



UNIVERSIDADE FEDERAL DO CEARÁ
CAMPUS DE SOBRAL
CURSO DE GRADUAÇÃO EM ENGENHARIA DA COMPUTAÇÃO

LUCAS NATANAEL LEITE FONTELES

**MODELO DE PREVISÃO DO MERCADO FINANCEIRO UTILIZANDO TÉCNICAS
DE APRENDIZADO DE MÁQUINA E ÁRVORES DE DECISÃO**

SOBRAL

2024

LUCAS NATANAEL LEITE FONTELES

MODELO DE PREVISÃO DO MERCADO FINANCEIRO UTILIZANDO TÉCNICAS DE
APRENDIZADO DE MÁQUINA E ÁRVORES DE DECISÃO

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia da Computação do Campus de Sobral da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Engenharia da Computação.

Orientador: Prof. Dr. José Cláudio do Nascimento

SOBRAL

2024

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

L554m Leite Fonteles, Lucas Natanael.
MODELO DE PREVISÃO DO MERCADO FINANCEIRO UTILIZANDO TÉCNICAS DE
APRENDIZADO DE MÁQUINA E ÁRVORES DE DECISÃO / Lucas Natanael Leite Fonteles. – 2024.
50 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Sobral,
Curso de Engenharia da Computação, Sobral, 2024.

Orientação: Prof. Dr. José Cláudio do Nascimento.

1. Bolsa de valores. 2. Árvore de decisão. 3. Aprendizado de máquina. I. Título.

CDD 621.39

LUCAS NATANAEL LEITE FONTELES

MODELO DE PREVISÃO DO MERCADO FINANCEIRO UTILIZANDO TÉCNICAS DE
APRENDIZADO DE MÁQUINA E ÁRVORES DE DECISÃO

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia da Computação do Campus de Sobral da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Engenharia da Computação.

Aprovada em:

BANCA EXAMINADORA

Prof. Dr. José Cláudio do Nascimento (Orientador)
Universidade Federal do Ceará (UFC)

Dr. Francisco Leonardo Bezerra Martins (Avaliador)
Universidade Federal do Ceará (UFC)

Me. Ítalo Rossi Araújo Costa (Avaliador)
Universidade Federal do Ceará (UFC)

Dedico a Deus, o criador dos céus e da terra. A mim por não ter desistido. À minha família, por sua capacidade de acreditar em mim. Em especial, à Mayara, Conceição, Valdecy, Lauanne, Layanne e Anderson.

AGRADECIMENTOS

Agradeço primeiramente a Deus. A mim por não ter desistido. À minha família, por sua capacidade de acreditar em mim. Em especial, à Mayara, Conceição, Valdecy, Lauanne, Layanne e Anderson.

Durante esse período de graduação algumas pessoas agregaram bastante em minha vida no processo de formação profissional e pessoal. Então, agradeço a todos os Profissionais da Universidade Federal do Ceará (Campus Mucambinho). Em especial, gostaria de agradecer ao Prof. Dr. José Cláudio do Nascimento por me orientar em minha tese de graduação e pelos conselhos que vão além da faculdade, agradeço pela sua companhia e pela paciência. Que o Criador possa te abençoar imensamente.

Agradeço também ao Prof. Dr. Jarbas Joaci De Mesquita Sá Junior, Michelle Pontes Fontenele Sousa, Felipe Gomes, Bryan S. Sousa, Victor Veríssimo Pereira Lopes, Augusto Coelho, Carlos Eduardo, Willian Praciano, Maxela Martins e ao José Paulo dos Santos Neto. Igualmente agradeço ao Edinardo Moreira Rodrigues (Doutorando em Engenharia Elétrica) e ao seu assistente, Alan Batista de Oliveira (Aluno de graduação em Engenharia Elétrica), pela adequação do *template* utilizado neste trabalho para que ficasse de acordo com as normas da biblioteca da Universidade Federal do Ceará (UFC).

"No mundo tereis aflições, mas tende bom
ânimo, eu venci o mundo." (João 16:33b).

RESUMO

Este trabalho verifica o desenvolvimento e a aplicação de um modelo preditivo baseado em aprendizado de máquina, utilizando árvores de decisão, para auxiliar investidores na tomada de decisões assertivas no mercado de ações. A ferramenta foi projetada para prever movimentos de preços de ativos financeiros, empregando técnicas de análise técnica e indicadores. O objetivo principal é oferecer uma base de suporte à decisão para investidores, especialmente para aqueles com menos experiência, contribuindo para reduzir erros e melhorar o desempenho de suas carteiras de investimento. Além disso, a ferramenta auxilia na gestão emocional do investidor, reduzindo vieses comportamentais que frequentemente levam a decisões impulsivas e potencialmente prejudiciais. O modelo utiliza dados históricos do ativo, processados com algoritmos de aprendizado de máquina para identificar padrões e tendências futuras. O desempenho do modelo foi avaliado em relação a uma estratégia passiva de investimento, demonstrando que a abordagem proposta gerou retornos superiores ao *benchmark*. A combinação de aprendizado de máquina com indicadores técnicos permitiu um aumento na acurácia das previsões, proporcionando uma vantagem competitiva em um mercado volátil. Ademais, o trabalho destaca a importância do uso de tecnologias como aprendizado de máquina e inteligência artificial no contexto financeiro atual, onde possui a capacidade de processar grandes volumes de dados e extrair *insights* valiosos que são cruciais para a tomada de decisões mais eficientes. A utilização desta ferramenta mostra-se promissora para democratizar o acesso a análise financeira, beneficiando investidores de todos os níveis de experiência.

Palavras-chave: Árvore de decisão. Aprendizado de máquina. Bolsa de valores.

ABSTRACT

This work examines the development and application of a predictive model based on machine learning, using decision trees, to assist investors in making assertive decisions in the stock market. The tool was designed to predict price movements of financial assets by employing technical analysis techniques and indicators. The main goal is to provide a decision support base for investors, particularly for those with less experience, contributing to the reduction of errors and improving the performance of their investment portfolios. Additionally, the tool helps in managing investor emotions, reducing behavioral biases that often lead to impulsive and potentially harmful decisions. The model uses historical asset data, processed with machine learning algorithms to identify patterns and future trends. The model's performance was evaluated against a passive investment strategy, demonstrating that the proposed approach generated returns superior to the benchmark. The combination of machine learning with technical indicators allowed for an increase in forecast accuracy, providing a competitive edge in a volatile market. Furthermore, the study highlights the importance of using technologies such as machine learning and artificial intelligence in the current financial context, where they have the capacity to process large volumes of data and extract valuable insights that are crucial for more efficient decision-making. The use of this tool proves promising for democratizing access to financial analysis, benefiting investors of all experience levels.

Keywords: Decision tree. Machine learning. Stock exchange.

LISTA DE FIGURAS

Figura 1 – Gráfico da evolução dos investidores.	14
Figura 2 – Consulta ao ticker VALE3.SA.	24
Figura 3 – Indicadores.	26
Figura 4 – Exemplo de implementação em uma árvore binária de busca.	28
Figura 5 – Exemplo de nível de uma árvore binária de busca.	29
Figura 6 – Exemplo de instanciação do modelo.	29
Figura 7 – Árvore.	36
Figura 8 – Parâmetros da árvore.	37
Figura 9 – Simulação do retorno.	39
Figura 10 – Simulação do retorno em código (vantagem).	39
Figura 11 – Visualização gráfica.	40
Figura 12 – Acurácia da VALE3.SA.	42
Figura 13 – Retorno Acumulado e <i>Benchmark</i>	43
Figura 14 – Gráfico da evolução VALE3.SA.	43

LISTA DE TABELAS

Tabela 1 – Perfil dos investidores.	15
Tabela 2 – Consulta ao código de negociação VALE3.SA.	25
Tabela 3 – Simulação do retorno.	39

LISTA DE ABREVIATURAS E SIGLAS

Adj. Close	Adjusted Close
AM	Aprendizado de Máquina
ANBIMA	Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais
B3	Bolsa de Valores do Brasil
BACEN	Banco Central do Brasil
CETIC BR	Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação
CR	<i>Cumulative Return</i> ou Retorno Cumulativo
EMA	<i>Exponential Moving Average</i> ou Média Móvel Exponencial
GI	Ganho de Informação
HMA	<i>Hull Moving Average</i> ou Média Móvel Hull
IA	Inteligência Artificial
MMP	Média Móvel Ponderada
PCA	<i>Principal Component Analysis</i> ou Análise de Componentes Principais
ROC	<i>Rate of Change</i> ou Taxa de Variação
RSI	<i>Relative Strength Index</i> ou Índice de Força Relativa
SMA	<i>Simple Moving Average</i> ou Média Móvel Simples
Taxa Selic	Taxa básica de juros da economia
Vale	Vale S.A.
yf	yfinance

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Motivação	17
1.2	Objetivos	18
1.2.1	<i>Objetivos gerais</i>	18
1.2.2	<i>Objetivos específicos</i>	18
2	FUNDAMENTAÇÃO TEÓRICA	19
2.1	Análises de mercado	19
2.1.1	<i>Análise fundamentalista</i>	19
2.1.2	<i>Análise Técnica</i>	19
2.1.3	<i>Indicadores</i>	20
2.1.3.1	<i>RSI - Relative Strength Index (Índice de Força Relativa)</i>	20
2.1.3.2	<i>ROC - Rate of Change (Taxa de variação)</i>	21
2.1.3.3	<i>SMA - Simple Moving Average (Média Móvel Simples)</i>	21
2.1.3.4	<i>EMA - Exponential Moving Average (Média Móvel Exponencial)</i>	22
2.1.3.5	<i>HMA - Hull Moving Average (Média Móvel Hull)</i>	22
2.2	Ferramentas utilizadas	23
2.2.1	<i>Python</i>	23
2.2.2	<i>Google Colab</i>	23
3	METODOLOGIA	24
3.1	Árvore de Decisão	28
3.1.1	<i>Fundamentos da Teoria da Informação</i>	30
3.1.2	<i>Ganho de Informação</i>	31
3.1.3	<i>Gini</i>	31
3.1.3.1	<i>Fórmula e cálculo</i>	32
3.1.3.2	<i>Seleção dos atributos em árvores de decisão</i>	32
3.1.3.3	<i>Comparação com Outros Critérios (Entropia)</i>	32
3.2	Dados	33
3.2.1	<i>Bibliotecas</i>	34
3.2.2	<i>Pré-processamento dos dados</i>	34
3.2.3	<i>DataFrame</i>	34

3.2.4	<i>Tipos de aprendizado</i>	35
3.2.4.1	<i>Aprendizado supervisionado</i>	35
3.2.4.2	<i>Aprendizado não supervisionado</i>	35
3.2.4.3	<i>Aprendizado semissupervisionado</i>	35
3.2.4.4	<i>Aprendizado por reforço</i>	36
3.2.5	<i>Treinamento do modelo</i>	36
4	ANÁLISE DOS DADOS	38
4.1	Avaliação dos dados	38
4.2	Visualização dos dados	40
5	RESULTADOS	42
6	DISCUSSÕES	45
7	CONCLUSÕES E TRABALHOS FUTUROS	47
	REFERÊNCIAS	48

