidade Federal de Goiás.

This study investigates the impact of incorporating fundamental and sentiment-based indicators, extracted from Portuguese news, into deep reinforcement learning (DRL) algorithms for portfolio optimization in the Brazilian financial market. The research involves collecting news articles, historical data, and financial indicators of assets, with sentiment and key entities extracted from the news using the Gemini Pro Large Language Model. To refine sentiment indicators, entity- and topic-based filtering techniques are applied to reduce informational noise. Statistical analyses using correlation coefficients show that entity-based filtering enhances the relationship between sentiment indicators and daily asset returns. Subsequently, sentiment and fundamental indicators are integrated into five DRL algorithms, tested across four distinct scenarios: (1) prices only, (2) prices and sentiment, (3) prices and fundamental indicators, and (4) all combined. A total of 44 samples were generated. Although the Kruskal-Wallis test did not reveal statistically significant differences between the scenarios, all DRL-based models outperformed baseline strategies such as Buy and Hold and the Ibovespa index in terms of Sharpe Ratio, indicating higher returns with better risk control. These findings suggest that the use of DRL algorithms can contribute to more effective risk management in portfolios composed of Brazilian financial market assets.

Keywords

Deep Reinforcement Learning, Brazilian Financial Market, Natural Language Processing, Fundamental Analysis and Large Language Models

A2C *Advantage Actor-Critic*
A3C *Asynchronous Advantage Actor-Critic*
AR *Aprendizado por Reforço*
ARP *Aprendizado por Reforço Profundo*
BAH *Buy and Hold*
COMPSAC *IEEE 49th Annual Computers, Software, and Applications Conference*
CSCI *International Conference on Computer Science and Computational Intelligence*
DL *Deep Learning*
DQN *Deep Q-Learning*
DDPG *Deep Deterministic Policy Gradient*
EMASI *Exponential Moving Average Sentiment Indicator*
HDBSCAN *Hierarchical Density-Based Spatial Clustering of Applications with Noise*
IA *Inteligência Artificial*
IBOV *Índice Ibovespa*
LLM *Large Language Model*
MA *Moving Average*
MASI *Moving Average Sentiment Indicator*
MD *Maximum Drawdown*
MDP *Markov Decision Process*
ML *Machine Learning*
PG *Policy Gradient*
PLN *Processamento de Linguagem Natural*
PPO *Proximal Policy Optimization*
RA *Retorno Anual*
RAG *Retrieval-Augmented Generation*
RSL *Revisão Sistemática da Literatura*
SAC *Soft Actor-Critic*
SR *Sharpe Ratio*
TD3 *Twin-Delayed Deep Deterministic Policy Gradient*
UMAP *Uniform Manifold Approximation and Projection for Dimension Reduction*

Sumário

Lista de Figuras	15
Lista de Tabelas	16
1 Introdução	17
1.1 Introdução	17
1.2 Problema	20
1.3 Objetivos e Hipótese	20
1.3.1 Geral	20
1.3.2 Específicos	21
1.3.3 Hipótese	21
1.4 Estrutura do trabalho	21
2 Fundamentação Teórica	22
2.1 Aprendizado por Reforço Profundo	22
2.1.1 Métodos baseados em função de valor	24
2.1.2 Métodos baseados em busca de política	26
2.1.3 Métodos ator-crítico	26
<i>Advantage Actor-Critic (A2C)</i>	27
<i>Proximal Policy Optimization (PPO)</i>	27
<i>Deep Deterministic Policy Gradient (DDPG)</i>	28
<i>Twin Delayed Deep Deterministic Policy Gradient (TD3)</i>	30
<i>Soft Actor-Critic (SAC)</i>	31
2.2 Análise fundamentalista	32
2.3 Embeddings	33
2.4 <i>Prompt Engineering</i>	35
2.4.1 Elementos de um prompt	35
2.4.2 Tipos de prompts	36
2.4.3 Técnicas	36
<i>Zero-shot learning</i>	36
<i>Few-shot learning</i>	37
<i>Retrieval Augmented Generation - RAG</i>	37
2.5 Modelagem de tópicos	39
3 Revisão da Literatura acerca da Integração de Aprendizado por Reforço Profundo com Indicadores baseados em Notícias Financeiras e Indicadores Fundamentalistas	41
3.1 Revisão Sistemática de Literatura	41
3.2 Protocolo de planejamento da RSL	42

3.3	Execução	45
3.4	Resultados da RSL	46
3.4.1	Análise dos artigos selecionados	47
	ARP	48
	Indicadores a partir de notícias	48
	Indicadores fundamentalistas	50
	ARP com indicadores baseados em notícias	50
	ARP com indicadores fundamentalistas	50
	Posicionamento da pesquisa	51
4	Integração de Aprendizado por Reforço Profundo com Indicadores baseados em Notícias Financeiras e Indicadores Fundamentalistas	52
4.1	Desenho do Experimento	52
4.1.1	Aquisição de dados	52
	Notícias financeiras	54
	B3	55
	Indicadores fundamentalistas	55
4.1.2	Extração de informações de notícias	56
	Prompt Engineering	56
	Modelagem de tópicos	57
4.1.3	Indicadores baseados em notícias	59
	Indicador de sentimento	59
	Indicador de Sentimento por Média Móvel	59
	Indicador de Sentimento por Média Móvel Exponencial	60
4.1.4	Filtros de artigos de notícias	60
4.1.5	Análise estatística dos indicadores de sentimento	62
4.1.6	Aprendizado por Reforço Profundo	65
	Ambiente de ARP	65
	Otimização de hiperparâmetros	67
4.1.7	Análise estatística dos resultados experimentais	68
5	Seleção de Indicador de Sentimento	72
5.1	Análise de Correlação	72
5.2	Relacionamento de notícias a cada entidade	74
6	Resultados e Discussões	76
6.1	Resultados	76
6.1.1	<i>Sharpe Ratio</i>	77
6.1.2	<i>Maximum Drawdown</i>	80
6.1.3	Retorno Anual	83
6.2	Discussões	85
7	Conclusão	87
	Referências Bibliográficas	88

Lista de Figuras

1.1	Fluxograma que descreve o problema de otimização de portfólio.	19
2.1	Estrutura de um algoritmo de ARP padrão [Yi e Liu 2023].	25
2.2	Exemplo de relação semântica capturada por embeddings [Google Cloud 2020].	34
2.3	Exemplo de resposta aplicando <i>Zero-shot learning</i> .	37
2.4	Exemplo de resposta aplicando <i>Few-shot learning</i> .	38
2.5	Estrutura da técnica <i>Retrieval Augmented Generation</i> .	39
3.1	Fases e passos da Revisão Sistemática da Literatura (Adaptado de [Nakagawa et al. 2017]).	43
3.2	Principais temas relacionados ao objetivo da Revisão Sistemática da Literatura.	44
3.3	Distribuição de publicações dos periódicos.	46
3.4	Distribuição de publicações das conferências.	47
4.1	Desenho do experimento.	53
4.2	Nuvem de palavras a partir das notícias coletadas.	54
4.3	Espaço vetorial dos <i>embeddings</i> das notícias.	57
4.4	Amostra do Indicador de Sentimento, MASÍ e EMASÍ ao longo do tempo.	61
4.5	Fluxo de geração dos indicadores de sentimento.	62
4.6	Distribuição de notícias por tópico.	63
5.1	Distribuição de notícias relativas a cada entidade.	75
5.2	Quantidade de notícias das ações selecionadas.	75
6.1	Distribuições do Sharpe Ratio: Análise por Algoritmo e Cenário.	77
6.2	Análise das distribuições do Sharpe Ratio.	78
6.3	Distribuições do Maximum Drawdown: Análise por Algoritmo e Cenário.	80
6.4	Análise das distribuições do Maximum Drawdown.	81
6.5	Distribuições do Retorno Anual: Análise por Algoritmo e Cenário.	83
6.6	Análise das distribuições do Retorno Anual.	84

Lista de Tabelas

3.1	Número de artigos selecionados com base em critérios de exclusão.	45
3.2	Sumário dos 11 artigos mais citados.	47
3.3	Sumário dos artigos selecionados.	49
4.1	Refinamentos do prompt	58
4.2	Escala de Cohen	65
4.3	Variáveis utilizadas em cada cenário	67
4.4	Hiperparâmetros dos algoritmos por cenário	69
4.5	P-valores para diferentes modelos em vários cenários e métricas de avaliação.	71
5.1	Número de notícias com base no filtro aplicado	72
5.2	Coefficientes de correlação de Spearman de cada indicador de sentimento	73
6.1	Métricas de desempenho das bases de comparação	77
6.2	Medidas estatísticas do Sharpe Ratio de cada algoritmo e cenário	79
6.3	Resultados do teste H de Kruskal-Wallis para o Sharpe Ratio.	80
6.4	Medidas estatísticas do Maximum Drawdown de cada algoritmo e cenário	82
6.5	Resultados do teste H de Kruskal-Wallis para o Maximum Drawdown.	82
6.6	Medidas estatísticas do Retorno Anual de cada algoritmo e cenário	85
6.7	Resultados do teste H de Kruskal-Wallis para o Retorno Anual.	85

Introdução

1.1 Introdução

O mercado de capitais desempenha um papel fundamental no desenvolvimento econômico de um país, conectando aqueles que têm capital e procuram lucro (os investidores) com aqueles que precisam de recursos (os tomadores de recursos). Esta dinâmica permite às empresas privadas e aos governos angariar fundos para projetos de investimento, enquanto os investidores beneficiam dos retornos dos seus investimentos [Lopes, Antunes e Cardoso 2007]. Essa troca é facilitada pelas bolsas de valores. Considerado a oitava maior economia do mundo em 2024, o Brasil segue esse modelo, com a B3 (Brasil, Bolsa, Balcão) servindo como a principal bolsa de valores para transações entre investidores e tomadores de recursos [International Monetary Fund 2021].

Uma estratégia amplamente implementada entre os investidores é a construção de portfólios, onde um portfólio é composto por “n” ativos financeiros. Um capital inicial é alocado entre esses ativos, de forma igualitária ou com pesos variáveis atribuídos a cada um. O objetivo é maximizar o lucro ao longo de um período definido.

Periodicamente, o portfólio passa por um rebalanceamento, reavaliando os pesos de cada ativo e decidindo se mantêm, substituem ou ajustam suas proporções. Essa periodicidade, frequentemente mensal, trimestral ou semestral, é determinada pelo investidor. Esse processo dá origem ao problema de otimização de portfólio, conforme ilustrado na Figura 1.1 [Behera et al. 2023].

O conceito de otimização permeia problemas de tomada de decisão em diversas áreas, da economia à engenharia, refletindo a busca pela melhor solução em cada contexto [Chong e Zak 2013]. Inicialmente, define-se uma função objetivo que quantifica a qualidade de uma solução. Essa função, que pode representar, por exemplo, a maximização do lucro ou a minimização do risco, depende das variáveis do problema. O objetivo da otimização é encontrar o conjunto de valores para essas variáveis que otimize (maximize ou minimize) a função objetivo [Nocedal e Wright 2006]. No caso da otimização de portfólio, busca-se obter uma divisão ótima de capital entre diferentes ativos, maximizando os retornos financeiros e minimizando o risco.

Como descrito por [Pinto 2022], a literatura econômica tradicional aborda o problema da otimização de portfólio utilizando modelos como CAPM, Markowitz e Black-Litterman. Este problema pode ser modelado como um processo de decisão de Markov, o que o torna adequado para a aplicação de algoritmos de aprendizado por reforço.

O aprendizado por reforço é um processo no qual um agente interage com um ambiente, realizando ações e recebendo recompensas com base nessas ações. Por meio de tentativa e erro, o agente ajusta seu comportamento para maximizar a recompensa total acumulada ao longo do tempo. Ao final do processo, ele se torna capaz de identificar a melhor ação a ser tomada em cada situação. O Aprendizado por Reforço Profundo (ARP), por sua vez, combina o aprendizado por reforço com técnicas de Aprendizado Profundo (*deep learning*) (AP), incorporando redes neurais ao processo de tomada de decisão.

A natureza imprevisível e volátil do mercado e dos preços das ações torna esse um problema complexo. Enquanto muitos estudos em finanças utilizam técnicas de aprendizado supervisionado para prever preços, pesquisas recentes têm demonstrado os benefícios dos algoritmos de Aprendizado por Reforço Profundo nesse domínio, especialmente na otimização de portfólios. Estes algoritmos se destacam por sua capacidade de lidar com problemas de alta dimensionalidade, característicos do mercado financeiro, e de se adaptar às dinâmicas incertas do mercado, aprendendo a ação ótima com base nos dados históricos de preços, em vez de tentar prever se um preço irá subir ou descer [Soleymani e Paquet 2020][Benhamou et al. 2021][Niu, Li e Li 2022]. O aprendizado por reforço permite ainda a incorporação de fatores como custos de transação e liquidez no processo de aprendizagem, aspectos muitas vezes difíceis de modelar como *features* [Pinto 2022].

Na aplicação de algoritmos de ARP para otimização de portfólios, a seleção de *features* relevantes para a tomada de decisão do agente é um fator a ser considerado. Geralmente, utilizam-se histórico de preços, volumes de transação e indicadores técnicos, como médias móveis. Entretanto, outras fontes de informação, como índices financeiros e notícias, também podem ser exploradas como indicadores úteis para o agente.

Conforme [Debastiani e Russo 2008], uma ação torna-se atrativa para a Escola Fundamentalista quando a empresa a qual ela retrata apresenta boas perspectivas futuras. Para avaliá-las, a análise fundamentalista examina os fundamentos da empresa, ou seja, os índices financeiros divulgados nos balanços trimestrais. Nesse contexto, as Demonstrações Financeiras (DFs) das empresas listadas na B3, publicadas trimestralmente em conformidade com a legislação vigente [B3 2024], fornecem dados essenciais para essa análise. Esses índices refletem a saúde financeira e o desempenho das companhias, auxiliando os investidores na tomada de decisões informadas e no aprimoramento do desempenho de seus portfólios [Edirisinghe e Zhang 2007]. Embora o termo original em inglês

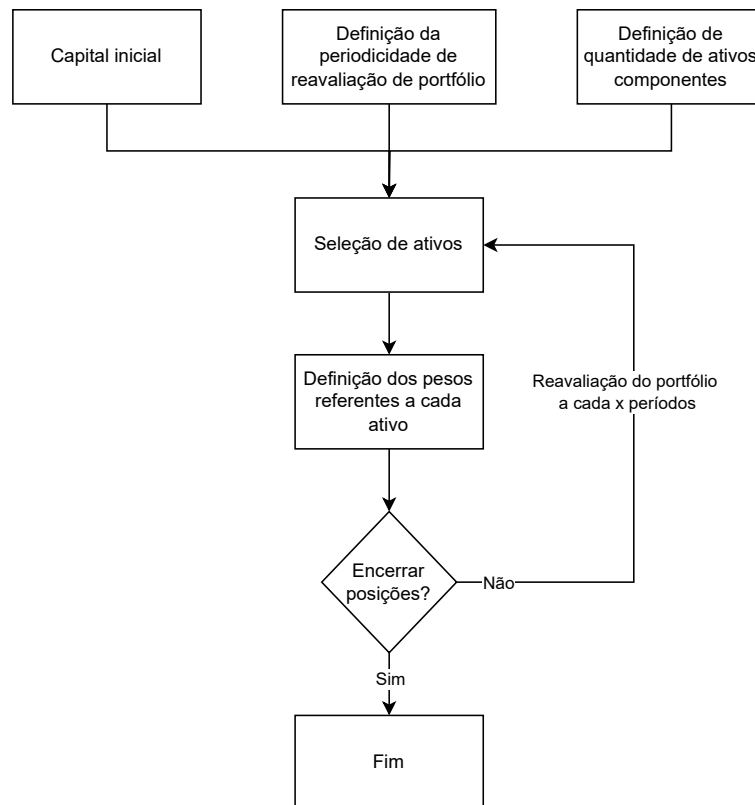


Figura 1.1: Fluxograma que descreve o problema de otimização de portfólio.

seja “*fundamental analysis*”, a tradução “análise fundamentalista” tornou-se amplamente difundida, levando à denominação dos índices como indicadores fundamentalistas.

Além disso, na era da sobrecarga de informações, uma ferramenta utilizada pelos investidores para auxiliar a tomada de decisões é a busca por conteúdos relacionados a ações, normalmente obtidos por meio de mecanismos de busca que muitas vezes levam a sites de notícias. Nos últimos anos, os avanços no processamento de linguagem natural e na aprendizagem automática permitiram a análise automática de sentimentos, uma técnica poderosa utilizada para avaliar o sentimento público e o seu impacto nos mercados financeiros [Fernandes et al. 2019]. Vários estudos empregam análise de sentimento de notícias para construir indicadores para aplicações como previsão de preços e gestão de portfólio com o objetivo de lucro, embora o indicador calculado seja normalmente baseado apenas na contagem de notícias positivas, negativas e neutras [Paiva et al. 2021][Agarwal et al. 2023].

No entanto, compreender artigos e manchetes de notícias financeiras em outros idiomas além do inglês apresenta desafios únicos. A maioria dos modelos de Processamento de Linguagem Natural (PNL) são treinados predominantemente em conjuntos de

dados em inglês [Wang 2022], devido à escassez de conjuntos de dados rotulados e recursos em outros idiomas. Esse viés linguístico dificulta o desempenho das técnicas de PNL na captura precisa das nuances e do contexto da linguagem financeira em contextos que não o inglês, especialmente considerando a importância de compreender variações linguísticas para tarefas eficazes de PNL [Tyagi e Bhushan 2023].

Contudo, há carência de estudos que exploram o mercado financeiro brasileiro e sua dinâmica, utilizando notícias em língua portuguesa e indicadores fundamentalistas. O mercado financeiro brasileiro vem crescendo de forma constante, atingindo um número total de 17,6 milhões de investidores em 2023, sendo os jovens maioria entre os investidores tanto em renda fixa quanto em renda variável [B3 2023]. A ampliação do acesso à informação e o crescimento exponencial de conteúdos voltados à educação financeira nas redes sociais são fatores que podem estar contribuindo significativamente para a tendência observada entre os jovens [Santos, Santos e Marques 2021].

1.2 Problema

Não obstante a vasta literatura sobre a aplicação de aprendizado por reforço profundo ao mercado financeiro, ainda são escassos os estudos voltados especificamente ao mercado brasileiro que utilizam notícias em língua portuguesa como fonte de informação. Ademais, no contexto da otimização de carteiras de ações, a maior parte das pesquisas concentra-se na utilização de indicadores técnicos, como detalhado no Capítulo 3, que apresenta a Revisão Sistemática da Literatura. Por outro lado, poucos estudos exploram indicadores fundamentalistas, os quais fornecem informações relevantes sobre a saúde financeira das empresas e são amplamente utilizados por investidores na tomada de decisão.

Diante desse cenário, o presente trabalho propõe a integração do aprendizado por reforço profundo para a otimização de portfólios, combinando indicadores de sentimento extraídos de notícias em língua portuguesa com indicadores fundamentalistas, visando preencher essa lacuna na literatura.

1.3 Objetivos e Hipótese

1.3.1 Geral

Avaliar a aplicação de algoritmos de aprendizado por reforço profundo na otimização de portfólio utilizando indicadores fundamentalistas e baseados em notícias para o mercado financeiro brasileiro.

1.3.2 Específicos

São pontuados os seguintes objetivos específicos:

- Construir de base de dados de notícias, indicadores fundamentalistas e histórico de preços;
- Rotular a base de dados de notícias;
- Realizar análise comparativa entre os resultados obtidos dos experimentos com diferentes agentes de ARP;
- Avaliar o impacto da utilização de indicadores fundamentalistas e baseados em notícias na otimização de portfólios de ativos financeiros no mercado brasileiro, aplicando aprendizado por reforço profundo (ARP) e comparando os resultados com métodos tradicionais, como Buy and Hold (BAH) e o índice Ibovespa, além de comparar o desempenho do ARP com e sem o uso desses indicadores.

1.3.3 Hipótese

A combinação de indicadores fundamentalistas e baseados em notícias, aplicando aprendizado por reforço profundo, pode gerar resultados favoráveis no cenário do mercado financeiro brasileiro.

1.4 Estrutura do trabalho

Esta dissertação está organizada da seguinte forma: o Capítulo 2 apresenta a fundamentação teórica, abordando os conceitos essenciais para a compreensão do estudo; o Capítulo 3 expõe a Revisão Sistemática da Literatura, com o objetivo de mapear o estado da arte relacionado ao tema da pesquisa; o Capítulo 4 descreve o método proposto; o Capítulo 5 detalha a análise estatística utilizada na seleção do indicador de sentimento; o Capítulo 6 discute os resultados obtidos; e o Capítulo 7 apresenta as considerações finais. Por fim, são listadas as referências bibliográficas.

Fundamentação Teórica

Para fornecer uma compreensão mais aprofundada da solução proposta a ser discutida posteriormente, é crucial familiarizar-se com os conceitos fundamentais. Ao modelar o problema de otimização de portfólio como um Processo de Decisão Markov (*Markov Decision Process* - MDP), é possível aplicar algoritmos de aprendizado por reforço profundo (ARP) para aprender estratégias de investimento com base em dados históricos, adaptando-se às mudanças nas condições de mercado. Além disso, existem diversos tipos de indicadores fundamentalistas que auxiliam na avaliação da saúde financeira de uma empresa. Portanto, as subseções seguintes descrevem os conceitos de aprendizado por reforço profundo e de análise fundamentalista, além de apresentar as técnicas de *embeddings*, *prompt engineering* e modelagem de tópicos, ferramentas essenciais para a extração de informação e aprimoramento da solução proposta.

2.1 Aprendizado por Reforço Profundo

Aprendizado por Reforço (AP) é um subcampo de *Machine Learning* (ML - aprendizagem de máquina) no qual um agente realiza ações em um ambiente e recebe recompensas baseadas na ação realizada em cada estado. O agente possui então o objetivo de maximizar a recompensa cumulativa recebida ao longo do tempo realizando as ações que resultam em recompensas maiores. Diferentemente do aprendizado supervisionado, o agente não recebe instruções de qual ação realizar, e necessita descobrir por meio de tentativas qual ação resulta em uma maior recompensa em cada estado [François-Lavet et al. 2018].

Frequentemente, os problemas abordados utilizando aprendizado por reforço são modelados como um processo de decisão de Markov (MDP - *Markov Decision Process*), que é um modelo matemático aplicado a problemas de decisão sequenciais [Jin 2023]. Esse modelo incorpora ações aleatórias realizadas por um agente, juntamente com as recompensas associadas, levando em consideração as propriedades de Markov do estado do sistema. Como descrito por [Arulkumaran et al. 2017], um MDP é definido da seguinte forma:

- \mathcal{S} representa o conjunto de estados possíveis em um processo;
- \mathcal{A} significa as ações disponíveis;
- $\mathcal{T}(s_{t+1}|s_t, a_t)$ é uma função que calcula a probabilidade de transição do estado s_t para $s_{t+1} \in \mathcal{S}$ ao realizar uma ação $a \in \mathcal{A}$ no instante t ;
- $\mathcal{R}(s_t, a_t, s_{t+1})$ é uma função que determina a recompensa ou punição associada a uma ação a_t realizada no estado $s_t \in \mathcal{S}$;
- γ é o fator de desconto que determina a importância das recompensas futuras, de modo que $\gamma \in [0, 1]$. Um valor mais próximo de 0 enfatiza recompensas imediatas, enquanto um valor mais próximo de 1 considera as recompensas a longo prazo.

Em tarefas episódicas, com um estado inicial e final bem definidos, o agente percorre uma trajetória de estados, ações e recompensas até atingir o estado final, quando o episódio se encerra e o processo reinicia. Essa trajetória, com tamanho finito T , representa um episódio completo. A função de recompensa é definida pela Equação 2-1.

$$R = \sum_{t=0}^{T-1} \gamma^t r_{t+1} \quad (2-1)$$

Em contrapartida, tarefas contínuas não possuem um estado final definido, fazendo com que T tenda ao infinito. Nesse caso, o retorno, ou seja, a soma das recompensas ao longo do tempo, também tenderia ao infinito. Para contornar esse problema, utiliza-se um fator de desconto $\gamma < 1$, que atribui um peso menor às recompensas futuras, garantindo que o retorno seja finito.

A política, definida como $\pi : \mathcal{S} \rightarrow p(\mathcal{A} = a|\mathcal{S})$, mapeia cada estado a uma distribuição de probabilidade sobre as ações possíveis. O objetivo do aprendizado por reforço é encontrar a política ótima, denotada por π^* (Eq. 2-2), que maximiza o retorno cumulativo esperado a partir de qualquer estado.

$$\pi^* = \arg \max_{\pi} \mathbb{E}[R(\pi)] \quad (2-2)$$

A modelagem de um problema como um Processo de Decisão de Markov geralmente assume a propriedade de Markov, que estabelece que a probabilidade de transição para o próximo estado depende apenas do estado atual. Em outras palavras, assume-se que o estado atual contém todas as informações relevantes para a tomada de decisão.

Na prática, essa suposição pode ser irrealista, já que em muitos cenários reais os estados não são totalmente observáveis. Para lidar com essa limitação, utiliza-se o Processo de Decisão de Markov Parcialmente Observável (POMDP - *partially observable MDP*). Em um POMDP, o agente recebe observações parciais do ambiente, representadas

por $o_t \in \Omega$. A distribuição de probabilidade da observação no tempo $t+1$, dada por $p(o_{t+1}|s_{t+1}, a_t)$, depende tanto do estado atual (s_{t+1}) quanto da ação anterior (a_t). Posto isso, enquanto um MDP assume observação completa do estado, um POMDP lida com a realidade de informações parciais, tornando-o mais adequado para cenários complexos do mundo real.

É interessante ressaltar que no âmbito do mercado financeiro, os agentes nem sempre têm acesso a todas as informações que afetam os preços futuros das ações. Como resultado, a condição característica de Markov dificilmente é firmada. No entanto, esta pressuposição simplifica a modelagem do problema.

Sendo assim, como a otimização de portfólio é um problema sequencial, geralmente é modelado como um MDP, possibilitando a aplicação de algoritmos de aprendizado por reforço.

Posto isso, Aprendizado por Reforço Profundo (ARP) combina aprendizado por reforço (AP) com arquiteturas de *deep learning* (aprendizado profundo), especialmente redes neurais. Essa integração proporciona ao agente uma notável capacidade de percepção advinda do *deep learning* e uma robusta habilidade de tomada de decisão proporcionada pelo aprendizado por reforço [Yi e Liu 2023].

A Figura 2.1 apresenta uma estrutura de algoritmo de ARP padrão, no qual um agente em um dado instante t realiza uma ação a . Então, o sistema transiciona para um estado s' e o agente recebe uma recompensa. Ao final do processo de aprendizagem, obtém-se uma política [Mao et al. 2016], e é representada pelas redes neurais. Na gestão de portfólio, o ARP é utilizado para otimizar estratégias de alocação de ativos, desta maneira o agente aprende uma política que identifica o melhor curso de ação com base na situação do mercado. A capacidade de aprender padrões complexos apresentada pelas redes neurais possibilita que o agente lide com alta dimensionalidade e espaços contínuos de estado [Yi e Liu 2023], que são desafios enfrentados quando se tem como objeto de estudo o mercado financeiro.

As abordagens para resolver problemas de aprendizado por reforço incluem métodos baseados em funções de valor, métodos baseados em busca de políticas e uma abordagem híbrida conhecida como ator-crítico, que combina funções de valor com busca de políticas.

