

UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"

FACULDADE DE CIÊNCIAS - CAMPUS BAURU

DEPARTAMENTO DE COMPUTAÇÃO

BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

JOÃO HENRIQUE Mouro SUIDEN

**DESENVOLVIMENTO DE ALGORITMOS DE APRENDIZADO DE
MÁQUINA PARA ANÁLISE E PREVISÃO DE ATIVOS
FINANCEIROS**

BAURU

Novembro/2024

JOÃO HENRIQUE MOURO SUAIDEN

**DESENVOLVIMENTO DE ALGORITMOS DE APRENDIZADO DE
MÁQUINA PARA ANÁLISE E PREVISÃO DE ATIVOS
FINANCEIROS**

Trabalho de Conclusão de Curso do Curso
de Ciência da Computação da Universidade
Estadual Paulista “Júlio de Mesquita Filho”,
Faculdade de Ciências, Campus Bauru.
Orientador: Prof. Dr. Douglas Rodrigues

BAURU
Novembro/2024

S939d

Suaíden, João Henrique Mouro

Desenvolvimento de Algoritmos de Aprendizado de Máquina para
Análise e Previsão de Ativos Financeiros / João Henrique Mouro

Suaíden. -- Bauru, 2024

40 p. : il., tabs., fotos

Trabalho de conclusão de curso (Bacharelado - Ciência da
Computação) - Universidade Estadual Paulista (UNESP), Faculdade
de Ciências, Bauru

Orientador: Douglas Rodrigues

1. Ciência da Computação. 2. Aprendizado de Máquina. 3.
Inteligência Artificial. 4. XGBoost. 5. Mercado Financeiro. I. Título.

João Henrique Mouro Suaiden

DESENVOLVIMENTO DE ALGORITMOS DE APRENDIZADO DE MÁQUINA PARA ANÁLISE E PREVISÃO DE ATIVOS FINANCEIROS

Trabalho de Conclusão de Curso do Curso de Ciência da Computação da Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Ciências, Campus Bauru.

Banca Examinadora

Prof. Dr. Douglas Rodrigues

Orientador

Universidade Estadual Paulista "Júlio de
Mesquita Filho"

Faculdade de Ciências

Departamento de Ciência da Computação

**Profa. Dra. Simone das Graças
Domingues Prado**

Universidade Estadual Paulista "Júlio de
Mesquita Filho"

Faculdade de Ciências

Departamento de Ciência da Computação

**Prof. Dr. Leandro Aparecido Passos
Junior**

Universidade Estadual Paulista "Júlio de
Mesquita Filho"

Faculdade de Ciências

Departamento de Ciência da Computação

Bauru, 12 de Novembro de 2024.

*Dedico este trabalho aos meus pais, motivo da minha mais profunda admiração e gratidão,
sem os quais nada disso teria sido possível.*

Agradecimentos

Quero agradecer primeiramente à minha família, em especial aos meus pais, Mário Henrique Suaiden e Karina Rodrigues Lima Mouro, a quem devo tudo. Obrigado por todo o esforço, trabalho e motivação que me fizeram chegar até aqui. Obrigado por toda conversa, ajuda e amor que me fizeram se manter aqui. Obrigado por todo conselho e ensinamento que me farão chegar onde um dia sei que chegarei. Obrigado às minhas tias, Natalina, Bel, Malu, Milena, Kelli e Kátia. Obrigado à minha vó Diva que sempre esteve ao meu lado e aos meus eternos avôs e avó Jorge, Façal e Maria, os quais que sempre sentirei saudades.

Quero agradecer à minha república Computos, meu eterno lar em Bauru, que me proporcionou momentos e memórias as quais levarei para o resto da vida. Aos meus amigos de república, Lucas Yuki, Gustavo Tandaya, Luca Munekata, João Zanqui, Lucas Suaiden, Vinícius Garcia, Marcos Ferrari, Gabriel Parizotto e Rafael Costa, obrigado por estarem presentes nessa jornada.