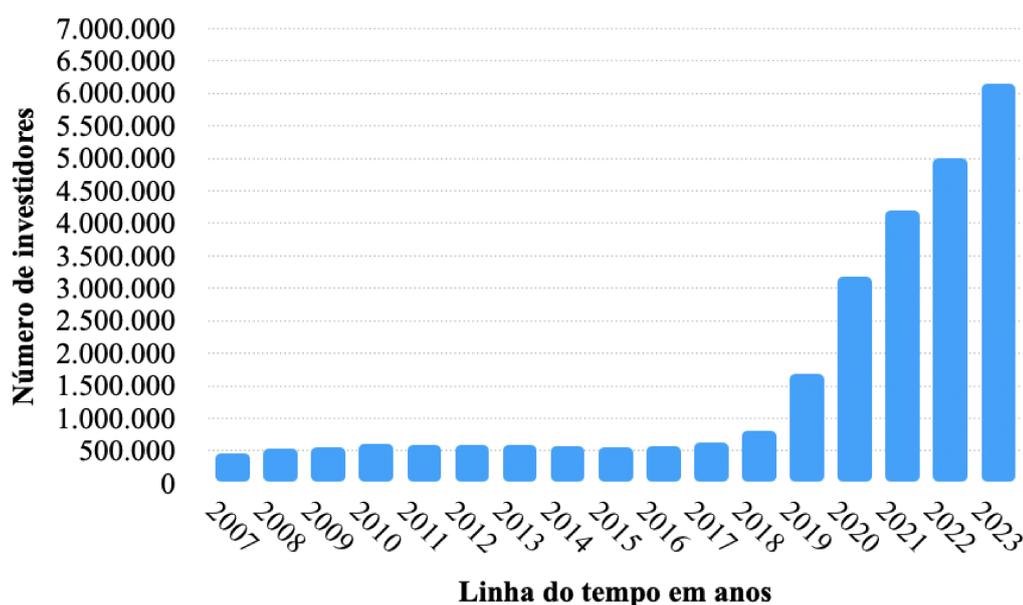
1 INTRODUÇÃO

Nos últimos anos, observa-se que o número de pessoas físicas investidoras na Bolsa de Valores do Brasil (B3) era significativamente baixo em relação à população total. Em 2007, segundo a Info Money (MONEY, 2019), havia pouco mais de 456.557 investidores. Vale ressaltar que, no passado, as negociações eram feitas por telefone (SUNO, 2019). Em meados de 2007, de acordo com o Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (CETIC BR), a internet só estava disponível em 17% do território brasileiro (CETIC, 2008).

Grande parte da população brasileira não tem conhecimento ou domínio sobre investimentos (USP, 2019). A popularização da Taxa básica de juros da economia (Taxa Selic) e suas variações pelo Banco Central do Brasil (BACEN), o avanço da internet, a ascensão dos influenciadores e a disseminação de conteúdo financeiro em plataformas digitais, principalmente as gratuitas como o YouTube, favorecem o aprendizado e incentivam, direta ou indiretamente, a população a investir.

Em pouco tempo, o número de investidores atingiu o primeiro milhão, e atualmente, esse número é de um pouco mais de 6 milhões, alocados em renda variável (B3, 2023). Segundo a Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais (ANBIMA), em um dado publicado em 2023 referente ao ano anterior, o volume dos investimentos dos brasileiros por parte das pessoas físicas no Brasil acumula 5 trilhões de reais (ANBIMA, 2023).

Figura 1 – Gráfico da evolução dos investidores.



Fonte: Elaborado pelo autor (2023).

Nota: Dados extraídos da B3.

Dentre os investidores que estão custodiados na B3, pode-se observar a faixa etária dos investidores através da Tabela 1 abaixo, que mostra o perfil dos investidores. Este dado é de junho de 2023.

Tabela 1 – Perfil dos investidores.

Faixa Etária	Homens	Mulheres
Até 15 anos	17.620	13.298
De 16 a 25 anos	635.494	119.267
De 26 a 35 anos	1.600.077	449.255
De 36 a 45 anos	1.389.437	411.648
De 46 a 55 anos	529.411	195.362
De 56 a 65 anos	306.021	122.386
Maior de 66 anos	210.225	101.990

Fonte: Elaborado pelo autor (2023).

Nota: Dados extraídos da B3.

Observa-se um crescimento no número de investidores a partir de 2018 (NETO *et al.*, 2022), o que pode implicar em um aumento no volume de transações e, conseqüentemente, no volume financeiro. O mercado da bolsa de valores brasileira possui grande relevância para o país. No âmbito internacional, a B3 é considerada a maior bolsa de valores da América Latina, segundo a Info Money (MONEY, 2023). A B3 tem chamado a atenção de muitos investidores, em especial aos investidores que buscam diversificar suas carteiras por meio da renda variável.

Com o aumento da base de investidores, como demonstrado na Figura 1, principalmente da faixa etária de 26 a 35 anos, na Tabela 1, surge a necessidade de utilizar ferramentas que auxiliam na tomada de decisões financeiras de maneira assertiva. Neste contexto, as técnicas de AM, como as árvores de decisão, oferecem uma oportunidade de apoiar esse crescente grupo de investidores com análises automatizadas e previsões baseadas em dados, auxiliando na tomada de decisões mais eficientes. De acordo com (MACHADO; CORRÊA, 2022), utilizando recursos de Inteligência Artificial é possível realizar análises preditivas de preços com base em séries históricas.

Com base em (GÉRON, 2021), AM é a ciência da programação de computadores de maneira que eles possam aprender com os dados. As ferramentas que simplificam o acesso as informações complexas podem não apenas ajudar a evitar erros de julgamento, mas também melhorar a performance das carteiras de investimento, especialmente para aqueles que estão começando no mercado de ações. Ainda mais, pode-se melhorar significativamente o desempenho

dos investimentos em suas carteiras, ao proporcionar uma base sólida para as decisões.

É importante salientar que investir na bolsa de valores exige dedicação, esforço e tempo para a avaliação criteriosa dos ativos e tomada de decisões adequadas. Segundo (PASSOS, 2018), é necessário ter conhecimento sobre o assunto para não acabar prejudicando-se. Um indivíduo com pouco conhecimento sobre um ativo específico e o mercado em geral corre o risco de tomar decisões equivocadas, seja na compra ou venda de um ativo, o que pode impactar negativamente sua carteira de investimentos. As decisões precipitadas podem gerar grandes prejuízos, principalmente em fundos geridos por gestores onde o impacto de uma má escolha pode ser ainda maior.

Para auxiliar nesse processo de tomada de decisões, surgem tecnologias como a Inteligência Artificial (IA). De acordo com (COPPIN, 2004), a IA é o estudo de sistemas que agem de uma maneira que, para qualquer observador, pareça ser inteligente, ou seja, a IA é capaz de realizar tarefas que exigem a inteligência humana. Tarefas nas quais podem exigir raciocínio, resolução de problemas, percepções, aprendizado e uso da linguagem.

Acredita-se que a ascensão da IA será cada vez mais relevante no decorrer dos anos. Segundo (LEE, 2019), ferramentas que usam IA mudarão a forma de como enxergamos o mundo, a forma como amamos, como nos relacionamos, trabalhamos e vivemos. Estima-se que haverá uma simbiose no mercado entre tarefas de otimização, que seria, o toque da máquina e toque humano, onde as ferramentas que utilizem técnicas de IA irão ficar cada vez mais comuns e essenciais.

Segundo (COPPIN, 2004), um subcampo da inteligência artificial é o Aprendizado de Máquina (AM), que é focado na construção de sistemas que podem aprender e melhorar a partir de seus próprios dados. O AM é, portanto, uma aplicação de IA que fornece aos sistemas a capacidade de aprender e melhorar. Entretanto, vale ressaltar que existem alguns tipos de aprendizados que serão mencionados em (3.2.4.1).

O AM envolve o desenvolvimento de algoritmos que têm a possibilidade de aprender com os dados e melhorar seu desempenho ao longo do tempo (GÉRON, 2021). No contexto da previsão de preços, os modelos de AM podem analisar grandes volumes de dados históricos, identificar padrões (que são componentes fundamentais na análise e interpretação de dados complexos), detectar tendências e fazer previsões com base nesses padrões. Uma análise gráfica de uma ação na bolsa de valores representa uma série temporal. De acordo com (MCKINNEY, 2018), qualquer dado medido em vários pontos no tempo compõe uma série temporal. No caso

das ações na bolsa de valores, as séries temporais geralmente possuem frequência fixa, em que a entrada de dados ocorre em intervalos regulares de acordo com alguma regra.

Há algumas técnicas de AM que podem ser aplicadas à previsão de preços. Os exemplos disso seriam: a regressão linear, redes neurais, árvores de decisão e algoritmos genéticos. Cada técnica tem suas próprias características e limitações. A escolha de uma técnica adequada dependerá do contexto e dos objetivos específicos. Além disso, a precisão e a eficiência das previsões dependem de vários fatores, como: a qualidade e a quantidade dos dados disponíveis, a escolha correta das variáveis de entrada, a escolha dos parâmetros do modelo e a validação adequada do desempenho do modelo (MÜLLER; SARAH, 2016).

Segundo (MACHADO; CORRÊA, 2022), o mercado financeiro é uma alternativa financeira para maiores ganhos, por outro lado também traz maiores riscos. Entretanto, é importante destacar que nem sempre uma técnica de previsão fornecerá informações precisas. Os mercados financeiros são caracterizados por sua complexidade e volatilidade, sendo influenciados por diversos fatores imprevisíveis, como notícias econômicas, eventos políticos, mudanças regulatórias e até mesmo aspectos comportamentais dos investidores. Esses fatores podem fazer com que os movimentos de preços se afastem dos padrões históricos, desafiando as previsões baseadas em análise técnica ou em modelos de aprendizado de máquina (AM). Desse modo, é fundamental compreender que, embora essas técnicas não garantam sucesso nos mercados financeiros, elas oferecem *insights* valiosos ao identificar padrões e tendências com base em dados históricos, contribuindo para a tomada de decisões.

1.1 Motivação

Investidores frequentemente se deparam com fortes emoções que podem influenciar negativamente suas decisões. A aplicação de árvores de decisão no aprendizado de máquina oferece uma solução eficaz para mitigar esses vieses emocionais e promover a racionalidade na tomada de decisões. Outrossim, pesquisas demonstram a viabilidade de construir e automatizar modelos matemáticos e probabilísticos para auxiliar na análise e previsão do comportamento do mercado financeiro. Diante dessa necessidade, o desenvolvimento de uma ferramenta que utilize aprendizado de máquina com árvores de decisão se configura como uma proposta promissora para aprimorar o desempenho de carteiras de investimentos.

1.2 Objetivos

1.2.1 *Objetivos gerais*

Este trabalho tem como objetivo principal a construção de uma ferramenta para auxiliar investidores na tomada de decisões assertivas em investimentos de renda variável. A ferramenta tem como objetivo principal identificar possíveis tendências e futuros movimentos no mercado financeiro.

A ferramenta apresentará indicadores relevantes, derivados de análises gráficas avançadas, que consideram o histórico de flutuações de preços nos ativos (*tickers*). Para tanto, será empregado processamento computacional robusto, utilizando algoritmos de aprendizado de máquina, com foco em árvores de decisão.

1.2.2 *Objetivos específicos*

- Coletar histórico de negociação do ativo de interesse.
- Pré-processar os dados.
- Aplicar indicadores técnicos relevantes.
- Dividir os dados em conjuntos de treinamento e teste.
- Treinar um modelo de árvore de decisão utilizando o conjunto de treinamento.
- Avaliar o desempenho do modelo utilizando métricas de acurácia e outras métricas relevantes.