2.1.1 Métodos baseados em função de valor

Os métodos baseados em função de valor concentram-se em aprender uma função que estima o valor, ou retorno esperado, de estar em um determinado estado (ou de executar uma ação em um estado), guiando o agente na escolha de ações que maximizam

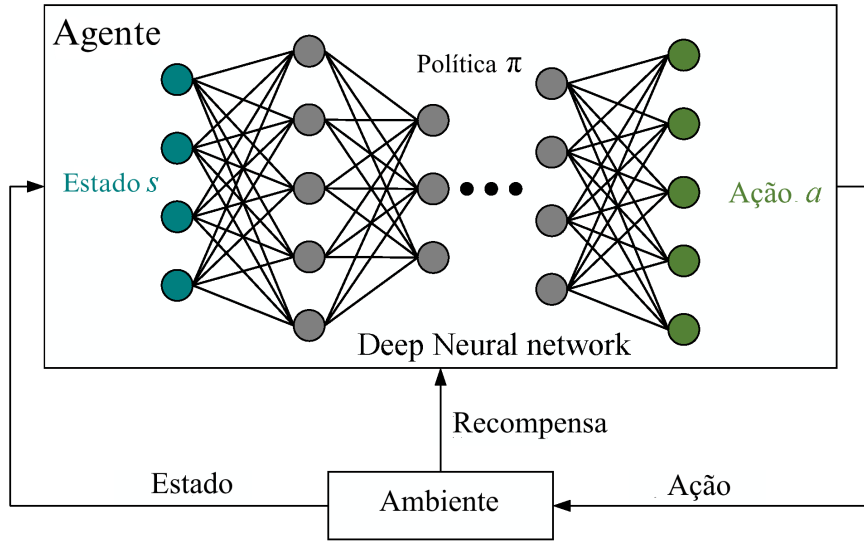


Figura 2.1: Estrutura de um algoritmo de ARP padrão [Yi e Liu 2023].

essa função de valor. A função de valor é determinada pela Eq. 2-3, de forma que a função de valor ótima é definida pela Eq. 2-4.

$$V^\pi(s) = \mathbb{E}[R|s, \pi] \quad (2-3)$$

$$V^* = \max_{\pi} V^\pi(s) \quad \forall s \in \mathcal{S} \quad (2-4)$$

No cenário de aprendizado por reforço, como as dinâmicas de transição T não são conhecidas, utiliza-se o cálculo da função de qualidade $Q^\pi(s, a)$ (Equação 2-5), sendo esta similar à função de valor de estado $V^\pi(s)$, com exceção de que é fornecida um ação inicial a . A política ótima pode ser obtida selecionando-se, em cada estado, a ação com o maior valor $Q(s, a)$.

$$Q^\pi(s, a) = \mathbb{E}[R|s, a, \pi] \quad (2-5)$$

Os algoritmos clássicos *Q-learning* e SARSA são métodos baseados em valor, os quais buscam estimar a função $Q(s, a)$ de forma recursiva. Essa recursão se baseia na equação de Bellman (Equação 2-6), que define a função $Q(s, a)$ em termos da recompensa imediata e da função Q no estado subsequente. Para aprimorar a estimativa da função $Q(s, a)$, os algoritmos *Q-learning* e SARSA empregam uma técnica de atualização baseada na diferença temporal, como demonstrado na Equação 2-7.

$$Q^\pi(s_t, a_t) = \mathbb{E}_{s_{t+1}}[r_{t+1} + \gamma Q^\pi(s_{t+1}, \pi(s_{t+1}))] \quad (2-6)$$

$$Q^\pi(s_t, a_t) \leftarrow Q^\pi(s_t, a_t) + \alpha \sigma \quad (2-7)$$

onde: α representa a taxa de aprendizado, que controla a velocidade de atualização da função $Q(s, a)$; σ corresponde ao erro de diferença temporal, calculado como $Y - Q^\pi(s_t, a_t)$, em que Y é o alvo da atualização, que representa uma estimativa mais precisa de $Q^\pi(s_t, a_t)$ e varia entre os algoritmos Q-learning e SARSA.

Enquanto *Q-learning* é um método *off-policy*, SARSA é um algoritmo *on-policy*. Segundo [Sutton e Barto 2018], métodos *on-policy* buscam avaliar ou melhorar a política que está sendo utilizada para tomar decisões e gerar os dados de treinamento. Métodos *off-policy* buscam avaliar ou melhorar uma política diferente daquela utilizada para gerar os dados de treinamento, permitindo que o agente aprenda com experiências geradas por outras políticas.

2.1.2 Métodos baseados em busca de política

Os métodos baseados em busca de política tem como foco a busca pela política ótima que mapeia a probabilidade de escolha de cada ação em cada estado, sem a necessidade explícita de calcular a função de valor.

Essa busca pode ser realizada utilizando algoritmos de otimização, como *Policy Gradient* (Gradiente de política) e algoritmos evolutivos, como o Hill-Climbing, que realizam a busca em um subspaço de políticas. No *Policy Gradient* (PG - Gradiente de política), os parâmetros da política são ajustados na direção que maximiza o retorno esperado, utilizando o gradiente da função objetivo em relação aos parâmetros da política.

2.1.3 Métodos ator-crítico

Os métodos Ator-Crítico representam uma abordagem híbrida no aprendizado por reforço, combinando elementos das abordagens anteriores. Deste modo, o componente “Ator” é responsável por aprender e executar a política, que determina a ação a ser tomada em cada estado. Enquanto isso, o componente “Crítico” aprende a função de valor, avaliando a qualidade da política adotada pelo Ator. Assim, o Crítico fornece *feedback* ao Ator, auxiliando-o na escolha de ações mais eficazes e refinando a política ao longo do tempo. A interação entre Ator e Crítico permite um aprendizado mais eficiente e estável, especialmente em problemas complexos que envolvem muitos estados e ações, como o mercado financeiro. Portanto, os algoritmos de ARP implementados neste estudo pertencem à categoria dos métodos ator-crítico e serão abordados detalhadamente nas subseções seguintes. Para descrições mais aprofundadas, consulte o trabalho de [Wang et al. 2024].

Advantage Actor-Critic (A2C)

O algoritmo *Advantage Actor-Critic* (A2C) é um método de aprendizado por reforço síncrono e on-policy derivado do algoritmo *Asynchronous Advantage Actor-Critic* (A3C) [Wang et al. 2024]. A2C emprega uma arquitetura ator-crítico, aprendendo simultaneamente uma política $\pi(a|s)$ (o ator) e uma função de valor de estado $V(s)$ (o crítico). O crítico avalia a política proposta pelo ator, guiando assim o processo de aprendizagem e reduzindo a variância. A2C opera com múltiplos atores paralelos coletando experiências simultaneamente. Uma distinção chave do A3C está nas atualizações de parâmetros síncronas do A2C, onde uma política central é atualizada após todos os atores terem completado sua coleta de experiências [Wang et al. 2024]. Essa sincronização contribui para um treinamento mais estável. A função vantagem $A(s, a)$ (Eq. 2-8) é usada para reduzir a variância.

$$A(s, a) = Q(s, a) - V(s) \quad (2-8)$$

Aqui, $Q(s, a)$ representa a função valor-ação, denotando a recompensa acumulada esperada ao se tomar a ação a no estado s e seguir a política ótima a partir daí, e $V(s)$ é a função de valor de estado, representando a recompensa acumulada esperada começando do estado s [Paiva et al. 2022]. A política é atualizada usando um método de subida de gradiente, com o gradiente escalado pela função vantagem, como é denotado na Eq. 2-9).

$$\nabla_{\theta} J(\theta) = \mathbb{E}t[\nabla \theta \log \pi(a_t | s_t; \theta) A(s_t, a_t)] \quad (2-9)$$

Em que: θ representa os parâmetros da política; $\pi(a_t | s_t; \theta)$ é a probabilidade de tomar a ação a_t no estado s_t dados os parâmetros θ ; $A(s_t, a_t)$ é a função de vantagem no tempo t ; $J(\theta)$ é a função objetivo que buscamos maximizar.

O mecanismo de atualização síncrona do A2C oferece uma maior estabilidade em comparação com a abordagem assíncrona do A3C, pois leva a dinâmicas de treinamento mais estáveis [Wang et al. 2024]. Ele utiliza eficientemente GPUs, permitindo um treinamento mais rápido através da paralelização [Wang et al. 2024]. O A2C atinge um desempenho competitivo, igualando ou superando o A3C, enquanto aprende simultaneamente as funções de valor e política. Isso reduz o viés típico dos métodos apenas críticos e a variância inerente às abordagens apenas de ator [Paiva et al. 2022].

Proximal Policy Optimization (PPO)

Proximal Policy Optimization (PPO) é um método de política de gradiente projetado para alcançar um equilíbrio entre a facilidade de implementação, a complexi-

dade das amostras e a eficiência das amostras, garantindo que as atualizações da política permaneçam próximas à política anterior [Khemlichi et al. 2023]. Sua estabilidade e relativa simplicidade contribuíram para sua popularidade no aprendizado por reforço [Byun, Kim e Wang 2020].

O PPO simplifica o algoritmo *Trust Region Policy Optimization* ao incorporar uma penalidade na função objetivo, restringindo efetivamente a atualização da política [Han et al. 2023]. Em vez de impor explicitamente uma restrição, o PPO otimiza uma função objetivo substituída com razões de probabilidade truncadas [Schulman et al. 2017]. A função objetivo no PPO é definida como [Wang et al. 2024]:

$$L(\theta) = \mathbb{E}_t \left[\min(r_t(\theta)A_t, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon)A_t) \right] \quad (2-10)$$

onde: $r_t(\theta)$ é a razão da probabilidade de uma ação sob a política atual em relação à sob a política antiga: $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)}$. A_t é a função de vantagem estimada no tempo t . ϵ é um hiperparâmetro que define o intervalo de recorte.

O mecanismo de recorte restringe a razão da política $r_t(\theta)$ ao intervalo $[1 - \epsilon, 1 + \epsilon]$, evitando atualizações de política excessivamente grandes que poderiam desestabilizar o treinamento [Wang et al. 2024]. O algoritmo incentiva o agente a alcançar de forma confiável a solução ótima fazendo pequenas e conservadoras atualizações de política.

O PPO se destaca pela simplicidade de implementação em comparação com outros algoritmos de otimização de políticas [Byun, Kim e Wang 2020]. Sua função objetivo ajustada impede mudanças bruscas, proporcionando um treinamento mais estável [Wang et al. 2024]. O algoritmo é eficaz com um número relativamente pequeno de amostras, alcançando bom desempenho [Byun, Kim e Wang 2020]. Além disso, oferece um equilíbrio ideal entre complexidade de implementação, eficiência de amostra e desempenho geral [Khemlichi et al. 2023]. Essa combinação de simplicidade, estabilidade e eficácia faz do PPO uma escolha popular para várias tarefas de aprendizado por reforço, sendo especialmente eficaz em tarefas de controle de sistemas contínuos [Liu 2024].

Deep Deterministic Policy Gradient (DDPG)

O *Deep Deterministic Policy Gradient* (DDPG) é um algoritmo ator-crítico, livre de modelo e fora da política, projetado para aprendizado em espaços de ação contínuos. DDPG combina a abordagem ator-crítico com insights de *Deep Q-Networks*, incorporando técnicas como *replay* de experiência e redes de alvo para estabilizar o aprendizado [Wang et al. 2024].

O DDPG utiliza duas redes neurais: uma rede ator $\pi(s; \theta)$ que aprende uma política determinística e uma rede crítica $Q(s, a; \omega)$ que estima a função do valor Q [Han et al. 2023]. A rede ator gera uma ação específica para um estado dado, ao con-

trário dos métodos de gradiente de política estocástica que geram uma distribuição de probabilidade sobre as ações [Khemlichi et al. 2023]. A rede crítica avalia a ação tomada pelo ator, fornecendo *feedback* para melhorar a política [Liu 2024].

A rede crítica é atualizada utilizando a equação de Bellman (Eq. 2-6). Para estabilizar o aprendizado, o DDPG emprega duas técnicas chave oriundas do DQN [Wang et al. 2024]:

- *Replay de Experiência*: DDPG usa um *buffer de replay* para armazenar transições passadas (s_t, a_t, r_t, s_{t+1}) . Durante o treinamento, um mini-lote de transições é aleatoriamente amostrado do *buffer de replay* para atualizar as redes do agente e do crítico. O uso de *replay* de experiência permite que o DDPG aprenda com experiências passadas, melhorando a eficiência de amostra.
- *Redes-Alvo (Target Networks)*: DDPG usa redes-alvo para calcular os Q-values alvo, semelhante ao DQN. No entanto, devido à maior complexidade da função valor do DDPG, faz-se necessária a técnica de substituição lenta (*soft replace*). Nesse processo, a rede-alvo se aproxima lentamente da função principal a cada iteração, tornando o treinamento mais estável. Redes-alvo são cópias atrasadas das redes de agente e crítico que são atualizadas lentamente usando uma regra de atualização suave:

$$\theta' \leftarrow \tau\theta + (1 - \tau)\theta' \quad (2-11)$$

$$\omega' \leftarrow \tau\omega + (1 - \tau)\omega' \quad (2-12)$$

onde θ' e ω' são os parâmetros das redes de ator e crítico de alvo, respectivamente, θ e ω são os parâmetros das redes principais de agente e crítico, e $\tau \ll 1$ é uma taxa de atualização pequena. Essa atualização lenta ajuda a estabilizar o processo de treinamento ao reduzir a variância dos valores alvo.

O DDPG (*Deep Deterministic Policy Gradient*) aprende uma política determinística e, em seguida, a expande em um espaço contínuo por meio da arquitetura ator-crítico [Wang et al. 2024]. Assim, o DDPG pode lidar efetivamente com tarefas em espaços de ações contínuas, onde métodos tradicionais de *Q-learning* não são diretamente aplicáveis [Koratomaddi et al. 2021]. Contudo, o DDPG pode ser sensível às configurações de hiperparâmetros, necessitando de ajustes cuidadosos para alcançar um bom desempenho [Han et al. 2023]. Ele também pode sofrer de viés de superestimação nos valores Q, o que pode levar a políticas subótimas, sendo assim o algoritmo *Twin Delayed DDPG* (TD3) foi desenvolvido para resolver este problema [Wang et al. 2024].

Twin Delayed Deep Deterministic Policy Gradient (TD3)

O *Twin Delayed Deep Deterministic Policy Gradient* (TD3) é uma melhoria do algoritmo Deep Deterministic Policy Gradient, projetado para resolver a tendência do DDPG de superestimar os valores de Q e sua sensibilidade às configurações de hiperparâmetros [Han et al. 2023]. O TD3 mantém a estrutura ator-crítico do DDPG, mas implementa diversas modificações essenciais para melhorar a estabilidade e o desempenho do treinamento [Wang et al. 2024]. Três características principais do TD3 ajudam a reduzir o viés de superestimação e a aumentar a estabilidade [Han et al. 2023]:

- ***Clipped Double Q-Learning***: O TD3 emprega duas redes críticas independentes, $Q_1(s, a)$ e $Q_2(s, a)$. Durante a atualização do crítico, utiliza-se o menor valor entre as duas estimativas de Q, o que ajuda a mitigar a superestimação dos valores de ação, promovendo alvos mais conservadores [Wang et al. 2024]. O valor-alvo y é definido por:

$$y = r + \gamma \min_{i=1,2} Q_{i,\text{alvo}}(s', a') \quad (2-13)$$

em que $a' = \pi_{\text{alvo}}(s')$ representa a ação gerada pela política-alvo no próximo estado s' .

- ***Atualizações de Política Atrasadas***: O TD3 atualiza a rede de política (ator) e as redes-alvo com menor frequência em relação às redes críticas [Wang et al. 2024, Han et al. 2023]. Geralmente, a política é atualizada a cada d atualizações do crítico (por exemplo, $d = 2$). Essa estratégia visa reduzir a propagação de erros nas atualizações da política, contribuindo para a estabilidade do treinamento e mitigando o sobreajuste.
- ***Suavização da Política-Alvo***: Para suavizar as estimativas da função Q, o TD3 adiciona ruído à ação-alvo durante o cálculo do valor-alvo [Han et al. 2023]. Essa abordagem dificulta que a política aprenda a explorar imperfeições nas estimativas de Q, promovendo maior robustez.

A ação-alvo a' é obtida por:

$$a'(s') = \text{clip}(\mu_{\theta_{\text{alvo}}}(s') + \text{clip}(\epsilon, -c, c), a_{\text{mín}}, a_{\text{máx}}), \quad \epsilon \sim \mathcal{N}(0, \sigma) \quad (2-14)$$

onde $\mu_{\theta_{\text{alvo}}}(s')$ representa a ação sugerida pela política-alvo parametrizada por θ_{alvo} , ϵ é um termo de ruído gaussiano com média zero e desvio padrão σ , e $a_{\text{mín}}$, $a_{\text{máx}}$ denotam os limites inferior e superior das ações permitidas. O ruído truncado contribui para a regularização da política-alvo e suavização da função Q [Liu 2024].

O algoritmo TD3 apresenta diversas melhorias em relação ao DDPG, destacando-se como uma abordagem mais robusta e eficaz para tarefas de controle

contínuo [Han et al. 2023]. Ao empregar o *Clipped Double Q-Learning*, o TD3 mitiga o viés de superestimação nas estimativas da função Q, resultando em um treinamento mais estável e confiável. Além disso, a utilização de atualizações de política atrasadas e a suavização da política-alvo contribuem significativamente para a estabilidade do processo de aprendizagem. A introdução de ruído truncado nas ações-alvo aumenta a robustez da política frente a imperfeições na função de valor. Como resultado, o TD3 tem demonstrado desempenho competitivo em diversos ambientes de controle contínuo, sendo amplamente reconhecido como um dos algoritmos *off-policy* mais eficazes em aprendizado por reforço [Liu 2024].

Soft Actor-Critic (SAC)

Soft Actor-Critic (SAC) é um algoritmo ator-crítico *off-policy* que opera no contexto de aprendizado por reforço com máxima entropia [Haarnoja et al. 2018]. Diferentemente de abordagens tradicionais, o SAC não depende das trajetórias passadas, buscando maximizar simultaneamente a recompensa esperada e a entropia da política [Han et al. 2023]. Esse critério de otimização promove uma maior exploração durante o treinamento, favorecendo o aprendizado de políticas mais robustas e generalizáveis [Khemlichi et al. 2023].

Uma característica fundamental do SAC é a utilização de um ator estocástico, em que a política retorna uma distribuição de probabilidade sobre as ações, em vez de uma ação determinística [Liu 2024]. A função objetivo do algoritmo incorpora um termo de entropia, conforme representado a seguir:

$$J(\pi) = \mathbb{E}_{\tau \sim \pi} \left[\sum_{t=0}^{T-1} \gamma^t (R(s_t, a_t) + \alpha H(\pi(\cdot | s_t))) \right] \quad (2-15)$$

onde $R(s_t, a_t)$ representa a recompensa no instante t , $H(\pi(\cdot | s_t))$ é a entropia da política no estado s_t , e α é o coeficiente de temperatura que controla o peso relativo da entropia na função objetivo [Liu 2024].

Assim como o TD3, o SAC emprega duas funções Q para reduzir o viés de superestimação. O valor mínimo entre as duas estimativas é utilizado para compor o valor-alvo durante o treinamento [Liu 2024]. Tanto a política quanto as funções Q são atualizadas de forma iterativa por meio de gradiente descendente estocástico [Han et al. 2023].

A inclusão do termo de entropia na função objetivo incentiva a exploração contínua, sendo especialmente útil em ambientes de alta complexidade [Han et al. 2023, Khemlichi et al. 2023]. Além disso, o SAC é reconhecido por sua robustez a diferentes configurações de hiperparâmetros e pela eficiência amostral superior em comparação a algoritmos baseados em trajetórias completas [Haarnoja et al. 2018].

2.2 Análise fundamentalista

A análise fundamentalista é um método de avaliação que considera fatores financeiros e não financeiros para determinar o verdadeiro valor intrínseco de uma empresa [Spooner 1984]. Isto, por sua vez, ajuda a identificar se o valor atual de uma empresa está sobrevalorizado ou subvalorizado, orientando decisões sobre quando comprar ou vender ativos. Também avalia se o preço é justo, indicando o alinhamento com o valor intrínseco [Barros 2015].

Quando se realiza uma análise fundamentalista, podem ser consideradas duas perspectivas: as abordagens *top-down* e *bottom-up*. Na abordagem *top-down*, os movimentos de preço são governados por fatores macroeconômicos, começando com uma análise da economia geral, depois o setor e, finalmente, a saúde da empresa. Por outro lado, a abordagem *bottom-up* concentra-se em fatores microeconômicos, enfatizando as expectativas e comportamentos individuais das empresas [Salvadori 2021][Malta e de Camargos 2016]. Para este projeto de pesquisa, será adotado o ponto de vista *bottom-up*, priorizando a análise detalhada dos atributos e desempenho específicos de cada empresa.

Ao analisar os dados contábeis divulgados pelas empresas, é possível derivar índices financeiros que oferecem um retrato da situação da empresa. Conforme observado por [Colnaghi 2021], estes indicadores fundamentalistas oferecem perspectivas sobre as condições atuais e futuras de uma empresa, tornando-os ferramentas cruciais para a realização de análises financeiras. Além disso, [Whittington 1980] identifica duas funções principais desses índices: normativa, onde medem o desempenho de uma empresa em relação a um *benchmark*, e preditiva, onde são utilizados para fins de previsão.

Conforme delineado por [Lagioia 2011], os índices fundamentalistas são divididos em quatro categorias principais: indicadores de mercado, que analisam o desempenho das ações; indicadores de liquidez, que avaliam a capacidade da empresa em cumprir obrigações com seus ativos; indicadores de endividamento, que medem a proporção de financiamento de terceiros em relação ao patrimônio; e indicadores de rentabilidade, que demonstram o retorno sobre o capital investido, incluindo métricas de giro. Em seguida, os indicadores utilizados neste trabalho serão descritos.

O *Dividend Yield* é um indicador de rentabilidade que avalia o desempenho de uma empresa com base nos dividendos distribuídos aos acionistas nos últimos 12 meses, como mostrado na Eq. 2-16. Os dividendos representam uma parcela dos lucros distribuída entre os acionistas, permitindo, assim, avaliar o retorno das ações para o investidor em função dos proventos pagos.

$$\text{Dividend Yield} = \frac{\text{Dividendos pagos por ação}}{\text{Preço unitário da ação}} \quad (2-16)$$

O Preço/Lucro é um indicador de mercado utilizado para verificar se o preço de uma ação está sobrevalorizada ou subvalorizada, sendo calculada de acordo com a Eq. 2-17.

$$P/L = \frac{\text{Preço}}{\text{Lucro}} \quad (2-17)$$

O Retorno sobre o Patrimônio Líquido (*Return on Equity* - ROE) é um índice de rentabilidade utilizado para identificar quão a empresa é eficiente em gerar lucro a partir de seus recursos. O ROE é calculado por meio da Eq. 2-18.

$$\text{ROE} = \frac{\text{Lucro líquido}}{\text{Patrimônio líquido}} \quad (2-18)$$

A Margem Bruta é uma métrica usada para avaliar a rentabilidade bruta de uma empresa. Ela é calculada como a razão entre o lucro bruto e a receita líquida, como definido pela Eq. 2-19. Essa métrica também indica o retorno percentual sobre o investimento realizado na empresa.

$$\text{Margem Bruta} = \frac{\text{Lucro bruto}}{\text{Receita líquida}} \quad (2-19)$$

2.3 Embeddings

Embedding de palavras é uma técnica de processamento de linguagem natural que transforma dados de texto categóricos em vetores numéricos. Esta abordagem é eficaz em fornecer contexto semântico aos modelos, posicionando palavras com significados semelhantes mais próximas umas das outras no espaço vetorial.

Como pode ser observado na Figura 2.2, a distância entre as palavras “rei” e “rainha” deve ser equivalente à distância entre “homem” e “mulher”, uma vez que ambas possuem a mesma relação semântica (gênero). Da mesma forma, as palavras “andou” e “andando”, assim como “nadou” e “nadando”, devem apresentar distâncias iguais, pois compartilham a mesma relação semântica (conjugação verbal).

Este mapeamento semântico permite a determinação da similaridade das sentenças calculando sua distância no espaço vetorial. Esta propriedade mostra-se valiosa em diversas aplicações, como sistemas de recomendação. Ao mapear usuários e itens dentro do espaço vetorial, o sistema pode identificar e recomendar efetivamente itens semelhantes às preferências de um usuário [Wang et al. 2023].

As formas mais comuns de calcular a distância entre vetores incluem a distância de Manhattan, também conhecida como Norma L1 (Eq. 2-20); a distância euclidiana, ou Norma L2 (Eq. 2-21); a distância de cosseno, também re-

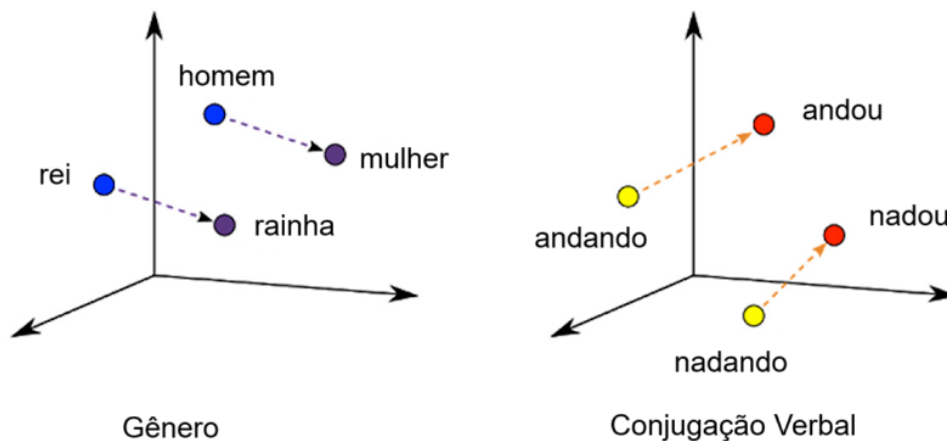


Figura 2.2: Exemplo de relação semântica capturada por embeddings [Google Cloud 2020].

ferida como similaridade de cosseno (Eq. 2-22); e o produto escalar (Eq. 2-23) [Morciano et al. 2024][Kurek et al. 2024][Zhu e Huang 2022].

A distância de Manhattan é a soma das diferenças absolutas entre pontos em todas as dimensões. Esta distância é nomeada em referência às ruas do distrito de Manhattan, Nova York. Observando seu cálculo em apenas duas dimensões, essa métrica propõe distâncias em grade, similares aos blocos e quadras de Manhattan. Esta métrica é utilizada para calcular a similaridade ou distância entre dois pontos, sendo aplicada em diversas áreas, como busca em vizinhança, problemas de clusterização ou agrupamento.

A distância euclidiana é fundamentada na obtenção da distância entre dois pontos por meio de um linha reta, representando intuitivamente a “menor distância” entre os pontos [Suwanda, Syahputra e Zamzami 2020].

A similaridade de cosseno é uma métrica amplamente utilizada em análise de texto e recuperação de informação. Baseia-se no cálculo do cosseno do ângulo entre dois vetores x e y , indicando a orientação de um em relação ao outro. O valor resultante varia de -1 a 1, onde -1 representa orientações opostas, 0 indica ortogonalidade (ausência de similaridade) e 1 representa que os vetores têm a mesma orientação [Kurek et al. 2024].

O produto escalar entre dois vetores x e y também mede a similaridade entre eles. Quanto maior o produto escalar, maior o grau de similaridade entre os vetores.

Estes métodos são definidos pelas seguintes equações:

$$d_{\text{Manhattan}}(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (2-20)$$

$$d_{\text{Euclidiana}}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2-21)$$

$$d_{\text{Cosseno}}(x, y) = 1 - \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \quad (2-22)$$

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^n x_i y_i = x_1 y_1 + x_2 y_2 + \dots + x_n y_n \quad (2-23)$$

onde n denota o número total itens, e $\mathbf{x} = (x_1, x_2, \dots, x_n)$ e $\mathbf{y} = (y_1, y_2, \dots, y_n)$ são vetores com n dimensões.

2.4 Prompt Engineering

De acordo com [Lee et al. 2023], os LLM's demonstraram a capacidade de extrair sinais comerciais lucrativos a partir do sentimento de notícias. Um aspecto essencial da utilização desses modelos envolve a elaboração de prompts eficazes para obter respostas úteis para aplicações específicas. Este processo, conhecido como *Prompt Engineering* (engenharia de prompt), abrange técnicas e habilidades destinadas a compreender as limitações do modelo e interagir efetivamente com eles para obter os resultados desejados [Saravia 2022].

Elaborar um *prompt* eficaz é um processo iterativo que começa com a definição da tarefa. Em seguida, um prompt inicial é construído, que serve de base para ajustes e refinamentos subsequentes com base nas respostas do modelo. Este ciclo de construção, avaliação de respostas e refinamento perdura até que se alcance um resultado satisfatório.

2.4.1 Elementos de um prompt

Um prompt serve como entrada para orientar a resposta de um modelo de linguagem. Pode incluir qualquer combinação dos seguintes elementos:

- Instrução: Uma descrição clara da tarefa ou instrução a ser executada pelo modelo;
- Contexto: Informações adicionais ou antecedentes que podem melhorar a compreensão e a resposta do modelo;
- Entrada: A pergunta, declaração ou ponto de dados específico que requer uma resposta;
- Saída: instruções sobre o formato, estilo ou tipo de saída desejado.

2.4.2 Tipos de prompts

Diferentes tipos de prompt provocam respostas variadas de modelos de linguagem. Os tipos comuns incluem:

- Abertos (*open-ended*): estes prompts proporcionam liberdade e flexibilidade, permitindo que o modelo gere respostas criativas e irrestritas;
 - Exemplo: Escreva um poema sobre a beleza da vida.
- Fechados (*close-ended*): estes restringem as opções de resposta, muitas vezes limitando-as a um conjunto específico de opções ou a um formato específico;
 - Exemplo: Qual é a capital da França? (a) Paris (b) Londres (c) Roma (d) Berlim
- Baseados em cenários: estes apresentam uma situação ou cenário específico, exigindo que o modelo responda dentro desse contexto;
 - Exemplo: Você é um entrevistador para uma vaga de cientista de dados. Quais perguntas faria ao candidato?
- Subjetivos: os quais incentivam o modelo a expressar opiniões, crenças ou interpretações, oferecendo perspectivas sobre seu raciocínio e considerações éticas.
 - Exemplo: Qual é a sua opinião sobre a importância da inteligência artificial na educação? Você acredita que ela pode substituir completamente os métodos de ensino tradicionais? Explique suas razões.