Meus agradecimentos a Associação Atlética Acadêmica Unesp Bauru, lugar que me ensinou sobre responsabilidade, trabalho em equipe, disciplina, e tantas outras coisas. Que me fez valorizar o esporte universitário e ter orgulho de ser unespiano. Instituição da qual tenho honra de ter feito parte da história e sei que ainda tem muito a conquistar.

Agradeço a todos que participaram dessa jornada de alguma maneira e contribuíram positivamente. Um agradecimento especial ao meu orientador Douglas Rodrigues, com quem pude contar e me auxiliou na conclusão deste trabalho.

Se você pode sonhar, você pode realizar.

Walt Disney

Resumo

Este trabalho tem como objetivo explorar o uso de algoritmos de aprendizado de máquina na análise e previsão de preços de ativos financeiros, contribuindo para estratégias de tomada de decisão mais precisas no mercado financeiro. A metodologia aplicada inclui a utilização do algoritmo XGBoost, além de indicadores financeiros técnicos como o RSI (*Relative Strength Index*) e MACD (*Moving Average Convergence Divergence*) que auxiliam na identificação de tendências e padrões de preços. Os resultados obtidos indicam que o XGBoost, em combinação com os indicadores financeiros, apresenta um desempenho promissor, com capacidade de prever movimentos de mercado em cenários de alta volatilidade e em diferentes intervalos de tempo. A avaliação de desempenho foi realizada com base em métricas como a Acurácia e o F1-Score, permitindo uma análise robusta da eficácia do modelo. Também foram feitas simulações de investimento, utilizando o modelo proposto, a fim de se observar o seu comportamento no mercado real.

Palavras-chave: XGBoost, aprendizado de máquina, indicadores financeiros, mercado financeiro.

Abstract

This work aims to explore the use of machine learning algorithms in the analysis and prediction of financial asset prices, contributing to more accurate decision-making strategies in the financial market. The applied methodology includes the use of the XGBoost algorithm, as well as technical financial indicators such as the RSI (Relative Strength Index) and MACD (Moving Average Convergence Divergence), which assist in identifying trends and price patterns. The results indicate that XGBoost, in combination with financial indicators, shows promising performance, with the ability to predict market movements in high-volatility scenarios and across different time intervals. The performance evaluation was based on metrics such as Accuracy and F1 Score, allowing for a robust analysis of the model's effectiveness. Investment simulations were also conducted using the proposed model to observe its behavior in real market conditions.

Keywords: XGBoost, machine learning, financial indicators, financial market.

Lista de figuras

Figura 1 – Tipos de AM	18
Figura 2 – Funcionamento de um classificador.	19
Figura 3 – Funcionamento do indicador RSI.	22
Figura 4 – Evolução do investimento (ativo x portfólio) - RSI	34
Figura 5 – Retorno percentual acumulado (ativo x portfólio) - RSI	34
Figura 6 – Drawdown (ativo x portfólio) - RSI	34
Figura 7 – Evolução do investimento (ativo x portfólio) - MACD	35
Figura 8 – Retorno percentual acumulado (ativo x portfólio) (ativo x portfólio) - MACD	35
Figura 9 – Drawdown (ativo x portfólio) - MACD	36
Figura 10 – Evolução do investimento (ativo x portfólio) - Modelo completo	37
Figura 11 – Retorno percentual acumulado (ativo x portfólio) - Modelo completo	38
Figura 12 – Drawdown (ativo x portfólio) - Modelo completo	38

Lista de quadros

Quadro 1 – Exemplo Acurácia 75%	30
Quadro 2 – Métricas do modelo com RSI	33
Quadro 3 – Métricas do modelo com MACD	35
Quadro 4 – Métricas do modelo completo	37