2 FUNDAMENTAÇÃO TEÓRICA

A ferramenta tem como objetivo auxiliar na análise financeira e tomada de decisões no mercado de ações. A aplicação combina análise técnica, aprendizado de máquina e visualização de dados para auxiliar na tomada de decisões financeiras. A sua estrutura é definida por bibliotecas, pré-processamento de dados, *DataFrame*, treinamento do modelo, avaliação dos resultados e a visualização dos resultados.

2.1 Análises de mercado

Existem algumas linhas de análises dentro deste universo de ativos de renda variável, exemplos disso seriam: a análise técnica e a análise fundamentalista.

2.1.1 *Análise fundamentalista*

Segundo (LEMOS, 2018), a análise fundamentalista utiliza modelos matemáticos, para calcular o fluxo de caixa futuro e trazê-lo a valor presente, permitindo assim projetar o preço-alvo de uma ação.

De acordo com os conceitos de (GRAHAM, 2020), a análise fundamentalista é uma ferramenta que tem como objetivo principal determinar o valor intrínseco de uma ação, com o objetivo de avaliar se o preço de mercado reflete esse valor. Graham defendeu a compra de ativos abaixo do seu valor intrínseco, proporcionando uma margem de segurança que poderia aumentar a possibilidade de rentabilidade a longo prazo.

Na análise fundamentalista faz-se uma avaliação geral da empresa em aspectos, como a situação financeira da empresa. Se ela tem bens, se sim quais seriam. Se ela tem dívidas de uma maneira detalhada. Quais os projetos que estão em execução e os planos futuros. Este tipo de análise considera os fundamentos da empresa.

2.1.2 *Análise Técnica*

De acordo com (SIQUEIRA, 2020), baseia seu estudo nos dados históricos das ações de uma determinada empresa no mercado, através de ferramentas gráficas e cálculos matemáticos, de forma a estabelecer padrões de comportamento nas variações de preço de um ativo. Em que *traders* e investidores buscam identificar oportunidades de lucro com base em

padrões e tendências históricas. Ao estudar gráficos de preços, indicadores técnicos e modelos matemáticos, os profissionais do mercado buscam prever movimentos futuros de preços e tomar decisões de compra ou venda de ativos. A análise técnica é uma abordagem baseada na premissa de que os preços dos ativos refletem todas as informações disponíveis sobre eles, portanto a análise pode ser usada para identificar padrões e tendências recorrentes. Os analistas técnicos usam uma variedade de indicadores e ferramentas, como médias móveis, osciladores, bandas de *Bollinger* e retrações de *Fibonacci*, para interpretar movimentos de preços e fazer previsões.

Ainda mais, a análise técnica leva em consideração conceitos como suporte e resistência, que são níveis de preço nos quais um ativo tende a inverter a direção. Esses níveis são estabelecidos com base em padrões anteriores e são considerados áreas de grande interesse para os *traders*, pois representam pontos potenciais de entrada ou saída do mercado.

É importante frisar que a análise técnica não se baseia em fundamentos econômicos ou informações financeiras das empresas. É concentrada principalmente na análise de gráficos de preços e padrões históricos, tendo uma possibilidade de ser considerado uma limitação, já que não leva em consideração variáveis econômicas, políticas ou outras externas que possam afetar os preços dos ativos.

2.1.3 Indicadores

De acordo com (MELO, 2022) os indicadores são utilizados para auxiliar os investidores nas análises gráficas, dando uma perspectiva diferente e, conseqüentemente, ajudando nas tomadas de decisões. Dentre vários indicadores, alguns mostram a tendência do mercado, probabilidade de reversão, volatilidade, retração e vários outros aspectos que auxiliam as análises técnicas dos investidores.

2.1.3.1 RSI - *Relative Strength Index* (Índice de Força Relativa)

O RSI é um indicador popular de *momentum* de preço e foi desenvolvido por Welles Wilder. Tem como objetivo medir a força ou fraqueza de um ativo, comparando a magnitude dos ganhos recentes com as perdas recentes ao longo de um período específico. De acordo com (AMARANTE, 2021), o seu objetivo consiste em "detectar" o mais cedo possível eventuais momentos de compra/venda. O RSI é um indicador que funciona melhor em momentos de volatilidade. Inicialmente, o *Relative Strength Index* ou Índice de Força Relativa (RSI) foi desenvolvido para avaliar um período de até 14 dias. Porém, ele pode avaliar outros períodos.

Define-se RSI:

$$RSI = 100 - \left[\frac{100}{1 + RS} \right] \quad (2.1)$$

$$RS = \frac{\text{Média de retornos positivos de } X \text{ dias}}{\text{Média de retornos negativos de } X \text{ dias}} \quad (2.2)$$

2.1.3.2 ROC - Rate of Change (Taxa de variação)

O indicador *Rate of Change* ou Taxa de Variação (ROC) é responsável por medir a variação percentual dos preços de fechamento, identifica tendências além de possíveis reversões, que podem ser usadas a favor do operador do ativo para tomadas de decisão. Segundo (COELHO, 2020), pode ser usado para se verificar divergências, como situações em que o ativo possa estar sobrecomprado ou sobrevendido. Os valores altos e baixos históricos do ROC funcionam como as regiões sobrecompradas e sobrevendidas, respectivamente.

Define-se o ROC:

$$ROC = \frac{(Pt - PT)}{PT} * 100 \quad (2.3)$$

Abaixo temos referência sobre o resultado do ROC. Vale ressaltar que as variações são limitadas a 100% abaixo e sem limites para cima de zero.

- I. ROC > 0 ascendente indica manutenção de uma tendência de alta.
- II. ROC < 0 descendente indica manutenção de uma tendência de baixa.
- III. ROC cruzando linha Zero de baixo para cima - oportunidade de compra.
- IV. ROC cruzando linha Zero de cima para baixo - oportunidade de venda.

2.1.3.3 SMA - Simple Moving Average (Média Móvel Simples)

Segundo (MENEZES *et al.*,), o indicador *Simple Moving Average* ou Média Móvel Simples (SMA), consegue atenuar as variações e produzir uma linha que nos aponta uma valorização ou uma desvalorização de uma maneira constante, e consiste na soma dos valores de um período, dividido pelo próprio período.

Define-se o SMA:

$$SMA_n = \frac{\sum_{i=1}^n \text{valor}_i}{n} \quad (2.4)$$

- I. SMA: será a média móvel.
- II. n: será o tamanho da janela que será utilizada
- III. valor: será o preço do ativo.

2.1.3.4 EMA - Exponential Moving Average (Média Móvel Exponencial)

Na média móvel exponencial, os dados recentes tem um peso maior quando comparado com os dados mais antigos. A *Exponential Moving Average* ou Média Móvel Exponencial (EMA), considerando o mesmo número de períodos, a média móvel exponencial acompanha as variações mais rapidamente do que a média móvel simples.

Define-se o EMA:

$$EMA = MME(n) = \alpha \times P(n) + (1 - \alpha) \times MME(n - 1) \quad (2.5)$$

- I. $P(n)$: é o preço do ativo no período n .
- II. $MME(n - 1)$: é a média móvel exponencial do período anterior $n - 1$.
- III. α : é o fator de suavização.

O fator de suavização, pode ser calculado:

$$\alpha = \frac{2}{N + 1} \quad (2.6)$$

- I. α : é o fator de suavização.
- II. N : é o número de períodos utilizados para calcular a MME.

Segundo (LOPES, 2024), EMA dá mais ênfase aos preços mais atuais, reagindo mais rapidamente às mudanças no mercado e atribui mais peso aos preços recentes, tornando-a mais sensível a mudanças recentes no preço. O (VOORSLUYS, 2006), acredita que o uso de EMA pode ser visto como uma forma de adicionar mais qualidade na informação.

2.1.3.5 HMA - Hull Moving Average (Média Móvel Hull)

Segundo o trader australiano (HULL, 1990) a *Hull Moving Average* ou Média Móvel Hull (HMA), faz o uso da média ponderada e da raiz quadrada do período analisado. Este

indicador é muito utilizado para identificar pontos de saída das operações, também considera-se o valor inteiro retornado da raiz quadrada e da divisão sem as porções decimais.

Define-se o HMA:

$$HMA_p = MMP \left(2 \times MMP \left(\text{Valor}, \frac{p}{2} \right) - MMP(\text{Valor}, p), \sqrt{p} \right) \quad (2.7)$$

- I. MMP: será a média móvel ponderada.
- II. Valor: será o resultado da Média Móvel Ponderada (MMP) no período.
- III. p: será o período.

2.2 Ferramentas utilizadas

Utilizou-se a linguagem de programação *Python* e a plataforma do *Google Colab* como ambiente de desenvolvimento. Nos pontos 2.2.1 e 2.2.2 encontra-se mais detalhes.

2.2.1 *Python*

Criada em 1991 por Guido Van Rossum, *Python* (PYTHON, 2023) é uma linguagem de programação interpretada, de alto nível, de código aberto e orientada a objetos. Amplamente usada, possuindo uma grande portabilidade com outras linguagens de programação sendo versátil e eficiente para o desenvolvimento de projetos.

De acordo com (MENEZES, 2014), existem várias vantagens em utilizar a linguagem *Python*, entre as quais se destacam sua simplicidade e clareza. É uma linguagem em constante crescimento em diversas áreas da computação e apresenta boa legibilidade nos programas escritos em *Python*. Por ser uma linguagem completa, permite: o uso de bibliotecas para acessar bancos de dados, processar arquivos, construir interfaces gráficas, entre outras funcionalidades. Desta maneira, possibilitando a obtenção de resultados em pouco tempo.

2.2.2 *Google Colab*

Construído com base no *Jupyter notebook*, o *Google Colab* (GOOGLE, 2023) é um ambiente interativo em nuvem da *Google* onde permite um trabalho remoto de maneira consistente. O *Google Colab* possibilita a execução e edição de células individuais de código em *Python* de forma facilitada.

3 METODOLOGIA

Baseia-se em técnicas de aprendizado de máquina e análise técnica para a previsão de preços de ativos financeiros. A linguagem adotada foi *Python* e o ambiente de desenvolvimento foi a plataforma do *Google Colab*. Inicialmente, é realizada a coleta de dados históricos dos ativos por meio da biblioteca *yfinance*, que fornece acesso a informações detalhadas do mercado financeiro.

Figura 2 – Consulta ao ticker VALE3.SA.

```
import yfinance as yf
import pandas as pd

# Definindo o ticker da VALE3. SA
ticker = "VALE3.SA"
# Coletando os dados históricos
dados_historicos = yf.download(ticker, period="5y")
# Formatando a tabela para exibição com pandas de forma organizada
dados_historicos = dados_historicos.head()
# Exibindo a tabela organizada
dados_historicos

✓ 0.6s
```

Fonte: Elaborado pelo autor (2024).

Na Figura 2, mostra-se um recorte de um algoritmo que tem como finalidade a coleta dos dados da empresa Vale S.A. (Vale), cujo código de negociação é VALE3.SA. Este recorte ajuda compreender como é realizada a coleta dos dados. Na linha 1, faz-se a importação da biblioteca, *yfinance* como citada anteriormente, na qual é renomeada como *yfinance* (*yf*), o qual será útil para a consulta e coleta dos dados.

Na Tabela 2, tem-se uma demonstração dos dados do código de negociação VALE3.SA com *Date*, *Open*, *High*, *Low*, *Close*, *Adjusted Close* (*Adj. Close*) e *Volume*. Formando as colunas da tabela, estas colunas significam respectivamente:

- *Date*: representa o dia da negociação.
- *Open*: representa o preço de abertura da ação no seu início de negociação.
- *High*: representa o pico do preço que a ação atingiu.
- *Low*: representa o menor preço que a ação atingiu.

