



Universidade de Brasília

Instituto de Ciências Exatas  
Departamento de Ciência da Computação

# Modelo Contextual de Negociação no Mercado de Ações Brasileiro com Algoritmo de Aprendizado por Reforço Profundo

Matheus Schmitz Oliveira

Dissertação apresentada como requisito parcial para  
conclusão do Mestrado em Informática

Orientador

Prof. Dr. Geraldo Pereira Rocha Filho

Brasília  
2023



# Dedicatória

Dedico esta dissertação a minha família, cujo apoio incondicional me proporcionou chegar ainda mais longe.

# Agradecimentos

Agradeço ao meu orientador, Prof. Dr. Geraldo Pereira da Rocha Filho, por todo apoio concedido durante o período do mestrado. Também agradeço ao Departamento de Ciência da Computação (CIC/UnB) pela constante contribuição em minha formação acadêmica.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), por meio do Acesso ao Portal de Periódicos.

# Resumo

As bolsas de valores estão presentes na sociedade ao longo dos últimos séculos, sendo fundamentais para a movimentação da economia e construção de grandes fortunas. No entanto, os preços das companhias oscilam, tornando desafiadora a tarefa de identificação das melhores oportunidades de compra e venda de suas ações. Nesse sentido, o uso de algoritmos para a negociação automática nas bolsas de valores ganhou evidência, apresentando características positivas como a velocidade e isenção de sentimentos na tomada de decisão. O Aprendizado por Reforço (RL) é aplicado em problemas envolvendo sequências de decisões em ambientes complexos, sendo promissor para a modelagem de ambientes de negociação de ativos. Apesar do grande avanço observado nas publicações recentes, foi identificada uma lacuna relacionada à combinação de dados numéricos do mercado e dados textuais provenientes de múltiplas fontes de informação. Dessa forma, o presente trabalho preenche a lacuna encontrada ao investigar, propor e validar o desenvolvimento de um modelo contextual baseado em Aprendizado por Reforço Profundo (DRL) para a negociação individualizada de ativos no mercado financeiro brasileiro. A proposta foi avaliada em quatro diferentes cenários, a partir de combinações entre a quantidade de dados utilizados e diferentes esquemas de recompensa adotado pelo agente de DRL do modelo contextual treinado. Para a avaliação foram escolhidos três *benchmarks*: investimento inicial, *Buy & Hold* da empresa específica e *Buy & Hold* do BOVA11. Os resultados mostraram que o modelo contextual desenvolvido superou o patrimônio líquido inicial investido em 94.5% dos casos no melhor caso. Além disso, os cenários que utilizaram o índice Sharpe como função de recompensa reportaram mais patrimônios líquidos finais acima dos *benchmarks* selecionados. Por fim, todos os cenários consideraram a simulação de taxas de transação cobradas pelas instituições financeiras, tornando o resultado ainda mais realista.

**Palavras-chave:** Aprendizado por Reforço, Análise de Sentimentos, Mercado de Ações Brasileiro, Negociação por Algoritmos

# Abstract

Stock exchanges have been present in society over the last few centuries, being fundamental for moving the economy and building great fortunes. However, company prices fluctuate, making the task of identifying the best opportunities for buying and selling their shares a challenge. In this sense, the use of algorithms for automatic trading on stock exchanges gained evidence, showing positive characteristics such as efficiency and support of feelings in decision-making. Reinforcement Learning is applied to problems involving sequences of decisions in complex environments, being promising for modeling asset trading environments. Despite the significant advance seen in recent news, a gap was identified related to the combination of numerical market data and textual data from multiple sources of information. Thus, the present work fills a gap found when investigating, proposing, and validating the development of a contextual model based on Deep Reinforcement Learning for the individualized trading of assets in the Brazilian financial market. The proposal was evaluated in four different scenarios, based on the combination between the amount of data used and various reward schemes adopted by the DRL agent of the trained contextual model. For the evaluation, three benchmarks were chosen: initial investment, Buy & Hold of the specific company and Buy & Hold of BOVA11. The results appreciated that the developed contextual model outperformed the invested initial equity in 94.5% of cases in the best case. Furthermore, the scenarios that used the Sharpe ratio as the reward function reported more net worth above the selected benchmarks. Finally, all scenarios consider the simulation of transaction fees charged by financial institutions, making the result even more realistic.

**Keywords:** Reinforcement Learning, Sentiment Analysis, Brazilian Stock Market, Algorithmic Trading

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Objetivos . . . . .	3
1.2	Contribuição . . . . .	3
1.3	Organização do Trabalho . . . . .	4
<b>2</b>	<b>Fundamentação Teórica</b>	<b>5</b>
2.1	Mercado de Ações . . . . .	5
2.1.1	Índices de Mercado . . . . .	6
2.1.2	Dados de Mercado . . . . .	8
2.1.3	Indicadores Técnicos . . . . .	9
2.1.4	Indicadores Macroeconômicos . . . . .	14
2.1.5	Negociação por Algoritmos . . . . .	14
2.2	Processamento de Linguagem Natural . . . . .	15
2.2.1	Reconhecimento de Entidades Nomeadas . . . . .	15
2.2.2	Análise de Sentimentos . . . . .	16
2.3	Séries Temporais . . . . .	17
2.3.1	Covariância e Correlação . . . . .	18
2.3.2	Análise de Risco . . . . .	19
2.4	Aprendizado de Máquina . . . . .	20
2.4.1	Redes Neurais Artificiais . . . . .	21
2.4.2	Redes Neurais Recorrentes . . . . .	23
2.5	Aprendizado por Reforço . . . . .	25
<b>3</b>	<b>Revisão Bibliográfica</b>	<b>30</b>
3.1	Aprendizado por Reforço no Mercado Financeiro . . . . .	30
3.2	Classificação de Textos no Mercado Financeiro . . . . .	33
3.3	Discussão . . . . .	34
<b>4</b>	<b>Metodologia</b>	<b>36</b>
4.1	Visão Geral . . . . .	36

4.2	Fontes de Dados . . . . .	37
4.2.1	Dados de Mercado . . . . .	37
4.2.2	Indicadores Técnicos . . . . .	38
4.2.3	Dados Textuais . . . . .	38
4.2.4	Dados Macroeconômicos . . . . .	43
4.3	Gerenciador de Dados . . . . .	44
4.3.1	Reconhecimento de Entidades Nomeadas . . . . .	45
4.3.2	Análise de Sentimentos . . . . .	46
4.4	Ambiente de Negociação . . . . .	49
4.5	Agente de DRL . . . . .	51
4.6	Formulação do Problema . . . . .	53
4.7	Considerações Finais . . . . .	54
<b>5</b>	<b>Resultados</b>	<b>55</b>
5.1	Experimentos . . . . .	55
5.2	Cenário (i): Todos os Dados e Recompensa por Lucro Simples . . . . .	57
5.3	Cenário (ii): Dados OHLCV e Recompensa por Lucro Simples . . . . .	58
5.4	Cenário (iii): Todos os Dados e Recompensa por Índice Sharpe . . . . .	59
5.5	Cenário (iv): Dados OHLCV e Recompensa por Índice Sharpe . . . . .	60
5.6	Análise dos Resultados por Ativo . . . . .	61
5.7	Considerações Finais . . . . .	63
<b>6</b>	<b>Conclusão</b>	<b>65</b>
6.1	Contribuições . . . . .	65
6.2	Trabalhos Futuros . . . . .	65
	<b>Referências</b>	<b>67</b>



# Lista de Figuras

2.1	Evolução da Taxa Selic e do IPCA desde 2018. . . . .	7
2.2	Evolução da Taxa Selic e do Ibovespa nos últimos 5 anos. . . . .	8
2.3	Agrupamento dos Contratos efetuados em Barras OHLCV de 1 minuto. . .	9
2.4	<i>Candlestick</i> OHLC Exemplificado (Adaptado de [1]). . . . .	9
2.5	Evolução do preço de fechamento diário e médias móveis para o papel EQTL3.	10
2.6	Comparação entre a SMA e EMA de 10 dias para o papel PETR4. . . . .	11
2.7	Média Móvel e Bandas de Bollinger para o papel VALE3. . . . .	12
2.8	Análise do indicador Índice de Força Relativa (RSI) nas cotações de fecha- mento diárias da PETR4. . . . .	13
2.9	Histórico de Preços de Fechamento das Ações Preferenciais da Petrobras (PETR4) entre Jan/2015 e Dez/2020. . . . .	17
2.10	Perceptron Simples (Frank Rosenblatt [2]). . . . .	22
2.11	Rede Neural do tipo <i>Feed Forward Neural Network</i> (FFNN). . . . .	22
2.12	Exemplificação de uma Rede Neural Recorrente (Figura Adaptada) <sup>1</sup> . . . .	23
2.13	Interior de uma Célula LSTM (Figura Adaptada) <sup>2</sup> . . . . .	24
2.14	Ciclo de Interação entre Agente e Ambiente. . . . .	26
4.1	Visão Geral dos Componentes do Modelo de Negociação por DRL. . . . .	37
4.2	Exemplo de Conteúdo Textual Extraído do Portal de Notícias <i>Suno Research</i> .	39
4.3	Exemplo de Conteúdo Textual Extraído do Portal de Notícias <i>Suno Research</i> .	40
4.4	Exemplo de <i>tweet</i> Extraído do Perfil do <i>Infomoney</i> . . . . .	42
4.5	Visão Geral do Componente Gerenciador de Dados. . . . .	44
4.6	Wordcloud gerada com títulos de notícias classificados como positivos. . . .	48
4.7	Primeiras Observações do Conjunto de Teste. . . . .	49
4.8	Interação entre o Agente de DRL e o Ambiente de Negociação. . . . .	52
5.1	Patrimônios líquidos finais no conjunto de teste para o <i>ticker</i> PETR4 em cada cenário de avaliação. . . . .	62

5.2 Patrimônios líquidos finais no conjunto de teste para o <i>ticker</i> ITUB4 em cada cenário de avaliação. . . . .	62
---	----

# Lista de Tabelas

3.1	Comparação de Características com os Trabalhos Relacionados. . . . .	35
4.1	Os 5 maiores setores que compõem o Ibovespa - Maio/22 a Agosto/22 . . . . .	38
4.2	Quantidade de Notícias por Empresa no portal Suno Research . . . . .	41
4.3	Quantidade de <i>tweets</i> nos Perfis Específicos . . . . .	42
4.4	Quantidade de <i>tweets</i> nos Perfis Relacionados . . . . .	43
4.5	Novos <i>tweets</i> com Referência às Companhias . . . . .	43
4.6	Divisão Treino/Teste dos Dados de Mercado e Indicadores Técnicos . . . . .	45
4.7	Quantidade de Notícias por Ano . . . . .	45
5.1	Hiperparâmetros testados e selecionados . . . . .	57
5.2	Patrimônios Líquidos Finais no Conjunto de Teste - Cenário (i) . . . . .	58
5.3	Patrimônios Líquidos Finais no Conjunto de Teste - Cenário (ii) . . . . .	59
5.4	Patrimônios Líquidos Finais no Conjunto de Teste - Cenário (iii) . . . . .	60
5.5	Patrimônios Líquidos Finais no Conjunto de Teste - Cenário (iv) . . . . .	60
5.6	Patrimônios finais agrupados por combinações entre as funções de recompensa e as fontes de dados . . . . .	63

# Lista de Abreviaturas e Siglas

**AI** Inteligência Artificial (do inglês, *Artificial Intelligence*).

**ANN** Redes Neurais Artificiais (do inglês, *Artificial Neural Networks*).

**API** Interface de Programação de Aplicações (do inglês, *Application Programming Interface*).

**AT** Negociação por Algoritmos (do inglês, *Algorithmic Trading*).

**BBANDS** Bandas de Bollinger (do inglês, *Bollinger Bands*).

**CNN** Redes Neurais Convolucionais (do inglês, *Convolutional Neural Networks*).

**DL** Aprendizado Profundo (do inglês, *Deep Learning*).

**DRL** Aprendizado por Reforço Profundo (do inglês, *Deep Reinforcement Learning*).

**EMA** Média Móvel Exponencial (do inglês, *Exponential Moving Average*).

**ETF** *Exchange Traded Fund*.

**FFNN** *Feed Forward Neural Networks*.

**GRU** Unidades Recorrentes Bloqueadas (do inglês, *Gated Recurrent Units*).

**LF** Função de Rotulagem (do inglês, *Labelling Function*).

**LSTM** *Long Short-Term Memory*.

**MACD** Média Móvel Convergente e Divergente (do inglês, *Moving Average Convergence Divergence*).

**MDP** Processo de Decisão de Markov (do inglês, *Markov Decision Process*).

**ML** Aprendizado de Máquina (do inglês, *Machine Learning*).

**MLP** Perceptron Multicamada (do inglês, *Multilayer Perceptron*).

**NER** Reconhecimento de Entidades Nomeadas (do inglês, *Named Entity Recognition*).

**NLP** Processamento de Linguagem Natural (do inglês, *Natural Language Processing*).

**OMS** Sistema de Gerenciamento de Ordens (do inglês, *Order Management System*).

**PGM** Métodos de Política por Gradiente (do inglês, *Policy Gradient Methods*).

**PPO** Otimização de Política Proximal (do inglês, *Proximal Policy Optimization*).

**RBM** Métodos Baseados em Regras (do inglês, *Ruled-Based Methods*).

**RL** Aprendizado por Reforço (do inglês, *Reinforcement Learning*).

**RNN** Redes Neurais Recorrentes (do inglês, *Recurrent Neural Networks*).

**RSI** Índice de Força Relativa (do inglês, *Relative Strength Index*).

**SA** Análise de Sentimentos (do inglês, *Sentiment Analysis*).

**ShR** Índice Sharpe (do inglês, *Sharpe Ratio*).

**SMA** Média Móvel Simples (do inglês, *Simple Moving Average*).

**SrR** Índice Sortino (do inglês, *Sortino Ratio*).

**WS** Supervisão Fraca (do inglês, *Weak Supervision*).

# Capítulo 1

## Introdução

Historicamente, a busca por riquezas foi responsável por grandes expedições, guerras e revoluções. De maneira gradual, alterações significativas ocorreram no modelo de transferência do capital, desde o escambo até a negociação de criptomoedas digitais. Em seu livro *best-seller Sapiens* [3], Yuval Noah Harari aponta as crenças coletivas como um dos pontos de sucesso da espécie humana, dentre as quais destacam-se o dinheiro e a religião. O sistema de confiança relativo ao dinheiro permite a sua adoção de forma universal, proporcionando a troca econômica entre diferentes sociedades, mesmo que estas apresentem divergências culturais, religiosas ou geográficas.

O mercado de ações é o ambiente no qual são negociadas frações do capital aberto das principais empresas globais. Pequenos investidores e grandes instituições financeiras o utilizam como um meio para tentar gerar lucros e multiplicar o capital investido [4]. Tal ambiente apresentou grande retorno acumulado comparado aos investimentos em ouro e títulos de renda fixa no decorrer do tempo, sendo considerado até os dias atuais como um dos principais meios de multiplicação de capital [4]. As bolsas de valores possibilitam aos investidores, dentre outras modalidades de investimento, a aquisição de participações em empresas por meio da compra de ações. É possível investir tanto em companhias mais consolidadas e tradicionais, quanto em pequenas e médias empresas com potencial de crescimento. Atualmente, a quase totalidade das ordens de negociação é executada de maneira eletrônica. O estudo realizado pelo *Kissell Research Group* destaca que, a partir de 2019, aproximadamente 99.9% do volume de contratos negociados foi realizado por meios eletrônicos, sendo que aproximadamente 92% do total utilizou algoritmos computacionais para a execução das ordens [5].

O avanço da ciência da computação, sobretudo na área de Aprendizado de Máquina (do inglês, *Machine Learning*) (ML), possibilitou a análise de grandes volumes de dados provenientes de diferentes fontes de informação para auxiliar na tomada de decisão final de compra e venda de ativos financeiros [6, 7]. No entanto, identificar e explorar a combinação

de diferentes fontes de informação em conjunto com dados históricos do mercado não é uma tarefa trivial. Atualmente, a predição da movimentação de preços no mercado de ações ainda é um grande desafio para as grandes instituições financeiras, analistas profissionais e investidores mais experientes [8, 9]. Além disso, o mercado é dinâmico e muitos são os fatores que influenciam na volatilidade e direcionamento diários, o que torna desafiador o desenvolvimento de sistemas computacionais que executam ordens de negociação em tempo real e de forma automatizada [10].

O domínio das finanças é extenso e proporciona a aplicação de técnicas computacionais para o aperfeiçoamento dos resultados em tarefas como a predição do movimento de mercado [11, 12] e a otimização do gerenciamento de ativos em carteira com controle de risco [13]. Soler-Dominguez *et al.* [14] apresentaram aplicações de metaheurísticas para a otimização de portfólios, precificação de opções e avaliação de risco de crédito. Kaczorowski *et al.* [15] aplicaram regressão logística múltipla e redes neurais para a predição do índice brasileiro de *small-caps*. Marcos López de Padro explorou o avanço das melhores práticas em computação para o desenvolvimento de aplicações com técnicas de ML e computação de alta performance em [16]. Nesse contexto, os algoritmos de ML combinados com os avanços conquistados na área de Processamento de Linguagem Natural (do inglês, *Natural Language Processing*) (NLP) apresentam soluções para resolver problemas de predição no mercado financeiro [17]. No entanto, utilizar somente os dados numéricos relacionados aos sentimentos extraídos das notícias pode trazer limitações consideráveis na construção de um mecanismo completo de negociação, tendo em vista os desafios enfrentados em obter classificações de qualidade em textos [18].

Existem diversos trabalhos na literatura que apresentam estratégias de negociação no mercado de ações com algoritmos [19, 20, 21, 22]. Vidyamurthy [19] apresentou uma estratégia determinística denominada *pairs trading*, em que dois ativos são negociados de forma conjunta, mas em direções opostas. Wu *et al.* [20] apresentaram uma estratégia de negociação com árvores de decisão. Outros trabalhos exploram somente os dados de mercado para a negociação de ativos na bolsa de valores [21, 22] ou investigam a relação do sentimento extraído nos textos no domínio das finanças com a movimentação de ativos e índices de mercado [23, 24]. Entretanto, existe uma lacuna pouco explorada nas publicações recentes, que envolve a combinação dos dados textuais de notícias sobre o mercado financeiro e os dados de mercado, tais como as cotações diárias, indicadores técnicos e dados macroeconômicos.

Estudos mais recentes apontam a área de Aprendizado por Reforço Profundo (do inglês, *Deep Reinforcement Learning*) (DRL) como uma alternativa aos métodos determinísticos utilizados na negociação de ativos do mercado financeiro [21, 25, 26, 27]. O Aprendizado por Reforço (do inglês, *Reinforcement Learning*) (RL) relaciona-se ao ML e

estuda a criação de agentes capazes de interagir em ambientes diversos, os quais realizam ações para maximizar uma recompensa acumulada ao longo do tempo [28].

Nesse sentido, esta abordagem entra em destaque no contexto da Negociação por Algoritmos (do inglês, *Algorithmic Trading*) (AT) por explorar padrões não triviais a partir de uma grande quantidade de dados de mercado. Além disso, com a ampla variedade de algoritmos determinísticos utilizados no mercado financeiro há anos, torna-se ainda mais relevante explorar novas metodologias de negociação. Nesse sentido, o DRL expande o RL tradicional ao utilizar as redes neurais provenientes do Aprendizado Profundo (do inglês, *Deep Learning*) (DL) para construir as funções que descrevem os principais componentes do agente (e. g. política de negociação) [29].

Apesar dos avanços reportados, poucos estudos apresentam uma abordagem contextual, que busca explorar a combinação de múltiplas fontes de informação na negociação de ativos do mercado financeiro brasileiro com o uso de agentes de Aprendizado por Reforço Profundo (do inglês, *Deep Reinforcement Learning*) (DRL) para a tomada de decisão de compra e venda de ativos selecionados, como explorado nesta dissertação.

## 1.1 Objetivos

O principal objetivo desta dissertação é investigar, propor e validar um modelo contextual que possibilite a negociação individualizada de ativos no mercado financeiro brasileiro utilizando um agente de DRL. Para alcançar o objetivo geral, os seguintes objetivos específicos foram definidos:

- Pesquisar trabalhos voltados às estratégias de negociação por algoritmos.
- Coletar dados do mercado financeiro brasileiro.
- Propor um modelo de DRL para a negociação de ativos.
- Definir o ambiente de negociação que o agente de DRL irá explorar.
- Analisar o desempenho da proposta para ativos da bolsa de valores brasileira em diferentes cenários de avaliação.
- Avaliar os resultados obtidos dentro das configurações propostas.

## 1.2 Contribuição

As principais contribuições deste trabalho são:



1. Criação de uma base de dados composta de diferentes tipos de dados do mercado financeiro.
2. Desenvolvimento de um classificador de sentimentos em textos curtos com abordagem supervisionada fraca.
3. Aplicação do DRL na negociação de ativos do mercado financeiro brasileiro.

### **1.3 Organização do Trabalho**

O Capítulo 2 apresenta os principais conceitos teóricos relacionados ao tema da pesquisa. São apresentados tópicos referentes aos mercados de ações, negociação por algoritmos, processamento de textos curtos, séries temporais, aprendizado de máquina e aprendizado por reforço.

No Capítulo 3 são exibidos os trabalhos relacionados a esta dissertação. As publicações são divididas em duas subseções: (i) aplicações do DRL na negociação de ativos; e (ii) classificação textual no mercado financeiro.

Em seguida, o Capítulo 4 apresenta a visão geral da proposta, detalhando o funcionamento dos módulos e componentes a serem desenvolvidos durante o projeto.

O Capítulo 5 expõe os resultados obtidos em diferentes cenários de avaliação, bem como as análises sobre os mesmos.

Por fim, o Capítulo 6 finaliza o documento, apresentando um breve resumo do que foi realizado e as principais conclusões obtidas ao final dos experimentos.

# Capítulo 2

## Fundamentação Teórica

Este capítulo apresenta a fundamentação teórica acerca do mercado de ações, abordando assuntos relacionados ao mercado de renda variável, os principais índices de mercado, dados típicos de mercado, indicadores técnicos, dados macroeconômicos e técnicas de negociação por algoritmos. Em seguida, são apresentadas técnicas de processamento de textos, como a classificação de sentimentos e reconhecimento de entidades nomeadas. Adiante, serão apresentados conceitos essenciais sobre dados relacionados a séries temporais. Por último, serão abordadas as áreas de aprendizado de máquina e aprendizado por reforço profundo.

### 2.1 Mercado de Ações

O mercado de ações desempenha importante papel na economia mundial. As bolsas de valores são responsáveis pela negociação de títulos públicos, ações de empresas, contratos de *commodities*, dentre outros. A [B]<sup>3</sup> (Brasil, Bolsa, Balcão) é a bolsa oficial do Brasil, com sede na cidade de São Paulo. Foi oficialmente estabelecida em seu formato atual a partir de 2017, após a realização de fusões entre outras bolsas nacionais existentes. Atualmente, existem mais de 400 empresas listadas, com potencial aumento quantitativo em virtude do futuro crescimento econômico do país e captação de investimentos do exterior.

A [B]<sup>3</sup> recebeu um número crescente de novos investidores nos últimos anos. Segundo consta no site oficial<sup>1</sup>, o número de CPFs cadastrados obteve um aumento expressivo quando comparado com o final do ano de 2018, em que passou de 813.291 para 3.970.384 investidores inscritos em 2021 (dados referentes ao acumulado até o final de Setembro de 2021). Embora seja um número de investidores pequeno em relação à população do país, há uma tendência crescente de ampliação nesse quantitativo. Tal movimento é justificado

---

<sup>1</sup><http://www.b3.com.br/>

pela busca de investimentos com retornos mais elevados em relação aos títulos de renda fixa tradicionais, que no período de pandemia apresentaram rendimentos pouco lucrativos devido a recorrentes quedas sofridas pela taxa básica de juros da economia brasileira. Além disso, a crescente popularização de conteúdos relacionados a educação financeira, por meio de portais de notícias, aplicativos, influenciadores digitais e propagandas também estimula a inserção de novos investidores no mercado de renda variável.

O mercado de ações proporciona ao investidor exposição a uma ampla possibilidade de investimentos. No entanto, o risco associado a cada investimento deve ser analisado previamente, o que nem sempre é uma tarefa trivial. Ao escolher as empresas para investir, é prudente realizar análises, que podem ser divididas em duas categorias principais: (i) *fundamentalista*: ao se avaliar aspectos como a governança corporativa, cultura organizacional, indicadores de balanço patrimonial e classificação setorial; e (ii) *técnica*., ao se avaliar o histórico de cotações e volume de negociações, a fim de identificar padrões e tendências na variação nos preços.

O valor de mercado das empresas leva em consideração a quantidade de ações disponíveis e o preço negociado para cada ação em um determinado momento. As variações nas cotações ocorrem conforme novas transações entre os títulos são realizadas, em que um possuidor do papel deseja vender sua participação na empresa para um comprador interessado. Muitos são os fatores que influenciam nessa variação de preços. Dentre eles, os fatores macroeconômicos tais como a taxa de juros, cotação do câmbio, inflação, taxa de desemprego, questões políticas e sanitárias afetam de forma sistêmica em todas as empresas. Outros fatores como as fusões entre empresas, aquisições e vendas específicas são classificados como não-sistêmicos e afetam somente um conjunto restrito de companhias.

### 2.1.1 Índices de Mercado

Os índices de mercado são referências genéricas ao investidor e representam uma carteira teórica composta por um número fixo de ações durante certo período. O principal índice brasileiro é o índice Bovespa (Ibovespa)<sup>2</sup>. O Ibovespa foi criado em 1968 e tornou-se referência para os investidores locais e internacionais. As ações listadas são atualizadas a cada quatro meses (na primeira segunda-feira de Janeiro, Maio e Setembro) e devem seguir critérios para a alocação no índice. No último rebalanceamento de 2021, o Ibovespa passou a ser composto por 91 ações de 87 empresas. No entanto, o índice brasileiro concentra grande parte de seu percentual em companhias que negociam *commodities* e em instituições do setor bancário.