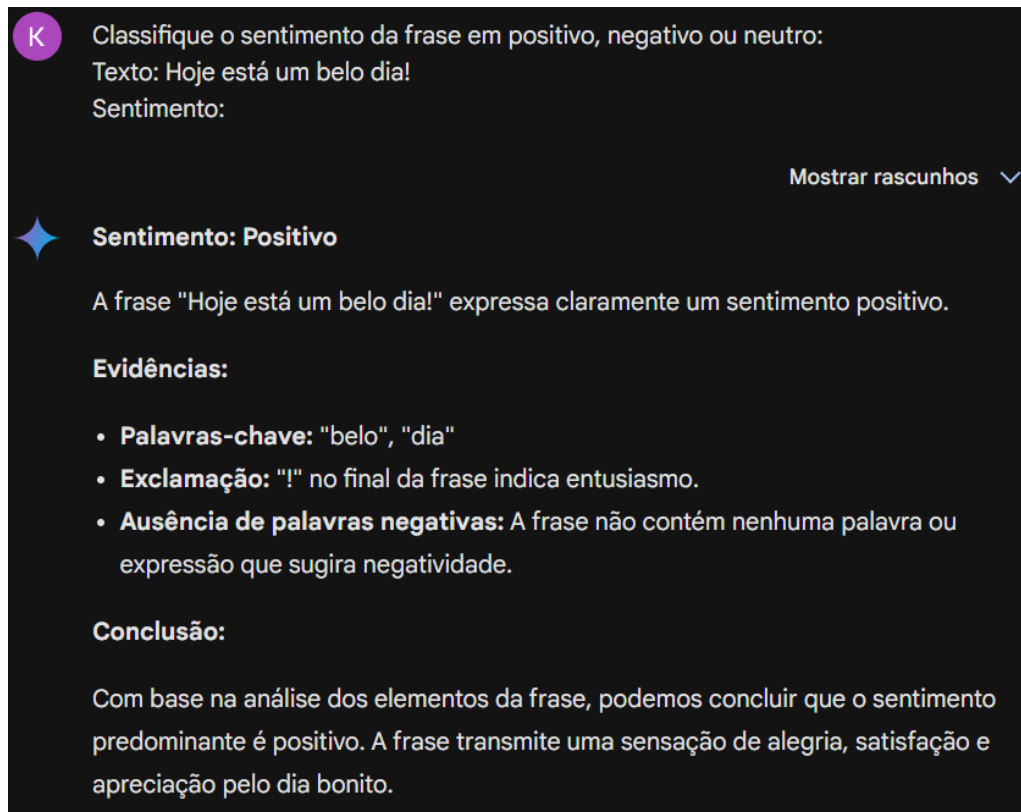
2.4.3 Técnicas

Algumas das principais técnicas de prompt *engineering* serão detalhadas nas subseções a seguir. As figuras apresentadas foram obtidas por meio da aplicação do Gemini disponibilizada pela Google¹.

Zero-shot learning

Na técnica *Zero-shot learning* é dado ao modelo apenas a tarefa a ser executada sem exemplos. A Figura 2.3 apresenta um exemplo de resposta obtida aplicando esta técnica.

¹<https://gemini.google.com/app>



K Classifique o sentimento da frase em positivo, negativo ou neutro:
Texto: Hoje está um belo dia!
Sentimento:

Mostrar rascunhos ▾

✦ **Sentimento: Positivo**

A frase "Hoje está um belo dia!" expressa claramente um sentimento positivo.

Evidências:

- **Palavras-chave:** "belo", "dia"
- **Exclamação:** "!" no final da frase indica entusiasmo.
- **Ausência de palavras negativas:** A frase não contém nenhuma palavra ou expressão que sugira negatividade.

Conclusão:

Com base na análise dos elementos da frase, podemos concluir que o sentimento predominante é positivo. A frase transmite uma sensação de alegria, satisfação e apreciação pelo dia bonito.

Figura 2.3: Exemplo de resposta aplicando *Zero-shot learning*.

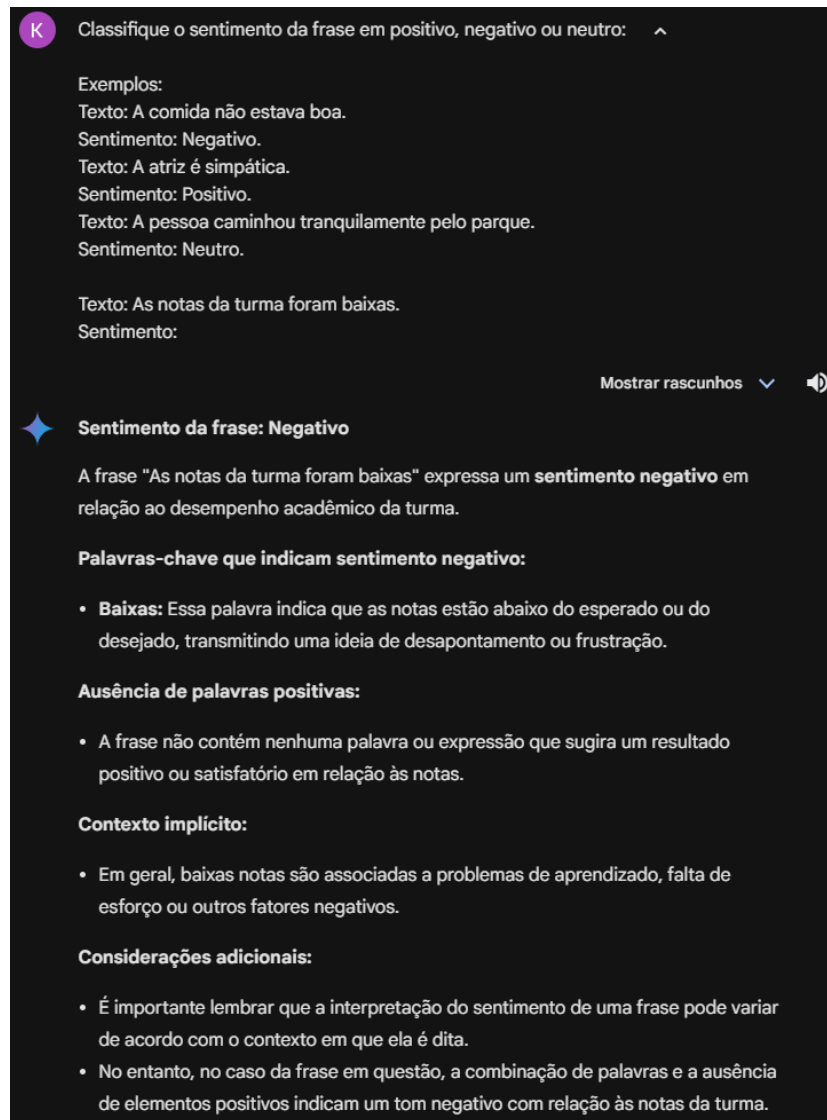
Few-shot learning

Quando a técnica *Zero-shot learning* apresenta resultados insatisfatórios, pode ser benéfico fornecer ao modelo alguns exemplos relevantes. Esta abordagem, conhecida como *Few-shot learning*, oferece contexto e exemplos adicionais, potencialmente levando a um melhor desempenho. Na Figura 2.4, é possível observar uma resposta obtida aplicando esta técnica.

Retrieval Augmented Generation - RAG

Large Language Models são ferramentas poderosas, no entanto às vezes podem gerar respostas que se desviam da realidade, um fenômeno conhecido como alucinação. Quando um LLM encontra uma consulta que não consegue responder com base em seus dados de treinamento, ele pode fabricar informações. Isto representa um desafio significativo, especialmente em aplicações como chatbots, onde a precisão das respostas e a confiança do utilizador são fundamentais.

Uma técnica eficaz para mitigar alucinações em LLMs é a *Retrieval Augmented Generation* (RAG), apresentada por pesquisadores da Meta AI. O RAG permite que o



K Classifique o sentimento da frase em positivo, negativo ou neutro: ^

Exemplos:
Texto: A comida não estava boa.
Sentimento: Negativo.
Texto: A atriz é simpática.
Sentimento: Positivo.
Texto: A pessoa caminhou tranquilamente pelo parque.
Sentimento: Neutro.

Texto: As notas da turma foram baixas.
Sentimento:

Mostrar rascunhos v 🔊

✦ **Sentimento da frase: Negativo**

A frase "As notas da turma foram baixas" expressa um **sentimento negativo** em relação ao desempenho acadêmico da turma.

Palavras-chave que indicam sentimento negativo:

- **Baixas:** Essa palavra indica que as notas estão abaixo do esperado ou do desejado, transmitindo uma ideia de desapontamento ou frustração.

Ausência de palavras positivas:

- A frase não contém nenhuma palavra ou expressão que sugira um resultado positivo ou satisfatório em relação às notas.

Contexto implícito:

- Em geral, baixas notas são associadas a problemas de aprendizado, falta de esforço ou outros fatores negativos.

Considerações adicionais:

- É importante lembrar que a interpretação do sentimento de uma frase pode variar de acordo com o contexto em que ela é dita.
- No entanto, no caso da frase em questão, a combinação de palavras e a ausência de elementos positivos indicam um tom negativo com relação às notas da turma.

Figura 2.4: Exemplo de resposta aplicando *Few-shot learning*.

modelo consulte um banco de dados externo para verificação de fatos, levando a respostas mais precisas e confiáveis.

A figura 2.5 ilustra o processo RAG. Primeiramente, os *embeddings* são gerados a partir de dados externos e armazenados em um banco de dados vetorial. Quando um usuário faz uma pergunta, o sistema de recuperação de informações utiliza a similaridade de sentenças, calculada a partir dos *embeddings* da pergunta e dos documentos, para recuperar documentos relevantes do banco de dados. O LLM então recebe a pergunta do usuário e esses documentos relevantes como entrada. Essa abordagem permite ao modelo compreender melhor o contexto da pergunta, verificar os fatos e, conseqüentemente, melhorar a qualidade das respostas.

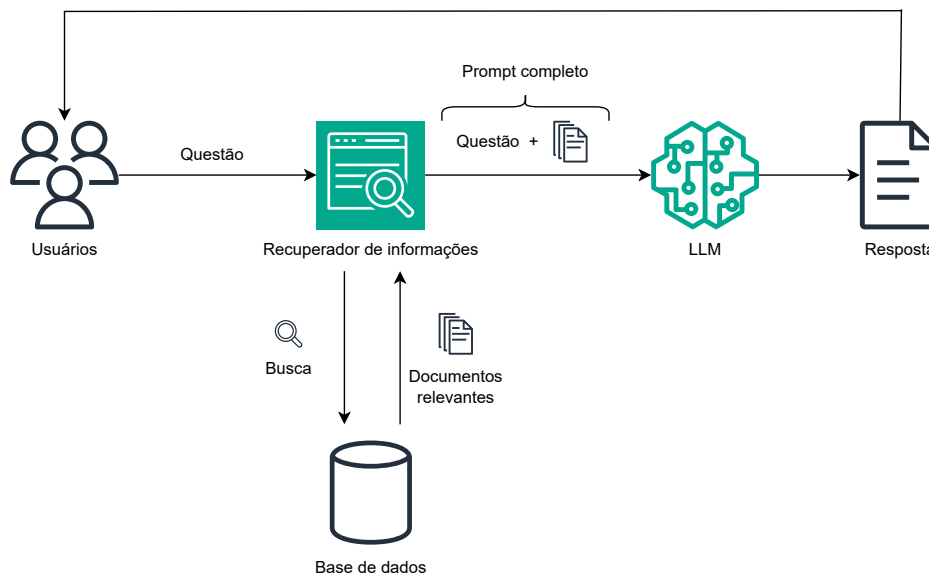


Figura 2.5: Estrutura da técnica *Retrieval Augmented Generation*.

2.5 Modelagem de tópicos

A modelagem de tópicos (*Topic Modeling*) é um método para descobrir padrões subjacentes em documentos extensos. O manuseio de documentos grandes com uma variedade de tópicos pode representar desafios, pois os métodos tradicionais de segmentação muitas vezes enfrentam dificuldades devido à dispersão criada por inúmeras palavras únicas que aparecem em apenas alguns documentos. Além disso, a falta de coocorrências de palavras torna um desafio para o sistema extrair informações semânticas das palavras. Desta forma, esta técnica consiste em gerar representações numéricas do conjunto de dados textual e utilizar um algoritmo de aprendizado não supervisionado de clusterização para dividir o conjunto em grupos. Os algoritmos de clusterização fazem o uso da representação vetorial e métricas de distância ou similaridade entre vetores para agrupar distintas entradas em *Clusters*. Considera-se um tópico, um agrupamento de textos similares. Espera-se que com uma representação vetorial adequada e a partir de uma modelagem de tópicos eficiente, os textos contidos em cada grupo contenham semelhanças semânticas ou sintáticas definidas a partir da construção dos *embeddings*.

Esta pesquisa utiliza uma metodologia de modelagem de tópicos aprimorada, conforme proposto em [Asyaky e Mandala 2021]. A técnica se baseia na representação textual por meio de *embeddings*, redução de dimensionalidade com UMAP (*Uniform Manifold Approximation and Projection for Dimension Reduction*) [McInnes, Healy e Melville 2020] e *clustering* (agrupamento) por meio do algoritmo HDBSCAN (*Hierarchical Density-Based Spatial Clustering of Applications with Noise*)

[McInnes e Healy 2017].

UMAP é uma técnica de redução de dimensionalidade cujo objetivo é representar um espaço de maior número de dimensões em um novo espaço com um número de dimensões reduzido, preservando características locais e globais. Técnicas de redução de dimensionalidade são utilizadas para diminuir o número de características em um conjunto de dados, auxiliando algoritmos de aprendizado de máquina na realização de tarefas de classificação, regressão e agrupamento. As principais vantagens da técnica UMAP em relação a outras técnicas semelhantes são a velocidade de execução, escalabilidade, flexibilidade em termos de parâmetros ajustáveis e sua característica de preservar a estrutura global e local presentes no conjunto de dados.

O algoritmo de agrupamento HDBSCAN foi utilizado nesta pesquisa em conjunto com o UMAP. Ele é aplicado em tarefas de aprendizado não supervisionado e utiliza o conceito de densidade para agrupar vetores semelhantes. Sua metodologia faz uso da ideia de hierarquia no espaço vetorial, promovendo a robustez dos resultados em espaços com densidades distintas. Suas principais características são: ser tolerante a ruídos e possuir um número reduzido de parâmetros para configurar sua execução. Este método é utilizado não apenas para tarefas de *clustering*, mas também para detecção de pontos destoantes (*outliers*) e reconhecimento de padrões.

Revisão da Literatura acerca da Integração de Aprendizado por Reforço Profundo com Indicadores baseados em Notícias Financeiras e Indicadores Fundamentalistas

A Revisão Sistemática da Literatura contida nesta seção foi publicada na *9th International Conference on Computer Science and Computational Intelligence (CSCI) 2024* com o título *Systematic Review of Portfolio Optimization in the Brazilian Financial Market: Integrating News and Fundamental Indicators with Deep Reinforcement Learning* [Silva et al. 2025].

3.1 Revisão Sistemática de Literatura

A Revisão Sistemática da Literatura (RSL) é um método de busca que visa identificar, analisar e sintetizar trabalhos relevantes sobre um tema [Kitchenham 2004]. Além disso, ao realizar a RSL é possível identificar os métodos utilizados pelos pares para aquele tema, bem como lacunas a serem exploradas, validando assim a importância do estudo. Diante disso, a RSL é um trabalho secundário que visa subsidiar outros estudos. Portanto, para a realização de levantamento bibliográfico com relação à temática de pesquisa apresentada para este trabalho, utilizou-se da ferramenta Parsifal (<https://parsifal/>), que se baseia no modelo apresentado por [Kitchenham 2004].

A Revisão Sistemática da Literatura compreende três fases, as quais são Planejamento, Condução e Publicação dos Resultados, conforme pode ser visualizado na Figura 3.1, a qual mostra as fases e passos da RSL, os quais foram empregados nesta RSL. Na fase de Planejamento, definiu-se o objetivo da revisão detalhado a seguir, em seguida estabeleceu-se um protocolo que compreende definir as questões de pesquisa, fontes de pesquisa, *string* de busca, critérios de qualidade e quais campos de dados seriam extraídos; por fim, foi feita avaliação do protocolo, que consiste em testar o protocolo. O teste

foi realizado verificando a viabilidade da execução da RSL a partir de um conjunto de estudos retornados pelo protocolo. A partir disso, os itens do protocolo puderam ser refinados iterativamente. Na fase de condução, foram identificados os estudos primários; posteriormente, selecionou-se estes estudos de acordo com os critérios de exclusão e qualidade estabelecidos, foram extraídos os dados estabelecidos para extração; em seguida, os dados foram sintetizados. Na fase de Publicação dos Resultados, realizou-se a descrição e avaliação dos resultados, os quais serão apresentados nas subseções seguintes.

A Revisão Sistemática da Literatura adotada neste estudo compreende três fases principais: Planejamento, Condução e Publicação dos Resultados, conforme ilustrado na Figura 3.1. A fase de Planejamento consistiu na definição do objetivo da revisão, detalhado posteriormente neste trabalho, e na elaboração de um protocolo detalhado. Este protocolo abrangeu a formulação das questões de pesquisa, a seleção das fontes de informação, a construção da *string* de busca, a definição dos critérios de inclusão/exclusão e qualidade, e a especificação dos dados a serem extraídos dos estudos selecionados. O protocolo foi avaliado e refinado iterativamente, buscando assegurar a viabilidade e a qualidade da RSL. A fase de Condução envolveu a busca pelos estudos primários nas fontes de informação selecionadas, seguida pela aplicação dos critérios de inclusão/exclusão e qualidade para a seleção dos estudos mais relevantes. Os dados relevantes foram extraídos dos estudos selecionados e, posteriormente, sintetizados. Finalmente, a fase de Publicação dos Resultados engloba a descrição e a análise crítica dos principais resultados da RSL, os quais serão apresentados nas subseções seguintes.

Esta revisão sistemática tem como objetivo analisar de forma abrangente a otimização de carteiras no mercado financeiro brasileiro, incorporando especificamente notícias e indicadores fundamentalistas por meio de técnicas de Aprendizado por Reforço Profundo (ARP). Visto que os indicadores financeiros utilizados para análise fundamentalista apontam a saúde e o desempenho financeiro de uma empresa e podem ajudar os investidores a tomar decisões informadas, ao mesmo tempo que gerem os riscos de forma eficaz para melhorar o desempenho geral da carteira [Edirisinghe e Zhang 2007]. Para melhor compreensão, a Figura 3.2 apresenta os principais temas relacionados ao objetivo da RSL. Ao examinar a literatura existente e os estudos empíricos, esta revisão procura identificar as principais tendências, desafios e oportunidades neste campo em evolução. Além disso, visa oferecer percepções sobre o impacto potencial da integração dessas tecnologias avançadas em estratégias de gestão de portfólio no mercado financeiro brasileiro.

3.2 Protocolo de planejamento da RSL

O protocolo de planejamento adotado para esta RSL foi o seguinte:

- *Questões de Pesquisa*

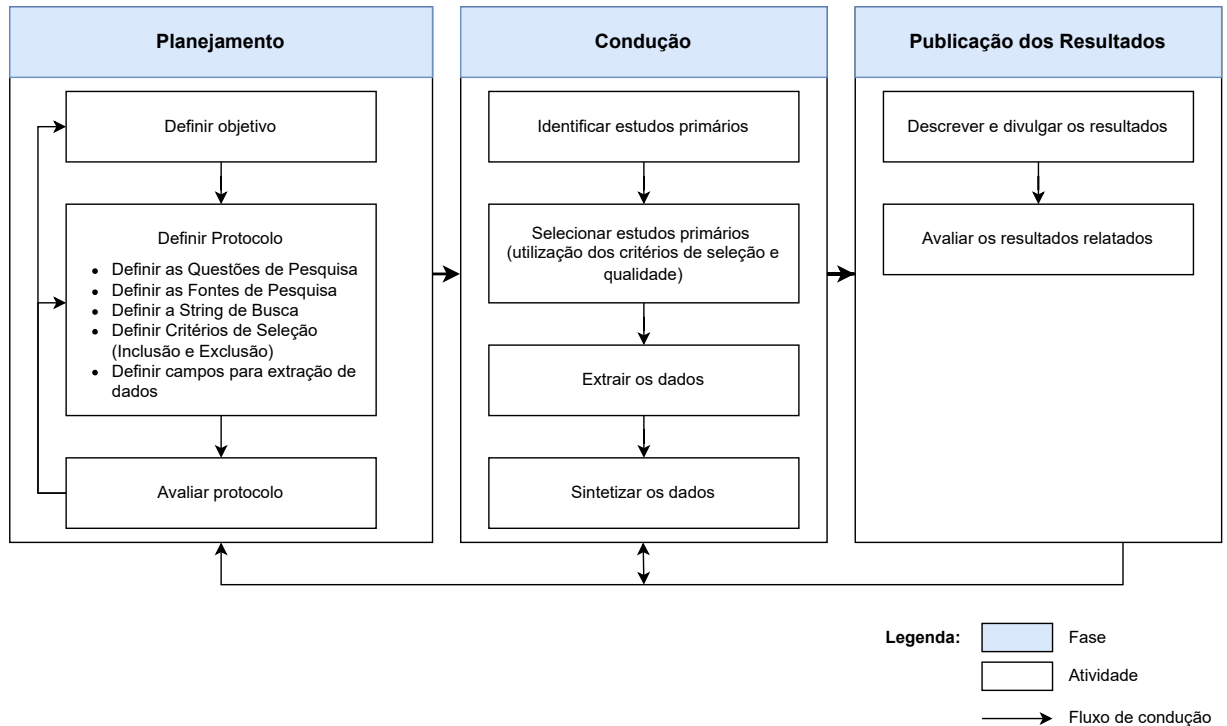


Figura 3.1: Fases e passos da Revisão Sistemática da Literatura (Adaptado de [Nakagawa et al. 2017]).

- *Q1*. Existem estudos que utilizam aprendizado por reforço profundo para otimização de portfólio de ativos financeiros?
 - *Q2*. Quais os métodos utilizados pelos estudos que aplicam aprendizado por reforço profundo?
 - *Q3*. Existem estudos que exploram a utilização de notícias para construção de portfólio?
 - *Q4*. Existem estudos que abordam a utilização de aprendizado por reforço profundo juntamente com indicadores baseados em notícias em língua portuguesa para otimização de portfólio de ativos financeiros?
 - *Q5*. Quais indicadores de análise quantitativa fundamentalista utilizados para otimização de portfólio de ativos financeiros?
- Bases utilizadas para busca: Quatro bases conhecidas foram selecionadas para a realização da busca para este estudo, sendo elas: IEEE Xplore Digital Library¹,

¹<https://ieeexplore.ieee.org>



Figura 3.2: Principais temas relacionados ao objetivo da Revisão Sistemática da Literatura.

ACM Digital Library², Scopus³, e Google Scholar⁴. É importante reconhecer que, embora o Google Scholar seja uma ferramenta abrangente e útil para a exploração inicial de um tema, sua utilização em revisões sistemáticas apresenta limitações. A natureza dinâmica dos resultados do Google Scholar, que se ajustam ao perfil de cada usuário, pode comprometer a reprodutibilidade da busca. No entanto, considerando a natureza exploratória desta dissertação e a busca por diferentes tipos de documentos acadêmicos, a utilização do Google Scholar se justifica como um complemento à busca em bases de dados tradicionais.

- *Crítérios de inclusão*
 - *I1*. Trabalhos que usam ARP para otimização de portfólio
 - *I2*. Trabalhos que usam notícias para otimização de portfólio
 - *I3*. Trabalhos que utilizam análise fundamentalista para otimização de portfólio
- *Crítérios de exclusão*
 - *E1*. Trabalhos fora do escopo

²<https://dl.acm.org/>

³<https://www.scopus.com/>

⁴<https://scholar.google.com/>

Tabela 3.1: Número de artigos selecionados com base em critérios de exclusão.

Fonte	Número de artigos	
	Coletados	Selecionados
ACM	40	2
Scopus	50	7
Google Scholar	614	15
IEEE Xplore	17	11

- E2. Trabalhos duplicados
- E3. Trabalhos anteriores a 2019
- E4. Trabalhos que não podem ser acessados gratuitamente
- E5. Trabalhos em línguas diferentes do inglês e do português
- *Critério de qualidade*
 - CQ1. Artigos que abordam a otimização de portfólio utilizando, pelo menos, um dos seguintes elementos: notícias, indicadores fundamentalistas ou ARP
 - CQ2. Artigos com descrição clara das técnicas utilizadas
 - CQ3. Artigos que realizam comparação do método proposto com outras abordagens, oferecendo uma análise crítica sobre a eficácia das técnicas utilizadas
- *Campos para extração de dados*
 - D1. Algoritmo de aprendizado por reforço profundo aplicado
 - D2. Método de uso das notícias
 - D3. Indicadores fundamentalistas empregados
 - D4. Mercado adotado para estudo

3.3 Execução

Em 4 de setembro de 2023, a busca resultou em 721 artigos, utilizando a seguinte *string* nas bases determinadas: (“*Financial market*” OR “*Finance*” OR “*Financial*”) AND (((“*News*” OR “*Portuguese news*” OR “*Portuguese*”) AND (“*Natural Language Processing*” OR “*NLP*”)) OR (“*Investment portfolio*” OR “*Portfolio management*” OR “*Portfolio optimization*”)) AND (“*Deep Reinforcement Learning*” OR “*DRL*” OR “*Fundamental analysis*” OR “*Quantitative analysis*” OR “*Quantitative indicators*”). Destes, 638 foram rejeitados com base em seus títulos, enquanto 17 eram duplicados e 12 não puderam ser acessados. Posto isto, foram escolhidos 35 artigos para posterior análise. Na Tabela 3.1 é possível observar a quantidade de trabalhos selecionados de acordo com os critérios de exclusão de cada base de dados.

3.4 Resultados da RSL

Com relação à distribuição de publicações ao longo dos anos, dentre os artigos selecionados, 6 foram publicados em 2020, 11 em 2021, 10 em 2022, e 8 em 2023. Vale destacar que esse tema específico tem recebido atenção significativa da comunidade científica.

Os artigos selecionados foram publicados em diversos jornais e conferências, a maioria deles específicos da área de inteligência artificial. Dos 35 trabalhos, 18 foram publicados em periódicos e 17 em anais de congressos. A revista IEEE Access teve o maior número de publicações, com 3, seguida pela conferência ICASP com 2. As Figuras 3.3 e 3.4 mostram a distribuição das publicações em periódicos e conferências, respectivamente.

Ademais, o número de citações de outros estudos é uma métrica significativa para aferir a relevância de um artigo, pois reflete seu impacto e influência na comunidade acadêmica [Chatterjee, Ghosh e Chakrabarti 2016]. Assim, a Tabela 3.2 apresenta o número de citações dos 11 artigos mais referenciados juntamente com seus respectivos anos de publicação. O estudo de [Soleymani e Paquet 2020], que propõe uma estrutura de gerenciamento de portfólio combinando aprendizado por reforço profundo com um autoencoder empilhado restrito e uma rede neural convolucional, destaca-se como o mais citado no período analisado, atraindo grande interesse da comunidade científica. Por outro lado, [Du e Tanaka-Ishii 2020] também foi considerado pertinente pelos estudiosos ao usar uma representação vetorial da influência das notícias sobre as ações por meio de uma estrutura de *deep learning* para otimização de portfólio. Observou-se, ainda, que a média de citações dos artigos selecionados é de 8.69.