- *Close*: representa o último preço onde a ação foi negociada.
- *Adj. Close*: representa o preço de fechamento ajustado, levando em, consideração eventos corporativos, dividendos, desdobramentos e fusões.
- *Volume*: representa o número total de ações da VALE3.SA que foram negociadas neste dia.

Tabela 2 – Consulta ao código de negociação VALE3.SA.

	Open	High	Low	Close	Adj Close	Volume
Date						
2019-08-19	44.290001	44.430000	43.209999	43.650002	27.495708	27533700
2019-08-20	43.509998	44.490002	42.900002	43.840000	27.615387	16865100
2019-08-21	43.599998	44.330002	43.070000	44.150002	27.810659	16171800
2019-08-22	44.380001	44.380001	43.549999	43.889999	27.646881	12254400
2019-08-23	43.830002	44.590000	42.900002	43.279999	27.262636	26087800

Fonte: Elaborado pelo autor (2024).

Em seguida, são aplicados indicadores técnicos, como RSI (*Relative Strength Index*), ROC (*Rate of Change*), SMA e médias exponenciais para extrair informações relevantes dos dados históricos.

Na Figura 3, tem-se uma função para adicionar indicadores técnicos ao *Dataframe*, no qual contém dados do último preço negociado durante o dia e volume de negociações, ou seja, preço de fechamento (*close*) e volume. Para o volume é aplicado o RSI nos parâmetros 2, 3, 5, 8 e 13 para o preço de fechamento é aplicado RSI (2.1.3.1), ROC (2.1.3.2), SMA (2.1.3.3), EMA (2.1.3.4) e HMA (2.1.3.5). Estes parâmetros selecionados foram os que obtiveram melhores resultados nos testes realizados. Na penúltima linha desta função, tem-se como objetivo remover as linhas que contêm valores nulos após o cálculo dos indicadores. Isto é importante porque muitos indicadores dependem de um certo número de períodos de dados para serem calculados e podem gerar valores nulos nas primeiras linhas.

Após a criação dos indicadores, é realizado o pré-processamento dos dados, que inclui a remoção de valores nulos e a codificação da variável de destino.

A variável de destino representa a direção do mercado no dia seguinte, e é codificada como 1 para movimento de alta, -1 para movimento de baixa e 0 para movimento lateral. Em seguida, os dados são divididos em conjuntos de treinamento e teste, permitindo a avaliação do desempenho do modelo em dados não vistos.

Figura 3 – Indicadores.

```

def addIndicadores(df):

    close = df['Close']
    volume = df['Volume']

    df.loc[:, 'RSI2_Vol'] = ta.rsi(volume, 2)
    df.loc[:, 'RSI3_Vol'] = ta.rsi(volume, 3)
    df.loc[:, 'RSI5_Vol'] = ta.rsi(volume, 5)
    df.loc[:, 'RSI8_Vol'] = ta.rsi(volume, 8)
    df.loc[:, 'RSI13_Vol'] = ta.rsi(volume, 13)
    df.loc[:, 'RSI2_Close'] = ta.rsi(close, 2)
    df.loc[:, 'RSI3_Close'] = ta.rsi(close, 3)
    df.loc[:, 'RSI5_Close'] = ta.rsi(close, 5)
    df.loc[:, 'RSI9_Close'] = ta.rsi(close, 9)
    df.loc[:, 'RSI14_Close'] = ta.rsi(close, 14)
    df.loc[:, 'RSI25_Close'] = ta.rsi(close, 25)
    df.loc[:, 'RSI35_Close'] = ta.rsi(close, 35)
    df.loc[:, 'RSI56_Close'] = ta.rsi(close, 56)
    df.loc[:, 'ROC2_Close'] = ta.roc(close, 2)
    df.loc[:, 'ROC3_Close'] = ta.roc(close, 3)
    df.loc[:, 'ROC5_Close'] = ta.roc(close, 5)
    df.loc[:, 'ROC9_Close'] = ta.roc(close, 9)
    df.loc[:, 'ROC14_Close'] = ta.roc(close, 14)
    df.loc[:, 'ROC25_Close'] = ta.roc(close, 25)
    df.loc[:, 'ROC35_Close'] = ta.roc(close, 35)
    df.loc[:, 'ROC56_Close'] = ta.roc(close, 56)
    df.loc[:, 'SMA9_Close'] = ta.sma(close, length=9)
    df.loc[:, 'SMA14_Close'] = ta.sma(close, length=14)
    df.loc[:, 'SMA25_Close'] = ta.sma(close, length=25)
    df.loc[:, 'SMA50_Close'] = ta.sma(close, length=50)
    df.loc[:, 'SMA100_Close'] = ta.sma(close, length=100)
    df.loc[:, 'SMA200_Close'] = ta.sma(close, length=200)
    df.loc[:, 'EMA9_Close'] = ta.ema(close, length=9)
    df.loc[:, 'EMA14_Close'] = ta.ema(close, length=14)
    df.loc[:, 'EMA25_Close'] = ta.ema(close, length=25)
    df.loc[:, 'EMA50_Close'] = ta.ema(close, length=50)
    df.loc[:, 'EMA100_Close'] = ta.ema(close, length=100)
    df.loc[:, 'EMA200_Close'] = ta.ema(close, length=200)
    df.loc[:, 'HMA14_Close'] = ta.hma(close, length=14)

    df.dropna(axis=0, inplace=True)
    return df

```

Fonte: Elaborado pelo autor (2024).

A árvore de decisão é escolhida como modelo de aprendizado de máquina para este trabalho, devido a facilidade de interpretação e implementação. Além disso, árvores de decisão lidam bem com dados categóricos e numéricos tornando-se uma ferramenta versátil. Existem outras alternativas de modelos de aprendizado de máquina, como redes neurais artificiais, máquinas de vetores de suporte e regressão logística, que também podem ser utilizados para previsão de preços no mercado financeiro. Há limitações ao utilizar árvores de decisão, pois podem se tornar muito complexas e ajustar-se demais aos dados de treinamento, prejudicando o desempenho em novos dados (teste). Além disso, para dados com padrões altamente não lineares e complexos, as árvores de decisão podem ser superadas por modelos mais sofisticados, como redes neurais.

O método de treinamento da árvore é baseado nos ajustes supervisionados dos dados históricos pré-processados e rotulados, sendo treinada com base nos dados de treinamento. O critério de divisão da árvore pode ser selecionado, como o critério de Gini (3.1.3), que mede a impureza dos dados. Uma vez treinada, a árvore de decisão é utilizada para fazer previsões de movimentos de alta, baixa ou lateralização da ação com base nos indicadores técnicos sobre os dados de teste.

A avaliação do modelo é realizada por meio da métrica de acurácia, que mede a proporção de previsões corretas em relação ao total de previsões realizadas. Além disso, é gerado um gráfico para visualizar o desempenho do modelo ao longo do tempo, comparando o retorno cumulativo do modelo com um padrão de referência.

É importante ressaltar que o sucesso da metodologia depende da qualidade dos dados utilizados, da escolha adequada dos indicadores técnicos e do ajuste dos parâmetros do modelo. A interpretação correta dos resultados e a combinação com outras formas de análise, como a análise fundamentalista, também são essenciais para a tomada de decisões informadas.

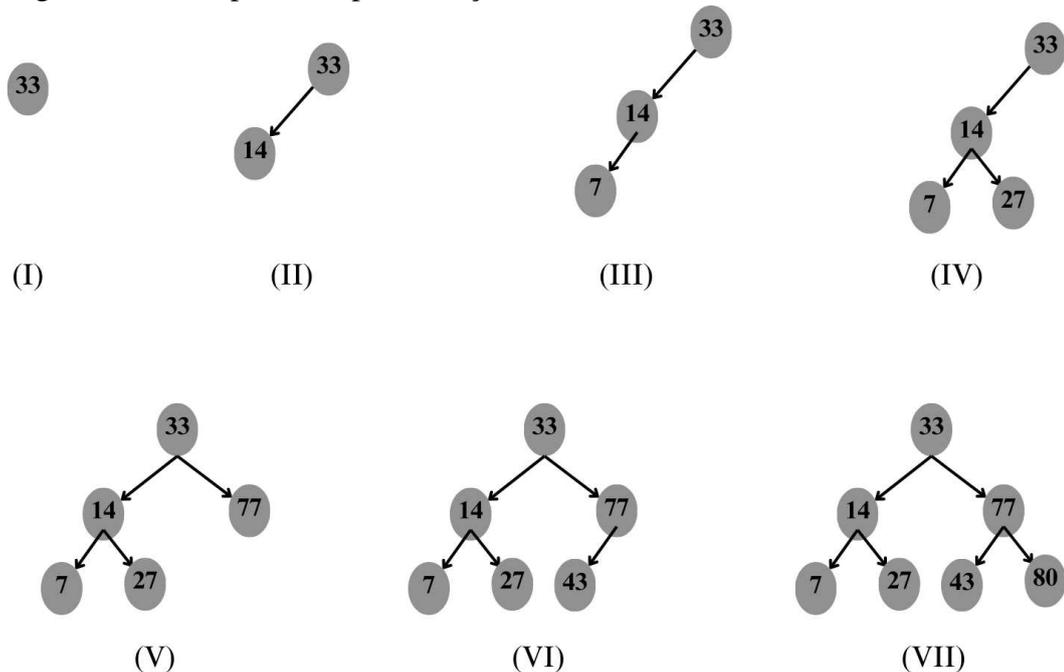
Em resumo, a metodologia adotada utiliza a classe *DecisionTreeClassifier* da biblioteca *scikit-learn*, que envolve a aplicação da técnica de aprendizado de máquina, baseadas em árvores de decisão juntamente com a utilização de indicadores técnicos e a análise de dados históricos para a previsão de preços de ativos financeiros. A árvore de decisão é utilizada como modelo de aprendizado de máquina, e a acurácia é empregada como métrica de avaliação. A metodologia oferece uma abordagem sistêmica e automatizada para auxiliar na tomada de decisões de compra e venda de ativos financeiros.

3.1 Árvore de Decisão

Uma árvore de decisão é uma estrutura semelhante a um fluxograma em que cada nó interno representa um teste de um atributo, cada ramo representa uma regra de decisão e cada nó folha representa um resultado. Salienta-se ainda que a árvore de decisão é uma técnica de aprendizado de máquina que usa modelos de decisão para tirar conclusões sobre o valor alvo de um projeto a partir de observações desse projeto. Eles são usados em problemas de classificação e regressão. Segundo (SIQUEIRA, 2020) as árvores de decisão dividem o problema em vários subproblemas a fim de chegar na melhor solução.

Com o objetivo de facilitar a compreensão do funcionamento de uma árvore binária de busca, será ilustrada na Figura 4 a implementação utilizando os números a seguir: 33, 14, 7, 27, 77, 43 e 80.

Figura 4 – Exemplo de implementação em uma árvore binária de busca.