---

<sup>2</sup>[http://www.b3.com.br/pt\\_br/market-data-e-indices/indices/indices-amplos/ibovespa.htm](http://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/ibovespa.htm)

Além do Ibovespa, os índices americanos S&P 500 (SPX) e *Dow Jones* (DJIA), o alemão DAX e o japonês *Nikkei* 255 são exemplos de índices relevantes na economia internacional. O índice VIX (do inglês *Chicago Board Options Exchange Volatility Index*), também referenciado por “índice do medo”, indica a volatilidade esperada para os próximos 30 dias, derivada das opções de compra e venda do índice americano S&P 500. Em momentos de pessimismo, por exemplo, os investidores podem comprar opções de venda (PUTs) para se protegerem de eventuais quedas no mercado. O aumento da cotação das opções é observado e representado na pontuação do índice VIX.

O Índice de Preços ao Consumidor Amplo (IPCA) é utilizado para acompanhar a inflação no Brasil. É representado por um número que indica a mudança de preços em produtos e serviços distribuídos em nove categorias (Alimentação e bebidas, Habitação, Artigos de residência, Vestuário, Transportes, Saúde e cuidados pessoais, Despesas pessoais, Educação e Comunicação). O Instituto Brasileiro de Geografia e Estatística (IBGE) é a instituição responsável por medir o IPCA. A variação positiva no índice indica a existência de inflação, que afeta diretamente no poder de compra da moeda. O investidor que deseja proteger o seu patrimônio do efeito da inflação pode investir no tesouro IPCA, que é um título público pré-fixado que paga a variação do IPCA na data de vencimento adicionado de um bônus. Com o objetivo de conter o efeito adverso da inflação, o Banco Central atualiza periodicamente a Selic<sup>3</sup>, que é a taxa de juros base da economia brasileira. O aumento da Selic encarece a concessão de empréstimos e financiamentos, o que reduz o consumo e, conseqüentemente, os preços no geral. A Figura 2.1 apresenta a evolução da Taxa Selic em paralelo com o IPCA acumulado nos últimos 12 meses.

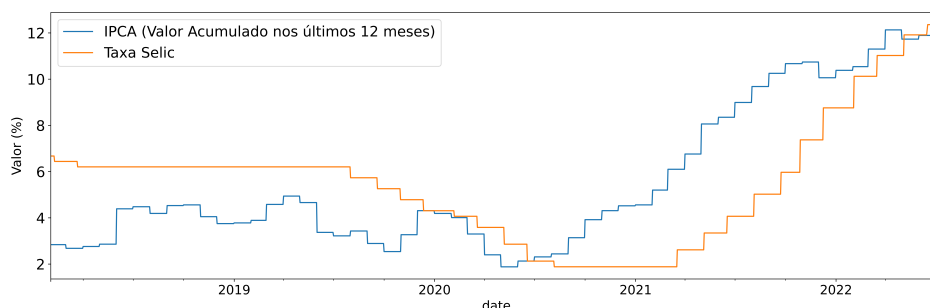


Figura 2.1: Evolução da Taxa Selic e do IPCA desde 2018.

A Figura 2.2 apresenta a evolução da taxa Selic em paralelo com o índice Bovespa. Observa-se que nos períodos de baixa taxa de juros, o índice Ibovespa tende a apresentar crescimento. Tendo em vista que os rendimentos de renda fixa atrelados à Selic também começam a render menos, o investidor pode buscar maiores retornos na bolsa de valores.

<sup>3</sup><https://www.bcb.gov.br/controleinflacao/taxaselic>

Os dados da Selic foram coletados a partir da API oficial do Banco Central do Brasil<sup>4</sup>. Os dados do Ibovespa foram coletados a partir da API do Yahoo! Finance.

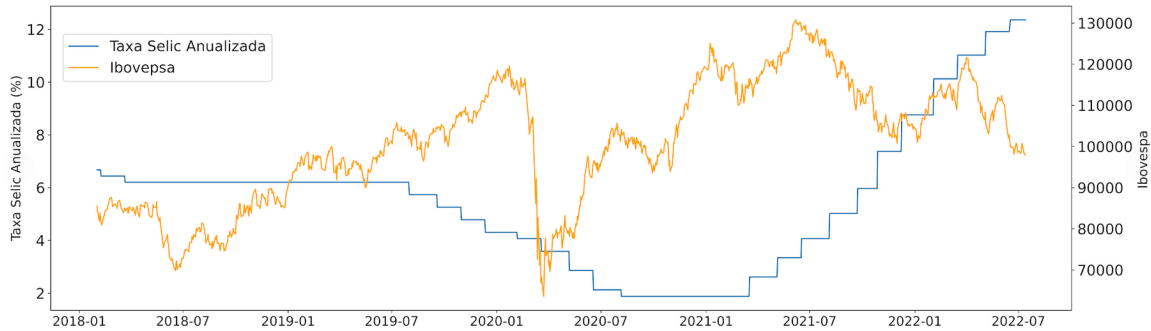


Figura 2.2: Evolução da Taxa Selic e do Ibovespa nos últimos 5 anos.

## 2.1.2 Dados de Mercado

As transações de ativos na [B]<sup>3</sup> são registradas por contratos denominados *trades*. Esses contratos contêm dados referentes ao momento da negociação (com precisão de milissegundos), ao *ticker* do ativo alvo, ao preço da compra, à quantidade de ações por contrato, ao identificador do comprador, ao identificador do vendedor e ao identificador único do *trade* no dia, sendo este iniciado pelo número 10 (para o primeiro *trade* do dia) e incrementado em múltiplos de 10 de forma sequencial. Em termos quantitativos, foram realizados mais de 500 mil *trades* para o *ticker* PETR4 na semana entre os dias 27 de Setembro de 2021 e 01 de Outubro de 2021, o que corresponde a um volume financeiro de aproximadamente R\$ 392.511.900 de acordo com o portal Yahoo! Finance<sup>5</sup>.

No entanto, em cada novo *trade* existe um valor base para o qual as ações são negociadas. O acompanhamento individual de cada transação é custoso em termos de armazenamento e processamento da informação. Em face desse desafio, os dados são comumente agrupados no tempo, em intervalos de minutos, horas ou dias. Os gráficos OHLCV (*Open, High, Low, Close and Volume*) são utilizados na análise gráfica de ativos do mercado financeiro, contendo informações referentes a um conjunto de negociações realizadas em períodos de tempo definidos.

A Figura 2.3 apresenta o agrupamento dos contratos processados pela [B]<sup>3</sup> em períodos de 1 minuto para o *ticker* PETR4 no dia 01 de Outubro de 2021.

A Figura 2.4 apresenta a representação gráfica de um *candlestick* OHLCV. A barra é representada pela cor verde pois o preço de fechamento foi superior ao preço de abertura.

<sup>4</sup><https://dadosabertos.bcb.gov.br/dataset/11-taxa-de-juros---selic>

<sup>5</sup><https://finance.yahoo.com/quote/PETR4.SA/history>

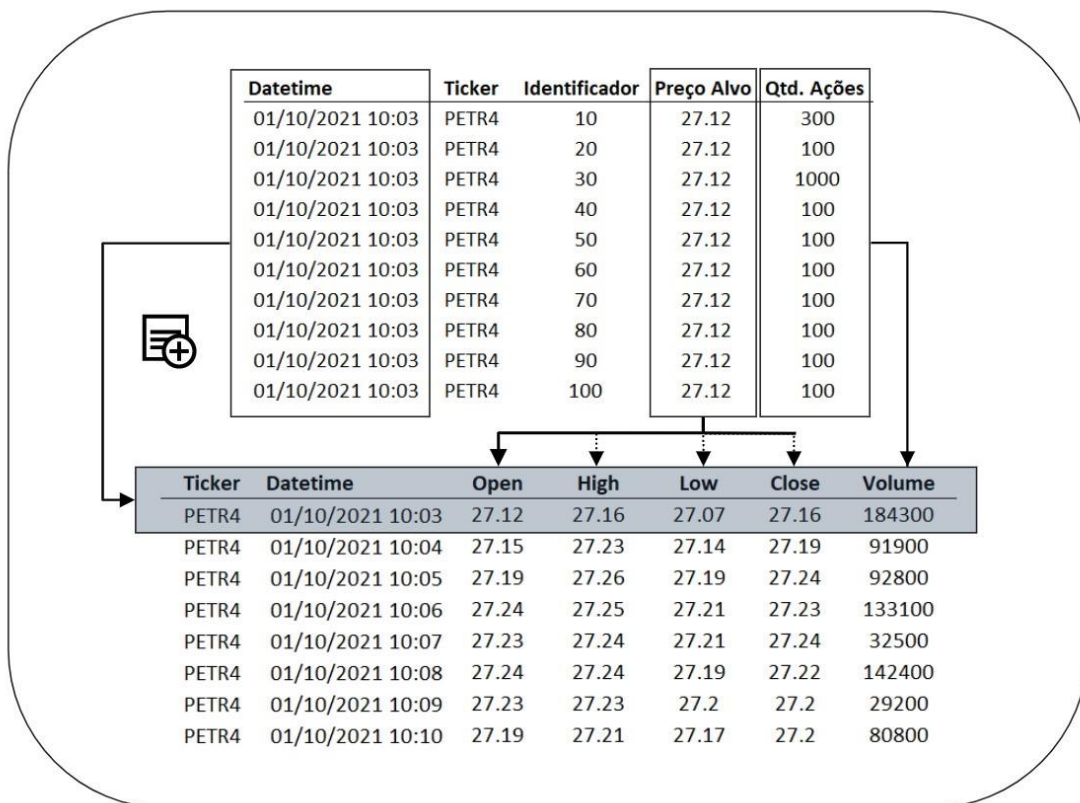


Figura 2.3: Agrupamento dos Contratos efetuados em Barras OHLCV de 1 minuto.

Do contrário, caso o preço de fechamento for menor que o preço de abertura, a barra seria representada pela cor vermelha.

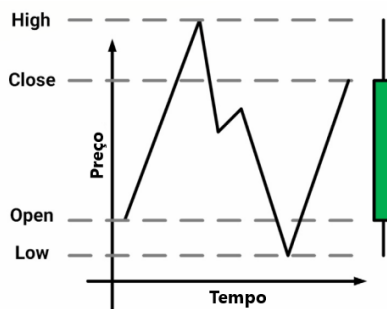


Figura 2.4: *Candlestick* OHLC Exemplificado (Adaptado de [1]).

### 2.1.3 Indicadores Técnicos

Os indicadores técnicos são dados numéricos obtidos a partir dos dados históricos de um determinado ativo. Os dados históricos contém preços e volume negociados. Estes

indicadores fornecem ao investidor sinais para a predição na movimentação de preços ou volatilidade futura, por meio de heurísticas ou padrões pré-determinados [30].

A Média Móvel Simples (do inglês, *Simple Moving Average*) (SMA) é um dos indicadores técnicos mais triviais, pois computa a média aritmética simples do preço de um ativo para um determinado número de observações. As SMAs calculadas com períodos menores (e. g. 10 dias) são mais voláteis, e apresentam a variação de curto prazo observada na série história. Já as SMAs com períodos maiores (e. g. 200 dias) apresentam maior resistência à variação, e são utilizadas para observar o comportamento de longo prazo da série de preços. A Figura 2.5 apresenta as curvas de preço de fechamento e as médias móveis curtas e longas para as ações da Equatorial Energia (EQTL3):

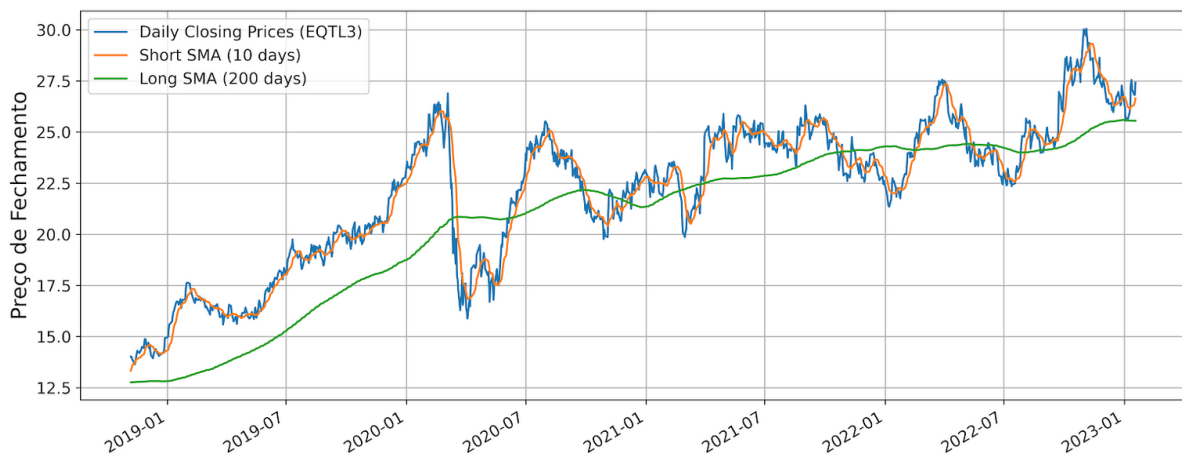


Figura 2.5: Evolução do preço de fechamento diário e médias móveis para o papel EQTL3.

A Equação 2.1 apresenta o cálculo da SMA:

$$SMA_t = \sum_{i=1}^n \frac{Close_{t-i}}{n} \quad (2.1)$$

Em que:

- $Close_t$  é o preço de fechamento do ativo no instante  $t$ ;
- $n$  é o tamanho da janela escolhida;

As SMAs podem ser utilizadas para produzir sinais de entrada e saída em posições de um determinado ativo. Por exemplo, ao utilizar duas SMAs, uma curta e uma longa, caso a SMA curta cruze a longa, pode ser um indicativo de tendência de alta no ativo, revelando um potencial momento de compra. A saída da posição pode ser efetuada em diferentes cenários: (i) atribuindo um limite máximo de perda (denominado *stop-loss*);

(ii) atribuindo um limite mínimo de ganho (denominado *take-profit*); ou (iii) ao cruzar novamente a SMA de longo prazo.

A Média Móvel Exponencial (do inglês, *Exponential Moving Average*) (EMA) também é um indicador de média móvel, similar às SMAs. No entanto, adiciona pesos maiores nas observações mais recentes. A Figura 2.6 apresenta a comparação entre a SMA e a EMA de mesma janela nos preços de fechamento diários do ativo PETR4. Observa-se que a linha vermelha (representando a EMA) apresenta mudanças mais rápidas em relação à linha verde (representando a SMA).

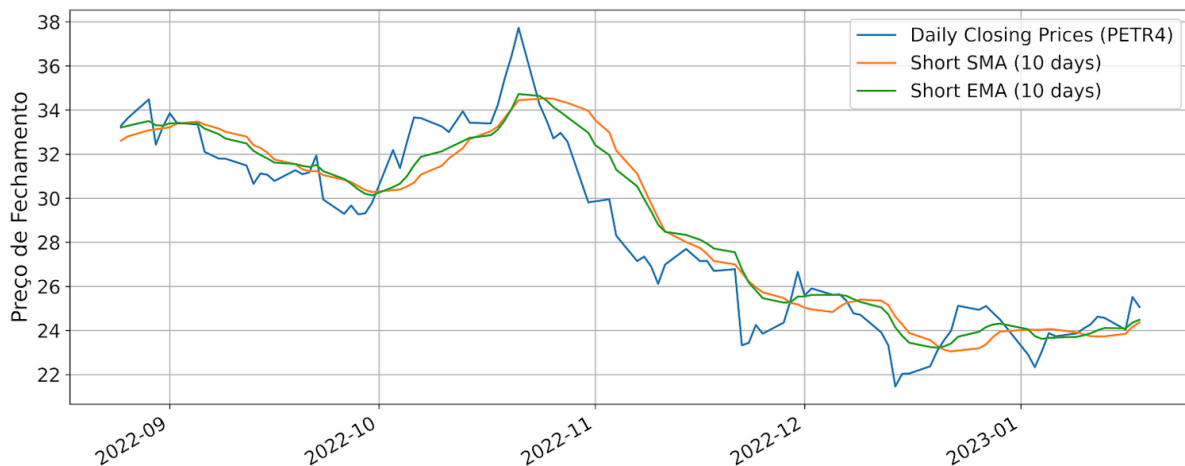


Figura 2.6: Comparação entre a SMA e EMA de 10 dias para o papel PETR4.

O indicador Média Móvel Convergente e Divergente (do inglês, *Moving Average Convergence Divergence*) (MACD) utiliza a combinação de duas EMAs para a identificação de tendências no movimento de preços dos ativos.

A Equação 2.2 apresenta o cálculo do indicador MACD:

$$\text{MACD}_t = \text{EMA Curta (12 dias)} - \text{EMA Longa (26 dias)} \quad (2.2)$$

As Bandas de Bollinger (do inglês, *Bollinger Bands*) (BBANDS) formam um indicador técnico bastante utilizado na análise gráfica que visa indicar momentos de reversão à média histórica. São construídas a partir de uma SMA, em que a banda inferior é obtida ao subtrair a quantidade de desvios padrões na janela de dias definida. Alternativamente, a banda superior é construída ao adicionar o desvio padrão. A Figura 2.7 apresenta as BBANDS em conjunto com uma SMA para as ações da VALE3.

A Equação 2.3 apresenta o cálculo dos componentes indicador BBANDS:

$$\text{BBANDS}_t = \text{SMA (n dias)} \pm \text{Bandas (x desvios)} \quad (2.3)$$





Figura 2.7: Média Móvel e Bandas de Bollinger para o papel VALE3.

Em que:

- SMA é a média móvel no período escolhido;
- Banda Superior é a SMA + Desvio padrão no período escolhido;
- Banda Inferior é a SMA - Desvio padrão no período escolhido;

O Índice de Força Relativa (do inglês, *Relative Strength Index*) (RSI) é um indicador de momento pertencente à classe dos osciladores. Os osciladores de momento medem a velocidade de movimento direcional nos preços [31]. As representações visuais produzidas podem indicar momentos de reversão nos preços, linhas de suporte e resistência e divergência de tendências.

A Equação 2.4 apresenta o cálculo do indicador RSI:

$$RSI = 100 - \frac{100}{1 + \left(\frac{G}{P}\right)} \quad (2.4)$$

Em que:

- G é o ganho médio no período escolhido;
- P é a perda média no período escolhido;

A Figura 2.8 apresenta uma análise do RSI. A imagem apresenta a interpretação do indicador para estratégias determinísticas de compra e venda do ativo PETR4. Os momentos em que o gráfico do indicador RSI cruza a linha vermelha indicam uma possível supervalorização do ativo, momento em que a venda é uma recomendação que pode gerar

lucros. Inversamente, os momentos em que o indicador RSI cruza a linha azul podem sugerir uma ordem de compra do ativo. No geral, a estratégia busca estabelecer marcas para “comprar barato e vender caro”.

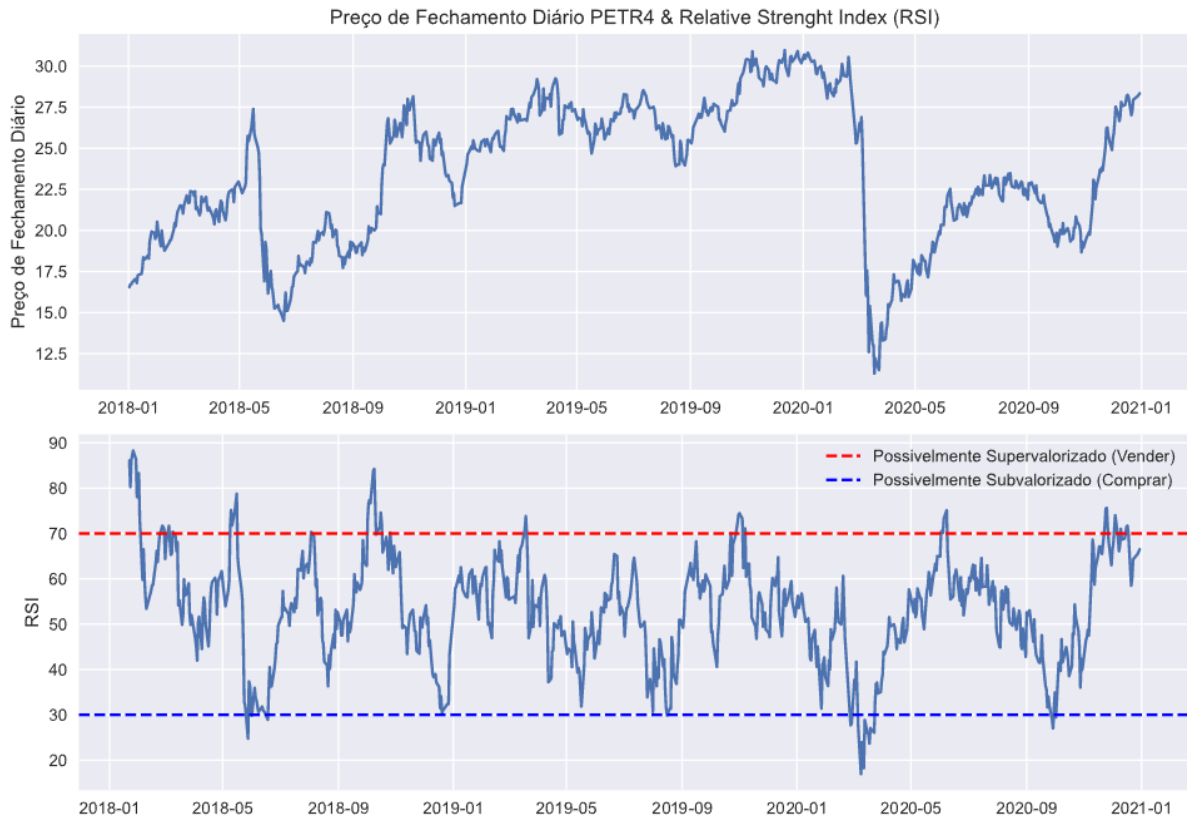


Figura 2.8: Análise do indicador Índice de Força Relativa (RSI) nas cotações de fechamento diárias da PETR4.

O *On-Balance Volume* (OBV) é um indicador de *momentum* proposto por Joseph Granville em [32]. O OBV utiliza o fluxo de volume para a predição de mudança de comportamento na série de preços de um ativo. De forma geral, o indicador busca identificar antecipadamente possíveis momentos de alta ou queda nos preços. O OBV pode ser utilizado para a tomada de decisão de compra e venda, em que os sinais são emitidos a partir do cruzamento entre o indicador atual e sua média móvel. No entanto, utilizar o OBV isoladamente pode apresentar sinais ruidosos em virtude de grande variação momentânea do volume, como uma divulgação de resultados ou adição da empresa em um índice relevante, que não necessariamente irá refletir em uma tendência.

A Equação 2.5 apresenta o cálculo do indicador OBV:

$$OBV_t = OBV_{t-1} + \begin{cases} \text{Volume}, & \text{se } Close_t > Close_{t-1} \\ 0, & \text{se } Close_t = Close_{t-1} \\ -\text{Volume}, & \text{se } Close_t < Close_{t-1} \end{cases} \quad (2.5)$$

Em que:

- $OBV_t$  é o valor do indicador na observação atual;
- $OBV_{t-1}$  é o valor do indicador na observação anterior;
- Volume é o volume negociado na observação atual;

### 2.1.4 Indicadores Macroeconômicos

Os dados (ou indicadores) macroeconômicos são métricas utilizadas para a avaliação periódica no contexto amplo da economia.

### 2.1.5 Negociação por Algoritmos

A negociação de forma eletrônica tornou-se comum com o avanço da computação, dos serviços de infraestrutura de rede e armazenamento de dados, que possibilitaram o acompanhamento imediato das cotações e o gerenciamento de grandes volumes de transações em frações de segundo. Nos dias de hoje, é comum que investidores iniciantes utilizem aplicativos *mobile* ou navegadores de *internet* no computador pessoal para acessar a conta, acompanhar o preço dos ativos e realizar operações de compra e venda através do *home broker* da corretora. Apesar da grande facilidade para executar as ordens, a tomada de decisão individual é limitada, principalmente quando o objetivo é a obtenção de ganhos em curto prazo [33]. O mercado de ações é influenciado por diversos fatores e torna difícil ao indivíduo o mapeamento de diferentes fontes de informação para a tomada de decisão. Para os investidores com foco no acúmulo de patrimônio a longo prazo, a análise fundamentalista das empresas em conjunto com aportes constantes e crescentes torna-se uma estratégia interessante e desvinculada do estresse em acompanhar ativamente as variações da economia.

De forma simplificada, a Negociação por Algoritmos (do inglês, *Algorithmic Trading*) (AT) é a metodologia de negociação na bolsa que utiliza computadores e um conjunto de regras matemáticas [34]. Apresenta benefícios como a velocidade, eficiência, escalabilidade e isenção da influência de sentimentos na tomada de decisão. Em geral, a AT é explorada em três etapas: (i) Desenvolvimento das estratégias: etapa em que são definidas as variáveis que irão indicar sinais de compra e venda dos ativos; (ii) Otimização de parâmetros: etapa de ajuste dos valores atribuídos aos parâmetros da estratégia inicial; e

(iii) Realização de testes em dados históricos (denominados *backtests*): etapa de avaliação dos resultados da estratégia caso fossem aplicados em dados reais do passado.

Dentre as estratégias de negociação por algoritmo, existem duas que merecem destaque: (i) Reversão à média (do inglês, *mean-reversion*): estratégia em que o investidor explora momentos de aparente ineficiência do mercado para a obtenção de lucro, apostando que os preços irão retornar a um patamar histórico médio; e (ii) Seguimento de tendência (do inglês, *trend-following*): estratégia em que são identificados fatores que apontam o direcionamento do mercado em relação ao ativo alvo. A estratégia de negociação simultânea de pares de ativos denominada “*Pairs Trading*” é um exemplo de técnica de reversão à média baseada em arbitragem estatística [35]. O algoritmo de negociação explora os desvios na diferença entre dos preços para simultaneamente exercer ordens de venda nos ativos supervalorizados e ordens de compra nos ativos subvalorizados. Essas estratégias são determinísticas, tendo em vista que os sinais de emissão de ordens são obtidos a partir de regras manualmente construídas.

A análise técnica é uma metodologia de predição de preços a partir do estudos de dados históricos de preço e volume. O objetivo é explorar padrões e tendências para realizar lucros. É uma análise que não considera a hipótese do mercado eficiente (EMH), cuja tese indica que os preços presentes refletem toda a informação disponível. Sendo assim, a AT não poderia ser utilizada para a realização de lucros.

## 2.2 Processamento de Linguagem Natural

Esta seção expõe ao leitor os tópicos referentes ao Processamento de Linguagem Natural utilizados durante o desenvolvimento do projeto. Primeiramente, será apresentado o Reconhecimento de Entidades Nomeadas, que é uma das principais tarefas de extração de informação em conteúdos textuais. Em seguida será exposto o tópico referente a Análise de Sentimentos.