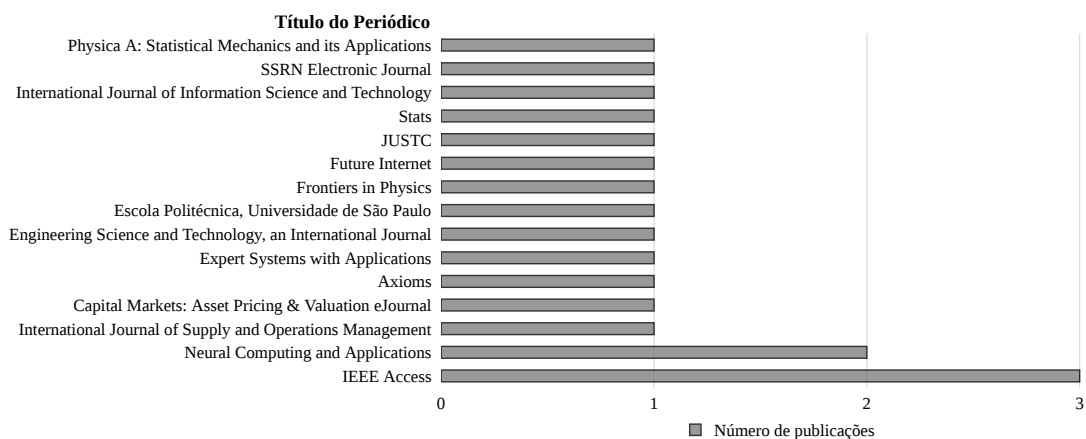


Figura 3.3: Distribuição de publicações dos periódicos.

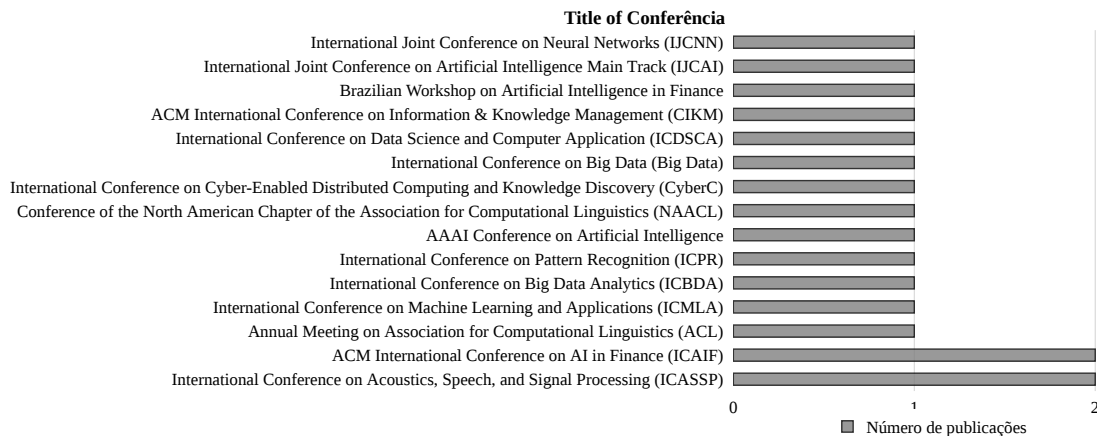


Figura 3.4: Distribuição de publicações das conferências.

Tabela 3.2: Sumário dos 11 artigos mais citados.

Artigo	Ano	Citações
<i>Financial portfolio optimization with online deep reinforcement learning and restricted stacked autoencoder—DeepBreath</i> [Soleymani e Paquet 2020]	2020	60
<i>Stock Embeddings Acquired from News Articles and Price History, and an Application to Portfolio Optimization</i> [Du e Tanaka-Ishii 2020]	2020	36
<i>Market sentiment-aware deep reinforcement learning approach for stock portfolio allocation</i> [Koratamaddi et al. 2021]	2021	33
<i>A deep Q-learning portfolio management framework for the cryptocurrency market</i> [Lucarelli e Borrotti 2020]	2020	31
<i>DeepTrader: A Deep Reinforcement Learning Approach for Risk-Return Balanced Portfolio Management with Market Conditions Embedding</i> [Wang et al. 2021]	2021	31
<i>AlphaPortfolio: Direct Construction Through Deep Reinforcement Learning and Interpretable AI</i> [Cong et al. 2020]	2020	24
<i>Deep Reinforcement Learning Agent for S&P 500 Stock Selection</i> [Huotari, Savolainen e Collan 2020]	2020	9
<i>Quantitative Day Trading from Natural Language using Reinforcement Learning</i> [Sawhney et al. 2021]	2021	9
<i>Application of Features and Neural Network to Enhance the Performance of Deep Reinforcement Learning in Portfolio Management</i> [Gu, Jiang e Su 2021]	2021	8
<i>Novel Deep Reinforcement Algorithm With Adaptive Sampling Strategy for Continuous Portfolio Optimization</i> [Huang, Miao e Hsiao 2021]	2021	8
<i>Detecting and adapting to crisis pattern with context based Deep Reinforcement Learning</i> [Benhamou et al. 2021]	2021	7

3.4.1 Análise dos artigos selecionados

A análise dos trabalhos foi conduzida com base em uma classificação que contempla o uso de aprendizado por reforço profundo, indicadores fundamentalistas, e indicadores baseados em notícias, bem como suas interseções. Nos artigos avaliados, não foram identificados estudos que combinem indicadores fundamentalistas com indicadores baseados em notícias. Todos os trabalhos atendem aos critérios de qualidade estabelecidos, utilizando pelo menos um dos elementos que contribuem para o objetivo geral desta RSL: notícias, indicadores fundamentalistas ou ARP. Além disso, eles descrevem as técnicas aplicadas de forma clara e comparam o método proposto com outras abordagens. A Tabela 3.3 oferece uma visão geral dos artigos analisados, apresentando para cada um deles: a classificação temática, os indicadores utilizados, o mercado estudado e um detalhamento dos indicadores fundamentalistas empregados. Essa tabela, portanto, fornece informações relevantes para responder à questão de pesquisa *Q5*, que busca identificar quais indicadores fundamentalistas têm sido utilizados em trabalhos sobre otimização de carteiras. Foi observada uma frequência do uso de algoritmos como *Deep Q-learning*

(DQN), *Deep Deterministic Policy Gradient* (DDPG), *Proximal Policy Gradient* (PPO) e suas variações, nos trabalhos que utilizam ARP.

ARP

A análise dos 19 artigos classificados na categoria de Aprendizado por Reforço Profundo confirma a pertinência da questão de pesquisa *Q1*, evidenciando que essa abordagem tem sido explorada para otimização de carteiras de ativos financeiros. Esses estudos demonstraram desempenho superior em comparação com métodos estado da arte e estratégias tradicionais como *Buy and Hold*. As métricas de avaliação mais utilizadas foram *Sharpe Ratio* e *Maximum Drawdown*. Conforme mostrado na Tabela 3.3, a maioria dos artigos enfocou o mercado dos Estados Unidos, seguido pelo mercado chinês. No entanto, outros trabalhos aplicaram esses métodos a mercados como da Nigéria, criptomoedas, Forex (mercado de câmbio) e *commodities* como ouro.

Concernente à questão *Q2*, a análise dos estudos revelou a utilização de diferentes classes de algoritmos de Aprendizado por Reforço Profundo para otimização de carteiras. As abordagens identificadas incluem algoritmos de gradiente de política, como o PPO, algoritmos baseados em valor, como o DQN, e algoritmos ator-crítico, como o DDPG.

O artigo [Costa e Costa 2023] foi o único desta classificação que empregou o mercado financeiro, apresentando o PortfolioOptimizationEnv (POE), um ambiente de otimização de portfólio *open-source* (aberto publicamente) que é compatível com métodos modernos *reinforcement learning*. Este ambiente foi desenvolvido com base no modelo FinRL e segue a estrutura da OpenAI Gym⁵. Além disso, possui integração com diversos repositórios existentes de algoritmos ARP.

Indicadores a partir de notícias

Os três trabalhos desta categoria analisam o sentimento em notícias para criar indicadores de gestão de portfólio, equilibrando risco e retorno. Com isso, eles respondem à questão *Q3*, que investiga a existência de estudos que utilizam notícias na construção de portfólios. Para isso, utilizam a ferramenta BERT (*Bidirectional Encoder Representations from Transformer*). Os resultados indicam que esses métodos superaram as bases de comparação. No estudo de [Colasanto et al. 2022], o modelo que incorpora o índice de polarização baseado em notícias obteve um *Sharpe Ratio* melhor em comparação com aquele que não o utilizou, potencialmente devido ao melhor controle de risco ao usar notícias como indicadores.

⁵<https://github.com/openai/gym>

Tabela 3.3: Sumário dos artigos selecionados.

Referência	ARP	Indicadores utilizados	Detalhamento de indicadores	Mercado aplicado
[Naseri, Rafee e Moghadam 2020]		Indicadores quantitativos fundamentalistas	Índice de Liquidez, Índice de Alavancagem (Estrutura de Capital), Cobertura de Juros, Dívida, Margem Líquida de Lucro, Retorno sobre o Patrimônio Líquido, Retorno sobre Ativos	-
[Du e Tanaka-Ishii 2020]		Indicadores a partir de notícias	Uso de notícias para determinar quando o preço da ação aumentará ou cairá de acordo com o preço do dia anterior	EUA
[Cong et al. 2020]	X	-	<i>Earnings to price ratio (EP)</i> , <i>Dividend yield (DY)</i> e <i>Trading volume (Vol)</i>	EUA
[Huotari, Savolainen e Collan 2020]	X	ARP com indicadores fundamentalistas		EUA
[Lucrelli e Borroiti 2020]	X	-	Vários indicadores técnicos	Criptomoedas
[Soleymani e Paquet 2020]	X	-	Vários indicadores técnicos	-
[Wang e Yu 2021]	X	-	ARP com indicadores fundamentalistas	EUA
[Gu, Jiang e Su 2021]	X	-	ARP com indicadores fundamentalistas	Criptomoedas
[Benhamou et al. 2021]	X	ARP com indicadores fundamentalistas		-
[Koratamaddi et al. 2021]	X	ARP com indicadores baseados em notícias		EUA
[Silva 2021]	X	ARP com indicadores baseados em notícias		Brasil
[Wang et al. 2021]	X	-	Indicadores quantitativos fundamentalistas	EUA e China
[Yan et al. 2021]		-	Indicadores quantitativos fundamentalistas	EUA
[Sawhney et al. 2021]	X	ARP com indicadores baseados em notícias		EUA e China
[Theodorou et al. 2021]	X	ARP com indicadores baseados em notícias		-
[Daiya e Lin 2021]	X	ARP com indicadores baseados em notícias		EUA
[Huang, Miao e Hsiao 2021]	X	-	Return on Equity (ROE), Marge de Lucro, Asset Turnover, Leverage Ratio, Operating Income, Lucro Bruto	EUA e China
[Zhuang, Chen e Qiu 2022]	X	-	Sentimento de notícias e de tweets	EUA
[Gao et al. 2022]	X	-	Sentimento de notícias e de tweets e Indicadores técnicos: Moving average convergence divergence (MACD), Average Directional Index (ADX), Relative Strength Index (RSI) e Average True Range (ATR)	EUA e China
[Colasanto et al. 2022]	X	Indicadores a partir de notícias		-
[Paiva et al. 2022]	X	ARP com indicadores baseados em notícias		EUA
[Zhu e Zhu 2022]	X	ARP com indicadores baseados em notícias		EUA
[Taguchi, Sakaji e Izumi 2022]	X	-	Indicadores a partir de notícias	EUA e Criptomoeda (Bitcoin)
[Khemilchi e al. 2022]	X	-	Indicadores a partir de notícias	EUA
[Liu et al. 2022]	X	-	Indicadores a partir de notícias	EUA
[Niu, Li e Li 2022]	X	Indicadores técnicos		Criptomoeda (Bitcoin) e commodity (ouro)
[Li et al. 2022]	X	-	Indicadores técnicos	EUA e China
[Costa e Costa 2023]	X	-	Indicadores técnicos	EUA
[Jin 2023]	X	-	Indicadores técnicos	EUA
[Yang et al. 2023]	X	ARP com indicadores baseados em notícias	Trajetória de estado orientada por notícias de destaque	EUA e Criptomoeda (Bitcoin)
[Olabode, Aremu e BAMISAIYE 2023]	X	-	ADR (Average Daily Range - Variação Diária Média), OBOS (Overbought/Oversold - Sobrecomprado/Sobrevendido) e ADL (Accumulation/Distribution Line - Linha de Acumulação/Distribuição)	Criptomoeda (Bitcoin) e commodity (ouro)
[Bat et al. 2023]	X	-	Indicadores técnicos	EUA e China
[Zhou et al. 2023]	X	-	Indicadores técnicos	Brasil
[Lin et al. 2023]	X	-	Indicadores técnicos	EUA
[Brimi e Tantiari 2023]	X	-	Indicadores técnicos	Nigéria
				China
				EUA e China
				Forex

Indicadores fundamentalistas

Dois dos artigos selecionados investigaram o uso da análise fundamentalista para otimização de portfólios, incorporando indicadores como Retorno sobre Patrimônio Líquido (ROE - *Return on Equity*), margem de lucro e índice de alavancagem. Em [Yan et al. 2021], os autores propõem o conceito de “Redes Fundamentais” (FN - *Fundamental Networks*) que integram indicadores fundamentais para melhorar o desempenho e a gestão de riscos. Esta abordagem aborda algumas limitações do modelo de média-variância amplamente utilizado, demonstrando maior eficiência, robustez e oferecendo portfólios bem diversificados. Como resultado, a FN superou o desempenho da carteira Markowitz utilizada como referência.

ARP com indicadores baseados em notícias

Este grupo de trabalhos utiliza técnicas de ARP em conjunto com indicadores derivados de notícias, em sua maioria gerados por meio de análise de sentimento. Três dos sete artigos desta classe também envolvem análise de tweets, enquanto três incorporam indicadores técnicos, como *Moving average convergence divergence* (MACD) e *Relative Strength Index* (RSI). As métricas mais recorrentes foram o *Sharpe Ratio* e o Retorno Total, evidenciando o uso da estratégia *Buy and Hold* para comparação. Observou-se que a integração de notícias leva a um melhor controle de risco [Silva 2021] e retornos possivelmente mais elevados [Koratamaddi et al. 2021].

O artigo [Silva 2021] é o único a focar no mercado brasileiro, explorando o uso de algoritmos de aprendizado por reforço profundo para desenvolver um sistema de negociação automatizado que combina previsão de preços e análise de sentimento de notícias, com o objetivo de maximizar retornos enquanto gerencia os riscos. Em relação à questão Q4, que investiga a existência de estudos que utilizam ARP juntamente com indicadores baseados em notícias em língua portuguesa para a otimização de ativos financeiros, este trabalho aplica essa abordagem em um sistema de negociação automatizada voltado para uma única ação, mas com potencial para adaptação a múltiplas ações.

ARP com indicadores fundamentalistas

Os três estudos que compõem esta classe demonstraram a superioridade dos métodos abordados, os quais integram ARP e indicadores fundamentalistas, em termos de *Sharpe Ratio* e retorno em relação aos *benchmarks* considerados. Isto posto, a abordagem expressa por [Huotari, Savolainen e Collan 2020] obteve um retorno estatisticamente significativo de 329.8% e *Sharpe Ratio* de 0.96.

Posicionamento da pesquisa

As pesquisas mais pertinentes ao objetivo geral deste estudo incluem o trabalho de [Koratamaddi et al. 2021], que emprega ARP com indicadores derivados de notícias e do Twitter para a alocação de carteira de ações, e o estudo de [Silva 2021], que utiliza ARP com indicadores provenientes de notícias em língua portuguesa para desenvolver um sistema de negociação automática de ações.

Integração de Aprendizado por Reforço Profundo com Indicadores baseados em Notícias Financeiras e Indicadores Fundamentalistas

4.1 Desenho do Experimento

O objetivo deste trabalho é avaliar uma abordagem para otimização de portfólio de ativos aplicando algoritmos de aprendizagem por reforço profundo utilizando indicadores fundamentalistas e baseados em notícias em língua portuguesa, com foco no mercado financeiro brasileiro.

O método proposto, conforme mostrado na Figura 4.1, é dividido em 6 fases. A fase I envolve a aquisição de dados de notícias em português, histórico de preços e indicadores fundamentalistas dos ativos do mercado financeiro brasileiro. Na fase II, é realizada a extração de informações de notícias, como análise de sentimentos e extração de entidades. A fase III concentra-se no cálculo de indicadores de sentimento padrão e variações com filtros, que serão detalhados nas subseções a seguir. A fase IV trata da análise estatística desses indicadores. Na fase V, são implementados e executados agentes de ARP utilizando os dados obtidos. Finalmente, a fase VI abrange a análise estatística dos resultados experimentais. Todas as atividades da pesquisa foram realizadas usando a linguagem Python¹ e todos os dados obtidos foram armazenados em um banco de dados PostgreSQL². Nas seções seguintes, cada fase e seus respectivos passos serão indicados pelo símbolo “()”.

4.1.1 Aquisição de dados

Na Fase I do experimento, coletou-se as notícias e dados da B3. Este processo será descrito nas subseções a seguir.

¹<https://www.python.org/>

²<https://www.postgresql.org/>

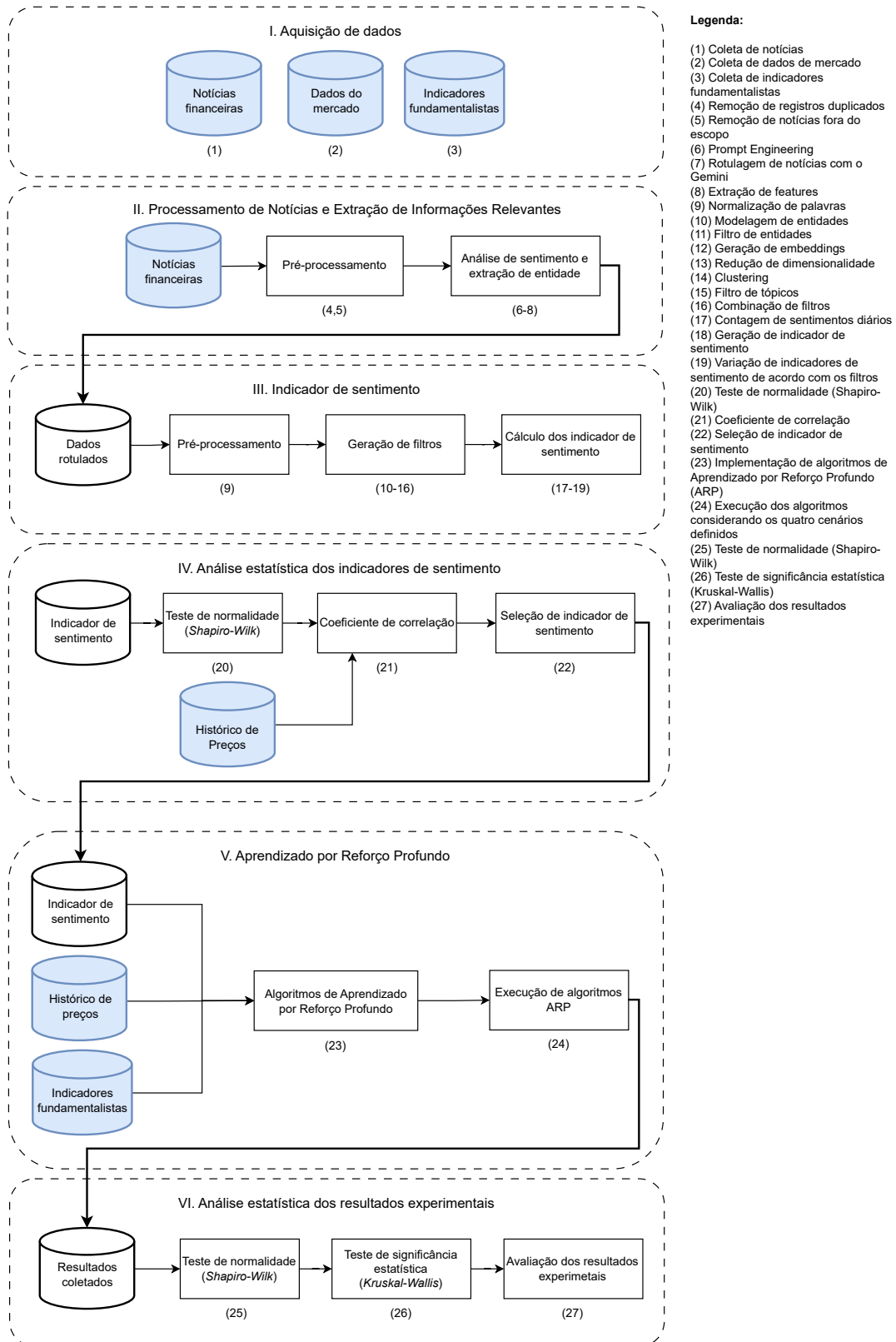


Figura 4.1: Desenho do experimento.

B3

Os dados foram extraídos da B3 por meio da biblioteca Yfinance⁵, que utiliza APIs públicas fornecidas pelo Yahoo, Inc (2). Foram adquiridos o volume de negociação e o histórico de preços de ações que integram o índice Ibovespa, principal índice do mercado de ações do Brasil, composto pelas ações mais negociadas da bolsa de valores brasileira (B3) [B3]. Isso inclui preço de abertura, preço máximo, preço mínimo, preço de fechamento e preço de fechamento ajustado. Esses dados foram obtidos para o intervalo de tempo referente à fonte da notícia. O preço de fechamento ajustado representa o valor de fechamento que contabiliza desdobramentos de ações, e dividendos. O desdobramento de uma ação, também conhecido como *stock split*, ocorre quando esta é dividida em “n” novas ações de modo que o valor do investimento total não seja alterado. Desta forma, ocorrendo um *split* de 2 por 1 em uma ação, cada ação será dividida em duas, e então a ação terá metade do valor inicial. A principal motivação para empresas realizarem um *split* reside em aumentar a liquidez das ações. Ao reduzir o preço por ação, a empresa torna o investimento mais acessível a um público mais amplo, incluindo pequenos investidores, o que pode contribuir para um aumento na demanda e na negociação das ações. Já os dividendos representam a distribuição de parte dos lucros apurados por uma empresa aos seus acionistas. Essa remuneração pode se dar na forma de dinheiro, a mais comum, ou através da emissão de novas ações, configurando-se como uma forma de recompensar os investidores pelo capital investido e incentivar a manutenção do investimento na empresa. Isso posto, os cálculos deste trabalho se baseiam exclusivamente no preço de fechamento ajustado.

Indicadores fundamentalistas

Os indicadores fundamentalistas das ações que compõem o índice Ibovespa foram obtidos do site Status Invest⁶ (3). Os dados coletados cobrem capitalização de mercado, liquidez, alavancagem e indicadores de rentabilidade para os constituintes do Ibovespa de 2011 a 2023. No entanto, algumas empresas têm dados incompletos para o período estudado, resultando em valores nulos no conjunto de dados. Os indicadores coletados são baseados em uma frequência trimestral. Para este estudo, utilizou-se os seguintes indicadores-chave: *Dividend Yield* (DY), indicando a renda gerada por ação; relação Preço/Lucro (P/L), mostrando a relação entre o preço atual da ação e o lucro da empresa; Retorno sobre o Patrimônio Líquido (ROE), medindo a rentabilidade da empresa em relação ao patrimônio líquido dos acionistas; e Margem Bruta, refletindo a eficiência da gestão de custos na produção e vendas.

⁵<https://pypi.org/project/yfinance/>

⁶<https://statusinvest.com.br/>

4.1.2 Extração de informações de notícias

Prompt Engineering

O *Large Language Model* (LLM) Gemini Pro, desenvolvido pelo Google [Google 2023], é um serviço avançado de Inteligência Artificial (IA) usado na Fase II para classificar o sentimento das notícias e extrair a principal entidade mencionada nelas. Neste estudo, utilizou-se prompts escritos em língua portuguesa, com o objetivo de avaliar o sentimento das notícias do cenário financeiro brasileiro e identificar a principal entidade a que essas notícias se referem, para posterior análise estatística (6).

O *prompt* inicial descreveu um cenário em que o modelo assumia o papel de um analista de investimentos atuando no setor financeiro brasileiro. Nesse cenário, o modelo forneceu sua resposta em forma de um vetor (*array*) rotulado como “label”. A Tabela 4.1 exibe as iterações realizadas para refinar o prompt. Para a entrada do modelo, foram concatenados tanto o título quanto o corpo de cada artigo de notícia. Enquanto o título oferece um resumo da notícia, a incorporação do corpo permite que o modelo compreenda as nuances contextuais, potencialmente aprimorando a precisão de suas respostas.

O prompt inicial apresentou resultados promissores em termos de análise de sentimento e do formato de saída solicitado. No entanto, ele teve dificuldades em identificar a entidade principal mencionada na notícia, frequentemente focando em entidades secundárias. Além disso, nem todas as notícias giravam em torno de empresas; algumas se concentravam em países, como o Brasil. Para resolver isso, foi introduzida uma condição na segunda iteração: se a notícia se referisse ao Brasil, mas não a uma empresa específica, a entidade seria rotulada como “Brasil”. Contudo, isso levou a classificações incorretas, com notícias não relacionadas sendo etiquetadas como brasileiras. Na terceira iteração, duas condições foram implementadas: se a notícia não mencionasse uma empresa específica, a entidade seria identificada pelo país mencionado; caso contrário, seria rotulada como “Não identificada”. No entanto, os problemas persistiram, com algumas notícias que mencionavam o Brasil recebendo a etiqueta “Não identificada”, apesar de se referirem a países ou empresas de outras nacionalidades. Além disso, as notícias podem se referir não apenas a países, mas também a regiões.

Em contraste com as iterações anteriores que basearam-se *zero-shot learning* (que não requer exemplos anteriores), a quarta iteração adotou a abordagem *few-shot learning*, fornecendo à IA um exemplo para ajudá-la a compreender o contexto e o formato da resposta desejada. O prompt foi ajustado para garantir que, caso a notícia fizesse referência a empresas ou regiões, a entidade refletiria essa informação. Todavia, estas modificações introduziram viés ao restringir o modelo a condições específicas, ignorando potenciais cenários fora delas. Assim, torna-se evidente a importância de elaborar um prompt sem viés e objetivo.

Na iteração final, as condições foram removidas e o prompt passou a solicitar apenas a classificação das notícias e a identificação da entidade principal em forma de uma matriz. Essa abordagem simplificada conseguiu gerar a resposta desejada, facilitando a análise subsequente e a manipulação da rotulação dos dados (7).

Modelagem de tópicos

A maioria dos modelos de embeddings é treinada em corpora de dados em inglês [Wang 2022]. No entanto, para expandir seu alcance para outros idiomas, os autores de [Wang et al. 2024] disponibilizaram o modelo Multilingual E5 (*multilingual-e5-large*), que gera *embeddings* de alta qualidade em um grande conjunto de línguas, incluindo a língua portuguesa. O *pipeline* (fluxo) de modelagem de tópicos empregado é representado pelas seguintes etapas:

1. Concatenar o título e a descrição da notícia; gerar a representação de incorporação (12);
2. Reduzir a dimensionalidade usando UMAP (13);
3. Agregar *embeddings* de notícias semelhantes usando HDBSCAN (14).

A saída do *pipeline* fornece uma atribuição de *cluster* do HDBSCAN, representando o tópico principal de cada artigo de notícia no conjunto de dados. A saída do *pipeline* é ilustrada na Figura 4.3, com 1.095 tópicos representados por cores distintas em uma visualização bidimensional criada usando UMAP para fins de visualização. Cada cor corresponde a um tópico, e aqueles semanticamente mais próximos têm intensidades de cor similares.

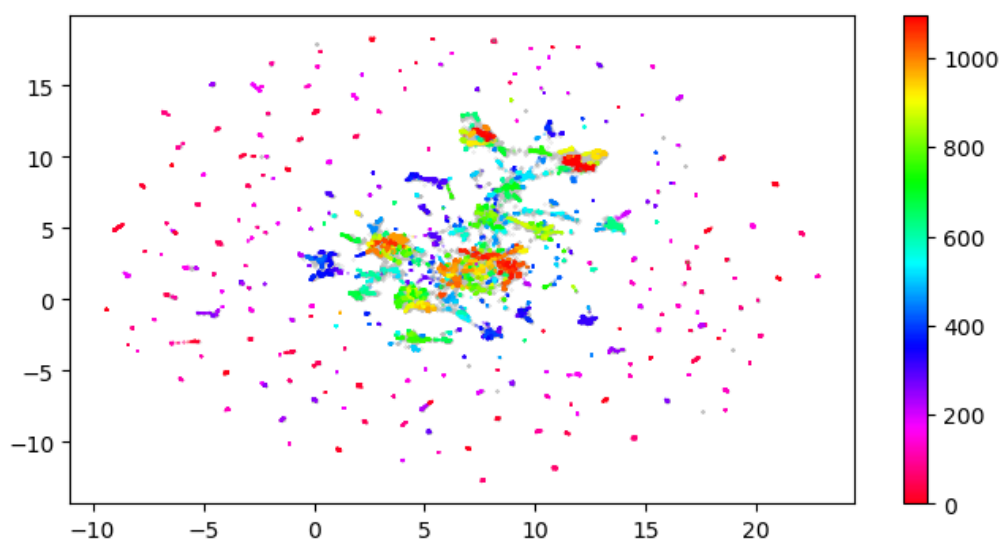


Figura 4.3: Espaço vetorial dos *embeddings* das notícias.