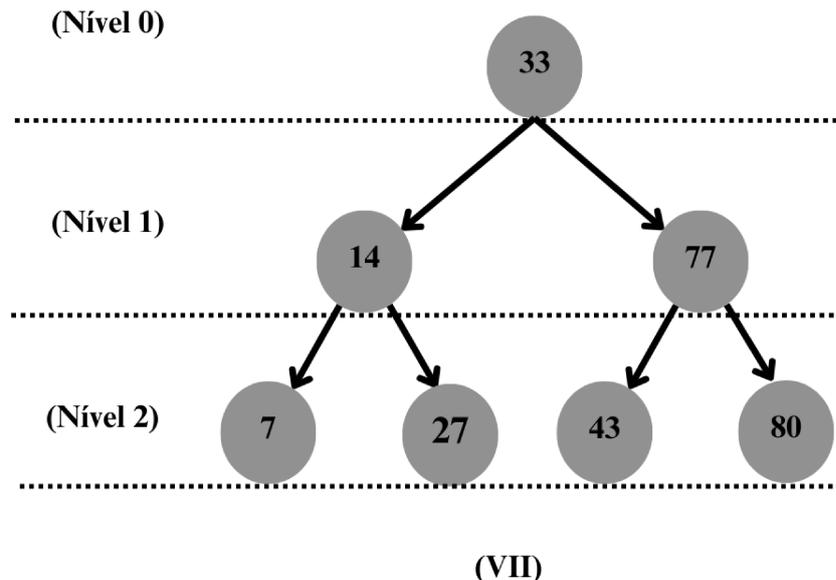
Fonte: Elaborado pelo autor (2024).

Na Figura 4, há um exemplo de uma árvore sendo implementada, onde o primeiro elemento será a raiz, neste exemplo temos o número 33. Seguindo a lista dos números para implementação, deve-se verificar se o próximo número da lista será maior ou menor em comparação com o seu pai. Neste primeiro momento usa-se o nó de número 14, por ser menor que seu nó 33 (nó pai) ficando alocado à esquerda como apresentado no item (II), se fosse maior que seu pai seria adicionado à direita. Ao adicionar o terceiro item da lista, observa-se que é um valor

menor que o nó folha atual, sendo então o novo nó folha adicionado a esquerda, pois é menor que seu pai e assim sucessivamente até o término da lista no item (VII) da Figura 4.

Ao inserir os números na ordem apresentada, iremos obter uma árvore de 2 níveis completa. Onde no nível 0 encontra-se o nó de número 33, no nível 1 tem-se os números 14 e 77 e no nível 2 os números 7, 27, 43 e 80, como pode-se observar na Figura 5.

Figura 5 – Exemplo de nível de uma árvore binária de busca.



Fonte: Elaborado pelo autor (2024).

A construção da árvore de decisão da aplicação segue o processo padrão de ajuste de um modelo em aprendizado de máquina. Na Figura 6 podemos observar a importação da biblioteca, criação do objeto árvore e a instanciação do modelo.

Figura 6 – Exemplo de instanciação do modelo.

```

from sklearn.tree import DecisionTreeClassifier

# Instanciando o modelo de árvore de decisão
arvore = DecisionTreeClassifier(criterion=criterio)
  
```

Fonte: Elaborado pelo autor (2024).

Ao construir uma árvore de decisão, é necessário escolher como dividir os dados em cada etapa. Critérios comuns para essa escolha são: entropia, ganho de informação e gini. A construção de uma árvore de decisão envolve a seleção iterativa de atributos para dividir o conjunto de dados em subconjuntos mais homogêneos. A parte crucial que permite a formação

dessas árvores é a análise estatística, que envolve diversos métodos e técnicas para extrair perspectivas valiosas dos dados históricos de mercado e usá-los para prever futuros resultados. A teoria da informação e do ganho de informação fornece uma base matemática para essa seleção, guiando o processo para a construção de árvores eficientes e precisas.

3.1.1 Fundamentos da Teoria da Informação

A teoria da informação, desenvolvida por Claude Shannon, quantifica a incerteza presente em um conjunto de dados. Este conceito pode ser aplicado em diversas áreas, entretanto na ciência da computação tem aplicações em processamento de linguagem natural onde extrai e analisa informações de textos, como tradução automática e reconhecimento de fala.

No aprendizado de máquina, treina-se modelos de aprendizado de máquina com base em grandes conjuntos de dados. Na inteligência artificial, desenvolve-se sistemas de inteligência artificial que podem aprender, raciocinar e tomar decisões. A teoria da informação tem conceitos básicos como a entropia, que é uma medida da incerteza ou imprevisibilidade de uma variável aleatória. Quanto maior a entropia, mais incerta ou imprevisível é a variável, assim mais informação a variável contém. Segundo (SHANNON, 1948), a equação (3.1) é referente à pureza da informação, ou seja, a entropia:

$$H(X) = - \sum_x p(x) \log_2(p(x)) \quad (3.1)$$

Aprofundando-se na equação acima, no qual representa a essência da entropia na Teoria da Informação, onde define-se a entropia (H) de uma variável aleatória discreta X, definida como a média ponderada da probabilidade de cada valor possível em um conjunto.

- $H(X)$: isso representa a entropia da variável X. É um valor não negativo.
- $p(x)$: representa a probabilidade de um valor específico (x) ocorrer. Por exemplo, se X representa o resultado do cara ou coroa de uma moeda, $p(\text{cara})$ seria a probabilidade de sair cara.
- $\log_2(p(x))$: este é o logaritmo (base 2) da probabilidade $p(x)$. O logaritmo ajuda a penalizar eventos altamente prováveis (onde $p(x)$ está próximo de 1) e enfatizar eventos menos prováveis (onde $p(x)$ está próximo de 0).

Para um melhor entendimento, imagine uma caixa cheia de bolas coloridas. Se todas as bolas forem vermelhas ($p(\text{vermelho}) = 1$), não há incerteza sobre a cor que você escolherá. A entropia (H) tende a zero nesse caso.

A entropia mede a incerteza ou imprevisibilidade de uma variável. O Ganho de Informação (GI) mede a redução da entropia ao dividir os dados com base em um atributo. Quanto maior o ganho de informação, mais útil é o atributo para separar os dados. Agora, imagine que a caixa tenha uma mistura de bolas vermelhas, azuis e verdes. Quanto mais uniformemente distribuídas as cores (o que significa que todas as probabilidades estão mais próximas umas das outras), mais difícil é prever qual cor você escolherá. Nesse cenário, a entropia (H) seria maior devido ao aumento da incerteza.

A equação basicamente calcula um fator médio de "surpresa" associado a cada resultado possível, ponderado por sua probabilidade. Quanto mais surpreendente (menos provável) for um resultado, maior sua contribuição para a incerteza geral (entropia). Esta fórmula é fundamental na Teoria da Informação e possui várias aplicações, como discutimos anteriormente. Ela nos permite quantificar o conteúdo de informação em uma mensagem, a eficiência de esquemas de compressão de dados e a complexidade de problemas de aprendizagem no aprendizado de máquina.

3.1.2 Ganho de Informação

O GI mede a redução da incerteza em um conjunto de dados X após a divisão em subconjuntos S_1, S_2, \dots, S_n , de acordo com os valores de um atributo. No âmbito do aprendizado de máquina, as árvores de decisão se consolidam como ferramentas indispensáveis para a classificação e o aprendizado a partir de dados. O cerne da construção robusta e eficaz dessas árvores reside na seleção criteriosa dos atributos que direcionam o processo de divisão. Neste contexto, o ganho de informação emerge como um conceito fundamental, fornecendo uma métrica para quantificar a relevância de cada atributo e, conseqüentemente, nortear a tomada de decisões na construção da árvore.

3.1.3 Gini

O índice de Gini é uma medida de impureza amplamente utilizada em algoritmos de aprendizado de máquina, especialmente em árvores de decisão. Ele quantifica o grau de mistura de classes em um determinado nó da árvore. No contexto das árvores de decisão, o objetivo é criar nós o mais puros possível, ou seja, nós onde as instâncias pertençam, preferencialmente, a uma única classe. Um nó é considerado puro quando todas as suas instâncias pertencem a uma única classe, e a impureza máxima ocorre quando as instâncias estão distribuídas igualmente

entre as classes. O índice de Gini é útil porque permite que o algoritmo de árvore de decisão escolha a melhor divisão de dados em cada estágio, buscando minimizar a impureza e criar divisões que separem de forma eficiente as diferentes classes. De acordo com (SHU, 2021) o índice de Gini também pode ser aplicado para avaliar se uma divisão é adequada ou não. O índice Gini pode ser calculado através da equação (3.2) do subtópico (3.1.3.1).

3.1.3.1 Fórmula e cálculo

O índice de Gini é dado por:

$$Gini(p) = 1 - \sum_{i=1}^C p(i)^2 \quad (3.2)$$

- $p(i)$ representa a probabilidade de um elemento ser classificado para uma classe i no nó.
- C é o número de classes.

O valor de $Gini(p)$ pode variar de 0, que seria o valor puro do nó e um valor máximo que vai depender do número de classes podendo ser 1. O objetivo da árvore de decisão é a minimização do índice Gini em cada divisão, para criar os nós mais homogêneos.

3.1.3.2 Seleção dos atributos em árvores de decisão

No processo de construção de uma árvore de decisão, o algoritmo utiliza o índice de Gini para decidir qual atributo dividir em cada nó. O atributo que resulta na maior redução de impureza, ou seja, que gera os nós mais puros após a divisão, é escolhido.

Durante a criação da árvore, cada vez que os dados são divididos, o índice de Gini é calculado para o nó original e para os nós resultantes da divisão. A árvore busca a divisão que minimize a impureza total, levando a uma melhor separação entre as classes. Esse processo continua até que todos os dados estejam perfeitamente classificados ou até que a árvore atinja um critério de parada, como profundidade máxima ou número mínimo de instâncias por nó.

3.1.3.3 Comparação com Outros Critérios (Entropia)

O índice de Gini é frequentemente comparado com a Entropia, outro critério de divisão comumente usado em árvores de decisão. Ambos medem a pureza dos nós, mas de maneiras ligeiramente diferentes. Porém, ambos os métodos geralmente resultam em divisões semelhantes, o Gini é ligeiramente mais rápido e, por isso, é o padrão em muitas implementações de árvores de decisão, incluindo o *DecisionTreeClassifier* da biblioteca *scikit-learn*.

- Índice de Gini: simples de calcular, mede a impureza de um nó. Menor valor indica maior pureza. Costuma ser mais eficiente em termos computacionais.
- Entropia: mede a incerteza em uma divisão e calcula a quantidade de "surpresa" ao fazer uma predição. A entropia tende a ser mais sensível a mudanças nas distribuições de probabilidade das classes.

O índice Gini fornece uma indicação de quão puros são os nós no final da árvore, sendo que um valor menor indica uma melhor divisão. No contexto das árvores de decisão, o índice Gini é usado como uma função de custo para avaliar as divisões nos dados. O objetivo é minimizar a probabilidade de classificação incorreta. De acordo com (SHU, 2021), o índice de Gini também pode ser aplicado para avaliar se uma divisão é adequada ou não.

O critério Gini é usado como a função para medir a qualidade de uma divisão no *DecisionTreeClassifier* da biblioteca *sklearn*. É um critério comumente usado em algoritmos de árvore de decisão devido à sua eficiência e simplicidade. No entanto, vale ressaltar que diferentes critérios podem levar a diferentes árvores e pode ser benéfico experimentar outros critérios como entropia para ver qual deles funciona melhor para o caso específico. Neste estudo, foi realizada uma comparação entre os critérios Gini e entropia em modelos de árvore de decisão, ambos treinados com o mesmo conjunto de dados do ativo VALE3.SA. Utilizou-se uma divisão de 80% para treinamento e 20% para teste. Além de o índice Gini apresentar um menor custo computacional, ele também demonstrou melhor acurácia em relação ao critério de entropia durante os testes realizados.