### 2.2.1 Reconhecimento de Entidades Nomeadas

O Reconhecimento de Entidades Nomeadas (do inglês, *Named Entity Recognition*) (NER) é o domínio de pesquisa na área de NLP que tem como objetivo identificar menções presentes nos textos [36]. As entidades a serem identificadas são palavras ou um conjunto de palavras pertencentes a um grupo limitado de classes definidas (e.g. nome de pessoas, lugares, instituições, empresas, datas e valores monetários) [37]. A partir de uma sequência de palavras  $s = \langle w_1, w_2, \dots, w_n \rangle$  o objetivo de um sistema de NER é identificar um conjunto de tuplas  $\langle I_s, I_e, t \rangle$ , em que  $I_s$  corresponde à posição inicial da entidade no texto,  $I_e$  corresponde à posição final, e  $t$  é uma das classes especificadas no problema [38].

Além da identificação de entidades genéricas, a aplicabilidade de NER pode ser ajustada de acordo com o domínio específico de interesse do projeto. Settles [39] utilizou técnicas de NER para a identificação de medicamentos, tratamentos, nomes de doenças e de substâncias em documentos textuais relacionados ao domínio médico. Li *et al.* [40] publicaram uma extensa revisão da literatura sobre a aplicação de técnicas de aprendizado profundo para NER, além de expor os principais desafios encontrados e propostas de trabalhos futuros na área.

Em domínios estáticos, cuja variação linguística é limitada, o NER pode ser obtido com o desenvolvimento de Métodos Baseados em Regras (do inglês, *Rule-Based Methods*) (RBM). Esses modelos são desenvolvidos a partir de descrições simples e construídos com regras manualmente fornecidas. Aggarwal [36] apresentou as metodologias recomendadas para a construção manual de algoritmos de NER. Tais métodos serão utilizados no atual projeto para a identificação de empresas mencionadas no conteúdo textual das notícias. A [B]<sup>3</sup> padroniza os códigos de negociação na bolsa com uma sequência de quatro letras e pelo menos um dígito (e.g. ABEV3, BBAS3, CYRE3). A criação de regras com expressões regulares, por exemplo, já soluciona o problema de identificar *tickers* válidos (e existentes) na bolsa de valores brasileira, desde que o mapeamento entre o código de negociação e o nome da empresa seja armazenado.

### 2.2.2 Análise de Sentimentos

A Análise de Sentimentos (do inglês, *Sentiment Analysis*) (SA) consiste em reunir e analisar as opiniões, pensamentos e impressões das pessoas sobre temas diversos, qualidade de produtos, serviços e assuntos gerais [41]. As classificações obtidas buscam atribuir um *score* de polaridade no intervalo  $[-1, +1]$  para as amostras de texto selecionadas. Dentre as abordagens existentes, duas ganham destaque: (i) baseada em léxicos: utiliza dicionários de termos positivos, negativos ou neutros e análise semântica ou estatística para determinar o sentimento final; e (ii) baseada em ML: utiliza algoritmos de ML em bases rotuladas para a criação de um classificador de sentimentos. As duas abordagens podem ser combinadas para produzir um classificador de sentimentos mais robusto.

O processo de classificação adotado em SA pode gerar informações relevantes para os investidores tomarem decisão de compra ou venda de ativos. O livro “*Trading on Sentiment*” entra em detalhes sobre como o sentimento percebido pelos investidores pode influenciar na negociação de ativos [42].

A classificação dos sentimentos apresenta abordagens de ML como destaque, sendo apresentadas as vantagens e desvantagens ao utilizar algoritmos como SVM, NB e FFNN para o processo de rotulagem dos conteúdos textuais [43].

## 2.3 Séries Temporais

Esta seção apresenta ao leitor os fundamentos relacionados às séries temporais, bem como a aplicabilidade dos conceitos abordados em diferentes áreas do conhecimento, especialmente no contexto dos mercados financeiros.

As séries temporais são definidas como uma sequência numérica de observações armazenadas de forma sucessiva [44]. As medições de temperatura diárias de uma determinada cidade, o registro mensal da quantidade de visitantes em um *site* de compras, o monitoramento contínuo de dispositivos computacionais e o histórico de preços de uma empresa listada na bolsa de valores são exemplos de bases de dados compostas por séries temporais.

A análise de séries temporais pode ser utilizada para identificar características como tendência e sazonalidade. A tendência representa uma direção de mudança na série de observações. A sazonalidade indica períodos em que um comportamento se repete nas séries temporais. É observada por hotéis que registram alta demanda de reservas nos períodos de férias, o que implica uma frequência semestral ou anual em sua taxa de ocupação, por exemplo. A análise de sazonalidade em séries temporais relacionadas ao mercado financeiro foi explorada em [45, 46, 47].

A Figura 2.9 apresenta o histórico de cotações das ações preferenciais da Petrobras (*ticker* PETR4). O gráfico expõe uma tendência de valorização na cotação do ativo entre os anos de 2016 e 2020. Observam-se, no entanto, períodos de queda na cotação, mas a tendência visualizada no longo prazo é de crescimento. O cenário apresentado, portanto, não é de expansão linear. Os momentos contrários à tendência de maior relevância são denominados momentos de retração. Além disso, a tendência de valorização indica uma maior força do lado comprador do ativo.

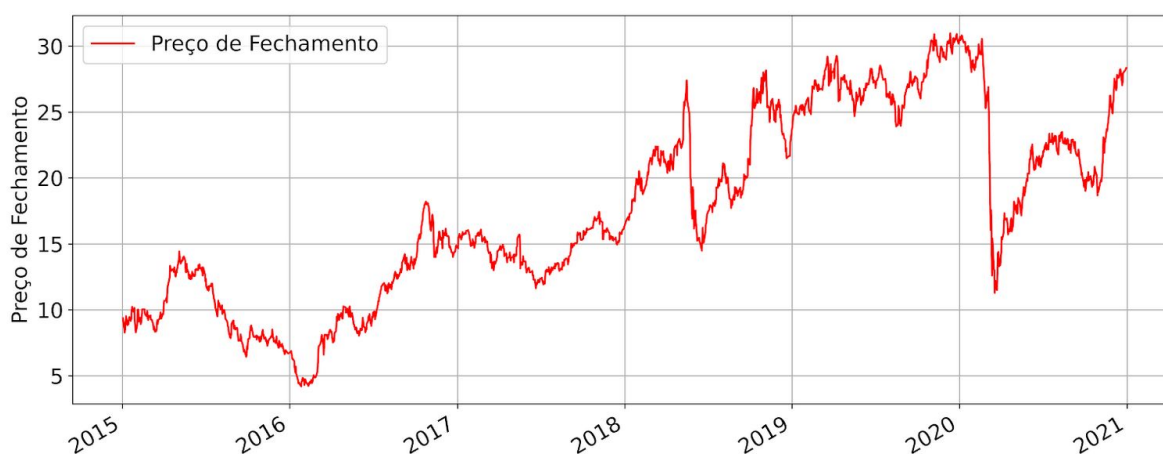


Figura 2.9: Histórico de Preços de Fechamento das Ações Preferenciais da Petrobras (PETR4) entre Jan/2015 e Dez/2020.

Os conceitos de suporte e resistência são fundamentais na dinâmica do movimento de preços dos ativos do mercado financeiro. As linhas de suportes, marcadas por determinado preço, são indicadores onde ocorreu uma maior demanda pelo ativo. Tal cenário é observado onde existe uma tendência de queda que é revertida após existir grandes compradores para o ativo. Do contrário, as linhas de resistência são encontradas nas tendências de alta, nos preços onde houve maior oferta pelo ativo, caracterizando um topo para a valorização das cotações. A identificação das linhas de suporte e resistência é feita através de análise gráfica ou ainda da utilização de médias móveis.

A predição em séries temporais é de grande interesse e vem sendo explorada com DL em publicações recentes [48]. Conforme exposto na Equação 2.6, o objetivo é determinar o valor da próxima observação  $\hat{y}_{t+1}$  a partir de uma função  $f(\cdot)$  que recebe intervalos históricos da série, além de poder incluir outra série temporal  $x_t$  e dados independentes do tempo  $s$ .

$$\hat{y}_{t+1} = f(y_{t-k:t}, x_{t-k:t}, s) \quad (2.6)$$

As séries temporais são construídas a partir de quatro componentes principais: (i) movimentos de longo prazo (tendência); (ii) movimentos sazonais de curto prazo; (iii) movimentos cíclicos de curto prazo; e (iv) aleatoriedades ou flutuações irregulares [49]. Os três primeiros itens são denominados sinais, ao passo que o último é denominado ruído. Analisar a relação sinal/ruído é de suma importância na construção de modelos para a predição de séries temporais [50].

A seguir, serão apresentados dois conceitos de interesse na análise comparativa entre diferentes séries temporais. A covariância e a correlação contribuem para o melhor entendimento do relacionamento conjunto entre duas séries de observações.

### 2.3.1 Covariância e Correlação

Analisar o comportamento de diferentes séries temporais é importante para identificar características em comum e calcular exposições ao risco. Solnik *et al.* [51] exploraram a correlação e volatilidade entre diferentes mercados internacionais.

A covariância é uma medida de variação entre duas variáveis aleatórias. A Equação 2.7 apresenta o cálculo da covariância para uma população, que é a soma do produto dos desvios entre a média de cada variável:

$$\sigma_{XY} = \frac{1}{n} \sum_{i=1}^n (X_i - \mu_X) \cdot (Y_i - \mu_Y) \quad (2.7)$$

Diferentemente do resultado da covariância, o coeficiente de correlação resulta em um valor no intervalo  $[-1; 1]$ . Os resultados próximos a 0 indicam baixa correlação entre

as variáveis. Os resultados em módulo próximos a 1 indicam alta correlação positiva entre as variáveis. Do contrário, os resultados em módulo próximos a -1 indicam alta correlação negativa entre as variáveis. A Equação 2.8 representa a correlação, que é a medida estatística que determina o grau em que duas séries temporais variam na mesma direção [52].

$$\rho_{XY} = \frac{1}{n} \sum_{t=1}^n \left( \frac{x_t - \bar{x}}{\sigma_x} \right) \left( \frac{Y_t - \bar{y}}{\sigma_y} \right) \quad (2.8)$$

No entanto, a correlação indica apenas o direcionamento comum entre duas séries temporais. Duas retas crescentes com coeficientes angulares distintos apresentam alta correlação positiva, mas nada é informado sobre a diferença entre as variáveis dependentes de cada equação. Nesse sentido, a cointegração, por exemplo, supera a limitação descrita.

A diversificação de ativos em portfólios como forma de minimização de riscos foi explorada por Markovitz em [53]. O portfólio ideal é aquele que maximiza os retornos ao passo que minimiza o risco associado. O balanceamento do portfólio ideal é realizado a partir da quantidade de cada ativo na carteira do investidor, levando em consideração as correlações presentes entre cada classe de ativos. A seleção de ativos menos correlacionados permite a redução do risco total associado ao portfólio final.

### 2.3.2 Análise de Risco

O risco é comumente associado na natureza a um contexto de medo e potencial perigo. A predisposição ao risco, no entanto, não é universal. Diferentes pessoas administram o sentimento de forma única, sendo umas mais propensas ao risco do que outras. Kahneman e Tversky iniciaram a teoria da perspectiva (do inglês *prospect theory*) ao realizarem uma pesquisa sobre a perspectiva do risco na tomada de decisão. Os resultados evidenciaram que os entrevistados atribuíram maior destaque às perdas relativas quando comparadas a ganhos de magnitude equivalente [54]. Mais recentemente, nos campos de estudo da psicologia e neurobiologia, Sokol-Hessner e Rutledge propuseram um novo *framework* para o entendimento da aversão ao risco [55].

Comparar apenas os retornos ao final de um período entre dois ativos resulta em uma análise incompleta de cada investimento, tendo em vista que duas curvas podem oscilar de maneiras diferentes e apresentarem o mesmo ponto de início e fim. Nessa situação, o investidor conservador terá mais afeição com aquele investimento similar a uma reta com coeficiente de angulação positivo e terá receio da aplicação que apresentar grandes oscilações, tanto positivas quanto negativas.

Ao construir uma carteira de ações, o investidor deve considerar dois tipos de risco que podem afetar o mercado: (i) riscos sistemáticos, que envolvem a economia por inteiro; e



(ii) riscos não-sistemáticos, que afetam companhias isoladas ou setores específicos. Com isso, a distribuição do capital investido em empresas não correlacionadas e em setores diversificados tenderá a diminuir o risco total do portfólio. A Equação 2.9 descreve o Índice Sharpe (do inglês, *Sharpe Ratio*) (ShR), que é uma das principais medidas de performance de investimentos ajustados ao risco, indicando o retorno obtido em relação ao risco tomado pelo investidor [56].  $R_p$  e  $\sigma_p$  representam, respectivamente, o retorno final e o desvio padrão do portfólio (ou de um ativo individual) e  $R_f$  representa o retorno “livre” de risco, em que geralmente é utilizada uma taxa referência para títulos de renda fixa (e.g. Selic, CDI). Valores negativos no ShR indicam que o investidor obteria melhores retornos alocando seu capital apenas no instrumento representado por  $R_f$ .

$$\text{Sharpe Ratio (ShR)} = \frac{R_p - R_f}{\sigma_p} \quad (2.9)$$

O índice Sharpe considera a volatilidade total no retorno dos investimentos, tanto positiva quanto negativa. Nesse sentido, o Índice Sortino (do inglês, *Sortino Ratio*) (SrR) foi desenvolvido com o objetivo de separar cada tipo de volatilidade, considerando apenas os desvios negativos do investimento. A Equação 2.10 descreve o SrR, em que  $R_p$  e  $\sigma_{neg}$  representam, respectivamente, o retorno final e o desvio padrão dos retornos negativos do portfólio e  $R_f$  representa o retorno “livre” de risco.

$$\text{Sortino Ratio (SrR)} = \frac{R_p - R_f}{\sigma_{neg}} \quad (2.10)$$

## 2.4 Aprendizado de Máquina

Esta seção apresenta ao leitor uma visão geral das principais arquiteturas de redes neurais artificiais e exemplos de aplicações do aprendizado profundo em diferentes contextos.

Definido por Russell e Norvig como “o estudo de agentes que recebem percepções do ambiente e executam ações” [57], o campo de estudos denominado Inteligência Artificial (do inglês, *Artificial Intelligence*) (AI) está em constante destaque no meio acadêmico e na automação industrial. As atividades realizadas por um sistema de AI pretendem reproduzir um comportamento similar ao humano na identificação, entendimento e execução de uma tarefa.

Com o avanço de tecnologias de melhor desempenho e mais acessíveis, tanto em *hardware* quanto em *software*, tornou-se possível implementar modelos de AI propostos no passado, cujo desenvolvimento imediato foi impossibilitado em larga escala em virtude da limitação computacional da época. O Aprendizado Profundo (do inglês, *Deep Learning*) (DL) aparece no contexto dos algoritmos de Aprendizado de Máquina (do inglês, *Machine Learning*) (ML), utilizados para a extração de informação em dados brutos e tomada de

decisão automática. Os algoritmos de DL obtiveram sucesso reconhecido em diversas aplicações recentes [58].

O desenvolvimento de carros autônomos é um tema em destaque em DL últimos anos. A fabricante americana de veículos elétricos Tesla, liderada pelo empresário Elon Musk, investe de forma expressiva em AI<sup>6</sup>. Os automóveis possuem uma grande quantidade de sensores, georreferenciamento e câmeras, responsáveis por coletar os dados do ambiente para a tomada de decisão de forma independente e em tempo real. O DL também obteve sucesso na assistência em diagnósticos médicos ao utilizar soluções de AI como o IBM Watson<sup>7</sup> para otimizar o tempo dos médicos e pacientes, além de detectar com maior precisão indícios de doenças em imagens de raios-X, ressonâncias magnéticas ou tomografias [59]. No contexto das redes sociais, o desenvolvimento de sistemas de recomendação é explorado com uso de DL [60]. Estes estão presentes em plataformas de comércio digital (Amazon), serviços de *streaming* de vídeo (Netflix), *streaming* de áudio (Spotify) e redes sociais com propostas variadas (Facebook, Instagram e TikTok).

### 2.4.1 Redes Neurais Artificiais

O desenvolvimento das Redes Neurais Artificiais (do inglês, *Artificial Neural Networks*) (ANN) mais complexas atualmente surgiu a partir da proposta de McCulloch e Pitts (1943) de um modelo matemático simples, sendo este uma representação numérica do neurônio biológico [61]. Posteriormente, as limitações do modelo inicial foram superadas com a criação do Perceptron por Frank Rosenblatt [62]. Em sua arquitetura básica, conforme exposto na Figura 2.10, o Perceptron é composto por um conjunto de entradas numéricas  $X = (x_1, x_2, \dots, x_m)$ , uma entrada de valor fixo 1 denominada *bias* (viés), um conjunto de pesos  $W = (w_1, w_2, \dots, w_m)$ , uma unidade somatória  $v$  e uma função de ativação  $\varphi$ , a qual irá fornecer a saída  $y$ . O processo de aprendizagem inerente ao algoritmo está na atualização do conjunto de pesos  $W$  com objetivo de mapear corretamente o conjunto de entradas  $X$  fornecidas no treinamento às suas respectivas saídas  $y$ .

Apesar de ser capaz de solucionar problemas linearmente separáveis com o Perceptron de maneira automática, isto é, sem o ajuste manual de pesos na rede, tal modelo ainda era limitado para a solução de problemas reais mais complexos, que demandam a separação de padrões com características não lineares. Quase duas décadas após a publicação do Perceptron Simples, o Perceptron Multicamada (do inglês, *Multilayer Perceptron*) (MLP) foi viabilizado com o ajuste do algoritmo de retropropagação, o que possibilitou a correção de pesos na rede em diferentes camadas ocultas e proporcionou a retomada do interesse na aplicação de redes neurais em problemas existentes [63].

---

<sup>6</sup><https://www.tesla.com/autopilotAI>

<sup>7</sup><https://www.ibm.com/br-pt/watson-health>

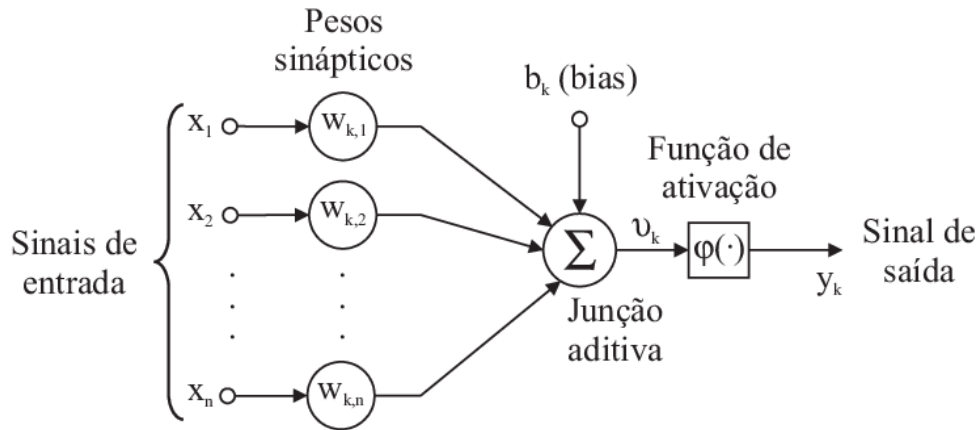


Figura 2.10: Perceptron Simples (Frank Rosenblatt [2]).

As ANNs são utilizadas como aproximadoras de funções ( $\hat{y} = f(x|w)$ ). As redes do tipo *Feed Forward Neural Networks* (FFNN) são coleções de neurônios conectados e organizados em diferentes camadas, sem ciclos existentes entre as conexões. Nesse tipo de arquitetura, os sinais são propagados para frente. A Figura 2.11 apresenta uma estrutura básica de uma ANN do tipo FFNN.

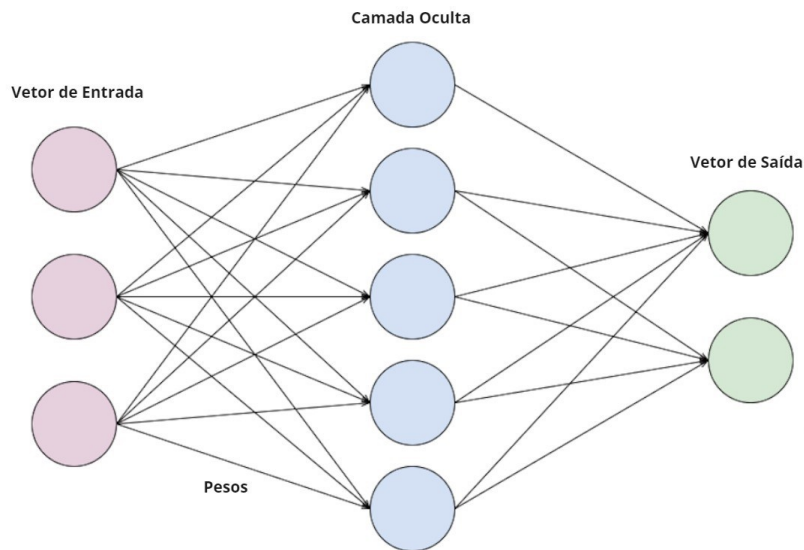


Figura 2.11: Rede Neural do tipo *Feed Forward Neural Network* (FFNN).

A primeira camada da rede é denominada camada de entrada, na qual os dados representados pelo vetor de entrada  $x = [x_1, x_2, x_3]$  são recebidos e propagados para as camada ocultas. Para cada neurônio na camada oculta, existe uma conexão associada a cada uma das entradas. As conexões são ponderadas pela matriz de pesos na rede. Cada conexão possui um valor  $w_{ij}$  associado (e.g. o valor  $w_{11}$  representa o peso associado ao primeiro

neurônio de entrada com o primeiro neurônio na camada oculta). A camada final da rede, representada pelo vetor de saída  $y$ , produz o valor aos dados de entrada fornecidos em  $x$ . A saída rede neural é resumida pela função  $y = \phi_2(\phi_1(x \cdot w_1) \cdot w_2)$ , em que  $\phi_1$  representa a função de ativação dos neurônios da camada oculta,  $w_1$  representa a matriz de pesos entre a entrada e a camada oculta,  $\phi_2$  representa a função de ativação dos neurônios de saída e  $w_2$  representa a matriz de pesos entre a camada oculta e a camada de saída. Os ajustes nas matrizes de pesos são explorados para aproximar a função de saída desejada.

## 2.4.2 Redes Neurais Recorrentes

As Redes Neurais Recorrentes (do inglês, *Recurrent Neural Networks*) (RNN) são formadas por arquiteturas mais complexas em relação às redes FFNN e desempenham a importante função de persistir informações pretéritas em conjunto com as entradas atuais [64]. Tal capacidade é de grande utilidade, visto que problemas de caráter sequencial podem ser explorados com maior robustez. Previsão de séries temporais, reconhecimento por voz e geração automática de textos são alguns exemplos de aplicação das RNN. A Figura 2.12 apresenta um exemplo da arquitetura em alto nível.

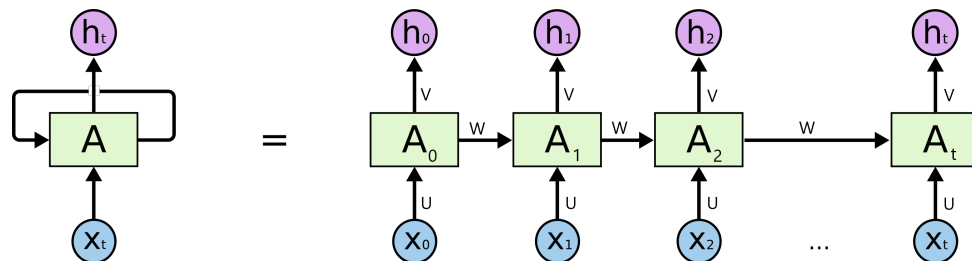


Figura 2.12: Exemplificação de uma Rede Neural Recorrente (Figura Adaptada)<sup>8</sup>.

Apesar de serem capazes de armazenar informações em marcas temporais passadas, estas são somente de curto prazo. A atualização do conjunto de pesos repassada na rede depende de um fator de conexão entre as camadas, que ao ser multiplicado em sequência, atualiza de forma inadequada as camadas anteriores. Com isso, descobriu-se que dependências mais antigas tornam-se pouco influentes, resultando no problema denominado dissipação do gradiente (do inglês, *vanishing gradient*).

Dentre as propostas de solução, uma de grande destaque e relevância foi a criação das redes *Long Short-Term Memory* (LSTM), que são amplamente utilizadas em problemas de caráter sequencial, principalmente no domínio textual, por serem capazes de corrigir o problema em armazenar informações de curto e longo prazo [65].

<sup>8</sup><https://colah.github.io/posts/2015-08-Understanding-LSTMs/>, criado em: 27/08/2015

As células de uma rede LSTM contém canais com funcionalidades específicas para filtrar as informações que serão incluídas ou descartadas em cada novo estado de uma célula  $C_t$ . A Figura 2.13 ilustra o fluxo de dados dentro de uma célula LSTM:

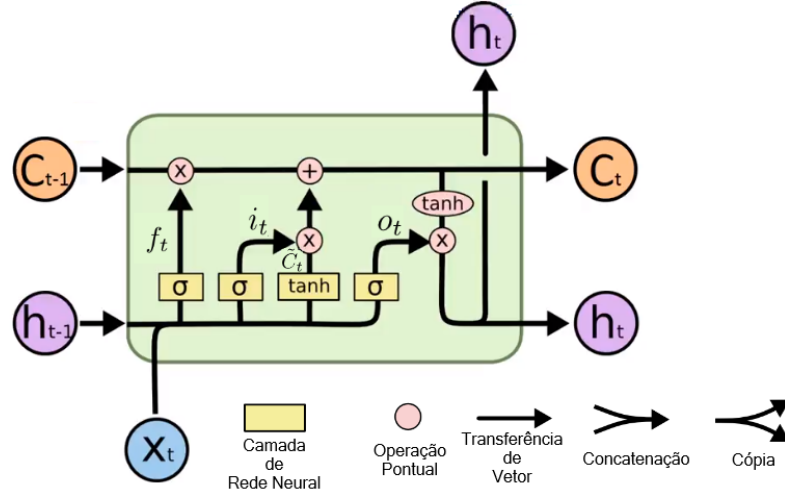


Figura 2.13: Interior de uma Célula LSTM (Figura Adaptada)<sup>9</sup>.

