

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ**

**BRUNO HIDEO KADOOKA**

**ANÁLISE DA EFICÁCIA DAS ESTRATÉGIAS BASEADAS EM PADRÕES DE  
CANDLESTICKS COM O USO DE MODELOS DE MACHINE LEARNING: UM  
ESTUDO SOBRE A LUCRATIVIDADE E EFETIVIDADE NO MERCADO  
FINANCEIRO**

**CORNÉLIO PROCÓPIO**

**2025**

**BRUNO HIDEO KADOOKA**

**ANÁLISE DA EFICÁCIA DAS ESTRATÉGIAS BASEADAS EM PADRÕES DE  
CANDLESTICKS COM O USO DE MODELOS DE MACHINE LEARNING: UM  
ESTUDO SOBRE A LUCRATIVIDADE E EFETIVIDADE NO MERCADO  
FINANCEIRO**

**Analysis of the Effectiveness of Strategies Based on Candlestick Patterns  
Using Machine Learning Models: A Study on Profitability and Performance  
in the Financial Market**

Trabalho de Conclusão de Curso de Graduação  
apresentado como requisito para obtenção  
do título de Bacharel em Engenharia de  
Computação da Universidade Tecnológica  
Federal do Paraná.

Orientador(a): Prof. Francisco Pereira Junior

**CORNÉLIO PROCÓPIO**

**2025**



[4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

Esta licença permite compartilhamento, remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

**BRUNO HIDEO KADOOKA**

**ANÁLISE DA EFICÁCIA DAS ESTRATÉGIAS BASEADAS EM PADRÕES DE  
CANDLESTICKS COM O USO DE MODELOS DE MACHINE LEARNING: UM  
ESTUDO SOBRE A LUCRATIVIDADE E EFETIVIDADE NO MERCADO  
FINANCEIRO**

Trabalho de Conclusão de Curso de Graduação  
apresentado como requisito para obtenção  
do título de Bacharel em Engenharia de  
Computação da Universidade Tecnológica  
Federal do Paraná.

Data de aprovação: 23 / Junho / 2025

---

Francisco Pereira Junior  
Titulação Mestrado  
Universidade Tecnológica Federal do Paraná

---

Henrique Yoshikazu Shishido  
Titulação Doutorado  
Universidade Tecnológica Federal do Paraná

---

Adriano Rivolli da Silva  
Titulação Doutorado  
Universidade Tecnológica Federal do Paraná

**CORNÉLIO PROCÓPIO**

**2025**

Dedico este trabalho ao meu Deus, Jeová, e à  
minha amada família.

## AGRADECIMENTOS

Primeiramente, gostaria de expressar minha profunda gratidão, respeito, fé e amor a Jeová, meu Deus, que me deu forças nos momentos difíceis e quando parecia não haver saída, além de me conceder muitas bênçãos em cada passo. Agradeço por ouvir minhas orações, seja de súplicas ou preces, e por me atrair novamente ao Senhor. Neste momento, considere meus singelos agradecimentos por tua bondade e amor leal.

Gostaria de agradecer imensamente à minha família, em especial ao meu pai, Sergio Satoru Kadooka, minha mãe, Edna Bernardino Kadooka, meu irmão, Hugo Kenji Kadooka, por estarem sempre dispostos a me apoiar e motivar. Agradeço especialmente ao meu tio, Edmilson Bernardino Gomes, por seus conselhos e inspiração ao longo deste período.

Aos meus amigos, tanto os de longa data quanto os que fiz durante essa jornada, agradeço por estarem ao meu lado, me apoiando e me levantando nos momentos em que minha família não pôde. Também agradeço aos amigos e irmãos da Congregação Rancharia por sua orientação, amor e carinho nos últimos anos.

Expresso minha gratidão ao meu professor e orientador, Francisco Pereira Junior, pela paciência e orientações ao longo deste trabalho, dedicando seu tempo e esforço para me ajudar.

Por último, mas com um lugar especial em meu coração, expresso uma gratidão profunda pelo amor, carinho e alegria que guardo da minha querida avó, Maria Aparecida Azarias Gomes (in memoriam), do meu querido tio, Elson Bernardino Gomes (in memoriam), e do meu grande amigo Lucas Eidy Tamura (in memoriam). Suas memórias e o impacto que tiveram em minha vida são tesouros que jamais esquecerei. Eles foram fontes inesgotáveis de apoio, orientação e inspiração, e sempre serão lembrados com um carinho imenso.

*Se não puder voar, corra. Se não puder correr, ande. Se não puder andar, rasteje. Mas continue em frente de qualquer jeito (Martin Luther King Jr).*

## RESUMO

Kadooka, Bruno Hideo. Análise da eficácia das estratégias baseadas em padrões de candlesticks com o uso de modelos de machine learning: Um estudo sobre a lucratividade e efetividade no mercado financeiro. 2025. 68 f. Trabalho de Conclusão de Curso, Engenharia de Computação, Universidade Tecnológica Federal do Paraná. Cornélio Procópio, 2025.

No mercado de investimentos brasileiro, observa-se um crescimento expressivo impulsionado pelo interesse cada vez maior dos investidores. Nesse contexto, a análise técnica é amplamente utilizada por aqueles que buscam oportunidades de curto prazo. Estratégias baseadas em padrões de *candlesticks* e em indicadores clássicos, como suportes, resistências e tendências, são comumente aplicadas, embora existam questionamentos quanto à sua eficácia e lucratividade, especialmente em ativos como contratos futuros e ações. Este trabalho teve como objetivo investigar a efetividade dessas estratégias, tanto de forma isolada quanto em combinação com indicadores técnicos tradicionais, integrando também modelos de *machine learning*. Foi conduzido um estudo abrangente sobre o gerenciamento de operações no *day trade*, com foco na identificação dos intervalos de tempo gráfico mais favoráveis. Entre os resultados, destaca-se a estratégia com 500 pontos de *stop gain* e *stop loss*, aliada ao modelo *Random Forest*, que apresentou desempenho eficaz e lucro consistente no contexto do *day trade*. Os achados contribuem para a avaliação crítica das estratégias tradicionais e demonstram o potencial de otimização por meio de abordagens computacionais.

**Palavras-chave:** Análise Técnica. Padrões de *Candlestick*. *Machine Learning*. *Day Trade*.

## ABSTRACT

Kadooka, Bruno Hideo. Analysis of the Effectiveness of Strategies Based on Candlestick Patterns Using Machine Learning Models: A Study on Profitability and Performance in the Financial Market. 2025. 68 f. Trabalho de Conclusão de Curso, Engenharia de Computação, Universidade Tecnológica Federal do Paraná. Cornélio Procópio, 2025.

The Brazilian investment market has shown significant growth, driven by the increasing interest of investors. In this context, technical analysis is widely used by those seeking short-term opportunities. Strategies based on candlestick patterns and classical indicators, such as support, resistance, and trends, are commonly applied, although their effectiveness and profitability are often questioned, particularly in assets such as futures contracts and stocks. This study aimed to evaluate the effectiveness of these strategies, both individually and in combination with traditional technical indicators, integrating machine learning models. A comprehensive analysis of trade management in the context of day trading was conducted, focusing on identifying the most favorable timeframes for achieving profitable results. Among the findings, the strategy using 500-point stop gain and stop loss parameters combined with the Random Forest model demonstrated strong performance and consistent profitability in day trading. The results contribute to a critical evaluation of traditional strategies and highlight the potential for optimization through computational approaches.

**Keywords:** Technical Analysis, Candlestick Patterns, Machine Learning, Day Trading.

## LISTA DE FIGURAS

Figura 1 – Resistências e Suportes . . . . .	5
Figura 2 – Tendência de Alta . . . . .	6
Figura 3 – Tendência de Baixa . . . . .	7
Figura 4 – Tendência Lateral . . . . .	7
Figura 5 – Gráfico de Barras . . . . .	8
Figura 6 – Estrutura de um candlestick . . . . .	9
Figura 7 – Estrutura do padrão Engolfo . . . . .	9
Figura 8 – Variações do padrão Martelo. . . . .	10
Figura 9 – Estrutura do padrão Harami . . . . .	11
Figura 10 – Código de Vencimento dos meses pares . . . . .	14
Figura 11 – Estrutura de um ticker de contrato futuro . . . . .	14
Figura 12 – Metodologia CRISP-DM. . . . .	16
Figura 13 – Visão geral da plataforma TRYD . . . . .	21
Figura 14 – Importação do <i>dataset</i> de <i>candles</i> . . . . .	25
Figura 15 – Cálculo das médias móveis e deslocadas do <i>Hi-Lo Activator</i> . . . . .	26
Figura 16 – Tendências identificadas pelo <i>Hi-Lo Activator</i> . . . . .	27
Figura 17 – Indicador <i>Hi-Lo Activator</i> em uso. . . . .	27
Figura 18 – Análise gráfica da variável "Tamanho_Candle". . . . .	30
Figura 19 – Estatística descritiva da variável "Tamanho_Candle". . . . .	31
Figura 20 – Modelo de <i>dashboard</i> das operações da estratégia <i>gain</i> e <i>loss</i> de 200 pontos. . . . .	32
Figura 21 – Resumo financeiro das estratégias no <i>dataset</i> de treino. . . . .	32
Figura 22 – Evolução da Carteira - Estratégia 5:1 nos dados de treino. . . . .	33
Figura 23 – Estatística descritiva da Carteira 5:1 nos dados de treino. . . . .	33
Figura 24 – Evolução da Carteira - Estratégia 5:1 nos dados de treino e de teste. . . . .	34
Figura 25 – Resumo financeiro das estratégias no <i>dataset</i> de teste. . . . .	34
Figura 26 – <i>Information Value</i> das variáveis categóricas na estratégia <i>gain</i> e <i>gain</i> de 500 pontos. . . . .	36
Figura 27 – Matriz de correlação variáveis numéricas na estratégia <i>gain</i> e <i>gain</i> de 500 pontos. . . . .	37
Figura 28 – Dataframe de <i>Outliers</i> na estratégia <i>gain</i> e <i>gain</i> de 500 pontos. . . . .	38
Figura 29 – Conceito da Validação Cruzada. . . . .	40
Figura 30 – Validação Cruzada com separação <i>K-Fold</i> . . . . .	41
Figura 31 – Validação Cruzada com separação <i>Stratified K-Fold</i> . . . . .	41
Figura 32 – <i>Rank</i> dos modelos de classificação na estratégia <i>gain</i> e <i>loss</i> de 500 pontos. . . . .	42
Figura 33 – Curvas <i>ROC</i> dos modelos de classificação na estratégia <i>gain</i> e <i>loss</i> de 500 pontos. . . . .	45
Figura 34 – Matriz de Confusão nos dados de teste na estratégia <i>gain</i> e <i>loss</i> de 500 pontos. . . . .	45
Figura 35 – Comparação de operações sem modelo e com modelo ao longo de 2023 na estratégia <i>gain</i> e <i>loss</i> de 500 pontos. . . . .	47
Figura 36 – Curvas <i>ROC</i> dos modelos de classificação na estratégia <i>gain</i> e <i>loss</i> de 200 pontos. . . . .	49
Figura 37 – Matriz de Confusão nos dados de teste na estratégia <i>gain</i> e <i>loss</i> de 200 pontos. . . . .	50

Figura 38 – Comparação de operações sem modelo e com modelo ao longo de 2023 na estratégia <i>gain</i> e <i>loss</i> de 200 pontos. . . . .	51
Figura 39 – Curvas <i>ROC</i> dos modelos de classificação na estratégia <i>gain</i> de 200 pontos e <i>loss</i> de 100 pontos. . . . .	52
Figura 40 – Matriz de Confusão nos dados de teste na estratégia <i>gain</i> de 200 pontos e <i>loss</i> de 100 pontos. . . . .	52
Figura 41 – Curvas <i>ROC</i> dos modelos de classificação na estratégia <i>gain</i> de 500 pontos e <i>loss</i> de 250 pontos. . . . .	54
Figura 42 – Matriz de Confusão nos dados de teste na estratégia <i>gain</i> de 500 pontos e <i>loss</i> de 250 pontos. . . . .	54
Figura 43 – Comparação de operações sem modelo e com modelo na estratégia <i>gain</i> de 500 pontos e <i>loss</i> de 250 pontos no ano de 2023. . . . .	55
Figura 44 – Curvas <i>ROC</i> dos modelos de classificação na estratégia <i>gain</i> de 500 pontos e <i>loss</i> de 100 pontos. . . . .	56
Figura 45 – Matriz de Confusão nos dados de teste na estratégia <i>gain</i> de 500 pontos e <i>loss</i> de 100 pontos. . . . .	57
Figura 46 – Curvas <i>ROC</i> dos modelos de classificação na estratégia <i>gain</i> de 250 pontos e <i>loss</i> de 500 pontos. . . . .	58
Figura 47 – Matriz de Confusão nos dados de teste na estratégia <i>gain</i> de 250 pontos e <i>loss</i> de 500 pontos. . . . .	58
Figura 48 – Comparação de operações sem modelo e com modelo na estratégia <i>gain</i> de 250 pontos e <i>loss</i> de 500 pontos no ano de 2023. . . . .	60

## LISTA DE QUADROS

Quadro 1 – Bibliotecas padrão de análise de dados . . . . .	22
---	----

## LISTA DE TABELAS

Tabela 1 – Posição dos ativos por volume financeiro médio mensal ao longo do ano (em bilhões de reais). . . . .	23
Tabela 2 – Quantidade de candles e padrões de candle e sua representatividade. . . .	29
Tabela 3 – Estratégias das operações e os 3 melhores modelos raqueados . . . . .	43
Tabela 4 – Resumo dos modelos raqueados . . . . .	43
Tabela 5 – Resumo dos modelos usados na estratégia <i>gain</i> e <i>loss</i> de 500 pontos . . . .	46
Tabela 6 – Resumo dos modelos usados com investimento inicial de R\$ 1.000,00 na estratégia <i>gain</i> e <i>loss</i> de 500 pontos . . . . .	48
Tabela 7 – Resumo dos modelos usados na estratégia <i>gain</i> e <i>loss</i> de 200 pontos . . . .	50
Tabela 8 – Resumo dos modelos usados na estratégia <i>gain</i> de 200 pontos e <i>loss</i> de 100 pontos . . . . .	53
Tabela 9 – Resumo dos modelos usados na estratégia <i>gain</i> de 500 pontos e <i>loss</i> de 250 pontos . . . . .	55
Tabela 10 – Resumo dos modelos usados na estratégia <i>gain</i> de 500 pontos e <i>loss</i> de 100 pontos . . . . .	57
Tabela 11 – Resumo dos modelos usados na estratégia <i>gain</i> de 250 pontos e <i>loss</i> de 500 pontos . . . . .	59
Tabela 12 – Resumo geral das estratégias com seus melhores modelos e resultados . . . .	61

## LISTA DE ABREVIATURAS E SIGLAS

ANBIMA	Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais
ARIMA	<i>AutoRegressive Integrated Moving Average</i>
AUC	<i>Area Under the Curve</i>
B3	Brasil, Bolsa, Balcão
BBDC4	Banco Bradesco ações preferenciais
BOVA11	Fundo de índice que replica o índice Ibovespa
CRISP-DM	<i>Cross-Industry Standard Process for Data Mining</i>
CSV	<i>Comma Separated Values</i>
DDE	<i>Dynamic Data Exchange</i>
DOL	Contrato Futuro de Dólar
FGV	Fundação Getulio Vargas
GRU	<i>Gated Recurrent Unit</i>
Hi-Lo	High-Low
IBM	<i>International Business Machines</i>
IND	Índice
IQR	Intervalo Interquartil
ITUB4	Itaú Unibanco ações preferenciais
IV	<i>Information Value</i>
KNN	<i>K-Nearest Neighbors</i>
MACD	<i>Moving Average Convergence Divergence</i>
ON	Ações Ordinárias
PETR4	Petrobras ações preferenciais
PN	Ações Preferenciais
ROC	<i>Rate of Change</i>
RSI	<i>Relative Strength Index</i>
RTD	<i>Real Time Data</i>
SARIMA	<i>Seasonal AutoRegressive Integrated Moving Average</i>

SMOTE	<i>Synthetic Minority Over-sampling Technique</i>
SVC	<i>Support Vector Classifier</i>
USIM5	Usiminas ações preferenciais
VALE3	Vale ações ordinárias
VWAP	<i>Volume Weighted Average Price</i>
WDO	Contrato Futuro de Dólar Mini
WIN	Contrato Futuro de Índice Mini
WOE	<i>Weight of Evidence</i>

# SUMÁRIO

<b>1 – INTRODUÇÃO</b>	<b>1</b>
1.1 PROBLEMATIZAÇÃO	1
1.2 JUSTIFICATIVA	2
1.3 OBJETIVOS	2
1.3.1 Objetivo Geral	3
1.3.2 Objetivos Específicos	3
1.4 ORGANIZAÇÃO DO TRABALHO	3
<b>2 – REVISÃO DE LITERATURA</b>	<b>4</b>
2.1 ANÁLISE TÉCNICA	4
2.1.1 Teoria de Dow	4
2.1.2 Suporte e Resistências	5
2.1.3 Tendências	6
2.1.4 Candlesticks	8
2.1.4.1 Engolfo	8
2.1.4.2 Martelo	10
2.1.4.3 Harami	11
2.2 ATIVOS FINANCEIROS	12
2.2.1 Ações	12
2.2.2 Mercado Futuro	12
2.2.2.1 Índice e Mini Índice Bovespa	13
2.2.2.2 Dólar e Mini Dólar	15
2.3 CRISP-DM	15
2.3.1 Entendimento de Negócios	16
2.3.2 Entendimento dos Dados	16
2.3.3 Preparação de Dados	17
2.3.4 Modelagem	17
2.3.5 Avaliação	18
2.3.6 Implementação	18
2.4 TRABALHOS RELACIONADOS	18
<b>3 – DESENVOLVIMENTO</b>	<b>20</b>
3.1 TECNOLOGIA E FERRAMENTAS	20
3.1.1 Plataforma TRYD	20
3.1.2 Click Up	21
3.1.3 Bibliotecas e Ferramentas Python	21
3.2 ENTENDIMENTO DO NEGÓCIO	22
3.3 ENTENDIMENTO DOS DADOS	24
3.3.1 <i>Dataset de Candlesticks</i>	24
3.3.2 <i>Dataset de Operações</i>	29
3.4 PREPARAÇÃO DOS DADOS	34
3.5 MODELAGEM	38
3.6 AVALIAÇÃO DOS MODELOS	43
<b>4 – ANÁLISE E DISCUSSÃO DOS RESULTADOS</b>	<b>44</b>

4.1	GAIN: 500 PONTOS - LOSS: 500 PONTOS . . . . .	44
4.2	GAIN: 200 PONTOS - LOSS: 200 PONTOS . . . . .	48
4.3	GAIN: 200 PONTOS - LOSS: 100 PONTOS . . . . .	51
4.4	GAIN: 500 PONTOS - LOSS: 250 PONTOS . . . . .	53
4.5	GAIN: 500 PONTOS - LOSS: 100 PONTOS . . . . .	56
4.6	GAIN: 250 PONTOS - LOSS: 500 PONTOS . . . . .	57
<b>5</b>	<b>– CONCLUSÃO . . . . .</b>	<b>61</b>
5.1	TRABALHOS FUTUROS . . . . .	63
5.2	CONSIDERAÇÕES FINAIS . . . . .	64
	<b>Referências . . . . .</b>	<b>66</b>

# 1 INTRODUÇÃO

Nos últimos anos tem havido um crescente interesse em investimentos por parte dos brasileiros, especialmente na bolsa de valores. De acordo com um relatório da [Ferrari \(2023\)](#), em 2012 havia pouco menos de 600 mil pessoas investindo na bolsa de valores, enquanto em 2021 esse número beirou 5 milhões, e em 2022 se aproximou dos 6 milhões.

Esse aumento significativo demonstra uma tendência de crescimento contínuo para os próximos anos. Vale ressaltar que, apesar desse crescimento, a parcela de investimento na B3 em ações representou apenas 3% do total da população brasileira em 2021, indicando um potencial de expansão muito grande, visto que, investimentos em poupança, embora ainda sejam a maior aplicação dos brasileiros, vêm diminuindo nos últimos anos, conforme os dados da [ANBIMA \(2022\)](#).

Neste tipo de investimento, destacam-se duas abordagens analíticas comumente adotadas: a análise fundamentalista e a análise técnica. Essas abordagens diferem em suas metodologias. Em resumo, a análise fundamentalista concentra-se em analisar os fundamentos da empresa, como sua saúde financeira, níveis de endividamento, crescimento de receita e lucro, entre outros indicadores relevantes. O objetivo é determinar o valor intrínseco ou justo da empresa, levando em consideração uma perspectiva de longo prazo ([PINTO, 2022a](#)).

A análise técnica adota uma abordagem distinta ao utilizar gráficos para identificar padrões relacionados às flutuações de preços. Nesse contexto, o interesse concentra-se exclusivamente no comportamento do mercado, envolvendo a compra e venda de ativos, como ações, visando obter lucro. Ao operar na posição de compra, o objetivo é adquirir ações a um preço baixo e vendê-las posteriormente a um preço mais elevado. Por outro lado, ao operar na posição de venda, busca-se vender ações a um preço alto para recomprá-las posteriormente a um preço mais baixo ([PINTO, 2022a](#)).

Não há um consenso sobre qual análise é melhor ou se uma supera a outra. O mais importante é que cada análise é adequada a um determinado perfil de investidor ([PINTO, 2022a](#)). É por esse motivo que profissões relacionadas, como *traders* que realizam operações de compra e venda na bolsa de valores, agentes de investimento que analisam e estudam a viabilidade de investimentos para seus clientes e gestores de investimento que tomam decisões sobre as carteiras de seus clientes, estão em alta no mercado. Esses profissionais representam diferentes abordagens de análise ou até mesmo utilizam ambas as análises para tomar suas decisões ([Educa Mais Brasil, 2023](#)).

Um dos grandes avanços recentes é o uso da inteligência artificial nas operações de *trading*, também conhecidas como operações da análise técnica. Esse crescimento se deve ao fato de as máquinas auxiliarem os profissionais em suas operações, uma vez que modelos de inteligência artificial bem treinados são capazes de analisar e compreender uma grande quantidade de variáveis que seria impossível para um ser humano. Além disso, uma grande vantagem é que esses modelos não sofrem com viés emocional, podendo executar operações de acordo com a estratégia adotada ([SMARTTBOT, 2023](#)).

## 1.1 PROBLEMATIZAÇÃO

O maior problema relacionado à análise técnica é a operação de *day trade*, que consiste em operações de *trading* realizadas no mesmo dia. Segundo um estudo realizado pela FGV por [Chague e Giovannetti \(2020\)](#), constatou-se que 99,43% dos indivíduos que tentaram viver de renda por meio do *day trade* falharam e acabaram desistindo dessa carreira. A lucratividade

obtida por aqueles que tiveram sucesso também não foi satisfatória.

Esse estudo expõe um problema que geralmente está associado à forma como os investidores interpretam o *day trade*. Muitos não consideram o esforço, o trabalho e os riscos envolvidos nesse tipo de investimento. Uma possível explicação para tantas perdas é que esses investidores encaram as operações como apostas, sem levar em conta todo o gerenciamento necessário por trás de uma operação. Eles podem desconhecer qual estratégia adotar, os riscos envolvidos e a relevância dos indicadores em determinadas situações (CAMPOLIM, 2022).

Outro fator relevante é a proliferação de "vendedores de cursos" ou "gurus de mercado". Esses indivíduos se aproveitam da inexperiência de investidores, promovendo estratégias supostamente infalíveis e empregando gatilhos mentais para atrair mais pessoas a adquirir seus cursos. Infelizmente, muitos compradores acabam sendo enganados, adotando estratégias sem fundamentos e sofrendo prejuízos significativos em suas operações (OLIVEIRA, 2020).

## 1.2 JUSTIFICATIVA

A motivação deste trabalho surge da constatação de que muitos analistas de mercado, especialmente os profissionais mais experientes, desenvolveram e utilizam padrões e indicadores para suas operações no *day trade*. No entanto, há dúvidas sobre a validade dessas estratégias atualmente. Por que tanto tempo e esforço foram investidos na reprodução dessas estratégias? Será que elas são falhas e já não são mais eficazes? É notório que viver exclusivamente do *day trade* é desafiador, devido ao grande número de fracassos observados no estudo de Chague e Giovannetti (2020). Diante disso, é crucial compreender os fatores que contribuem para essas falhas recorrentes. Será que as estratégias simples, como os padrões de *candlesticks*, tendências e outros indicadores amplamente utilizados, são falhas? Será possível determinar ou prever os preços dos ativos? Ou será possível buscar uma capacidade preditiva das tendências de mercado?

Outro fator que impulsiona este estudo é a exploração do potencial da inteligência artificial como ferramenta de suporte para análise e tomada de decisões no *day trade*. Embora o estudo de Chague e Giovannetti (2020) não forneça informações sobre o uso dessa tecnologia pelos *traders* analisados, os avanços contínuos e melhorias nos modelos de inteligência artificial sugerem que sua aplicação possa trazer aprimoramentos nos resultados das operações de *day trade*. Portanto, é relevante investigar se o emprego da inteligência artificial pode contribuir para o aperfeiçoamento das estratégias adotadas no *day trade*, considerando os dados e conhecimentos disponíveis.

Assim, este estudo pretende contribuir para a investigação da análise técnica no contexto do *day trade*, utilizando modelos de *machine learning*. Avaliando se estratégias baseadas em padrões de *candlesticks* e outros indicadores são adequadas para o gerenciamento de *trades*, buscando compreender seu impacto no sucesso ou fracasso das operações. A utilização de modelos de *machine learning* proporciona uma abordagem inovadora e aprofundada, permitindo avaliar a efetividade e relevância dessas estratégias tradicionais no atual cenário de *trading*.

## 1.3 OBJETIVOS

Nesta seção apresentam-se os objetivos que norteiam este estudo, levando em consideração a problematização e as motivações descritas na seção de justificativa. São apresentados o objetivo geral e os objetivos específicos que guiam esta pesquisa.

### 1.3.1 Objetivo Geral

O objetivo deste trabalho é investigar a eficácia das estratégias que envolvem padrões de *candlesticks*, em uma combinação com indicadores tradicionais, como RSI, Estocástico, Hi-Lo e médias móveis, aliadas ao uso de modelos de *machine learning*. Será realizado um estudo sobre o gerenciamento de *trades* no *day trade*, com o intuito de identificar se essas estratégias apresentam lucratividade e avaliar a continuidade da eficácia das estratégias tradicionais.

### 1.3.2 Objetivos Específicos

Os objetivos específicos direcionam a pesquisa, auxiliando na definição das etapas necessárias para alcançar o objetivo geral estabelecido:

- Encontrar um conjunto de dados adequado que atenda às operações de *day trade*;
- Classificar corretamente os padrões de *candlesticks*;
- Classificar corretamente os indicadores de suportes e resistências, e as tendências de alta, baixa e lateralidade;
- Aprender e aplicar estratégias tradicionais utilizando padrões de *candlesticks* isoladamente e combinados com os indicadores mencionados anteriormente;
- Aprofundar o conhecimento nas áreas de investimento, análise de dados e ciência de dados;
- Identificar o modelo de aprendizagem mais adequado para a situação do trabalho, comparando os modelos de classificação.

## 1.4 ORGANIZAÇÃO DO TRABALHO

Este capítulo descreveu o contexto e as motivações, além de estabelecer o objetivo geral e os objetivos específicos que se espera alcançar com o desenvolvimento do trabalho. O [Capítulo 2](#) abordará os principais conceitos da análise técnica, incluindo características fundamentais como suporte e resistência, tendências, gráficos e padrões utilizados. Além disso, discutirá os ativos financeiros, como ações e contratos futuros, e apresentará a metodologia CRISP-DM que será utilizada, bem como revisará trabalhos relacionados e estudos correlatos.

O [Capítulo 3](#) detalhará os procedimentos do trabalho, incluindo a seleção dos ativos utilizados e a justificativa por trás dessas escolhas, além de abordar o período de análise. Também serão discutidas as ferramentas e tecnologias empregadas no trabalho, como foi realizado o trabalho, as estratégias abordadas, quais modelos foram utilizados. Já no [Capítulo 4](#) apresentará as métricas e a visualização da performance dos modelos em prever as operações lucrativas de cada estratégia, qual o melhor modelo de cada estratégia, como se saíram os modelos frente aos dados de treino e teste, como se saíram financeiramente e como é a volatilidade dessas operações. Por fim, no [Capítulo 5](#), será avaliado, no contexto geral, se é possível obter lucratividade com os modelos e se é possível viver desses lucros. Serão apresentados também trabalhos que possam estender as pesquisas. é

## 2 REVISÃO DE LITERATURA

Neste capítulo será abordado os conceitos que envolvem a proposta deste trabalho, bem como a análise técnica em geral e seus componentes, ativos financeiros como ações e contratos futuros, a metodologia que será utilizada, CRISP-DM voltada para ciência de dados e os trabalhos correlatos.

### 2.1 ANÁLISE TÉCNICA

A origem da análise técnica remonta ao Japão do século XVIII, especificamente à Bolsa de Arroz de Dojima, onde comerciantes utilizavam gráficos conhecidos como *candlestick* para analisar o mercado de arroz. Esses gráficos, desenvolvidos por Munehisa Homma, permitiam visualizar padrões de preços e volumes, facilitando a tomada de decisões comerciais. A técnica foi introduzida no Ocidente por Steve Nison, que publicou o livro *Japanese Candlestick Charting Techniques*, popularizando o uso desses gráficos entre os traders ocidentais (NISON, 1991).

Em 1889, Charles Dow iniciou a análise dos movimentos de ações e publicou artigos no *The Wall Street Journal* que, posteriormente, foram compilados e deram origem à Teoria de Dow. Essa teoria propõe que os mercados se movem em tendências e que essas tendências podem ser identificadas e utilizadas na previsão de movimentos futuros dos preços. A Teoria de Dow permanece como um dos pilares da análise técnica moderna (INVESTOPEDIA, 2023).

Atualmente, a análise técnica é amplamente utilizada por *traders* e investidores para especular ou prever o movimento dos preços de um determinado ativo financeiro, apoiados em um período gráfico, volume, tendências e padrões do passado que podem se repetir novamente (PINTO, 2022b).