3.2 Dados

Os dados constituem um conjunto de informações que podem ser desconexas e aleatórias, representando fatos brutos e discretos que, isoladamente, podem não possuir um significado específico. Podem ser números, textos, imagens, sons ou qualquer outra forma de representação de informação. Pense nos dados como peças de um quebra-cabeça. Cada peça pode não revelar muito por si só, mas quando reunidas de forma organizada formam uma imagem ou mensagem importante.

Neste trabalho, os dados são um conjunto de informações numéricas onde refere-se aos registros históricos de um ativo financeiro de interesse do usuário. Os dados são extraídos através de uma biblioteca chamada *yfinance*. É possível encontrar informações, como: o preço

de abertura, o preço de fechamento, o volume de negociação, o preço mais baixo e o preço mais alto.

Os dados serão usados para treinar o modelo, podendo ser dividido em treino e teste. Ademais, os dados podem ser temporais ou aleatórios. Na divisão temporal preserva-se a ordem cronológica dos dados e na divisão aleatória os dados são embaralhados antes de serem divididos em treino e teste.

3.2.1 Bibliotecas

Para prosseguir com o processo criativo da ferramenta, será necessária a instalação das seguintes bibliotecas: *pandas* (MCKINNEY, 2018), responsável pela manipulação e análise dos dados financeiros. *Sklearn* (PEDREGOSA *et al.*, 2011), é utilizada para a implementação de algoritmos de aprendizado de máquina, como o modelo de árvore de decisão. *Yfinance* (AROUSSE, 2011), é responsável pela coleta de dados financeiros históricos, enquanto *pandas_ta* é utilizada para o cálculo de indicadores técnicos. A biblioteca *ipywidgets* permite a criação de interfaces interativas. *Graphviz* é usada para a visualização das árvores de decisão, e *pydot* é uma biblioteca complementar ao *Graphviz*. Por fim, *plotly* é utilizada para criar gráficos interativos. Essas bibliotecas fornecem funcionalidades específicas necessárias para a análise de dados financeiros e a visualização dos resultados.

3.2.2 Pré-processamento dos dados

Os dados coletados continham alguns valores ausentes, os quais foram removidos para evitar erros durante o treinamento do modelo. Além disso, para garantir que todas as variáveis estivessem na mesma escala, foi aplicada a normalização dos dados utilizando o *MinMaxScaler*, transformando todos os valores para a faixa de 0 a 1. Adicionalmente, indicadores técnicos como RSI (2.1.3.1) e SMA (2.1.3.3) foram calculados e adicionados ao conjunto de dados para enriquecer as previsões.

3.2.3 DataFrame

Com os dados pré-processados cria-se um *DataFrame* a partir dos dados históricos de um ativo específico. Assim, o *DataFrame* é uma estrutura de dados organizada em linhas e colunas. Além das informações tradicionais, o *DataFrame* utilizado também define a direção do

mercado no dia seguinte, assim como o alvo (target) a ser previsto pelo modelo. Os dados são pré-processados e limpos antes de serem utilizados.

3.2.4 Tipos de aprendizado

Existem alguns tipos de aprendizado que são de suma importância para o entendimento, estes conceitos são definidos nos subtópicos abaixo:

3.2.4.1 Aprendizado supervisionado

No aprendizado supervisionado é usado sempre que queremos prever um determinado resultado. O modelo é treinado com um conjunto de dados rotulados, ou seja, os dados de entrada são acompanhados das respostas corretas. O objetivo é que o modelo aprenda a mapear entradas para saídas corretas. Segundo (MÜLLER; SARAH, 2016), o aprendizado supervisionado é um dos tipos de aprendizado mais promissores, sendo classificação e regressão exemplos válidos de um modelo com aprendizado supervisionado.

3.2.4.2 Aprendizado não supervisionado

No aprendizado não supervisionado, o modelo é treinado com dados que não possuem rótulos. O objetivo é encontrar padrões ou estruturas ocultas nos dados. Técnicas comuns incluem *clustering* (agrupamento de dados semelhantes) e redução de dimensionalidade (como a *Principal Component Analysis* ou Análise de Componentes Principais (PCA)). Segundo (MÜLLER; SARAH, 2016), um exemplo disso seria a extração de tópicos em coleções de documentos de texto.

3.2.4.3 Aprendizado semissupervisionado

Segundo (BRAGA, 2010), os algoritmos semissupervisionados também produzem classificadores. O aprendizado semissupervisionado é uma técnica de aprendizado de máquina que combina uma pequena quantidade de dados rotulados com uma grande quantidade de dados não rotulados para treinar modelos.

3.2.4.4 Aprendizado por reforço

O modelo aprende através de interações com o ambiente, recebendo recompensas ou punições com base em suas ações. O objetivo é maximizar a recompensa total ao longo do tempo. O aprendizado por reforço é amplamente utilizado em robótica, jogos e sistemas de recomendação. De acordo com (JUNIOR, 2021), este tipo de aprendizado possui três componentes principais: o agente (aprendiz ou tomador de decisão), o ambiente (tudo com que o agente interage) e ações (o que o agente pode fazer). Um exemplo famoso é o *AlphaGo*, o programa de inteligência artificial que venceu campeões humanos no jogo de *Go*.

3.2.5 Treinamento do modelo

O modelo adotado para este trabalho foi o de AM supervisionado, utilizou-se o algoritmo de árvore de decisão, implementada por meio da classe *DecisionTreeClassifier* da biblioteca *scikit-learn*. Este modelo realiza tarefas de classificação com boa interpretação, além de resolver bem problemas com dados estruturados. Este modelo é baseado em uma estrutura hierárquica que se assemelha a um fluxograma. Cada nó interno da árvore representa uma decisão baseada em uma característica dos dados e cada nó folha representa uma classe predita. Neste caso: movimentos de alta, baixa ou lateralidade nos preços das ações.

Como mencionado na subseção 3.1.3.2, ao instanciar o modelo adotou-se o critério Gini como medida de impureza, que avalia a qualidade das divisões dos nós da árvore.

Figura 7 – Árvore.

```
def criaArvore(df, criterio, divisaoDados, testSize):

    if divisaoDados == 'temporal':
        x_train, x_test, y_train, y_test = train_test_split(df.drop('Classe', axis=1), df['Classe'], test_size=testSize/100, shuffle=False)
    elif divisaoDados == 'aleatoria':
        x_train, x_test, y_train, y_test = train_test_split(df.drop('Classe', axis=1), df['Classe'], test_size=testSize/100)

    arvore = DecisionTreeClassifier(criterion=criterio)
    arvore.fit(x_train, y_train)
    resultado = arvore.predict(x_test)

    print('\t\t\tACURRACY\n')
    print(metrics.classification_report(y_test, resultado))

    return arvore, resultado
```

✓ 0.0s

Fonte: Elaborado pelo autor (2024).

Sabe-se que também é necessário treinar um modelo de árvore de decisão com base nos dados fornecidos. Na Figura 7, o critério de divisão e a divisão dos dados podem ser escolhidos pelo usuário. Observe a Figura 8.

Figura 8 – Parâmetros da árvore.

```
arvore, resultado = criaArvore(df, 'gini', 'temporal', 20)
```

Fonte: Elaborado pelo autor (2024).

Na Figura 8, observa-se que, ao utilizar o critério Gini, os dados são organizados de forma temporal, preservando sua ordem cronológica. A divisão dos dados é feita em dois conjuntos: 80% dos dados são destinados ao treinamento e os 20% restantes ao teste. Após a divisão, o modelo é ajustado com base no conjunto de treinamento.

4 ANÁLISE DOS DADOS

Esta seção, tem o objetivo de examinar como os dados foram avaliados. Nas subseções abaixo constam mais detalhes.

4.1 Avaliação dos dados

Calcula-se os retornos financeiros com base nas decisões tomadas pelo modelo. Os retornos cumulativos são calculados e comparados com um *benchmark*. Isso permite avaliar a eficácia do modelo em relação a uma estratégia de investimento passiva. Na Figura 8, é utilizado um comando que retorna as métricas de acurácia, que são essenciais para avaliar a eficiência da ferramenta. As métricas apresentadas incluem:

- Acurácia: proporção de previsões corretas sobre o total de exemplos.
- Precisão: proporção de exemplos classificados como positivos que realmente são positivos (verdadeiros positivos / (verdadeiros positivos + falsos positivos)).
- *Recall* (Sensibilidade): proporção de exemplos positivos que foram corretamente identificados (verdadeiros positivos / (verdadeiros positivos + falsos negativos)).
- *F1-Score*: média harmônica entre precisão e *recall*, oferecendo uma métrica que leva em consideração tanto os falsos positivos quanto os falsos negativos.

Essas métricas avaliam a capacidade do modelo de classificar corretamente as direções de movimento do mercado (alta, baixa ou lateralização) com base nos dados históricos de preços e indicadores técnicos. Com base nas decisões de compra e venda é calculado o retorno financeiro simulado, através da função *calculaRetorno* na Figura 9, que compara o desempenho da estratégia baseada nas previsões do modelo com um *benchmark*.

- *benchmark*: a linha de base (ou *benchmark*) é o retorno cumulativo de uma estratégia passiva de manter a ação (baseado no retorno percentual diário da ação).
- *CR* (*Cumulative Return* ou *Retorno Cumulativo* (*CR*)): este é o retorno cumulativo da estratégia baseada nas previsões do modelo. Ele é calculado como o retorno cumulativo ponderado pelas decisões do modelo *Decision*. Se o modelo previu uma alta, a estratégia "compra" a ação, e se previu uma queda, a estratégia "vende" ou "shorta" a ação.

Figura 9 – Simulação do retorno.

```
df.loc[:, 'Benchmark'] = (df['Change'] + 1).cumprod()
df.loc[:, 'Rt'] = 1 + df.Decision * df.Change
df.loc[:, 'CR'] = df.Rt.cumprod()
```

Fonte: Elaborado pelo autor (2024).

Na Figura 9, está representada a função *calculaRetorno*, que retorna a diferença entre o retorno cumulativo (*CR*) e o *benchmark*. Este valor mostra a vantagem que o modelo oferece em relação a uma estratégia passiva.

Figura 10 – Simulação do retorno em código (vantagem).

```
vantagem = df.CR.iloc[-1] - df.Benchmark.iloc[-1]
```

Fonte: Elaborado pelo autor (2024).

A Figura 9 e a Figura 10 se complementam, neste trecho do algoritmo, testa-se a métrica onde o modelo mostra uma vantagem em termos de retorno financeiro. A variável *vantagem* é usada para calcular a vantagem da estratégia de investimento baseada nas previsões do modelo em relação a uma estratégia de *benchmark* passiva.

Tabela 3 – Simulação do retorno.