A Equação 2.11 apresenta a função de esquecimento da célula, na qual as saídas próximas a zero indicam que a informação deve ser descartada ao passo que as saídas próximas a 1 indicam que deve ser persistida.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2.11)$$

Em seguida, os próximos portões  $i_t$  e  $\tilde{C}_t$ , indicam quais valores serão atualizados e quais novas informações devem ser armazenadas no estado atual da célula, respectivamente. Em ambas as equações são utilizadas as entradas atuais em conjunto com a saída da célula anterior. A Equação 2.12 e a Equação 2.13 apresentam o cálculo realizados nos portões mencionados

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2.12)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (2.13)$$

O estado atual da célula ( $C_t$ ) é descrito pela Equação 2.14, contendo relação com as demais funções expostas.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (2.14)$$

<sup>9</sup><https://colah.github.io/posts/2015-08-Understanding-LSTMs/>, criado em: 27/08/2015

Por fim, a saída da célula, codificada no vetor de saída  $h_t$ , será uma versão filtrada do estado atual, definida pela Equação 2.15 e pela Equação 2.16:

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \quad (2.15)$$

$$h_t = o_t * \tanh(C_t) \quad (2.16)$$

## 2.5 Aprendizado por Reforço

Esta seção fornece ao leitor os fundamentos relacionados ao aprendizado por reforço, seus componentes principais, objetivos específicos e casos recomendados de uso.

A tomada de decisão em ambientes de incerteza é um dos grandes desafios enfrentados em ML [66]. Em cenários dinâmicos, com atualizações frequentes e em tempo real, nota-se a necessidade de criar mecanismos de decisão adaptativos, capazes de aprender a executar sequências de boas ações e otimizar o desempenho com a experiência adquirida. O Aprendizado por Reforço (do inglês, *Reinforcement Learning*) (RL) é uma das metodologias de aprendizado computacional no contexto da AI. Os algoritmos de RL são adequados para problemas de natureza sequencial que envolvem decisões complexas e possuem um objetivo de longo prazo [67]. Ao serem comparadas com os métodos supervisionados de ML, as tarefas que envolvem RL não recebem dados rotulados e exploram além de decisões pontuais de classificação. Através de tentativa e erro e maximização de um sinal numérico de recompensa, o algoritmo de RL aprende ao avaliar o *feedback* gerado por suas ações e experiências. O uso de algoritmos de RL ganhou destaque em uma série de aplicações: (i) desenvolvimento de jogadores artificiais de *Atari* e Xadrez [68, 69]; (ii) simuladores de locomoção em robôs [70]; e (iii) aperfeiçoamento de carros autônomos [71].

Os problemas de RL podem ser descritos por um Processo de Decisão de Markov (do inglês, *Markov Decision Process*) (MDP). Um MDP é um processo de controle estocástico de tempo discreto. Sendo assim, o processo é baseado em decisões tomadas até a chegada de um objetivo, em que os estados futuros dependem parcialmente das ações tomadas e as etapas de tempo avançam em intervalos finitos (i.e.  $t \in 1, 2, 3, \dots$ ).

Considerando os conjuntos finitos de estados ( $S$ ), ações ( $A$ ) e recompensas ( $R$ ) e cada etapa sendo executada de forma sequencial, o agente percebe o estado atual  $s_t \in S$  e executa uma ação  $a_t \in A$ , resultando no par estado-ação  $(s_t, a_t)$ . Um novo estado  $s_{t+1} \in S$  é atingido e o agente recebe uma recompensa  $r_t \in R$ , obtida a partir da ação  $a_t$  executada em  $s_t$ . Em geral, o MDP assume que o estado atual ( $s_t$ ) somente depende do estado imediatamente anterior ( $s_{t-1}$ ), conforme exposto na Equação 2.17, sendo que a adição do histórico de observações é desnecessária. No entanto, o estado pode ser modelado

como um conjunto de observações anteriores, aumentando a quantidade de informação disponível ao agente.

$$P[S_{t+1}|S_t = s_t] = P[S_{t+1}|S_t = s_t, S_{t-1}, \dots, S_0 = s_0] \quad (2.17)$$

O RL dispõe de elementos básicos, independente do problema em que seja aplicado. O **agente** é a entidade que deve observar e interagir com o cenário em que atua, denominado **ambiente**, tomando as **ações** corretas para maximizar o sinal de **recompensa**. O objetivo do agente é maximizar a somatória das recompensas, denominada **retorno**. A execução de ações pelo agente resulta em uma modificação no ambiente, que retornará um valor numérico de recompensa imediata e avançará para um novo estado. O **estado** é uma descrição completa do ambiente em determinado momento de tempo  $t$ , enquanto que as *observações* são descrições parciais do estado. A Figura 2.14 exemplifica o relacionamento entre os principais componentes do RL.

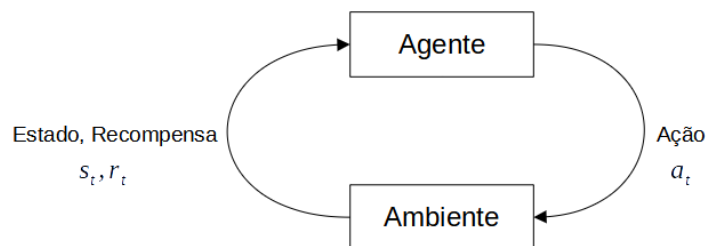


Figura 2.14: Ciclo de Interação entre Agente e Ambiente.

A **política** ( $\pi$ ) é definida como o conjunto de regras que mapeia quais ações serão tomadas por um agente a partir do estado atual em que se encontra. A função que representa a política pode ser ou determinística ( $a_t = \mu(s_t)$ ) ou estocástica ( $a_t \sim \pi(\cdot|s_t)$ ), sendo que o objetivo do RL identificar qual é a política que maximiza o retorno (denominada política ótima  $\pi_*$ ). As ANNs são comumente aplicadas para a criação de políticas, com parâmetros sendo representados por pesos e vieses da rede.

Uma **trajetória** (ou episódio) ( $\tau$ ) é definida como a sequência de estados e ações executadas no ambiente até um estado terminal. A transição de estados depende exclusivamente da ação mais recente, sendo que estas são executadas pelo agente de acordo com a política adotada.

A **recompensa** ( $r_t$ ) é a resposta numérica do ambiente ao agente a cada ação tomada. A Equação 2.18 retrata a recompensa como uma função do estado atual, da ação tomada pelo agente e do estado destino. O **retorno** ( $G_t$ ) constitui a somatória das recompensas

obtidas pelo agente em sua trajetória. O retorno é representado pela Equação 2.19, em que  $\gamma \in [0, 1]$  corresponde a um fator de desconto aplicado nas recompensas futuras. O fator  $\gamma$  é utilizado para medir a quantidade de etapas futuras que o agente irá observar ao planejar suas ações. Ao considerar  $\gamma = 0$ , todas as recompensas futuras serão iguais a zero, levando o agente a tomar decisões que o proporcionem recompensas imediatas. No caso em que  $\gamma = 1$  as recompensas futuras não serão reduzidas, o que permite maior planejamento na tomada de decisões. O objetivo final é maximizar a somatória das recompensas descontadas.

$$r_t = R(s_t, a_t, s_{t+1}) \quad (2.18)$$

$$G_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \dots = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \quad (2.19)$$

A **função valor** ( $v_\pi(s)$ ) representa o valor do retorno esperado, ao considerar uma política  $\pi$ , iniciando as etapas a partir do estado  $s$ . A função valor representa o valor do estado  $s_t$  e é descrita pela Equação 2.20. Observa-se que  $G_t$  pode ser expandido a partir da Equação 2.19.

$$v_\pi(s) = E_\pi[G_t | S_t = s] \quad (2.20)$$

Em seguida, a **função ação-valor** ( $q_\pi(s, a)$ ) representa o retorno esperado ao iniciar no estado  $s$ , selecionar uma ação específica  $a$  e seguir de acordo com a política  $\pi$  até o final do episódio. É utilizada para avaliar o retorno de uma ação em um determinado estado. A função ação-valor é representada pela Equação 2.21.

$$q_\pi(s, a) = E_\pi[G_t | S_t = s, A_t = a] \quad (2.21)$$

Adiante, existem as funções de vantagem ( $A_\pi(s, a)$ ), que descrevem a qualidade de uma ação específica em relação à média de uma amostra das demais ações possíveis. A função de vantagem, descrita pela Equação 2.22 é formada a partir da relação entre a Equação 2.20 e Equação 2.21. Um valor positivo em  $A_\pi(s, a)$  indica que tomar uma ação específica  $a$  é melhor do que seguir a política estabelecida. Nesse sentido, o objetivo é reforçar as ações que levariam a melhores resultados e mitigar os efeitos das ações que resultariam em piores métricas.

$$A_\pi(s, a) = q_\pi(s, a) - v_\pi(s) \quad (2.22)$$

Considerando a propriedade recursiva da Equação 2.19, pode-se expandir a função valor  $v_\pi$ , resultando na Equação de Bellman para  $v_\pi$ , representada pela Equação 2.23.



Nota-se uma relação recursiva entre o valor de um determinado estado  $s$  e os valores dos outros estados  $s'$ .

$$\begin{aligned}
v_\pi(s) &= E_\pi[G_t | S_t = s] \\
&= E_\pi[R_{t+1} + \gamma G_{t+1} | S_t = s] \\
&= \sum_a \pi(a|s) \sum_{s',r} p(s', r | s, a) [r + \gamma v_\pi(s')]
\end{aligned} \tag{2.23}$$

De forma similar, a equação de Bellman também pode ser obtida para a Equação 2.21, representada pela Equação 2.24:

$$\begin{aligned}
q_\pi(s, a) &= E_\pi[G_t | S_t = s, A_t = a] \\
&= E_\pi[R_{t+1} + \gamma G_{t+1} | S_t = s, A_t = a] \\
&= \sum_{s',r} p(s', r | s, a) [r + \gamma \sum_{a'} \pi(a' | s') q_\pi(s', a')]
\end{aligned} \tag{2.24}$$

Para solucionar um MDP deve-se considerar que o valor ótimo de um estado  $s$  é o retorno esperado ao seguir uma política  $\pi_*$ . A política ótima é aquela que seleciona as ações  $a$  que maximizam  $v(s)$  ou  $q(s, a)$ . Sendo assim,  $\pi_*$  é descrita pela Equação 2.25:

$$\begin{aligned}
\pi_*(s) &= \arg \max_a \sum_{s',r} p(s', r | s, a) [r + \gamma v_*(s)] \\
&= \arg \max_a q_*(s, a)
\end{aligned} \tag{2.25}$$

No entanto, para solucionar a Equação 2.25 é necessário saber os valores ótimos de  $v(s)$  e  $q(s, a)$ . Com isso, pode-se reescrever a Equação 2.23 e a Equação 2.24 em suas versões otimizadas. As versões otimizadas de  $v(s)$  e  $q(s, a)$  são representadas pelas Equação 2.26 e Equação 2.27, respectivamente.

$$v_*(s) = \max_a \sum_{s',r} p(s', r | s, a) [r + \gamma v_*(s')] \tag{2.26}$$

$$q_*(s, a) = \sum_{s',r} p(s', r | s, a) [r + \gamma \max_{a'} q_*(s', a')] \tag{2.27}$$

Os métodos de Monte Carlo podem ser utilizados para obter os valores de  $v_*(s)$  e  $q_*(s, a)$  a partir de amostras no conjunto de observações do ambiente, por exemplo. Os métodos de diferença temporal, que combinam as características dos métodos de Monte Carlo e de programação dinâmica, também são utilizados para alcançar o mesmo objetivo.

O Aprendizado por Reforço Profundo (do inglês, *Deep Reinforcement Learning*) (DRL) expande o RL tradicional ao utilizar Aprendizado Profundo (do inglês, *Deep Learning*)

(DL) para determinar as funções que descrevem os principais componentes do agente [29]. Nesse sentido, as ANN podem ser utilizadas para aproximar  $v_\pi$  e  $Q_\pi(s, a)$ .

Estudos recentes apontam a aplicação de Aprendizado por Reforço (do inglês, *Reinforcement Learning*) (RL) como alternativa aos métodos determinísticos utilizados na negociação de ativos nas bolsas de valores internacionais [21, 25, 72]. O RL é aplicado para obter a política que maximiza a recompensa especificada em um determinado ambiente, sem a necessidade de criar estratégias manuais e determinísticas.

Proposto por pesquisadores da OpenAI, o algoritmo Otimização de Política Proximal (do inglês, *Proximal Policy Optimization*) (PPO) [73] pertence aos Métodos de Política por Gradiente (do inglês, *Policy Gradient Methods*) (PGM) em RL, que são métodos cujo objetivo é determinar uma política  $\pi$  que represente um mapeamento direto entre estados e as melhores ações a serem tomadas em cada estado [74]. Os PGMs são métodos que utilizam aproximadores de funções para estimar as probabilidades de tomar uma determinada ação dentro do conjunto de possíveis ações ( $\pi(a|s, \theta) \in [0, 1]$ ). Nesse cenário, o conjunto de entradas é representado pelo estado  $s = [s_1, s_2]$ . O vetor de saída é representado pelas probabilidades de cada ação ser tomada no estado  $s$ ,  $\pi(s|\theta) = [p(a_1), p(a_2), p(a_3)]$ .

O PPO é um algoritmo de RL livre de modelo (do inglês, *model-free*) e pode ser aplicado em ambientes com ações discretas ou contínuas. No domínio das finanças, o PPO foi utilizado para a negociação de ativos e balanceamento de portfólio, apresentando estabilidade, baixa variância e simplicidade de implementação e otimização como características positivas [75]. O aprendizado do agente de RL é feito de maneira *online* e baseado em política estocástica, em que as ações a serem tomadas a partir do estado atual possuem diferentes probabilidades. Ao invés de calcular as funções ação-valor ou estado-valor, os PGM calculam um estimador de uma função de política parametrizada utilizando algoritmos de gradiente descendente. O PPO melhora a estabilidade do treinamento das redes que definem as políticas ao restringir a atualização da política em cada etapa de treinamento [75]. O PPO é um algoritmo de RL do tipo ator-crítico, cujo objetivo é otimizar sua performance a partir de atualizações na rede do ator. O ator é responsável por mapear as observações às ações que o agente pode realizar. O crítico é o componente responsável por avaliar as ações que o agente tomou, retornando uma expectativa da recompensa para o conjunto de observações em análise.

# Capítulo 3

## Revisão Bibliográfica

Este capítulo apresenta os principais trabalhos que envolvem o uso de algoritmos de AI com foco na negociação de ativos do mercado financeiro. Primeiramente são expostas as aplicações do DRL na negociação por algoritmos. Em seguida, são apresentados os trabalhos que utilizam a extração de informação em dados textuais obtidos nos diferentes portais de divulgação de conteúdos relacionados ao domínio das finanças. Por fim, um quadro comparativo entre a proposta desenvolvida neste documento e os trabalhos relacionados é apresentado com o objetivo de destacar os pontos de interseção.

A escolha dos trabalhos relacionados teve como início a busca por publicações que apresentassem em seu conteúdo os principais termos de interesse deste trabalho. As pesquisas foram realizadas em acervos de publicações científicas em destaque (*ACM Digital Library*, *IEEE Xplore* e *ScienceDirect*) e tiveram como filtro inicial palavras-chave como “Inteligência Artificial” (do inglês, *Artificial Intelligence*), “Aprendizado por Reforço” (do inglês, *Reinforcement Learning*) e “Mercado de Ações” (do inglês, *Stock Market*). Além disso, as publicações resultantes foram filtradas por ano de publicação e relevância, tendo como preferência publicações mais atuais e mais referenciada por outros autores.

### 3.1 Aprendizado por Reforço no Mercado Financeiro

A literatura que envolve o processo de desenvolvimento de técnicas e algoritmos para a negociação na bolsa de valores enfrenta certos impasses na divulgação do estado-da-arte [76]. Dado que boas estratégias movimentam grandes quantidades monetárias, existe uma escassez no detalhamento e publicidade dos algoritmos utilizados e dos resultados adquiridos em ambientes não simulados. Além disso, diferentes modelagens do ambiente de negociação, períodos de treinamento variados, diferentes técnicas de pré-processamento de dados e a escolha de diferentes métricas de avaliação dos resultados são alguns dos fatores que geram uma dificuldade na comparação dos trabalhos existentes na linha de pesquisa

envolvendo o uso de DRL aplicado ao mercado financeiro [27]. O estudo comparativo explorado em [77] destaca a utilização de fontes alternativas de dados como um ponto a ser melhor explorado para potencializar os resultados da aplicação de DRL na negociação de ativos.

Théate e Ernst [21] apresentaram uma nova abordagem de DRL para explorar a negociação de ativos de forma individualizada, denominada TDQN (*Trading Deep Q-Network*). Os autores exploraram ações de diferentes empresas, setores e regiões, mas não consideraram uma análise de carteira na qual os ativos apresentam relações de risco entre si. Os dados de treinamento contêm apenas o histórico diário dos preços de abertura, fechamento, valor máximo, valor mínimo e volume de contratos negociados (OHLCV). Diferentemente do explorado por Yang *et al.* em [78], o estudo não acrescentou indicadores técnicos ou dados alternativos ao conjunto de informações avaliado pelo agente, embora mencionem a possibilidade de uso desses conjuntos de dados. A estratégia de negociação com o TDQN apresentou resultados promissores nas ações da Apple, em que obteve o maior ShR e segundo melhor retorno monetário acumulado durante o período de avaliação. No entanto, não obteve boas métricas para a negociação das ações da Tesla, em que apresentou um ShR baixo e obteve lucro de somente 98 dólares, resultado muito abaixo da estratégia de “comprar e manter” (do inglês, *Buy & Hold*), que obteve lucro de 29.847 dólares a partir de um investimento inicial de 100.000 dólares.

Carta *et al.* [25] exploraram uma abordagem *ensemble* com diferentes agentes baseados em *Q-Learning* para a negociação de ativos durante o horário oficial de mercado. Os agentes foram treinados com diferentes janelas temporais, sendo que a decisão final consiste na combinação das decisões de cada agente individual. Os autores utilizaram os mercados futuros dos índices S&P 500 e DAX, com as marcações de preço agregadas em períodos de uma hora. O modelo proposto atingiu resultados promissores, com retornos acumulados superiores à estratégia passiva de comprar e manter as ações no período de avaliação.

Wu *et al.* [72] utilizaram Unidades Recorrentes Bloqueadas (do inglês, *Gated Recurrent Units*) (GRU) para a extração de características dos dados de mercado e indicadores técnicos. O Índice Sortino (do inglês, *Sortino Ratio*) (SrR) [79] foi utilizado como função de recompensa. Tal medida avalia o retorno de um investimento de acordo com o risco associado, considerando apenas o desvio padrão dos retornos negativos. Foram desenvolvidos dois algoritmos de DRL: (i) GDQN, do tipo crítico com *Q-networks* profundas; e (ii) GDQN, do tipo ator-crítico com DDPG [80]. Os resultados reportados indicaram que as estratégias apresentaram retornos mais estáveis nos períodos de alta volatilidade do mercado. No entanto, o estudo apresenta limitações como a ausência de custos de transação e a negociação de apenas um ativo por vez.

Yang *et al.* [78] propuseram um método de agrupamento de três algoritmos de DRL para o gerenciamento do portfólio composto pelas 30 ações do índice *Dow Jones*. Os algoritmos escolhidos pelos autores foram o PPO, DDPG e A2C. Além de utilizarem dados de OHLCV, os autores também forneceram indicadores técnicos ao agente de negociação. A avaliação foi feita com os algoritmos de forma individual e em *ensemble*. Os resultados evidenciaram que o algoritmo PPO proporcionou o maior retorno acumulado no período de avaliação, mas a combinação dos agentes pelo método de *ensemble* dos três algoritmos resultou no maior ShR, que leva em consideração o retorno do investimento comparado ao risco total associado.

Taewook e Ha Young [81] aplicaram DRL para a negociação de pares de ativos com a técnica de arbitragem estatística denominada *pairs-trading*. Os autores utilizaram dados de fechamento diário ajustados de uma amostra das ações constituintes do índice S&P500. Os ativos foram selecionados após a realização de análises do volume negociado, da liquidez e testes de cointegração para verificar o equilíbrio de longo prazo entre os pares formados. A estratégia utiliza limiares dinâmicos para emitir os sinais de ordem. O sinal acompanhado é a diferença ajustada entre os preços dos ativos. A metodologia proposta foi comparada com métodos determinísticos que adotam limiares fixos para a abertura e fechamento de posições. Os resultados reportados superaram as estratégias tradicionais, apresentado rendimento superior em todos os 11 pares selecionados.

Chen e Huang [82] desenvolveram um sistema de negociação de ativos multimodal ao combinarem dados de mercado (e. g. preço de fechamento dos ativos) e dados textuais de notícias. Os autores utilizaram a ferramenta VADER [83] para a extração de sentimentos nas notícias coletadas. Além disso, propuseram um modelo de influência que relaciona a data de publicação da mensagem e o impacto causado ao longo do tempo. Os dados de mercado foram agregados em períodos de 1 (um) minuto. As notícias coletadas são dos portais Bloomberg e Reuters e foram organizadas de forma diária. Os resultados experimentais reportados pelo sistema multimodal superaram o desempenho do modelo utilizando somente o histórico de preços do ativos em pelo menos 13.26%.

Taggian *et al.* [1] utilizaram DRL para a criação de um modelo de negociação individual de ativos com módulos dedicados à extração de características nos dados de mercado. Os módulos desenvolvidos obtém características a partir do histórico de preços utilizando diferentes arquiteturas de redes neurais (e. g. MLPs, CNNs e GRUs). Os resultados superaram outras configurações estado-da-arte para o mesmo período de avaliação no índice americano *Dow Jones*, indicando que o aprendizado de *features* mais complexas melhora a performance do modelo de negociação.

Koratomaddi *et al.* [84] exploraram o uso de DRL para a alocação de um portfólio das ações que compõem o índice *Dow Jones*. Os autores propuseram uma abordagem que

combina dados de mercado e dados referentes ao sentimento extraído de textos do mercado financeiro. A estratégia de alocação de ativos busca maximizar o retorno enquanto minimiza o risco associado ao investimento. Os dados textuais foram extraídos a partir do Google News e também de publicações do Twitter. Os sentimentos foram extraídos com o auxílio da ferramenta VADER [83]. O melhor resultado final obtido para o valor do portfólio utilizou a combinação dos dados de mercado e dos sentimentos extraídos, superando todos os *benchmarks* selecionados pelos autores. A melhor configuração apresentou um retorno anualizado de 22.05%, com um ShR de 2.07.

Lin e Beling propuseram um *framework* de negociação utilizando o algoritmo de DRL PPO em duas metodologias: (i) Com redes LSTM e FFNN; e (ii) Com dados de compra/venda do livro de ordens limitadas sendo fornecidos como entradas para uma FFNN [85]. Os resultados reportados superaram estratégias com sinais determinísticos utilizando os indicadores técnicos TWAP e VWAP.

## 3.2 Classificação de Textos no Mercado Financeiro

Sert *et al.* [86] exploraram a combinação de diferentes tarefas de classificação em NLP para analisar a movimentação do índice *Dow Jones*. Os autores combinaram atividades como extração de tópicos, análise de sentimentos e NER em textos extraídos do jornal *New York Times* e da rede social *Twitter*. Os resultados indicam que as notícias exercem impacto em uma janela futura de dias. No entanto, o processo é dependente de ferramenta para a classificação prévia dos sentimentos e das entidades nomeadas presentes nos dados textuais. Zhao *et al.* [87] propuseram uma nova arquitetura de rede neural composta por redes CNNs a nível de caracteres e GRUs para classificar textos do mercado financeiro.

Uma outra linha de pesquisa aplicada aos textos do mercado financeiro explora a classificação de sentimentos encontrados em notícias, fóruns de discussão, *blogs* e publicações realizadas em redes sociais [88]. Nti, Adeykoya e Weyori [11] desenvolveram um modelo para a predição da movimentação de preços na bolsa de valores de Gana (GSE). O trabalho utiliza dados textuais extraídos do Twitter, de fóruns e portais de notícias, além de aplicar métricas do *Google Trends*. Os resultados indicaram que a combinação das fontes de dados aumenta a acurácia do modelo proposto. Também foi observado que uma janela com um maior número de dias é recomendada para agregar o sentimento geral, pois o resultado pode continuar influenciando o mercado em dias futuros. No entanto, um aspecto negativo observado é que a proposta é dependente de ferramenta para a classificação dos sentimentos nos textos.

Ainda com foco na análise de sentimentos em publicações do Twitter, Naeem *et al.* [89] utilizaram métodos estatísticos para a identificação de uma relação de causalidade

entre o índice de felicidade nas publicações do Twitter<sup>1</sup> e o índice VIX de diferentes países. Também com dados do Twitter, Kinyua *et al.* exploraram o impacto das publicações do ex-presidente Donald Trump [12] na movimentação dos índices SPX e DJIA.

Vargas *et al.* [90] utilizaram os sentimentos extraídos de notícias de empresas específicas da bolsa de valores brasileira como atributos de entrada para uma rede LSTM Bidirecional treinada para prever a movimentação da cotação dos ativos da empresa analisada. Os autores combinaram dados históricos de preços, variação cambial do dólar, taxa Selic e valor do índice *Dow Jones* com os sentimentos extraídos de notícias do Twitter.

Januário *et al.* [91] construíram uma base composta de 828 notícias extraídas de diferentes portais de divulgação brasileiros com rótulos manualmente atribuídos. Os textos coletados mencionam empresas de grande volume negociado na [B]<sup>3</sup>. Em adição, os autores realizaram experimentos com SA envolvendo diferentes classificadores, tais como abordagem léxica, Naive Bayes e MLP. Os resultados reportados indicam que os classificadores envolvendo ML obtiveram melhor performance ao serem comparados com a abordagem léxica para o domínio específico da pesquisa.