Os gráficos de análise técnica seguem diferentes períodos de tempo, sendo o mais comum o tempo diário, em que cada elemento do gráfico corresponde a um dia de pregão do mercado. No entanto, também existem tempos maiores, como semanais e mensais, e outros menores, como horas e minutos (LANA, 2016).

#### 2.1.1 Teoria de Dow

A Teoria de Dow é a base da análise técnica, que afirma que o mercado e seus ativos são cíclicos e podem estimar tendências de preços com base no passado. Essa teoria é sustentada por três conceitos básicos, segundo (PINTO, 2022b):

1. **O mercado desconta tudo.** Isso significa que a precificação do ativo já reflete todas as informações negativas ou positivas, e o analista deve se concentrar nessa precificação e não em notícias externas ou balanços da empresa. Além disso, de acordo com (CARVALHO, 2021) certos imprevistos não estão precificados, mas o mercado rapidamente ajusta esses imprevistos nos preços, como ocorreu com a pandemia do COVID-19.
2. **Movimentos de preços baseiam-se em tendências.** Os preços dos ativos não se movem de forma imprevisível, mas possuem três tendências: primária, secundária e terciária. A tendência primária é a mais duradoura e pode durar anos, sendo usada em análises de longo prazo. A tendência secundária geralmente dura meses e é aplicada para análise de médio a longo prazo. A tendência terciária tem uma duração menor em relação às outras e é usada para o curtíssimo ou curto prazo (PINTO, 2022b; CARVALHO, 2021).

3. **Repetição é usada como um padrão de análise.** Compreender os movimentos do mercado e os comportamentos humanos associados a eles é essencial para operar de forma eficiente tanto na compra quanto na venda de ativos. Nesse sentido, Dow apresentou exemplos de três comportamentos na tendência primária: acumulação, participação pública e distribuição. É importante ressaltar que esses comportamentos estão relacionados à psicologia humana e não representam um padrão fixo do mercado (PINTO, 2022b; CARVALHO, 2021).

A Teoria de Dow ainda é utilizada para a variação do preço e volume de negociação. No entanto, os analistas sugerem que essa teoria seja usada como um dos alicerces para análise, sem confiar totalmente nela. A principal crítica voltada para a Teoria de Dow é justamente o reconhecimento dos sinais de tendência. Segundo alguns especialistas, "o atraso desse reconhecimento pode levar a perder 25% do movimento do mercado", o que é muito quando se leva em consideração o curtíssimo, curto e médio prazo de uma negociação (CARVALHO, 2021).

### 2.1.2 Suporte e Resistências

Em qualquer mercado financeiro, ocorre uma batalha entre compradores e vendedores, ambos buscando obter lucro, porém, um dos lados inevitavelmente sairá em desvantagem. Os compradores desejam adquirir o ativo a um preço mais baixo para vendê-lo posteriormente a um preço mais elevado, enquanto os vendedores têm interesse em vender o ativo a um preço mais alto para recomprá-lo a um valor mais baixo. Essa dinâmica pode ser comparada a um cabo de guerra em constante movimento (CLEARCORRETORA, 2020, p.14).

Figura 1 – Resistências e Suportes



Fonte: (CLEARCORRETORA, 2020)

As tendências de preço não seguem uma trajetória linear e podem apresentar três direções distintas. O mercado exibe formações de fundos (pontos de preço mais baixos) e topos (pontos de preço mais altos) .

Os suportes representam fundos nos quais os vendedores não conseguem exercer pressão sobre os compradores. Essas regiões geralmente atuam como pontos de reversão, pois os compradores estabelecem níveis de preço ideais para compra, enquanto os vendedores podem

considerar a realização de lucros. Em essência, os suportes agem como uma base ou um apoio para impedir que o preço caia ainda mais (CLEARCORRETORA, 2020, p.15).

As resistências são os topos nos quais os compradores não conseguem exercer força sobre os vendedores. Novamente, diversas situações podem ocorrer nesses pontos, como os vendedores entrando no mercado acreditando que o preço atingiu seu máximo, ou os compradores realizando lucros. De qualquer forma, o preço apresenta uma resistência ou uma barreira em relação à sua valorização (CLEARCORRETORA, 2020, p.15).

Compreender adequadamente a identificação de suportes e resistências é de extrema importância, uma vez que esses níveis podem definir tendências, sinalizar reversões de preço e auxiliar na elaboração de estratégias de negociação. É relevante destacar que os suportes e resistências costumam apresentar um volume de negociação mais elevado, uma vez que representam áreas de disputa em que uma das partes pode prevalecer. Esses padrões podem ser respeitados ou rompidos, e quando um suporte é quebrado, ele se torna uma resistência, enquanto uma resistência rompida passa a funcionar como um suporte. Essa dinâmica ocorre devido aos operadores testarem essas regiões ou as utilizarem como pontos de *stop* para limitar ganhos ou perdas (CLEARCORRETORA, 2020, p.15).

### 2.1.3 Tendências

Conforme mencionado, o mercado exhibe tendências, sendo que na Teoria de Dow são abordados três tipos de tendências em relação ao tempo: primária, secundária e terciária. Cada uma dessas tendências pode seguir uma direção específica, que pode ser classificada como alta, baixa e lateralização ou consolidação.

Isso indica que o mercado não segue movimentos lineares, mas sim uma sequência de movimentos dentro de uma tendência. O mercado pode avançar em uma direção, corrigir esse movimento e retornar ao movimento anterior, lembrando a forma de um zigue-zague.

A tendência de alta é caracterizada por topos e fundos ascendentes, evidenciando que a força dos compradores está se fortalecendo progressivamente. Em tendências de médio ou curto prazo, as resistências são superadas e novos níveis de resistência são estabelecidos acima dos níveis anteriores, tendências de longo prazo tem poucas alterações, não sendo interessante neste estudo (CLEARCORRETORA, 2020, p.18).

Figura 2 – Tendência de Alta



Fonte: (CLEARCORRETORA, 2020)

Por sua vez, a tendência de baixa é caracterizada por topos e fundos descendentes, indicando que os vendedores estão exercendo pressão para reduzir os preços. Novamente, em

tendências de médio ou curto prazo, é comum a quebra dos suportes anteriores e a formação de novos suportes em níveis ainda mais baixos (CLEARCORRETORA, 2020, p.18).

Figura 3 – Tendência de Baixa



Fonte: (CLEARCORRETORA, 2020)

E, por fim, na tendência lateral, também conhecida como congestão ou caixas devido à forma gráfica que apresenta, os suportes e resistências são respeitados. Nesse caso, os topos e fundos não conseguem impulsionar o mercado em uma direção específica, indicando uma indecisão entre compradores e vendedores e uma falta de tendência clara (CLEARCORRETORA, 2020, p.19).

Figura 4 – Tendência Lateral



Fonte: (CLEARCORRETORA, 2020)

### 2.1.4 Candlesticks

Os investidores costumavam utilizar o gráfico de barras para obter informações sobre os preços dos ativos. Essas barras exibiam dados como o valor máximo, valor de fechamento, valor de abertura e valor mínimo. No entanto, esse tipo de gráfico apresentava dificuldades na visualização dos valores de abertura e fechamento. Esses valores eram representados por barras horizontais, sendo a abertura à esquerda e o fechamento à direita. As barras verticais indicavam os valores máximo e mínimo, seguindo uma estrutura semelhante ao gráfico de velas (GONSALES, 2022).

Figura 5 – Gráfico de Barras



Fonte: (GONSALES, 2022)

Por esse motivo, Steve Nison substituiu os gráficos de barras pelos gráficos de velas, que também têm origem japonesa. O gráfico de velas ou *candlesticks* apresenta as mesmas informações do gráfico de barras, porém de forma ainda mais intuitiva. As barras horizontais são preenchidas, e para indicar a posição da abertura e do fechamento, esse preenchimento é representado por cores branca ou preta. Atualmente, *softwares* como plataformas de negociação utilizam, com frequência, as cores vermelha e azul para essa finalidade. (GONSALES, 2022).

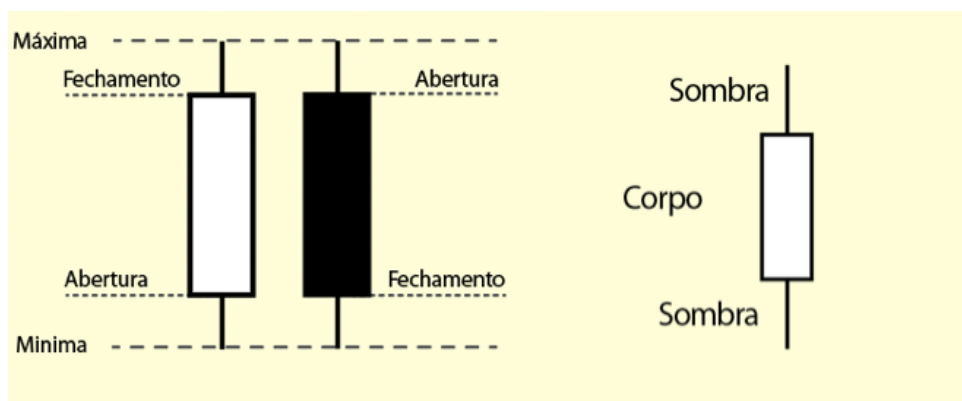
Os outros elementos que compõem o *candle* são a sombra ou pavio superior, que representa o valor máximo, a sombra ou pavio inferior, que representa o valor mínimo, e o corpo, que representa os valores entre abertura e fechamento (GONSALES, 2022). Esses elementos são de suma importância para classificar os *candlesticks* em seus padrões. O próprio Nison (1991) classificou os *candles* com base nas características desses elementos, ou seja, nomeou os padrões de *candles*.

Os *candles* de reversão indicam uma possível mudança de direção do mercado. No entanto, Steve Nison discordava do nome "reversão", pois esse termo sugere que um movimento será abruptamente encerrado e outro movimento oposto começará, o que nem sempre é o caso na prática. Mesmo que o padrão apareça, pode ser que o mercado não defina exatamente a mudança de direção esperada (NISON, 1991, p.27).

#### 2.1.4.1 Engolfo

O primeiro padrão de reversão é o engolfo, que recebe esse nome porque o segundo *candle* "engole" o primeiro, indicando uma possível mudança de direção no mercado. Se o movimento do mercado era de queda e o engolfo aparece, ele recebe o nome de "Engolfo de Alta", pois sugere uma reversão para a alta. Em contrapartida, se o mercado estava em alta e o padrão aparece, ele recebe o nome de "Engolfo de Baixa", pois sugere uma reversão para a baixa.

Figura 6 – Estrutura de um candlestick



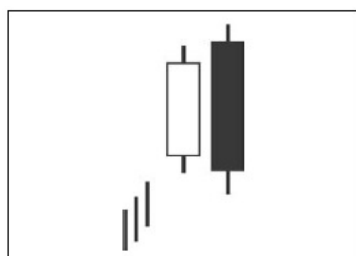
Fonte: (ADVFN, 2022)

Nison (1991, p.38-39) estabelece três critérios para identificar um padrão de engolfo:

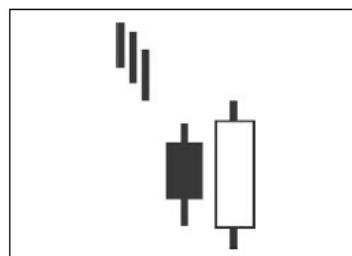
1. O mercado deve apresentar uma tendência clara de alta ou de baixa, pois o engolfo é um padrão de reversão da tendência atual e não funciona bem em mercados laterais ou sem tendência definida;
2. O padrão precisa ter duas velas: o corpo do segundo *candle* deve envolver o corpo inteiro do primeiro *candle*, mas não é necessário que as sombras sejam envolvidas;
3. O segundo *candle* deve ser de cor oposta ao primeiro.

Figura 7 – Estrutura do padrão Engolfo

(a) Engolfo de Baixa



(b) Engolfo de Alta



Fonte: (BULKOWSKI, 2008, p.308, 317)

Além desses critérios, Nison (1991, p.39-40) também sugere outros fatores que podem dar ainda mais força ao padrão de engolfo:

- O primeiro *candle* tende a ser menor, enquanto o segundo *candle* tende a ser maior, e quanto maior a diferença entre os corpos, mais forte é o sinal de reversão;
- O movimento anterior a ser revertido deve estar prolongado ou esticado, o que sugere uma exaustão da tendência atual;
- Maior volume no segundo *candle* em relação ao primeiro;
- Embora o padrão seja formado por apenas duas velas, o segundo *candle* pode envolver dois ou mais *candles* anteriores, o que aumenta ainda mais sua força como sinal de reversão.

Quanto aos pontos de entrada e saída para operar com o padrão de engolfo, Bulkowski (2008, p.314-316) sugere que o ponto de entrada ou confirmação é o ponto máximo para o engolfo de alta e o ponto mínimo para o engolfo de baixa. Já o ponto de proteção ou *stop*

*loss* é o contrário: o ponto mínimo para o engolfo de alta e o ponto máximo para o engolfo de baixa. O alvo da operação pode ser até a chegada de uma resistência ou suporte, ou até encontrar um novo sinal de reversão (BULKOWSKI, 2008, p.314-316, 324-325).

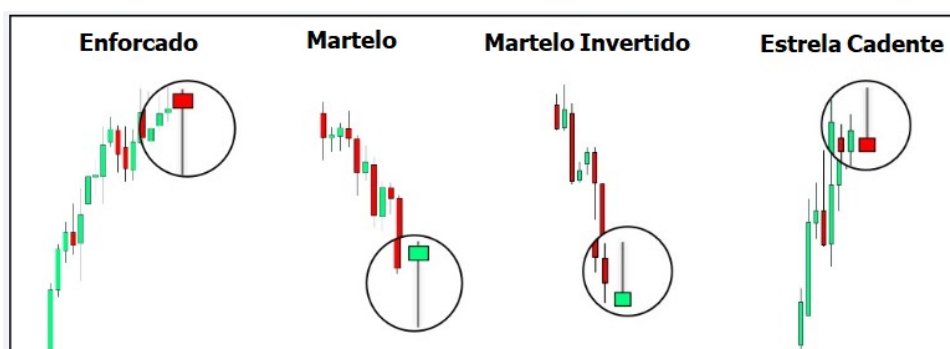
#### 2.1.4.2 Martelo

O martelo apresenta algumas variações em seu formato. Entre as variações mais conhecidas estão o martelo tradicional e o martelo invertido. Quando o padrão está seguindo uma tendência de baixa, é chamado de martelo, enquanto em uma tendência de alta, é chamado de Enforcado. O martelo invertido, é uma versão do martelo de cabeça para baixo, ele também aparece em movimentos de baixa, caso o movimento é de alta, o nome deste martelo invertido é Estrela Cadente.

Assim como em outros padrões, Nison (1991, p.29) cita três características obrigatórias para identificar um martelo:

1. O corpo do *candle* está localizado na parte superior do *candle*, apresentando, portanto, uma sombra superior muito curta ou inexistente. No caso do martelo invertido, o corpo está na parte inferior, com pouca ou nenhuma sombra inferior;
2. A sombra inferior deve ter o dobro da altura do corpo;
3. O padrão é formado por apenas um *candle*, e a cor do corpo não é relevante.

Figura 8 – Variações do padrão Martelo.



Fonte: Traduzido de (AWAIS, 2020)

Embora o padrão do martelo seja flexível em relação às suas características, é importante não perder a sua forma. Portanto, a característica obrigatória é a posição do corpo na parte superior. Quanto à sombra inferior, pode ser que não atinja o dobro do corpo, mas quanto maior essa sombra, mais força o padrão ganhará. Nison também menciona que a cor do corpo não é importante para identificar o padrão, mas em certos momentos pode fortalecer a sua interpretação. Por exemplo, em um movimento de baixa, um martelo com corpo verde pode indicar uma maior força para a reversão, enquanto um corpo vermelho pode sugerir uma indecisão no mercado (NISON, 1991, p.29).

A forma de operar com o padrão do martelo é diferente dos outros padrões. É necessário aguardar a confirmação do movimento por meio de outro *candle*, ou seja, quando um martelo aparece, é preciso que o próximo *candle* tenha um fechamento acima dos pontos máximo ou mínimo do padrão. Dessa forma, o analista pode entrar na operação na abertura do próximo *candle*. Os pontos máximo e mínimo do martelo são utilizados como níveis de proteção. Quanto ao alvo e objetivo da operação, isso dependerá de outras análises, como padrões de topos e fundos ou até mesmo linhas de tendência (BULKOWSKI, 2008, p.352-355).

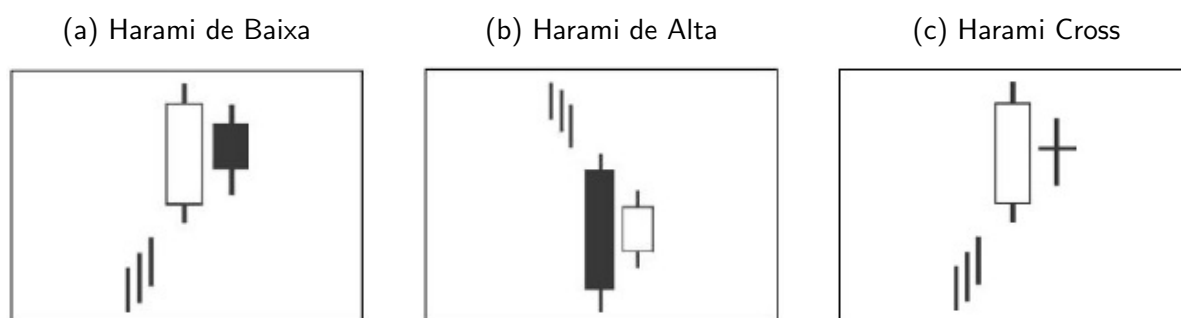
### 2.1.4.3 Harami

O padrão Harami é o inverso do engolfo. Curiosamente, "Harami" pode ser uma palavra japonesa que significa "grávida", e a aparência do padrão realmente lembra esse termo. Ao contrário do engolfo, no Harami é o primeiro *candle* que envolve o segundo *candle*. Quando o padrão ocorre em um movimento de alta, é chamado de Harami de Baixa, e quando ocorre em um movimento de queda, é chamado de Harami de Alta. Além disso, quando o segundo *candle* é um Doji, temos a variação chamada de Harami Cross.

As características do Harami são praticamente opostas às do engolfo. Nison (1991, p.80,81) define as seguintes características:

1. É necessário ter dois *candles*, onde o primeiro *candle* envolve o segundo. Essa é a principal característica que diferencia o Harami do Engolfo;
2. As cores dos *candles* podem ser iguais ou diferentes, não há regras definidas, embora na maioria das vezes as cores sejam opostas;
3. A presença de sombras não interfere no padrão, e o segundo *candle* pode ser um Doji, o que pode fortalecer o padrão.

Figura 9 – Estrutura do padrão Harami



Fonte: (BULKOWSKI, 2008, p.374, 383, 392)

Nison (1991, p.80) reconhece que o padrão Harami tem menos força em comparação aos três padrões mencionados anteriormente. Na verdade, exceto pelo Harami Cross, esse padrão geralmente indica mais o final de um movimento do que uma reversão propriamente dita. O primeiro *candle*, por ter um volume maior, pode indicar um possível sinal de reversão, mas o segundo *candle*, por ser um pouco menor, sugere que o movimento pode ter um breve impulso antes de se reverter, o que o torna um sinal de possível fim do movimento.

No entanto, a versão Harami Cross, em que o segundo *candle* é um Doji, é considerado um forte padrão de reversão, pois o mercado dá um sinal claro de mudança com o primeiro *candle*, e o segundo, por ser um Doji, geralmente indica indecisão ou incerteza em relação ao movimento anterior do mercado. Portanto, pode indicar uma iminente reversão (NISON, 1991, p.81).

A forma de operar o Harami é exatamente a mesma do engolfo. Os pontos de entrada devem ser no *candle* que ultrapassa os pontos máximo ou mínimo do primeiro *candle* do Harami. Às vezes, espera-se que o *candle* feche antes de entrar na abertura do próximo *candle*, assim como é feito no martelo. Os pontos de proteção devem estar próximos aos pontos máximo ou mínimo opostos à entrada. É importante lembrar que quanto mais o primeiro *candle* envolver ou quanto maior o número de *candles* envolvidos, mais forte será o sinal de reversão (BULKOWSKI, 2008, p.380-391).

## 2.2 ATIVOS FINANCEIROS

Um ativo financeiro é um instrumento ou contrato que representa um direito de propriedade sobre um recurso econômico e é negociado em um mercado financeiro. Esses ativos desempenham um papel fundamental no mundo dos investimentos, permitindo aos investidores adquirir e negociar direitos sobre diferentes classes de ativos.

### 2.2.1 Ações

Ações são ativos financeiros que representam uma fração de propriedade em uma empresa ou sociedade anônima. Os detentores dessas ações, conhecidos como acionistas, têm direitos e deveres semelhantes aos demais sócios da empresa. A negociação de ações é aberta a qualquer pessoa física, jurídica ou instituição financeira, realizada por meio de corretoras de valores (WILTGEN, 2019).

As ações podem ser classificadas em três categorias: Ordinárias, Preferenciais e *Units*. As ações Ordinárias conferem direito a voto nas assembleias de acionistas, permitindo que os acionistas participem das decisões da empresa. Já as ações preferenciais possuem vantagens na distribuição de dividendos ou na recuperação de ativos em caso de falência da empresa. As ações *Units* combinam uma quantidade de ações preferenciais e ordinárias em uma única unidade de negociação (WILTGEN, 2019).

Cada categoria de ações é identificada por um código específico. Por exemplo, o código 3 representa ações ON (Ordinárias), enquanto o código 4 representa ações PN (Preferenciais). Os códigos de 5 a 8 também representam ações preferenciais, classificadas de acordo com o valor dos dividendos. Por sua vez, o código 11 é atribuído às ações *Units*. O ticket de ação é composto por quatro letras seguidas por um código identificador. Por exemplo, o ticker PETR4 representa a ação preferencial da empresa Petrobras, enquanto o ticker USIM5 representa a ação preferencial de classe A da empresa Usiminas (B3, 2022a).

As ações são consideradas ativos de renda variável e estão sujeitas à flutuação de preços no mercado financeiro. Por esse motivo, são amplamente utilizadas para realizar análises técnica e fundamentalista. No entanto, investir em ações também envolve riscos, incluindo o risco de mercado, relacionado à volatilidade dos preços, e o risco de liquidez, relacionado à dificuldade de vender ações com baixo volume de negociação (WILTGEN, 2019).

É importante ressaltar que investir em ações implica alto risco, devido à natureza volátil do mercado, o que pode resultar em perda de capital investido. Esse risco, conhecido como risco de mercado, é influenciado por diversos fatores externos, como políticas governamentais, flutuações econômicas e mudanças nas condições de mercado. Além disso, existe o risco de liquidez, uma vez que a liquidação da venda de ações na B3, por exemplo, ocorre em um prazo de dois dias após a transação (B3, 2022a). No entanto, algumas empresas possuem baixo volume de negociação, o que pode dificultar a venda de ações. Portanto, é essencial que os investidores estejam cientes dos riscos envolvidos na negociação de ações e adotem medidas para mitigá-los, como diversificação de investimentos e análise criteriosa das empresas em que desejam investir (WILTGEN, 2019).

### 2.2.2 Mercado Futuro

Este tipo de investimento consiste na compra e venda de contratos que representam *commodities*, índices, metais preciosos e moedas, visando uma perspectiva futura. O investidor pode lucrar tanto com a compra como com a venda desses contratos (TOROINVESTIMENTOS, 2022).

Nas operações de compra, o investidor adquire um contrato atual, esperando que seu valor aumente no futuro ou na data de vencimento. Ao adquirir o contrato por um valor inferior ao valor de venda futuro, a diferença entre os dois valores será o lucro. No entanto, é importante ressaltar que também existem riscos, e prejuízos podem ocorrer caso o valor de compra seja superior ao valor de venda no futuro. Por outro lado, nas operações de venda, o investidor vende um contrato que deverá ser comprado no futuro. Se a venda for realizada por um valor superior ao valor de compra futuro, o investidor terá lucro. Novamente, é importante considerar os riscos, pois prejuízos podem ocorrer se a venda for feita por um valor inferior ao valor de compra futuro (TOROINVESTIMENTOS, 2022).

Uma das principais vantagens de operar contratos futuros é a alta liquidez, permitindo que o investidor encerre suas posições a qualquer momento. Isso torna esse tipo de investimento adequado para operações de curto prazo, como *day trade*, ou de médio prazo, como *swing trade*. Além disso, a facilidade e flexibilidade são atrativos, pois é possível operar tanto na compra como na venda sem a necessidade de possuir o contrato ou recorrer ao aluguel de ativos, como ocorre no caso das ações. A alavancagem é outra vantagem significativa, principalmente em contratos futuros de índices ou moedas, permitindo operar com quantias de dinheiro superiores ao valor total necessário, uma vez que o lucro é obtido a partir da diferença entre os valores dos contratos. Dessa forma, é possível obter lucros potencialmente maiores em comparação com outros tipos de investimento, como no caso das ações (TOROINVESTIMENTOS, 2022).

Por outro lado, é crucial ressaltar que esse tipo de investimento apresenta um risco significativo, que pode ser ainda maior do que o das ações. Enquanto no mercado de ações o pior cenário envolve a falência da empresa investida, levando o valor da ação a zero e resultando na perda total do investimento, nos contratos futuros o risco está associado a possíveis perdas que excedem o valor inicialmente investido, podendo gerar dívidas. Para mitigar esses riscos, as corretoras exigem a chamada Margem de Garantia, que consiste em valores depositados na conta do investidor para cobrir eventuais prejuízos das operações. Caso as perdas ultrapassem o valor da margem, a corretora pode acionar a chamada zeragem compulsória, com o objetivo de proteger o cliente, que, nesse caso, será obrigado a arcar com as perdas da operação (TOROINVESTIMENTOS, 2022).

#### 2.2.2.1 Índice e Mini Índice Bovespa

O índice Bovespa é o termômetro do mercado acionário do Brasil. Esse índice reúne diversas ações de empresas brasileiras, sendo que metade de sua composição está em empresas gigantes do setor bancário e de commodities, como Vale, Petrobras, Banco do Brasil, Banco Bradesco e Banco Itaú. Por esse motivo, ele é o principal indicador de desempenho usado para analisar a situação do mercado em geral e é uma ótima referência para verificar a economia do país e dos investidores. Tornando-se também um ótimo ativo de especulação para o futuro, seja para alta ou queda, cria-se grande expectativa em torno desse índice (B3, 2022b).

Na B3, são negociados dois contratos: o IND e o WIN. Esses contratos possuem algumas características em comum que facilitam sua visualização e cálculos durante as operações. A cotação de ambos os contratos varia em múltiplos de 5, o que proporciona uma fácil compreensão dos valores envolvidos. Além disso, os contratos têm vencimento sempre em meses pares, na quarta-feira mais próxima do dia 15 do mês de vencimento (B3, 2022b; B3, 2022e).

Para identificar cada contrato, é utilizado um código que é adicionado ao ticket do contrato. Esse código é composto pelo código do contrato, seguido pelo mês de vencimento e o ano abreviado do vencimento. Dessa forma, é possível distinguir e acompanhar os diferentes contratos ao longo do tempo (B3, 2022b; B3, 2022e).

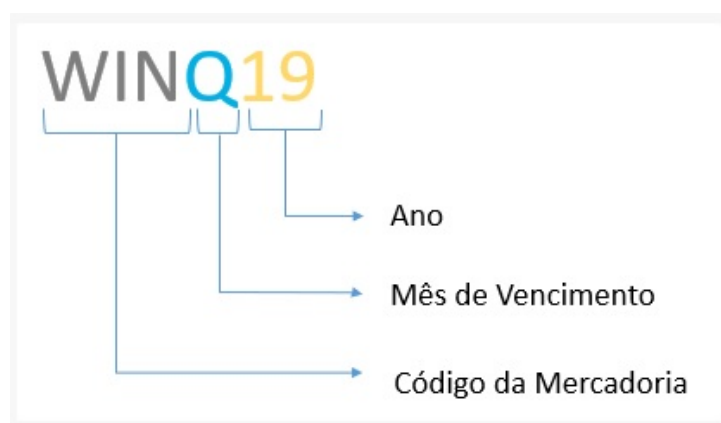
Referente as diferenças temos no contrato IND, conhecido como contrato cheio,

Figura 10 – Código de Vencimento dos meses pares

Vencimento	Letra
Fevereiro	G
Abril	J
Junho	M
Agosto	Q
Outubro	V
Dezembro	Z

Fonte: (B3, 2022c)

Figura 11 – Estrutura de um ticker de contrato futuro



Fonte: (B3, 2022c)

representa o índice Bovespa e possui uma cotação em que cada ponto equivale a 1 real. Além disso, é obrigatório operar em lotes de 5 contratos, resultando em uma flutuação mínima de 25 reais por operação. Essas características tornam o contrato IND mais adequado para pessoas jurídicas ou institucionais, que devem estar preparadas para uma margem de garantia mais elevada (B3, 2022b).

O contrato WIN, por sua vez, foi criado como uma opção ao contrato IND, buscando reduzir riscos e valores envolvidos nas operações. Conhecido como mini índice, ele possui um tamanho menor em relação ao contrato cheio. Cada ponto do contrato WIN tem um valor de 20 centavos, o que resulta em uma flutuação de 1 real (B3, 2022c). Uma diferença significativa entre o contrato WIN e o IND é a ausência de restrição quanto ao lote mínimo no WIN, permitindo que os investidores operem com a quantidade desejada do contrato. Isso possibilita uma flutuação mínima de 1 real, o que pode ser atrativo para diferentes perfis de investidores (B3, 2022b; B3, 2022c).

Além disso, o contrato WIN é acessível a pessoas físicas, jurídicas e institucionais, e não exige uma margem de garantia elevada. Essa característica amplia a participação de investidores individuais e contribui para uma maior liquidez nas operações envolvendo o contrato

WIN. Pessoas físicas também têm a opção de operar o contrato IND, caso desejem (B3, 2022b; B3, 2022c).