	Date	Close	Decision	Change	Benchmark	Rt	CR
1	2022-03-18 00:00:00-03:00	89.088669	-1.0	0.018982	1.018982	0.981018	0.981018
2	2022-03-21 00:00:00-03:00	91.610481	-1.0	0.028307	1.047826	0.971693	0.953249
3	2022-03-22 00:00:00-03:00	89.561516	-1.0	-0.022366	1.024390	1.022366	0.974569
4	2022-03-23 00:00:00-03:00	89.422440	-1.0	-0.001553	1.022799	1.001553	0.976082
5	2022-03-24 00:00:00-03:00	89.848923	-1.0	0.004769	1.027678	0.995231	0.971427
...
269	2023-04-13 00:00:00-03:00	79.510002	-1.0	-0.013524	0.909423	1.013524	2.217268
270	2023-04-14 00:00:00-03:00	78.809998	-1.0	-0.008804	0.901416	1.008804	2.236789
271	2023-04-17 00:00:00-03:00	77.870003	-1.0	-0.011927	0.890665	1.011927	2.263468
272	2023-04-18 00:00:00-03:00	78.540001	-1.0	0.008604	0.898328	0.991396	2.243993
273	2023-04-19 00:00:00-03:00	76.250000	-1.0	-0.029157	0.872135	1.029157	2.309421

273 rows x 7 columns

Fonte: Elaborado pelo autor (2024).

A Tabela 3 apresenta os resultados de uma simulação de retorno baseada em um modelo preditivo de aprendizado de máquina aplicado ao ativo financeiro. Nesta simulação:

- *Date*: esta coluna representa a data em que o ativo foi negociado, fornecendo o contexto temporal para cada transação ou previsão do modelo.
- *Rt*: mostra o retorno diário do ativo em termos percentuais. Ele é calculado com base na diferença entre o preço de fechamento do dia e o preço de fechamento do dia anterior, indicando o lucro ou perda do dia.
- *Decision*: refere-se à previsão do modelo de aprendizado de máquina para o próximo movimento do mercado (compra, venda ou manter). Onde, 1 indica uma decisão de compra, -1 indica venda e 0 indica que a posição será mantida
- *Change*: refere-se à variação percentual do preço do ativo em relação ao dia anterior, indicando o movimento real do preço do ativo.
- *CR*: mostra o retorno acumulado da estratégia preditiva ao longo do tempo, ou seja, a soma dos retornos diários ao longo de todo o período de análise.
- *Benchmark*: Representa o retorno acumulado de uma estratégia passiva (*benchmark*), onde o ativo é comprado e mantido durante todo o período, sem realizar operações de compra e venda com base em previsões.

4.2 Visualização dos dados

Figura 11 – Visualização gráfica.

```
def Grafico(df):
    fig = make_subplots(rows=2, cols=1, shared_yaxes=True)

    #####
    # Para plotar as linhas. Cada linha precisa de um código fig.add_trace. No exemplo abaixo, temos duas linhas do df.
    fig.add_trace(go.Scatter(x=df.index, y=df.Benchmark, name='Benchmark'),
                  row=1, col=1)
    fig.add_trace(go.Scatter(x=df.index, y=df.CR, name='Cumulative Return - CR'),
                  row=1, col=1)
    #####
    # --- barras para plotar o delta
    fig.add_trace(go.Bar(x=df.index, y=df.Decision, name='Decision'),
                  row=2, col=1)

    fig.update_layout(height=600, width=600,
                       title_text="Análise Financeira", xaxis_title="Tempo",
                       yaxis_title="Retorno(%)")

    fig.update_xaxes(title_text="Tempo", showgrid=True, gridwidth=1, gridcolor='lightgray', showline=True, linewidth=1, row=2, col=1)
    fig.update_yaxes(title_text="Decision", showgrid=True, gridwidth=1, gridcolor='lightgray', showline=True, linewidth=1, row=2, col=1)

    return fig
✓ 0.0s
```

Fonte: Elaborado pelo autor (2024).

Para uma melhor visualização dos resultados financeiros é necessário gerar um gráfico interativo. O gráfico mostra o retorno cumulativo do modelo e do *benchmark* ao longo do tempo bem como as decisões tomadas pelo modelo. Essa visualização auxilia na análise e interpretação dos resultados obtidos. O código da Figura 11 irá ajudar neste passo de visualização.

5 RESULTADOS

Na seção 4.1 encontra-se detalhes dos indicadores de avaliação dos dados como *precision*, *recall*, *F1-score*. Ao analisar o ativo de código de negociação VALE3.SA a partir do ano de 2017 até 2023, obtém-se uma acurácia satisfatória de 78%. Foi utilizado o critério Gini (3.1.3), a divisão de dados ocorreu de modo temporal, divididos em dois conjuntos, sendo eles 80% para treino e 20% para teste. Ocorreram outras tentativas de testes, dentre elas 70% de treino e 30% de teste. Entretanto, não apresentou uma acurácia satisfatória, dentre todos os testes o que apresentou uma melhor acurácia, foi 80% de treino e 20% de teste, que foi o modelo utilizado neste trabalho. Na Figura 12, observa-se com mais detalhes a acurácia.

Figura 12 – Acurácia da VALE3.SA.

```
[ ] arvore, resultado = criaArvore(tabela, 'gini', 'temporal', 20)
```

ACURRACY				
	precision	recall	f1-score	support
-1.0	0.88	0.65	0.75	152
1.0	0.72	0.91	0.80	148
accuracy			0.78	300
macro avg	0.80	0.78	0.78	300
weighted avg	0.80	0.78	0.78	300

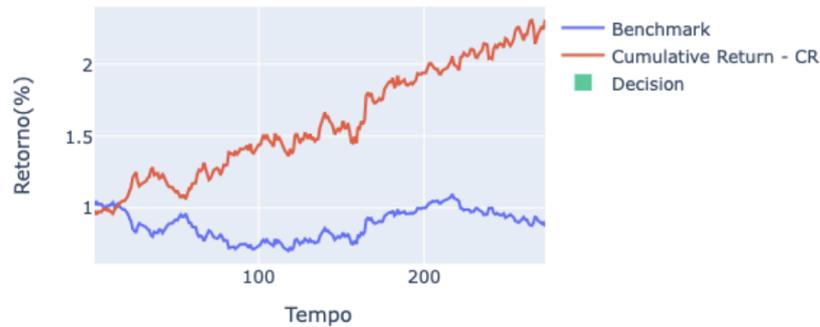
Fonte: Elaborado pelo autor (2023).

Observa-se a coluna *support* definida pela função *classification_report()* da biblioteca *sklearn.metrics*. O *support* informa as quantidades de amostras das classes presentes no conjunto de dados de teste. Ele ajuda a contextualizar o desempenho do modelo para cada classe, o *support* refere-se ao número de instâncias verdadeiras. O modelo apresenta dados de *precision*, *recall*, *F1-score* e *support* satisfatórios, capaz de identificar tanto movimentos de alta, como movimentos de baixa.

Após a análise, o ativo com *ticker* VALE3.SA, é representado no gráfico da Figura 13, demonstrando um comparativo entre o retorno acumulado gerado por uma estratégia preditiva baseada em aprendizado de máquina e o *benchmark* de uma estratégia passiva. A análise é dividida em dois subgráficos que evidenciam o comportamento da estratégia ao longo do tempo.

Figura 13 – Retorno Acumulado e *Benchmark*.

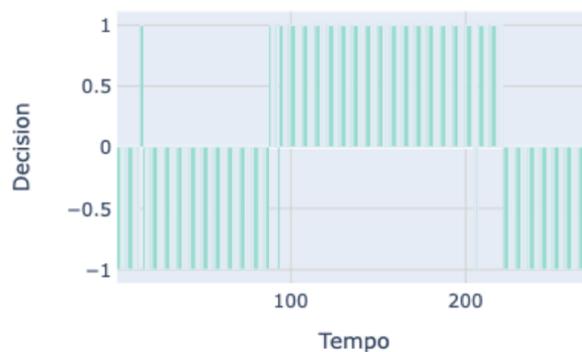
Análise Financeira



Fonte: Elaborado pelo autor (2024).

A linha vermelha do gráfico de retorno por tempo, representa o retorno acumulado da estratégia preditiva, que foi gerada a partir das decisões de compra e venda previstas pelo modelo de árvore de decisão. Ao longo do tempo, a linha vermelha indica um crescimento sustentado do retorno da estratégia, o que sugere a eficácia do modelo preditivo em identificar padrões no comportamento dos preços das ações. Observa-se uma inclinação ascendente consistente, sinalizando um aumento contínuo do retorno percentual ao longo do período analisado. A linha Azul (*benchmark*) corresponde ao retorno acumulado de uma estratégia passiva apenas com a manutenção do ativo ao longo do tempo, sem alterações baseadas em previsões do mercado. O modelo funciona para todos os ativos, porém nos testes alguns ativos tiveram resultados ruins, a escolha da árvore de decisão foi pela sua simplicidade de implementação e apresentou resultados satisfatórios.

Figura 14 – Gráfico da evolução VALE3.SA.



Fonte: Elaborado pelo autor (2024).

Nota: O tempo está descrito em dias.

Na Figura 14, tem-se um gráfico de barras que exhibe as decisões de compra e venda tomadas pela estratégia preditiva ao longo do tempo. O valor 1 indica uma decisão de compra,

sugerindo que o modelo previu uma alta no preço do ativo, enquanto o valor -1 indica uma decisão de venda, sugerindo uma previsão de queda. O eixo vertical representa o tipo de decisão (-1 para venda, 1 para compra), e o eixo horizontal indica o tempo. A Figura 14 mostra como as decisões foram distribuídas ao longo do período de análise, complementando o gráfico superior ao demonstrar a relação entre essas decisões e o desempenho da estratégia.

6 DISCUSSÕES

Os resultados obtidos ao longo do desenvolvimento deste trabalho indicam que o uso de árvores de decisão como modelo preditivo para o mercado de ações é uma abordagem viável e promissora. A estratégia preditiva baseada em aprendizado de máquina demonstrou, por meio de simulações, a capacidade de superar o benchmark de uma estratégia passiva de "*buy and hold*", gerando retornos superiores ao longo do tempo. Essa constatação reforça a aplicabilidade de técnicas de aprendizado supervisionado, especialmente árvores de decisão, em ambientes complexos como o mercado financeiro.

A acurácia do modelo, alcançando 78% nos testes realizados com o ativo VALE3.SA, mostra-se satisfatória quando comparada a modelos passivos. No entanto, é importante considerar que a acurácia por si só não é suficiente para garantir uma vantagem contínua no mercado. A análise dos retornos acumulados indicou que o modelo obteve melhor desempenho em momentos de alta volatilidade, sugerindo que a combinação de algoritmos de aprendizado de máquina com indicadores técnicos pode oferecer uma vantagem competitiva em períodos de instabilidade.

Apesar dos resultados positivos, algumas limitações foram observadas. O modelo de árvore de decisão, embora simples e fácil de interpretar, pode apresentar dificuldades ao lidar com dados altamente dinâmicos e não-lineares, como os do mercado de ações. Isso sugere que modelos mais complexos, como redes neurais ou algoritmos baseados em *ensemble* (por exemplo, *Random Forests*), poderiam ser explorados em trabalhos futuros para melhorar a robustez e precisão das previsões. Além disso, a divisão dos dados em 80% para treinamento e 20% para teste pode não capturar adequadamente as flutuações do mercado, já que o volume de dados históricos e as características específicas de cada ativo podem influenciar a capacidade preditiva do modelo.

Outro ponto relevante discutido é a dependência do modelo em relação à qualidade dos dados históricos utilizados. Como o desempenho da árvore de decisão depende diretamente das variáveis de entrada e dos indicadores técnicos selecionados, variações nos dados, como erros ou omissões, podem impactar negativamente os resultados. Ademais, fatores exógenos, como eventos macroeconômicos, crises políticas e novas regulamentações, não foram considerados no escopo do modelo, o que poderia comprometer sua precisão em determinados cenários.