Tabela 4.1: Refinamentos do prompt

Iteração	Prompt
1	<p>Você é um economista e analista de investimentos. Classifique essa notícia do mercado financeiro em positivo, negativo e neutro, considerando apenas o contexto da notícia, e especifique de qual empresa a notícia se refere. Retorne apenas a resposta (sem justificção ou explicação) na forma de um array chamado “label”, no qual a primeira posição é a classificação da notícia e a segunda posição é qual empresa a notícia se refere.</p>
2	<p>Você é um economista e analista de investimentos. Classifique essa notícia do mercado financeiro em positivo, negativo e neutro, considerando apenas o contexto da notícia, e especifique de qual empresa a notícia se refere. Retorne apenas a resposta (sem justificção ou explicação) na forma de um array chamado “label”, no qual a primeira posição é a classificação da notícia e a segunda posição é qual empresa a notícia se refere (Caso não se refira a nenhuma empresa, mas se refira ao Brasil atribua o valor “Brasil”).</p>
3	<p>Como economista e analista de investimentos, avalie o sentimento expresso na seguinte notícia do mercado financeiro brasileiro. Realize uma classificação em positivo, negativo ou neutro, considerando exclusivamente o contexto da notícia. Identifique também a empresa mencionada na notícia. Retorne a resposta de forma precisa em um array intitulado “label”, onde a primeira posição indica a classificação da notícia (positivo, negativo ou neutro) e a segunda posição especifica a empresa referida. Se a notícia não mencionar uma empresa específica, atribua o nome do país ao qual a notícia se refere. Caso não haja menção a uma empresa específica nem a um país, atribua o valor “Não identificado”.</p>
4	<p>Como economista e analista de investimentos do mercado financeiro brasileiro, avalie o sentimento expresso na seguinte notícia. Realize uma classificação em positivo, negativo ou neutro, considerando exclusivamente o contexto da notícia. Identifique também as empresas ou regiões da qual a notícia tem como foco. Retorne apenas a resposta na forma de um array chamado “label”, onde a primeira posição indica a classificação da notícia (positivo, negativo ou neutro) e, caso a notícia tenha como foco empresas ou regiões, a segunda posição do array recebe essa informação. Exemplo: Texto: A Hypera (BVMF:HYPE3) registrou lucro líquido das operações continuadas de 307,8 milhões de reais no quarto trimestre do ano passado, queda de 28,7% frente ao mesmo período de 2022, informou a companhia nesta quarta-feira. Resposta: label = [“Negativo”, “Hypera”]</p>
5	<p>Como economista e analista de investimentos do mercado financeiro brasileiro, avalie o sentimento expresso na seguinte notícia. Realize uma classificação em positivo, negativo ou neutro, considerando exclusivamente o contexto da notícia. Retorne apenas a resposta na forma de um array, onde a primeira posição indica a classificação da notícia (positivo, negativo ou neutro) e a segunda posição indica a principal entidade a qual a notícia se refere. Exemplo: Texto: A Hypera (BVMF:HYPE3) registrou lucro líquido das operações continuadas de 307,8 milhões de reais no quarto trimestre do ano passado, queda de 28,7% frente ao mesmo período de 2022, informou a companhia nesta quarta-feira. Resposta: [“Negativo”, “Hypera”]</p>

4.1.3 Indicadores baseados em notícias

A partir dos métodos de *prompt engineering* descritos anteriormente, o LLM Gemini utilizado para analisar cada artigo de notícias gerou como saída com um *array* (vetor) com dois elementos. O primeiro elemento representa o sentimento e o segundo identifica a entidade principal da notícia. No entanto, foi necessário realizar um processamento na saída para extrair o campo correto e remover qualquer ruído dos dados gerados.

O processamento envolveu as seguintes etapas: obter a posição do sentimento no *array* (primeira posição), converter o texto para letras minúsculas e, em seguida, eliminar pontuação, símbolos, números e *stopwords* usando a biblioteca Spacy [Honnibal et al. 2020] em português (8)(9). Se o resultado for uma *string* vazia, o texto é reprocessado removendo apenas símbolos e pontuação. Estas etapas são repetidas para a extração de entidade contida no *array* (segunda posição) (8). Este processamento da saída garante que o sentimento e a entidade principal extraída da notícia sejam precisos e livres de qualquer ruído ou interferência desnecessária que possa afetar a confiabilidade do indicador de sentimento. Ao extrair o sentimento, os resultados podem ser “Positivo”, “Negativo” ou “Neutro”. As entidades extraídas (segunda posição do *array*) incluem resultados como nomes de empresas, figuras políticas, organizações e outras entidades relevantes mencionadas na notícia.

O campo de sentimento é essencial para a construção do indicador, e a entidade pode ser utilizada para filtrar quais notícias devem ser incluídas no indicador.

Indicador de sentimento

O indicador de sentimento S foi construído contabilizando quantas notícias positivas, negativas e neutras ocorrem em um dia específico (17). A ausência de notícias em um determinado dia é representada por zero. A equação do indicador padrão é dada pela Eq. 4-1 (18).

$$S = \frac{\text{positivas} - \text{negativas}}{\text{positivas} + \text{negativas} + \text{neutras}} \quad (4-1)$$

Indicador de Sentimento por Média Móvel

Um indicador resistente à flutuação foi criado a partir do indicador de sentimento original, aplicando uma média móvel de 7 dias para suavizar as flutuações diárias. O indicador de média móvel de sentimento (*Moving Average Sentiment Indicator* - MASI) (Eq. 4-2) não apenas suaviza o sinal de sentimento original, mas também preserva o impacto dos sentimentos de notícias anteriores dentro do período de tempo definido (19).

$$\text{MASI}_t = \frac{1}{7} \sum_{i=t-6}^t S_i \quad (4-2)$$

Indicador de Sentimento por Média Móvel Exponencial

A média móvel ponderada exponencial (*Exponential Moving Average* - EMA) é uma técnica estatística empregada para suavizar dados de séries temporais e é especialmente valiosa para detectar tendências ao longo do tempo. Esta abordagem atribui mais peso aos pontos de dados recentes no processo de cálculo, sendo utilizada para construir o indicador de sentimento de média móvel exponencial (*Exponential Moving Average Sentiment Indicator* - EMASI) conforme mostrado pela Eq. 4-3 (19). No cálculo da média móvel ponderada exponencial, o parâmetro α especifica diretamente o fator de suavização. Este fator determina a rapidez com que as observações mais antigas perdem a sua influência. Com α definido como 0,3 na configuração experimental, o peso de cada observação diminui em 30% para cada aumento de unidade no tempo, priorizando observações recentes e dando-lhes mais peso no cálculo da média em comparação com as mais antigas. Um valor de α definido como 0,3 propõe um equilíbrio capturando a tendência do indicador de notícias e também suas flutuações de curto prazo. Valores menores de α como 0,1 produzem médias mais suaves, menos reativas ao curto prazo, enquanto valores maiores como 0,5 produzem médias mais reativas a flutuações recentes.

$$\text{EMASI}_t = \frac{S + (1 - \alpha) \cdot \text{EMASI}_{t-1}}{1 - \alpha} \quad (4-3)$$

As principais características dos três indicadores de sentimento são visíveis na amostra representada pela Figura 4.4. O indicador de sentimento padrão responde à variação diária do sentimento, enquanto o **MASI** incorpora informações de dias anteriores e o **EMASI** atribui maior importância às notícias recentes no cálculo. Pode-se observar em 4.4 o comportamento das curvas, onde o indicador de sentimento puro possui maior variação, capturando todas as flutuações de valores presentes ao passar dos dias, o indicador **MASI** produz uma curva mais suave, trazendo consigo a tendência das notícias em um prazo maior, e o indicador **EMASI** produz uma curva com características balanceadas entre a tendência central, e as flutuações de curto prazo dos sentimentos das notícias.

4.1.4 Filtros de artigos de notícias

O site Investing.com, utilizado como fonte de notícias para este estudo, agrega notícias de diversas plataformas de publicação, oferecendo perspectivas não apenas sobre o mercado de ações brasileiro, mas também sobre os mercados globais, incluindo, mas não se limitando aos Estados Unidos, Europa e Ásia. Embora essas atualizações globais

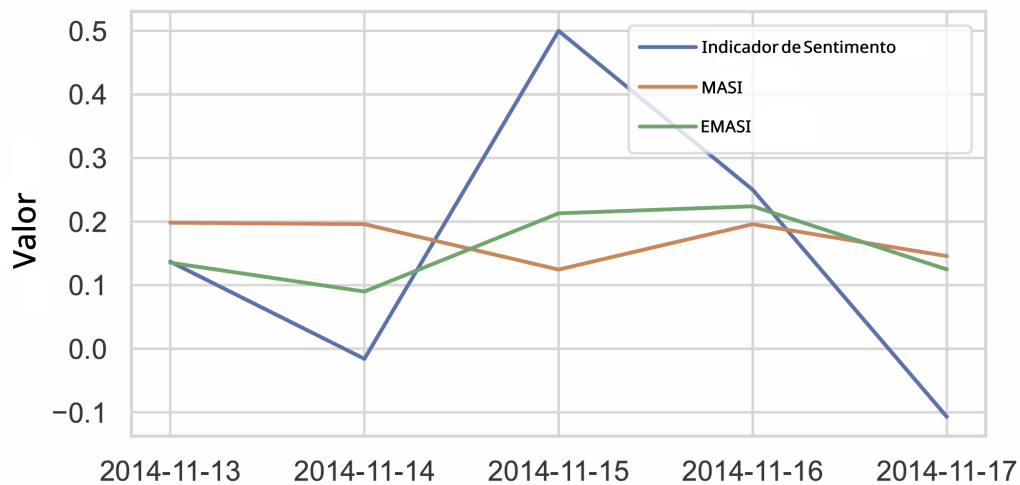
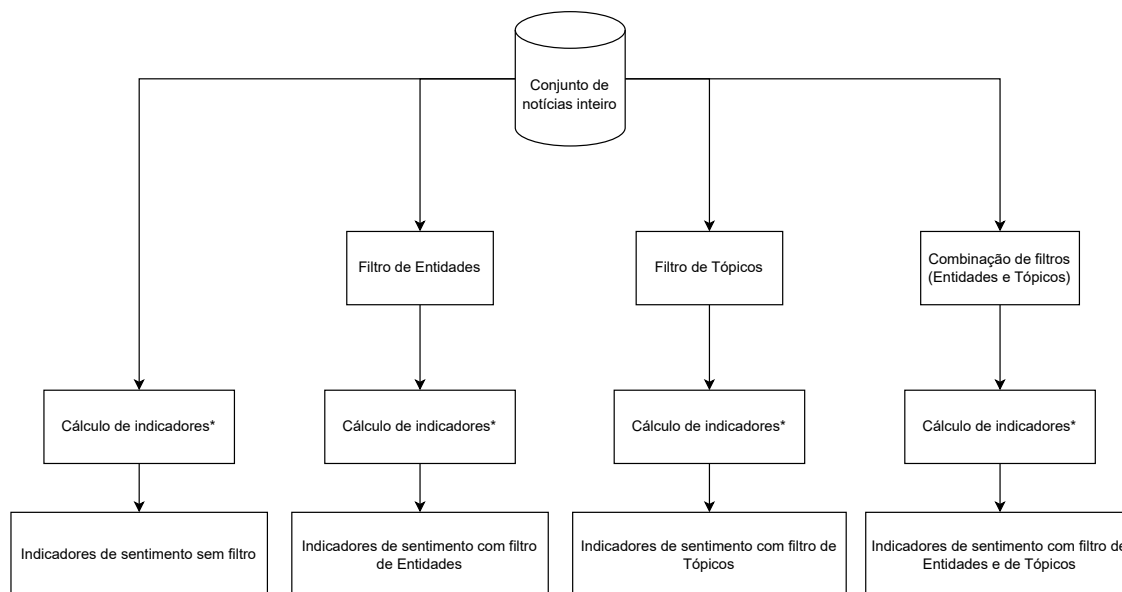


Figura 4.4: Amostra do Indicador de Sentimento, MASI e EMASI ao longo do tempo.

sejam valiosas, elas nem sempre impactam diretamente o mercado brasileiro. Para refinar o conjunto de dados e focar exclusivamente em notícias relevantes para o mercado brasileiro, implementou-se filtros a partir das entidades identificadas através do modelo Gemini Pro e os tópicos mais prevalentes identificados através de esforços anteriores de modelagem de tópicos. Consequentemente, foram gerados os indicadores de sentimento Puro, MASI e EMASI, abordados neste capítulo, para os seguintes cenários: uso do conjunto completo de dados; aplicação de um filtro com base nas entidades principais; uso de um filtro baseado na modelagem de tópicos descrita anteriormente; e, finalmente, combinação de filtros das entidades principais com a modelagem de tópicos (11)(16). A Figura 4.5 ilustra o fluxo de geração desses indicadores.

O filtro de entidade selecionou a entidade primária denominada “Ibovespa” e excluiu outras notícias (10). O filtro de tópicos foi desenvolvido a partir de técnicas de modelagem de tópicos sendo estas: geração de *Embeddings* para a representação vetorial dos textos em um espaço “n” dimensional, redução da dimensionalidade do vetor de características, e agregação do vetor obtido em tópicos, de modo que os pontos em cada tópico possuem características semelhantes no espaço vetorial, assim como descrito no Capítulo 2. Na construção dos indicadores baseados em tópicos, foi adotado um critério de relevância: somente tópicos com um mínimo de 100 notícias associadas foram considerados na análise (15). Este critério teve como objetivo reduzir o impacto de temas com baixo volume de notícias, considerados como “ruído”. Como resultado,



* Na etapa de cálculo de indicadores de sentimento, são computados três indicadores distintos: Puro, MASI e EMASI.

Figura 4.5: Fluxo de geração dos indicadores de sentimento.

foram selecionados 288 tópicos, os quais foram utilizados como base para o cálculo dos indicadores. A distribuição das notícias por tópico pode ser vista na Figura 4.6.

Os *embeddings* que não são semelhantes o suficiente para fazer parte de um tópico são definidos como *outliers* (valores em um conjunto de dados que destoam significativamente dos demais, podendo representar eventos incomuns, variáveis extremas ou erros). O filtro de tópicos remove as notícias atípicas segregadas pelo HDBSCAN. No HDBSCAN, uma atribuição de cluster de -1 indica pontos de ruído ou valores discrepantes. Esses são pontos de dados que o algoritmo não atribui a nenhum cluster específico porque não atendem aos critérios de densidade para fazer parte de um cluster. Esses pontos são normalmente considerados ruído ou anomalias no conjunto de dados. O último filtro foi criado combinando os dois filtros anteriores, o filtro de entidade e o filtro de modelagem de tópicos.

4.1.5 Análise estatística dos indicadores de sentimento

A etapa de análise estatística dos indicadores de sentimento deste estudo visa determinar a existência de relação entre os indicadores obtidos e o desempenho do mercado, bem como quantificar a força dessa relação por meio do coeficiente de correlação. É fundamental salientar que a correlação, ainda que indique uma relação positiva ou negativa, não implica em causalidade entre as variáveis.

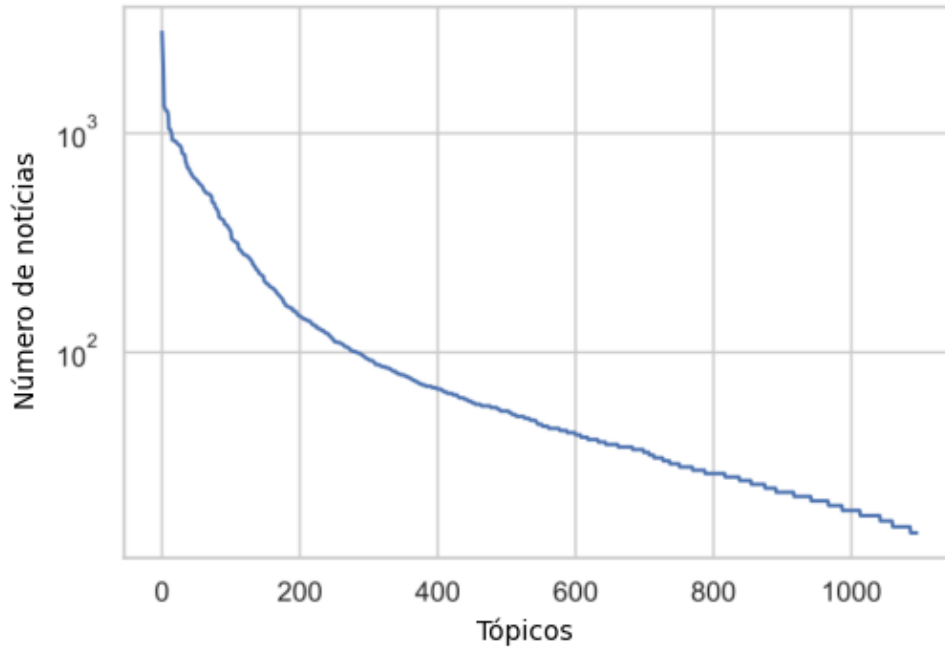


Figura 4.6: Distribuição de notícias por tópico.

A escolha do método de cálculo do coeficiente de correlação depende da distribuição dos dados. Enquanto diversos métodos assumem a normalidade dos dados, a violação desse pressuposto exige a utilização de métodos não paramétricos. Nesse contexto, o teste Shapiro-Wilk, recomendado por autores como [Thode 2002], destaca-se como uma ferramenta eficaz para a verificação da normalidade dos dados.

Neste trabalho, o teste Shapiro-Wilk, disponível na biblioteca Scipy [Virtanen et al. 2020], foi aplicado para verificar a normalidade dos dados antes da condução das análises de correlação (20). A estatística do teste é dada pela Eq. 4-4, onde W é a estatística de Shapiro-Wilk, $x_{(i)}$ são os valores de amostra ordenados, a_i são constantes calculadas a partir dos valores esperados de $x_{(i)}$ e sua matriz de covariância, x_i são os valores de amostra observados, \bar{x} é a média da amostra, e n é o tamanho da amostra.

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (4-4)$$

O teste de Shapiro-Wilk avalia a hipótese nula (H_0) de que a população da qual a amostra é retirada segue uma distribuição normal. O valor estatístico do teste W é utilizado para calcular o valor-p, que é retornado pela função correspondente na biblioteca SciPy.

Em um nível de significância (α) de 0,05, H_0 é rejeitado se o valor-p for

menor do que α , indicando que os dados se desviam significativamente da normalidade. Por outro lado, se o valor-p exceder α , não há evidência suficiente para rejeitar H_0 , sugerindo que os dados não se desviam significativamente de uma distribuição normal [Bakar e Rosbi 2018].

Regra de decisão:

- Se $p \geq 0.05$, os dados são considerados normalmente distribuídos, e testes paramétricos são apropriados.
- Se $p < 0.05$, os dados não são normalmente distribuídos, e testes não paramétricos devem ser usados.

Essa avaliação de normalidade informa a seleção de métodos estatísticos apropriados, contribuindo para a validade e robustez da análise.

Como nem todos os dados possuíam uma distribuição normal, utilizou-se o método não paramétrico de correlação de postos de Spearman [Spearman 1904] (Eq. 4-5), um método não paramétrico que não requer a normalidade dos dados, assim como realizado em [Alanyali, Moat e Preis 2013] (21). A biblioteca Pandas foi utilizada para calcular o coeficiente. Na Eq. 4-5, ρ é o coeficiente de correlação de classificação de Spearman, d_i representa a diferença entre as classificações dos pares de variáveis correspondentes nas duas amostras e n é o número de observações.

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n} \quad (4-5)$$

O coeficiente de Spearman assume valores entre -1 e 1, onde:

- $\rho = -1$ indica uma correlação negativa perfeita;
- $\rho = 0$ aponta ausência de correlação;
- $\rho = 1$ revela uma correlação positiva perfeita.

Para análise de correlação, gerou-se *features* que representam o retorno diário (variação percentual) e suas médias ao longo de diversos períodos futuros (1, 7, 30 e 60 dias) com base no preço de fechamento (22). Esta abordagem permite avaliar a importância de cada combinação de indicadores de sentimento e filtrar, examinando as suas correlações com tendências futuras do mercado ao longo de diferentes horizontes temporais. A variação percentual do preço de uma ação VP_t , calculada de um dia para o outro, indica a diferença proporcional no preço entre esses dois dias, expressa como uma porcentagem representada na Equação 4-6, onde PF_t é o preço de fechamento do dia t e PF_{t-1} é o preço de fechamento do dia anterior. Os resultados da análise de correlação dos indicadores de sentimento são apresentados no Capítulo 5.

$$VP_t = \left(\frac{PF_t - PF_{t-1}}{PF_{t-1}} \right) \times 100 \quad (4-6)$$

A Escala de Cohen, definida em [Cohen et al. 2003], foi utilizada (Tabela 4.2) para avaliar o grau de correlação obtido.

Tabela 4.2: Escala de Cohen

Coefficiente de Correlação	Grau de Correlação
0,0 a 0,1	muito pequeno
0,1 a 0,3	pequeno
0,3 a 0,5	moderado
0,5 a 0,7	grande
0,7 a 0,9	muito grande
0,9 a 1,0	próximos

4.1.6 Aprendizado por Reforço Profundo

Na Fase IV, que se concentrou na implementação de algoritmos de Aprendizado por Reforço Profundo (DRL) e na execução de experimentos, os seguintes métodos foram utilizados: *Advantage Actor-Critic* (A2C), *Proximal Policy Optimization* (PPO), *Deep Deterministic Policy Gradient* (DDPG), *Twin Delayed Deep Deterministic Policy Gradient* (TD3) e *Soft Actor-Critic* (SAC) (23). A implementação desses algoritmos foi conduzida usando a biblioteca FinRL [Liu et al. 2022], que fornece interfaces padronizadas e invólucros baseados na estrutura Stable Baselines3 [Raffin et al. 2025]. Além disso, o ambiente usado para otimização de portfólio foi o StockPortfolioEnv, também fornecido pela FinRL e anteriormente adotado no estudo por [Guan e Liu 2021].

Ambiente de ARP

O ambiente simula interações no mercado financeiro para otimização de portfólio, permitindo a interação do agente. A cada passo de tempo discreto t , o agente seleciona um vetor de ação $\mathbf{a}_t \in \mathbb{R}^n$, representando preferências não normalizadas para n ativos, incluindo dinheiro. Essas ações são transformadas em uma alocação de portfólio válida $\mathbf{w}_t \in \mathbb{R}^n$ usando a função *softmax*:

$$\mathbf{w}_t = \text{softmax}(\mathbf{a}_t) = \frac{\exp(\mathbf{a}_t)}{\sum_{i=1}^n \exp(a_{t,i})} \quad (4-7)$$

Onde, $w_t^{(\text{dinheiro})}$ denota a proporção do portfólio alocada para a posição em caixa, a qual é assumida como tendo um retorno zero. Os pesos restantes são distribuídos entre os ativos de risco.

O ambiente calcula o retorno diário do portfólio r_t usando a seguinte formulação:

$$r_t = (1 - w_t^{\text{(dinheiro)}}) \cdot \left(\sum_{i=1}^{n-1} \hat{w}_{t,i} \cdot \left(\frac{P_{t+1,i}}{P_{t,i}} - 1 \right) \right) \quad (4-8)$$

em que:

- $\hat{w}_{t,i} = \frac{w_{t,i}}{\sum_{j=1}^{n-1} w_{t,j}}$ é o peso normalizado do ativo i , excluindo o componente de caixa;
- $P_{t,i}$ é o preço do ativo i no momento t .

O valor do portfólio é atualizado de acordo com o retorno realizado:

$$V_{t+1} = V_t \cdot (1 + r_t) \quad (4-9)$$

Para incorporar a gestão de risco, o ambiente pode opcionalmente impor um limite de turbulência τ_{thresh} . Se a turbulência observada no mercado τ_t exceder esse limite, o portfólio é totalmente liquidado em dinheiro, incorrendo em um custo de transação:

$$V_{t+1} = V_t - c, \quad \text{onde} \quad c = \gamma \cdot V_t \quad (4-10)$$

com γ denotando a taxa proporcional de custo de transação. Uma taxa de custo de transação de 0,03 é utilizada, refletindo a taxa aplicada pela B3 [B3].

Na etapa final (ou seja, $t = T$), o ambiente avalia o desempenho usando o *Sharpe Ratio*, definido como:

$$\text{Sharpe Ratio} = \frac{\sqrt{252} \cdot \mathbb{E}[r_t]}{\text{Std}[r_t] + \varepsilon} \quad (4-11)$$

onde $\mathbb{E}[r_t]$ e $\text{DP}[r_t]$ denotam a média empírica e o desvio padrão dos retornos diários, e ε é uma pequena constante para garantir a estabilidade numérica.

A recompensa escalar retornada ao agente a cada etapa é definida como o retorno do portfólio normalizado pelo valor atual do portfólio:

$$\text{recompensa}_t = \frac{r_t}{V_t} \quad (4-12)$$

O ambiente retorna o próximo estado, a recompensa, uma bandeira booleana indicando se o episódio terminou, e variáveis auxiliares no formato exigido pelas interfaces no estilo OpenAI Gym.

Assim, quatro cenários distintos foram avaliados para cada algoritmo (24). O primeiro cenário considerou apenas os preços e o volume de negociação como variáveis. O segundo cenário incluiu preços, volume e um indicador de sentimento. O terceiro cenário incorporou preços, volume e indicadores fundamentais. Finalmente, o quarto cenário utilizou a combinação mais abrangente: preços, volume, indicador de sentimento e indicadores fundamentais. A Tabela 4.3 delinea as variáveis usadas em cada um dos cenários mencionados.

Tabela 4.3: Variáveis utilizadas em cada cenário

Categoria	Variáveis Utilizadas	Cenário 1	Cenário 2	Cenário 3	Cenário 4
Dados de Mercado	Preços (abertura, fechamento, mínimo, máximo); Volume.	✓	✓	✓	✓
Indicadores de Sentimento	Sentimento de notícias financeiras.	✗	✓	✗	✓
Indicadores Fundamentalistas	Dividend Yield (DY); P/L (Preço/Lucro); ROE (Return on Equity - Retorno sobre o Patrimônio Líquido); Margem Bruta.	✗	✗	✓	✓

Otimização de hiperparâmetros

Algoritmos baseados em redes neurais, como os de aprendizado por reforço profundo, dependem de diversos hiperparâmetros, que são variáveis de configuração fundamentais para o sucesso do treinamento e o desempenho final do modelo. A escolha adequada desses hiperparâmetros pode impactar significativamente a eficiência e a qualidade do agente treinado. O processo de encontrar os melhores valores para esses hiperparâmetros é conhecido como otimização de hiperparâmetros.

Neste trabalho, foi utilizada a biblioteca Optuna para conduzir essa otimização. Diferentemente de métodos mais simples, como a Busca Aleatória (*Random Search*), que avalia combinações de hiperparâmetros de maneira totalmente aleatória, Optuna adota uma abordagem inteligente. Ela emprega técnicas de amostragem adaptativa e aprendizado baseado nas iterações anteriores, o que permite explorar de forma mais eficiente o espaço de soluções e concentrar os esforços em regiões mais promissoras. Essa estratégia reduz o tempo necessário para encontrar configurações de hiperparâmetros que maximizem o desempenho do modelo, tornando-a mais eficaz em comparação com abordagens tradicionais.

Para cada um dos quatro cenários descritos na subseção anterior, realizou-se um estudo com a biblioteca Optuna, compreendendo 50 iterações para cada algoritmo: A2C, PPO, DDPG, TD3 e SAC. A função objetivo maximizada foi o índice *Sharpe Ratio*.

Os melhores conjuntos de hiperparâmetros obtidos nesses estudos estão detalhados na Tabela 4.4, cujos elementos representam:

- `n_steps`: Quantidade de passos executados em cada ambiente antes de realizar uma atualização. O tamanho do lote de treinamento é calculado como $n_steps * n_env$, onde `n_env` representa o número de instâncias do ambiente executando em paralelo;
- `gamma`: Fator de desconto aplicado às recompensas futuras, determinando o peso relativo dessas recompensas no cálculo do valor atual;
- `learning_rate`: Taxa de aprendizado, que pode ser definida como um valor fixo ou como uma função que varia de acordo com o progresso do treinamento;
- `vf_coef`: Peso dado à função de valor no cálculo da perda total do modelo;

- `ent_coef`: Peso dado ao termo de entropia na função de perda, usado para incentivar a exploração de diferentes ações;
- `max_grad_norm`: Limite máximo para o *clipping* de gradientes, sendo este um método para evitar que valores excessivamente altos prejudiquem a estabilidade do treinamento;
- `gae_lambda`: Parâmetro que controla o equilíbrio entre viés e variância no Estimador de Vantagem Generalizado (GAE). Quando configurado como 1, equivale ao cálculo de vantagem tradicional;
- `net_arch`: Configuração das camadas das redes de política e de valor, podendo ser definida como uma lista de inteiros ou um dicionário especificando ambas;
- `clip_range`: Parâmetro que define o intervalo de clipping dos valores, podendo ser um valor fixo ou uma função que varia conforme o progresso do treinamento (de 1 a 0);
- `batch_size`: Número de amostras incluídas em cada *minibatch* (lote) de treinamento;
- `tau`: Fator de atualização suave para os pesos das redes-alvo (*Polyak update*), com valores entre 0 e 1;
- `buffer_size`: Tamanho do *replay buffer* (memória para amostras de experiências de treinamento);
- `policy_delay`: Intervalo entre as atualizações da política e das redes-alvo. Durante esse intervalo, os valores de Q são atualizados com maior frequência (a cada passo de treinamento).