#### 2.2.2.2 Dólar e Mini Dólar

Diversas moedas, como o Euro, Yuan, Iene e Bitcoin, funcionam como índices em diferentes mercados. No entanto, o Dólar continua sendo a principal moeda global para transações e serve como referência não apenas para o mercado americano, mas também para outras moedas. Sendo assim, o Dólar é fundamental para medir a economia brasileira, especialmente devido à exposição das exportações a essa moeda (B3, 2022d).

Na B3, assim como no índice Bovespa, são oferecidos dois contratos de Dólar: DOL e WDO. O valor do Dólar Comercial é utilizado como tamanho do contrato, determinado pela taxa de câmbio entre o Real e o Dólar Comercial. Os contratos possuem vencimento mensal, sempre no primeiro dia útil do mês, em contraste com os contratos do Índice Bovespa. A flutuação desse contrato é decimal, com variação de 0,5 pontos. Os tickets desses contratos seguem a mesma lógica dos contratos do Índice Bovespa, com o código do contrato, mês de vencimento e ano abreviado do vencimento. Geralmente, o código de vencimento é igual ao do Índice Bovespa, mas, por ser mensal, pode utilizar códigos diferentes (B3, 2022d; B3, 2022f).

O contrato DOL é o contrato cheio de Dólar, operado em lotes de 5 contratos. Ele segue as mesmas regras do contrato cheio do Índice Bovespa. No entanto, esse contrato requer um investimento maior, o que resulta em lucros e prejuízos mais significativos em comparação ao contrato cheio do Índice Bovespa. Cada ponto do contrato DOL equivale a 50 reais. Com a flutuação mínima de 0,5 pontos e a obrigatoriedade de operar lotes de 5 contratos, a flutuação mínima de valor é de 125 reais. Esse contrato é voltado principalmente para pessoas jurídicas e institucionais, devido aos custos e margens de garantia mais elevados (B3, 2022d).

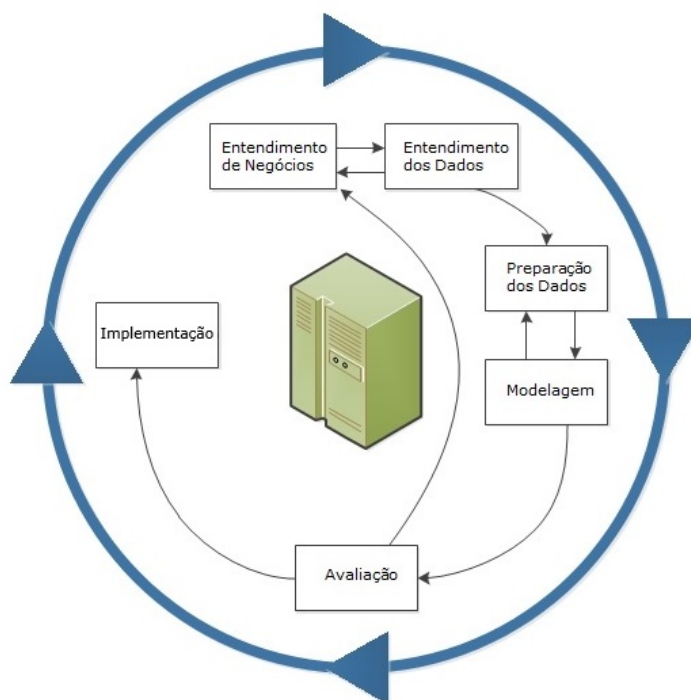
O mini Dólar WDO tem 1/5 do tamanho do contrato cheio DOL. Ele não possui as mesmas restrições do contrato cheio, permitindo que os investidores operem com a quantidade desejada do contrato, sem restrição quanto ao lote. Cada ponto no contrato WDO equivale a 10 reais, resultando em uma flutuação mínima de valor de 5 reais. Além disso, o contrato WDO apresenta alta liquidez em comparação ao contrato cheio, devido aos valores de custo e margem mais acessíveis. Esse contrato é adequado para investidores que desejam operar com volumes menores e desfrutar de maior flexibilidade (B3, 2022f).

### 2.3 CRISP-DM

O CRISP-DM (*Cross-Industry Standard Process for Data Mining*), ou simplesmente Processo Padrão da Indústria Cruzada para Mineração de Dados, é uma metodologia ou modelo de processo utilizado para orientar projetos de análise de dados de forma estruturada e eficiente, composta por seis etapas: entendimento do negócio, entendimento dos dados, preparação dos dados, modelagem, avaliação e implantação (IBM, 2023).

Essa metodologia é altamente flexível e iterativa, permitindo que cada etapa seja revisada e refinada ao longo do processo de análise de dados. Isso significa que é possível retornar a uma etapa anterior a qualquer momento para ajustar os detalhes. Em termos de flexibilidade, é possível aplicar modelos de análise diferentes de acordo com as necessidades específicas de cada projeto. É fundamental garantir uma preparação adequada nas primeiras etapas, como o entendimento do negócio e dos dados (IBM, 2023).

Figura 12 – Metodologia CRISP-DM.



Fonte: Traduzido de (IBM, 2023).

### 2.3.1 Entendimento de Negócios

O entendimento do negócio é considerado como a etapa inicial e crítica no processo de mineração de dados ou análise de dados. Nessa fase, o objetivo é obter uma compreensão profunda dos objetivos e requisitos do negócio, a fim de orientar a análise de dados de forma eficaz. Isso envolve identificar os objetivos do projeto, definir o escopo da análise, e entender os requisitos de informação necessários para atingir os objetivos (IBM, 2023, p.5).

Durante o entendimento do negócio, é essencial realizar uma análise completa das necessidades da organização, considerando os aspectos estratégicos, operacionais e técnicos. Isso inclui avaliar os desafios enfrentados, as oportunidades de negócio, as restrições orçamentárias e de tempo, bem como compreender os requisitos específicos dos usuários finais e das partes interessadas. Essa etapa é fundamental para garantir que a análise de dados seja alinhada com as metas e prioridades do negócio (IBM, 2023, p.6-9).

Ao concluir o processo de entendimento do negócio, espera-se ter um conhecimento abrangente dos objetivos, restrições e requisitos do projeto. Isso fornecerá uma base sólida para o planejamento das próximas etapas do processo de mineração de dados, permitindo que sejam tomadas decisões informadas sobre as técnicas e abordagens de análise mais adequadas para alcançar os resultados desejados. O entendimento claro do negócio também ajuda a estabelecer expectativas realistas e a criar um alinhamento entre a equipe de análise de dados e os *stakeholders* do projeto (IBM, 2023, p.11).

### 2.3.2 Entendimento dos Dados

Essa etapa é fundamental para o sucesso da análise de dados, pois envolve a exploração detalhada dos conjuntos de dados disponíveis. Durante o processo de entendimento dos dados,

os analistas devem examinar a estrutura, o conteúdo e as características dos dados (IBM, 2023, p.13).

Uma parte importante do entendimento dos dados é identificar e compreender as relações entre as variáveis. Os analistas devem examinar as associações, correlações e padrões nos dados, buscando entender as interações entre as diferentes variáveis e o impacto que elas podem ter nos resultados da análise. Além disso, é necessário avaliar a qualidade e a integridade dos dados, verificando se existem valores ausentes, *outliers* ou inconsistências que possam afetar os resultados (IBM, 2023, p.14-15).

Durante o processo de entendimento dos dados, é comum utilizar técnicas de visualização, como gráficos e tabelas, para ajudar na exploração e na compreensão dos dados. Essas representações visuais facilitam a identificação de padrões, tendências e anomalias nos dados, permitindo uma análise mais eficaz. Além disso, a documentação detalhada dos resultados e das observações feitas durante o entendimento dos dados é essencial para fornecer um registro claro e reproduzível das etapas realizadas (IBM, 2023, p.13-15).

### 2.3.3 Preparação de Dados

Essa fase envolve a transformação e a limpeza dos conjuntos de dados, a fim de garantir que eles estejam prontos para a modelagem. Durante a preparação dos dados, os analistas realizam uma série de tarefas, incluindo a seleção das variáveis relevantes, o tratamento de valores ausentes e *outliers*, e a criação de novas variáveis, "estima-se que a preparação de dados geralmente leve de 50 a 70% do tempo e do esforço de um projeto" (IBM, 2023, p.19).

Uma das principais atividades na preparação dos dados é a seleção das variáveis que serão usadas na análise preditiva. Os analistas devem identificar quais variáveis são relevantes para o objetivo da análise preditiva e descartar aquelas que não contribuem significativamente. Isso envolve a compreensão das relações entre as variáveis e a consideração dos critérios estabelecidos para a análise (IBM, 2023, p.19-20).

Durante a preparação dos dados também é comum a criação de novas variáveis derivadas. Isso envolve a combinação ou transformação de variáveis existentes para criar novos indicadores ou métricas relevantes para a análise. Por exemplo, pode-se calcular a média, o desvio padrão ou a diferença entre variáveis para gerar *insights* adicionais. Essas variáveis derivadas podem fornecer uma perspectiva mais abrangente dos dados e facilitar a modelagem (IBM, 2023, p.21-22).

### 2.3.4 Modelagem

Durante a etapa de modelagem, os analistas selecionam e aplicam os algoritmos adequados para os seus objetivos de análise. Esses algoritmos podem variar desde modelos lineares, como regressão linear, até algoritmos mais avançados, como árvores de decisão, redes neurais ou algoritmos de clusterização. A escolha do algoritmo depende do tipo de problema, das características dos dados, das metas da análise e os tipos de análise seja descritiva, preditiva ou prescritiva. O objetivo principal da modelagem é desenvolver modelos que possam capturar os padrões e relacionamentos nos dados e fornecer visões diferentes para a tomada de decisões (IBM, 2023, p.25).

Após a criação dos modelos, é essencial avaliar sua eficácia e precisão para determinar quais são adequados para uso final. A avaliação envolve a aplicação de critérios predefinidos, como métricas de desempenho, para determinar quais modelos são mais adequados. Além disso, é importante analisar a lógica dos resultados obtidos e considerar a facilidade de interpretação e implantação dos modelos. Esta tarefa não deve ser confundida com etapa posterior, esta tarefa

visa usar métricas para avaliar o modelo criado, as métricas comuns são: Acurácia, Precisão, *Recall* e *F1-Score* (IBM, 2023, p.28).

### 2.3.5 Avaliação

A etapa de avaliação tem como objetivo validar a eficácia do modelo em relação aos objetivos estabelecidos. Conforme mencionado anteriormente, a tarefa de avaliação na etapa de modelagem tem como objetivo aprimorar o próprio modelo, ajustando seus parâmetros. Nessa etapa, a avaliação é realizada para comparar os resultados do modelo com os objetivos e metas estabelecidos na primeira etapa, o Entendimento do Negócio (IBM, 2023, p.31).

Essa etapa é crucial, pois um modelo que atende às metas e objetivos pode prosseguir para a etapa final e ser concluído. No entanto, se o modelo não atender às metas e objetivos estabelecidos, pode ser necessário retornar até mesmo ao início da metodologia, revisando o entendimento do negócio, obtendo uma nova compreensão dos dados, realizando uma nova preparação dos dados e uma nova modelagem, ou seja, refazendo todo o processo (IBM, 2023, p.31-32).

Com o objetivo de esclarecer esses conceitos, as métricas empregadas nessa fase apresentam uma abordagem mais abrangente em relação à etapa anterior. Essa confusão surge devido à tradução dos termos "Assess" e "Evaluate" para o português, ambos sendo traduzidos como "Avaliação", embora seus significados sejam opostos. O "Assess" é realizado durante a fase de modelagem e geralmente envolve uma avaliação quantitativa, expressa em números. Já na etapa seguinte, utilizamos o termo "Evaluate" para descrever uma avaliação qualitativa, verificando se o modelo atende aos requisitos estabelecidos e se é considerado satisfatório.

### 2.3.6 Implementação

A etapa de Implementação, é a última fase do processo de mineração de dados. Nessa etapa, o foco está em colocar o modelo em uso prático, integrando-o ao ambiente de negócios e garantindo sua efetiva aplicação (IBM, 2023, p.35).

Durante a implementação, é fundamental realizar a transferência do modelo para a produção, o que pode envolver a configuração de sistemas e a preparação dos dados em escala real. Além disso, é importante garantir que os *stakeholders* estejam adequadamente capacitados para utilizar e interpretar os resultados do modelo. A documentação completa do processo de implementação, incluindo todas as etapas e decisões tomadas, é essencial para permitir a reprodução e manutenção adequada do modelo no futuro (IBM, 2023, p.35-38).

## 2.4 TRABALHOS RELACIONADOS

O livro de Bulkowski (2008) aborda uma variedade de padrões de *candle*, fornecendo suas características principais e realizando análises estatísticas sobre cada um deles. Por exemplo, no (BULKOWSKI, 2008, cap.2) apresenta um *ranking* baseado em informações estatísticas, como a frequência com que um padrão reverteu o movimento do mercado, seu desempenho ao longo de determinados períodos e qual padrão se saiu melhor em diferentes movimentos de mercado. Além disso, o autor define estratégias para operar os padrões de *candle*, incluindo pontos de entrada e saída, *stop loss* e metas de lucro, ou *stop gain*. Ele também calcula a lucratividade dos padrões com base nessas estratégias, considerando a média de lucro e avaliando a viabilidade das estratégias. O livro é amplamente reconhecido entre analistas devido à teoria abordada pelo autor, bem como ao estudo prático da aplicação das estratégias que envolvem exclusivamente os padrões de *candle*. É importante ressaltar que o livro abrange

ações do S&P500 ao longo de aproximadamente 10 anos de análise, incluindo a análise de cerca de 5 milhões de *candles* (BULKOWSKI, 2008, p.21). Os exemplos nos capítulos incluem velas de diferentes períodos, como diário, semanal e até mesmo de horas.

A dissertação de mestrado de Maia (2018) consistiu em um estudo no qual ele classificou 24 padrões de *candle* para o movimento de baixa e 21 padrões de *candle* para o movimento de alta no ativo BOVA11, no período de 2005 a 2010. Em seguida, ele testou duas estratégias utilizando algoritmos como Monte Carlo, Teste T, Teste Z, entre outros, durante o período de 2011 a 2016. Por meio de resultados apresentados em tabelas e gráficos, ele pôde avaliar a eficácia de cada estratégia e o desempenho de cada padrão nos algoritmos desenvolvidos. Com base nesse estudo, ele conseguiu identificar os momentos ideais de entrada e saída para cada padrão, bem como avaliar a lucratividade desses pontos de entrada. Através das estratégias abordadas, ele pôde determinar quais padrões de *candle* apresentavam maior potencial de lucro e qual era o desempenho de cada um deles nos algoritmos utilizados (MAIA, 2018, p.43-73). Essa pesquisa identificou oportunidades de negociação e aprimorar as duas estratégias de investimento usadas com base nos padrões de *candle* analisados no ativo BOVA11 (MAIA, 2018, p.74-77).

O trabalho de conclusão de curso de Mateus (2022) consistiu no treinamento de um modelo de rede neural para identificar topos e fundos duplos em 6 ações de alta liquidez no mercado da B3. Além de identificar esses padrões, a rede neural também aplicava estratégias de negociação. Com base nesse modelo, desenvolveu-se uma aplicação que fornecia resultados das estratégias adotadas, permitindo a comparação entre os ativos e identificando a melhor estratégia a ser utilizada. Através dessa aplicação, ele foi capaz de ajustar os parâmetros das estratégias e determinar qual ativo apresentava os melhores resultados. Ele também calculou medidas importantes, como a média de ganhos, perdas e o número de entradas realizadas. Com essas informações, (MATEUS, 2022, p.36-40) obteve um lucro significativo ao operar os padrões de topo e fundo duplo identificados pela rede neural. Esse trabalho de conclusão de curso contribuiu para aprimorar a compreensão desses padrões de mercado com uso de imagens para identificar topos e fundos duplos e fornecer *insights* para a tomada de decisões de investimento (MATEUS, 2022, p.41-42) .

### 3 DESENVOLVIMENTO

Este capítulo apresenta os processos envolvidos na elaboração do trabalho, seguindo os princípios estabelecidos pela metodologia CRISP-DM. Serão discutidas as estratégias analisadas, bem como as ferramentas e tecnologias utilizadas no desenvolvimento do estudo.

#### 3.1 TECNOLOGIA E FERRAMENTAS

Nesta seção, são apresentadas as tecnologias e ferramentas que foram empregadas no desenvolvimento do trabalho. Primeiramente, foi utilizada a plataforma de negociação TRYD, que oferece recursos para análise e extração de dados do mercado financeiro. Na sequência, a ferramenta ClickUp foi adotada para o controle e organização das atividades, permitindo o acompanhamento e gerenciamento eficiente das tarefas. Quanto à linguagem de programação, foi utilizado o Python, amplamente empregado em ciência de dados devido à sua simplicidade e às poderosas bibliotecas disponíveis para análise e criação de modelos de *machine learning*.

##### 3.1.1 Plataforma TRYD

Existem várias plataformas de negociação disponíveis no mercado, cada uma buscando simplificar o processo de operar na bolsa de valores. Uma das principais e amplamente utilizadas é a plataforma TRYD<sup>1</sup>, que possui integração com diversas corretoras. Uma vantagem do TRYD é a sua facilidade de uso e o fato de ser gratuita ao abrir uma conta em uma corretora. A plataforma TRYD está disponível tanto para *desktop* quanto para dispositivos móveis, podemos ver sua interface na [Figura 13](#).

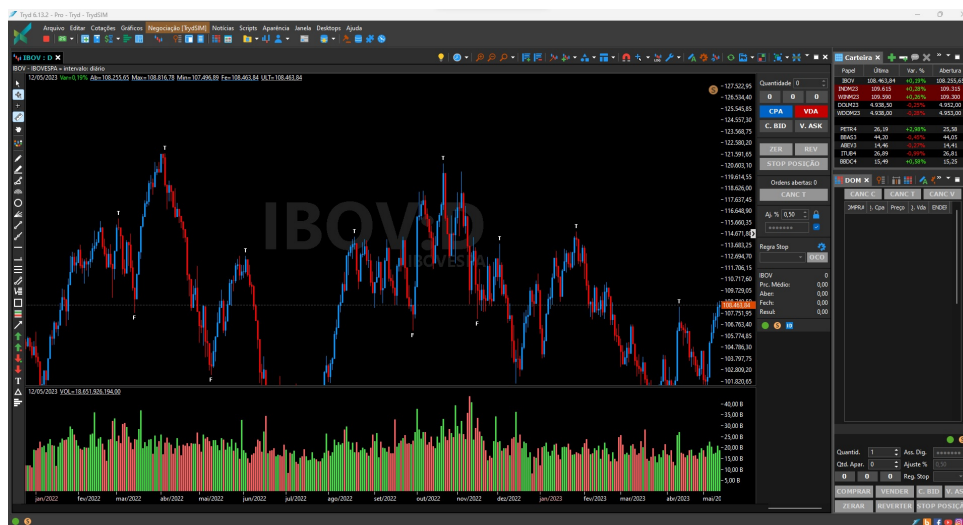
Uma funcionalidade interessante do TRYD é o seu gráfico simulador, que permite simular operações sem se expor aos riscos reais do mercado. Além disso, a plataforma permite exportar dados históricos de diferentes intervalos de tempo e também indicadores de análise técnica. Esses indicadores podem auxiliar na identificação de padrões gráficos, topos, fundos e tendências.

No entanto, uma desvantagem do TRYD é a falta de integração direta com a linguagem Python. Isso significa que modelos criados e treinados em Python não podem ser utilizados em tempo real na plataforma TRYD. A integração em tempo real com o TRYD é possível, porém, de forma atrasada, através do uso de DDE (*Dynamic Data Exchange*) e RTD (*Real-Time Data*), conectados ao Excel ([TRYD, 2023](#)).

---

<sup>1</sup><https://www.tryd.com.br/>

Figura 13 – Visão geral da plataforma TRYD



Fonte: Autoria própria

### 3.1.2 Click Up

O ClickUp<sup>2</sup> é uma ferramenta web para o gerenciamento de projetos e tarefas, semelhante ao Trello e Asana, que são concorrentes conhecidos. Trata-se de um software baseado em nuvem e altamente personalizável. No desenvolvimento deste trabalho, o ClickUp foi utilizado para o controle e organização das atividades.

Uma característica interessante do ClickUp é a oferta gratuita de muitas funcionalidades que, em outras ferramentas, normalmente estão disponíveis apenas nas versões *premium*. Isso inclui recursos como Diagramas de Gantt, integrações com planilhas Excel ou Google Planilhas, e várias opções de personalização nas listas de tarefas, automações e notificações. Essas funcionalidades adicionais foram úteis para melhorar a organização e a eficiência do gerenciamento do projeto (GOMES, 2022).

Os quadros Kanban são compostos por colunas que representam os estágios do fluxo de trabalho, e as tarefas são movidas de uma coluna para outra à medida que avançam no processo. Isso oferece uma melhor visualização do fluxo de trabalho, permitindo a identificação de gargalos e uma distribuição equilibrada de tarefas entre os membros da equipe (GOMES, 2022).

Uma das vantagens do ClickUp é que seus quadros Kanban são altamente adaptáveis e personalizáveis. É possível classificar as tarefas por níveis de prioridade, estabelecer cores para diferenciar tipos de tarefas ou anexar arquivos relevantes em cada quadro. Essa flexibilidade permite que os usuários personalizem os quadros Kanban de acordo com suas necessidades específicas e melhorem a eficiência do gerenciamento de tarefas (GOMES, 2022).

### 3.1.3 Bibliotecas e Ferramentas Python

A linguagem de programação escolhida para o projeto foi Python, amplamente utilizada em projetos que envolvem análise de dados e ciência de dados. As bibliotecas listadas no Quadro 1 desempenharam um papel fundamental no desenvolvimento do trabalho, principalmente na análise exploratória dos dados.

<sup>2</sup><https://clickup.com/>

Quadro 1 – Bibliotecas padrão de análise de dados.

Nome da Biblioteca	Descrição
Pandas	Uma poderosa biblioteca para manipulação e análise de dados em formato de <i>DataFrames</i> ou <i>Series</i> .
NumPy	Esta biblioteca fornece inúmeras funções para cálculos matemáticos e algébricos, criação e manipulação de <i>arrays</i> unidimensionais e multidimensionais, entre outras funcionalidades.
Matplotlib	Uma ferramenta essencial para visualização, plotagem e configuração de gráficos na linguagem Python.
Seaborn	Esta biblioteca é baseada no Matplotlib e fornece visualizações gráficas voltadas para análises estatísticas, como <i>boxplot</i> , histogramas, regressão e outros modelos estatísticos.
Plotly	Fornece recursos interativos para a criação de gráficos e <i>dashboards</i> , como zoom no gráfico e recorte nos eixos de forma intuitiva. Essa interatividade não é encontrada nas bibliotecas de visualização mencionadas anteriormente.
Scikit-Learn	Essa biblioteca oferece desde técnicas para tratamento de dados, como a normalização dos dados, até recursos para todas as etapas do processo de modelagem. Além de disponibilizar modelos de aprendizado, também inclui métricas para avaliação, <i>pipelines</i> e validação cruzada.
Scikit-Optimize	Essa biblioteca oferece recursos para aprimorar os modelos de aprendizagem, incluindo a busca bayesiana para a otimização de hiperparâmetros. Além disso, disponibiliza técnicas de <i>over sample</i> e <i>under sample</i> para o balanceamento dos dados.

Fonte: Elaborado pelo próprio autor

Além disso, os códigos e o uso dessas bibliotecas foram realizados em um ambiente interativo baseado na *web*, o Jupyter Notebook<sup>3</sup>. Essa ferramenta permite controlar e organizar o fluxo de trabalho das tarefas de maneira eficiente.

Todas essas bibliotecas e ferramentas podem ser instaladas e gerenciadas facilmente por meio da plataforma Anaconda<sup>4</sup>. Essa ferramenta é capaz de fornecer um ambiente unificado, o que facilita a configuração e o acesso às bibliotecas necessárias. Com o uso do Anaconda, é possível agilizar o processo de instalação das ferramentas e bibliotecas, evitando a necessidade de realizar o *download* e a instalação individual de cada componente.

### 3.2 ENTENDIMENTO DO NEGÓCIO

Conforme mencionado, a motivação deste trabalho é verificar se as estratégias de *day-trade* que envolvem padrões de *candlesticks* e outros indicadores são lucrativas, bem como avaliar se os modelos de *machine learning* podem contribuir para aprimorar essa lucratividade. Nesse contexto, os seguintes pontos foram levantados:

<sup>3</sup><https://jupyter.org/>

<sup>4</sup><https://www.anaconda.com/>

- Qual ativo financeiro será analisado?
- Qual será o período de análise?
- Como será realizada a coleta dos dados?
- Baseando-se no *day-trade*, qual será o tempo gráfico considerado?
- Quais estratégias serão empregadas?

A escolha dos ativos financeiros deve considerar os princípios da Teoria de Dow relacionados ao volume. Este está intrinsecamente ligado às tendências de mercado, uma vez que os movimentos dos preços são influenciados por elas. Em outras palavras, ativos com maior volume tendem a seguir mais consistentemente os indicadores da análise técnica. Com base nessa premissa, foi utilizada a plataforma TRYD para analisar os dados de diferentes ativos financeiros, cujos resultados estão apresentados na [Tabela 1](#).

Tabela 1 – Posição dos ativos por volume financeiro médio mensal ao longo dos anos (em bilhões de reais).

Posição	Ativo	2018	2019	2020	2021	2022	2023	Média Total
1	WIN	47,50	131,60	233,53	436,54	363,44	324,58	256,20
2	WDO	40,28	54,03	146,92	179,38	179,46	156,02	126,01
3	DOL	67,67	74,38	81,27	75,91	69,53	66,55	72,55
4	IND	7,92	14,54	16,51	23,06	18,38	14,18	15,77
5	PETR4	1,44	1,43	1,57	2,10	2,52	1,60	1,77
6	VALE3	0,88	1,01	1,71	2,58	2,52	1,92	1,77
7	ITUB4	0,65	0,73	1,03	1,04	0,99	0,87	0,88
8	BBDC4	0,45	0,64	0,97	0,99	0,81	0,82	0,78

Fonte: Elaborado pelo próprio autor

Observou-se que os ativos com maior volume estavam concentrados no mercado futuro, destacando-se o WIN (Mini Índice Ibovespa), que apresentou mais do que o dobro do volume registrado pelo segundo ativo mais negociado, o WDO (Mini Dólar). Essa diferença foi atribuída à menor alavancagem relativa do Mini Índice em comparação ao Mini Dólar, proporcionando maior segurança. Embora os contratos cheios (IND e DOL) também apresentassem alavancagem elevada, o volume desses contratos superou significativamente o das ações. Entre os anos de 2018 e 2023, as ações com maior volume foram PETR4 (Petrobrás), VALE3 (Vale), ITUB4 (Itaú) e BBDC4 (Bradesco). Dessa forma, a escolha do ativo financeiro recaiu WIN, em razão de seu alto volume financeiro. Esse ativo apresentou maior aderência às análises e estratégias da análise técnica aplicadas no *day-trade*.

O período de análise compreendeu os anos de 2018 a 2023, conforme apresentado na [Tabela 1](#). No entanto, para a aplicação de modelos de *machine learning*, é necessário separar os dados em conjuntos de treino e teste. Assim, o período de treino abrange os anos de 2018 a 2022, enquanto o período de teste corresponde ao ano de 2023.

Esse período de análise apresenta desafios significativos devido a diversas interferências, como eventos políticos e os impactos da pandemia de Covid-19. Essas circunstâncias resultaram na ausência de uma tendência clara no mercado durante o intervalo analisado, caracterizado por alta volatilidade e movimentos abruptos. Embora um período mais longo para treino e teste fosse ideal para uma análise mais robusta, amenizando estas interferências, a plataforma utilizada não disponibiliza dados anteriores a 2018.

É fundamental esclarecer qual *dataset* será analisado. A plataforma utilizada permite a exportação de dados de *candlesticks* em diversos tempos gráficos, resultando em um *dataset*

contendo informações sobre esses *candles*, incluindo data e hora, valores de abertura, máxima, mínima, fechamento e volume financeiro. Esse conjunto de dados será denominado *dataset* dos *candles*.

Embora esse *dataset* receba tratamentos específicos, ele não será diretamente analisado. Sua principal função é servir de base para a criação de simulações de operações. A partir das variáveis contidas no *dataset* dos *candles*, serão geradas operações simuladas, cujos resultados serão coletados para compor um novo *dataset*, denominado *dataset* de operações. É esse *dataset* de operações que será analisado, e com ele será possível verificar, por meio de modelos de classificação, a capacidade de prever operações lucrativas. Todo o processo de construção desses dois *datasets* será detalhado na próxima seção.

Com relação ao tempo gráfico, por se tratar de *day-trade*, é necessário utilizar um tempo gráfico intradiário, não superior a 1 hora. Assim, restam as opções de 15 minutos, 5 minutos e 1 minuto. Considerando que o tempo gráfico de 15 minutos possui um período maior e proporciona uma boa quantidade de padrões de *candlestick*, optou-se pela análise com *candles* de 15 minutos.

Apesar do tempo gráfico de 1 hora oferecer maior estabilidade, a formação de padrões de *candlestick* nesse intervalo é menos frequente, dificultando sua identificação. Por outro lado, tempos gráficos menores, como 5 minutos e 1 minuto, aumentam significativamente a quantidade de padrões devido à maior granularidade dos dados, mas reduzem o período de análise efetiva, tornando-os menos adequados para este estudo. Dessa forma, os *candles* de 15 minutos foram definidos como o tempo gráfico ideal, equilibrando frequência de padrões e abrangência do período analisado.

As estratégias empregadas têm como base o gerenciamento de *trade*, utilizando *stop loss* e *stop gain*. Os pontos de entrada e confirmação serão definidos com base no livro de [Nison \(1991\)](#) e [Bulkowski \(2008\)](#). No entanto, os pontos de saída, que frequentemente envolvem a identificação de fatores como tendência, suporte e resistência, apresentam maior complexidade na análise técnica quando baseados apenas em dados brutos, tornando sua implementação mais desafiadora. Esses elementos de análise técnica, como suporte e resistência, seriam mais facilmente identificados com o uso de técnicas de processamento de imagem.