### 3.3 Discussão

A Tabela 3.1 apresenta uma análise comparativa entre os principais pontos de destaque explorados nesta dissertação e os recursos utilizados pelos trabalhos mencionados. Em sua maioria, a literatura atual abrange trabalhos que somente utilizam dados de mercado para a negociação de ativos na bolsa de valores [21, 25, 72] ou ainda trabalhos que exploram a relação do sentimento extraído nos textos no domínio das finanças com a movimentação de índices de mercado [23, 11, 24]. Entretanto, existe uma lacuna nas publicações que envolve a combinação dos dados textuais de notícias sobre o mercado financeiro e os dados de mercado, como as cotações diárias e os indicadores técnicos. Nesse sentido, esta dissertação apresenta uma abordagem contextual, que busca explorar a combinação de múltiplas fontes de informação na negociação de ativos do mercado financeiro brasileiro com o uso de agentes de DRL para a tomada de decisão de compra e venda de ativos selecionados.

---

<sup>1</sup>[http://hedonometer.org/timeseries/en\\_all](http://hedonometer.org/timeseries/en_all)

Tabela 3.1: Comparação de Características com os Trabalhos Relacionados.

<b>Trabalho</b>	<b>DRL</b>	<b>Dados OHLCV</b>	<b>Indicadores Técnicos</b>	<b>Dados Textuais</b>
Théate e Ernst [21]	✓	✓		
Carta <i>et al.</i> [25]	✓	✓	✓	
Taggian <i>et al.</i> [1]	✓	✓	✓	
Yang <i>et al.</i> [78]	✓	✓	✓	
Koratomaddi <i>et al.</i> [84]	✓	✓		✓
Sert <i>et al.</i> [86]				✓
Vargas <i>et al.</i> [90]				✓
Januário <i>et al.</i> [91]				✓
<b>Este trabalho</b>	✓	✓	✓	✓



# Capítulo 4

## Metodologia

Este capítulo apresenta uma descrição dos módulos e componentes desenvolvidos para a construção de um modelo contextual que permite a negociação individualizada de ativos no mercado financeiro brasileiro com base em DRL. A seguir será apresentada a visão geral da proposta, contendo a descrição das fontes de dados utilizadas, as transformações efetuadas, o direcionamento das informações trocadas entre os módulos e seus respectivos componentes e a funcionalidade específica de cada item.

### 4.1 Visão Geral

A visão geral dos componentes utilizados para desenvolver o modelo contextual de negociação de ativos com DRL é ilustrada na Figura 4.1. O modelo contextual proposto é formado por três módulos: (i) **Fontes de Dados**: módulo responsável por fornecer ao módulo Ambiente de Negociação os dados de mercado (e.g. cotações e indicadores técnicos), dados alternativos (e.g. sentimento extraído de textos de notícias e publicações em redes sociais) e dados macroeconômicos (e.g. taxa Selic e IPCA) que serão utilizados para a criação do modelo contextual de negociação a partir do treinamento do módulo Agente de DRL; (ii) **Ambiente de Negociação**: módulo responsável por organizar os diferentes tipos de dados recebidos, realizar transformações e fornecer as informações necessárias ao módulo Agente de DRL (e.g. quantidade monetária disponível em carteira, quantidade de papéis adquiridos, conjunto atual dos dados selecionados); e (iii) **Agente de DRL**: algoritmo de DRL modelado para a negociação individualizada de ativos do mercado financeiro a partir da combinação de dados estruturados e não-estruturados.

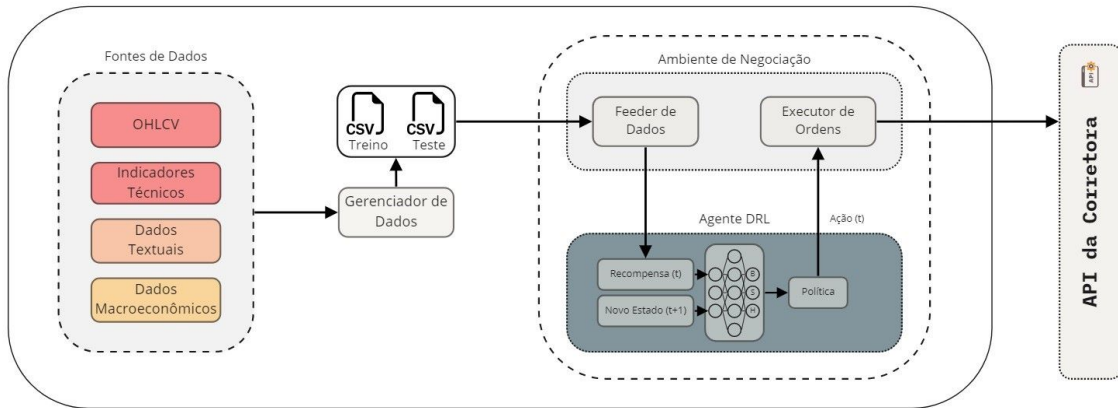


Figura 4.1: Visão Geral dos Componentes do Modelo de Negociação por DRL.

## 4.2 Fontes de Dados

Os dados utilizados para o desenvolvimento do modelo serão apresentados nas próximas subseções. Os dados são classificados em quatro categorias: (i) **Dados de Mercado**: São os dados compostos pela série de preços dos ativos, agregados em períodos diários; (ii) **Indicadores Técnicos**: São dados obtidos pela série de representações numéricas baseadas em heurísticas obtidas a partir dos dados mercado; (iii) **Dados Textuais**: São as polaridades dos sentimentos extraídos das notícias e publicações em redes sociais que mencionam as empresas de interesse; e (iv) **Dados Macroeconômicos**: São os dados relacionados à economia do país.

### 4.2.1 Dados de Mercado

Os dados de mercado utilizados no modelo contextual correspondem aos dias de negociação na [B]<sup>3</sup> entre o início de 2018 (02/01/2018) e início do mês de Julho de 2022 (01/07/2022). A seleção das empresas foi realizada de acordo com maior volume médio de contratos negociados no período, agrupados por setor de atuação, a fim de realizar o estudo entre as principais empresas de maior liquidez em cada setor. Para ilustrar a importância de cada setor, a Tabela 4.1 apresenta a composição percentual de cada um dos cinco maiores setores participantes no Ibovespa, com dados obtidos no dia 03/08/2022 diretamente do *site* oficial da [B]<sup>3</sup>. A composição sofre pequenas variações diariamente, de acordo com a flutuação do preço das ações e outros fatores. Observa-se que mais da metade do índice (50.46%) é composto por bancos e empresas ligadas a *commodities*. Com isso, para o presente trabalho, foram selecionadas as seguintes empresas em cada se-

tor: (i) Bancos: **ITUB4**; (ii) Exploração. Refino e Distribuição: **PETR4**; (iii) Minerais Metálicos: **VALE3**; (iv) Energia Elétrica: **EQTL3**; e (v) Serviços Financeiros Diversos: **B3SA3**.

Tabela 4.1: Os 5 maiores setores que compõem o Ibovespa - Maio/22 a Agosto/22

Setor	Participação no Índice (%)
Bancos	18.75%
Exploração. Refino e Distribuição	17.44%
Minerais Metálicos	14.27%
Energia Elétrica	9.27%
Serviços Financeiros Diversos	3.63%
<b>Total</b>	<b>63.36%</b>

Além disso, para o desenvolvimento do modelo contextual foram explorados os dados OHLCV de cada empresa, agregados em períodos diários, que correspondem ao preço de abertura, preço máximo, preço mínimo, preço de fechamento e volume de contratos negociados, respectivamente. Como o acompanhamento em tempo real é custoso, e o modelo contextual é focado para o uso de investidores comuns, esta dissertação utilizou os dados agregados em períodos diários.

## 4.2.2 Indicadores Técnicos

A adição dos indicadores técnicos complementa a base de dados construída para o desenvolvimento do modelo contextual proposto. Nesse sentido, os dados numéricos fornecidos pelos indicadores aumentam a informação total acerca do momento do ativo em relação à sua série histórica. Para o cálculo dos indicadores técnicos considerados, foi utilizada a biblioteca TA-Lib<sup>1</sup>.

A seleção das métricas considerou diferentes técnicas de análise, considerando sinais de tendência, volatilidade, direcionamento e volume. Dessa forma, os seguintes indicadores de análise técnica foram selecionados: (i) Indicadores de Tendência: *Average Directional Movement Index* (ADX), *Simple Moving Average* (SMA) e *Moving Average Convergence Divergence* (MACD); (ii) Indicadores de *Momentum*: *Relative Strength Index* (RSI); e (iii) Indicadores de Volatilidade: *Bollinger Bands* (BBANDS).

## 4.2.3 Dados Textuais

Ao seguir com a construção das bases de dados fornecidas ao modelo contextual de negociação, os dados textuais foram utilizados como fonte adicional de informação relativa

<sup>1</sup><https://github.com/mrjbq7/ta-lib>

aos ativos negociados. As notícias publicadas sobre a temática do mercado financeiro em portais especializados e redes sociais contribuem com a disseminação da informação para os investidores. No mundo das finanças, o conhecimento de fatos relevantes em tempo real pode proporcionar grandes oportunidades. Por determinação da CVM (Comissão de Valores Mobiliários), as empresas listadas na [B]<sup>3</sup> são obrigadas a disponibilizar uma área de RI (Relações com Investidores), ambiente em que ocorre a divulgação dos fatos relevantes por parte da empresa.

No entanto, acompanhar ativamente o RI de cada companhia registrada não é uma tarefa trivial. Além de ser um trabalho moroso, os portais de divulgação não apresentam uma estrutura de consulta padronizada. Além disso, a obtenção e organização de dados não-estruturados provenientes de diferentes fontes é desafiadora [92]. Ainda, as notícias divulgadas dentro do horário de mercado são propagadas em diferentes portais, o que pode prejudicar os investidores mais especializados a tomarem decisões em tempo real. Em face dos desafios mencionados, empresas como a Bloomberg<sup>2</sup> oferecem soluções de fornecimento centralizado de notícias para os profissionais do mercado financeiro.

A Figura 4.2<sup>3</sup> apresenta um exemplo de título de notícia extraído do portal *Suno Research*. Apesar do texto curto, o exemplo contém informações relevantes para o investidor na tomada de decisão. A manchete apresenta menção a três empresas distintas, seguidas pelo código de negociação (sublinhado em vermelho): Petrorio, BRF e Marfrig. Além disso, indica a data de publicação inicial da notícia, destacada pelo retângulo verde. A notícia foi publicada em um sábado, após o encerramento das negociações da Bovespa. Além disso, é possível identificar um sentimento positivo relacionado ao texto, no qual as palavras “lídera”, “ganhos” e “anima” evidenciam tal característica. O módulo Gerenciador de Dados será responsável por extrair informações dos dados textuais de notícias, transformar os dados e encaminhá-los ao Agente de DRL.

## NEGÓCIOS

# Ibovespa: Petrorio (PRIO3) lidera ganhos da semana; fusão de BRF (BRFS3) e Marfrig (MRFG3) anima mercado



por Bruno Galvão

02/10/2021 17:05

Atualizado em: 02/10/2021 17:09

Figura 4.2: Exemplo de Conteúdo Textual Extraído do Portal de Notícias *Suno Research*.

<sup>2</sup><https://www.bloomberg.com.br/>

<sup>3</sup><https://www.suno.com.br/noticias/ibovespa-petrorio-prio3-brf3-marfrig/>

A extração dos dados textuais provenientes dos portais de notícias foi viabilizada a partir da criação de *crawlers*, implementados com as funcionalidades do *framework* de código aberto Scrapy<sup>4</sup>. Os *scripts* desenvolvidos percorrem a lista de notícias publicadas pelos portais e armazena o conteúdo de interesse em formato JSON para futura manipulação. Os seguintes dados foram coletados: (i) **Tópico da notícia**: É a classificação que cada portal atribui ao conteúdo explorado na publicação; (ii) **Título da notícia**; (iii) **Data e Hora de publicação**; (iv) **Texto completo da notícia**; (v) **Link**: URL para a página original da publicação; e (vi) **Tags**: Assuntos relacionados à notícia (e.g. *ticker* das empresas mencionadas e principais assuntos), sendo que cada portal tem um padrão de preenchimento deste campo específico. A Figura 4.3 apresenta o conteúdo extraído da Figura 4.2, em formato JSON.

```

"topic": "Negócios",
"title": "Ibovespa: Petrorio (PRIO3) lidera ganhos da semana; fusão de BRF(BRFS3) e Marfrig (MRFG3) anima mercado",
"date": "02/10/2021 17:05",
"full_text": "O Ibovespa inaugurou bem o mês de outubro, fechou em forte alta seu primeiro pregão. (...)",
"url": "https://www.suno.com.br/noticias/ibovespa-petrorio-prio3-brf3-marfrig/",
"tags": [
  "Bradesco (BBDC3/BBDC4)",
  "BRF (BRFS3)",
  "Ibovespa",
  "Itaú Unibanco (ITUB3/ITUB4)",
  "Marfrig (MRFG3)",
  "PetroRio (PRIO3)"
]

```

Figura 4.3: Exemplo de Conteúdo Textual Extraído do Portal de Notícias *Suno Research*.

Os resumos de desempenho semanal das ações do Ibovespa geralmente são apresentados em notícias que contém múltiplas referências aos *tickers* negociados, tanto no título da notícia quanto nas *tags* relacionadas. O módulo Gerenciador de Dados será responsável por atribuir um menor valor às empresas identificadas neste tipo de notícia. Alternativamente, será responsável por atribuir um maior valor às notícias que mencionam uma empresa específica.

A obtenção dos dados textuais de múltiplos portais de notícias foi realizada e publicada em [92]. Além disso, os autores propuseram uma solução que categoriza textos curtos para retratar a situação atual do mercado financeiro. A solução proposta obtém os dados e seus respectivos rótulos dos portais de notícias que, quando combinados com um mecanismo de incorporação de palavras, são usados como entrada para a rede LSTM Bidirecional realizar a classificação dos textos de notícias do mercado financeiro.

<sup>4</sup><https://scrapy.org/>

A Tabela 4.2 apresenta a quantidade de notícias por empresa no portal *Suno Research*. A plataforma possibilita a criação de *crawlers* que fazem a busca por empresa sendo mencionada na notícia. Sendo assim, foram realizadas cinco buscas distintas, uma para cada empresa, resultando no quantitativo de notícias exposto a seguir, após filtrar as datas de publicação entre 02/01/2018 e 01/07/2022:

Tabela 4.2: Quantidade de Notícias por Empresa no portal Suno Research

Empresa (Ticker)	Quantidade
B3 (B3SA3)	605
EQUATORIAL (EQTL3)	43
ITAUUNIBANCO (ITUB4)	588
PETROBRAS (PETR4)	1.344
VALE (VALE3)	1.012
<b>Total</b>	<b>3.592</b>

Em adição aos títulos de notícias, os dados textuais do Twitter<sup>5</sup> também foram coletados para a construção da base de dados. O *Twitter* é uma rede social americana criada em 2006 que teve como foco inicial a troca de mensagens textuais curtas, denominadas *tweets*. Com a popularização do uso da plataforma, novos recursos foram adicionados, possibilitando a inclusão de imagens, vídeos, *links*, *retweets* e curtidas nas publicações. O uso do *Twitter* tornou-se popular como meio de divulgação de notícias por parte das autoridades internacionais, artistas, jornais e influenciadores. A Figura 4.4 apresenta um exemplo de *tweet* extraído do perfil do portal de notícias *Infomoney*. O conteúdo textual é curto, no entanto, pode-se identificar menção a 4 empresas listadas na bolsa (MGLU3, VIIA3, CASH3 e MRVE3) e um sentimento positivo relacionado à notícia ao mencionar que as “ (...) ações de varejistas e construtoras saltam mais de 10% (...)”. Além disso, a publicação apresenta uma quantidade não nula de curtidas e *retweets*, que são compartilhamentos da notícia por outros perfis da rede social.

Foram mapeados os perfis oficiais de cada companhia, bem como perfis alternativos relacionados a publicações do mercado financeiro. O histórico de publicações foi coletado com a biblioteca Tweepy<sup>6</sup>, que realiza requisições ao Twitter via API utilizando credenciais de desenvolvedor fornecidas gratuitamente pela rede social. A Tabela 4.3 apresenta a quantidade de *tweets* coletados via API do Twitter para cada uma das empresas analisadas no projeto. O filtro de data aplicado corresponde ao início do treinamento do modelo e ao final do período de teste. As datas utilizadas para o corte foram 02/01/2018 e 01/07/2022, respectivamente.

O quantitativo de publicações obtidos na Tabela 4.3 mostrou-se pouco expressivo em relação ao número total de dias de negociação avaliados durante o período de treino e

<sup>5</sup><https://twitter.com/>

<sup>6</sup><https://www.tweepy.org/>



Figura 4.4: Exemplo de *tweet* Extraído do Perfil do *Infomoney*.

Tabela 4.3: Quantidade de *tweets* nos Perfis Específicos

Nome da Perfil (Ticker)	Usuário do Twitter	Quantidade de <i>tweets</i>
B3 (B3SA3)	@B3_Oficial	630
CEA Grupo Equatorial (EQTL3)	@cea_equatorial	9
Itaú (ITUB4)	@itau	0
Postos Petrobras (PETR4)	@postospetrobras	22
Vale no Brasil (VALE3)	@valenobrasil	57

teste. Por esse motivo foi incentivada a busca por outros perfis relacionados ao mercado financeiro que divulgam notícias sobre as empresas analisadas durante o projeto. A Tabela 4.4 apresenta a lista dos 13 perfis selecionados para a extração de *tweets* relacionados ao mercado financeiro, bem como a quantidade de *tweets* extraídos de cada um. As datas de seleção dos *tweets* são as mesmas da tabela Tabela 4.3.

O conjunto de dados textuais extraído dos perfis relacionados ao mercado financeiro foi pré-processado e encaminhado para o módulo de NER. O módulo consiste em identificar quais empresas, dentre as exploradas nesse projeto, que estão sendo mencionadas no texto completo do *tweet*. Nesse sentido, uma publicação é considerada de interesse se menciona pelo menos uma das empresas. As regras de identificação buscam encontrar uma ocorrência exata ou do *ticker* ou do nome do pregão da empresa (listado no *site* oficial da [B]<sup>3</sup>). A Tabela 4.5 apresenta a quantidade de novos *tweets* extraídos dos perfis relacionados com referência explícita às companhias de interesse após a aplicação do

Tabela 4.4: Quantidade de *tweets* nos Perfis Relacionados

Nome do Perfil	Usuário do Twitter	Quantidade de <i>tweets</i>
Acionista.com.br	@acionistacombr	2.724
Arko Advice	@arkoadvice	585
Bloomberg Brasil	@BloombergBrasil	1.287
Brazil Journal	@BrazilJournal	3.133
Bússola do Investidor	@bussoladoinves	2.890
Economia Estadão	@EstadaoEconomia	3.167
Fatos Relevantes S.A.	@Fatosdomercado	1.310
Folha Mercado	@folha_mercado	2.354
InfoMoney   IM+	@infomoney	2.105
Investidores Brasil Oficial	@InvestidoresO	1.430
Investing.com Brasil	@InvestingBrasil	1.497
Valor Econômico	@valoreconomico	1.680
Valor Investe	@valorinveste	1.722

módulo de NER. Novamente, os dados foram filtrados de acordo com a data de publicação.

Tabela 4.5: Novos *tweets* com Referência às Companhias

Ticker	Quantidade
B3 (B3SA3)	908
EQUATORIAL (EQTL3)	53
ITAUUNIBANCO (ITUB4)	152
PETROBRAS (PETR4)	1.240
VALE (VALE3)	598
<b>Total</b>	<b>2.951</b>

Por fim, os dados textuais do Twitter foram agrupados. Cada *tweet* publicado ou pelo perfil oficial da empresa ou mencionado em uma das contas relacionadas ao mercado financeiro será atribuído ao *ticker* de referência (regras apresentadas na subseção 4.4.1). A quantidade de *tweets* para cada empresa é a soma simples da quantidade de publicações contidas na Tabela 4.3 e na Tabela 4.5. A extração dos sentimentos a partir dos dados textuais coletadas será explorada na subseção 4.3, em que será descrita a metodologia adotada para o cálculo da polaridade nas publicações.

#### 4.2.4 Dados Macroeconômicos

Com o objetivo de fornecer dados relacionados ao desempenho geral da economia do país ao modelo contextual de negociação foram coletados dois dos principais dados macroeconômicos nacionais via API do Banco Central do Brasil<sup>7</sup>. Os indicadores macroeconômicos escolhidos foram: (i) Taxa Selic; e (ii) IPCA Acumulado de 12 meses. Os dados numéricos referentes à taxa Selic foram convertidos para variações diárias, sendo que

<sup>7</sup><https://dadosabertos.bcb.gov.br/dataset/11-taxa-de-juros---selic>



este é fixado durante determinado período de tempo. O IPCA acumulado de 12 meses é atualizado de forma mensal e foi atribuído o mesmo valor ao mês de referência.

### 4.3 Gerenciador de Dados

Esta seção irá apresentar o componente Gerenciador de Dados, que fornece ao *Feeder* de dados do modelo contextual os dados tratados e transformados de acordo com a necessidade de ajuste em cada tipo de dado. O algoritmo de NER é responsável por identificar as empresas mencionadas nas notícias e atribuir a cada conteúdo textual a referência de cada companhia mencionada. O algoritmo de SA é utilizado após a identificação das empresas nos conteúdos textuais, gerando representações numéricas referentes às polaridades das notícias agrupadas pelo algoritmo de NER. O componente Transformação dos Dados é responsável por selecionar e transformar as variáveis recebidas. A Figura 4.5 apresenta uma visão geral dos recursos deste componente.

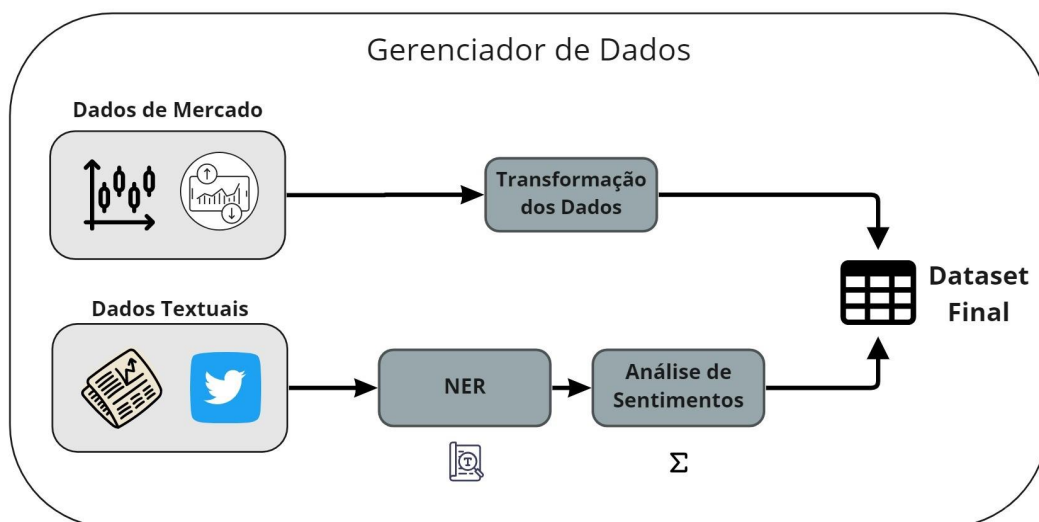


Figura 4.5: Visão Geral do Componente Gerenciador de Dados.

Em sequência, os dados são separados em dois conjuntos: (i) treino; e (ii) teste. A Tabela 4.6 apresenta a divisão treino/teste estabelecida para o treinamento e avaliação do modelo de DRL escolhido. Os dados mais recentes, que correspondem ao primeiro semestre do ano de 2022, foram utilizados para a avaliação do modelo contextual. Os dados de avaliação contém um total de 125 dias de negociação na bolsa, o que corresponde a pouco mais de 10% dos dados diários totais utilizados para o desenvolvimento da pesquisa.

Tal separação em treino e teste teve como objetivo fornecer ao modelo contextual um maior quantitativo de observações para o seu treinamento. Observa-se na Tabela 4.7

Tabela 4.6: Divisão Treino/Teste dos Dados de Mercado e Indicadores Técnicos

Conjunto	Datas Início e Fim	Quantidade de dias de <i>trades</i>
Treino	02/01/2018 - 30/12/2021	989 (88.78%)
Teste	03/01/2022 - 01/07/2022	125 (11.22%)

que os dados textuais começam a apresentar quantitativo relevante de notícias a partir do ano de 2019. Com isso, um período maior com as informações extraídas dos dados textuais foi utilizado, resultando em quase 90% das observações para o conjunto de treino. Para a avaliação, o período mais recente disponível no desenvolvimento da dissertação foi escolhido.

Tabela 4.7: Quantidade de Notícias por Ano

Ano de Publicação	Quantidade
2018	39
2019	475
2020	891
2021	1.544
2022* (Até 01/07)	640
<b>Total</b>	<b>3.589</b>

O algoritmo de NER dispõe do mecanismo de extração de informação nas notícias armazenadas em formato JSON e nos *tweets*. Para isso, o componente foi desenvolvido com regras manualmente fornecidas para identificar menções às empresas nos títulos das notícias, nas *tags* relacionadas e no texto completo dos *tweets*. Nesse sentido, se uma mesma notícia ou *tweet* menciona duas empresas, tal conteúdo textual será utilizado na construção da base de cada uma das duas empresas mencionadas. As regras manuais buscam a ocorrência exata do *ticker* ou o nome de pregão oficial da empresa analisada.

O algoritmo de SA recebe os textos já agrupados por empresa e realiza a classificação da notícia de acordo com a polaridade identificada pelas LFs desenvolvidas. A representação dos sentimentos foi realizada de forma diária, sendo esta uma das *features* do conjunto de dados final fornecido para o treinamento do Agente de DRL. Os detalhes da extração de menções e do cálculo dos sentimentos diários são descritos nas Subseções 4.3.1 e 4.3.2, respectivamente.

### 4.3.1 Reconhecimento de Entidades Nomeadas

O componente de Reconhecimento de Entidades Nomeadas foi desenvolvido com o objetivo de filtrar as publicações por empresa, possibilitando a seleção mais assertiva dos conteúdos textuais de cada ativo analisado. Para tanto, foram utilizadas expressões regulares e

identificação de palavras-chave para atribuir os rótulos nas publicações de acordo com a empresa que foi mencionado no conteúdo textual proveniente do *Twitter*.

O pseudocódigo apresentado abaixo descreve o processo de identificação das empresas mencionadas no conteúdo textual proveniente do *Twitter*. Inicialmente, são coletados os históricos de publicação dos perfis mapeados. Em seguida, para cada *tweet*, as palavras são isoladas em uma lista, que passa por um processo de limpeza dos dados. Na sequência, para cada palavra da lista já processada, é realizada a busca pela correspondência exata do *ticker* ou do nome da empresa. Caso pelo menos uma menção for encontrada, o texto será associado ao *ticker* da empresa identificada. Por fim, será retornado uma base de dados contendo somente os *tweets* em que houve pelo menos uma menção identificada.