Em cada cenário, os algoritmos foram treinados utilizando hiperparâmetros otimizados ao longo de 100.000 iterações (*timesteps*), considerando as limitações dos recursos computacionais.

4.1.7 Análise estatística dos resultados experimentais

Uma análise estatística dos resultados foi realizada na fase VI, com o objetivo de fornecer uma base rigorosa para comparações entre os métodos avaliados. Os dados analisados correspondem às métricas de desempenho dos modelos de Aprendizado por Reforço Profundo em relação aos diferentes cenários de treinamento. As métricas observadas foram: *Sharpe Ratio* (SR), *Maximum Drawdown* (MD) e Retorno Anual (RA) [Lin et al. 2023].

O índice de Sharpe avalia a relação entre o retorno anual obtido pelo investidor e o risco, expresso pela volatilidade do portfólio. Esta métrica permite aos investidores comparar diferentes ativos ou portfólios, considerando não apenas os retornos absolutos, mas também a incerteza associada a esses retornos. O índice é calculado pela Eq. 4-13,

Tabela 4.4: Hiperparâmetros dos algoritmos por cenário

Algoritmo	Hiperparâmetro	Cenário 1	Cenário 2	Cenário 3	Cenário 4
A2C	n_steps	8	5	7	10
	gamma	0.9049	0.9033	0.9815	0.9926
	learning_rate	5.36e-4	8.04e-4	6.75e-4	1.70e-4
	vf_coef	0.4469	0.6146	0.8076	0.2403
	ent_coef	2.35e-6	6.17e-8	9.12e-4	5.46e-6
	max_grad_norm	0.7976	0.6159	0.9628	0.4818
	gae_lambda	0.9603	0.9041	0.9722	0.9793
	net_arch	[32, 32]	[32, 32]	[64, 64]	[64, 64]
PPO	n_steps	20	14	12	6
	gamma	0.9761	0.9972	0.9406	0.9232
	learning_rate	8.66e-4	7.68e-5	8.06e-4	1.20e-4
	vf_coef	0.5085	0.6408	0.5999	0.3181
	ent_coef	4.77e-7	3.60e-4	4.25e-3	1.30e-8
	max_grad_norm	0.5875	0.6867	0.5567	0.6481
	gae_lambda	0.9532	0.9694	0.8510	0.9834
	clip_range	0.2221	0.2308	0.3729	0.2247
	net_arch	[8, 8]	[32, 32]	[32, 32]	[64, 64]
DDPG	gamma	0.9843	0.9990	0.9033	0.9181
	learning_rate	3.02e-4	1.06e-4	9.32e-4	6.88e-4
	batch_size	68	123	84	144
	tau	0.0748	0.0697	0.0928	0.0128
	buffer_size	186,902	753,558	640,845	563,062
	net_arch	[32]	[16, 16]	[32]	[32, 32]
TD3	gamma	0.9049	0.9106	0.9040	0.9589
	learning_rate	1.74e-4	6.63e-5	8.50e-4	5.33e-4
	batch_size	171	49	256	116
	tau	0.0034	0.0842	0.0718	0.0925
	buffer_size	773,075	500,358	101,700	146,213
	policy_delay	2	2	3	2
	net_arch (pi)	[32]	[32, 32]	[16, 16]	[16, 16]
	net_arch (qf)	[32]	[32, 32]	[16, 16]	[16, 16]
SAC	gamma	0.9822	0.9750	0.9996	0.9701
	learning_rate	2.14e-5	2.06e-5	5.63e-4	2.29e-5
	batch_size	230	65	154	130
	tau	0.0475	0.0744	0.0181	0.0460
	buffer_size	387,664	783,029	417,874	219,532
	net_arch (pi)	[16, 16]	[16, 16]	[16, 16]	[16, 16]
	net_arch (qf)	[16, 16]	[16, 16]	[16, 16]	[16, 16]

onde \bar{R} representa o retorno anualizado da ação ou portfólio, e σ_R denota o desvio padrão do retorno anualizado.

$$SR = \frac{\bar{R}}{\sigma_R} \quad (4-13)$$

O *Maximum Drawdown* (MD) indica a maior queda registrada de um ativo ou portfólio, desde o seu valor máximo até o mínimo, durante um determinado período. Esta métrica é utilizada para prever o risco de perdas significativas [Lin et al. 2023]. A fórmula do MD é expressa pela Eq. 4-14, onde p_t é o valor do pico e p_τ é o valor do vale.

$$MD = \max_{\tau \geq t} \frac{p_t - p_\tau}{p_t} \quad (4-14)$$

O retorno anual de um ativo ou portfólio é a porcentagem de lucro que um investidor obtém ao final de um ano específico. Ele é calculado pela Eq. 4-15, onde V_i representa o valor inicial e V_f o valor final do portfólio. Os números de dias de negociação em um ano e no total são representados por T_{year} e T_{all} , respectivamente.

$$RA = \frac{V_f - V_i}{V_i} \times \frac{T_{year}}{T_{all}} \quad (4-15)$$

Para cada combinação modelo-cenário, as respectivas métricas de desempenho foram coletadas. Para estabelecer significância estatística nas análises, 44 execuções foram realizadas para cada combinação avaliada.

A escolha dos métodos de teste de significância estatística depende da distribuição dos dados. Embora muitos métodos clássicos assumam normalidade, violações dessa suposição requerem o uso de alternativas não paramétricas. Nesse contexto, o teste de Shapiro-Wilk, conforme recomendado por autores como [Thode 2002], destaca-se como uma ferramenta eficaz para avaliar a normalidade dos dados.

A escolha dos métodos de teste de significância estatística depende da distribuição dos dados. Neste estudo, o teste de Shapiro-Wilk, descrito anteriormente na seção 4.1.5, foi aplicado para avaliar a normalidade dos dados antes de conduzir análises de significância estatística. As amostras obtidas durante a fase experimental indicam que os dados não são normalmente distribuídos, como evidenciado por várias combinações de modelo-métrica-cenário mostradas na Tabela 4.5. Consequentemente, métodos não-paramétricos são necessários para conduzir uma análise estatística apropriada.

Para comparar mais de dois grupos independentes quando a suposição de normalidade é violada, o teste de Kruskal-Wallis H [Kruskal e Wallis 1952] serve como uma alternativa não-paramétrica robusta ao ANOVA unidirecional [Kaufmann e Schering 2014] (26). Este teste baseado em postos avalia se há diferenças estatisticamente significativas na tendência central (medianas) entre k grupos sem assumir uma distribuição normal ou variâncias iguais.

Tabela 4.5: P-valores para diferentes modelos em vários cenários e métricas de avaliação.

Modelo	Cenário	P-valor	Métrica
A2C	1	0.016529	Maximum Drawdown
PPO	1	0.017902	Maximum Drawdown
DDPG	1	0.021922	Maximum Drawdown
SAC	2	0.000179	Sharpe Ratio
SAC	2	0.000609	Annual Return
TD3	2	0.022228	Maximum Drawdown
PPO	3	0.040507	Maximum Drawdown
SAC	4	0.002576	Maximum Drawdown
PPO	4	0.006846	Maximum Drawdown
DDPG	4	0.005140	Maximum Drawdown

O estatístico de teste H é calculado como:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1) \quad (4-16)$$

onde k é o número de grupos, n_i é o tamanho do grupo i , R_i é a soma das posições no grupo i , e N é o número total de observações. Sob a hipótese nula (H_0), que assume que todos os grupos compartilham a mesma distribuição, H segue uma distribuição qui-quadrado com $k - 1$ graus de liberdade.

Regra de Decisão:

- Rejeitar H_0 se o p-valor $< \alpha$ (0,05), indicando que pelo menos um grupo difere significativamente.
- Falhar em rejeitar H_0 se o p-valor $\geq \alpha$, indicando que não há diferença significativa entre os grupos.

Em situações onde $p < 0,05$, a hipótese nula (H_0) é rejeitada, indicando que pelo menos um grupo apresenta uma diferença estatisticamente significativa em sua mediana. Para determinar quais grupos específicos diferem, recomenda-se comparações *post-hoc* pareadas. Essa metodologia permite comparações estatísticas rigorosas entre múltiplos grupos, mesmo na presença de dados ordinais, enviesados ou não normalmente distribuídos.

A avaliação dos resultados experimentais envolveu a comparação das métricas de cada algoritmo em quatro cenários e a verificação da significância estatística de suas diferenças. Esses resultados são apresentados no Capítulo 6 (27). Além disso, esses resultados foram comparados com pontos de referência tradicionais, como *Buy and Hold* (BAH), conforme realizado por [Li et al. 2022] em seu estudo sobre aprendizado por reforço profundo, e o Índice Ibovespa (IBOV) para avaliar a eficácia.

Seleção de Indicador de Sentimento

A análise dos indicadores de sentimento descrita nesta seção foi publicada na *9th International Conference on Computer Science and Computational Intelligence (CSCI) 2024* com o título *Development and Evaluation of a Sentiment Indicator based on Financial News in Portuguese* [Silva et al. 2025].

5.1 Análise de Correlação

A aplicação dos filtros de notícias, desenvolvidos nas etapas (10)-(16), levou a uma redução significativa no volume de dados, conforme mostrado na Tabela 5.1 (11)(15)(16). A porcentagem de notícias removidas variou entre 40% a 96,3%, dependendo do filtro aplicado. A tabela exibe o número de notícias restantes após a aplicação de cada filtro, assim como a porcentagem em relação ao conjunto inicial de notícias (sem filtro).

Tabela 5.1: Número de notícias com base no filtro aplicado

	Contagem de notícias	Porcentagem
Sem filtro	177155	100%
Filtro de entidade	11005	6,21%
Filtro de tópicos	105911	59,78%
Filtro de entidade e tópicos	6541	3,69%

Para analisar as correlações, calculou-se os coeficientes de correlação entre os indicadores de sentimento (puro, MASI e EMASI) e suas correspondentes características de mercado, como volume, preço de abertura, preço máximo, preço mínimo e preço de fechamento ajustado. Também calculou-se esses coeficientes para o retorno diário (variação percentual) e suas médias para janelas futuras de 1, 7, 30 e 60 dias. Esses cálculos levaram em consideração diferentes cenários de geração de indicadores de sentimento: sem filtro, com filtro de entidade, com filtro de tópicos, e com filtros de entidade e de tópicos. Contudo, apenas as correlações consideradas mais significativas são apresentadas na Tabela 5.2, destacando aquelas classificadas como “grande”, “muito

Tabela 5.2: Coeficientes de correlação de Spearman de cada indicador de sentimento

Filtro	Indicador	Retorno diário	D + 1	Média D + 7	Média D + 30	Média D + 60
Sem filtro	Indicador de sentimento	0,483	0,359	0,233	0,166	0,259
	MASI	0,113	0,091	0,335	0,231	0,251
	EMASI	0,254	0,193	0,346	0,252	0,152
Filtro de entidade	Indicador de sentimento	0,815	0,561	0,311	0,195	0,304
	MASI	0,268	0,194	0,731	0,471	0,330
	EMASI	0,553	0,390	0,632	0,435	0,132
Filtro de tópicos	Indicador de sentimento	0,433	0,315	0,197	0,123	0,208
	MASI	0,098	0,081	0,319	0,219	0,208
	EMASI	0,230	0,173	0,325	0,235	0,083
Filtro de entidade e tópicos	Indicador de sentimento	0,729	0,491	0,247	0,144	0,194
	MASI	0,268	0,194	0,640	0,366	0,207
	EMASI	0,520	0,368	0,538	0,339	0,089

grande” ou “próximo” segundo a Escala de Cohen (Tabela 4.2). As demais correlações calculadas que não estão na Tabela 5.2 apresentaram coeficientes baixos.

Ao realizar a comparação entre os indicadores de sentimento sem filtro e os gerados com filtro de tópicos, verifica-se que as correlações mais fortes são entre o indicador de sentimento puro e o retorno diário, indicando níveis moderados de correlação de 0.483 e 0.433, respectivamente. Ao aplicar o filtro de entidade, foram observados coeficientes notáveis, classificados como “muito grandes”, em correlações como a do indicador puro com a retorno diário (0,815), MASI com a média da janela de 7 dias futuros (0,731) e EMASI com a média da mesma janela (0,632). Além disso, a combinação dos filtros de entidade e de tópico também resultou em correlações “muito grandes” entre o indicador puro e a retorno diário (0,729), assim como entre MASI e a média da janela de 7 dias futuros (0,640).

A forte associação entre o indicador de sentimento da Média Móvel de 7 dias e os preços futuros das ações, juntamente com a sua correlação significativa com o indicador de sentimento da média móvel exponencial, pode ser atribuída a vários elementos. Ambas as médias móveis eliminam flutuações de curto prazo nos dados de sentimento, expondo assim padrões subjacentes que antecipam mudanças futuras no sentimento. Além disso, ao integrar valores de sentimento passados, estes indicadores reconhecem tendências de sentimento a longo prazo e compensam as respostas atrasadas dos investidores às mudanças no sentimento. Além disso, a incorporação de valores de sentimento recentes juntamente com dados históricos aumenta a sua capacidade de prever prontamente os movimentos do mercado, ao mesmo tempo que considera tendências mais amplas no sentimento. Em resumo, estas características tornam ambas as médias móveis instrumentos valiosos para capturar e prever a dinâmica dos sentimentos nos mercados financeiros.

A abordagem de filtro de entidades revelou-se a mais promissora, apresentando os maiores coeficientes de correlação. O indicador de sentimento padrão mostrou a maior correlação com o mercado, enquanto o filtro de tópicos resultou nas menores correlações. Na Fase V do experimento, os testes iniciais com a entidade Ibovespa não trouxeram resultados satisfatórios. Assim, cada notícia foi associada à sua entidade correspondente

para calcular o indicador de sentimento padrão, conforme Eq. 4-1, melhorando as métricas nos testes (22).

5.2 Relacionamento de notícias a cada entidade

A rotulagem de entidades em notícias, realizada pelo modelo Gemini Pro, apresenta inconsistências. Notícias sobre a Petrobras, por exemplo, podem ser rotuladas como “Petrobras”, “PETR4” (seu código na bolsa de valores) ou, ainda, “Petrobrás” (com erro de acentuação). Essa falta de padronização impede a agregação por correspondência exata, exigindo técnicas mais sofisticadas.

A correspondência *fuzzy* é uma técnica para encontrar similaridades entre textos não idênticos. Esse método utiliza medidas de distância para calcular a diferença entre os textos e estabelecer um índice de similaridade. Para agrupar as notícias por empresa, utilizou-se a biblioteca RapidFuzz, baseada na FuzzyWuzzy. Ambas são destinadas à correspondência *fuzzy* de strings, visando encontrar ocorrências aproximadas de um padrão. O cálculo de similaridade (Eq. 5-1) empregado neste trabalho, baseado no algoritmo de Ratcliff e Obershelp [Ratcliff e Metzener 1988], calcula a similaridade entre duas strings, S_1 e S_2 , de comprimentos m e n , respectivamente. O algoritmo busca a quantidade de caracteres correspondentes em ambas as strings, dividindo-a pela soma dos comprimentos ($m + n$). Caracteres correspondentes são aqueles presentes na maior subsequência comum contínua. O processo é recursivo, considerando as substrings restantes à esquerda e à direita como novas strings até que todas as partes sejam analisadas [Nalawati e Yuntari 2021]. O resultado é multiplicado por 2 para normalizar o valor entre 0 e 1 e, posteriormente, por 100, resultando em um índice de similaridade de 0 a 100, onde 0 indica nenhuma similaridade e 100 indica strings idênticas.

$$D_{ro} = 100 * 2 * \frac{C}{m + n} \quad (5-1)$$

sendo: C é o número de caracteres correspondentes na subsequência contígua mais longa entre as duas strings de entrada, m e n representam os tamanhos das strings S_1 e S_2 , respectivamente.

Assim, foram obtidos os agrupamentos de notícias para cada entidade, totalizando 4.813 entidades. A Figura 5.1 apresenta as 15 entidades com maior volume de notícias.

Para a etapa de experimentação, o portfólio foi composto pelas cinco ações do índice Ibovespa com maior volume de notícias associadas, conforme os dados coletados, excluindo-se aquelas sem dados fundamentalistas históricos suficientes. As ações selecionadas foram: Eletrobras (ELET3), Embraer (EMBR3), JBS (JBSS3), Banco do Brasil

(BBAS3) e Braskem (BRKM5). A Figura 5.2 ilustra a quantidade de notícias associada a cada uma dessas ações.

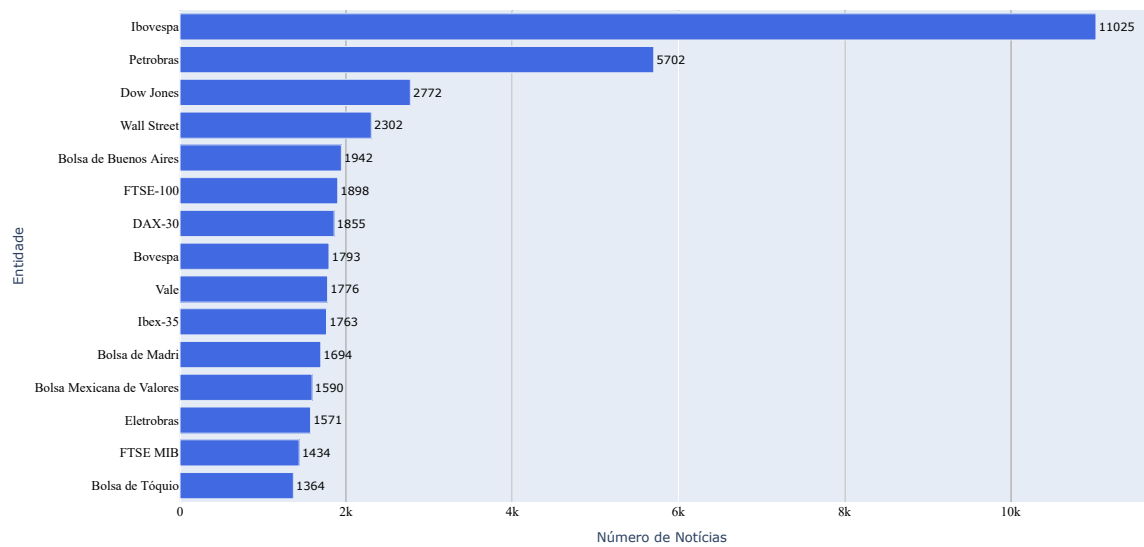


Figura 5.1: Distribuição de notícias relativas a cada entidade.

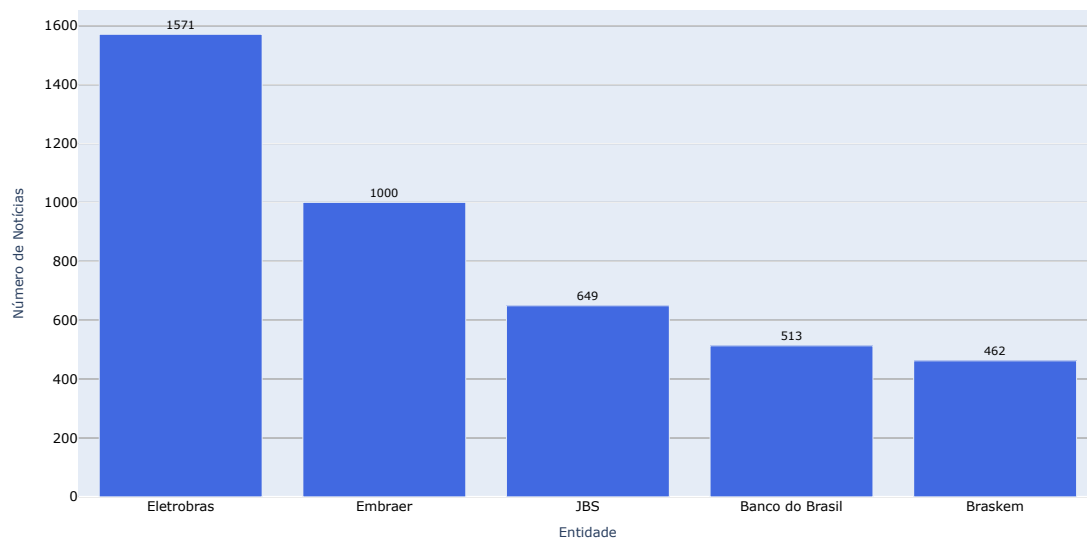


Figura 5.2: Quantidade de notícias das ações selecionadas.

Embora as entidades Petrobras e Vale apresentassem maior volume de notícias, seus dados fundamentalistas históricos trimestrais estavam disponíveis apenas a partir de 2019, enquanto as demais ações selecionadas possuíam dados desde 2013, sendo este o período de interesse para o estudo.

Resultados e Discussões

Os resultados apresentados nesta seção foram publicados na *IEEE 49th Annual Computers, Software, and Applications Conference (COMPSAC) 2025* com o título *A Deep Reinforcement Learning Approach for Portfolio Optimization of Brazilian Assets Using Fundamental and Sentiment Indicators* [Silva et al. 2025].

6.1 Resultados

Conforme ilustrado na Figura 4.1, esta seção apresenta os resultados obtidos a partir das etapas desenvolvidas. Inicialmente, na Fase I, procedeu-se à aquisição dos dados, incluindo notícias financeiras, histórico de preços e indicadores fundamentalistas. Em seguida, na Fase II, realizou-se a extração do sentimento expresso nas notícias e identificação das principais entidades mencionadas. A Fase III foi dedicada à geração de indicadores de sentimento, utilizando diferentes abordagens de filtragem: sem filtro, filtro de entidades, filtro de tópicos e uma combinação destes. Na Fase IV, realizou-se uma análise estatística dos indicadores de sentimento para verificar sua correlação com o desempenho do índice Ibovespa mediante o cálculo do coeficiente de correlação. A Fase V consistiu na implementação e execução de cinco algoritmos de ARP utilizando os dados coletados nas fases anteriores. Por último, na Fase VI, foi efetuada uma análise estatística dos resultados experimentais, a qual iniciou com um teste de normalidade para decidir o método de significância estatística adequado, seguida pelo teste de Kruskal-Wallis para avaliar a presença de diferenças estatísticas entre as amostras obtidas.

Para avaliar o desempenho dos modelos propostos, seus resultados foram comparados com duas estratégias de referência: o índice Ibovespa (IBOV) e a estratégia Buy and Hold (BAH). A Tabela 6.1 apresenta o Sharpe Ratio (SR), Maximum Drawdown (MD) e Retorno Anual (RA) dessas referências. A estratégia BAH supera o IBOV em SR e RA, enquanto o IBOV demonstra um melhor controle de queda.

As subseções a seguir apresentam os resultados dos algoritmos, organizados de acordo com as métricas avaliadas: *Sharpe Ratio*, *Maximum Drawdown* e *Retorno Anual*. Cada métrica é analisada com base nas 44 amostras consideradas no estudo.

Tabela 6.1: Métricas de desempenho das bases de comparação

Referência	SR	MD	RA
IBOV	0,5461	-0,2650	0,0957
BAH	0,5867	-0,3702	0,1275

6.1.1 Sharpe Ratio

A Figura 6.1 apresenta as distribuições do *Sharpe Ratio* para cada algoritmo nos quatro cenários analisados. A linha vermelha tracejada indica a média de cada distribuição, enquanto as linhas azul e verde representam, respectivamente, os valores do *Sharpe Ratio* do Ibovespa e da estratégia *Buy and Hold*. Observa-se que os boxplots de todos os algoritmos situam-se predominantemente acima dessas duas referências, o que sugere um desempenho superior em termos de controle de risco. Complementarmente, a Figura 6.2 exibe um gráfico de Ridgeline, que permite visualizar com mais clareza a distribuição dos valores de *Sharpe Ratio* por algoritmo e cenário, facilitando a comparação entre os grupos.

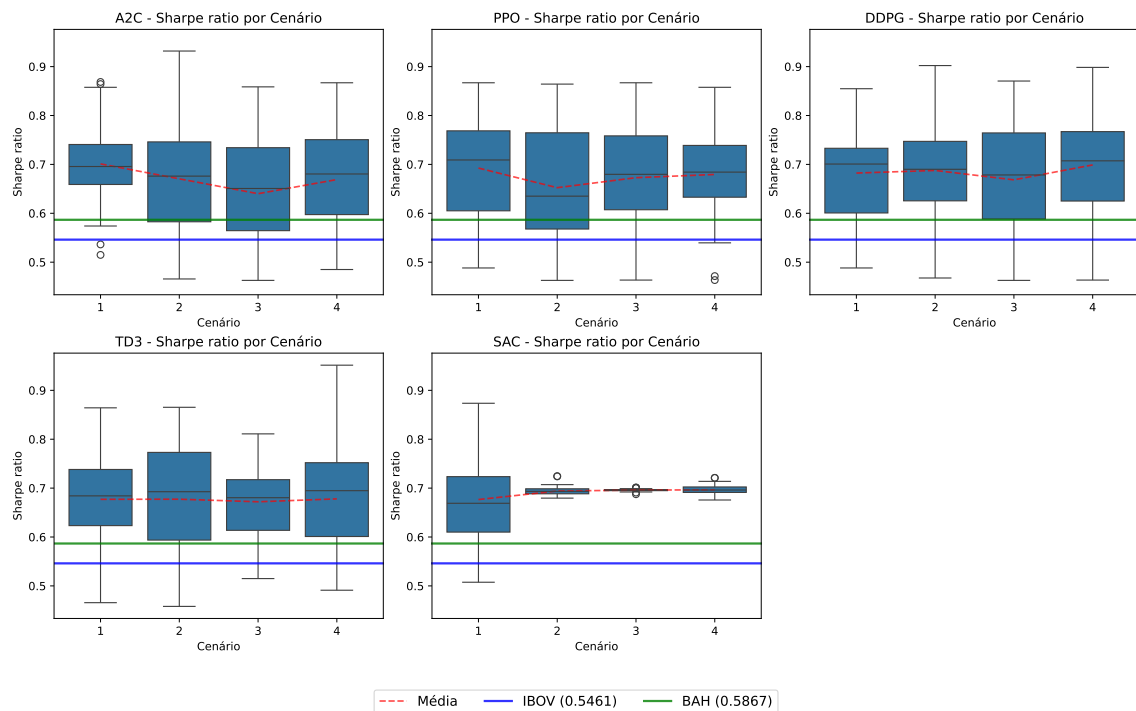


Figura 6.1: Distribuições do Sharpe Ratio: Análise por Algoritmo e Cenário.