Dada essa dificuldade, as estratégias de saída adotadas neste estudo utilizarão métodos mais simples, baseando-se em *stop loss* e *stop gain*, com pontos pré-definidos de ganho ou perda. Essas estratégias podem seguir proporções como 1:1, 2:1, ou outras configurações preestabelecidas, possibilitando a definição clara dos limites de risco e retorno de cada operação.

### 3.3 ENTENDIMENTO DOS DADOS

Na seção anterior, foram mencionados dois *datasets*: o de *candlesticks* e o de operações. Nesta seção, será feita a distinção entre eles, detalhando a relação existente. Em seguida, serão apresentadas as análises exploratórias do *dataset* de operações.

#### 3.3.1 *Dataset* de *Candlesticks*

O *dataset* de *candlesticks* é gerado pela plataforma utilizada e representa os dados brutos, os quais podem ser exportados em formato CSV (*Comma-separated values*). A [Figura 14](#) ilustra os dados brutos exportados. Durante o processo de exportação, é essencial realizar a separação entre os dados de treino e teste antes de qualquer tratamento, para evitar o vazamento de informações, *outdata leakage*, que ocorre quando dados do conjunto de teste influenciam o treino do modelo, comprometendo a avaliação realista de seu desempenho.

Figura 14 – Importação do *dataset* de *candles*

Out[5]:

	Data	Hora	Abertura	Máxima	Mínima	Fechamento	VOL
0	28/12/2023	18:30:00	135850	135850	135850	135850	236732210
1	28/12/2023	18:15:00	135855	135935	135830	135890	1351739856
2	28/12/2023	18:00:00	135795	135880	135725	135850	3520213939
3	28/12/2023	17:45:00	135855	135910	135775	135795	4350075596
4	28/12/2023	17:30:00	135725	135870	135710	135850	4509936054
...	...	...	...	...	...	...	...
54030	28/02/2018	10:00:00	87580	87680	87330	87345	15793272910
54031	28/02/2018	09:45:00	87505	87590	87495	87575	7356277630
54032	28/02/2018	09:30:00	87475	87540	87415	87505	6684590490
54033	28/02/2018	09:15:00	87600	87640	87475	87475	5988311960
54034	28/02/2018	09:00:00	87570	87620	87400	87600	7230584370

54035 rows × 7 columns

Fonte: Autoria própria

Conforme apresentado na Figura 14, o período analisado contém 54.035 *candles*, distribuídos em sete variáveis. Por meio de técnicas de *feature engineering*, foram adicionadas novas variáveis a esses *candles*, como a classificação dos *candles* como Martelo e suas variações, Harami e suas variações, além de Engolfo e suas variações. Além disso, foram calculados os valores de indicadores técnicos e incluídas outras variáveis adicionais, conforme detalhado a seguir:

- **"Índice\_Diario"**: variável quantitativa discreta que identifica a ordem do *candle* dentro de um dia. Como, em média, um dia contém 36 *candles*, essa variável permite marcar o tempo de maneira estruturada;
- **"Volume\_Dia"**: variável quantitativa contínua que representa o volume acumulado de todos os *candles* de um mesmo dia até o momento de sua formação;
- **"Percentual\_Dia"**: variável que calcula o percentual da diferença entre o preço de abertura do dia e o preço de abertura de cada *candle*, proporcionando uma análise da volatilidade diária;
- **"Percentual\_Candle"**: variável semelhante à anterior, mas que calcula a variação percentual entre o preço de abertura do *candle* atual e o preço de abertura do *candle* anterior;
- **"Tamanho\_Candle"**: variável que representa a amplitude do *candle*, calculada pela diferença entre o preço máximo e o preço mínimo;
- **"Período\_Dia"**: variável categórica que classifica o *candle* como pertencente ao período da manhã ou da tarde;
- **"Eventos\_Dia"**: variável categórica que identifica o contexto do *candle*, indicando se ele está associado a períodos normais de negociação, abertura do índice, abertura do mercado de ações e títulos, horário de almoço ou proximidade do fechamento do mercado.

Após a criação dessas variáveis, foi necessário realizar um tratamento específico nos dados. Durante o período analisado, ocorreram momentos de *circuit breaker*, que correspondem a interrupções abruptas no mercado em momentos de crise, com o objetivo de proteger investidores e *traders* contra quedas acentuadas ou volatilidade extrema (ANVERSA, 2024). Essas interrupções, bem como o encerramento regular do mercado, geram *candles* de tamanho

zero, nos quais os preços de abertura, máxima, mínima e fechamento são iguais. Esses *candles* são comumente classificados como "doji dos quatro preços" e indicam a ausência de movimentação de mercado no período.

Considerando que esses *dojis* representam apenas o fechamento formal do pregão e não trazem informações úteis para a análise técnica, decidiu-se pela exclusão dessas linhas do conjunto de dados. Essa remoção não compromete a estrutura temporal nem a sequência dos dados, pois esses *candles* representam períodos sem movimentação real e geralmente repetem o preço do *candle* anterior.

A classificação dos *candles* foi realizada utilizando a biblioteca *talib*, que, além de fornecer classificações de *candles*, também disponibiliza indicadores técnicos já calculados. Por essa razão, foram utilizados os indicadores disponíveis na biblioteca, uma vez que eles representam, de maneira matemática, as características da análise técnica, como tendências, suportes e resistências.

Entretanto, essa abordagem apresenta desafios, não apenas relacionados à escolha adequada dos parâmetros dos indicadores, mas também à identificação de uma tendência, suporte ou resistência. Essa tarefa é complexa porque a identificação visual desses elementos nem sempre é clara. Assim, os indicadores utilizam estatísticas e cálculos para estimar a possibilidade de ocorrência desses eventos na análise técnica.

Para rastrear ou identificar tendências, foi utilizado o *Hi-Lo Activator*, que não está implementado na biblioteca *talib*, mas cujo cálculo é relativamente simples, porém trabalhoso. São necessárias duas médias móveis simples: uma calculada com os valores máximos e outra com os valores mínimos. O período dessas médias é escolhido conforme necessário. Após o cálculo das médias, estas são deslocadas por um período, que corresponde a 15 minutos, já que todo o conjunto de dados possui *candles* de 15 minutos. Ou seja, o *candle* atual utiliza os valores das médias calculados no *candle* anterior.

Esses cálculos estão ilustrados na [Figura 15](#). As linhas vermelhas representam a média móvel dos valores máximos, sendo a linha tracejada a média sem deslocamento e a linha contínua a média com deslocamento. O mesmo ocorre com as linhas verdes, que representam as médias dos valores mínimos, sendo a linha contínua a média já deslocada ([SMARTTBOT, 2016](#)).

Figura 15 – Cálculo das médias móveis e deslocadas do *Hi-Lo Activator*.



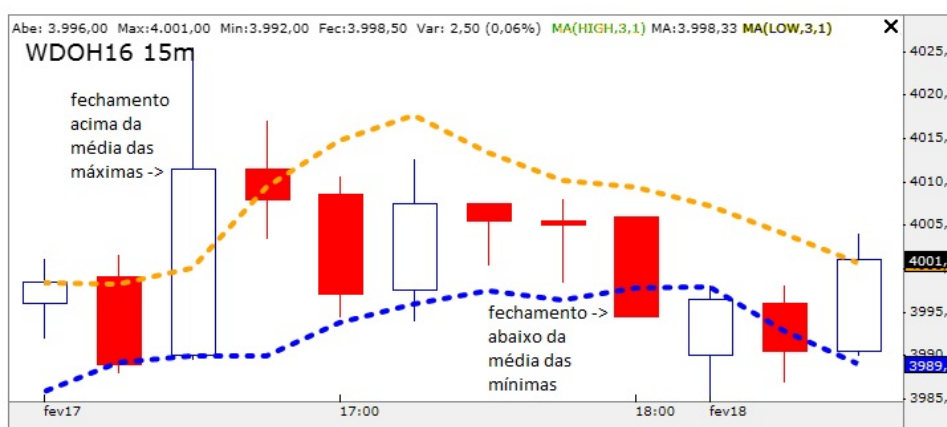
Fonte: ([SMARTTBOT, 2016](#))

Após o cálculo, foi necessário definir a tendência. Como há duas médias, uma acima e outra abaixo dos *candles*, a mudança de tendência ocorre quando uma das médias atravessa um *candle*. Por exemplo, se a média dos valores máximos atravessa um *candle*, identifica-se uma tendência de alta, e a média dos valores mínimos passa a ser a referência. De forma análoga, se a média dos valores mínimos atravessa um *candle*, ocorre uma mudança para uma tendência de baixa, e a média dos valores máximos passa a ser a referência.

Essa situação está representada na Figura 16, que é uma continuação da Figura 15. Nesta figura, a média dos máximos está em laranja, indicando uma tendência de baixa, enquanto a média dos mínimos está em azul, indicando uma tendência de alta. As mudanças de tendência são determinadas com base no fechamento dos *candles*.

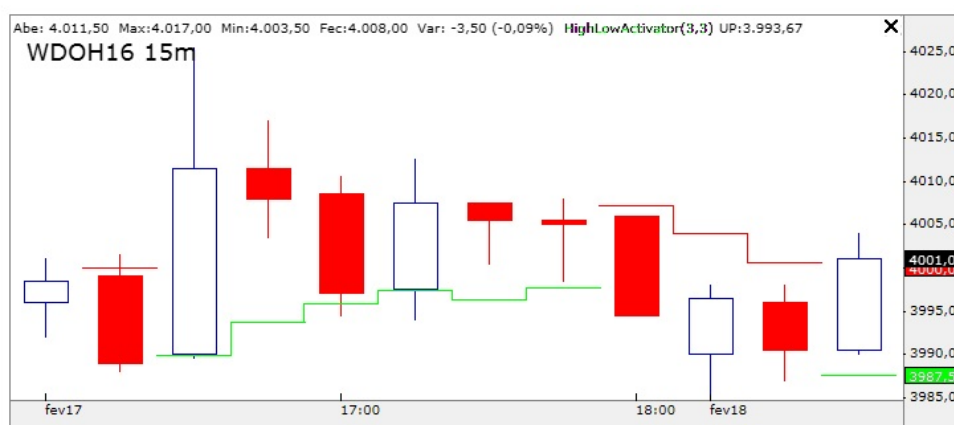
Para simplificar a identificação da tendência, as plataformas e o código utilizam uma marcação visual, como ilustrado na Figura 17. Nesse caso, os *candles* são considerados em alta quando há uma marcação verde abaixo deles, e em baixa quando há uma marcação vermelha acima. Para gerar essas linhas, os valores das médias são arredondados, visto que o índice encerra com valores múltiplos de 5 (SMARTTBOT, 2016).

Figura 16 – Tendências identificadas pelo *Hi-Lo Activator*.



Fonte: (SMARTTBOT, 2016)

Figura 17 – Indicador *Hi-Lo Activator* em uso.



Fonte: (SMARTTBOT, 2016)

O principal objetivo do indicador *Hi-Lo Activator* é rastrear tendências. Contudo, os valores das marcações vermelhas e verdes apresentados na [Figura 17](#) também podem indicar suportes e resistências. Por esse motivo, no *dataset* de *candles*, optou-se por manter não apenas a indicação da tendência (alta ou baixa), mas também os valores das médias do *Hi-Lo*, para que possam ser utilizados como suporte e resistência. O período das médias móveis utilizado foi de 10 *candles*, sendo esse o padrão ([SMARTTBOT, 2016](#)).

As operações com padrões de *candles* requerem uma indicação de tendência. Por exemplo, no padrão de *candles* Martelo, após uma tendência de baixa, ocorre o padrão e pode-se esperar uma reversão para uma tendência de alta, caracterizando-o como um padrão de reversão. Por outro lado, há uma variação do Martelo, conhecida como Estrela Cadente, que atua de forma contrária: após uma tendência de alta, espera-se uma reversão para tendência de baixa, como pode ser visto na [Figura 8](#).

Assim, a identificação das tendências é essencial para as operações, e o indicador *Hi-Lo Activator* desempenha justamente esse papel, sendo o principal indicador utilizado neste *dataset*. Embora outros métodos, como o cruzamento de médias móveis exponenciais, possam ser utilizados para identificar tendências, neste trabalho, eles servirão como suporte para o *Hi-Lo*.

Os indicadores auxiliares utilizados, já calculados pela biblioteca *talib*, são apresentados a seguir, com seus objetivos e parâmetros padronizados, conforme usualmente empregados:

- **MACD**: utiliza duas médias móveis exponenciais de curto e médio prazo, juntamente com uma linha de sinal, que representa a média da diferença entre elas. Este indicador é usado para identificar a força de uma tendência.
- **RSI**: mede a força da tendência, indicando se o preço está em condições de sobrecompra ou sobrevenda, ou seja, se está caro ou barato demais.
- **Estocástico**: tem o mesmo objetivo do *RSI*, verificando condições de sobrecompra ou sobrevenda. A diferença está no cálculo: enquanto o *RSI* utiliza médias de ganhos e perdas, o Estocástico analisa os valores de fechamento, máximos e mínimos em relação ao preço atual.
- **VWAP**: é uma média ponderada do preço pelo volume, utilizada para identificar suportes e resistências de maneira dinâmica, variando conforme o preço médio negociado pelos *traders*.
- **Médias móveis exponenciais**: utilizadas para identificar tendências, principalmente por meio do cruzamento de uma ou mais médias. Este conceito é a base de outros indicadores, como o *MACD*. A diferença em relação à média móvel simples está no fato de que a exponencial atribui mais peso aos preços recentes, sendo, portanto, mais sensível a mudanças. Neste trabalho, foram utilizadas médias móveis exponenciais de 9 períodos (curto prazo) e 21 períodos (médio e longo prazo). Para operações *day trade*, 21 *candles* de 15 minutos representam um período considerado longo.

Após a realização da *feature engineering* e o filtro de algumas linhas, o *dataset* de *candlesticks* para treino foi definido com 43.670 *candles*, distribuídos em 37 variáveis, enquanto o *dataset* de teste contém 9.049 *candles*. A [Tabela 2](#) resume as informações relativas aos *datasets* de treino e teste no que diz respeito à quantidade de padrões de *candles*, tais como martelo, engolfo e harami.

Tabela 2 – Quantidade de candles e padrões de candle e sua representatividade.

Dataset	Total Candles	Engolfo	Peso	Martelo	Peso	Harami	Peso
Treino	43.670	2.362	5,41%	1.673	3,83%	1.492	3,42%
Teste	9.049	514	5,68%	404	4,46%	313	3,46%

Fonte: Elaborado pelo próprio autor

### 3.3.2 Dataset de Operações

Com o *dataset* de *candles* pronto, são simuladas as operações seguindo a estratégia previamente definida na [Seção 3.2](#). Essa estratégia baseia-se, inicialmente, em identificar um padrão de *candle* (Martelo, Harami ou Engolfo) e aguardar a confirmação desse padrão.

A confirmação geralmente ocorre no próximo *candle*, que precisa ultrapassar o preço de entrada. No caso do padrão Martelo e suas variações, é necessário aguardar o fechamento do *candle* para validar a confirmação. Dessa forma, a estratégia de operar padrões não se limita à simples identificação do padrão, mas inclui a espera por sua confirmação, momento em que o padrão de *candle* se torna um *candle* de referência.

Enquanto o *candle* de confirmação não aparece, a referência do padrão não é alterada. No entanto, se surgir um novo padrão de *candle*, ele passa a ser a nova referência. Caso chegue ao final do dia sem confirmação ou sem um novo padrão, a referência é descartada.

No código implementado em *Python* para este trabalho, responsável por simular as operações com base nos padrões de *candles*, o padrão identificado é denominado *candle* de referência. Quando uma operação é realizada, as 37 variáveis desse *candle* são utilizadas. Assim, boa parte das variáveis do *dataset* de operações é composta pelas informações do *candle* de referência que originou uma operação, seja ela vencedora ou perdedora.

Outras regras aplicadas no código de operação dizem respeito ao encerramento das operações e ao método de validação dos preços. Todas as operações são encerradas às 17:45, ou seja, no momento em que o *candle* de 17:45 é fechado, utiliza-se o valor desse fechamento para verificar o lucro ou prejuízo em relação ao ponto de entrada. Portanto, o horário de operação da simulação vai das 9:00 às 17:45.

Outro ponto importante está relacionado à natureza estática dos dados utilizados, que não são em tempo real. Assim, a maneira de verificar se o preço atual atingiu o valor de entrada, o *stop loss* ou o *stop gain* é comparar esses valores com o preço máximo e mínimo do *candle*. Vale ressaltar que os pontos de entrada, *stop loss* e *stop gain* são pré-estabelecidos pelo *candle* de referência. Por exemplo, para um ganho de 500 pontos, o valor-alvo será o ponto de entrada somado a 500 em caso de compra ou subtraído 500 em caso de venda. Os valores máximos e mínimos do *candle* de confirmação são utilizados apenas para verificar se esses pontos foram atingidos.

Outro desafio desse método está em determinar qual ponto foi atingido primeiro: o *stop loss* ou o *stop gain*. Como os dados utilizados são compostos por *candles* de 15 minutos, não há informação sobre a sequência exata dos preços dentro de cada intervalo. Portanto, quando um mesmo *candle* atinge tanto a máxima quanto a mínima correspondentes aos níveis de *stop*, não é possível identificar com precisão qual dos dois foi acionado primeiro.

Para evitar ambiguidade e superestimação de resultados, optou-se por considerar o pior caso: sempre que o *stop loss* estiver dentro do intervalo do *candle*, a operação será registrada como perdedora, mesmo que o *stop gain* também tenha sido atingido. Essa abordagem reflete uma estratégia conservadora e privilegia a segurança dos resultados simulados. Embora existam situações em que o ganho possa ter ocorrido antes da perda, elas são exceções, já que o

mercado geralmente segue uma tendência clara dentro do intervalo analisado, atingindo apenas um dos extremos.

Nas operações envolvendo mercado futuro, é possível realizar tanto operações de compra quanto de venda. O lucro em uma operação de compra consiste em adquirir um contrato, esperar sua valorização e vendê-lo por um preço mais alto. Já na operação de venda, o objetivo é vender um contrato por um preço elevado e, posteriormente, recomprá-lo por um valor mais baixo.

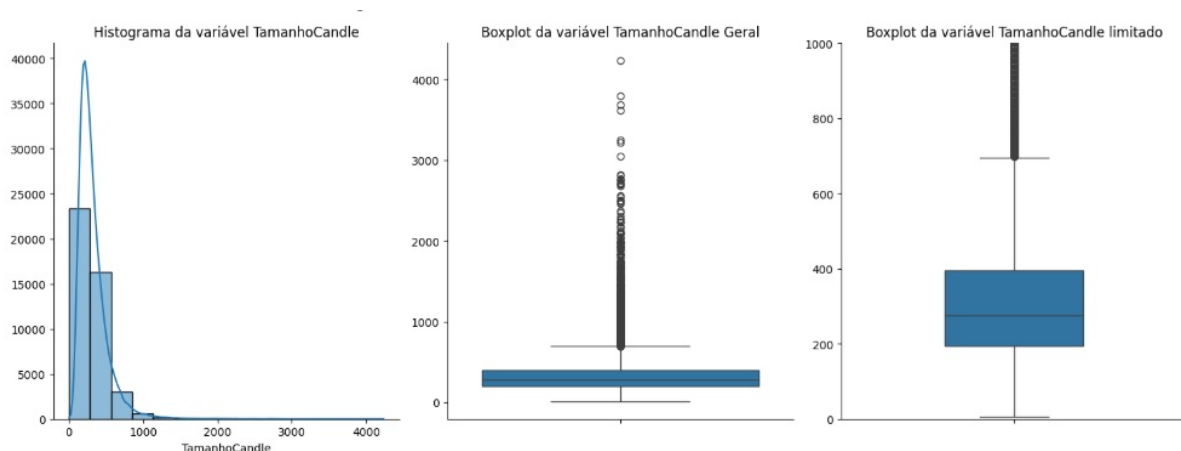
A decisão de comprar ou vender exige atenção aos padrões de *candles*, e duas situações principais podem ocorrer:

- **estar em uma tendência de baixa e aparecer um padrão que pode reverter para uma tendência de alta** — neste caso, será realizada uma operação de compra. Os padrões que indicam essa operação são: engolfo de alta, martelo, martelo invertido, harami de alta e harami cross;
- **estar em uma tendência de alta e aparecer um padrão que pode reverter para uma tendência de baixa** — será realizada uma operação de venda. Os padrões que indicam essa operação são: engolfo de baixa, estrela cadente, enforcado, harami de baixa e harami cross.

Com base na análise da variável *Tamanho\_Candle* do *dataset* de *candles*, podem ser estabelecidos valores para o *stop loss* e *stop gain*. De acordo com a [Figura 18](#), observa-se, por meio do histograma, uma assimetria positiva, ou seja, a frequência dos *candles* está concentrada em tamanhos abaixo de 1000. Além disso, os *boxplots* indicam que *candles* com tamanho superior a 700 são considerados outliers ou dados discrepantes. A variabilidade dos tamanhos dos *candles*, em termos de frequência, está entre 200 e 400 pontos. As estatísticas da [Figura 19](#) confirmam essa análise visual, mostrando que a mediana dos dados está em 275, o percentil 25% está em 195 e o percentil 75% em 395, o que reforça a variabilidade. Vale lembrar que o tamanho se refere a cada *candle* de 15 minutos.

Dessa forma, seria razoável adotar uma estratégia que busque 200 pontos de *gain*, já que este valor é atingível. Da mesma forma, um *gain* prolongado de 500 pontos também seria viável.

Figura 18 – Análise gráfica da variável "Tamanho\_Candle".



Fonte: Autoria própria

Figura 19 – Estatística descritiva da variável "Tamanho\_Candle".

```
Out[22]:
```

count	43670.00
mean	324.38
std	207.92
min	5.00
25%	195.00
50%	275.00
75%	395.00
max	4235.00
Name: TamanhoCandle, dtype: float64	

Fonte: Autoria própria

Logo, as estratégias serão baseadas nestes dois alvos, considerando diferentes proporções entre *gain* e *loss*. Para uma proporção 1:1, são utilizados 200 pontos de *gain* e 200 pontos de *loss*, assim como 500 pontos de *gain* e 500 pontos de *loss*.

Para uma proporção de 2:1, que implica dois ganhos para uma perda, temos 200 pontos de *gain* para 100 pontos de *loss*, e 500 pontos de *gain* para 250 pontos de *loss*.

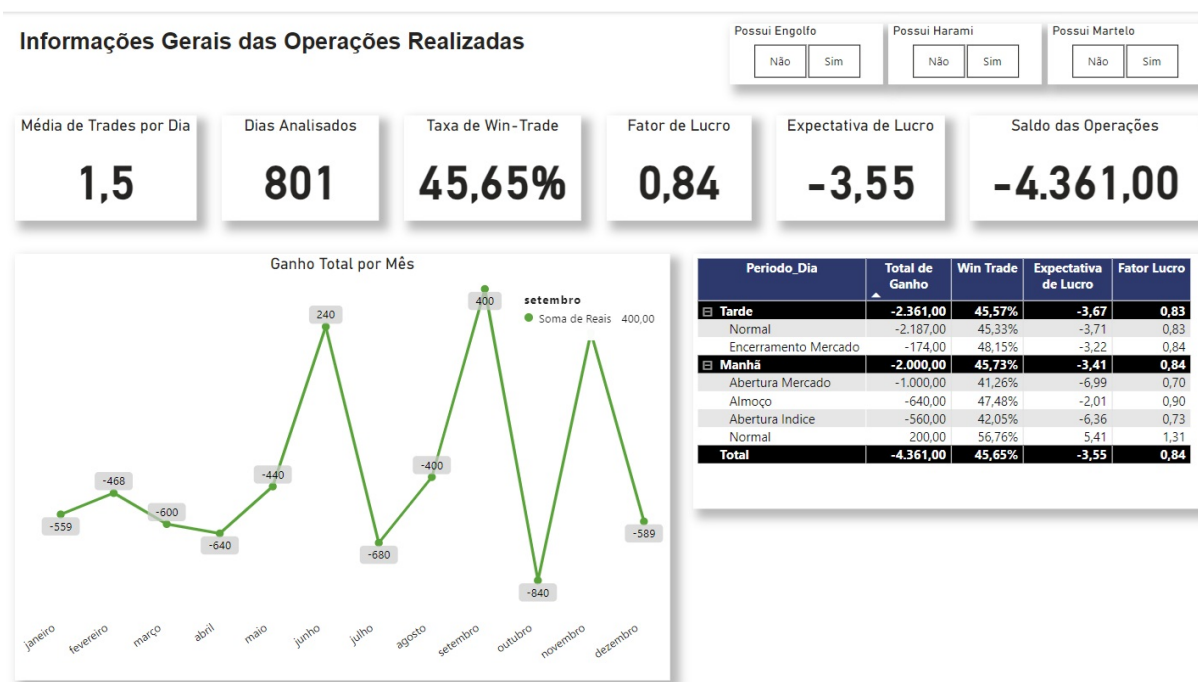
Além disso, existem duas estratégias incomuns. A primeira prioriza o *gain*, com uma proporção 5:1 (5 ganhos para 1 perda), onde o *gain* é de 500 pontos e o *loss* de 100 pontos. A segunda é o inverso, com proporção 1:2 (1 ganho para 2 perdas), cujo objetivo é dar mais tempo para alcançar o *gain*. Nessa estratégia, o *gain* é de 250 pontos, considerado razoavelmente alcançável, enquanto o *loss* é de 500 pontos, que exige mais *candles*.

Com as regras da operação definidas e as estratégias já desenvolvidas, foi mencionado que o *dataset* de operação contém 37 colunas, correspondentes às características do padrão de *candle* que originou cada operação. Além dessas, há outras variáveis relevantes. Neste trabalho, são explicadas aquelas que exigem uma descrição mais detalhada, enquanto outras, como pontos, valores em reais, valor de entrada e saída, são autoexplicativas.

- **"Operação"**: identifica o tipo de operação realizada, compra ou venda.
- **"Trading"**: identifica o tipo de *trading* que ocorreu: stop loss, stop gain ou encerramento de fechamento.
- **"Natureza"**: identifica se a operação foi de lucro ou de perda, sendo esta a variável **target** dos modelos de classificação.
- **"Tempo Encontra Trading"**: esta variável conta quanto tempo foi aguardado até ocorrer a operação. ela conta os *candles*, sendo que o número 10 representa 10 *candles* de 15 minutos, ou seja, 150 minutos. vale ressaltar que valores maiores que 36 indicam que levou um dia completo.
- **"Tempo Total"**: indica quanto tempo durou a operação, desde o *candle* de referência, passando pela confirmação, até atingir o alvo de *gain*, *loss* ou o encerramento pelo horário. essa variável também é medida em *candles* de 15 minutos.
- **"Drawdown"**: indica o maior risco de perda naquele momento, sem contar o *stop loss*. refere-se ao risco que a operação enfrentava caso o *stop loss* não fosse acionado.

Deste modo, com as operações simuladas, podemos analisar o *dataset* de operações. Utilizando o Power BI, cria-se um *dashboard* para cada estratégia, conforme ilustrado na [Figura 20](#). Todos seguem o mesmo modelo. Como há 6 estratégias, a [Figura 21](#) resume as análises realizadas.

Figura 20 – Modelo de *dashboard* das operações da estratégia *gain* e *loss* de 200 pontos.



Fonte: Autoria própria.

Figura 21 – Resumo financeiro das estratégias no dataset de treino.

Estratégia	Nro Tradings	Trade por Dia	Win Trade	Saldo	Lucro por Operacao	Fator Lucro
Gain: 200 Pontos - Loss: 200 Pontos	1229	1,5	45,65%	-4.361,00	-3,55	0,84
Gain: 200 Pontos - Loss: 100 Pontos	1267	1,6	34,89%	1.034,00	0,82	1,06
Gain: 500 Pontos - Loss: 500 Pontos	905	1,3	49,50%	-888,00	-0,98	0,98
Gain: 500 Pontos - Loss: 250 Pontos	1080	1,4	33,15%	-1.681,00	-1,56	0,95
Gain: 500 Pontos - Loss: 100 Pontos	1207	1,5	20,13%	3.862,00	3,20	1,20
Gain: 250 Pontos - Loss: 500 Pontos	1055	1,4	63,89%	-2.745,00	-2,60	0,92

Fonte: Autoria própria.

Quando se analisam os valores positivos na [Figura 21](#), principalmente na estratégia de 500 pontos de *gain* e 100 de *loss*, e a [Figura 22](#), que mostra a evolução desta carteira, é fácil acreditar que vencer o mercado é simples: basta aumentar a proporção de ganho em relação à perda.

Contudo, dois pontos devem ser considerados. Primeiro, o saldo apresentado reflete o lucro bruto, ou seja, a diferença entre ganhos e perdas, sem considerar taxas e tarifas da B3, impostos sobre operações vencedoras e eventuais taxas de corretagem cobradas por algumas corretoras. Portanto, os números reais não são tão elevados, e é importante deixar isso claro para evitar ilusões com margens aparentemente altas.

Figura 22 – Evolução da Carteira - Estratégia 5:1 nos dados de treino.



Fonte: Autoria própria.

Figura 23 – Estatística descritiva da Carteira 5:1 nos dados de treino.

```
Out[24]: count    1207.000000
         mean     3073.686827
         std      1226.556021
         min       400.000000
         25%      2402.000000
         50%      3526.000000
         75%      4026.000000
         max      5022.000000
         Name: Lucro Acumulado Reais, dtype: float64
```

Fonte: Autoria própria.

O segundo ponto é a volatilidade. Para exemplificar, considere a estratégia 5:1 ou 500 pontos de ganhos para 100 pontos de perdas, embora a carteira tenha iniciado em 2018 e, ao final de 2022, o lucro bruto fosse de R\$ 3.862, é importante observar o período de 2018 a 2019 na [Figura 22](#). Considerando um investimento inicial de R\$ 1.000, boa parte desse período apresentou perdas, com a carteira atingindo R\$ 400 no valor mínimo, conforme indicado na [Figura 23](#). Isso representa um prejuízo de 60%. Surge, então, a pergunta: o *trader* continuaria com essa abordagem até alcançar o período de 2022?