---

**Algorithm 1** Componente de Reconhecimento de Entidades Nomeadas

---

```
1:  $\mathbf{T}$  = Download dos tweets de uma conta específica.
2:  $\mathbf{F}$  = {}
3: for  $t \in T$  do
4:    $\mathbf{W}$  = Lista de palavras em t.
5:    $\mathbf{W}'$  = Lista de palavras preprocessada.
6:    $\mathbf{MT}$  = Set com os tickers das empresas mencionadas.
7:   for  $w \in W'$  do
8:     if  $w \in tickers \vee w \in nome\ da\ empresa$  then  $\mathbf{MT} = \mathbf{MT} \cup \{w\}$ 
9:     end if
10:  end for
11:  for  $tk \in MT$  do  $\mathbf{F} = \mathbf{F} \cup \{(t, tk)\}$ 
12:  end for
13:   $\mathbf{MT} = \{\}$ 
14: end for
15: return  $F$ 
```

---

Os dados textuais provenientes do portal de notícia *Suno Research* foram devidamente categorizados por empresa pelo *crawler* desenvolvido, o que dispensa a busca por menções da empresa em todos os textos coletados deste portal.

### 4.3.2 Análise de Sentimentos

Um dos maiores desafios encontrados é a escassez de ferramentas disponíveis para a classificação de sentimentos em textos específicos do mercado financeiro. Alguns trabalhos utilizam a ferramenta VADER [83] para a classificação de sentimentos em textos diversos. No entanto, tal recurso foi desenvolvido para textos em inglês, mais especificamente para conteúdos publicados em redes sociais. Além disso, a falta de bases cuidadosamente rotuladas dificulta a comparação do desempenho entre diferentes abordagens de classificação textual no domínio financeiro.

Em face dos desafios mencionados anteriormente, uma abordagem de rotulagem dos dados com Supervisão Fraca (do inglês, *Weak Supervision*) (WS) foi adotada na classificação dos sentimentos das notícias coletadas para o treinamento do modelo contextual. A classificação dos sentimentos nos títulos de notícias e nos textos do Twitter foi realizada a partir da criação de heurísticas e regras manualmente fornecidas. Por exemplo, a presença das palavras *lucro*, *ótimo* e *melhor* indicam um sentimento positivo no texto. Por outro lado, palavras como *prejuízo*, *péssimo* e *pior* indicam um sentimento negativo no texto analisado. O *framework* Snorkel AI<sup>8</sup> foi utilizado para a criação das regras de classificação. Cada regra é representada por uma Função de Rotulagem (do inglês, *Labelling Function*) (LF). A partir das LFs desenvolvidas um modelo de classificação é construído, capaz de lidar com conflitos resultantes da rotulagem simultânea de diferentes LFs [93]. As LFs desenvolvidas incluem regras:

- **Baseadas em Dicionários:** Heurísticas baseadas em busca de palavras-chave em dicionários de termos específicos (adjetivos e substantivos); Foram definidas duas listas iniciais para os adjetivos: (i) adjetivos positivos; e (ii) adjetivos negativos. Cada lista foi preenchida com alguns dos adjetivos mais observados nos títulos das notícias (e.g destaque, maior, positivo, otimista e ótimo, as publicações com polaridade positiva). Em seguida, a lista foi aumentada a partir da consulta de sinônimos para as palavras inicialmente listadas, resultando em mais adjetivos e substantivos relacionados às palavras inicialmente propostas.
- **Baseadas em Expressões Regulares:** Heurísticas criadas a partir do desenvolvimento de expressões regulares comumente encontradas nos títulos de notícias; No domínio das finanças, expressões como “distribuição de dividendos” e “fechamento em alta” são caracterizadas como positivas. Já expressões como “previsão de queda” e “resultado negativo” são caracterizadas como negativas. Dessa forma, as LFs foram construídas com o objetivo de identificar a ocorrência de tais expressões nos textos.

A composição de diferentes LFs acelera o processo de rotulagem dos dados, reduzindo o custo de criação de base por especialistas no domínio específico. O novo paradigma de rotulagem programática permite a criação de grandes bases de treinamento para modelos de ML, com destaque na escalabilidade e adaptabilidade das LFs. Outra vantagem evidente é a capacidade de gerenciar o controle de qualidade sobre as regras criadas. O modelo treinado lida com conflitos existentes entre as diferentes classificações propostas pelas LFs. Os dados textuais que apresentarem conflitos na rotulagem final serão classificados como neutros pelo algoritmo de SA.

---

<sup>8</sup><https://snorkel.ai/>



Os resultados agregados do Twitter foram calculados de maneira similar, com a adição de um fator multiplicativo em decorrência da quantidade de *retweets* na publicação analisada. A Equação 4.3 apresenta o cálculo do sentimento geral para os *tweets* coletados, em que  $s(x)$  representa a média aritmética do sentimento associado a todos os *tweets* rotulados no dia  $d_i$  após a classificação da função  $S(\cdot)$ .

$$s(x) = \frac{1}{N} \sum_i^N S(d_i) * \text{RTs Count} \quad (4.3)$$

Em seguida, os resultados foram ajustados de acordo com três categorias possíveis: (i) POSITIVO: o valor 1 foi atribuído aos dias cujo resultado de  $s(x)$  foi positivo; (ii) NEGATIVO: o valor  $-1$  foi atribuído aos dias cujo resultado de  $s(x)$  foi negativo; e (iii) NEUTRO: o valor 0 foi atribuído aos dias cujo resultado de  $s(x)$  foi 0 ou não houve notícia publicada. O valor final do sentimento agregado de forma diária é expresso pela Equação 4.4:

$$\text{Sentimento Final}_{\text{Twitter}} = \begin{cases} 1, & \text{se } s(x) > 0 \\ -1, & \text{se } s(x) < 0 \\ 0, & \text{se } s(x) = 0 \end{cases} \quad (4.4)$$

O conjunto de dados completo para treino/teste é constituído de 38 vetores de características, fornecendo ao ambiente de negociação os dados históricos de preço e volume, indicadores técnicos, dados numéricos sobre os sentimentos extraídos dos dados textuais e os dados macroeconômicos. A Figura 4.7 apresenta uma amostra dos vetores de características presentes nas observações iniciais do conjunto de teste.

date	open	high	low	close	volume	volume_obv	rsi	adx	ma_5	ma_20	sent_score_news	sent_score_twitter	selic	ipca
03/01/2022	76,08	77,01	75,32	75,70	14999400	1,29E+09	57,09	21,85	75,25	74,66	1	0	8,76	10,38
04/01/2022	75,60	76,14	74,05	74,28	15146000	1,28E+09	52,37	20,89	74,85	74,95	0	0	8,76	10,38
05/01/2022	74,51	75,77	74,27	74,82	18215200	1,30E+09	53,93	20,00	75,02	75,05	0	0	8,76	10,38
06/01/2022	75,73	77,38	75,41	76,60	19267500	1,31E+09	58,75	19,90	75,44	75,18	0	0	8,76	10,38
07/01/2022	77,42	79,54	76,84	79,47	21407000	1,34E+09	65,09	20,60	76,18	75,51	0	0	8,76	10,38

Figura 4.7: Primeiras Observações do Conjunto de Teste.

## 4.4 Ambiente de Negociação

Esta seção irá apresentar o Ambiente de Negociação, que é o módulo do modelo contextual responsável por simular a interação de um investidor com a corretora. Nesse sentido, o Ambiente de Negociação fornece ao módulo do Agente de DRL os dados em formato estruturado e prepara o envio de ordens para a bolsa de valores, a partir das instruções

de compra e venda emitidas pelo Agente de DRL. Para a transformação dos dados desenvolvidos, o componente Gerenciador de Dados, cujo objetivo é tratar os dados recebidos pela Fonte de Dados e enviar ao Agente de DRL uma versão estruturada das informações contidas tanto nos dados numéricos quanto nos dados textuais relacionados ao mercado financeiro. A comunicação com a bolsa de valores pode ser simulada pelo componente Executor de Ordens, que tem por finalidade transformar os comandos de compra e venda emitidos pelo Agente de DRL em requisições para uma API de uma corretora. Para a modelagem do Ambiente de Negociação o *framework* TensorTrade<sup>9</sup> foi utilizado, seguindo a especificação dos ambientes de RL estabelecida pela biblioteca OpenAI Gym<sup>10</sup>.

A criação de um ambiente com o TensorTrade apresenta como requisito a definição de diferentes componentes para representar as informações sendo trocadas entre o Ambiente de Negociação e o Agente de DRL, que será descrito em mais detalhes na próxima Seção. A seguir serão listados os principais componentes do Ambiente de Negociação do TensorTrade:

- **Bolsa de Valores:** É a bolsa de valores fictícia a ser utilizada no Ambiente de Negociação, responsável por enviar as ordens. A bolsa simulada utiliza uma taxa de comissão de 0.35% por *trade*. É representada pela Classe “*Exchange*”.
- **Instrumentos:** Os instrumentos representam os ativos financeiros a serem negociados pelo ambiente. São identificados por uma representação textual (e.g. “BRL” para a quantidade monetária e “ASSET” para a quantidade de papéis) e pela precisão do valor numérico (e.g. 2 casas decimais para ambos os instrumentos criados). São representados pela Classe “*Instrument*”.
- **Carteira:** A carteira armazena o balanço (quantidade) de cada instrumento vinculado a uma bolsa de valores. É representada pela Classe “*Wallet*”.
- **Portfólio:** O portfólio é representação da composição das diferentes carteiras. É representado pela Classe “*Portfolio*”.
- **Fonte de Dados:** É o componente responsável por agrupar diferentes séries de dados numéricos de forma organizada. Contém todas as *features* selecionadas para o treinamento do Agente de DRL. É representado pela Classe “*DataFeed*”.
- **Recompensa:** A recompensa é a representação numérica em cada etapa do tempo durante um episódio. Foram utilizados dois esquemas de recompensa: (i) Lucro Simples: esquema de recompensa que utiliza somente variações monetárias na carteira. É representado pela Classe “*default.rewards.SimpleProfit*”; e (ii) Lucro ajustado

---

<sup>9</sup><https://www.tensortrade.org/en/latest/index.html>

<sup>10</sup><https://gym.openai.com/>

pelo Risco: esquema de recompensa que utiliza o ShR nas transações realizadas. É representado pela Classe “*default.rewards.RiskAdjustedReturns*”.

## 4.5 Agente de DRL

Esta seção irá apresentar o módulo Agente de DRL, que tem como objetivo aprender regras para a negociação individualizada de ativos na bolsa de valores a partir da combinação de diferentes tipos de dados do mercado financeiro. O Agente de DRL será a representação virtual de um *trader*, que é um investidor cujo propósito é obter lucro na bolsa de valores ao aproveitar os momentos de volatilidade do mercado. O objetivo do Agente de DRL é criar uma política que maximize a soma descontada das recompensas durante a etapa de treinamento.

Para atingir o objetivo proposto, o Agente de DRL recebe os dados devidamente tratados fornecidos pelo componente Gerenciador de Dados do Ambiente de Negociação. A partir dos dados recebidos, o Agente de DRL é treinado para formular uma política de negociação que execute ordens de compra e venda de um ativo específico. O retorno do Agente de DRL leva em consideração não somente o lucro total acumulado, mas também métricas de risco associado à volatilidade da série de preços do ativo.

Para a modelagem do agente de DRL foi utilizada a biblioteca RLLib<sup>11</sup>, presente no *framework* Ray [94]. A RLLib oferece grande quantidade de implementações para os mais variados algoritmos de DRL, permitindo o treinamento dos agentes de forma *offline* e também a execução em ambientes de produção. Os principais componentes de RL no contexto da AT serão apresentados a seguir:

- **Agente:** É o módulo responsável pela tomada de decisão final que será enviada ao Ambiente de Negociação em cada novo estado. O algoritmo de DRL irá determinar a política que o agente irá seguir.
- **Política:** É a estratégia de negociação no estado atual, representada por uma distribuição de probabilidade das possíveis ações a serem tomadas a partir do contexto observado pelo Agente.
- **Ações:** É o conjunto de ações possíveis a serem realizadas pelo agente. As ações possíveis são: (i) comprar; (ii) vender; e (iii) não mudar posição (permanecer comprado ou não comprar ativos), caracterizando um cenário discreto. As duas primeiras ações envolvem alteração na quantidade de ações do ativo na carteira e no saldo disponível.

---

<sup>11</sup><https://docs.ray.io/en/latest/rllib/index.html>



- **Estado:** É o vetor que contém as informações essenciais para o agente: (i) dados OHLCV; (ii) indicadores técnicos; (iii) representação numérica dos dados textuais; (iv) quantidade de ações em carteira; e (v) quantidade monetária em carteira.
- **Recompensa:** É a resposta numérica imediata do ambiente após o agente executar uma ação em seu estado atual. Serão avaliadas, separadamente, dois tipos de recompensas: (i) o Lucro/Perda simples; e (ii) Índice Sharpe (ShR).

A Figura 4.8 apresenta a relação entre o Agente de DRL e o Ambiente de Negociação desenvolvido. O algoritmo PPO [73] foi selecionado para a modelagem do Agente de DRL em virtude das características reportadas pelas aplicações em problemas envolvendo negociação de ativos no mercado financeiro em [75, 78]. O Ambiente de Negociação fornecerá ao Agente de DRL as informações disponíveis em cada instante de tempo. Os dados organizados pelo Gerenciador de Dados, bem como as informações sobre o saldo disponível e os títulos adquiridos são apenas uma amostra das informações observadas por um investidor no contexto da negociação de ativos. O Agente de DRL é responsável por perceber o estado atual, avaliar a recompensa recebida a partir da ação executada pela política atual e decidir se irá ajustar a política vigente. As ações tomadas pelo Agente de DRL serão enviadas ao componente Executor de Ordens, que irá formatar a mensagem e enviar o comando a uma Interface de Programação de Aplicações (do inglês, *Application Programming Interface*) (API) externa.

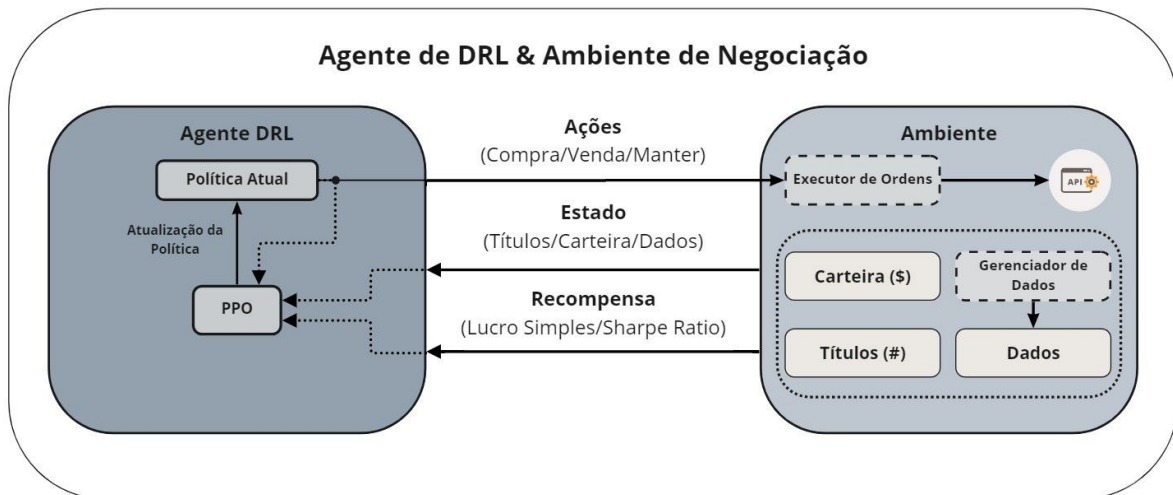


Figura 4.8: Interação entre o Agente de DRL e o Ambiente de Negociação.

## 4.6 Formulação do Problema

Nesta seção são descritos os componentes utilizados para a criação de um modelo contextual de negociação de ativos individualizados na bolsa de valores. São apresentadas as ações e recompensas, o estado e o algoritmo de DRL utilizado para o desenvolvimento dos experimentos.

O desenvolvimento dos componentes e a realização dos experimentos foi utilizado o *framework open-source* Tensortrade<sup>12</sup>. O TensorTrade é constituído de componentes modulares que, quando utilizados em conjunto, podem construir uma estratégia de negociação completa. As estratégias combinam agentes de DRL e lógica de negociação no formato de um ambiente *gym*. Um ambiente de negociação é criado a partir de um conjunto de componentes modulares que podem ser misturados e compostos para criar uma grande diversidade de estratégias de investimento.

Alguns dos principais componentes são listados a seguir:

- **Dados:** O conjunto de dados  $d \in D$  é composto da tupla  $d_t = (M_t, T_t, S_t, MC_t)$ , em que  $M_t$  é representa os dados de mercado,  $T_t$  são os indicadores técnicos,  $S_t$  são os sentimentos de notícias e de redes sociais e  $MC_t$  são os dados macroeconômicos.
- **Observações:** O conjunto de observações enviado ao agente  $o \in O$  é composto pela tupla  $o_t = (t_t, c_t, d_t, r_{t-1})$ , em que  $t_t$  representa a quantidade de títulos (i. e. número de ações) em posse do agente,  $c_t$  representa o saldo monetário atual do agente,  $d_t$  são os dados em posse do agente no instante de tempo  $t$  e  $r_{t-1}$  representa a recompensa obtida a partir da última ação realizada.
- **Estado:** Cada estado é composto das 10 observações mais recentes fornecidas pelo ambiente. Deste modo, o estado  $s_t \in S$  é representado por  $s_t = o_t, o_{t-1}, o_{t-2}, \dots, o_{t-9}$ . A utilização de uma janela composta dos últimos 10 dias úteis de negociação permite uma visualização mais completa por parte do agente na hora da tomada de decisão.
- **Ações:** As ações  $a \in A$  tomadas pelo agente são simples: (i) Compra: converte a totalidade do saldo disponível ( $S_t$ ) em títulos ( $T_t$ ); (ii) Venda: converte a totalidade dos títulos ( $T_t$ ) em saldo monetário ( $S_t$ ); e (iii) Manter: não realiza a compra/venda de títulos no instante em avaliação. O esquema de ações definido dispõe de mecanismos de *stop-loss* e *take-profit* para controlar perdas e realizar ganhos predefinidos, respectivamente. O valor de saída de 7.5% foi estabelecido no caso de perdas. No caso de ganhos, o valor de 10% foi definido como ponto de saída do *trade*.
- **Recompensa:** As recompensas, representadas por  $r$ , são calculadas a cada etapa de tempo. No esquema de recompensa do lucro simples, o cálculo é simplesmente

---

<sup>12</sup><https://github.com/tensortrade-org/tensortrade>

a variação do preço de fechamento do ativo em relação ao momento da compra. No esquema de lucro ajustado pelo risco, o cálculo da recompensa utiliza o ShR, com taxa livre de risco sendo a média da taxa Selic no período de treinamento, convertida para uma variação diária.

Em síntese, no modelo contextual, o agente de DRL recebe as observações extraídas do ambiente como entrada, encaminha os dados recebidos através de modelos internos (e. g. Redes Neurais Artificiais) e produz uma ação como saída. As observações são constituídas por dados OHLCV, indicadores técnicos, sentimentos extraídos de notícias e de redes sociais de um determinado ativo e, por fim, dados macroeconômicos. O agente de DRL irá receber esses dados e determinar se, em determinado momento, irá vender, comprar ou manter os títulos do ativo em questão. O agente não tem intuição acerca dos preços ou *trades* sendo representados por esses valores. Ao invés disso, o modelo simplesmente aprende quais valores irá retornar na saída a partir dos valores de entrada (no caso do agente de DRL do modelo contextual, uma sequência dos valores de entrada  $s_t$ ) a fim de obter a maior recompensa possível.

## 4.7 Considerações Finais

Conforme exposto durante o capítulo, nesta dissertação é proposto um modelo contextual de negociação de ativos com agente de DRL. A implementação do modelo contextual utilizou um agente de aprendizado por reforço profundo, que utiliza redes neurais para a construção das regras de tomada de decisão. Além disso, o componente de análise de sentimentos adotou a abordagem supervisionada fraca para extrair a representação numérica do sentimento identificado nos textos.

# Capítulo 5

## Resultados

Este capítulo tem como objetivo validar o modelo contextual com agente de DRL proposto para a negociação de ativos individualizados do mercado de ações brasileiro, além de analisar o desempenho do modelo em quatro cenários de avaliação com configurações específicas. Inicialmente, serão expostas as descrições dos cenários e os hiperparâmetros selecionados. Em seguida, serão apresentados os resultados nos quatro cenários de avaliação, sendo estes diferentes entre si pelas quantidades de dados disponíveis e pelas funções de recompensa dos agentes de DRL. Os resultados serão agrupados e quantitativamente comparados com os *benchmarks* selecionados. O objetivo da avaliação por cenário é verificar o desempenho do modelo contextual em situações onde mais ou menos informações a respeito do ativo alvo são fornecidas, além de explorar o uso de função de recompensa que considera a volatilidade associada ao investimento. Adiante, os desempenhos do melhor e pior modelo treinado por ativo serão apresentados. Por fim, serão exibidas as considerações finais da seção.

### 5.1 Experimentos

Os experimentos realizados buscaram verificar o desempenho do modelo contextual utilizando um agente de DRL na negociação individualizada de ativos selecionados da [B]<sup>3</sup> em quatro distintos cenários de avaliação. Para tanto, diferentes combinações entre as fontes de dados e funções de recompensa dos agentes de DRL foram exploradas. As avaliações em cenários distintos tem como objetivo comparar os resultados em que mais ou menos dados são fornecidos ao modelo contextual, além de esquemas de recompensa que consideram e desconsideram o risco associado à negociação dos ativos. A seguir, cada um dos cenários é detalhado:

- (i). Cenário que utiliza o esquema de recompensa sem considerar o risco associado (i.e. Lucro Simples) e todas as modalidades de dados disponíveis (i.e. Dados OHLCV, Indicadores Técnicos, Sentimentos dos Dados Textuais e Dados Macroeconômicos);
- (ii). Cenário mais simples. Neste cenário é utilizado o esquema de recompensa sem considerar o risco associado (i.e. Lucro Simples) e somente os dados OHLCV;
- (iii). Cenário mais completo. Neste cenário é utilizado o esquema de recompensa considerando o risco associado (i. e. Índice Sharpe) e todas as modalidades de dados disponíveis (i.e. Dados OHLCV, Indicadores Técnicos, Sentimentos dos Dados Textuais e Dados Macroeconômicos);
- (iv). Cenário que utiliza o esquema de recompensa considerando o risco associado (i.e. Índice Sharpe) e somente os dados OHLCV;

Em cada um dos quatro cenários descritos acima foi utilizada a mesma configuração de ambiente de execução e também os mesmos recursos computacionais. Os *scripts* envolvendo o treinamento dos modelos contextuais com agentes de DRL foram implementados no *Google Colab Pro*. Além disso, todo o código desenvolvido pode ser encontrado no repositório do Github<sup>1</sup> dessa dissertação.

A Tabela 5.1 apresenta os hiperparâmetros otimizados para o treinamento do Agente de DRL dedicado para a negociação de ativos de forma individualizada no modelo contextual. As células LSTM utilizaram as configurações padrão do Tensorflow<sup>2</sup>. Por questões de tempo de treinamento, os resultados destacados na tabela são os valores pertencentes à melhor configuração obtida após execução de *grid-search* no treinamento do modelo contextual com agente de DRL na negociação das ações do *ticker* B3SA3, utilizando a configuração reportada no cenário (i). A melhor configuração selecionada foi aquela que obteve maior métrica de recompensa média por episódio ao final das iterações. As recompensas são calculadas em relação a ação tomada na etapa de tempo imediatamente anterior ao estado atual. No cenário (i), que foi utilizado para a otimização dos hiperparâmetros, a função de recompensa é calculada a partir da diferença de preços entre o momento da compra do ativo e o preço do ativo na etapa atual.

Os resultados a serem apresentados representam o patrimônio líquido final no conjunto de teste após as ações de compra e venda realizadas pelo modelo contextual com agente de DRL, considerando um investimento inicial de R\$ 100.000 livres para a tomada de decisão do modelo. Os resultados foram coletados em 100 avaliações do agente de DRL treinado. Em cada uma dos cenários de avaliação, o agente treinado é restaurado e utilizado para a realização de *trades* com dados externos ao conjunto de treinamento para cada empresa

---

<sup>1</sup>[https://github.com/mso13/Mestrado\\_PPGI](https://github.com/mso13/Mestrado_PPGI)

<sup>2</sup>[https://www.tensorflow.org/api\\_docs/python/tf/keras/layers/LSTM](https://www.tensorflow.org/api_docs/python/tf/keras/layers/LSTM)

Tabela 5.1: Hiperparâmetros testados e selecionados

Descrição	Configurações Possíveis	Configuração Selecionada
Janela de Observações Passadas (em dias)	[5, 10]	<b>10</b>
Estrutura da FFNN	[[256, 256], [128, 64, 32]]	<b>[256, 256]</b>
Quantidade de Células LSTM	[64, 128]	<b>64</b>
Taxa de Aprendizado	[0.001, 0.0001, 0.00001]	<b>0.0001</b>
Gamma ( $\gamma$ )	[0.9, 0.95, 0.99]	<b>0.9</b>

selecionada. A avaliação no conjunto de teste foi realizada em múltiplas passagens com o objetivo de evitar resultados espúrios e analisar a variabilidade da evolução do saldo inicial da carteira. Os resultados finais obtidos consideram uma simulação das taxas cobradas pela bolsa de valores (0.35% por *trade*), o que diminui o patrimônio líquido final, mas o torna mais próximo à realidade dos custos operacionais existentes nas bolsas de valores e corretoras de ativos.

Foram selecionados três *benchmarks* principais para a avaliação do desempenho do modelo contextual: (i) Investimento Inicial: é o valor adotado para todos os cenários de avaliação, composto de R\$ 100.000 disponíveis para a aquisição das ações das empresas; (ii) *Buy & Hold* de cada empresa: é o valor acompanhado desde o primeiro dia de negociação no cenário de avaliação, após a conversão do investimento inicial de R\$ 100.000 em ações de cada empresa (e. g. caso o investimento inicial fosse capaz de adquirir uma única ação da empresa no primeiro dia, o valor final seria a cotação desta mesma ação no último dia disponível no conjunto de avaliação); e (iii) *Buy & Hold* do BOVA11: é o *Exchange Traded Fund* (ETF) que visa replicar os retornos do principal indicador do mercado de ações brasileiro, o Ibovespa.