Em termos práticos, a implementação de uma ferramenta preditiva como a desenvolvida neste trabalho tem o potencial de beneficiar investidores de todos os níveis de experiência, especialmente aqueles com menos familiaridade com o mercado de ações. Ao automatizar o

processo de análise e recomendação de compra e venda, o modelo oferece suporte para tomadas de decisões mais racionais, mitigando o impacto de vieses emocionais que frequentemente afetam investidores inexperientes. No entanto, é essencial que os investidores entendam que nenhum modelo preditivo pode garantir o sucesso no mercado financeiro, sendo sempre necessário um gerenciamento de risco adequado.

Por fim, os resultados deste estudo evidenciam que o aprendizado de máquina, quando aplicado ao mercado de ações, pode proporcionar uma ferramenta eficaz para identificar padrões e tendências. No entanto, a natureza volátil e imprevisível do mercado sugere que futuras pesquisas poderiam explorar a combinação de diferentes abordagens preditivas, como a inclusão de dados de análise fundamentalista, para refinar ainda mais a acurácia e a robustez do modelo.

7 CONCLUSÕES E TRABALHOS FUTUROS

Este trabalho teve como objetivo desenvolver e avaliar um modelo preditivo baseado em aprendizado de máquina, especificamente utilizando árvores de decisão para auxiliar investidores nas tomadas de decisões de compra e venda de ações. O modelo foi aplicado a dados históricos do ativo VALE3.SA, utilizando indicadores técnicos amplamente reconhecidos, como RSI, ROC, e médias móveis para otimizar as previsões e melhorar a acurácia das decisões.

Ademais, os resultados demonstraram que a metodologia adotada foi eficaz em gerar retornos superiores à estratégia passiva de *"buy-and-hold"*. Como evidenciado pela linha de retorno acumulado consistentemente acima do *benchmark* no período analisado. A combinação dos indicadores técnicos com o modelo de árvore de decisão possibilitou identificar padrões relevantes no comportamento dos preços das ações, resultando em decisões preditivas que proporcionaram uma vantagem financeira em comparação à manutenção passiva do ativo.

Adicionalmente, a divisão temporal dos dados mostrou-se uma abordagem adequada para simular um cenário real de mercado, onde o modelo é treinado com dados históricos e avaliado em dados futuros, refletindo de forma mais precisa o comportamento do ativo em situações de incerteza e volatilidade.

Para melhorias futuras, poderia implementar modelos mais avançados de AM. Como por exemplo, redes neurais recorrentes que são projetados para lidar com dados sequenciais, aumentando a precisão das previsões em séries temporais. Também, técnicas de *ensemble learning* podem ser implementadas para aumentar a robustez das previsões.

Outra melhoria interessante é a implementação de relatórios diários que são enviados por e-mail em horário predefinido. A geração de relatórios diários pode ser automatizada por meio de *scripts* que consolidam informações relevantes, como: previsões de mercado, retornos diários e análise de desempenho. Estes relatórios podem ser gerados em formatos de PDF ou de Excel e enviados automaticamente por e-mail em dias ou horários predefinidos, utilizando a biblioteca *smtplib* em *Python* ou integrando plataformas de notificação. Este processo garante que os usuários recebam atualizações periódicas e consistentes. Facilitando o acompanhamento contínuo da estratégia preditiva, permitindo que as tomadas de decisões sejam embasadas e estratégicas.

REFERÊNCIAS

AMARANTE, A. N. A. D. A. S. E. E. M. D. **TÉCNICOS DA TEORIA A PRÁTICA**. [S.l.]: UM FÍSICO NA BOLSA, 2021.

ANBIMA. **Investimentos dos brasileiros somam R\$ 5 trilhões em 2022**. 2023. Disponível em: <https://www.anbima.com.br/pt_br/noticias/investimentos-dos-brasileiros-somam-r-5-trilhoes-em-2022.htm#:~:text=O%20volume%20financeiro%20investido%20por.> Acesso em: 25 maio. 2023.

AROUBI, R. **Documentação oficial de yfinance**. 2011. Disponível em: <<https://ypi.org/project/yfinance/>>. Acesso em: 23 out. 2024.

B3. **Perfil pessoas físicas**. 2023. Disponível em: <https://www.b3.com.br/pt_br/market-data-e-indices/servicos-de-dados/market-data/consultas/mercado-a-vista/perfil-pessoas-fisicas/faixa-etaria/> Acesso em: 01 julho. 2023.

BRAGA, A. **Aprendizado semissupervisionado multidescrição em classificação de textos**. Dissertação (Dissertação de Mestrado) — Universidade de São Paulo (USP), São Carlos, 2010.

CETIC. **Uso e Posse de Computador e Internet, Barreiras de Acesso, Uso do Celular, Intenção de aquisição**. 2008. Disponível em: <<https://cetic.br/media/analises/destaques-tic-2007.pdf>> Acesso em: 25 maio. 2023.

COELHO, F. F. **MACHINE LEARNING E ANÁLISE TÉCNICA COMO FERRAMENTAS PARA CONSTRUÇÃO DE PORTFÓLIOS DE RENDA VARIÁVEL NO MERCADO BRASILEIRO**. Dissertação (Mestrado) — FUNDAÇÃO GETÚLIO VARGAS (FGV), Rio de Janeiro, 2020.

COPPIN, B. **TÉCNICOS DA TEORIA A PRÁTICA**. [S.l.]: JONES AND BARTLETT PUBLISHERS, 2004. v. 1st ED.

GOOGLE. **Colaboratory**. 2023. Disponível em: <<https://research.google.com/colaboratory/intl/pt-BR/faq.html>> Acesso em: 28 maio. 2023.

GRAHAM, B. **O INVESTIDOR INTELIGENTE**. [S.l.]: HARPER COLLINS, 2020.

GÉRON, A. **Mãos à Obra: Aprendizado de Máquina com Scikit-Learn, Keras TensorFlow**. [S.l.]: Alta Books, 2021. v. 2ª Edição.

HULL, A. **Hull Moving Average (HMA)**. 1990. Disponível em: <<https://ajuda.nelogica.com.br/hc/pt-br/articles/4413356467227-Hull-Moving-Average-HMA-1>>. Acesso em: 25 maio. 2023.

JUNIOR, P. I. F. D. S. **Estrutura a termo da taxa de juros no Brasil: Projeções utilizando aprendizado de máquina**. Dissertação (Dissertação de Mestrado) — Universidade de Brasília(UNB), Brasília, 2021.

LEE, K.-F. **Inteligência Artificial**. [S.l.]: Globo Livros, 2019. volume 1.

LEMO, F. **Análise Técnica dos Mercados Financeiros Um guia completo e definitivo dos métodos de negociação de ativos**. São paulo, SP: Saraiva Uni, 2018. p.23 – 25 p.

LOPES, M. F. **A Introdução de Indicadores Técnicos em Políticas Paramétricas de Portefólio**. Dissertação (Mestrado) — UNIVERSIDADE D COIMBRA, Coimbra - Portugal, 2024.

MACHADO, A. A. de A.; CORRÊA, G. N. Avaliação de modelos de previsão dos valores das ações no mercado financeiro usando aprendizado de máquina. **Retec**, Ourinhos, São Paulo, 2022.

MCKINNEY, W. **Pandas Documentation**. 2018. Disponível em: <https://pandas.pydata.org/docs/user_guide/index.html#user-guide>. Acesso em: 23 out. 2024.

MCKINNEY, W. **PYTHON PARA ANÁLISE DE DADOS: Tratamento de dados com pandas, numpy e ipython**. São Paulo, SP: O' Reilly Novatec, 2018. 42 p.

MELO, R. M. **Previsão da variação de preços aplicada ao mercado de opções binárias com o uso de redes neurais LSTM**. João Monlevade–MG: Universidade Federal de Ouro Preto (UFOP), 2022.

MENEZES, A.; EVANGELISTA, I.; BARBOSA, R.; ESKENAZI, Y.; MENEZES, M. O. de. Inteligência artificial na bolsa de valores. **Universidade Presbiteriana Mackenzie (UPM)**, São Paulo, SP.

MENEZES, N. N. C. **INTRODUÇÃO À PROGRAMAÇÃO COM PYTHON**. SÃO PAULO: NOVATEC, 2014.

MONEY, I. **Bolsa se aproxima de 1 milhão de investidores, mas ainda é pouco**. 2019. Disponível em: <<https://www.infomoney.com.br/onde-investir/bolsa-se-aproxima-de-1-milhao-de-investidores-mas-ainda-e-pouco/>>. Acesso em: 25 maio. 2023.

MONEY, I. **O que é a B3 e como investir na bolsa brasileira**. 2023. Disponível em: <<https://www.infomoney.com.br/guias/o-que-e-b3/#:~:text=Apesar%20de%20nÃo%20estar%20entre,investidores%20na%20bolsa%20aumentou%2040%25.>> Acesso em: 25 maio. 2023.

MÜLLER, A. C.; SARAH, S. **Introduction To Machine Learning With Python A Guide For Data Scientists**. [S.l.]: O' Reilly Media, 2016. volume 3.

NETO, C.; SOUZA, D. de; FERREIRA; SUMAQUEIRO, A.; MASCARENHAS; CESAR, C.; MAY; MENDES, A.; VASSÃO; CAVALHEIRO, T.; RÁO; MARTINS, E. A bolsa de valores e os novos investidores nos tempos atuais. **Revista Gestão em Foco**, 2022.

PASSOS, A. **Investindo na bolsa de valores. O que você precisa saber antes de investir**. [S.l.]: Novatec, 2018. Primeira edição.

PEDREGOSA, F.; VAROQUAUX; G.; GRAMFORT; A.; MICHEL; V.; THIRION; B.; GRISEL; O.; BLONDEL; M. **Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830**. 2011. Disponível em: <<https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>>. Acesso em: 23 out. 2024.

PYTHON. **General Python FAQ**. 2023. Disponível em: <<https://docs.python.org/3/faq/general.html#general-information>> Acesso em: 28 maio. 2023.

SHANNON, C. E. **A Mathematical Theory of Communication**. [S.l.]: Bell System Technical Journal, 1948. v. 27. 379–423, 623–656 p.

SHU, F. Dissertação de mestrado, **APLICANDO APRENDIZADO DE MÁQUINA À SUPERVISÃO DO MERCADO DE CAPITAIS Classificação e extração de informações de documentos financeiros**. RIO DE JANEIRO: [s.n.], 2021.

SIQUEIRA, G. D. S. Dissertação de graduação, **ANÁLISE DE APLICAÇÃO DE ALGORITMOS DE APRENDIZADO DE MÁQUINA E ANÁLISE TÉCNICA NA PREDIÇÃO DE TENDÊNCIA DE AÇÕES DA BOLSA DE VALORES**. NITERÓI, RJ: HARPER COLLINS, 2020.

SUNO. **Pregão viva voz: conheça como era o funcionamento da bolsa no passado**. 2019. Disponível em: <<https://www.suno.com.br/artigos/pregao-viva-voz/>> Acesso em: 25 maio. 2023.

USP, J. da. **Brasileiro não sabe investir nem se organizar financeiramente**. 2019. Disponível em: <<https://jornal.usp.br/atualidades/brasileiro-nao-sabe-investir-nem-se-organizar-financeiramente/>> Acesso em: 25 maio. 2023.

VOORSLUYS, W. **A Introdução de Indicadores Técnicos em Políticas Paramétricas de Portefólio**. Dissertação (Mestrado) — (USP), São Carlos, 2006.