Caso a operação fosse mantida, com investimento inicial de R\$ 1.000, o lucro bruto ao final de 2022 seria de R\$ 4.862. No entanto, mesmo separando os dados de teste dos de treino, a [Figura 24](#), que inclui os dados de 2018 a 2022 e a continuação da operação em 2023, mostra uma queda acentuada. Em 2023, houve uma perda superior a R\$ 1.000. Quem começou entre 2020 e 2022 teria pouco lucro, enquanto quem iniciou em 2023 não apenas perderia todo o capital investido, mas também sairia devendo dinheiro à corretora. Em mercados futuros, é possível acumular dívidas após perder todo o capital.

Além disso, esta abordagem é apenas um exemplo. A Figura 25 demonstra que todas as outras estratégias também enfrentaram dificuldades em 2023, evidenciando que o cenário adverso impactou todas de forma semelhante.

Figura 24 – Evolução da Carteira - Estratégia 5:1 nos dados de treino e de teste.



Fonte: Autoria própria.

Figura 25 – Resumo financeiro das estratégias no *dataset* de teste.

Estratégia	Nro Tradings	Trade por Dia	Win Trade	Saldo	Lucro por Operacao	Fator Lucro
Gain: 200 Pontos - Loss: 200 Pontos	242	0,98	45,87%	-815,00	-3,37	0,84
Gain: 200 Pontos - Loss: 100 Pontos	250	1,01	30,40%	-461,00	-1,84	0,87
Gain: 500 Pontos - Loss: 500 Pontos	171	0,69	47,37%	-1.080,00	-6,32	0,87
Gain: 500 Pontos - Loss: 250 Pontos	208	0,84	30,29%	-1.457,00	-7,00	0,80
Gain: 500 Pontos - Loss: 100 Pontos	234	0,95	14,10%	-1.110,00	-4,74	0,72
Gain: 250 Pontos - Loss: 500 Pontos	200	0,81	63,00%	-850,00	-4,25	0,88

Fonte: Autoria própria.

### 3.4 PREPARAÇÃO DOS DADOS

Com o *dataset* de operações criado, a Figura 21 e a Figura 25, na coluna "Nro Tradings", mostram a quantidade de linhas ou operações presentes em cada *dataset* de treino e teste dessas estratégias. Cada operação contém, ao todo, 49 variáveis ou colunas.

Nem todas essas variáveis podem ser utilizadas nos modelos de classificação, pois é necessário tratá-las para evitar *data leakage* (vazamento de informações), o que poderia enviesar os modelos. Além disso, é essencial aprimorar algumas variáveis, verificar seu poder de predição, e analisar problemas relacionados a *outliers*, multicolinearidade e tipos de dados presentes em cada variável.

O primeiro tratamento realizado consistiu na exclusão de variáveis que poderiam causar *data leakage*. A variável *target* é representada pela coluna "Natureza", que contém informações sobre se a operação gerou lucro ou prejuízo. Entretanto, outras variáveis indicam, de forma explícita, o resultado da operação. Por exemplo, a coluna "Pontos" informa a quantidade de

pontos ganhos ou perdidos. No entanto, para determinar se houve ganho ou perda, a operação precisa estar concluída.

Ao concluir a operação, já se sabe se houve lucro ou prejuízo, tornando a variável "Pontos" inviável para coleta antes do início da operação. Dessa forma, essa variável não pode ser utilizada para prever se a operação será lucrativa.

Nesse contexto, é fundamental identificar quais variáveis podem ser coletadas em um cenário simulado em tempo real, antes do início de uma operação, e utilizar essas informações para prever o resultado da operação. Após a análise, identificou-se que um total de 12 variáveis causava vazamento de informações. Dessa forma, o *dataset* final foi reduzido para 37 variáveis.

Outro tratamento relevante diz respeito aos indicadores *MACD*, *RSI*, *HiLo*, Estocástico e Médias Móveis. A maioria desses indicadores apresentava apenas valores numéricos no *dataset*. Naturalmente, o cálculo desses indicadores é baseado nos preços dos *candles*. No entanto, para facilitar a identificação de relações pelos modelos, optou-se por simplificar esses indicadores em estratégias comumente utilizadas.

Por exemplo, o *MACD* era originalmente representado por duas variáveis no *dataset*: "MACD Line" e "MACD Signal". Esse indicador é amplamente utilizado para identificar tendências e fornecer sinais de compra e venda. Para evitar que o modelo precisasse analisar relações internas entre as duas variáveis e com outras informações, essas foram substituídas por novas variáveis categóricas, permitindo que o modelo trabalhasse diretamente com as condições interpretadas.

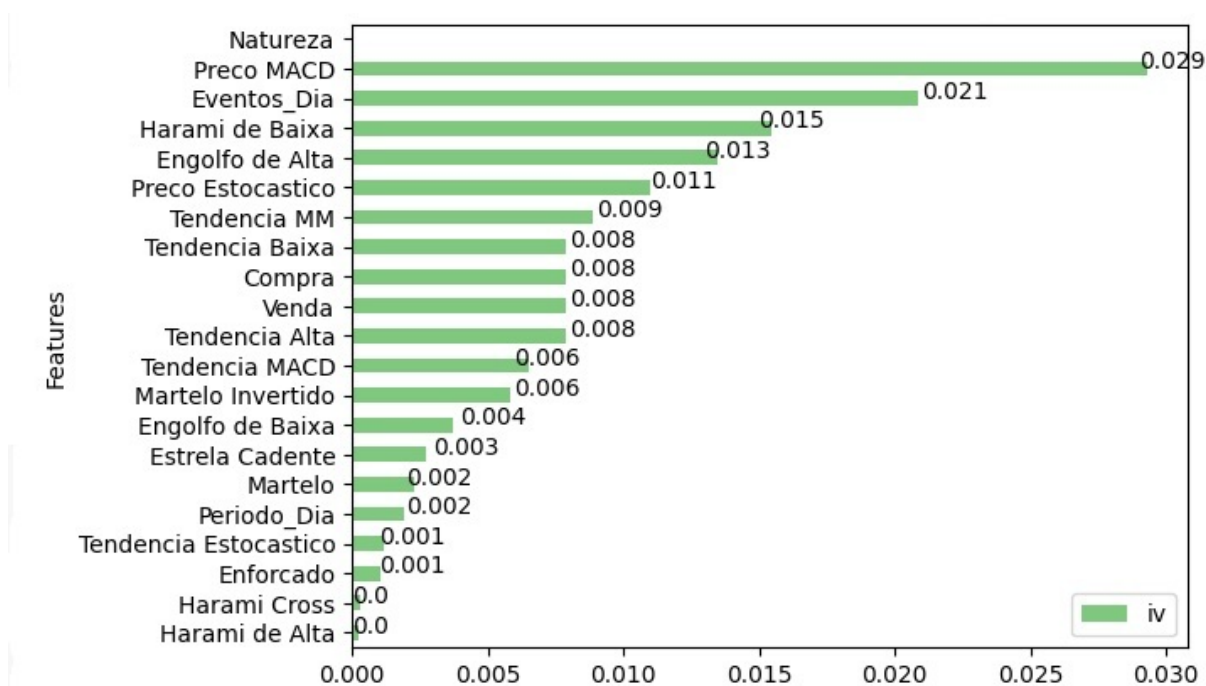
Conforme descrito no artigo da [Nelogica \(2020\)](#), "MACD Line" maior que zero indica uma tendência de alta, enquanto valores menores que zero indicam uma tendência de baixa. Assim, os muitos valores únicos dessa variável foram transformados em apenas dois, correspondentes às tendências identificadas.

Além disso, segundo o mesmo artigo, "MACD Line" maior que "MACD Signal" representa um sinal de compra, enquanto "MACD Line" menor que "MACD Signal" indica um sinal de venda. Com isso, os valores únicos das variáveis originais foram consolidados em uma nova variável com apenas dois estados: compra ou venda. Esse procedimento foi replicado para os demais indicadores, reduzindo a complexidade dos dados e tornando-os mais adequados para os modelos de classificação.

Separando as variáveis, vamos analisar somente as variáveis categóricas. Temos, no total, 21 variáveis categóricas. Utilizando as medidas *IV (Information Value)* e *WOE (Weight of Evidence)*, é possível verificar se essas variáveis categóricas influenciam os modelos, conforme representado na [Figura 26](#).

Dentre as variáveis, somente "Preço Estocástico" apresenta uma predição considerada fraca; todas as outras variáveis são classificadas como insignificantes. Contudo, é importante considerar as diferentes estratégias e verificar quais variáveis apresentam boa predição em cada caso.

De modo geral, manteremos as variáveis que identificam os padrões de *candle*, além de pelo menos uma variável que represente os indicadores. Nesta análise do *IV*, considerando as seis estratégias, cinco variáveis tiveram menor peso total.

Figura 26 – *Information Value* das variáveis categóricas na estratégia *gain* e *gain* de 500 pontos.

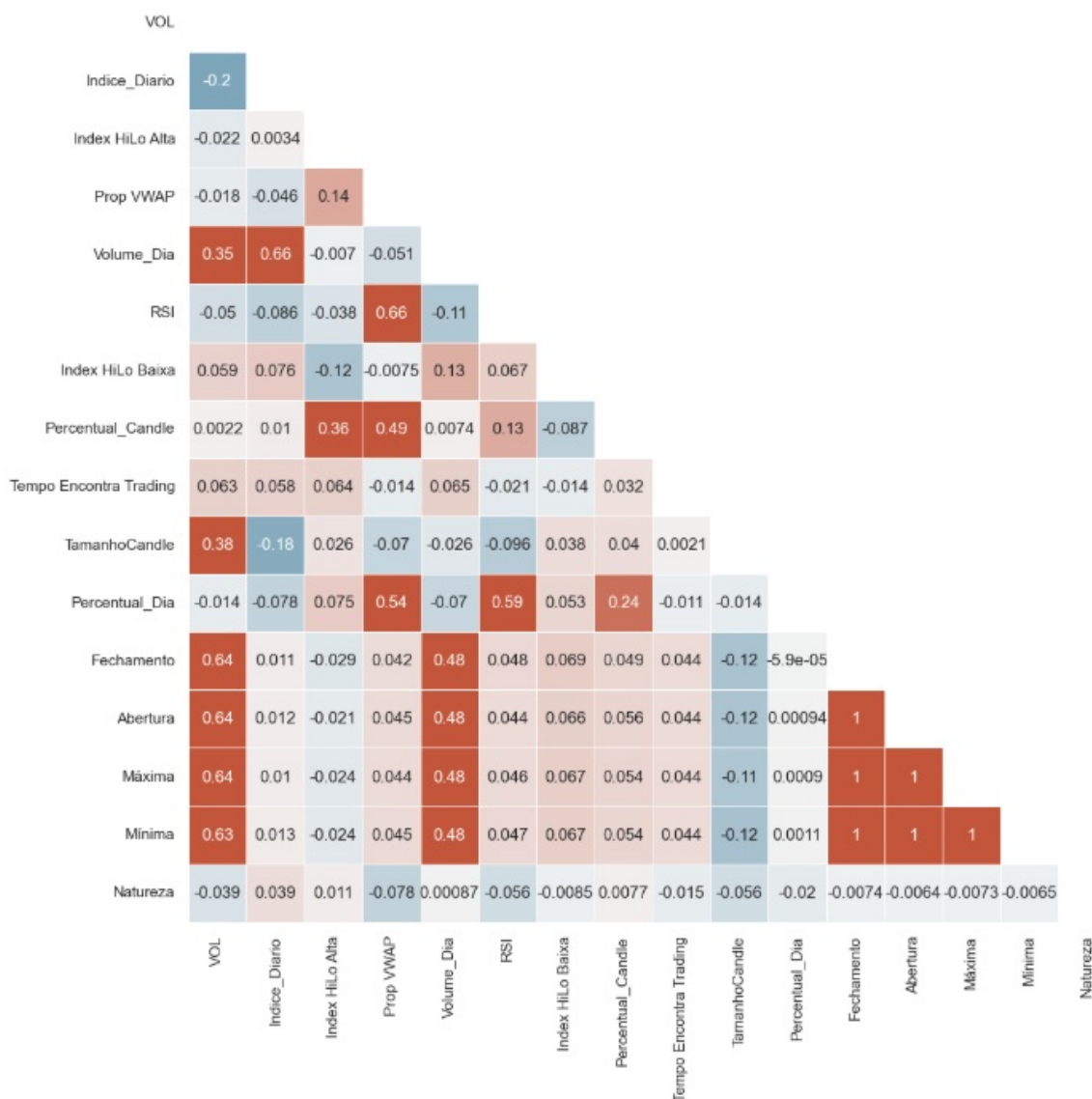
Fonte: Autoria própria.

Para as variáveis numéricas, a primeira análise realizada foi a correlação entre elas. Utilizando a correlação de Pearson, obteve-se a matriz de correlação mostrada na [Figura 27](#). A principal observação refere-se à multicolinearidade entre as variáveis "Abertura", "Máxima", "Mínima" e "Fechamento". Essas variáveis apresentam correlação perfeita, com valor igual a 1, devido ao fato de refletirem a valorização do preço do *candle*, resultando em comportamentos muito similares.

Manter as quatro variáveis no modelo não traria benefícios, podendo até tornar o algoritmo mais lento, especialmente em modelos baseados em árvores de decisão, que são sensíveis à presença de muitas variáveis. Por essa razão, optou-se por manter apenas a variável "Abertura", que é a primeira disponível em tempo real e, portanto, fácil de ser coletada.

Ressalta-se que essas características se referem ao *candle* de referência; além dele, uma confirmação é necessária. Assim, antes do início de uma operação, essas informações já estão disponíveis, o que significa que as características dos *candles* não configuram vazamento de dados.

Figura 27 – Matriz de correlação variáveis numéricas na estratégia *gain* e *gain* de 500 pontos.



Fonte: Autoria própria.

Outra análise importante para as variáveis numéricas refere-se à detecção de *outliers*. A Figura 28 mostra um *dataframe* criado para identificar *outliers* de maneira semelhante ao *boxplot*. Utilizou-se o *IQR* (Intervalo Interquartil), calculado pela subtração entre o terceiro e o primeiro quartil. Tradicionalmente, o *IQR* é multiplicado por 1,50 para definir os limites superior e inferior de detecção. No entanto, considerando que algumas variáveis analisadas são percentuais e que, no mercado financeiro, variações superiores a 10% já podem ser consideradas atípicas, adotou-se, de forma metodológica, um fator mais restritivo de 1,10. Essa decisão visa tornar o critério de detecção mais sensível ao contexto específico dos dados financeiros.

Figura 28 – Dataframe de *Outliers* na estratégia *gain* e *gain* de 500 pontos.

Nome de Coluna	Index HiLo Alta	Percentual_Candle	Index HiLo Baixa	Percentual_Dia	Tempo Encontra Trading	TamanhoCandle	Prop VWAP	Volume_Dia	VOL	RSI	Abertura	Indice_
Min	0.0	-3.31	0.0	-5.55	2.0	25.0	-0.067	0.0	0.0	15.0	63700.0	
Q1	0.0	-0.11	0.0	-0.51	11.0	225.0	-0.003	3.0	3.0	41.0	93295.0	
Median	0.0	-0.02	2.0	0.0	23.0	310.0	0.0	8.0	7.0	50.0	104815.0	
Q3	6.0	0.08	10.0	0.49	44.0	450.0	0.003	18.0	11.0	59.0	114080.0	
Max	55.0	1.36	970.0	7.17	204.0	2090.0	0.047	54.0	30.0	87.0	130905.0	
IQR	6.0	0.19	10.0	1.0	33.0	225.0	0.006	15.0	8.0	18.0	20785.0	
Lower fence	-6.6	-0.319	-11.0	-1.61	-25.3	-22.5	-0.0096	-13.5	-5.8	21.2	70431.5	
Upper fence	12.6	0.289	21.0	1.59	80.3	697.5	0.0096	34.5	19.8	78.8	136943.5	
Skew	2.57603	-4.26181	6.751053	0.362016	2.088376	2.157277	-1.54807	1.093622	0.916465	-0.006088	-0.450425	0.0
Num_Outliers	118	83	82	81	75	65	55	42	35	20	10	
Percent_Outliers	0.08%	0.06%	0.05%	0.05%	0.05%	0.04%	0.04%	0.03%	0.02%	0.01%	0.01%	

Fonte: Autoria própria.

No tratamento de valores discrepantes, optou-se por não removê-los, visto que representam uma proporção muito pequena do conjunto de dados, inferior a 0,1%, além de serem considerados relevantes para a análise. Destacam-se, nesse contexto, as variáveis "Index Hilo Alta" e "Index Hilo Baixa", que indicam a duração de uma tendência de alta ou baixa, respectivamente. Merece atenção também a variável "Tempo Encontra Trading", que informa o número de *candles* entre o encerramento de uma operação e o início da próxima. Essas variáveis foram convertidas em categóricas, com quatro faixas de classificação: "rápido", "normal", "longa" e "muito longa".

A etapa final do pré-processamento envolveu a normalização dos dados numéricos, também conhecida como *scaling*. Esse procedimento foi necessário devido à ampla variação nos valores das variáveis, com alguns extremamente altos e outros muito baixos. Utilizou-se a função `RobustScaler` da biblioteca `scikit-learn`, que utiliza os quartis em vez da média e do desvio padrão, oferecendo maior robustez diante de valores discrepantes. Esse processo foi aplicado exclusivamente às variáveis numéricas, excluindo a variável `target`.

Com os dados tratados, os *datasets* foram divididos entre treino e teste, mantendo as separações feitas anteriormente. Após o pré-processamento, as bases mantiveram o número original de linhas, mas passaram a conter 28 variáveis. As variáveis categóricas foram convertidas para valores numéricos, permitindo a aplicação dos modelos de classificação.

### 3.5 MODELAGEM

Nesta seção, são apresentadas as etapas de escolha do modelo de aprendizagem, a estratégia adotada para o balanceamento das classes e a definição da validação cruzada utilizada na avaliação dos modelos. Os algoritmos testados incluem: KNN (*K-Nearest Neighbors*), SVC (*Support Vector Machine Classifier*), *Decision Tree*, *Random Forest*, *ExtraTrees*, *AdaBoost*, *GradientBoost*, *XGBoost* e *LightGBM*.

Após a importação dos dados já preparados, estes são divididos entre as variáveis preditoras, representadas por  $X$ , e a variável alvo, representada por  $y$ . Em seguida, é realizada a verificação do balanceamento das classes na variável alvo, ou seja, da proporção entre operações com lucro e com prejuízo.

Das seis estratégias analisadas, apenas uma apresenta balanceamento de 50% entre as classes — a estratégia com 500 pontos de *gain* e *loss*. As demais estratégias apresentam

desbalanceamento, o que demanda a adoção de uma abordagem específica para o tratamento desse problema.

O balanceamento das classes é uma etapa fundamental na construção de modelos de classificação, especialmente quando há predominância de uma classe sobre a outra. Duas abordagens principais são comumente utilizadas para lidar com esse problema: o ajuste de pesos nos algoritmos de classificação e o uso de técnicas de *resampling*.

A primeira abordagem consiste em recorrer aos mecanismos internos de balanceamento oferecidos por alguns algoritmos, os quais ajustam os pesos das classes, atribuindo maior peso à classe minoritária. No entanto, essa funcionalidade não está disponível em todos os modelos — algoritmos como KNN, *AdaBoost* e *GradientBoost* não oferecem esse recurso. Além disso, mesmo entre os modelos que suportam esse tipo de balanceamento, foi constatado que, em todos os casos analisados, o classificador gerado tendia a prever apenas uma das classes.

Como alternativa, podem ser aplicadas técnicas de *under resample* ou *over resample*. A técnica de *under resample* busca reduzir a classe majoritária até que ambas as classes tenham quantidades equivalentes de amostras. Entre as estratégias disponíveis, destacam-se o uso das funções `RandomUnderSampler` e `NearMiss`. A primeira realiza a remoção aleatória de registros da classe majoritária, enquanto a segunda aplica critérios baseados na distância entre os dados para selecionar quais registros manter.

Já o *over resample* consiste em aumentar a quantidade de amostras da classe minoritária. As técnicas mais utilizadas nesse caso incluem `RandomOverSampler`, que replica aleatoriamente registros existentes, e *Synthetic Minority Over-sampling Technique* (SMOTE), que gera registros sintéticos com base na interpolação entre vizinhos próximos. No entanto, em contextos como o mercado financeiro, marcado por alta volatilidade e ruído nos dados, o uso de registros artificiais pode representar um risco, uma vez que os padrões gerados podem não refletir situações reais observadas no histórico.

Considerando essas características, optou-se pelo uso do `RandomUnderSampler`, com o objetivo de equilibrar o conjunto de dados por meio da redução da classe majoritária. Essa escolha também foi motivada por testes preliminares com o SMOTE, que indicaram tendência à superestimação de padrões irreais, comprometendo o desempenho dos modelos.

Diante desse contexto, diversos testes indicaram que a técnica de *under resample* proporciona um aprendizado mais eficiente para os modelos. As métricas associadas às operações financeiras geralmente não apresentam resultados expressivos devido à alta volatilidade do mercado. Portanto, mesmo com grande volume de dados, o modelo não seria capaz de capturar todas as possíveis situações. Nesse sentido, a desvantagem do *under resample* torna-se menos relevante em comparação ao *over resample*, uma vez que modelos treinados com dados reais, ainda que em menor quantidade, apresentam desempenho mais confiável nos dados de teste. Por outro lado, modelos treinados com grande volume de dados duplicados ou sintéticos podem exibir resultados artificialmente elevados nas métricas de treino, porém tendem a falhar na generalização para novos dados. Dessa forma, nesta análise específica, foi adotada a abordagem que prioriza dados reais, mesmo que em menor volume.

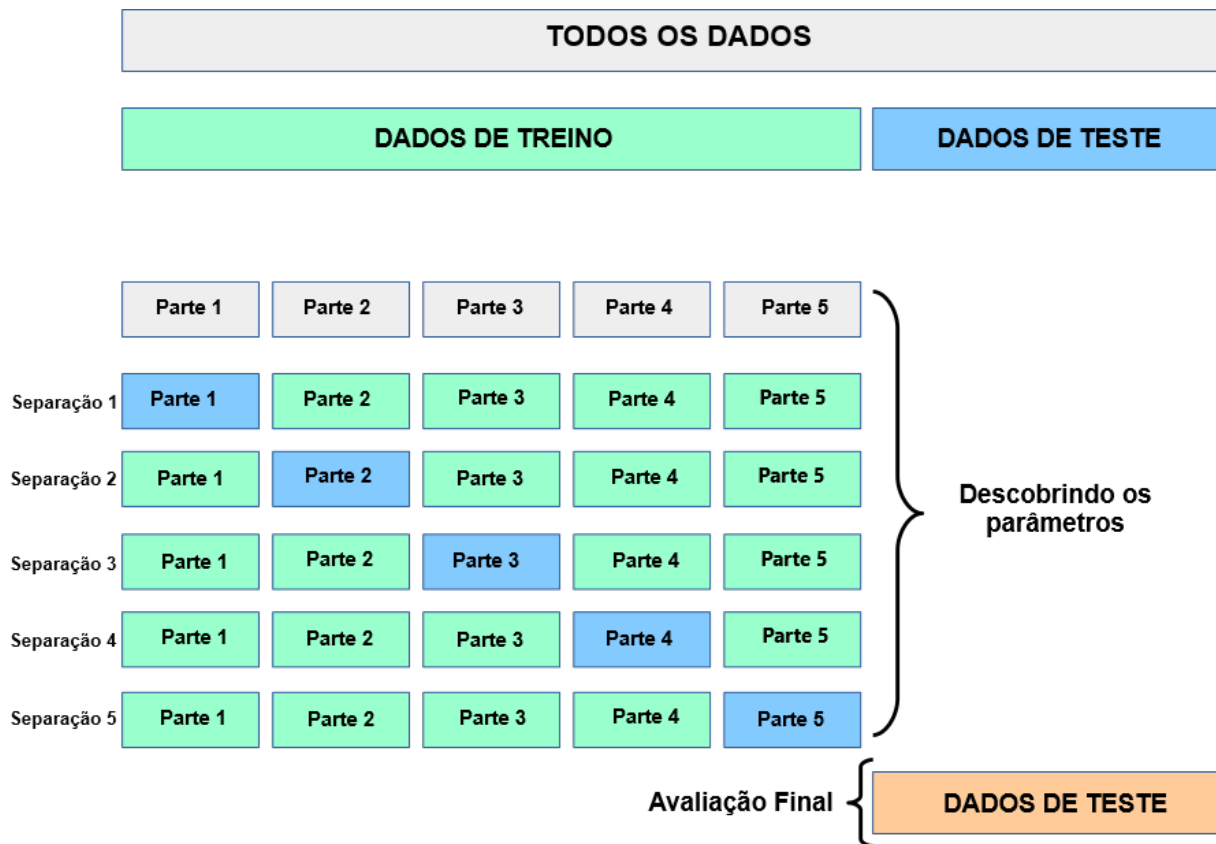
Para a avaliação dos modelos, foram consideradas as técnicas de *hold-out* e validação cruzada. A técnica de *hold-out* consiste na divisão dos dados brutos em três conjuntos: treino, validação e teste. Entretanto, essa abordagem é sensível à aleatoriedade, o que pode comprometer sua eficácia em datasets com elevado ruído e volatilidade. Isso ocorre porque a amostra selecionada pode ser muito favorável ou, inversamente, desfavorável a uma situação específica. Em contraste, a validação cruzada divide o conjunto de treino em cinco ou dez partes, mitigando essa sensibilidade e proporcionando maior robustez e confiança na avaliação dos modelos.

Existem diferentes técnicas de validação cruzada que podem ser aplicadas, entre elas

o *KFold*, o *Stratified KFold* e a validação cruzada para séries temporais. O método *KFold* divide o conjunto de dados em partes iguais, usando uma delas para teste e as demais para treino. Já o *Stratified KFold* mantém a proporção original das classes em cada divisão, sendo recomendado para conjuntos de dados desbalanceados, como no caso desta análise.

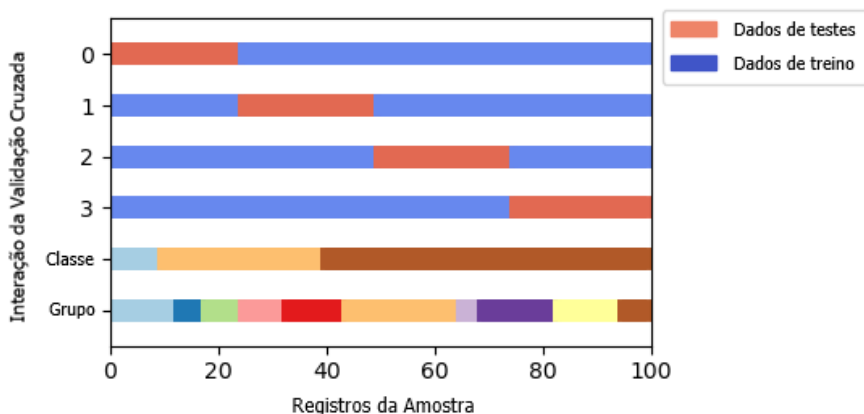
No presente estudo, o conjunto de dados possui algumas características temporais, mas não configura uma série temporal estrita. Por isso, não se aplicou a validação cruzada específica para séries temporais, que preserva a ordem cronológica dos dados para evitar o vazamento de informações. Essa distinção é importante para garantir a validade dos resultados, especialmente em situações onde a sequência temporal dos dados impacta diretamente a modelagem.

Figura 29 – Conceito da Validação Cruzada.



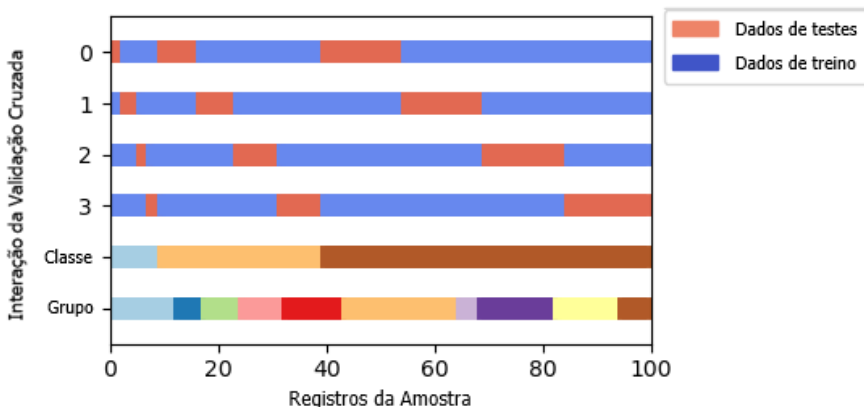
Fonte: (SCIKIT-LEARN, 2025).

Figura 30 – Validação Cruzada com separação *K-Fold*.



Fonte: (SCIKIT-LEARN, 2025).

Figura 31 – Validação Cruzada com separação *Stratified K-Fold*.



Fonte: (SCIKIT-LEARN, 2025).

Nesta etapa, os modelos foram inicialmente avaliados utilizando a validação cruzada do tipo *Stratified K-Fold*, com o objetivo de lidar com o desbalanceamento das classes presentes no conjunto de dados. Contudo, os resultados obtidos não foram satisfatórios, o que motivou a aplicação da técnica de *under-sampling* para balanceamento das classes.

Após o balanceamento dos dados, empregou-se a validação cruzada simples do tipo *K-Fold* para avaliar os modelos treinados. Com base nas métricas de desempenho obtidas, foi elaborado um ranking dos modelos para cada uma das seis estratégias analisadas. A seguir, os melhores modelos de cada estratégia foram selecionados para a etapa de otimização de hiperparâmetros.

Vale destacar que, para cada estratégia, há um ranking específico de modelos. A Figura 32 exemplifica o ranking referente à estratégia com 500 pontos de *gain* e *loss*.

Figura 32 – Rank dos modelos de classificação na estratégia *gain* e *loss* de 500 pontos.