Os patrimônios líquidos finais reportados após 100 iterações do agente de DRL do modelo contextual treinado e aplicado aos dados do ambiente de teste foram comparadas com os três *benchmarks* descritos. Além disso, a análise das medianas e da distância interquartil dos resultados finais possibilitou identificar valores pertencentes a performances muito distantes do esperado (conhecidas por *outliers*). Ainda, a análise da amplitude interquartil também possibilita verificar a estabilidade buscada ao utilizar recursos computacionais modernos para a realização de compra e venda de ativos no mercado financeiro.

## 5.2 Cenário (i): Todos os Dados e Recompensa por Lucro Simples

Esta seção tem como objetivo avaliar o desempenho do modelo contextual com o fornecimento de todos os tipos de dados disponíveis e uma função de recompensa de lucro simples para o agente de DRL.

A Tabela 5.2 apresenta os resultados de cada um dos ativos no cenário (i), no qual o modelo contextual utiliza todas as modalidades de dados disponíveis e o cálculo da recompensa do agente de DRL envolve apenas o lucro simples. Em apenas 2 dos 5 ativos selecionados observou-se mais da metade das avaliações acima do investimento inicial de R\$ 100.000: B3SA3 e PETR4. Além disso, nenhum dos agentes treinados superou o *benchmark* de *Buy & Hold* das respectivas empresas durante o período de avaliação. Por fim, o desempenho obtido nas avaliações do modelo contextual, em todos os ativos selecionados, mostrou que as estratégias desenvolvidas para o cenário (i) foram superiores ao desempenho médio do mercado, que é representado pelo ETF do Ibovespa em mais da metade das avaliações. De fato, o valor do índice durante o período de avaliação apresentou um cenário de queda. Entretanto, as cinco empresas selecionadas obtiveram valorização em suas ações no mesmo período. Tal observação justifica um melhor desempenho do modelo contextual em superar o *benchmark* (#3) e uma dificuldade em superar o *benchmark* (#2).

Tabela 5.2: Patrimônios Líquidos Finais no Conjunto de Teste - Cenário (i)

Métrica	B3SA3	EQTL3	ITUB4	PETR4	VALE3
Mínimo nas 100 Iterações (R\$)	R\$ 83645.29	R\$ 84058.74	R\$ 84843.17	R\$ 92777.03	R\$ 81515.75
Máximo nas 100 Iterações (R\$)	R\$ 139130.17	R\$ 111566.02	R\$ 128921.04	R\$ 130898.36	R\$ 140267.79
25º Percentil nas 100 Iterações (R\$)	R\$ 95921.07	R\$ 93510.71	R\$ 93292.84	R\$ 103922.82	R\$ 84876.99
50º Percentil nas 100 Iterações (R\$)	R\$ 100946.85	R\$ 100000.00	R\$ 97132.41	R\$ 105742.90	R\$ 98216.93
75º Percentil nas 100 Iterações (R\$)	R\$ 109384.10	R\$ 101918.71	R\$ 103512.80	R\$ 114787.83	R\$ 102990.24
(#1) Acima do Investimento Inicial	54%	38%	35%	97%	37%
(#2) Acima do <i>Buy &amp; Hold</i> da Empresa	30%	13%	14%	12%	45%
(#3) Acima do <i>Buy &amp; Hold</i> do BOVA11	77%	66%	58%	97%	69%

A pequena variação observada nos resultados finais reportados apontaram uma estabilidade no desempenho dos agentes de DRL no modelo em relação ao investimento inicial disponibilizado para a compra e venda dos ativos selecionados. Tal característica é justificada pelo *stop-loss* utilizado e pela dinâmica de treinamento do agente de DRL selecionado.

### 5.3 Cenário (ii): Dados OHLCV e Recompensa por Lucro Simples

Nesta seção é apresentado os resultados do segundo cenário de avaliação, em que é mantido o esquema de recompensa do agente de DRL do modelo contextual no cenário (i), mas limita a quantidade de dados disponibilizados para o treinamento dos agentes. Nesse sentido, busca-se avaliar o desempenho em um cenário menos informativo, o qual dispõe somente dos dados OHLCV dos ativos.

A Tabela 5.3 apresenta os resultados de cada um dos ativos no cenário (ii), no qual o agente de DRL utiliza somente os dados OHLCV e o cálculo da recompensa envolve o lucro simples. No pior caso, apenas dois dos ativos listados superaram o *benchmark* de *Buy & Hold* da empresa específica: B3SA3 e VALE3. No entanto, em 4 dos 5 ativos selecionados observou-se mais da metade das avaliações acima do *benchmark* (#1). Por fim, assim como reportado no cenário (i), mais da metade dos patrimônios líquidos finais dos ativos permaneceram acima do desempenho atingindo por uma estratégia passiva de replicação do Ibovespa.

Tabela 5.3: Patrimônios Líquidos Finais no Conjunto de Teste - Cenário (ii)

Métrica	B3SA3	EQTL3	ITUB4	PETRA4	VALE3
Mínimo nas 100 Iterações (R\$)	R\$ 83870.50	R\$ 84090.14	R\$ 84066.89	R\$ 90985.82	R\$ 81857.29
Máximo nas 100 Iterações (R\$)	R\$ 152703.23	R\$ 122149.61	R\$ 123134.42	R\$ 133408.64	R\$ 134583.33
25º Percentil nas 100 Iterações (R\$)	R\$ 103397.83	R\$ 100989.14	R\$ 91648.12	R\$ 102160.69	R\$ 91079.08
50º Percentil nas 100 Iterações (R\$)	R\$ 112576.31	R\$ 102669.61	R\$ 98005.01	R\$ 105921.55	R\$ 100302.70
75º Percentil nas 100 Iterações (R\$)	R\$ 120342.70	R\$ 104733.69	R\$ 103429.9	R\$ 113995.56	R\$ 109810.53
(#1) Acima do Investimento Inicial	79%	83%	37%	83%	51%
(#2) Acima do <i>Buy &amp; Hold</i> da Empresa	69%	18%	18%	12%	54%
(#3) Acima do <i>Buy &amp; Hold</i> do BOVA11	85%	88%	57%	83%	62%

No cenário (ii), as medianas dos patrimônios finais reportados também apontaram uma estabilidade no desempenho do modelo contextual em relação ao investimento inicial disponibilizado para a compra e venda dos ativos. Além disso, os modelos contextuais treinados de forma individualizada também mostraram-se superiores à média do mercado, que é representada pelo ETF do Ibovespa.

## 5.4 Cenário (iii): Todos os Dados e Recompensa por Índice Sharpe

No terceiro cenário avaliado, uma nova função de recompensa é atribuída ao agente de DRL, em que agora o risco associado ao investimento é considerado no momento da troca de recursos entre a carteira e o número de ações compradas. O cenário (iii) considera todas as modalidades de dados disponíveis.

A Tabela 5.4 expõe os resultados de cada um dos ativos no terceiro cenário de avaliação, no qual o modelo contextual utiliza todas as modalidades de dados coletados e o cálculo da recompensa do agente de DRL envolve o lucro ajustado ao risco em relação ao preço de aquisição do ativo. Em 4 dos 5 ativos selecionados observou-se mais da metade das avaliações acima do *benchmark* (#1). No entanto, apenas um dos ativos superou o *benchmark* de *Buy & Hold* da empresa específica: B3SA3. Por fim, os patrimônios líquidos finais de todos os ativos superaram em mais da metade das avaliações a estratégia passiva de replicação do Ibovespa, representada pelo *benchmark* (#3).



Tabela 5.4: Patrimônios Líquidos Finais no Conjunto de Teste - Cenário (iii)

Métrica	B3SA3	EQTL3	ITUB4	PETR4	VALE3
Mínimo nas 100 Iterações (R\$)	R\$ 81572.43	R\$ 91918.71	R\$ 90439.88	R\$ 90785.97	R\$ 84805.29
Máximo nas 100 Iterações (R\$)	R\$ 164111.79	R\$ 125097.52	R\$ 117318.99	R\$ 149043.77	R\$ 117418.20
25º Percentil nas 100 Iterações (R\$)	R\$ 103580.93	R\$ 100888.67	R\$ 96973.08	R\$ 111024.63	R\$ 92026.92
50º Percentil nas 100 Iterações (R\$)	R\$ 113346.17	R\$ 103276.20	R\$ 99421.95	R\$ 116831.72	R\$ 98424.83
75º Percentil nas 100 Iterações (R\$)	R\$ 123349.25	R\$ 106291.42	R\$ 102595.44	R\$ 125855.69	R\$ 101596.54
(#1) Acima do Investimento Inicial	87%	78%	46%	98%	34%
(#2) Acima do <i>Buy &amp; Hold</i> da Empresa	69%	24%	7%	33%	41%
(#3) Acima do <i>Buy &amp; Hold</i> do BOVA11	91%	86%	81%	98%	61%

O terceiro cenário avaliado apresentou todas as medianas dos patrimônios líquidos finais superiores aos dois cenários anteriores, com exceção da mediana reportada para o ativo VALE3 no cenário (ii). No modelo treinado para PETR4, a mediana foi superior por uma diferença monetária de mais de R\$ 10.000. Tal resultado indica uma preferência em adotar o índice Sharpe como função de recompensa para os agentes de DRL.

## 5.5 Cenário (iv): Dados OHLCV e Recompensa por Índice Sharpe

Esta seção apresenta os resultados obtidos pelo modelo contextual no último cenário de avaliação, em que é mantida a função de recompensa por lucro ajustado ao risco, mas é limitada a quantidade de dados que os agentes de DRL possuem para o treinamento da sua política de negociação.

Na Tabela 5.5 são apresentados os resultados de cada um dos ativos neste último cenário, no qual o modelo contextual com agente de DRL utiliza somente os dados OHLCV e o cálculo da recompensa envolve o lucro ajustado ao risco em relação ao preço de aquisição do ativo. Em 4 dos 5 ativos selecionados observaram-se mais da metade das avaliações acima do montante inicial de R\$ 100.000 investidos: EQTL3 e PETR4. No entanto, nenhum dos ativos superou o *benchmark* (#2). Por fim, todos os modelos apresentaram resultados finais de avaliação superiores ao *benchmark* (#3) e mais da metade das avaliações.

Tabela 5.5: Patrimônios Líquidos Finais no Conjunto de Teste - Cenário (iv)

Métrica	B3SA3	EQTL3	ITUB4	PETR4	VALE3
Mínimo nas 100 Iterações (R\$)	R\$ 83590.33	R\$ 91952.82	R\$ 89398.16	R\$ 101566.34	R\$ 84254.66
Máximo nas 100 Iterações (R\$)	R\$ 134401.79	R\$ 121691.33	R\$ 133335.53	R\$ 138513.09	R\$ 127731.64
25º Percentil nas 100 Iterações (R\$)	R\$ 94531.36	R\$ 101364.12	R\$ 98523.52	R\$ 105154.73	R\$ 91335.92
50º Percentil nas 100 Iterações (R\$)	R\$ 100603.45	R\$ 105546.50	R\$ 101057.99	R\$ 114585.99	R\$ 97753.76
75º Percentil nas 100 Iterações (R\$)	R\$ 111675.28	R\$ 110503.86	R\$ 109399.94	R\$ 117799.99	R\$ 102233.39
(#1) Acima do Investimento Inicial	53%	80%	52%	100%	39%
(#2) Acima do <i>Buy &amp; Hold</i> da Empresa	37%	43%	34%	21%	41%
(#3) Acima do <i>Buy &amp; Hold</i> do BOVA11	69%	90%	85%	100%	56%

O último cenário avaliado não superou as medianas reportadas pelo cenário anterior em três ativos: B3SA3, PETR4 e VALE. No entanto, estes são os ativos com maior quantitativo de informações textuais disponíveis, o que indica uma vantagem para o cenário (iii) em relação ao cenário de avaliação atual.

## 5.6 Análise dos Resultados por Ativo

A seguinte seção tem como objetivo analisar o desempenho do modelo contextual em seu melhor e pior cenário de avaliação, destacando os cenários em destaque pela respectiva performance reportada nas seções anteriores. Ao analisar os resultados finais, o desempenho dos modelos contextuais com agentes de DRL que utilizaram o lucro simples como recompensa foram ligeiramente menores do que os resultados obtidos pelos agentes que utilizaram o índice Sharpe como métrica de avaliação das ações tomadas. Nesse sentido, a busca por estratégias que utilizem informação de risco associado à tomada de decisão é preferível. A negociação individualizada superou, para a maior parte dos ativos, a estratégia passiva de replicação do Ibovespa, o que mostrou ser relevante identificar oportunidades nos momentos de volatilidade do mercado com algoritmos computacionais.

A Figura 5.1 apresenta os patrimônios líquidos finais para os modelos contextuais com agentes de DRL especializados na negociação das ações da Petrobras (representada pelo *ticker* PETR4) em cada cenário de avaliação. Os agentes treinados para a PETR4 obtiveram melhor desempenho geral dentro dos quatro cenários avaliados em comparação com a quantidade de avaliações acima dos *benchmarks* selecionados. A melhor combinação pertence ao cenário (iii), com o uso de todos os dados e função de recompensa do agente de DRL por índice Sharpe. Tendo em vista que PETR4 é um dos ativos mais líquidos da bolsa brasileira e que diversas fontes de notícias publicam ativamente a seu respeito, a adoção de um modelo capaz de lidar com múltiplas fontes de dados é benéfica nesse tipo de ativo. Além disso, os cenários (iii) e (iv), que utilizam função de recompensa com lucro ajustado ao risco mostraram-se muito superiores aos cenários (i) e (ii), que utilizam o lucro simples como função de recompensa.

Por fim, a Figura 5.2 apresenta os patrimônios líquidos finais para as ações do banco Itaú (representada pelo *ticker* ITUB4). Esta configuração obteve o pior desempenho geral nos quatro cenários avaliados. A pior configuração foi relativa ao cenário (i), com o uso de todos os dados e função de recompensa por lucro simples. Para este ativo, os cenários com função de recompensa por índice Sharpe apresentaram maior performance aos cenários com lucro simples. No entanto, a adição dos indicadores técnicos e dados relativos aos sentimentos das publicações não impactou diretamente em melhores resultados na negociação dos ativos do *ticker* ITUB4, sendo que as medianas dos patrimônios

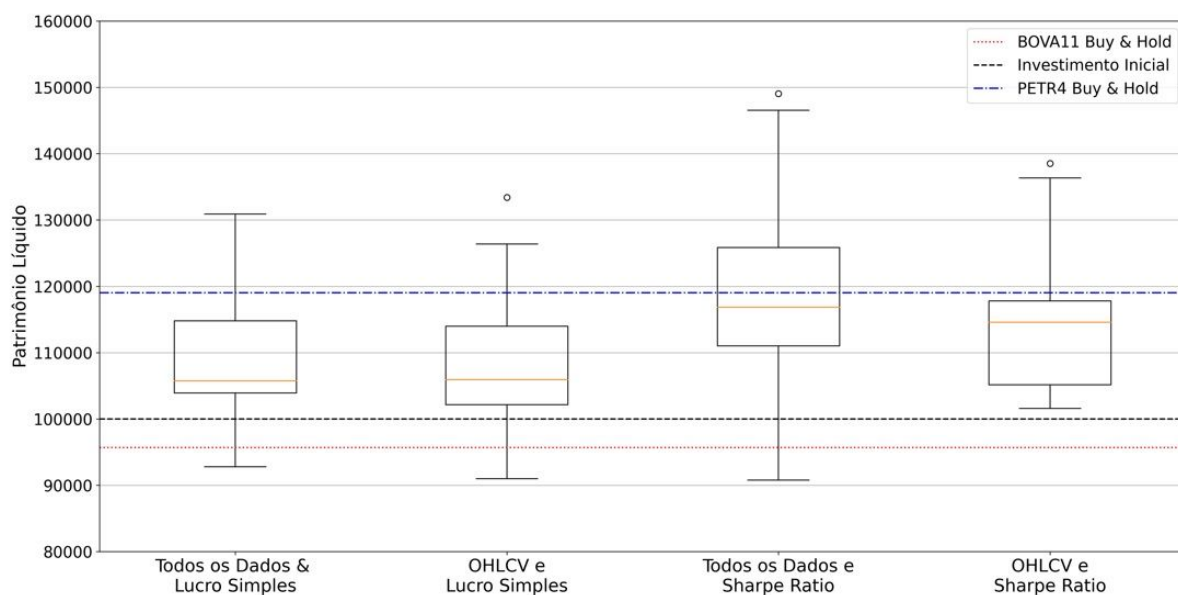


Figura 5.1: Patrimônios líquidos finais no conjunto de teste para o *ticker* PETR4 em cada cenário de avaliação.

líquidos finais nos cenários (i) e (iii) foram ligeiramente menores que os cenários (ii) e (iv), respectivamente.

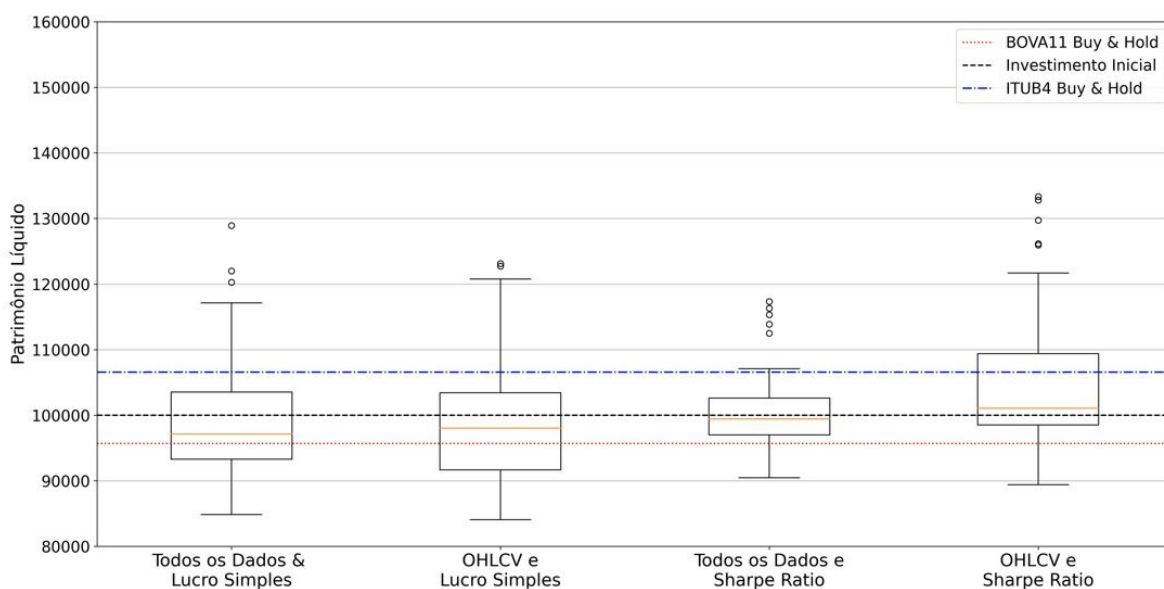


Figura 5.2: Patrimônios líquidos finais no conjunto de teste para o *ticker* ITUB4 em cada cenário de avaliação.

A Tabela 5.6 apresenta o compilado dos patrimônios finais obtidos ao executar o

modelo contextual com agentes de DRL no conjunto de teste, considerando as quatro diferentes combinações entre os esquemas de recompensa e as diferentes fontes de dados. A avaliação teve como objetivo acompanhar a evolução dos patrimônios líquidos das carteiras ao final do período de teste. Para tanto, cada empresa (representada pelo respectivo *Ticker*) foi avaliada de forma individualizada e os resultados foram organizados de acordo com cada cenário de avaliação.

Tabela 5.6: Patrimônios finais agrupados por combinações entre as funções de recompensa e as fontes de dados

Recompensa & Dados	Ticker	Mínimo	Máximo	Percentil	Percentil	Percentil	Investimento Inicial	<i>Buy &amp; Hold</i>
		(R\$)	(R\$)	(25°)	(50°)	(75°)	(#)	(#)
Lucro Simples & Todas as Fontes	B3SA3	R\$ 83645.29	R\$ 139130.17	R\$ 95921.07	R\$ 100946.85	R\$ 109384.10	54%	30%
	EQTL3	R\$ 84058.74	R\$ 111566.02	R\$ 93510.71	R\$ 100000.00	R\$ 101918.71	38%	13%
	ITUB4	R\$ 84843.17	R\$ 128921.04	R\$ 93292.84	R\$ 97132.41	R\$ 103512.80	35%	14%
	PETR4	R\$ 92777.03	R\$ 130898.36	R\$ 103922.82	R\$ 105742.90	R\$ 114787.83	97%	12%
	VALE3	R\$ 81515.75	R\$ 140267.79	R\$ 84876.99	R\$ 98216.93	R\$ 102990.24	37%	45%
Lucro Simples & OHLCV	B3SA3	R\$ 83870.50	R\$ 152703.23	R\$ 103397.83	R\$ 112576.31	R\$ 120342.70	79%	69%
	EQTL3	R\$ 84090.14	R\$ 122149.61	R\$ 100989.14	R\$ 102669.61	R\$ 104733.69	83%	18%
	ITUB4	R\$ 84066.89	R\$ 123134.42	R\$ 91648.12	R\$ 98005.01	R\$ 103429.96	37%	18%
	PETR4	R\$ 90985.82	R\$ 133408.64	R\$ 102160.69	R\$ 105921.55	R\$ 113995.56	83%	12%
	VALE3	R\$ 81857.29	R\$ 134583.33	R\$ 91079.08	R\$ 100302.70	R\$ 109810.53	51%	54%
Índice Sharpe & Todas as Fontes	B3SA3	R\$ 81572.43	R\$ 164111.79	R\$ 103580.93	R\$ 113346.17	R\$ 123349.25	87%	69%
	EQTL3	R\$ 91918.71	R\$ 125097.52	R\$ 100888.67	R\$ 103276.20	R\$ 106291.42	78%	24%
	ITUB4	R\$ 90439.88	R\$ 117318.99	R\$ 96973.08	R\$ 99421.95	R\$ 102595.44	46%	7%
	PETR4	R\$ 90785.97	R\$ 149043.77	R\$ 111024.63	R\$ 116831.72	R\$ 125855.69	98%	33%
	VALE3	R\$ 84805.29	R\$ 117418.20	R\$ 92026.92	R\$ 98424.83	R\$ 101596.54	34%	41%
Índice Sharpe & OHLCV	B3SA3	R\$ 83590.33	R\$ 134401.79	R\$ 94531.36	R\$ 100603.45	R\$ 111675.28	53%	37%
	EQTL3	R\$ 91952.82	R\$ 121691.33	R\$ 101364.12	R\$ 105546.50	R\$ 110503.86	80%	43%
	ITUB4	R\$ 89398.16	R\$ 133335.53	R\$ 98523.52	R\$ 101057.99	R\$ 109399.94	52%	34%
	PETR4	R\$ 101566.34	R\$ 138513.09	R\$ 105154.73	R\$ 114585.99	R\$ 117799.99	100%	21%
	VALE3	R\$ 84254.66	R\$ 127731.64	R\$ 91335.92	R\$ 97753.76	R\$ 102233.39	39%	41%

## 5.7 Considerações Finais

O modelo contextual desenvolvido foi avaliado para diferentes ativos da bolsa de valores brasileira. A avaliação em diferentes cenários possibilitou verificar o desempenho do modelo em situações de total disposição dos dados coletados e também com a limitação dos dados, resumindo-se apenas às barras OHLCV. Além disso, a troca do esquema de recompensa dos agentes de DRL também foi explorada, o que ocasionou em diferentes políticas de investimentos treinadas para os agentes, cada uma com um maior ou menor exposição ao risco associado. Dessa forma, a pequena variabilidade interquartil reportada, bem como a superação dos *benchmarks* selecionados mostrou que o desenvolvimento do modelo contextual é aplicável para investimentos reais.

Nos dois primeiros cenários, em que ambos utilizam o lucro simples como função de recompensa nos agentes de DRL do modelo contextual, o cenário (i) mostrou-se mais vantajoso para as ações de EQTL3 e VALE3. Já o cenário (ii) indicou maior quantitativo de patrimônios finais acima dos *benchmarks* para as ações de B3SA3, ITUB4 e PETR4.

Por fim, ao analisar os dois últimos cenários, em que ambos utilizam o índice Sharpe como função de recompensa nos agentes de DRL do modelo contextual, o cenário (iii) mostrou-se mais vantajoso somente para as ações de B3SA3. Já o cenário (iv) indicou maior quantitativo de patrimônios finais acima dos *benchmarks* para os demais ativos.

Portanto, para os dois esquemas de recompensa selecionados, os cenários com menor quantitativo de informações disponibilizadas obtiveram mais avaliações acima dos *benchmarks* selecionados. No entanto, como visto para o caso das ações de PETR4, que representa o melhor modelo contextual reportado, a mediana das avaliações foi superior para no cenário (iii), sendo este o cenário de avaliação mais completo.

# Capítulo 6

## Conclusão

Os mercados de ações exercem grande influência na economia geral de um país e também nas finanças pessoais dos investidores. Nesse sentido, utilizar mecanismos computacionais modernos capazes de auxiliar e otimizar a tomada de decisão com a utilização de grandes quantidades de dados é de extremo benefício aos investidores, tendo em vista a grande volatilidade e incerteza atrelada às movimentações dos preços das ações de cada empresa. Com isso, os resultados finais obtidos pelos modelos contextuais com agentes de DRL desenvolvidos mostraram-se promissores, tendo aplicabilidade direta nos investimentos de renda variável ao indicar momentos de compra e venda de cada ativo de forma individualizada.

### 6.1 Contribuições

O desenvolvimento do trabalho produziu as seguintes contribuições:

- O aceite do artigo “*Towards the Categorization of Brazilian Financial Market Headlines*” para o IEEE Latin America 2021<sup>1</sup>.
- A submissão do artigo “*A Deep Reinforcement Learning based Trading Agent for the Brazilian Stock Market*” para o IEEE Latin America 2022;

### 6.2 Trabalhos Futuros

Em possíveis trabalhos futuros pretende-se explorar a comparação dos patrimônios líquidos finais obtidos por diferentes algoritmos de DRL, cuja forma de treinamento é diferente da adotada pelo PPO. Alternativamente, as estratégias determinísticas tradicionais com

---

<sup>1</sup><https://latamt.ieee9.org/index.php/transactions/article/view/5751>

indicadores técnicos podem ser utilizadas para analisar o *trade-off* entre a complexidade computacional exigida pelos agentes de DRL e as regras manualmente fornecidas por essas estratégias. Outro ponto de interesse futuro é avaliação do desempenho do agente de DRL com dados intradiários, isto é, dados de *trades* agregados em menores intervalos de tempo, em tempo real, e sem ambiente de simulação com dados históricos.