Os valores do Sharpe Ratio para cada algoritmo e cenário estão detalhados na Tabela 6.2. Esta tabela inclui a média, mediana, máximo, e mínimo. O tamanho do efeito (d de Cohen) foi empregado na tabela para quantificar a magnitude das diferenças entre o desempenho do índice de Sharpe de cada algoritmo de aprendizado por reforço

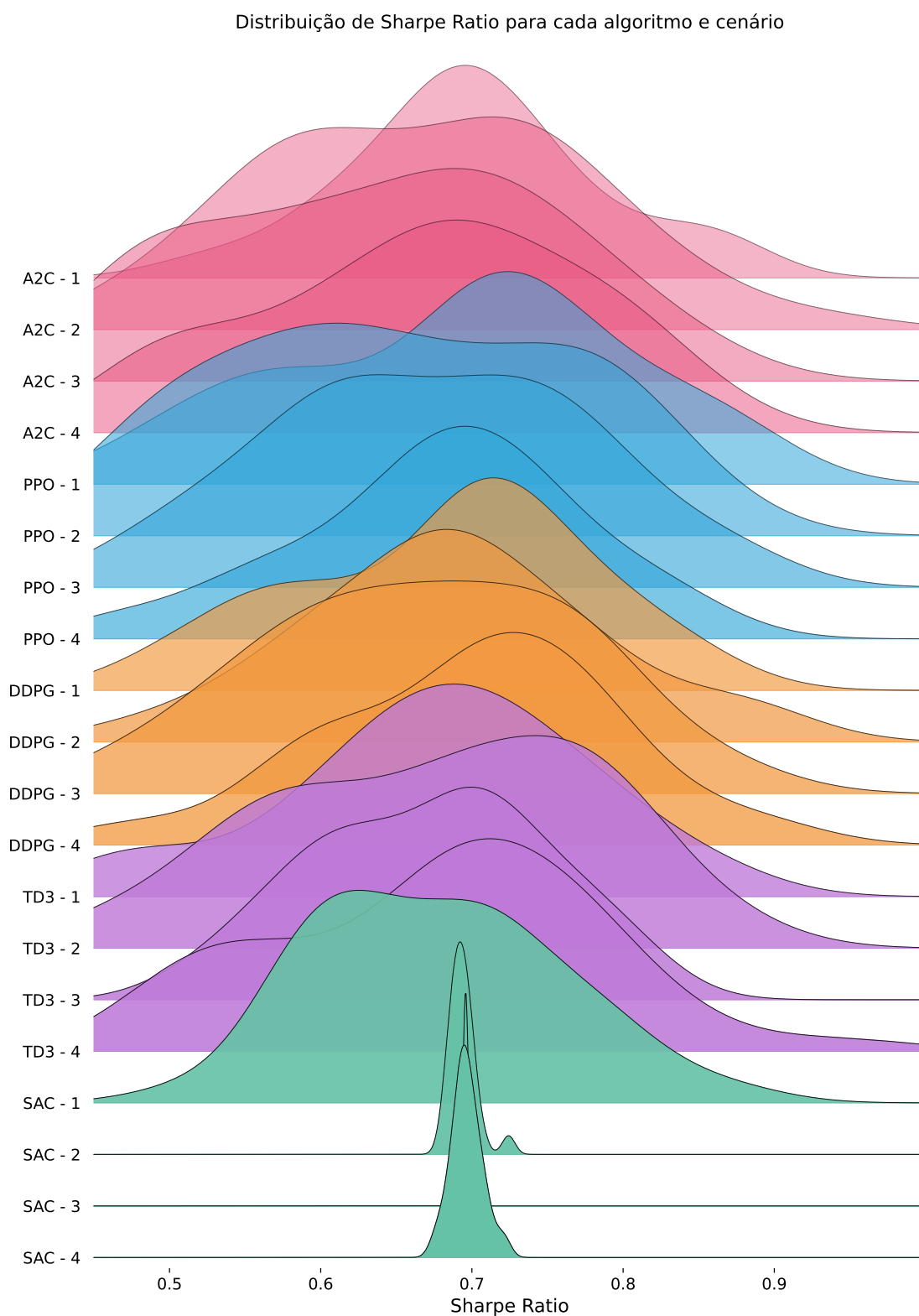


Figura 6.2: Análise das distribuições do Sharpe Ratio.

e as bases de comparação (IBOV e BAH). Esta medida padronizada considera tanto a tendência central quanto a variabilidade das distribuições de desempenho dos algoritmos, proporcionando uma comparação mais robusta do que simples diferenças percentuais. O cálculo do tamanho do efeito normaliza a diferença entre a mediana do índice de Sharpe do algoritmo e o valor benchmark pela desvio padrão agrupado, resultando em uma métrica adimensional que facilita comparações entre algoritmos. A interpretação segue os limiares convencionais de Cohen (Tabela 4.2), permitindo avaliar tanto a significância estatística quanto prática no desempenho do algoritmo.

Tabela 6.2: Medidas estatísticas do Sharpe Ratio de cada algoritmo e cenário

Algoritmo	Cenário	Média	Mediana	Máximo	Mínimo	Efeito (Ibov)	Efeito (BAH)
A2C	1	0,701	0,696	0,869	0,515	Muito grande	Muito grande
	2	0,671	0,676	0,932	0,466	Muito grande	Muito grande
	3	0,640	0,651	0,859	0,463	Muito grande	Muito grande
	4	0,669	0,680	0,867	0,485	Muito grande	Muito grande
PPO	1	0,693	0,709	0,867	0,488	Muito grande	Muito grande
	2	0,652	0,635	0,864	0,463	Muito grande	Muito grande
	3	0,673	0,680	0,867	0,463	Muito grande	Muito grande
	4	0,679	0,684	0,858	0,463	Muito grande	Muito grande
DDPG	1	0,682	0,701	0,855	0,488	Muito grande	Muito grande
	2	0,688	0,690	0,902	0,468	Muito grande	Muito grande
	3	0,668	0,678	0,871	0,463	Muito grande	Muito grande
	4	0,699	0,707	0,898	0,463	Muito grande	Muito grande
TD3	1	0,677	0,684	0,864	0,466	Muito grande	Muito grande
	2	0,677	0,693	0,865	0,458	Muito grande	Muito grande
	3	0,672	0,680	0,811	0,515	Muito grande	Muito grande
	4	0,678	0,695	0,951	0,491	Muito grande	Muito grande
SAC	1	0,677	0,669	0,874	0,507	Muito grande	Muito grande
	2	0,694	0,693	0,724	0,680	Muito grande	Muito grande
	3	0,696	0,696	0,701	0,688	Muito grande	Muito grande
	4	0,697	0,696	0,721	0,676	Muito grande	Muito grande

Os resultados do teste H de Kruskal-Wallis para o *Sharpe Ratio* são apresentados na Tabela 6.3. Este teste foi utilizado para avaliar a existência de diferenças estatisticamente significativas entre os cenários para cada algoritmo de aprendizado por reforço, tendo o *Sharpe Ratio* como métrica. Para uma análise mais completa, foi calculado o tamanho do efeito (Eta quadrado) para quantificar a magnitude das diferenças, categorizadas como negligenciáveis ($< 0,01$), pequenas (0,01-0,06), médias (0,06-0,14) ou grandes ($> 0,14$). Os resultados mostraram que nenhum algoritmo apresentou diferenças significativas entre os cenários ($p > 0,05$), com Eta quadrado próximo de zero. Isso sugere que a variação entre cenários não teve impacto relevante no desempenho dos algoritmos em relação ao *Sharpe Ratio* neste experimento.

Tabela 6.3: Resultados do teste H de Kruskal-Wallis para o Sharpe Ratio.

Algoritmo	Estatística H	P-valor	Eta quadrado	Interpretação do P-valor	Interpretação do tamanho do efeito
A2C	6,400	0,094	0,020	Não significativo	Pequeno
PPO	2,935	0,402	0,000	Não significativo	Desprezível
DDPG	2,379	0,497	-0,004	Não significativo	Desprezível
TD3	0,214	0,975	-0,016	Não significativo	Desprezível
SAC	5,264	0,153	0,013	Não significativo	Pequeno

6.1.2 Maximum Drawdown

A Figura 6.3 exibe as distribuições do *Maximum Drawdown* para cada algoritmo nos quatro cenários analisados. A linha tracejada vermelha representa a média de cada distribuição, enquanto as linhas azul e verde indicam, respectivamente, os valores de *Maximum Drawdown* do Ibovespa e da estratégia *Buy and Hold*. Nota-se que os boxplots de todos os algoritmos, em todos os cenários, situam-se acima do BAH e apresentam mediana próxima ao IBOV, indicando um controle de perdas semelhante ao do índice e superior ao da estratégia passiva. A Figura 6.4 complementa essa análise por meio de um gráfico de Ridgeline, que evidencia a distribuição dos valores de *Maximum Drawdown* entre os grupos avaliados.

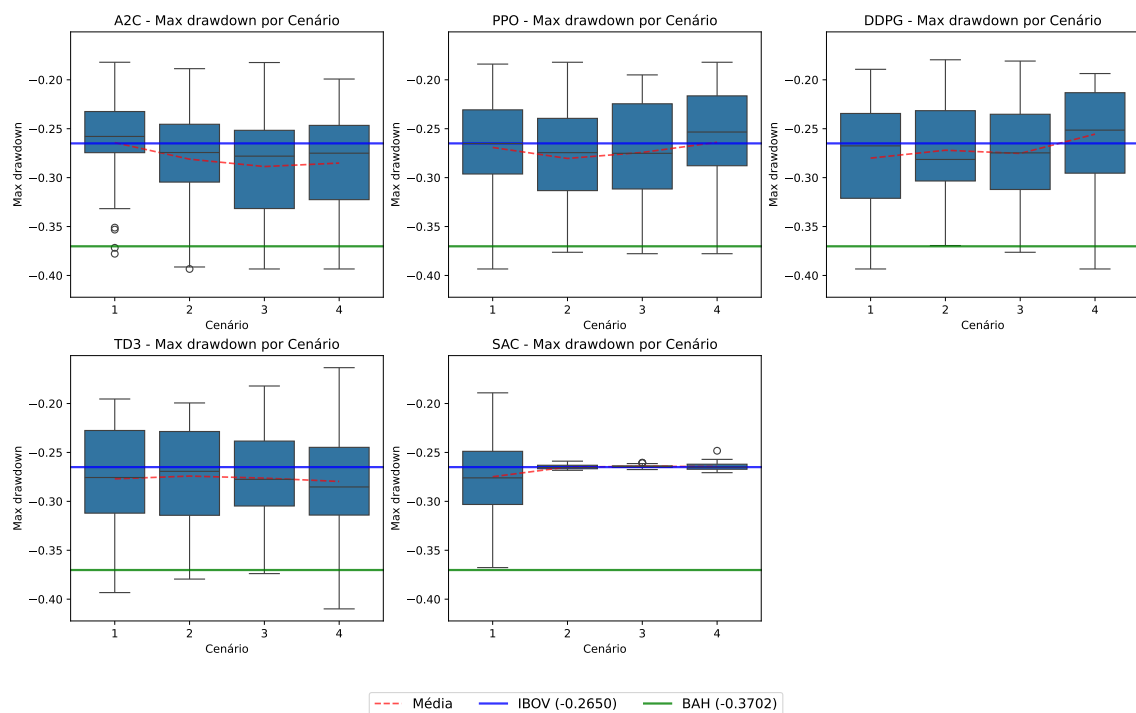


Figura 6.3: Distribuições do Maximum Drawdown: Análise por Algoritmo e Cenário.

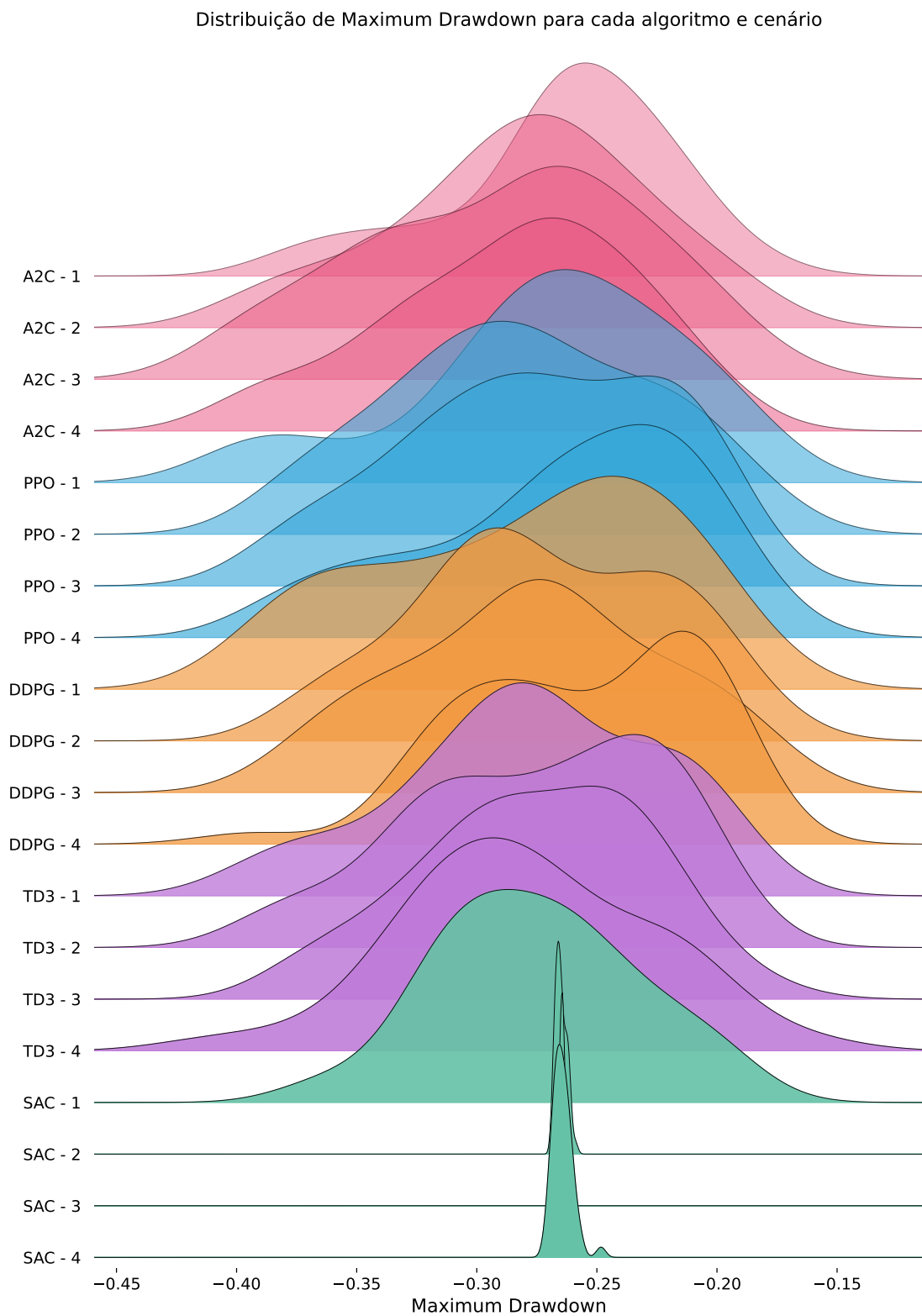


Figura 6.4: Análise das distribuições do Maximum Drawdown.

De forma similar à subseção anterior, a Tabela 6.4 apresenta os valores de *Maximum Drawdown* para cada algoritmo em cada cenário, incluindo as estatísticas de média, mediana, valor máximo, valor mínimo, além do tamanho do efeito da diferença de desempenho em relação ao IBOV e ao BAH.

Tabela 6.4: Medidas estatísticas do Maximum Drawdown de cada algoritmo e cenário

Algoritmo	Cenário	Média	Mediana	Máximo	Mínimo	Efeito (Ibov)	Efeito (BAH)
A2C	1	-0.264	-0.258	-0.182	-0.378	Muito grande	Muito grande
	2	-0.281	-0.274	-0.189	-0.393	Muito grande	Muito grande
	3	-0.289	-0.278	-0.182	-0.393	Muito grande	Muito grande
	4	-0.285	-0.275	-0.199	-0.393	Muito grande	Muito grande
PPO	1	-0.269	-0.265	-0.184	-0.393	Muito grande	Muito grande
	2	-0.28	-0.274	-0.182	-0.376	Muito grande	Muito grande
	3	-0.274	-0.275	-0.195	-0.378	Muito grande	Muito grande
	4	-0.264	-0.253	-0.182	-0.378	Muito grande	Muito grande
DDPG	1	-0.28	-0.268	-0.189	-0.393	Muito grande	Muito grande
	2	-0.272	-0.281	-0.18	-0.369	Muito grande	Muito grande
	3	-0.275	-0.275	-0.181	-0.376	Muito grande	Muito grande
	4	-0.256	-0.251	-0.194	-0.393	Muito grande	Muito grande
TD3	1	-0.277	-0.276	-0.195	-0.393	Muito grande	Muito grande
	2	-0.274	-0.269	-0.199	-0.379	Muito grande	Muito grande
	3	-0.276	-0.278	-0.182	-0.374	Muito grande	Muito grande
	4	-0.28	-0.285	-0.163	-0.41	Muito grande	Muito grande
SAC	1	-0.275	-0.276	-0.189	-0.368	Muito grande	Muito grande
	2	-0.265	-0.265	-0.259	-0.268	Muito grande	Muito grande
	3	-0.264	-0.264	-0.261	-0.267	Muito grande	Muito grande
	4	-0.264	-0.265	-0.248	-0.271	Muito grande	Muito grande

Os resultados do teste H de Kruskal-Wallis para o *Maximum Drawdown* são apresentados na Tabela 6.5. O teste foi aplicado para investigar a presença de diferenças estatisticamente significativas entre os cenários para cada algoritmo, tendo o *Maximum Drawdown* como métrica de avaliação. Assim como na análise anterior, foi calculado o tamanho do efeito (Eta quadrado) para mensurar a magnitude das diferenças observadas. Os valores encontrados indicam efeitos negligenciáveis ou pequenos, com todos os p-valores acima do nível de significância de 0,05. Esses resultados apontam que os diferentes cenários não impactaram significativamente o controle de perdas dos algoritmos, segundo a métrica de *Maximum Drawdown*.

Tabela 6.5: Resultados do teste H de Kruskal-Wallis para o Maximum Drawdown.

Algoritmo	Estatística H	P-valor	Eta quadrado	Interpretação do P-valor	Interpretação do tamanho do efeito
A2C	6,446	0,092	0,020	Não significativo	Pequeno
PPO	2,865	0,413	-0,001	Não significativo	Desprezível
DDPG	4,402	0,221	0,008	Não significativo	Desprezível
TD3	0,449	0,930	-0,015	Não significativo	Desprezível
SAC	6,165	0,104	0,018	Não significativo	Pequeno

6.1.3 Retorno Anual

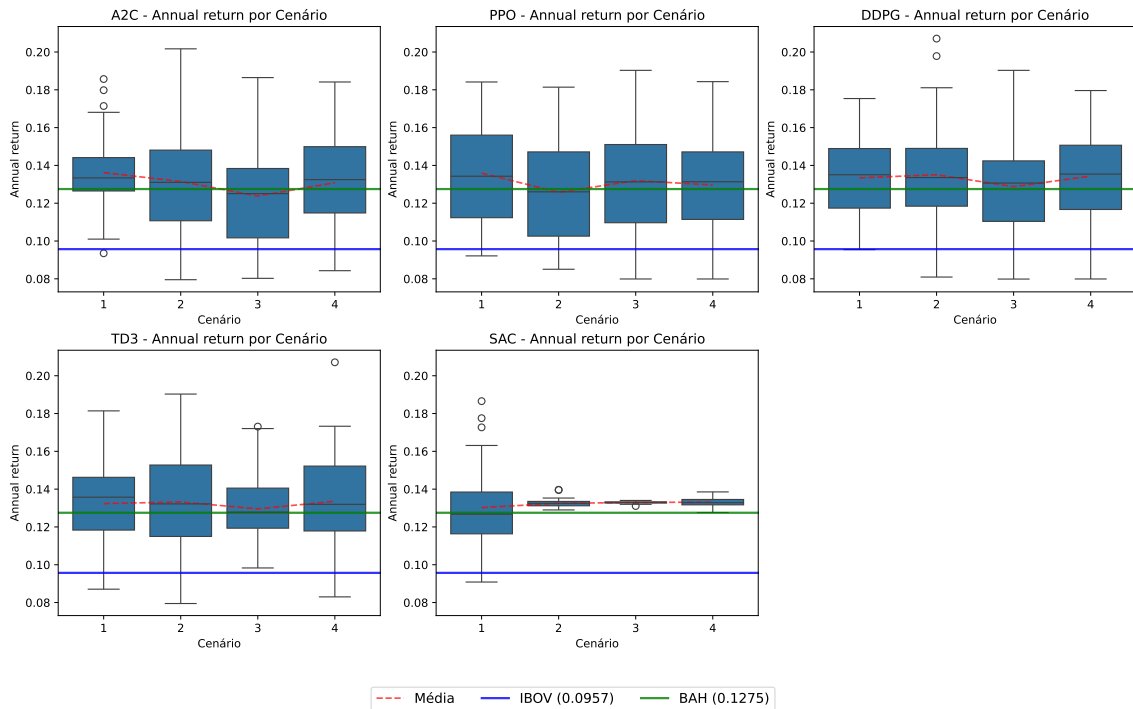


Figura 6.5: Distribuições do Retorno Anual: Análise por Algoritmo e Cenário.

A Tabela 6.6 resume os resultados de Retorno Anual obtidos por cada algoritmo nos diferentes cenários. São apresentados os valores de média, mediana, máximo e mínimo, bem como os tamanhos dos efeitos em comparação aos desempenhos do IBOV e da estratégia BAH.

Os resultados do teste H de Kruskal-Wallis para o *Retorno Anual* são evidenciados na Tabela 6.7. O objetivo da análise foi verificar se há diferenças estatisticamente significativas entre os cenários, considerando o retorno médio anual obtido por cada algoritmo. Também nesta métrica, os valores de Eta quadrado permaneceram próximos de zero, e os p-valores não indicaram significância estatística ($p > 0,05$). Dessa forma, conclui-se que, no contexto deste experimento, a variação entre os cenários não influenciou de forma relevante o desempenho dos algoritmos em termos de *Retorno Anual*.

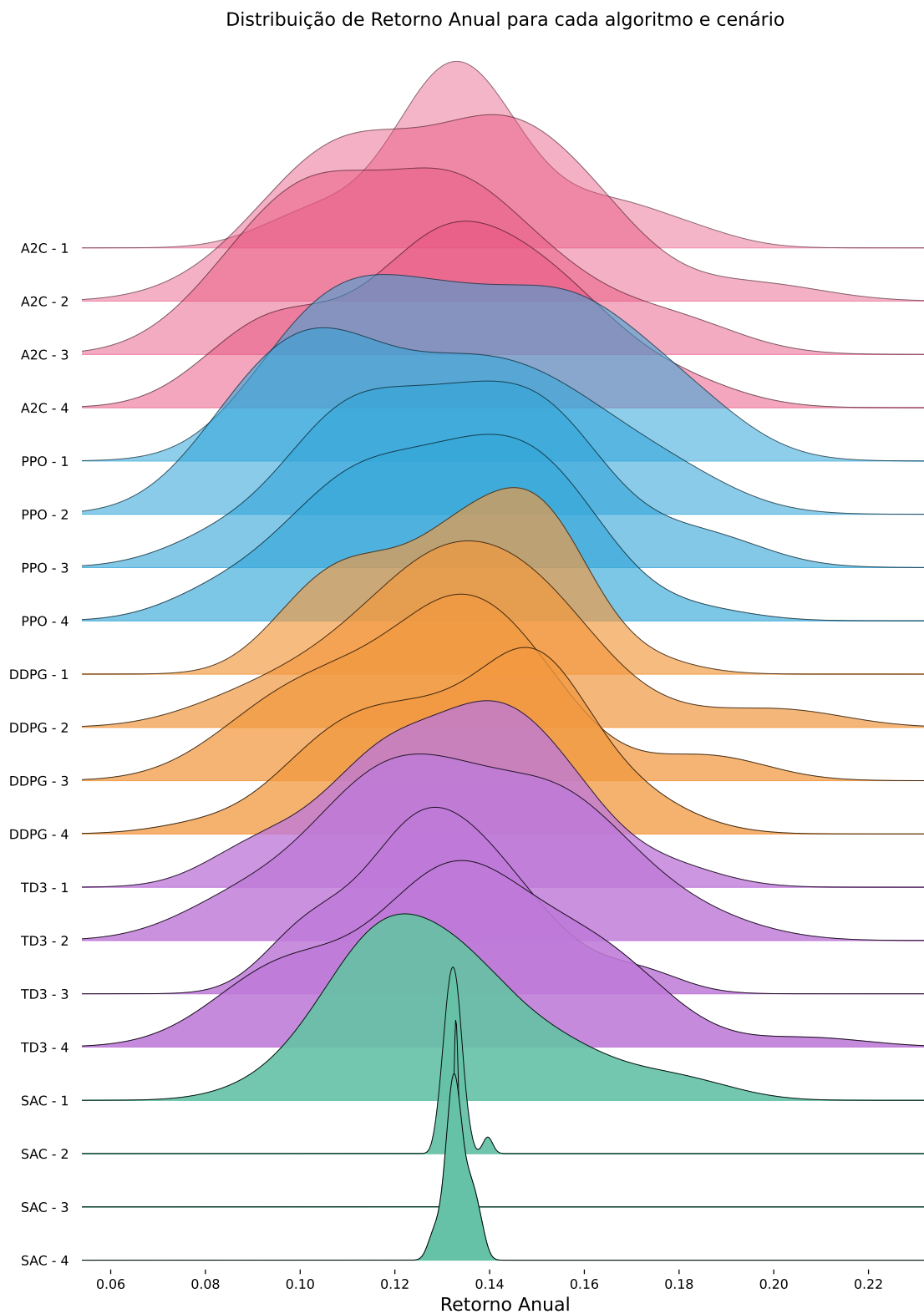


Figura 6.6: Análise das distribuições do Retorno Anual.

Tabela 6.6: Medidas estatísticas do Retorno Anual de cada algoritmo e cenário

Algoritmo	Cenário	Média	Mediana	Máximo	Mínimo	Efeito (Ibov)	Efeito (BAH)
A2C	1	0.136	0.133	0.186	0.094	Muito grande	Médio
	2	0.131	0.131	0.202	0.08	Muito grande	Pequeno
	3	0.124	0.125	0.186	0.08	Muito grande	Pequeno
	4	0.131	0.132	0.184	0.084	Muito grande	Pequeno
PPO	1	0.136	0.134	0.184	0.092	Muito grande	Médio
	2	0.126	0.126	0.181	0.085	Muito grande	Pequeno
	3	0.132	0.131	0.19	0.08	Muito grande	Pequeno
	4	0.13	0.131	0.184	0.08	Muito grande	Pequeno
DDPG	1	0.133	0.135	0.175	0.095	Muito grande	Efeito médio
	2	0.135	0.134	0.207	0.081	Muito grande	Médio
	3	0.129	0.131	0.19	0.08	Muito grande	Pequeno
	4	0.134	0.135	0.18	0.08	Muito grande	Médio
TD3	1	0.132	0.136	0.181	0.087	Muito grande	Médio
	2	0.133	0.132	0.19	0.079	Muito grande	Pequeno
	3	0.13	0.128	0.173	0.098	Muito grande	Pequeno
	4	0.134	0.132	0.207	0.083	Muito grande	Pequeno
SAC	1	0.13	0.127	0.187	0.091	Muito grande	Pequeno
	2	0.133	0.133	0.14	0.129	Muito grande	Muito grande
	3	0.133	0.133	0.134	0.131	Muito grande	Muito grande
	4	0.133	0.133	0.139	0.128	Muito grande	Muito grande

Tabela 6.7: Resultados do teste H de Kruskal-Wallis para o Retorno Anual.

Algoritmo	Estatística H	P-valor	Eta quadrado	Interpretação do P-valor	Interpretação do tamanho do efeito
A2C	5,433	0,143	0,014	Não significativo	Pequeno
PPO	3,473	0,324	0,003	Não significativo	Desprezível
DDPG	2,600	0,458	-0,002	Não significativo	Desprezível
TD3	0,966	0,809	-0,012	Não significativo	Desprezível
SAC	7,067	0,070	0,024	Não significativo	Pequeno

6.2 Discussões

No Cenário 1, PPO alcançou o maior Sharpe Ratio (0,709), enquanto TD3 destacou-se com o melhor Retorno Anual (0,136) e controle de queda. No Cenário 2, SAC registrou o maior *Sharpe Ratio* (0,693), e DDPG liderou em Retorno Anual (0,134). O Cenário 3 reforçou o forte desempenho do SAC, com o maior *Sharpe Ratio* (0,696) e RA (0,133), enquanto PPO demonstrou a menor queda. No Cenário 4, que incorporou todos os tipos de variáveis, DDPG superou outros modelos, alcançando o maior *Sharpe Ratio* (0,707) e Retorno Anual (0,135).

Em vários cenários, PPO, DDPG e SAC consistentemente superaram as referências. A inclusão de indicadores de sentimento e fundamentais pode ter apoiado o desempenho aprimorado do modelo, fornecendo um espaço de características mais rico para a tomada de decisões. Embora o teste de Kruskal-Wallis não tenha revelado diferença estatisticamente significativa entre os grupos, os modelos DRL superaram consistentemente ambas as referências, com *Sharpe Ratios* e Retornos Anuais mais elevados e quedas mais

controladas, potencialmente indicando um controle de risco superior devido a algoritmos adaptativos mais adequados às condições de mercado. Conforme detalhado na Tabela 6.6, a métrica de Retorno Anual era comparável entre os modelos treinados e as referências, particularmente BAH. A melhoria mais notável foi no *Sharpe Ratio*, principalmente devido aos modelos gerarem retornos semelhantes com menor risco, conforme evidenciado pela métrica de *Maximum Drawdown*. Esse aspecto os torna atraentes para investidores avessos ao risco que buscam retornos atrativos.