Modelo	AUC-ROC	Acuracia
KNN	0.5534	0.5470
ExtraTrees	0.5377	0.5293
RandomForest	0.5291	0.5238
GradientBoost	0.5216	0.5127
AdaBoost	0.5162	0.5116
LightGBM	0.5112	0.4972
DecisionTree	0.4983	0.4983
XGBoost	0.4933	0.4906
NaiveBayes	0.4610	0.5094

Fonte: Autoria própria.

A otimização dos hiperparâmetros do modelo foi realizada por meio da otimização bayesiana, que utiliza inferência bayesiana para identificar os melhores parâmetros, tais como taxa de aprendizado, regularização e número de variáveis, entre outros, com o objetivo de aprimorar o desempenho do modelo.

A implementação inicial envolveu a definição da função objetivo, na qual o modelo é avaliado com seus parâmetros por meio de validação cruzada. A biblioteca *scikit-optimize* disponibiliza a função `gp_minimize` para essa finalidade. Nessa função, uma métrica de avaliação é selecionada; neste trabalho, a métrica escolhida foi a *auc-roc*. Em casos de falha de aprendizagem, quando o modelo tende a prever predominantemente uma única classe, podem ser utilizadas métricas alternativas, como a F1 ou a acurácia balanceada, esta última calculada pela média dos *recall* das duas classes.

A métrica selecionada é fornecida como valor negativo à função `gp_minimize`, uma vez que esta busca minimizar o valor da função objetivo. Essa função utiliza um método de otimização baseado em processos gaussianos para explorar o espaço dos hiperparâmetros e identificar combinações que promovam o melhor desempenho do modelo.

A segunda etapa consiste na definição do espaço de busca, especificando intervalos de valores ou categorias para os parâmetros a serem testados. Em seguida, invoca-se a função `gp_minimize`, estabelecendo o número de iterações, que, neste caso, foi definido como 100. Isso significa que a função realizará 100 avaliações de combinações de parâmetros dentro do espaço de busca definido. As primeiras 70 avaliações são realizadas de forma aleatória, enquanto as últimas 30 são conduzidas com base na inferência bayesiana, concentrando-se em regiões do espaço de parâmetros que demonstraram maior potencial de desempenho.

Na [Tabela 3](#), apresentam-se as seis estratégias e os três melhores modelos ranqueados para cada uma delas. Já a [Tabela 4](#) resume a frequência com que cada modelo foi selecionado. Observa-se que o modelo *LightGBM* não foi utilizado em nenhuma das estratégias analisadas. Por outro lado, os modelos que mais apareceram foram KNN, *Extra Trees* e *Random Forest*.

Modelos baseados em *ensemble* costumam apresentar bom desempenho. Entretanto, destaca-se a presença do algoritmo KNN entre os mais selecionados, ocupando, na maioria dos casos, a primeira posição no ranking. Essa ocorrência, mesmo após cuidados para evitar enviesamento em algoritmos baseados em vizinhos próximos, pode ser explicada pelo fato de os conjuntos de dados serem relativamente pequenos e conterem muitas variáveis categóricas. Adicionalmente, modelos *ensemble* tendem a apresentar melhor desempenho em conjuntos de dados maiores e com maior quantidade de variáveis numéricas.

Tabela 3 – Estratégias das operações e os 3 melhores modelos raqueados.

Estratégia	1º Modelo	2º Modelo	3º Modelo
Gain: 200 Pontos - Loss: 200 Pontos	KNN	XGBoost	Random Forest
Gain: 200 Pontos - Loss: 100 Pontos	Ada Boost	Gradient Boost	Random Forest
Gain: 500 Pontos - Loss: 500 Pontos	KNN	Extra Trees	Random Forest
Gain: 500 Pontos - Loss: 250 Pontos	Extra Trees	KNN	SVC
Gain: 500 Pontos - Loss: 100 Pontos	Ada Boost	Decision Tree	Gradient Boost
Gain: 250 Pontos - Loss: 500 Pontos	KNN	Extra Trees	XGBoost

Fonte: Elaborado pelo próprio autor

Tabela 4 – Resumo dos modelos raqueados

Modelo	1º Modelo	2º Modelo	3º Modelo	Total
KNN	3	1	0	4
Extra Trees	1	2	0	3
Random Forest	0	0	3	3
Ada Boost	2	0	0	2
Gradient Boost	0	1	1	2
XGBoost	0	1	1	2
SVC	0	0	1	1
Decision Tree	0	1	0	1
LightGBM	0	0	0	0

Fonte: Elaborado pelo próprio autor

### 3.6 AVALIAÇÃO DOS MODELOS

A métrica principal utilizada para avaliação dos modelos foi a *roc-auc score*, devido à sua capacidade de fornecer uma visão abrangente da performance considerando diferentes limiares de classificação. Além disso, as métricas de *precision* e *recall* foram analisadas de forma complementar, uma vez que, no contexto das estratégias avaliadas, o *recall* é especialmente relevante para a classe positiva, enquanto a *precision* assume maior importância para a classe negativa.

Para complementar a avaliação quantitativa, foram utilizadas análises visuais, incluindo as curvas ROC e as matrizes de confusão, que possibilitam uma interpretação mais detalhada do desempenho dos modelos.

No capítulo seguinte, será realizada uma análise aprofundada dos resultados obtidos, destacando o impacto das diferentes métricas na escolha do modelo mais adequado para cada estratégia.

## 4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Nesta seção, são apresentados os resultados obtidos pelos modelos em cada uma das estratégias, com o objetivo de identificar qual modelo apresentou o melhor desempenho, se houve melhorias nas operações e se a adoção da estratégia é justificada com base nos resultados obtidos.

Ao final, é realizada uma síntese comparativa entre os melhores modelos de cada estratégia, avaliando qual delas demonstrou maior eficácia, bem como a viabilidade de sua aplicação em tempo real.

### 4.1 GAIN: 500 PONTOS - LOSS: 500 PONTOS

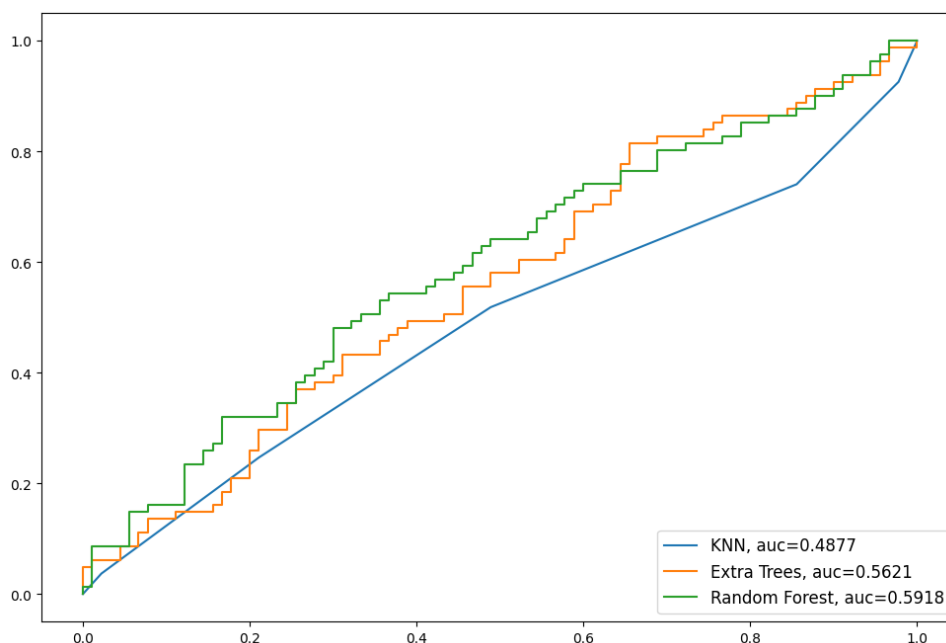
A estratégia de 500 pontos no WIN foi previamente analisada ao longo do trabalho. Em termos de lucro bruto, esse valor equivale a aproximadamente R\$100, o que torna os cálculos mais diretos. Embora se trate de um alvo elevado, sua adoção como meta diária é plausível, considerando a volatilidade característica do ativo.

A [Figura 32](#) apresenta o desempenho dos modelos com base nos dados de treino e validação cruzada, enquanto a [Figura 33](#) exibe os resultados obtidos com os dados de teste. Na curva ROC, valores próximos de 0,50 indicam desempenho comparável ao de um classificador aleatório; abaixo desse limiar, o modelo se aproxima do comportamento de um classificador *dummy*. O modelo *KNN*, por exemplo, apresentou desempenho satisfatório nos dados de treino, mas ficou abaixo desse limite nos testes. Em contrapartida, o *Random Forest* obteve um valor de *AUC-ROC* igual a 0,5918.

Esse resultado pode ser considerado aceitável, dadas as características do problema e as dificuldades associadas à previsão desse tipo de operação. A elevada variabilidade dos dados e a limitação na capacidade preditiva das variáveis dificultam a obtenção de valores de *AUC* próximos de 1,0. Além do bom desempenho na curva ROC, o modelo *Random Forest* também apresentou superioridade em métricas como Acurácia e Precisão. A matriz de confusão dos três modelos analisados é apresentada na [Figura 34](#).

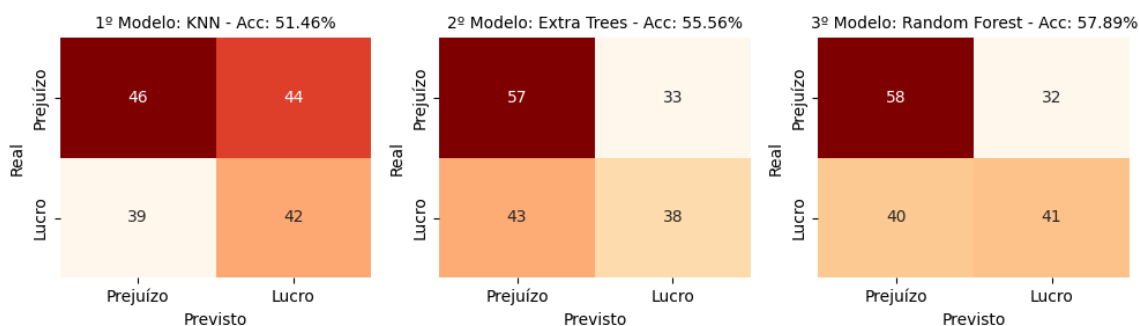
Adicionalmente, é importante observar que, quando o modelo indica uma previsão de prejuízo, a operação tende a ser evitada. Esse comportamento, por si só, já representa uma vantagem em relação à execução de operações sem apoio de modelos preditivos. Outro fator relevante é que as taxas operacionais não foram consideradas diretamente nas análises. Como essas taxas incidem sobre cada operação realizada, estratégias que resultam em menor volume de operações podem, indiretamente, reduzir custos operacionais e, assim, melhorar a rentabilidade líquida.

Figura 33 – Curvas ROC dos modelos de classificação na estratégia *gain* e *loss* de 500 pontos.



Fonte: Autoria própria.

Figura 34 – Matriz de Confusão nos dados de teste na estratégia *gain* e *loss* de 500 pontos.



Fonte: Autoria própria.

A análise da [Figura 34](#) revela que o modelo *Extra Trees* apresenta boa acurácia geral, atribuída, em grande parte, à sua capacidade de prever corretamente as operações de prejuízo. No entanto, esse modelo demonstrou desempenho inferior na identificação de operações lucrativas, sendo o que obteve menor acerto nessa categoria.

O modelo *KNN*, por sua vez, evidenciou maior capacidade de identificar corretamente operações de lucro, superando os demais modelos nesse aspecto, ainda que tenha apresentado menor equilíbrio entre as classes.

O modelo *Random Forest* obteve o melhor desempenho global, destacando-se pelo equilíbrio nas previsões das duas classes. Esse modelo foi o que mais se aproximou de reduzir simultaneamente os erros de falsos positivos e falsos negativos, mostrando-se o mais eficaz na tarefa de classificação entre os avaliados.

A [Tabela 5](#) apresenta a comparação dos resultados financeiros das estratégias, considerando apenas a diferença entre o lucro e o prejuízo bruto. A tabela fornece uma visão

clara da rentabilidade obtida por cada modelo. Em termos de desempenho tanto no processo de aprendizagem quanto nos resultados financeiros, o modelo *Random Forest* apresentou os melhores resultados entre os avaliados.

Tabela 5 – Resumo dos modelos usados na estratégia *gain* e *loss* de 500 pontos.

Operação	Tradings	Win-Rate	Saldo	Lucro por Operação
Sem Modelo	171	47,37%	- 1.080,00	- 6,32
Com KNN	86	48,84%	-382,00	-4,44
Com Extra Trees	71	53,52%	386,00	5,44
<b>Com Random Forest</b>	<b>73</b>	<b>56,16%</b>	<b>734,00</b>	<b>10,05</b>

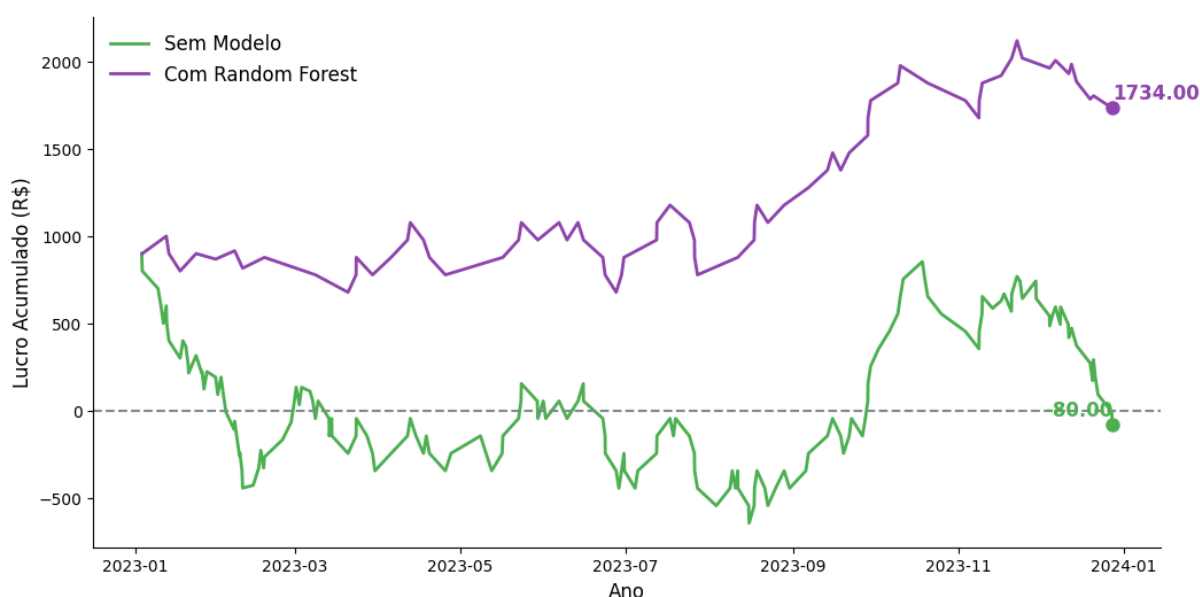
Fonte: Elaborado pelo próprio autor

Na [Figura 34](#), os valores indicam a quantidade de operações previstas, enquanto a [Tabela 5](#) apresenta um resumo financeiro referente apenas às operações consideradas lucrativas. Por exemplo, o modelo *Random Forest* previu lucro para 73 operações, das quais 41 resultaram em lucro verdadeiro, correspondendo a um *win-rate* de 56,16%. Isso significa que, entre as operações previstas como lucro, mais da metade foi efetivamente lucrativa.

A tabela financeira complementa essa análise ao apresentar o saldo final das operações, que foi de R\$ 734,00, representando um lucro médio bruto de aproximadamente R\$ 10,05 por operação, sem considerar taxas ou impostos. Esse resumo possibilita avaliar não apenas a quantidade de acertos do modelo, mas também seu impacto econômico, facilitando a comparação entre as estratégias.

E com relação à volatilidade, observa-se que o saldo da estratégia utilizando o modelo é positivo. Considerando um investimento inicial de R\$ 1.000,00, analisou-se o comportamento das operações ao longo do ano de 2023. Conforme indicado na [Tabela 5](#), a estratégia sem o uso do modelo resultaria em perda desse investimento. Em contrapartida, a aplicação do modelo *Random Forest* permitiria alcançar um saldo final de R\$ 1.734,00. Ressalta-se que esses valores são brutos, não considerando taxas ou impostos.

Figura 35 – Comparação de operações sem modelo e com modelo ao longo de 2023 na estratégia *gain* e *loss* de 500 pontos.



Fonte: Autoria própria.

Ao analisar a [Figura 35](#), observa-se a volatilidade em ação e como ela se intensifica significativamente na ausência do modelo. Sem sua utilização, o capital inicial teria sido completamente consumido já em fevereiro, exigindo novos aportes para continuidade das operações. Além disso, entre os meses de junho e setembro, registrou-se uma queda superior a R\$ 500,00, o que poderia comprometer novamente todo o capital disponível. Mesmo com uma recuperação parcial nos meses de outubro e novembro, o valor inicialmente investido não seria recuperado, sendo seguida por uma nova perda acentuada no mês seguinte.

Por outro lado, com o uso do modelo, embora também tenham ocorrido períodos de retração, nenhum deles resultou na perda total do capital durante o intervalo analisado. Evidentemente, trata-se de um período relativamente curto, mas é possível perceber que o modelo contribuiu para suavizar a volatilidade das operações em comparação à estratégia sem modelo. A [Tabela 6](#) reforça essa observação ao apresentar o desempenho financeiro mensal, evidenciando que, mesmo nos meses com prejuízo, as perdas foram substancialmente menores.

Por exemplo, em janeiro, o prejuízo sem o modelo foi de R\$ 909,00, enquanto com o modelo esse valor foi reduzido para R\$ 132,00 — uma diferença de quase sete vezes. Em dezembro, a perda sem o modelo atingiu R\$ 722,00, ao passo que, com o modelo, limitou-se a R\$ 287,00, correspondendo a uma proporção aproximada de 2,5 vezes. Essa diferença ilustra a principal vantagem da abordagem proposta: embora os ganhos não sejam extraordinários, o controle das perdas representa um benefício expressivo para a gestão do capital.

Tabela 6 – Resumo dos modelos usados com investimento inicial de R\$ 1.000,00 na estratégia *gain* e *loss* de 500 pontos.

Mês/Ano	Operações	Saldo Mensal	Saldo Total
01/2023	Sem Modelo	-909,00	91,00
01/2023	Com Modelo	-132,00	868,00
02/2023	Sem Modelo	-58,00	33,00
02/2023	Com Modelo	10,00	878,00
03/2023	Sem Modelo	-379,00	-346,00
03/2023	Com Modelo	-100,00	778,00
04/2023	Sem Modelo	100,00	-246,00
04/2023	Com Modelo	0,00	778,00
05/2023	Sem Modelo	300,00	54,00
05/2023	Com Modelo	200,00	978,00
06/2023	Sem Modelo	-400,00	-346,00
06/2023	Com Modelo	-100,00	878,00
07/2023	Sem Modelo	-100,00	-446,00
07/2023	Com Modelo	-100,00	778,00
08/2023	Sem Modelo	0,00	-446,00
08/2023	Com Modelo	400,00	1178,00
09/2023	Sem Modelo	700,00	254,00
09/2023	Com Modelo	600,00	1778,00
10/2023	Sem Modelo	300,00	554,00
10/2023	Com Modelo	100,00	1878,00
11/2023	Sem Modelo	88,00	642,00
11/2023	Com Modelo	143,00	2021,00
12/2023	Sem Modelo	-722,00	-80,00
12/2023	Com Modelo	-287,00	1734,00

Fonte: Elaborado pelo próprio autor

#### 4.2 GAIN: 200 PONTOS - LOSS: 200 PONTOS

Esta estratégia tem como objetivo reduzir o tempo das operações em comparação à abordagem anterior, baseada em alvos de 500 pontos para *gain* e *loss*. Conforme discutido anteriormente, os *candles* de 15 minutos apresentam, em média, cerca de 275 pontos, o que permite alcançar o alvo de forma relativamente rápida. No entanto, essa mesma característica também aumenta a suscetibilidade a perdas, como evidenciado pelo desbalanceamento de classes nesta configuração, em que aproximadamente 54% das operações resultam em prejuízo.

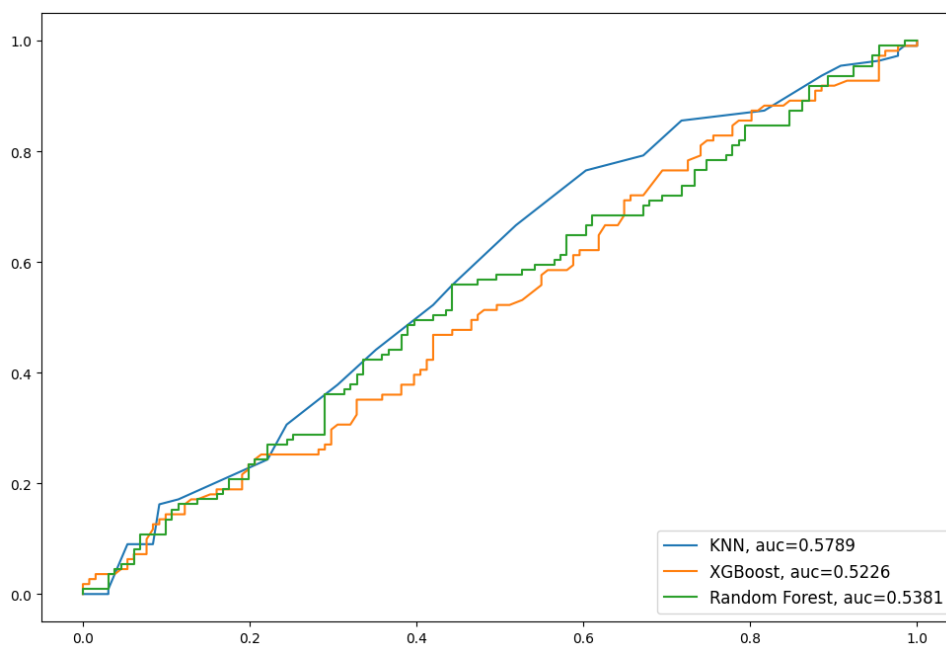
Em relação à estratégia anterior, uma diferença relevante é a substituição do modelo *Extra Trees* pelo *XGBoost*, embora a hierarquia de desempenho entre os modelos tenha se mantido semelhante. Outro aspecto a destacar é observado na [Figura 36](#), na qual o modelo *KNN* apresentou desempenho superior aos demais. Considerando a complexidade das operações e a elevada quantidade de variáveis envolvidas, os valores da AUC tendem a se concentrar

próximos de 0,50, o que torna inviável alcançar valores significativamente altos de forma consistente. Dessa forma, embora o modelo não apresente uma curva ROC ideal, o valor de AUC de 0,5789 pode ser considerado aceitável dentro desse contexto.

A Figura 37 mostra que, mesmo após a aplicação de técnicas de *under resampling*, os modelos continuaram a concentrar suas previsões em operações com prejuízo. Embora tenham apresentado bom desempenho na identificação de perdas, observou-se uma quantidade considerável de falsos positivos — casos em que o modelo previu lucro, mas a operação resultou em prejuízo. Esse tipo de erro foi particularmente recorrente no modelo *XGBoost*, que registrou 63 operações de falso lucro, superando as 57 de verdadeiro lucro, o que compromete a confiabilidade da estratégia e pode indicar um risco elevado de perdas.

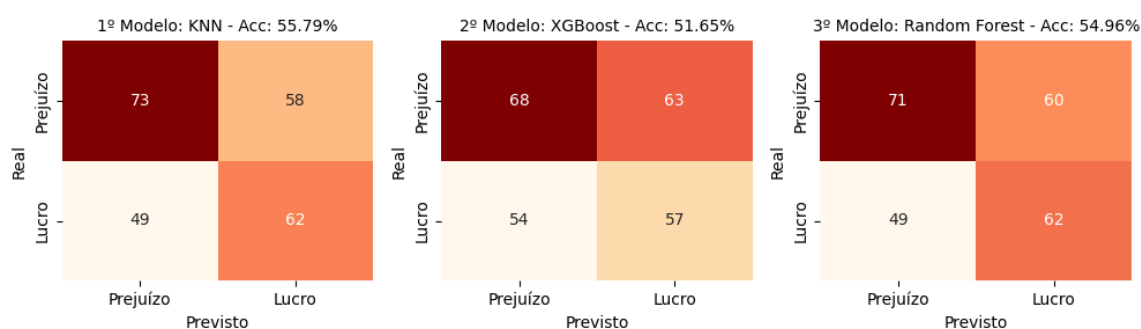
Em contraste, os modelos *Random Forest* e *KNN* apresentaram maior equilíbrio. Ambos mantiveram o número de falsos positivos inferior ao de verdadeiros positivos, além de registrarem alta acurácia na previsão de operações com prejuízo. O *KNN* apresentou desempenho ligeiramente superior, com 58 operações de falso lucro frente a 62 de verdadeiro lucro, enquanto o *Random Forest* contabilizou 60 falsos lucros e também 62 verdadeiros lucros.

Figura 36 – Curvas ROC dos modelos de classificação na estratégia *gain* e *loss* de 200 pontos.



Fonte: Autoria própria.

Figura 37 – Matriz de Confusão nos dados de teste na estratégia *gain* e *loss* de 200 pontos.



Fonte: Autoria própria.

A comparação entre os modelos, sob a perspectiva financeira, indica um desempenho insatisfatório. A Tabela 7 mostra que, embora o modelo *KNN* tenha registrado lucro, o valor provavelmente não seria suficiente para cobrir os custos operacionais, como taxas e impostos. O impacto dos falsos positivos comprometeu a rentabilidade das estratégias, tornando esta abordagem menos eficiente.

Entre os modelos com saldo positivo, o *Random Forest* apresentou lucro de R\$ 27,00, enquanto o *KNN* alcançou R\$ 132,00. No entanto, esses valores são baixos frente ao risco assumido. No caso do *KNN*, o lucro médio por operação foi de apenas R\$ 1,10 em um total de 120 operações, o que reforça a limitação da estratégia em termos de retorno financeiro.

Tabela 7 – Resumo dos modelos usados na estratégia *gain* e *loss* de 200 pontos.

Operação	Tradings	Win-Rate	Saldo	Lucro por Operação
Sem Modelo	242	45,87%	-815,00	-3,37
<b>Com KNN</b>	<b>120</b>	<b>51,67%</b>	<b>132,00</b>	<b>1,10</b>
Com XGBoost	120	47,50%	-255,00	-2,13
Com Random Forest	122	50,82%	27,00	0,22

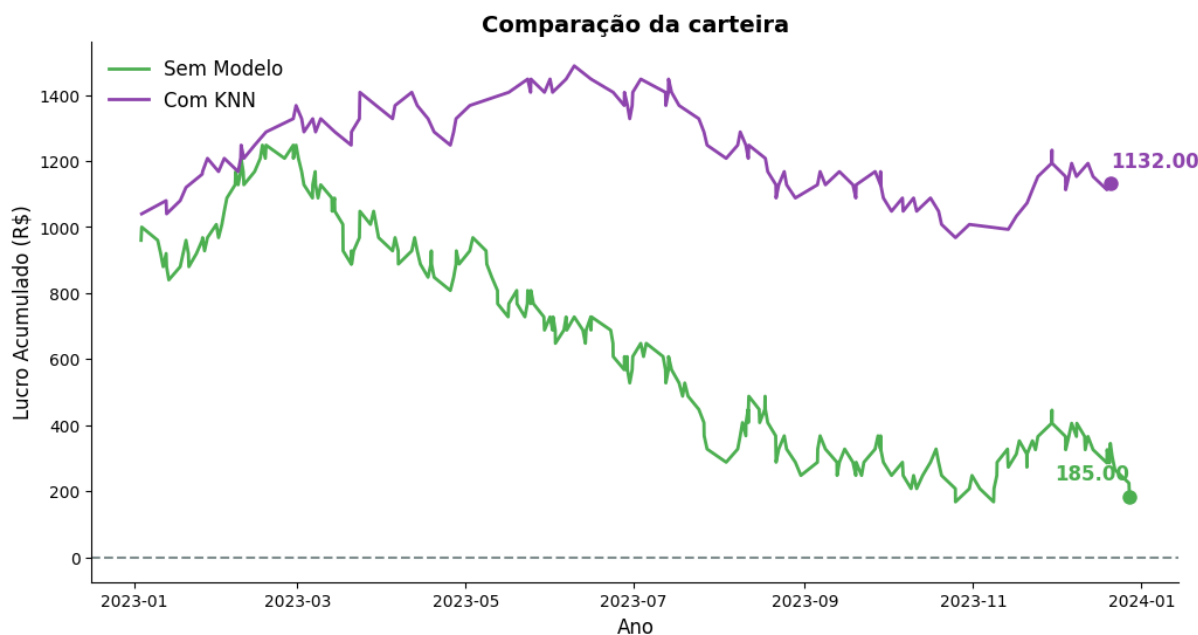
Fonte: Elaborado pelo próprio autor

Em relação à volatilidade, a Figura 38 ilustra o comportamento das operações ao longo do tempo. Supondo um investimento inicial de R\$ 1.000,00, observa-se que o melhor modelo, o *KNN*, não eliminou completamente as oscilações, mas conseguiu preservar o capital por um período mais prolongado. No final do ano, houve uma queda que levou o saldo abaixo do valor inicial, indicando uma perda parcial, seguida por uma leve recuperação.

Ainda assim, o modelo ofereceu certa proteção ao capital, especialmente quando comparado às operações realizadas sem o apoio de modelos preditivos, que sofreram quedas expressivas a partir de março. Embora nenhuma das abordagens tenha levado à exaustão total do capital no período analisado, nota-se uma tendência mais acentuada de deterioração nas operações sem modelo, sugerindo um risco maior de esgotamento do saldo ao longo do tempo.

Não foi incluída a tabela mensal para esta estratégia, uma vez que o resultado financeiro, embora positivo, é muito reduzido. O objetivo principal da tabela seria identificar os meses com lucro, pois estes implicam em tributação. No entanto, considerando os custos operacionais, é possível que a estratégia deixe de ser lucrativa após o desconto de taxas e impostos.