Além disso, a construção de um *pipeline* completo de execução em tempo real possibilitaria analisar os resultados de forma mais robusta. A exposição aos riscos de liquidez dos ativos, às taxas reais cobradas pelas instituições financeiras e diferenças entre preço de envio de ordem e preço executado são pontos a se considerar no desenvolvimento de um sistema mais completo de negociação de ações no mercado financeiro. No entanto, os resultados obtidos são promissores e indicam a possibilidade real de obter lucro no mercado financeiro com algoritmos mais sofisticados de DRL.

Por fim, o desenvolvimento do componente executor de ordens, que será responsável por receber os sinais do agente de DRL e preparar a mensagem a ser enviada para a bolsa de valores, contendo informações como a quantidade de ações por contrato e o valor alvo a ser negociado. Esse componente poderá interagir com um Sistema de Gerenciamento de Ordens (do inglês, *Order Management System*) (OMS) especializado na comunicação com a [B]<sup>3</sup>, como o fornecido pela Market Data Cloud<sup>2</sup>. Os dados de envio de ordem serão armazenados em formato JSON, simulando os campos do protocolo FIX<sup>3</sup> e integrados com a API de uma corretora específica.

---

<sup>2</sup><https://www.marketdatacloud.com.br/>

<sup>3</sup><https://www.investopedia.com/terms/f/financial-information-exchange.asp>

# Referências

- [1] Taghian, Mehran, Ahmad Asadi e Reza Safabakhsh: *Learning Financial Asset-Specific Trading Rules via Deep Reinforcement Learning*. Expert Systems with Applications, 195:116523, 2022, ISSN 0957-4174. <https://www.sciencedirect.com/science/article/pii/S0957417422000239>. ix, 9, 32, 35
- [2] Haykin, S. S.: *Neural Networks and Learning Machines*. Pearson Education, 3ª edição, 2009, ISBN 9780131471399. ix, 22
- [3] Harari, Yuval Noah: *Sapiens: A Brief History of Humankind*. Dvir Publishing House Ltd, 2015. 1
- [4] Siegel, Jeremy James: *Stocks for the Long Run: The Definitive Guide to Financial Market Returns & Long-Term Investment Strategies*. McGraw Hill, 2014. 1
- [5] Kissell, Robert L.: *Algorithmic Trading Methods (Second Edition)*. páginas 1–21. Academic Press, second edition edição, 2021, ISBN 978-0-12-815630-8. <https://www.sciencedirect.com/science/article/pii/B9780128156308000016>. 1
- [6] Nobre, Renato A., Khalil C. do Nascimento, Patricia A. Vargas, Alan Demétrius Baria Valejo, Gustavo Pessin, Leandro A. Villas e Geraldo P. Rocha Filho: *Aurora: an autonomous agent-oriented hybrid trading service*. Neural Computing and Applications, Sep 2021, ISSN 1433-3058. <https://doi.org/10.1007/s00521-021-06508-3>. 1
- [7] Cagliero, Luca, Jacopo Fior e Paolo Garza: *Shortlisting machine learning-based stock trading recommendations using candlestick pattern recognition*. Expert Systems with Applications, 216:119493, 2023, ISSN 0957-4174. <https://www.sciencedirect.com/science/article/pii/S095741742202512X>. 1
- [8] Prado, Marcos López de: *The 10 Reasons Most Machine Learning Funds Fail*. Journal of Portfolio Management, 2018. 2
- [9] Polamuri, Subba Rao, Dr. Kudipudi Srinivas e Dr. A. Krishna Mohan: *Multi-model generative adversarial network hybrid prediction algorithm (mmgan-hpa) for stock market prices prediction*. Journal of King Saud University - Computer and Information Sciences, 34(9):7433–7444, 2022, ISSN 1319-1578. <https://www.sciencedirect.com/science/article/pii/S1319157821001683>. 2



- [10] Chun, Dohyun, Hoon Cho e Doojin Ryu: *Discovering the drivers of stock market volatility in a data-rich world*. Journal of International Financial Markets, Institutions and Money, 82:101684, 2023, ISSN 1042-4431. <https://www.sciencedirect.com/science/article/pii/S1042443122001561>. 2
- [11] Nti, Isaac Kofi, Adebayo Felix Adekoya e Benjamin Asubam Weyori: *Predicting Stock Market Price Movement Using Sentiment Analysis: Evidence From Ghana*. Appl. Comput. Syst., 25(1):33–42, may 2020, ISSN 2255-8691. <https://doi.org/10.2478/acss-2020-0004>. 2, 33, 34
- [12] Kinyua, Johnson D., Charles Mutigwe, Daniel J. Cushing e Michael Poggi: *An Analysis of the Impact of President Trump’s Tweets on the DJIA and S&P 500 Using Machine Learning and Sentiment Analysis*. Journal of Behavioral and Experimental Finance, 29:100447, 2021, ISSN 2214-6350. <http://www.sciencedirect.com/science/article/pii/S2214635020303762>. 2, 34
- [13] Barroso, B.C., R.T.N. Cardoso e M.K. Melo: *Performance analysis of the integration between Portfolio Optimization and Technical Analysis strategies in the Brazilian stock market*. Expert Systems with Applications, 186:115687, 2021, ISSN 0957-4174. <https://www.sciencedirect.com/science/article/pii/S0957417421010721>. 2
- [14] Soler-Dominguez, Amparo, Angel A. Juan e Renatas Kizys: *A Survey on Financial Applications of Metaheuristics*. ACM Comput. Surv., 50(1), abril 2017, ISSN 0360-0300. <https://doi.org/10.1145/3054133>. 2
- [15] Kaczorowski, Bianca and Kleina, Mariana and Augusto Mendes Marques, Marcos and de Assis Silva, Wiliam: *Artificial intelligence and the multivariate approach in predictive analysis of the small cap index of the brazilian stock exchange*. IEEE Latin America Transactions, 19(11):1924–1932, Apr. 2021. <https://latamt.ieeeer9.org/index.php/transactions/article/view/4808>. 2
- [16] Prado, Marcos Lopez de: *Advances in Financial Machine Learning*. Wiley Publishing, 1st edição, 2018, ISBN 1119482089. 2
- [17] Xing, Frank Z., E. Cambria e R. Welsch: *Natural language based financial forecasting: A survey*. Artificial Intelligence Review, 50:49–73, 2017. 2
- [18] Xu, Qianwen Ariel, Victor Chang e Chrisina Jayne: *A systematic review of social media-based sentiment analysis: Emerging trends and challenges*. Decision Analytics Journal, 3:100073, 2022, ISSN 2772-6622. <https://www.sciencedirect.com/science/article/pii/S2772662222000273>. 2
- [19] Vidyamurthy, Ganapathy: *Pairs Trading : Quantitative Methods and Analysis / G. Vidyamurthy*. Wiley, janeiro 2004, ISBN 978-0-471-46067-1. 2
- [20] Wu, Muh Cherng, Sheng Yu Lin e Chia Hsin Lin: *An Effective Application of Decision Tree to Stock Trading*. Expert Systems with Applications, 31(2):270–274, 2006, ISSN 0957-4174. <https://www.sciencedirect.com/science/article/pii/S0957417405002137>. 2

- [21] Théate, Thibaut e Damien Ernst: *An Application of Deep Reinforcement Learning to Algorithmic Trading*. Expert Systems with Applications, 173:114632, 2021, ISSN 0957-4174. <https://www.sciencedirect.com/science/article/pii/S0957417421000737>. 2, 29, 31, 34, 35
- [22] Li, Yuming, Pin Ni e Victor Chang: *Application of Deep Reinforcement Learning in Stock Trading Strategies and Stock Forecasting*. Computing, 102(6):1305–1322, Jun 2020, ISSN 1436-5057. <https://doi.org/10.1007/s00607-019-00773-w>. 2
- [23] O. Carosia, Arthur E. de, Guilherme P. Coelho e Ana E. A. da Silva: *The Influence of Tweets and News on the Brazilian Stock Market through Sentiment Analysis*. Em *Proceedings of the 25th Brazillian Symposium on Multimedia and the Web*, página 385–392. Association for Computing Machinery, 2019, ISBN 9781450367639. <https://doi.org/10.1145/3323503.3349564>. 2, 34
- [24] Khan, Wasiat, Usman Malik, Mustansar Ali Ghazanfar, Muhammad Awais Azam, Khaled H. Alyoubi e Ahmed S. Alfakeeh: *Predicting Stock Market Trends Using Machine Learning Algorithms via Public Sentiment and Political Situation Analysis*. Soft Comput., 24(15):11019–11043, aug 2020, ISSN 1432-7643. <https://doi.org/10.1007/s00500-019-04347-y>. 2, 34
- [25] Carta, Salvatore, Anselmo Ferreira, Alessandro Sebastian Podda, Diego Reforgiato Recupero e Antonio Sanna: *Multi-DQN: An ensemble of Deep Q-learning agents for stock market forecasting*, *journal = Expert Systems with Applications*. 164:113820, 2021, ISSN 0957-4174. <https://www.sciencedirect.com/science/article/pii/S0957417420306321>. 2, 29, 31, 34, 35
- [26] Serrano, Will: *Deep Reinforcement Learning with the Random Neural Network*. Engineering Applications of Artificial Intelligence, 110:104751, 2022, ISSN 0952-1976. <https://www.sciencedirect.com/science/article/pii/S0952197622000501>. 2
- [27] Millea, Adrian: *Deep Reinforcement Learning for Trading - A Critical Survey*. Data, 6(11), 2021, ISSN 2306-5729. <https://www.mdpi.com/2306-5729/6/11/119>. 2, 31
- [28] Sutton, Richard S e Andrew G Barto: *Reinforcement Learning: An Introduction*. MIT press, 2018. 3
- [29] Wang, Hao nan, Ning Liu, Yi yun Zhang, Da wei Feng, Feng Huang, Dong sheng Li e Yi ming Zhang: *Deep Reinforcement Learning: A Survey*. Frontiers of Information Technology & Electronic Engineering, 21(12):1726–1744, Dec 2020, ISSN 2095-9230. <https://doi.org/10.1631/FITEE.1900533>. 3, 29
- [30] Murphy, John J.: *Technical Analysis of the Financial Markets*. New York Institute of Finance, Fishkill, N.Y., 1999, ISBN 0735200661 9780735200661. 10
- [31] Wilder, J.W.: *New Concepts in Technical Trading Systems*. Trend Research, 1978, ISBN 9780894590276. <https://books.google.com.br/books?id=WesJAQAAMAAJ>. 12

- [32] Granville, J.E.: *Granville's New Key to Stock Market Profits*. Papamoa Press, 2018, ISBN 9781789126037. <https://books.google.com.br/books?id=21ukDwAAQBAJ>. 13
- [33] Chague, Fernando, Rodrigo De-Losso e Bruno Giovannetti: *Day Trading for a Living?* junho 2020. 14
- [34] Kissell, Robert L.: *Algorithmic Trading Methods: Applications Using Advanced Statistics, Optimization, and Machine Learning Techniques*. Academic Press, 2020, ISBN 978-0-12-815630-8. 14
- [35] Krauss, Christopher: *Statistical Arbitrage Pairs Trading Strategies: Review and Outlook*. *Journal of Economic Surveys*, 31(2):513–545, 2017. 15
- [36] Aggarwal, C. C.: *Machine Learning for Text*. Springer Publishing Company, Incorporated, 1ª edição, 2018, ISBN 3319735306, 9783319735306. 15, 16
- [37] Grishman, R. e Sundheim, B.: *Message Understanding Conference-6: A Brief History*. Em *Proceedings of the 16th Conference on Computational Linguistics - Volume 1*, COLING, páginas 466–471, Stroudsburg, PA, USA, (1996). Association for Computational Linguistics. <https://doi.org/10.3115/992628.992709>. 15
- [38] Jing Li, J., Sun, A., Han, J. e Li, C.: *A survey on deep learning for named entity recognition*. *CoRR*, abs/1812.09449, (2018). 15
- [39] Settles, B.: *Biomedical Named Entity Recognition Using Conditional Random Fields and Rich Feature Sets*. Em *Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and Its Applications*, JNLPBA, páginas 104–107, Stroudsburg, PA, USA, 2004. Association for Computational Linguistics. <http://dl.acm.org/citation.cfm?id=1567594.1567618>. 16
- [40] Li, J., A. Sun, J. Han e C. Li: *A Survey on Deep Learning for Named Entity Recognition*. *IEEE Transactions on Knowledge & Data Engineering*, páginas 1–1, 2020, ISSN 1558-2191. 16
- [41] Wankhade, Mayur, Annavarapu Chandra Sekhara Rao e Chaitanya Kulkarni: *A survey on sentiment analysis methods, applications, and challenges*. *Artificial Intelligence Review*, Feb 2022, ISSN 1573-7462. <https://doi.org/10.1007/s10462-022-10144-1>. 16
- [42] Peterson, Richard L.: *Trading on Sentiment*, capítulo 1, páginas 1–18. John Wiley & Sons, Ltd, 2016, ISBN 9781119219149. <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119219149.ch1>. 16
- [43] Birjali, Marouane, Mohammed Kasri e Abderrahim Beni-Hssane: *A comprehensive survey on sentiment analysis: Approaches, challenges and trends*. *Knowledge-Based Systems*, 226:107134, 2021, ISSN 0950-7051. <https://www.sciencedirect.com/science/article/pii/S095070512100397X>. 16
- [44] George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel e Greta M. Ljung: *Time Series Analysis: Forecasting and Control*. John Wiley and Sons Inc., 2016. 17

- [45] Sarma, S N: *Stock Market Seasonality in an Emerging Market*. Vikalpa, 29(3):35–42, 2004. <https://doi.org/10.1177/0256090920040303>. 17
- [46] Wasiuzzaman, Shaista: *Seasonality in the Saudi stock market: The Hajj effect*. The Quarterly Review of Economics and Finance, 67:273–281, 2018, ISSN 1062-9769. <https://www.sciencedirect.com/science/article/pii/S1062976917302247>. 17
- [47] Li, Fengyun, Huacheng Zhang e Dazhi Zheng: *Seasonality in the cross section of stock returns: Advanced markets versus emerging markets*. Journal of Empirical Finance, 49:263–281, 2018, ISSN 0927-5398. <https://www.sciencedirect.com/science/article/pii/S092753981830077X>. 17
- [48] Mahmoud, Amal e Ammar Mohammed: *A Survey on Deep Learning for Time-Series Forecasting*, páginas 365–392. Springer International Publishing, Cham, 2021, ISBN 978-3-030-59338-4. [https://doi.org/10.1007/978-3-030-59338-4\\_19](https://doi.org/10.1007/978-3-030-59338-4_19). 18
- [49] Lazzeri, Francesca: *Machine Learning for Time Series Forecasting with Python*. Wiley, 2020, ISBN 978-1-119-68236-3. 18
- [50] Leng, Jinsong: *Modelling and analysis on noisy financial time series*. Journal of Computational Chemistry, 02:64–69, 2014. 18
- [51] Solnik, Bruno, Cyril Boucrelle e Yann Le Fur: *International market correlation and volatility*. Financial Analysts Journal, 52(5):17–34, 1996, ISSN 0015198X. <http://www.jstor.org/stable/4479942>. 18
- [52] Moretin, Pedro Alberto e Wilton de Oliveira Bussab: *Estatística Básica*. Saraiva, 9.ed. edição, 2017. 19
- [53] Markowitz, Harry: *Portfolio Selection*. The Journal of Finance, 7(1):77–91, março 1952. <https://www.jstor.org/stable/2975974>. 19
- [54] Kahneman, Daniel e Amos Tversky: *Prospect Theory: An Analysis of Decision under Risk*. Econometrica, 47(2):263–291, 1979, ISSN 00129682, 14680262. <http://www.jstor.org/stable/1914185>. 19
- [55] Sokol-Hessner, Peter e Robb B. Rutledge: *The Psychological and Neural Basis of Loss Aversion*. Current Directions in Psychological Science, 28(1):20–27, 2019. <https://doi.org/10.1177/0963721418806510>. 19
- [56] Sharpe, William F.: *The Sharpe Ratio*. The Journal of Portfolio Management, 21(1):49–58, 1994, ISSN 0095-4918. <https://jpm.pm-research.com/content/21/1/49>. 20
- [57] Russell, Stuart e Peter Norvig: *Artificial Intelligence: A Modern Approach*. Prentice Hall Press, USA, 3rd edição, 2009, ISBN 0136042597. 20
- [58] Dong, Shi, Ping Wang e Khushnood Abbas: *A survey on deep learning and its applications*. Computer Science Review, 40:100379, 2021, ISSN 1574-0137. <https://www.sciencedirect.com/science/article/pii/S1574013721000198>. 21

- [59] Ahmed, M. N., A. S. Toor, K. O’Neil e D. Friedland: *Cognitive Computing and the Future of Health Care Cognitive Computing and the Future of Healthcare: The Cognitive Power of IBM Watson Has the Potential to Transform Global Personalized Medicine*. IEEE Pulse, 8(3):4–9, 2017. 21
- [60] Zhang, Shuai, Lina Yao, Aixin Sun e Yi Tay: *Deep Learning Based Recommender System: A Survey and New Perspectives*. ACM Comput. Surv., 52(1), fevereiro 2019, ISSN 0360-0300. <https://doi.org/10.1145/3285029>. 21
- [61] Mcculloch, W. e Pitts, W.: *A Logical Calculus of Ideas Immanent in Nervous Activity*. Bulletin of Mathematical Biophysics, 5:127–147, (1943). 21
- [62] Rosenblatt, F.: *The perceptron: A probabilistic model for information storage and organization in the brain*. Psychological Review, 65(6):386–408, (1958). <http://dx.doi.org/10.1037/h0042519>. 21
- [63] Rumelhart, D. E., Hinton, G. E. e Williams, R. J.: *Learning Representations by Back-propagating Errors*. Nature, 323(6088):533–536, (1986). 21
- [64] Goodfellow, Ian, Yoshua Bengio e Aaron Courville: *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>. 23
- [65] Hochreiter, S. e Schmidhuber, J.: *Long Short-Term Memory*. Neural Comput., 9(8):1735–1780, (November 1997). <http://dx.doi.org/10.1162/neco.1997.9.8.1735>. 23
- [66] Dulac-Arnold, Gabriel, Daniel Mankowitz e Todd Hester: *Challenges of Real-World Reinforcement Learning*, 2019. 25
- [67] Winder, Phil: *Reinforcement Learning*. O’Reilly Media, Inc., 1st edição, 2020, ISBN 9781098114831. 25
- [68] Mnih, Volodymyr, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra e Martin Riedmiller: *Playing Atari with Deep Reinforcement Learning*. 2013. 25
- [69] Campbell, Murray, A. Joseph Hoane e Feng hsiung Hsu: *Deep Blue*. Artif. Intell., 134(1–2):57–83, janeiro 2002, ISSN 0004-3702. [https://doi.org/10.1016/S0004-3702\(01\)00129-1](https://doi.org/10.1016/S0004-3702(01)00129-1). 25
- [70] Haarnoja, Tuomas, Sehoon Ha, Aurick Zhou, Jie Tan, George Tucker e Sergey Levine: *Learning to Walk via Deep Reinforcement Learning*, 2019. 25
- [71] Kiran, Bangalore Ravi, Ibrahim Sobh, Victor Talpaert, Patrick Mannion, Ahmad A. Al Sallab, Senthil Kumar Yogamani e Patrick Pérez: *Deep Reinforcement Learning for Autonomous Driving: A Survey*. CoRR, abs/2002.00444, 2020. <https://arxiv.org/abs/2002.00444>. 25
- [72] Wu, Xing, Haolei Chen, Jianjia Wang, Luigi Troiano, Vincenzo Loia e Hamido Fujita: *Adaptive Stock Trading Strategies with Deep Reinforcement Learning Methods*. Information Sciences, 538:142–158, 2020, ISSN 0020-0255. <https://www.sciencedirect.com/science/article/pii/S0020025520304692>. 29, 31, 34

- [73] Schulman, John, Filip Wolski, Prafulla Dhariwal, Alec Radford e Oleg Klimov: *Proximal Policy Optimization Algorithms*, 2017. 29, 52
- [74] Hariom Tatsat, Sahil Puri, Brad Lookabaugh: *Machine Learning and Data Science Blueprints for Finance*. O'Reilly Media, Inc., 2020, ISBN 9781492073055. 29
- [75] Liu, Xiao Yang, Hongyang Yang, Qian Chen, Runjia Zhang, Liuqing Yang, Bowen Xiao e Christina Dan Wang: *FinRL: A Deep Reinforcement Learning Library for Automated Stock Trading in Quantitative Finance*, 2020. 29, 52
- [76] Pricope, Tidor-Vlad: *Deep Reinforcement Learning in Quantitative Algorithmic Trading: A Review*. CoRR, abs/2106.00123, 2021. <https://arxiv.org/abs/2106.00123>. 30
- [77] Sun, Shuo, Rundong Wang e Bo An: *Reinforcement Learning for Quantitative Trading*, 2021. <https://arxiv.org/abs/2109.13851>. 31
- [78] Yang, Hongyang, Xiao Yang Liu, Shan Zhong e Anwar Walid: *Deep Reinforcement Learning for Automated Stock Trading: An Ensemble Strategy*. 2020. 31, 32, 35, 52
- [79] Sortino, Frank A. e Lee N. Price: *Performance Measurement in a Downside Risk Framework*. The Journal of Investing, 3(3):59–64, 1994, ISSN 1068-0896. 31
- [80] Lillicrap, Timothy P., Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver e Daan Wierstra: *Continuous control with deep reinforcement learning*, 2019. 31
- [81] Kim, Taewook e Ha Young Kim: *Optimizing the Pairs-Trading Strategy Using Deep Reinforcement Learning with Trading and Stop-Loss Boundaries*. Complexity, 2019:3582516, Nov 2019, ISSN 1076-2787. <https://doi.org/10.1155/2019/3582516>. 32
- [82] Chen, Yu Fu e Szu Hao Huang: *Sentiment-Influenced Trading System Based on Multimodal Deep Reinforcement Learning*. Applied Soft Computing, 112:107788, 2021, ISSN 1568-4946. 32
- [83] Hutto, Clayton J. e Eric Gilbert: *VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text*. Em Adar, Eytan, Paul Resnick, Munmun De Choudhury, Bernie Hogan e Alice H. Oh (editores): *ICWSM*. The AAAI Press, 2014, ISBN 978-1-57735-659-2. <http://dblp.uni-trier.de/db/conf/icwsm/icwsm2014.html#HuttoG14>. 32, 33, 46
- [84] Koratamaddi, Prahlad, Karan Wadhwani, Mridul Gupta e Sriram G. Sanjeevi: *Market Sentiment-aware Deep Reinforcement Learning Approach for Stock Portfolio Allocation*. Engineering Science and Technology, an International Journal, 24(4):848–859, 2021, ISSN 2215-0986. <https://www.sciencedirect.com/science/article/pii/S2215098621000070>. 32, 35, 48

- [85] Lin, Siyu e Peter A. Beling: *An End-to-End Optimal Trade Execution Framework Based on Proximal Policy Optimization*. Em *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI'20*, 2021, ISBN 9780999241165. 33
- [86] Sert, Onur Can, Salih Doruk Şahin, Tansel Özyer e Reda Alhajj: *Analysis and Prediction in Sparse and High Dimensional Text Data: The Case of Dow Jones Stock Market*. *Physica A: Statistical Mechanics and its Applications*, 545:123752, 2020, ISSN 0378-4371. 33, 35
- [87] Zhao, W., G. Zhang, G. Yuan, J. Liu, H. Shan e S. Zhang: *The Study on the Text Classification for Financial News Based on Partial Information*. *IEEE Access*, 8:100426–100437, 2020. 33
- [88] Xing, Frank Z., Lorenzo Malandri, Yue Zhang e Erik Cambria: *Financial Sentiment Analysis: An Investigation into Common Mistakes and Silver Bullets*. páginas 978–987. International Committee on Computational Linguistics, 2020. 33
- [89] Naeem, Muhammad Abubakr, Saqib Farid, Balli Faruk e Syed Jawad Hussain Shahzad: *Can Happiness Predict Future Volatility in Stock Markets?* *Research in International Business and Finance*, 54:101298, 2020, ISSN 0275-5319. 33
- [90] Vargas, Gabriel, Leonardo Silvestre, Luís Rigo Júnior e Helder Rocha: *B3 Stock Price Prediction Using LSTM Neural Networks and Sentiment Analysis*. *IEEE Latin America Transactions*, 20(7):1067–1074, May 2022. <https://latamt.ieee9.org/index.php/transactions/article/view/6236>. 34, 35
- [91] Januário, Brenda A., Arthur Emanuel de O. Carosia, Ana Estela A. da Silva e Guilherme P. Coelho: *Sentiment Analysis Applied to News from the Brazilian Stock Market*. *IEEE Latin America Transactions*, 20(3):512–518, Sep. 2021. <https://latamt.ieee9.org/index.php/transactions/article/view/5977>. 34, 35
- [92] Schmitz, Matheus, Roger Immich, Gustavo Pessin e Geraldo Pereira Rocha Filho: *Towards the Categorization of Brazilian Financial Market Headlines*. *IEEE Latin America Transactions*, 100(XXX), Sep. 2021. <https://latamt.ieee9.org/index.php/transactions/article/view/5751>. 39, 40
- [93] Ratner, Alexander, Braden Hancock, Jared Dunnmon, Frederic Sala, Shreyash Pandey e Christopher Ré: *Training Complex Models with Multi-Task Weak Supervision*, 2018. <https://arxiv.org/abs/1810.02840>. 47
- [94] Moritz, Philipp, Robert Nishihara, Stephanie Wang, Alexey Tumanov, Richard Liaw, Eric Liang, Melih Elibol, Zongheng Yang, William Paul, Michael I. Jordan e Ion Stoica: *Ray: A Distributed Framework for Emerging AI Applications*, 2017. <https://arxiv.org/abs/1712.05889>. 51