Conclusão

Neste estudo, uma abordagem de Aprendizado por Reforço Profundo foi empregada para otimizar portfólios de ativos brasileiros, integrando dados de mercado com indicadores de sentimento advindos de notícias em português, além de informações fundamentais. O desempenho dos modelos A2C, PPO, DDPG, TD3 e SAC foi comparado, e verificou-se que, embora os testes estatísticos não indicassem diferenças significativas entre eles, todos consistentemente superaram os *benchmarks* do Ibovespa e *Buy and Hold*. Vantagens notáveis foram observadas nos retornos ajustados pelo risco (*Sharpe Ratio*) e no controle do *Maximum Drawdown*, apontando para uma gestão de risco eficaz adquirida pelos agentes de Aprendizado por Reforço Profundo. A incorporação de dados de sentimento e indicadores fundamentalistas mostrou-se benéfica para lidar com as particularidades do mercado brasileiro. Este trabalho valida o Aprendizado por Reforço Profundo como uma estratégia promissora para otimização de portfólios no Brasil, especialmente quando enriquecida com informações qualitativas e fundamentalistas locais.

Há várias direções promissoras para a ampliação desta pesquisa. Primeiramente, aumentar o número de execuções por experimento poderia fornecer uma validação estatística mais robusta dos resultados. Além disso, a ampliação das iterações de treinamento dentro do ambiente pode permitir que os modelos alcancem uma convergência mais eficiente, potencialmente melhorando suas métricas de desempenho. Estudos futuros poderiam também investigar uma gama maior de algoritmos de aprendizado por reforço, a fim de avaliar diferenças em termos de generalização e desempenho. Outro ponto a considerar seria a inclusão de um número maior de ações na composição do portfólio, criando um cenário de investimento mais complexo que poderia aumentar os rendimentos potenciais. Estas extensões contribuiriam para um entendimento mais profundo das capacidades e limitações dos modelos em ambientes financeiros dinâmicos.

Referências Bibliográficas

- [Agarwal et al. 2023]AGARWAL, A. et al. Sentiment analysis in stock price prediction: A comparative study of algorithms. In: *2023 10th International Conference on Computing for Sustainable Global Development (INDIACom)*. [S.l.: s.n.], 2023. p. 1403–1407.
- [Alanyali, Moat e Preis 2013]ALANYALI, M.; MOAT, H. S.; PREIS, T. Quantifying the relationship between financial news and the stock market. *Scientific Reports*, v. 3, 2013. Disponível em: <<https://api.semanticscholar.org/CorpusID:12405643>>.
- [Arulkumaran et al. 2017]ARULKUMARAN, K. et al. Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine*, Institute of Electrical and Electronics Engineers (IEEE), v. 34, n. 6, p. 26–38, nov. 2017. ISSN 1053-5888. Disponível em: <<http://dx.doi.org/10.1109/MSP.2017.2743240>>.
- [Asyaky e Mandala 2021]ASYAKY, M. S.; MANDALA, R. Improving the performance of hdbscan on short text clustering by using word embedding and umap. In: *2021 8th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*. [S.l.: s.n.], 2021. p. 1–6.
- [B3]B3. *Ibovespa B3* / *B3*. https://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/ibovespa.htm. [Accessed 09-06-2024].
- [B3]B3. *Tarifas de Ações e Fundos de Investimento*. https://www.b3.com.br/pt_br/produtos-e-servicos/tarifas/listados-a-vista-e-derivativos/renda-variavel/tarifas-de-acoes-e-fundos-de-investimento/a-vista/. B3 - Brasil, Bolsa, Balcão. Accessed: 2025-04-17.
- [B3 2023]B3. *Número de investidores na B3 cresce 34% em renda fixa e 23% em renda variável em 12 meses*. 2023. [Accessed 30-04-2024]. Disponível em: <https://www.b3.com.br/pt_br/noticias/numero-de-investidores-na-b3-cresce-34-em-renda-fixa-e-23-em-renda-variavel-em-12-meses.htm>.
- [B3 2024]B3. *Balanço de empresa: como analisar para investir melhor*. 2024. Acessado em: 30 de junho de 2024. Disponível em:

<<https://borainvestir.b3.com.br/noticias/empresas/balanco-de-empresa-como-analisar-para-investir-melhor/>>.

[Bai et al. 2023]BAI, Z.-L. et al. Mercury: A deep reinforcement learning-based investment portfolio strategy for risk-return balance. *IEEE Access*, v. 11, p. 78353–78362, 2023. ISSN 2169-3536.

[Bakar e Rosbi 2018]BAKAR, N. A.; ROSBI, S. Diversification diagnostics for portfolio investment using combination of cryptocurrency and stock price. *International Journal of Advanced Research*, v. 6, p. 2320–5407, 10 2018.

[Barros 2015]BARROS, T. Análise técnica e fundamentalista: Ensaio sobre os métodos de análise. *Revista de Administração e Negócios da Amazônia*, v. 7, p. 39–63, 09 2015.

[Behera et al. 2023]BEHERA, J. et al. Prediction based mean-value-at-risk portfolio optimization using machine learning regression algorithms for multi-national stock markets. *Engineering Applications of Artificial Intelligence*, v. 120, p. 105843, 2023. ISSN 0952-1976. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0952197623000271>>.

[Benhamou et al. 2021]BENHAMOU, E. et al. Detecting and adapting to crisis pattern with context based deep reinforcement learning. In: *2020 25th International Conference on Pattern Recognition (ICPR)*. [S.l.: s.n.], 2021. p. 10050–10057. ISSN 1051-4651.

[Brini e Tantari 2023]BRINI, A.; TANTARI, D. Deep reinforcement trading with predictable returns. *Physica A: Statistical Mechanics and its Applications*, v. 622, 2023. Cited by: 0; All Open Access, Green Open Access.

[Byun, Kim e Wang 2020]BYUN, J.-S.; KIM, B.; WANG, H. *Proximal Policy Gradient: PPO with Policy Gradient*. 2020. Disponível em: <<https://arxiv.org/abs/2010.09933>>.

[Chatterjee, Ghosh e Chakrabarti 2016]CHATTERJEE, A.; GHOSH, A.; CHAKRABARTI, B. K. Universality of citation distributions for academic institutions and journals. *PLOS ONE*, Public Library of Science (PLoS), v. 11, n. 1, p. e0146762, jan. 2016. ISSN 1932-6203.

[Chong e Zak 2013]CHONG, E.; ZAK, S. *An Introduction to Optimization*. Wiley, 2013. (Wiley Series in Discrete Mathematics and Optimization). ISBN 9781118279014. Disponível em: <https://books.google.com.br/books?id=8J_ev5ihKEoC>.

[Cohen et al. 2003]COHEN, J. et al. *Applied multiple regression/correlation analysis for the behavioral sciences*. 3rd ed.. ed. Mahwah, N.J.: L. Erlbaum Associates, 2003. ISBN 0805822232.

- [Colasanto et al. 2022]COLASANTO, F. et al. Bert's sentiment score for portfolio optimization: a fine-tuned view in black and litterman model. *Neural Computing and Applications*, v. 34, 05 2022.
- [Colnaghi 2021]COLNAGHI, N. P. *A parametric portfolio optimization model using signals generated via machine learning*. 87 p. Dissertação (Dissertação de Mestrado) — Insper Instituto de Ensino e Pesquisa, São Paulo, SP, 2021. Disponível em: <<https://repositorio.insper.edu.br/handle/11224/2856>>.
- [Cong et al. 2020]CONG, L. W. et al. Alphaportfolio: Direct construction through deep reinforcement learning and interpretable ai. *Capital Markets: Asset Pricing & Valuation eJournal*, 2020.
- [Costa e Costa 2023]COSTA, C.; COSTA, A. Poe: A general portfolio optimization environment for finrl. In: *Anais do II Brazilian Workshop on Artificial Intelligence in Finance*. Porto Alegre, RS, Brasil: SBC, 2023. p. 132–143. ISSN 0000-0000.
- [Daiya e Lin 2021]DAIYA, D.; LIN, C. Stock movement prediction and portfolio management via multimodal learning with transformer. In: *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.: s.n.], 2021. p. 3305–3309. ISSN 2379-190X.
- [Debastiani e Russo 2008]DEBASTIANI, C.; RUSSO, F. *Avaliando Empresas, Investindo em Ações: A aplicação prática da análise fundamentalista na avaliação de empresas*. Novatec Editora, 2008. ISBN 9788575221792. Disponível em: <<https://books.google.com.br/books?id=jm-c9PY-QNsC>>.
- [Du e Tanaka-Ishii 2020]DU, X.; TANAKA-ISHII, K. Stock embeddings acquired from news articles and price history, and an application to portfolio optimization. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics, 2020. p. 3353–3363.
- [Edirisinghe e Zhang 2007]EDIRISINGHE, N.; ZHANG, X. Generalized dea model of fundamental analysis and its application to portfolio optimization. *Journal of Banking & Finance*, v. 31, n. 11, p. 3311–3335, 2007. ISSN 0378-4266. Risk Management and Quantitative Approaches in Finance. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0378426607001288>>.
- [Fernandes et al. 2019]FERNANDES, D. S. A. et al. Decision-making simulator for buying and selling stock market shares based on twitter indicators and technical analysis. In: *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. [S.l.: s.n.], 2019. p. 2626–2632.

- [François-Lavet et al. 2018]FRANÇOIS-LAVET, V. et al. An introduction to deep reinforcement learning. *Foundations and Trends® in Machine Learning*, Now Publishers, v. 11, n. 3–4, p. 219–354, 2018. ISSN 1935-8245. Disponível em: <<http://dx.doi.org/10.1561/22000000071>>.
- [Gao et al. 2022]GAO, R. et al. A novel densenet-based deep reinforcement framework for portfolio management. In: *2022 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*. [S.l.: s.n.], 2022. p. 158–165. ISSN 2833-8898.
- [Google Cloud 2020]Google Cloud. *Example-Based Explanations to Build Better AI/ML Models*. 2020. <https://cloud.google.com/blog/products/ai-machine-learning/example-based-explanations-to-build-better-aiml-models>. Accessed: 2024-07-04.
- [Google 2023]GOOGLE, G. T. Gemini: A family of highly capable multimodal models. *ArXiv*, abs/2312.11805, 2023. Disponível em: <<https://api.semanticscholar.org/CorpusID:266361876>>.
- [Gu, Jiang e Su 2021]GU, F.; JIANG, Z.; SU, J. Application of features and neural network to enhance the performance of deep reinforcement learning in portfolio management. In: *2021 IEEE 6th International Conference on Big Data Analytics (ICBDA)*. [S.l.: s.n.], 2021. p. 92–97.
- [Guan e Liu 2021]GUAN, M.; LIU, X.-Y. Explainable deep reinforcement learning for portfolio management: an empirical approach. *Proceedings of the Second ACM International Conference on AI in Finance*, 2021. Disponível em: <<https://api.semanticscholar.org/CorpusID:243848035>>.
- [Haarnoja et al. 2018]HAARNOJA, T. et al. *Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor*. 2018. Disponível em: <<https://arxiv.org/abs/1801.01290>>.
- [Han et al. 2023]HAN, D. et al. A survey on deep reinforcement learning algorithms for robotic manipulation. *Sensors*, v. 23, n. 7, 2023. ISSN 1424-8220. Disponível em: <<https://www.mdpi.com/1424-8220/23/7/3762>>.
- [Honnibal et al. 2020]HONNIBAL, M. et al. spacy: Industrial-strength natural language processing in python. 2020.

- [Huang, Miao e Hsiao 2021]HUANG, S.-H.; MIAO, Y.-H.; HSIAO, Y.-T. Novel deep reinforcement algorithm with adaptive sampling strategy for continuous portfolio optimization. *IEEE Access*, v. 9, p. 77371–77385, 2021. ISSN 2169-3536.
- [Huotari, Savolainen e Collan 2020]HUOTARI, T.; SAVOLAINEN, J.; COLLAN, M. Deep reinforcement learning agent for s&p 500 stock selection. *Axioms*, v. 9, n. 4, 2020. ISSN 2075-1680.
- [International Monetary Fund 2021]International Monetary Fund. *IMF Data Mapper*. 2021. <https://www.imf.org/external/datamapper/profile/WEOWORLD>. Accessed: 2024-06-30.
- [Jin 2023]JIN, B. A mean-var based deep reinforcement learning framework for practical algorithmic trading. *IEEE Access*, v. 11, p. 28920–28933, 2023. ISSN 2169-3536.
- [Kaufmann e Schering 2014]KAUFMANN, J.; SCHERING, A. Analysis of variance anova. In: _____. *Wiley StatsRef: Statistics Reference Online*. John Wiley & Sons, Ltd, 2014. ISBN 9781118445112. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat06938>>.
- [Khemlichi e al. 2022]KHEMLICHI, F.; AL. et. Deep deterministic policy gradient based portfolio management system. *International Journal of Information Science and Technology*, v. 6, n. 3, p. 29–39, September 2022. ISSN 2550-5114.
- [Khemlichi et al. 2023]KHEMLICHI, F. et al. Actor-critic methods in stock trading : A comparative study. In: *2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*. [S.l.: s.n.], 2023. p. 1–5.
- [Kitchenham 2004]KITCHENHAM, B. Procedures for performing systematic reviews. *Keele, UK, Keele Univ.*, v. 33, 08 2004.
- [Koratamaddi et al. 2021]KORATAMADDI, P. et al. Market sentiment-aware deep reinforcement learning approach for stock portfolio allocation. *Engineering Science and Technology, an International Journal*, v. 24, n. 4, p. 848–859, 2021. ISSN 2215-0986.
- [Kruskal e Wallis 1952]KRUSKAL, W. H.; WALLIS, W. A. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, [American Statistical Association, Taylor & Francis, Ltd.], v. 47, n. 260, p. 583–621, 1952. ISSN 01621459, 1537274X. Disponível em: <<http://www.jstor.org/stable/2280779>>.
- [Kurek et al. 2024]KUREK, J. et al. Zero-shot recommendation ai models for efficient job–candidate matching in recruitment process. *Applied Sciences*, v. 14, n. 6, 2024. ISSN 2076-3417. Disponível em: <<https://www.mdpi.com/2076-3417/14/6/2601>>.

- [Lagioia 2011]LAGIOIA, U. C. T. *Fundamentos do mercado de capitais*. 3. ed. [S.l.]: Editora Atlas, 2011.
- [Lee et al. 2023]LEE, G.-G. et al. *Gemini Pro Defeated by GPT-4V: Evidence from Education*. 2023.
- [Li et al. 2022]LI, Z. et al. Finrl-podracr: High performance and scalable deep reinforcement learning for quantitative finance. In: *Proceedings of the Second ACM International Conference on AI in Finance*. New York, NY, USA: Association for Computing Machinery, 2022. (ICAIF '21). ISBN 9781450391481.
- [Lin et al. 2023]LIN, R. et al. Dynamic portfolio optimization via augmented ddpq with quantum price levels-based trading strategy. In: *2023 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2023. p. 1–8. ISSN 2161-4407.
- [Liu et al. 2022]LIU, R. et al. Computer intelligent investment strategy based on deep reinforcement learning and multi-layer lstm network. In: *2022 IEEE 2nd International Conference on Data Science and Computer Application (ICDSCA)*. [S.l.: s.n.], 2022. p. 1006–1015.
- [Liu 2024]LIU, S. An evaluation of ddpq, td3, sac, and ppo: Deep reinforcement learning algorithms for controlling continuous system. In: *Proceedings of the 2023 International Conference on Data Science, Advanced Algorithm and Intelligent Computing (DAI 2023)*. Atlantis Press, 2024. p. 15–24. ISBN 978-94-6463-370-2. ISSN 1951-6851. Disponível em: <https://doi.org/10.2991/978-94-6463-370-2_3>.
- [Liu et al. 2022]LIU, X.-Y. et al. *FinRL: A Deep Reinforcement Learning Library for Automated Stock Trading in Quantitative Finance*. 2022. Disponível em: <<https://arxiv.org/abs/2011.09607>>.
- [Lopes, Antunes e Cardoso 2007]LOPES, C. F. L. R.; ANTUNES, L. P.; CARDOSO, M. A. S. Financiamento de longo prazo: mercado de debêntures e programa de emissão da bndespar. *Revista do BNDES*, BNDES, Rio de Janeiro, v. 14, n. 27, p. 43–70, Jun 2007.
- [Lucarelli e Borrotti 2020]LUCARELLI, G.; BORROTTI, M. A deep q-learning portfolio management framework for the cryptocurrency market. *Neural Computing and Applications*, v. 32, n. 23, p. 17229 – 17244, 2020. Cited by: 27; All Open Access, Hybrid Gold Open Access.
- [Malta e de Camargos 2016]MALTA, T. L.; de Camargos, M. A. Variáveis da análise fundamentalista e dinâmica e o retorno acionário de empresas brasileiras entre 2007 e 2014. *REGGE - Revista de Gestão*, v. 23, n. 1, p. 52–62, 2016. ISSN 1809-2276. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1809227616300066>>.

- [Mao et al. 2016]MAO, H. et al. Resource management with deep reinforcement learning. In: . New York, NY, USA: Association for Computing Machinery, 2016. (HotNets '16), p. 50–56. ISBN 9781450346610.
- [McInnes e Healy 2017]MCINNES, L.; HEALY, J. Accelerated hierarchical density based clustering. In: *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2017. Disponível em: <<http://dx.doi.org/10.1109/ICDMW.2017.12>>.
- [McInnes, Healy e Melville 2020]MCINNES, L.; HEALY, J.; MELVILLE, J. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. 2020.
- [Morciano et al. 2024]MORCIANO, G. et al. Use of recommendation models to provide support to dyslexic students. *Expert Systems with Applications*, v. 249, 2024. Cited by: 1; All Open Access, Green Open Access.
- [Nakagawa et al. 2017]NAKAGAWA, E. et al. *Revisão Sistemática da Literatura em Engenharia de Software: Teoria e Prática*. [S.l.]: Elsevier Brasil, 2017. ISBN 9788535285970.
- [Nalawati e Yuntari 2021]NALAWATI, R. E.; YUNTARI, A. D. Ratcliff/obershelp algorithm as an automatic assessment on e-learning. In: *2021 4th International Conference of Computer and Informatics Engineering (IC2IE)*. [S.l.: s.n.], 2021. p. 244–248.
- [Naseri, Rafiee e Moghadam 2020]NASERI, M. R. A.; RAFIEE, F. M.; MOGHADAM, S. K. Modeling portfolio optimization based on fundamental analysis using an expert system in the real estate industry. *International Journal of Supply and Operations Management*, v. 7, n. 1, p. 39 – 50, 2020. Cited by: 1.
- [Niu, Li e Li 2022]NIU, H.; LI, S.; LI, J. Metatrader: An reinforcement learning approach integrating diverse policies for portfolio optimization. In: *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. New York, NY, USA: Association for Computing Machinery, 2022. (CIKM '22), p. 1573–1583. ISBN 9781450392365.
- [Nocedal e Wright 2006]NOCEDAL, J.; WRIGHT, S. *Numerical Optimization*. Springer New York, 2006. (Springer Series in Operations Research and Financial Engineering). ISBN 9780387227429. Disponível em: <<https://books.google.com.br/books?id=7wDpBwAAQBAJ>>.
- [Olabode, Aremu e BAMISAIYE 2023]OLABODE, O.; AREMU, D.; BAMISAIYE, F. Stock trading optimization in emerging market of nigeria using deep reinforcement machine learning and equity market neutral strategy. *SSRN Electronic Journal*, 01 2023.

- [Paiva et al. 2021]PAIVA, F. C. L. et al. Intelligent trading systems: a sentiment-aware reinforcement learning approach. In: *Proceedings of the Second ACM International Conference on AI in Finance*. ACM, 2021. (ICAIF'21). Disponível em: <<http://dx.doi.org/10.1145/3490354.3494445>>.
- [Paiva et al. 2022]PAIVA, F. C. L. et al. Intelligent trading systems: A sentiment-aware reinforcement learning approach. In: *Proceedings of the Second ACM International Conference on AI in Finance*. New York, NY, USA: Association for Computing Machinery, 2022. (ICAIF '21). ISBN 9781450391481.
- [Pinto 2022]PINTO, V. M. M. *Estudo de aplicação de técnicas de aprendizado por reforço no problema de otimização de portfólio*. Dissertação (Dissertação de Mestrado) — Universidade Federal de Minas Gerais, Brasil, abril 2022. Disponível em: <http://hdl.handle.net/1843/49786>.
- [Raffin et al. 2025]RAFFIN, A. et al. *Stable-Baselines3 Documentation*. [S.l.], 2025. Accessed: 2025-01-19. Disponível em: <<https://stable-baselines3.readthedocs.io/en/master/>>.
- [Ratcliff e Metzener 1988]RATCLIFF, J. W.; METZENER, D. E. Pattern matching: The gestalt approach. *Dr. Dobb's Journal*, p. 46, July 1988.
- [Salvadori 2021]SALVADORI, B. L. *Seleção e alocação de ações baseadas em análise fundamentalista e técnica e MCDA*. 100 p. Dissertação (Dissertação (Mestrado em Produção)) — Instituto Tecnológico de Aeronáutica e Universidade Federal de São Paulo, São José dos Campos, 2021.
- [Santos, Santos e Marques 2021]SANTOS, G. F. da S.; SANTOS, C. A. da S.; MARQUES, N. L. A estratégia de buy and hold na composição da carteira de jovens investidores da baixada fluminense. *Caderno De Administração*, v. 29, n. 2, p. 6–31, 2021. Disponível em: <<https://doi.org/10.4025/cadadm.v29i2.57318>>.
- [Saravia 2022]SARAVIA, E. Prompt engineering guide. *GitHub Repository*, 12 2022. Disponível em: <<https://github.com/dair-ai/Prompt-Engineering-Guide>>.
- [Sawhney et al. 2021]SAWHNEY, R. et al. Quantitative day trading from natural language using reinforcement learning. In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Online: Association for Computational Linguistics, 2021. p. 4018–4030.
- [Schulman et al. 2017]SCHULMAN, J. et al. Proximal policy optimization algorithms. *CoRR*, abs/1707.06347, 2017. Disponível em: <<http://arxiv.org/abs/1707.06347>>.

- [Silva et al. 2025]SILVA, K. C. et al. Systematic review of portfolio optimization in the brazilian financial market: Integrating news and fundamental indicators with deep reinforcement learning. In: ARABNIA, H. R. et al. (Ed.). *Computational Science and Computational Intelligence*. Cham: Springer Nature Switzerland, 2025. p. 19–32. ISBN 978-3-031-94940-1.
- [Silva et al. 2025]SILVA, K. C. et al. Development and evaluation of a sentiment indicator based on financial news in portuguese. In: ARABNIA, H. R. et al. (Ed.). *Computational Science and Computational Intelligence*. Cham: Springer Nature Switzerland, 2025. p. 144–157. ISBN 978-3-031-94940-1.
- [Silva et al. 2025]SILVA, K. C. et al. A deep reinforcement learning approach for portfolio optimization of brazilian assets using fundamental and sentiment indicators. In: *2025 IEEE 49th Annual Computers, Software, and Applications Conference (COMPSAC)*. [S.l.: s.n.], 2025. p. 1879–1884.
- [Silva 2021]SILVA, R. F. da. *Automated stock trading system using deep reinforcement learning and price and sentiment prediction modules*. Tese (Doutorado), 2021.
- [Soleymani e Paquet 2020]SOLEYMANI, F.; PAQUET, E. Financial portfolio optimization with online deep reinforcement learning and restricted stacked autoencoder—deepbreath. *Expert Systems with Applications*, v. 156, 2020. Cited by: 53; All Open Access, Hybrid Gold Open Access.
- [Spearman 1904]SPEARMAN, C. The proof and measurement of association between two things. *The American Journal of Psychology*, University of Illinois Press, v. 15, n. 1, p. 72–101, January 1904. Disponível em: <<https://www.jstor.org/stable/1412159>>.
- [Spooner 1984]SPOONER, M. C. Origin of fundamental analysis. *Financial Analysts Journal*, Routledge, v. 40, n. 4, p. 79–80, 1984.
- [Sutton e Barto 2018]SUTTON, R. S.; BARTO, A. G. *Reinforcement Learning: An Introduction*. 2nd. ed. Cambridge, MA: MIT Press, 2018. ISBN 978-0262039246.
- [Suwanda, Syahputra e Zamzami 2020]SUWANDA, R.; SYAHPUTRA, Z.; ZAMZAMI, E. Analysis of euclidean distance and manhattan distance in the k-means algorithm for variations number of centroid k. *Journal of Physics: Conference Series*, v. 1566, p. 012058, 06 2020.
- [Taguchi, Sakaji e Izumi 2022]TAGUCHI, R.; SAKAJI, H.; IZUMI, K. Ssaam: Sentiment signal-based asset allocation method with causality information. In: *2022 IEEE International Conference on Big Data (Big Data)*. [S.l.: s.n.], 2022. p. 2373–2376.

- [Theodorou et al. 2021]THEODOROU, T.-I. et al. An ai-enabled stock prediction platform combining news and social sensing with financial statements. *Future Internet*, v. 13, n. 6, 2021. ISSN 1999-5903.
- [Thode 2002]THODE, H. C. *Testing for normality*. Hoboken, NJ: Taylor and Francis, 2002. Disponível em: <<https://cds.cern.ch/record/1989990>>.
- [Tyagi e Bhushan 2023]TYAGI, N.; BHUSHAN, B. Demystifying the role of natural language processing (nlp) in smart city applications: Background, motivation, recent advances, and future research directions. *Wireless Personal Communications*, v. 130, p. 857–908, 2023. Disponível em: <<https://doi.org/10.1007/s11277-023-10312-8>>.
- [Virtanen et al. 2020]VIRTANEN, P. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, v. 17, p. 261–272, 2020. Disponível em: <<https://doi.org/10.1038/s41592-019-0686-2>>.
- [Wang et al. 2023]WANG, D. et al. Multi-view enhanced graph attention network for session-based music recommendation. *ACM Trans. Inf. Syst.*, Association for Computing Machinery, New York, NY, USA, v. 42, n. 1, aug 2023. ISSN 1046-8188. Disponível em: <<https://doi-org.ez49.periodicos.capes.gov.br/10.1145/3592853>>.
- [Wang e Yu 2021]WANG, H.; YU, S. Robo-advising: Enhancing investment with inverse optimization and deep reinforcement learning. In: *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*. [S.l.: s.n.], 2021. p. 365–372.
- [Wang et al. 2024]WANG, L. et al. Multilingual e5 text embeddings: A technical report. *arXiv preprint arXiv:2402.05672*, 2024.
- [Wang et al. 2024]WANG, X. et al. Deep reinforcement learning: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, v. 35, n. 4, p. 5064–5078, 2024.
- [Wang 2022]WANG, Z. *Modern Question Answering Datasets and Benchmarks: A Survey*. 2022.
- [Wang et al. 2021]WANG, Z. et al. Deept trader: A deep reinforcement learning approach for risk-return balanced portfolio management with market conditions embedding. *Proceedings of the AAAI Conference on Artificial Intelligence*, v. 35, n. 1, p. 643–650, May 2021.
- [Whittington 1980]WHITTINGTON, G. Some basic properties of accounting ratios. *Journal of Business Finance & Accounting*, v. 7, p. 219–232, 1980. Disponível em: <<https://api.semanticscholar.org/CorpusID:154206719>>.

- [Yan et al. 2021]YAN, X. et al. A network view of portfolio optimization using fundamental information. *Frontiers in Physics*, v. 9, 2021. Cited by: 0; All Open Access, Gold Open Access.
- [Yang et al. 2023]YANG, M. et al. Spotlight news driven quantitative trading based on trajectory optimization. In: ELKIND, E. (Ed.). *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*. [S.l.]: International Joint Conferences on Artificial Intelligence Organization, 2023. p. 4930–4939.
- [Yi e Liu 2023]YI, J.; LIU, X. Deep reinforcement learning for intelligent penetration testing path design. *Applied Sciences*, v. 13, n. 16, 2023. ISSN 2076-3417.
- [Zhou et al. 2023]ZHOU, Y. et al. A deep temporal factor analysis method for large scale financial portfolio selection. In: *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.: s.n.], 2023. p. 1–5.
- [Zhu e Huang 2022]ZHU, H.; HUANG, J. A new method for determining the embedding dimension of financial time series based on manhattan distance and recurrence quantification analysis. *Entropy*, v. 24, n. 9, 2022. ISSN 1099-4300. Disponível em: <<https://www.mdpi.com/1099-4300/24/9/1298>>.
- [Zhu e Zhu 2022]ZHU, T.; ZHU, W. Quantitative trading through random perturbation q-network with nonlinear transaction costs. *Stats*, v. 5, n. 2, p. 546 – 560, 2022. Cited by: 4; All Open Access, Gold Open Access.
- [Zhuang, Chen e Qiu 2022]ZHUANG, W.; CHEN, C.; QIU, G. A new deep reinforcement learning model for dynamic portfolio optimization. *Journal of University of Science and Technology of China*, v. 52, n. 11, 2022. Cited by: 0; All Open Access, Hybrid Gold Open Access.