Figura 38 – Comparação de operações sem modelo e com modelo ao longo de 2023 na estratégia *gain* e *loss* de 200 pontos.



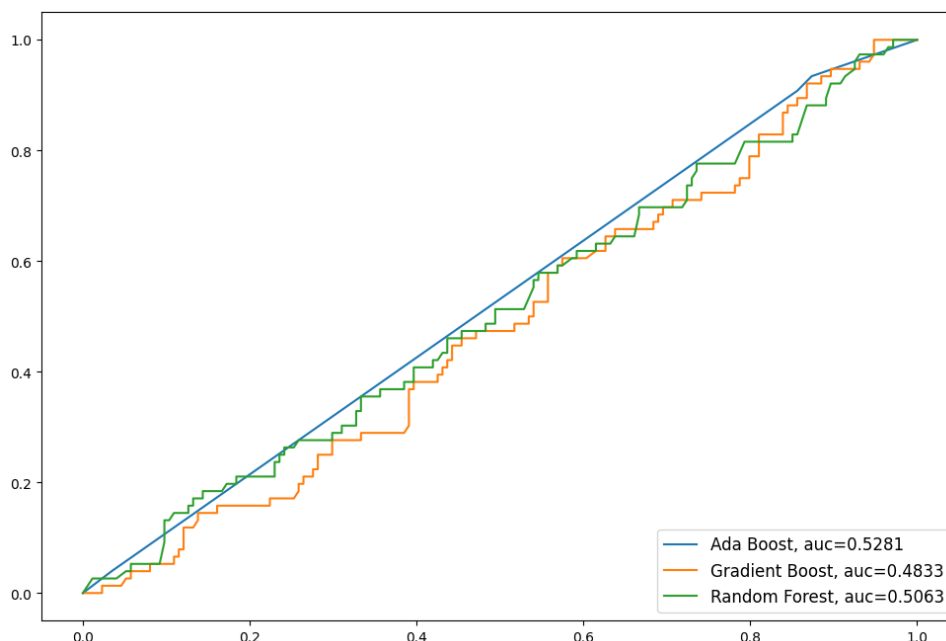
Fonte: Autoria própria.

#### 4.3 GAIN: 200 PONTOS - LOSS: 100 PONTOS

Ao analisar a matriz de confusão apresentada na [Figura 40](#), observa-se que os modelos não conseguiram filtrar adequadamente as operações lucrativas. O modelo *AdaBoost*, em particular, apresentou um número expressivo de falsos positivos — 152 operações previstas como lucro que, na realidade, resultaram em prejuízo. Esse valor é consideravelmente superior ao número de acertos na classe de lucro, com apenas 71 operações corretamente classificadas. Apesar do uso de técnicas de balanceamento de classes no treinamento, o modelo demonstrou baixa capacidade de distinguir operações efetivamente lucrativas, indicando desempenho insatisfatório nesta estratégia.

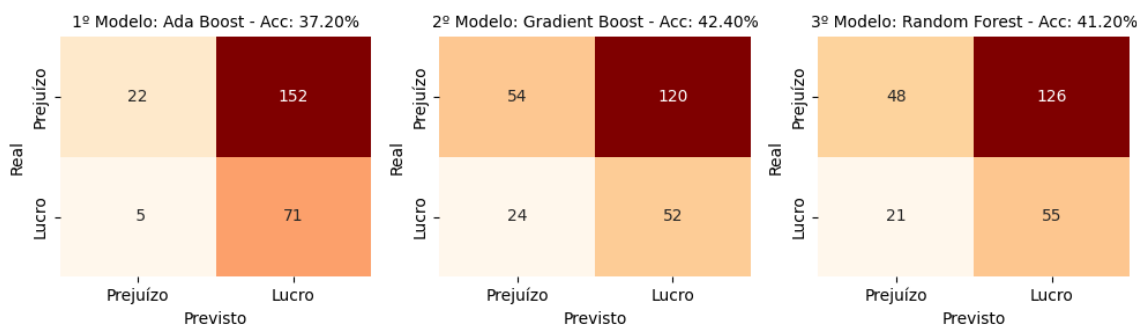
Outros modelos, como *Gradient Boost* e *Random Forest*, tiveram desempenho relativamente superior ao *AdaBoost*, mas ainda limitado. Embora tenham apresentado maior assertividade na identificação de prejuízos, o número de falsos positivos permaneceu elevado — 120 para o *Gradient Boost* e 126 para o *Random Forest*. Isso evidencia que, mesmo com alguma melhoria, os modelos ainda apresentaram dificuldades relevantes em filtrar corretamente as operações de lucro, comprometendo a eficácia da estratégia como um todo.

Figura 39 – Curvas ROC dos modelos de classificação na estratégia *gain* de 200 pontos e *loss* de 100 pontos.



Fonte: Autoria própria.

Figura 40 – Matriz de Confusão nos dados de teste na estratégia *gain* de 200 pontos e *loss* de 100 pontos.



Fonte: Autoria própria.

Do ponto de vista financeiro, a Tabela 8 confirma que a estratégia baseada em modelos não apresentou resultados satisfatórios. Todos os três algoritmos analisados resultaram em prejuízo, ainda que com perdas inferiores às observadas nas operações sem modelo. Sem o uso de modelos, o saldo negativo foi de R\$ 461,00, enquanto os modelos apresentaram perdas de R\$ 341,00 para o *Random Forest* e o *Gradient Boost*, e de R\$ 221,00 para o *AdaBoost*. Outro ponto relevante é a quantidade de operações realizadas, variando entre 180 e 250 *tradings*, o que pode aumentar consideravelmente os custos operacionais.

Diante desse cenário, a análise da volatilidade, do gráfico de evolução da carteira e da tabela de lucro mensal torna-se dispensável, uma vez que os resultados consolidados não demonstram viabilidade econômica.

Tabela 8 – Resumo dos modelos usados na estratégia *gain* de 200 pontos e *loss* de 100 pontos.

Operação	Tradings	Win-Rate	Saldo	Lucro por Operação
Sem Modelo	250	30,40%	-461,00	-1,84
Com Ada Boost	223	31,84%	-221,00	-0,99
Com Gradient Boost	172	30,23%	-341,00	-1,98
Com Random Forest	181	30,39%	-341 ,00	-1,88

Fonte: Elaborado pelo próprio autor

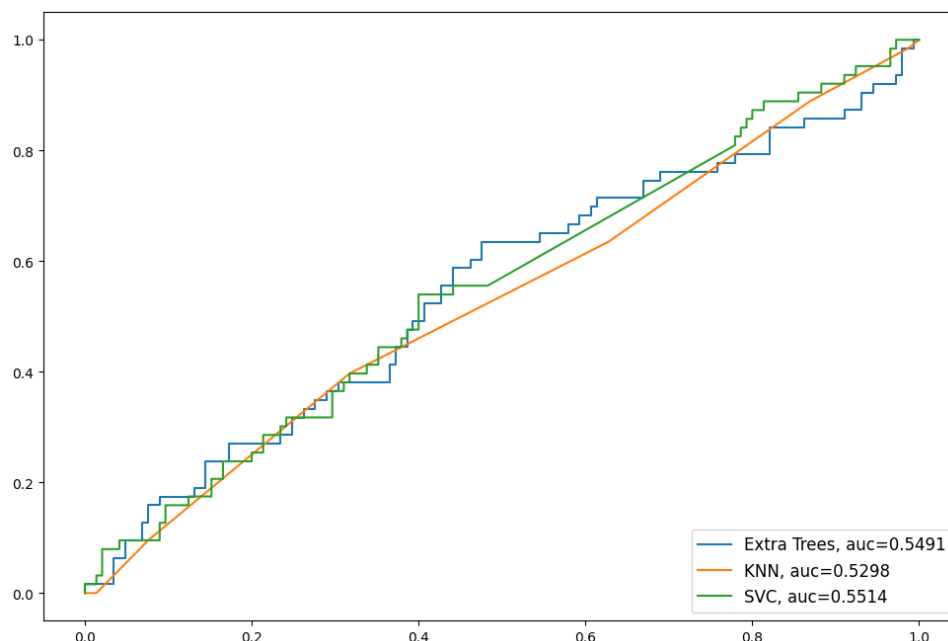
#### 4.4 GAIN: 500 PONTOS - LOSS: 250 PONTOS

O objetivo desta estratégia permanece o mesmo da versão anterior, porém com valores mais elevados. Embora o *loss* ainda possa ser alcançado rapidamente, a obtenção de lucros superiores a 500 pontos tende a ser mais demorada, dada a nova proporção. Dessa forma, a proposta busca apenas tornar o prejuízo um pouco menos recorrente, embora ainda se espere que as perdas sejam frequentes. No entanto, os ganhos obtidos, quando ocorrem, podem compensar o valor das perdas acumuladas. Entre os modelos avaliados, *Extra Trees*, *KNN* e *SVC* apresentaram os melhores resultados nos dados de treinamento.

A [Figura 41](#) evidencia uma melhora em relação à estratégia anterior com a mesma proporção. Nessa figura, os modelos *SVC* e *Extra Trees* alcançam valores de *AUC* superiores aos da versão anterior da estratégia, embora ainda inferiores aos observados nas configurações com proporção 1:1.

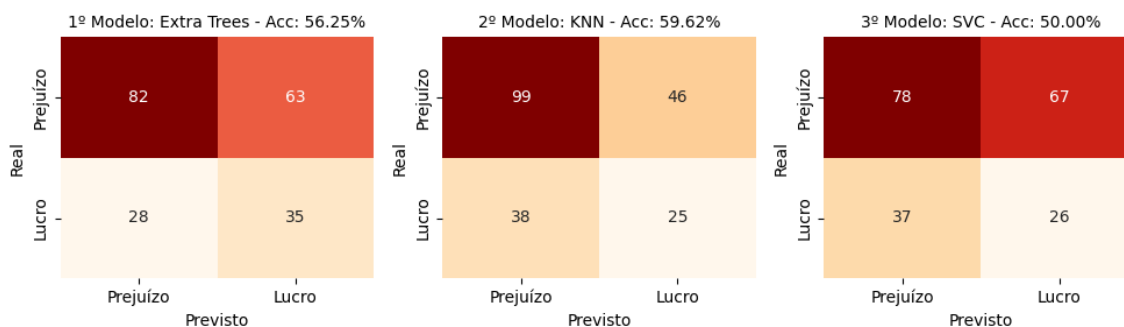
A matriz de confusão, apresentada na [Figura 42](#), também indica progresso. Os modelos foram capazes de filtrar melhor os dados, embora os falsos positivos ainda persistam. O aumento na precisão ao identificar operações com prejuízo pode ter implicações positivas na gestão financeira. Observa-se, ainda, uma mudança no desempenho relativo dos modelos: nesta configuração, o *KNN* superou o *SVC*. Ressalta-se que as curvas *ROC* e os valores de *AUC* foram obtidos com dados de treinamento, enquanto a matriz de confusão foi gerada a partir dos dados de teste. O bom desempenho do *KNN*, portanto, está associado principalmente à sua capacidade de prever corretamente operações que resultam em prejuízo.

Figura 41 – Curvas ROC dos modelos de classificação na estratégia *gain* de 500 pontos e *loss* de 250 pontos.



Fonte: Autoria própria.

Figura 42 – Matriz de Confusão nos dados de teste na estratégia *gain* de 500 pontos e *loss* de 250 pontos.



Fonte: Autoria própria.

A Tabela 9 destaca, em primeiro lugar, o desempenho do modelo *KNN*. Conforme ilustrado na Figura 42, o modelo classificou 25 operações como lucrativas e 46 como prejuízo. Em um cenário ideal, com proporção de 2:1 entre ganho e perda, poderia-se inferir que o total obtido nas 25 operações positivas compensaria as perdas nas 46 negativas. No entanto, essa análise simplificada não se confirma na prática.

A própria Tabela 9 demonstra que a rentabilidade do modelo não segue essa lógica. Isso ocorre porque, além das metas definidas por *stop gain* e *stop loss*, as operações também podem ser encerradas automaticamente no final do pregão. Esse fator impacta diretamente o resultado financeiro, reduzindo o potencial de lucro esperado.

Ainda assim, diferentemente da outra estratégia com proporção 2:1, esta apresentou ao menos um modelo com resultado positivo. Contudo, ao considerar custos operacionais e

incidência de impostos, a viabilidade econômica continua questionável.

Tabela 9 – Resumo dos modelos usados na estratégia *gain* de 500 pontos e *loss* de 250 pontos.

Operação	Tradings	Win-Rate	Saldo	Lucro por Operação
Sem Modelo	208	30,29%	-1.457,00	-7,00
<b>Extra Trees</b>	<b>98</b>	<b>35,71%</b>	<b>154,00</b>	<b>1,57</b>
KNN	71	35,21%	-136,00	-1,92
SVC	93	27,96%	-971,00	-10,44

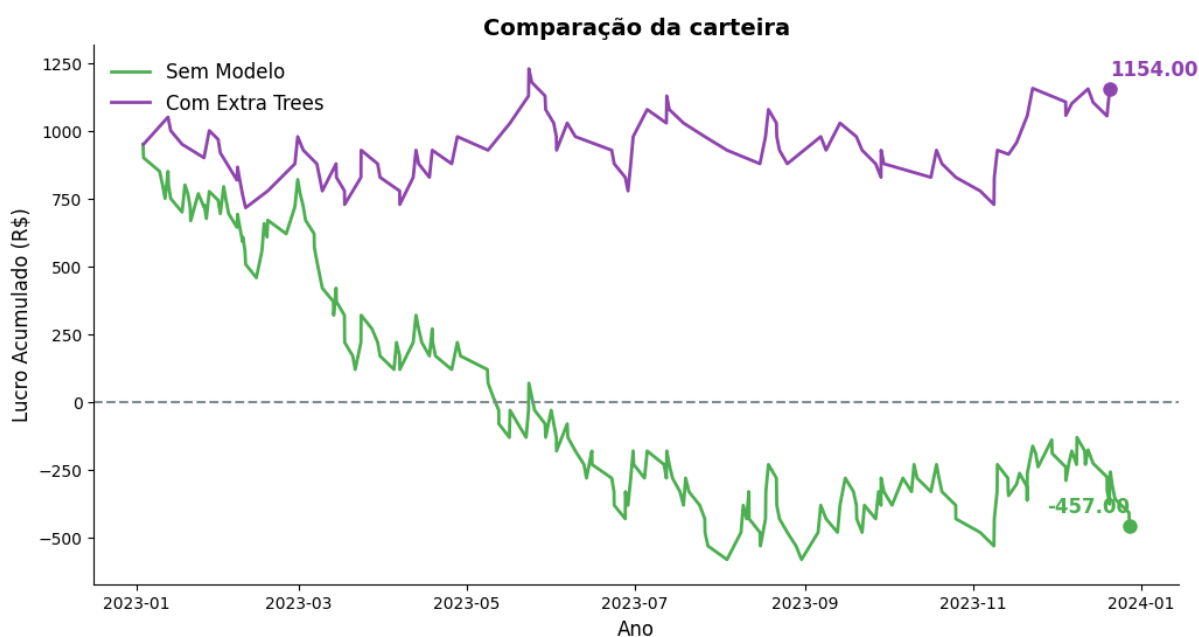
Fonte: Elaborado pelo próprio autor

A [Figura 43](#) revela que a volatilidade da carteira não foi suavizada com o uso do modelo, apresentando oscilações frequentes ao longo do tempo. Ainda assim, observa-se certa estabilidade dentro de uma faixa delimitada, com um suporte em torno de R\$ 750,00 e uma resistência próxima de R\$ 1.250,00. O saldo oscilou entre esses limites, sem apresentar tendência clara de valorização ou depreciação acentuada.

Embora o modelo não tenha gerado lucros expressivos — especialmente quando considerados os custos operacionais e tributários —, seu desempenho foi significativamente superior ao das operações realizadas sem o uso de modelos. Nesse caso, a carteira entrou em colapso já em maio, sem qualquer sinal de recuperação, mantendo-se em níveis baixos e voláteis até o final do período analisado. Esse comportamento pode ser interpretado como uma armadilha financeira, na qual aportes adicionais modestos tendem a ser consumidos rapidamente, ampliando o prejuízo acumulado.

Por fim, a tabela de lucros mensais foi omitida, uma vez que os valores obtidos foram reduzidos e não contribuiriam para uma análise mais aprofundada da estratégia.

Figura 43 – Comparação de operações sem modelo e com modelo na estratégia *gain* de 500 pontos e *loss* de 250 pontos no ano de 2023.



Fonte: Autoria própria.

#### 4.5 GAIN: 500 PONTOS - LOSS: 100 PONTOS

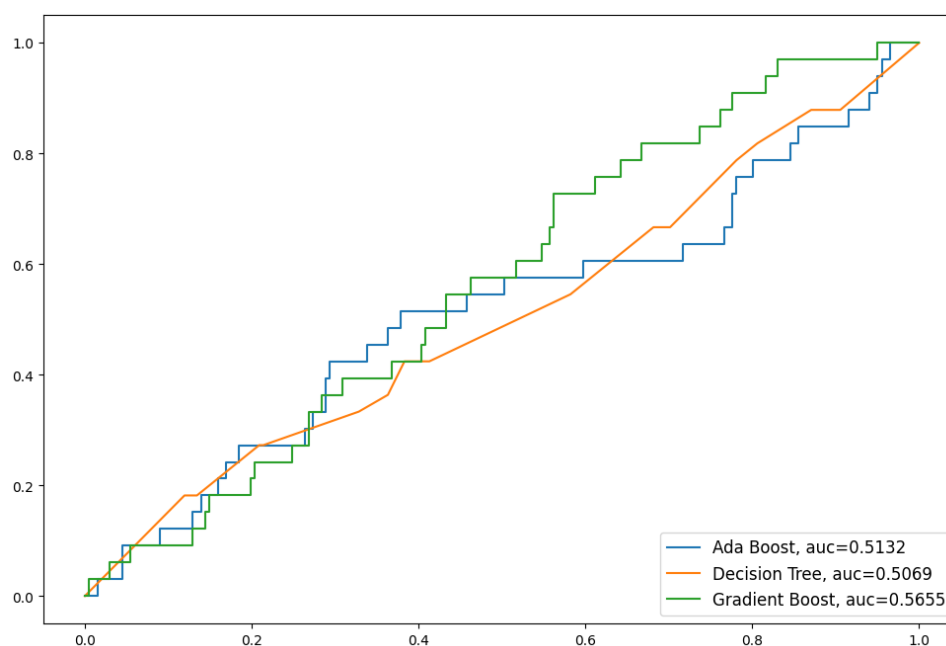
Nesta estratégia, buscou-se ampliar a relação entre ganho e perda, operando com uma proporção mais agressiva. Como já discutido, o principal desafio reside na previsão de operações com prejuízo. O desbalanceamento da base é marcante: cerca de 80% das operações pertencem à classe de perda, o que é coerente com a configuração adotada — *stop loss* curto e *stop gain* longo. Nesse cenário, a expectativa é que um único ganho compense aproximadamente cinco perdas consecutivas, o que reforça a aposta no valor absoluto do lucro, e não na sua frequência.

A Figura 44 mostra que apenas o modelo *Gradient Boost* apresentou desempenho razoável durante o treinamento. Os demais algoritmos demonstraram dificuldades diante do forte desbalanceamento, mesmo com a aplicação de técnicas como o *under sampling*, que, apesar de reduzir a quantidade de dados disponíveis, buscou equilibrar a distribuição entre as classes. Essa abordagem contribuiu para melhorar a previsibilidade dos modelos em cenários reais, mas os efeitos ainda foram limitados.

A Figura 45 revela que o principal problema persiste: os modelos continuam com dificuldades em prever corretamente as operações de perdas, resultando em um elevado número de falsos positivos. Isso reforça a fragilidade da estratégia, uma vez que, além da dificuldade em identificar corretamente as classes, a quantidade de operações lucrativas é, por natureza, reduzida.

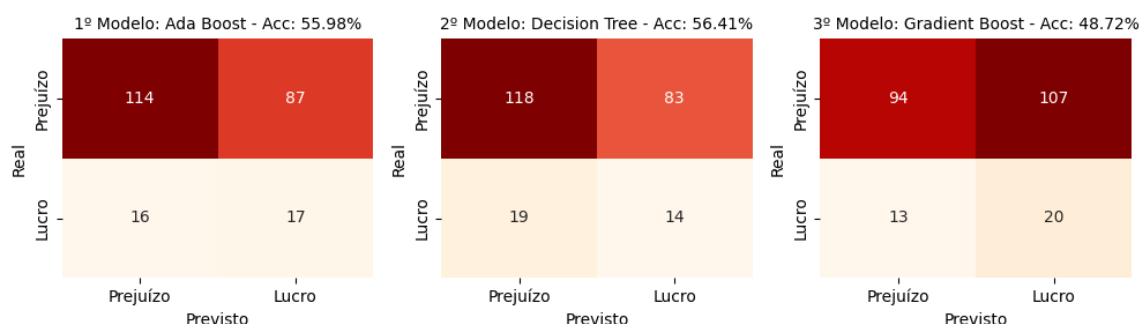
Embora o *Gradient Boost* tenha apresentado métricas aceitáveis nos dados de treino, seu desempenho nos dados de teste foi o pior entre os modelos, indicando problemas de generalização. Por outro lado, os demais algoritmos atingiram acurácias mais elevadas apenas por classificarem a maioria das operações como prejuízo — comportamento típico em cenários desbalanceados, mas que não representa um bom desempenho preditivo.

Figura 44 – Curvas ROC dos modelos de classificação na estratégia *gain* de 500 pontos e *loss* de 100 pontos.



Fonte: Autoria própria.

Figura 45 – Matriz de Confusão nos dados de teste na estratégia *gain* de 500 pontos e *loss* de 100 pontos.



Fonte: Autoria própria.

Já nos dados financeiros apresentados na Tabela 10, observa-se que nenhum modelo alcançou desempenho satisfatório. As taxas de *win-rate* foram muito baixas, o que é esperado considerando que o *stop loss* curto atua como uma proteção para limitar perdas.

Um ponto relevante é o desempenho do modelo *Decision Tree*, que apresentou resultados inferiores até mesmo às operações sem modelo. Embora a maior quantidade de operações possa aumentar os custos e impactar negativamente o resultado das operações sem modelo, esta é a primeira ocorrência em que um modelo apresenta desempenho inferior à operação aleatória.

Diante desse cenário, conclui-se que a estratégia não é viável com os parâmetros adotados. Considerando as perdas significativas em todas as posições, não se justifica a análise da volatilidade nem do lucro mensal.

Tabela 10 – Resumo dos modelos usados na estratégia *gain* de 500 pontos e *loss* de 100 pontos.

Operação	Tradings	Win-Rate	Saldo	Lucro por Operação
Sem Modelo	234	26,92%	-1.110,00	-4,74
Ada Boost	104	16,35%	-237,00	-2,28
Decision Tree	97	14,43%	-533,00	-5,49
Gradient Boost	127	15,75%	-418,00	-3,29

Fonte: Elaborado pelo próprio autor

#### 4.6 GAIN: 250 PONTOS - LOSS: 500 PONTOS

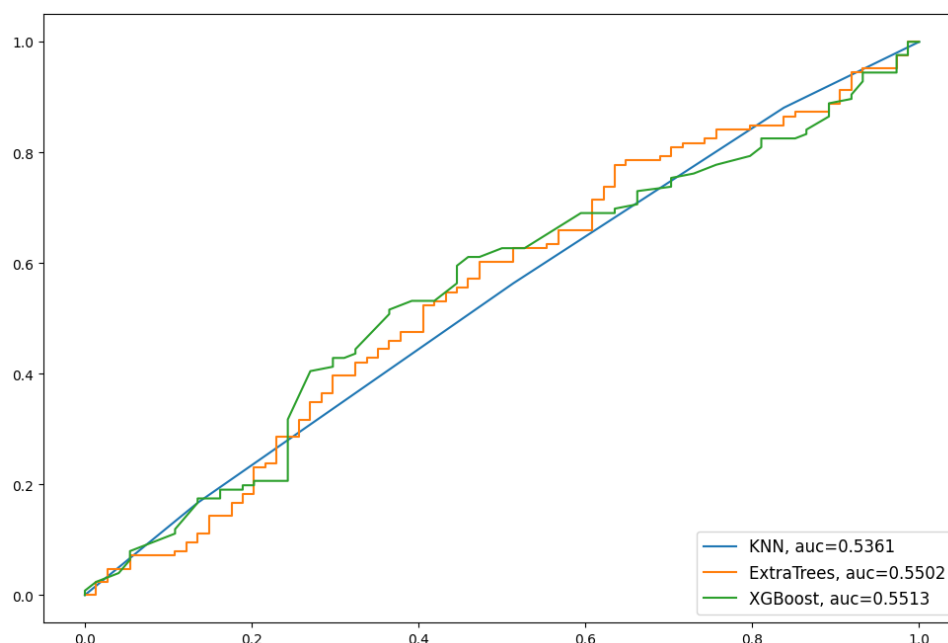
Na última estratégia analisada, o objetivo difere significativamente das anteriores, pois busca-se obter um lucro considerável, mesmo que o tempo para alcançá-lo seja maior. Dessa forma, a estratégia prioriza dar mais tempo para que o lucro seja alcançado, enquanto o *loss* demora mais a ser atingido. Como nas outras estratégias, há um desbalanceamento nos dados, e é importante verificar se o volume de lucros supera o de perdas.

Conforme ilustrado na Figura 46, os modelos apresentaram desempenho semelhante, com exceção do *KNN*, que teve um desempenho ligeiramente inferior. Os modelos *XGBoost* e *Extra Trees* alcançaram valores de *AUC* próximos de 0,55, considerados aceitáveis para esta análise. Vale destacar que vários modelos testados nesta seção apresentaram valores

semelhantes de *AUC*. O maior valor registrado até o momento foi 0,59, obtido na estratégia com *gain* e *loss* de 500 pontos. Isso evidencia a dificuldade inerente à tarefa de predição, embora os modelos *Extra Trees* e *XGBoost* tenham se destacado nos dados de treinamento.

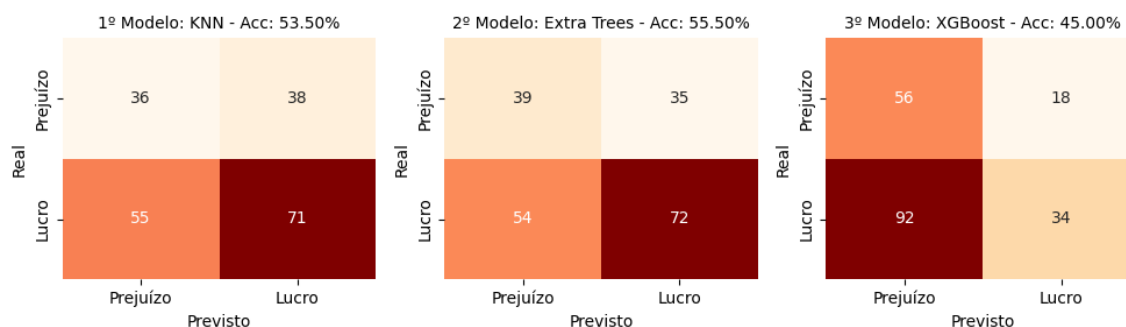
Ao analisar a matriz de confusão mostrada na [Figura 47](#), observa-se um problema na predição do lucro. Especificamente, há um número considerável de falsos negativos, ou seja, casos em que o modelo prevê prejuízo, mas a operação resultou em lucro. Esse tipo de erro impacta negativamente o desempenho dos modelos. Entretanto, exceto pelo *XGBoost*, que demonstra um viés em priorizar a previsão de prejuízo — classificando poucas operações como lucro —, os modelos conseguiram prever mais corretamente as operações lucrativas do que os falsos negativos. Em comparação com os valores de *AUC* das curvas *ROC*, apenas o modelo *Extra Trees* manteve um desempenho consistente entre os dados de treino e teste.

Figura 46 – Curvas *ROC* dos modelos de classificação na estratégia *gain* de 250 pontos e *loss* de 500 pontos.



Fonte: Autoria própria.

Figura 47 – Matriz de Confusão nos dados de teste na estratégia *gain* de 250 pontos e *loss* de 500 pontos.



Fonte: Autoria própria.

A [Tabela 11](#) mostra que, do ponto de vista financeiro, o modelo *Extra Trees* apresentou o melhor desempenho entre os modelos testados. Nesta estratégia, cada predição correta de lucro é muito relevante, pois as perdas têm grande impacto no resultado final.

Embora a diferença de *win-rate* entre os modelos *KNN* e *Extra Trees* seja pequena, a diferença no saldo financeiro é bastante expressiva. Isso indica que, mesmo com taxas de acerto semelhantes, a qualidade das predições do *Extra Trees* gerou resultados mais favoráveis.

É importante ressaltar que parte das operações foi encerrada automaticamente no final do dia. Isso interfere na proporção ideal entre ganho e perda, contribuindo para a variação dos saldos observados.

Tabela 11 – Resumo dos modelos usados na estratégia *gain* de 250 pontos e *loss* de 500 pontos.

Operação	Tradings	Win-Rate	Saldo	Lucro por Operação
Sem Modelo	200	63,00%	-850,00	-4,25
KNN	109	65,14%	0,00	0,00
<b>Extra Trees</b>	<b>107</b>	<b>67,29%</b>	<b>236,00</b>	<b>2,21</b>
XGBoost	52	65,38%	-136,00	-2,62

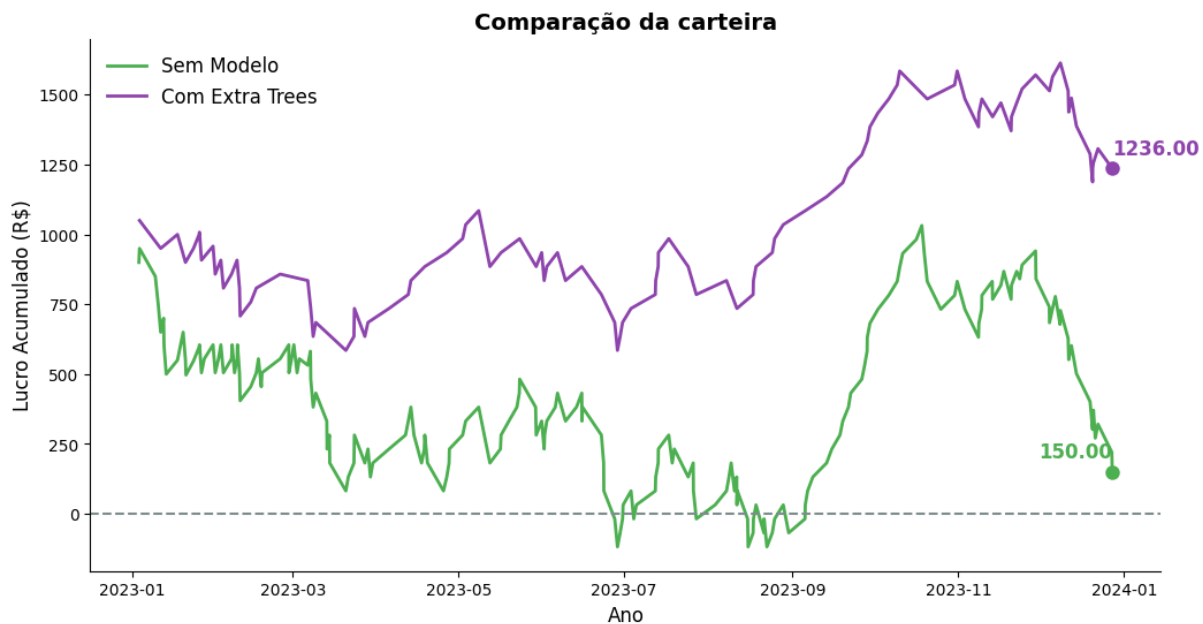
Fonte: Elaborado pelo próprio autor

Com relação à volatilidade, a [Figura 48](#) ilustra a evolução da carteira ao longo do tempo. Observa-se que o modelo utilizado não foi capaz de proteger integralmente contra as oscilações significativas. Apesar de a carteira apresentar lucro ao final do período e não ter ocorrido falência com o investimento inicial de R\$ 1.000,00, houve uma queda acentuada nos primeiros meses do ano, chegando a valores inferiores a R\$ 750,00, seguida de recuperação.

Esse comportamento de forte oscilação se repetiu em diversos momentos ao longo do ano. Em setembro, por exemplo, verificou-se um lucro expressivo, mas, nos meses seguintes, a carteira entrou novamente em trajetória de queda. Ainda assim, vale destacar que, em comparação às operações sem o uso de modelo, a estratégia demonstrou certa resiliência, uma vez que a carteira sem modelo foi completamente comprometida já em julho.

Caso tenha havido um novo aporte em setembro, o modelo permitiu uma valorização significativa no curto prazo, embora essa valorização tenha sido novamente comprometida pelas quedas subsequentes.

Figura 48 – Comparação de operações sem modelo e com modelo na estratégia *gain* de 250 pontos e *loss* de 500 pontos no ano de 2023.



Fonte: Autoria própria.

## 5 CONCLUSÃO

Nesta seção, apresenta-se uma análise crítica do trabalho, com destaque para os modelos que obtiveram os melhores desempenhos, as principais dificuldades enfrentadas ao longo do processo e as contribuições potenciais para a área de investimentos. Também são discutidas perspectivas futuras que podem ampliar o escopo da pesquisa e aprofundar a compreensão dos resultados obtidos.

Ao iniciar este trabalho, tomando como base o artigo de [Chague e Giovannetti \(2020\)](#), partiu-se da hipótese de que as estratégias de *day trade* adotadas por operadores do mercado resultariam, em sua maioria, em desempenho negativo. No entanto, há diferenças importantes entre as abordagens: a primeira refere-se ao ativo analisado — o artigo avaliou operações com ações, enquanto este estudo foca no mercado futuro WIN. A segunda diz respeito ao escopo da análise — enquanto o artigo concentra-se no perfil dos profissionais que operam, este trabalho examina o desempenho de estratégias específicas frequentemente divulgadas ou utilizadas por esses operadores.

Tabela 12 – Resumo geral das estratégias com seus melhores modelos e resultados.

Estratégia	Modelo	Tradings	Valor	Lucro por Operação
<b>Gain: 500 Pontos - Loss: 500 Pontos</b>	<b>Random Forest</b>	<b>73</b>	<b>734,00</b>	<b>10,05</b>
Gain: 200 Pontos - Loss: 200 Pontos	KNN	120	132,00	1,10
Gain: 200 Pontos - Loss: 100 Pontos	Ada Boost	223	-221,00	-0,99
Gain: 500 Pontos - Loss: 250 Pontos	Extra Trees	98	154,00	1,57
Gain: 500 Pontos - Loss: 100 Pontos	Ada Boost	104	-237,00	-2,28
Gain: 250 Pontos - Loss: 500 Pontos	Extra Trees	107	236,00	2,21

Fonte: Elaborado pelo próprio autor

De fato, a maioria das estratégias avaliadas resultou em perdas. Ao longo do trabalho, foram definidas seis estratégias, testadas tanto com nove modelos diferentes quanto sem modelo, totalizando 60 carteiras simuladas. Conforme apresentado na [Tabela 12](#), apenas três carteiras obtiveram lucro positivo ao final do período analisado.

Entre essas, a que apresentou os melhores indícios de lucratividade foi a carteira baseada na estratégia com 500 pontos de *gain* e *loss*, treinada com o modelo *Random Forest*. Essa estratégia apresentou bons resultados em diversos aspectos. O lucro médio por operação foi de R\$ 10,05, um valor significativamente superior às taxas e impostos envolvidos nas transações. Além disso, entre todas as estratégias testadas, essa registrou o menor número de operações, o que contribui para a redução dos custos operacionais e minimiza impactos emocionais associados a sucessivas perdas.

A volatilidade dessa estratégia ao longo de 2023 também se manteve controlada, conforme ilustrado na [Figura 35](#). Embora tenham ocorrido perdas, o que é natural no mercado, a carteira esteve longe de zerar o capital disponível e, nos últimos meses do período analisado, apresentou uma tendência de aumento nos lucros. Comparado aos demais modelos, esse foi o que obteve o melhor desempenho, sendo o único com chances reais de lucratividade.

Além disso, mais três estratégias apresentaram lucro. No entanto, o lucro médio por operação foi inferior a R\$ 3,00. Considerando as taxas da B3, taxas de corretagem e impostos que variam conforme o número de contratos, volume de negociação, vencimento de contrato,

descontos e estratégias das corretoras, acredita-se que um lucro por operação superior a R\$ 5,00 seja necessário para garantir uma margem de segurança.

Em relação à volatilidade, a estratégia de 200 pontos de *gain* e *loss* apresentou um comportamento mais estável, sem perdas significativas em relação ao capital inicial. Essa estratégia chegou a apresentar ganhos consideráveis no início do ano, mas posteriormente reverteu para perdas, conforme a [Figura 38](#).

Enquanto isso, outras estratégias lucrativas, embora tenham gerado maiores retornos ao final do período, inicialmente registraram prejuízos antes de se recuperarem. Esse fator é relevante, pois afeta diretamente o aspecto emocional de um operador que utilizaria tais estratégias. No caso dos modelos mais voláteis, que chegaram a perder mais de 25% do capital inicial nos primeiros meses, surge o questionamento: será que um *trader* suportaria essas perdas iniciais sem abandonar a estratégia antes da recuperação?

Por essa razão, a proporção 1:1 apresentou os melhores resultados neste trabalho, principalmente na estratégia de 500 pontos de *gain* e *loss*. Dois dos melhores desempenhos financeiros pertencem a essa configuração. A [Tabela 5](#) mostrou que, com o modelo *Extra Trees*, o lucro por operação foi superior a R\$ 5,00 — um valor considerado suficiente para cobrir os custos operacionais. Embora a volatilidade desse modelo não tenha sido discutida em detalhes, em termos de saldo, ele se destacou como o segundo melhor desempenho geral, ficando atrás apenas do modelo *Random Forest*. Em seguida, o modelo treinado com a estratégia de 200 pontos de *gain* e *loss* ocupou a terceira colocação em desempenho financeiro.

Portanto, é possível identificar estratégias com potencial de lucratividade no mercado ao utilizar padrões de *candles* combinados com indicadores financeiros. No entanto, tais abordagens são complexas e apresentam desempenho modesto. Em geral, os modelos analisados nos dados de treino apresentaram valores de *AUC* próximos de 0,50, o que equivale ao desempenho de modelos *dummy*, que classificam de forma aleatória. Apenas um modelo alcançou um *AUC* de 0,59, valor considerado atípico, já que a maioria ficou entre 0,50 e 0,55.

Além da complexidade técnica, o aspecto emocional desempenha um papel decisivo no desempenho operacional. A disciplina é fundamental, sobretudo no cumprimento dos limites de *stop loss* e *stop gain*. O *stop loss* demanda paciência, pois, se a operação for encerrada prematuramente, pode haver reversão positiva. Por outro lado, a permanência excessiva em uma operação com perdas pode aumentar ainda mais o prejuízo. Já o *stop gain* requer confiança no planejamento para evitar saídas antecipadas que comprometam o desempenho. A ganância, por sua vez, pode anular lucros obtidos, caso a operação não seja encerrada no momento adequado.

Operar nesse tipo de mercado exige esforço contínuo, mesmo com o uso de *bots*, pois o modelo precisa ser constantemente reavaliado e retreinado, dada a alta volatilidade e a natureza mutável do ambiente financeiro. Apesar disso, não há garantias de sucesso. Este estudo utilizou um período de cinco anos para treinamento e um ano para teste. Entre as 60 carteiras avaliadas, apenas duas apresentaram indícios de lucratividade consistente, o que representa pouco mais de 3% do total. Ainda assim, a rentabilidade observada não é definitiva: não se sabe como tais estratégias se comportariam em horizontes mais longos, como 10, 15 ou 30 anos, nem se manteriam desempenho positivo diante de eventos inesperados ou choques de mercado. Essas são questões que exigem estudos futuros mais aprofundados.

Outro ponto relevante discutido por [Chague e Giovannetti \(2020\)](#) refere-se à viabilidade de obter um lucro médio diário de R\$ 100,00 e, assim, viver exclusivamente de operações no mercado financeiro. Tal meta, no entanto, está distante da realidade observada neste estudo. É importante destacar que todas as simulações foram realizadas neste estudo com um capital inicial de apenas R\$ 1.000,00, o que impõe uma limitação significativa ao potencial de retorno.

O melhor desempenho obtido neste trabalho resultou em um saldo de R\$ 734,00 ao

longo de 2023, o que corresponde a uma média mensal de aproximadamente R\$ 61,00 — ou pouco mais de R\$ 3,00 por dia. Apesar de se tratar de uma média, o impacto da volatilidade é substancial: em um dia pode-se obter R\$ 100,00 de ganho, e no seguinte, perder R\$ 200,00. A Tabela 5 apresenta o lucro mensal real dessa carteira, evidenciando inclusive meses com prejuízo, o que exige resiliência para atravessar períodos desfavoráveis.

Além disso, em um cenário no qual se desejasse viver dos rendimentos dessas operações, seria necessário realizar retiradas regulares do capital acumulado. Esse fator comprometeria ainda mais a sustentabilidade da carteira, já que a diminuição do capital reduziria a capacidade de absorver perdas futuras e de gerar novos lucros.

## 5.1 TRABALHOS FUTUROS

Como este estudo foi limitado a um único tempo gráfico de 15 minutos, seria pertinente aplicar as mesmas estratégias em outros tempos gráficos intradiários para verificar se a lucratividade se mantém. Isso permitiria avaliar se as estratégias com proporção 1:1 continuam apresentando bons resultados ou se outras configurações, como 2:1 ou 1:2, teriam desempenho superior.

Além da proporção entre *gain* e *loss*, seria necessário ajustar a quantidade de pontos alvo conforme o tempo gráfico utilizado. Em tempos gráficos menores, a amplitude média dos *candles* tende a ser menor, exigindo uma redução proporcional nos pontos de entrada e saída. Por outro lado, em tempos gráficos maiores, como 30 minutos ou 1 hora, os *candles* geralmente apresentam maior variação de preços, o que justificaria o aumento na quantidade de pontos visados, alinhando o alvo ao comportamento típico do preço nesse intervalo.

Outra possibilidade de ampliação do estudo envolve o uso exclusivo de indicadores técnicos. Estratégias baseadas apenas em indicadores poderiam resultar em um maior número de operações e ganhos mais expressivos, além de facilitar o estabelecimento de um *stop loss* de proteção. Isso permitiria operações ancoradas em sinais objetivos de compra ou venda, possibilitando uma análise mais direta da lucratividade dessas estratégias. Vale destacar que a utilização de padrões de *candles*, apesar de útil, limitou a frequência das operações neste estudo.

No que diz respeito aos ativos financeiros, seria interessante aplicar as estratégias a outros contratos futuros, como o Mini Dólar (WDO), o contrato cheio de Ibovespa (IND) e o contrato cheio de Dólar (DOL), com as devidas adaptações na pontuação e nos parâmetros operacionais. Também é viável explorar a aplicação dessas estratégias em ativos com alta liquidez, como ações negociadas nas bolsas americanas. Além disso, criptomoedas com grande volume de negociação, como o Bitcoin, representam um mercado promissor para testar a robustez dos modelos em diferentes contextos de volatilidade e comportamento de preços.

Outra possibilidade de aprimoramento do modelo está relacionada à previsão de tendências de mercado, elemento essencial para estratégias de *day trade*. Embora este estudo tenha utilizado indicadores como médias móveis, Hi-Lo, MACD e RSI na tentativa de capturar sinais de direção dos preços, os resultados mostraram que a identificação precisa de tendências ainda é limitada. De acordo com a Teoria de Dow, essas tendências podem ser classificadas como primárias, secundárias e terciárias, o que reforça a complexidade do problema.

Essa limitação impacta diretamente os resultados, uma vez que muitas estratégias operacionais dependem de movimentos consistentes de alta ou baixa para alcançar lucratividade. Estratégias que buscam antecipar a direção do ativo — em vez de prever valores exatos — podem ser mais eficazes em determinados contextos. Assim, aperfeiçoar a detecção de tendências, seja por meio de técnicas de *machine learning*, indicadores combinados ou até

processamento de imagens, pode representar um avanço relevante na eficácia dos modelos utilizados.

Também é recomendável aplicar os modelos e estratégias analisadas neste estudo em contextos com perfis de investimento distintos, como o modelo *Buy and Hold*, amplamente utilizado por investidores com foco em análise fundamentalista. Essa comparação pode proporcionar uma avaliação mais ampla da eficácia das estratégias baseadas em análise técnica e *machine learning*, especialmente em horizontes de longo prazo. A realização desse tipo de experimento permitiria investigar se os padrões e modelos que se mostraram eficazes no *day trade* mantêm sua relevância em estratégias de investimento orientadas por fundamentos econômicos e financeiros.

Por fim, a adoção de modelos mais sofisticados demandaria uma reformulação metodológica, com o tratamento do problema sob a perspectiva de séries temporais, em vez de uma tarefa de classificação supervisionada. Neste estudo, optou-se pela construção de um *dataset* estruturado a partir de operações de mercado, adaptado para modelos classificatórios. No entanto, técnicas como ARIMA, SARIMA, *Prophet*, redes neurais recorrentes e modelos baseados em *Reinforcement Learning* apresentam potencial para capturar padrões temporais e dinâmicos mais complexos do mercado financeiro. Dessa forma, investigações futuras podem explorar a modelagem sequencial dos dados, considerando aspectos como dependência temporal, autocorrelação e sazonalidade, com o objetivo de aprimorar a capacidade preditiva e a robustez das estratégias desenvolvidas.

## 5.2 CONSIDERAÇÕES FINAIS

Existem estratégias potencialmente lucrativas no *day trade*, mas sua execução está longe de ser simples. A complexidade técnica envolvida, o controle emocional necessário, além da criação de automações e modelos de aprendizado de máquina, exigem um elevado nível de preparo por parte do operador. Por esse motivo, é comum que muitos desistam após sucessivas falhas.

Apesar de todos os tipos de investimento envolverem incertezas, essa modalidade, especialmente no mercado futuro, apresenta riscos significativamente mais elevados. Nessas operações, as perdas podem ultrapassar o valor inicialmente investido, resultando em saldos negativos. Em contrapartida, em investimentos como ações — também considerados de risco — as perdas normalmente se limitam ao capital aplicado, sem a possibilidade de endividamento adicional.

As operações de *day trade* avaliadas neste estudo podem apresentar valorização do capital ao longo do tempo; no entanto, isso não deve ser interpretado como uma possibilidade real de obtenção de renda recorrente, mesmo em cenários com maior alavancagem.

Com base nos resultados apresentados, este estudo oferece evidências de que há possibilidade de obter lucros em algumas estratégias. No entanto, ao se comparar o retorno observado com o nível de risco e esforço exigido, constata-se que a rentabilidade tende a ser limitada.

Do ponto de vista computacional, este trabalho explorou um conjunto expressivo de dados históricos do mercado futuro, estruturando um *dataset* específico para o treinamento e a avaliação de modelos preditivos. Por meio de linguagens de programação, foi possível simular, a partir de um *dataset* de *candles*, diversas operações e realizar testes ao longo do período estudado. A análise dos dados possibilitou a criação de estratégias e a aplicação de tratamentos adequados, como exemplificado pela análise univariada apresentada na [Subseção 3.3.1](#), que permitiu determinar o número ideal de pontos para o tempo gráfico considerado.

Além da análise de estratégias operacionais, este estudo proporcionou uma compreensão aprofundada dos desafios recorrentes na área de ciência de dados, tais como *data leakage* em tarefas de previsão, tratamento de valores ausentes e identificação e correção de *outliers*. Para assegurar a integridade do processo de modelagem, adotou-se uma abordagem robusta de pré-processamento dos dados, validação cruzada e avaliação criteriosa do desempenho dos modelos.

Foram aplicadas técnicas de aprendizado de máquina supervisionado, com destaque para os algoritmos KNN, *Random Forest* e *Extra Trees*, os quais demonstraram desempenho significativo durante os testes. A busca por melhores configurações foi conduzida por meio de técnicas avançadas de otimização de hiperparâmetros, destacando-se a busca bayesiana, que mostrou-se eficiente no ajuste fino dos modelos.

Adicionalmente, este trabalho contribui para a demonstração da aplicabilidade de soluções computacionais avançadas no contexto do mercado financeiro, reforçando a interdisciplinaridade entre ciência de dados, inteligência artificial e engenharia de software. Ressalta-se, ainda, a importância da análise crítica e contínua dos dados para a construção de modelos preditivos confiáveis, bem como a necessidade de constante atualização frente às dinâmicas e particularidades do mercado estudado.

Por fim, os resultados alcançados evidenciam o potencial da integração entre técnicas computacionais e conhecimento financeiro para o desenvolvimento de ferramentas capazes de auxiliar na tomada de decisão, abrindo caminhos para futuras pesquisas e aprimoramentos neste campo.

## Referências

- ADVFN. **Como funcionam os gráficos de candlestick?** 2022. Disponível em: <<https://br.advfn.com/educacional/candlestick/funcionamento-dos-candles>>. Acesso em: 06 de Abril de 2023. Citado na página 9.
- ANBIMA. **RAIO X DO INVESTIDOR BRASILEIRO 4ª EDIÇÃO**. 2022. Disponível em: <[https://www.anbima.com.br/pt\\_br/especial/raio-x-do-investidor-2021.htm](https://www.anbima.com.br/pt_br/especial/raio-x-do-investidor-2021.htm)>. Acesso em: 30 de Maio de 2023. Citado na página 1.
- ANVERSA, L. **O que é circuit breaker? Entenda e relembre as ‘paradas’ da bolsa brasileira**. 2024. Disponível em: <<https://exame.com/invest/guia/o-que-e-circuit-breaker-entenda-e-relembre-as-paradas-da-bolsa-brasileira/>>. Acesso em: 01 de Janeiro de 2025. Citado na página 25.
- AWAIS, M. **HOW TO TRADE THE HANGING MAN CANDLESTICK**. 2020. Disponível em: <<https://forexboat.com/hanging-man-candlestick/>>. Acesso em: 01 de Maio de 2023. Citado na página 10.
- B3. **Ações**. 2022. Disponível em: <[https://www.b3.com.br/pt\\_br/produtos-e-servicos/negociacao/renda-variavel/acoes.htm](https://www.b3.com.br/pt_br/produtos-e-servicos/negociacao/renda-variavel/acoes.htm)>. Acesso em: 06 de Abril de 2023. Citado na página 12.
- B3. **Futuro de Ibovespa**. 2022. Disponível em: <[https://www.b3.com.br/pt\\_br/produtos-e-servicos/negociacao/renda-variavel/futuro-de-ibovespa.htm](https://www.b3.com.br/pt_br/produtos-e-servicos/negociacao/renda-variavel/futuro-de-ibovespa.htm)>. Acesso em: 28 de Abril de 2023. Citado 3 vezes nas páginas 13, 14 e 15.
- B3. **Futuro de Ibovespa**. 2022. Disponível em: <[https://www.b3.com.br/pt\\_br/produtos-e-servicos/negociacao/renda-variavel/mercado-de-acoes/futuro-de-ibovespa.htm](https://www.b3.com.br/pt_br/produtos-e-servicos/negociacao/renda-variavel/mercado-de-acoes/futuro-de-ibovespa.htm)>. Acesso em: 28 de Abril de 2023. Citado 2 vezes nas páginas 14 e 15.
- B3. **Futuro de Taxa de Câmbio de Reais por Dólar Comercial**. 2022. Disponível em: <[https://www.b3.com.br/pt\\_br/produtos-e-servicos/negociacao/moedas/futuro-de-taxa-de-cambio-de-reais-por-dolar-comercial.htm](https://www.b3.com.br/pt_br/produtos-e-servicos/negociacao/moedas/futuro-de-taxa-de-cambio-de-reais-por-dolar-comercial.htm)>. Acesso em: 28 de Abril de 2023. Citado na página 15.
- B3. **Futuro Mini de Ibovespa**. 2022. Disponível em: <[https://www.b3.com.br/pt\\_br/produtos-e-servicos/negociacao/renda-variavel/futuro-mini-de-ibovespa.htm](https://www.b3.com.br/pt_br/produtos-e-servicos/negociacao/renda-variavel/futuro-mini-de-ibovespa.htm)>. Acesso em: 28 de Abril de 2023. Citado na página 13.
- B3. **Futuro Míni de Taxa de Câmbio de Reais por Dólar Comercial**. 2022. Disponível em: <[https://www.b3.com.br/pt\\_br/produtos-e-servicos/negociacao/moedas/futuro-mini-de-taxa-de-cambio-de-reais-por-dolar-comercial.htm](https://www.b3.com.br/pt_br/produtos-e-servicos/negociacao/moedas/futuro-mini-de-taxa-de-cambio-de-reais-por-dolar-comercial.htm)>. Acesso em: 28 de Abril de 2023. Citado na página 15.
- BULKOWSKI, T. N. **Encyclopedia of Candlestick Charts**. 1. ed. [S.l.]: Wiley, 2008. Citado 6 vezes nas páginas 9, 10, 11, 18, 19 e 24.
- CAMPOLIM, A. **Por que você não deve investir tudo no day trade**. 2022. Disponível em: <<https://www.infomoney.com.br/colunistas/tudo-clear/por-que-voce-nao-deve-investir-tudo-no-day-trade/>>. Acesso em: 30 de Maio de 2023. Citado na página 2.

- CARVALHO, C. **Teoria de Dow: o que é e como funciona?** 2021. Disponível em: <<https://yubb.com.br/artigos/acoes/teoria-de-dow-o-que-e-e-como-funciona>>. Acesso em: 04 de Abril de 2023. Citado 2 vezes nas páginas 4 e 5.
- CHAGUE, F.; GIOVANNETTI, B. É possível viver de day-trade em ações? **Brazilian Review Of Finance**, 2020. Disponível em: <<https://bibliotecadigital.fgv.br/ojs/index.php/rbfin/article/download/81949/78263/176074>>. Acesso em: 30 de Maio de 2023. Citado 4 vezes nas páginas 1, 2, 61 e 62.
- CLEARCORRETORA. **Análise Técnica - Operando através de gráficos**. 2020. Disponível em: <<https://www.clear.com.br/site/Content/pdf/download-ebooks/eBook-AnaliseTecnica.pdf>>. Acesso em: 20 de Maio de 2023. Citado 3 vezes nas páginas 5, 6 e 7.
- Educa Mais Brasil. **Quais profissões estão em alta no mercado financeiro?** 2023. Disponível em: <<https://www.educamaisbrasil.com.br/educacao/carreira/quais-profissoes-estao-em-alta-no-mercado-financeiro>>. Acesso em: 30 de Maio de 2023. Citado na página 1.
- FERRARI, H. **Número de investidores na B3 sobe para 5,8 milhões em 2022**. 2023. Disponível em: <<https://www.poder360.com.br/economia/numero-de-investidores-na-b3-sobe-para-58-milhoes-em-2022/>>. Acesso em: 30 de Maio de 2023. Citado na página 1.
- GOMES, P. **TRELLO X CLICK UP: VEJA SUAS CARACTERÍSTICAS**. 2022. Disponível em: <[https://organizenapratica.com.br/trello\\_ou\\_clickup/](https://organizenapratica.com.br/trello_ou_clickup/)>. Acesso em: 12 de Maio de 2023. Citado na página 21.
- GONSALES, T. **O que é Análise Técnica de Ações: Como Operar na Bolsa usando Análise Gráfica**. 2022. Disponível em: <<https://exame.com/invest/guia/analise-tecnica/>>. Acesso em: 06 de Abril de 2023. Citado na página 8.
- IBM. **IBM SPSS Modeler CRISP-DM Guide**. 2023. Versão 18.4. Disponível em: <[https://www.ibm.com/docs/en/SS3RA7\\_18.4.0/pdf/ModelerCRISPDM.pdf](https://www.ibm.com/docs/en/SS3RA7_18.4.0/pdf/ModelerCRISPDM.pdf)>. Acesso em: 01 de Maio de 2023. Citado 4 vezes nas páginas 15, 16, 17 e 18.
- INVESTOPEDIA. **Who Was Charles Dow?** 2023. 14 de Junho de 2025. Disponível em: <<https://www.investopedia.com/articles/financial-theory/08/charles-dow.asp>>. Citado na página 4.
- LANA, M. **Tudo Sobre Tempo Gráfico Na Análise Técnica**. 2016. Disponível em: <<https://www.bussoladoinvestidor.com.br/tempo-grafico-na-analise-tecnica/>>. Acesso em: 01 de Abril de 2023. Citado na página 4.
- MAIA, A. V. **Timing no mercado de ações no Brasil com padrões candlesticks e indicadores associados**. Dissertação (Mestrado) — Universidade Estadual de Feira de Santana, Feira de Santana, 2018. Citado na página 19.
- MATEUS, L. de S. **Técnicas de inteligência artificial e análise gráfica sobre o mercado de ações brasileiro**. Monografia (Graduação) — Universidade Tecnológica Federal do Paraná, 2022. Citado na página 19.
- NELOGICA. **O Indicador MACD**. 2020. Disponível em: <<https://www.nelogica.com.br/conhecimento/artigos/indicadores-estudo/indicador-macd>>. Acesso em: 01 de Janeiro de 2025. Citado na página 35.

NISON, S. **Japanese Candlestick Charting Techniques: A Contemporary Guide to the Ancient Investment Techniques of the Far East**. 1. ed. New York: [s.n.], 1991. Citado 6 vezes nas páginas 4, 8, 9, 10, 11 e 24.

OLIVEIRA, I. de. **Day trade: “gurus” iludem investidores com promessa de 1% ao dia de retorno**. 2020. Disponível em: <<https://einvestidor.estadao.com.br/comportamento/cursos-day-trade/>>. Acesso em: 30 de Maio de 2023. Citado na página 2.

PINTO, L. **Análise Fundamentalista x Análise Técnica: Entenda as diferenças entre os métodos para investir na Bolsa**. 2022. Disponível em: <<https://conteudos.xpi.com.br/aprenda-a-investir/relatorios/analise-fundamentalista-x-analise-tecnica-entenda-diferencas/>>. Acesso em: 30 de Maio de 2023. Citado na página 1.

PINTO, L. **O que é análise técnica: confira o guia completo**. 2022. Disponível em: <<https://conteudos.xpi.com.br/aprenda-a-investir/relatorios/o-que-e-analise-tecnica/>>. Acesso em: 01 de Abril de 2023. Citado 2 vezes nas páginas 4 e 5.

SCIKIT-LEARN. **3.1. Cross-validation: evaluating estimator performance**. 2025. Disponível em: <[https://scikit-learn.org/stable/modules/cross\\_validation.html](https://scikit-learn.org/stable/modules/cross_validation.html)>. Acesso em: 05 de Janeiro de 2025. Citado 2 vezes nas páginas 40 e 41.

SMARTTBOT. **HiLo Activator: como usá-lo em uma estratégia de investimento**. 2016. Disponível em: <<https://smarttbot.com/trader/indicador-hilo-activator-daytrade/>>. Acesso em: 20 de Maio de 2023. Citado 3 vezes nas páginas 26, 27 e 28.

SMARTTBOT. **Guia completo sobre robôs traders**. 2023. Disponível em: <<https://smarttbot.com/trader/guia-completo-sobre-robos-traders/>>. Acesso em: 23 de Maio de 2025. Citado na página 1.

TOROINVESTIMENTOS. **O que é o Mercado Futuro?** 2022. <<https://artigos.toroinvestimentos.com.br/trading/mercado-futuro>>. Acesso em: 28 de Abril de 2023. Citado 2 vezes nas páginas 12 e 13.

TRYD. **Automatizador**. 2023. Disponível em: <[https://www.tryd.com.br/manual/HTML/index.html?automatizador\\_do\\_tryd.htm](https://www.tryd.com.br/manual/HTML/index.html?automatizador_do_tryd.htm)>. Acesso em: 12 de Maio de 2023. Citado na página 20.

WILTGEN, J. **O que são ações e quais as vantagens e riscos desse investimento**. 2019. Disponível em: <<https://www.seudinheiro.com/2019/acoes/o-que-sao-acoes/>>. Acesso em: 04 de Abril de 2023. Citado na página 12.