Lista de abreviaturas e siglas

AM	Aprendizado de Máquina
SVM	<i>Support Vector Machines</i>
IA	Inteligência Artificial
RSI	<i>Relative Strength Index</i>
MME	Média Móvel Exponencial
MMS	Média Móvel Simples
MACD	<i>Moving Average Convergence Divergence</i>
CCI	<i>Commodity Channel Index</i>
ATR	<i>Average True Range</i>
SAR	<i>Stop and Reverse</i>
OBV	<i>On-Balance Volume</i>
PPO	<i>Percentage Price Oscillator</i>
ADX	<i>Average Direcional Index</i>

Sumário

1	INTRODUÇÃO	14
1.1	Problemática	15
1.2	Justificativa	15
1.3	Objetivos	16
1.3.1	Objetivo Geral	16
1.3.2	Objetivos Específicos	16
2	FUNDAMENTAÇÃO TEÓRICA	17
2.1	Aprendizado de Máquina	17
2.1.1	Aprendizado Supervisionado	18
2.1.2	Aprendizado Não Supervisionado	19
2.1.3	Aprendizado por Reforço	19
2.2	XGBoost (<i>eXtreme Gradient Boosting</i>)	20
2.3	Indicadores Financeiros	21
2.3.1	RSI (<i>Relative Strength Index</i>)	22
2.3.2	MME (Média Móvel Exponencial)	23
2.3.3	MACD (<i>Moving Average Convergence Divergence</i>)	23
2.3.4	Oscilador Estocástico	24
2.3.5	Cálculo das Bandas de Bollinger	24
2.3.6	Índice de Canal de Commodities (CCI)	25
2.3.7	Cálculo do <i>Average True Range</i> (ATR)	25
2.3.8	Cálculo do Parabolic SAR (<i>Stop and Reverse</i>)	26
2.3.9	Cálculo do <i>On-Balance Volume</i> (OBV)	26
2.3.10	Cálculo do <i>Percentage Price Oscillator</i> (PPO)	26
2.3.11	Cálculo do <i>Average Directional Index</i> (ADX)	27
2.3.12	Cálculo do Williams %R	27
3	METODOLOGIA	28
3.1	Bibliotecas e Ferramentas	28
3.1.1	YahooQuery	28
3.1.2	Pandas	29
3.1.3	Matplotlib	29
3.1.4	Jupyter Notebook	29
3.2	Métricas Utilizadas	29
3.2.1	Acurácia	29
3.2.2	F1 Score	30

3.2.3	Rendimento Final	30
3.2.4	Drawdown	30
3.3	Implementação do Modelo	31
3.3.1	Conjunto de Dados	31
3.3.2	Pré-processamento dos Dados	31
3.3.3	Divisão dos Dados	31
3.3.4	Treinamento do Modelo	32
4	RESULTADOS	33
4.1	Modelo com RSI	33
4.2	Modelo com MACD	35
4.3	Modelo completo	36
5	CONCLUSÃO	39
5.1	Trabalhos futuros	39
	REFERÊNCIAS	40

1 Introdução

Nas últimas décadas, a expansão do mercado financeiro foi acompanhada por uma demanda crescente por ferramentas de previsão mais precisas e acessíveis. Entre os diversos fatores que explicam essa expansão, o avanço da tecnologia tem desempenhado um papel central. Plataformas digitais e aplicativos de investimento possibilitaram que pessoas comuns, sem experiência prévia, realizassem investimentos de forma prática e segura, democratizando o acesso ao mercado de capitais. Além disso, a disseminação rápida de informações financeiras por meio da internet e das redes sociais tem promovido uma maior educação financeira entre a população, aspecto essencial para a inclusão de novos investidores no mercado (AMORIM et al., 2018). Esse crescimento do setor pode ainda ser percebido tendo em vista o número de investidores pessoa física no Brasil, que cresceu cerca de 3.500% entre 2002 e 2020 (NETO et al., 2022), um reflexo da popularização dessas plataformas e da maior conscientização financeira.

Entretanto, apesar desse acesso ampliado, a tomada de decisão no mercado financeiro continua sendo uma tarefa complexa, especialmente para investidores que precisam lidar com grandes volumes de dados e incertezas econômicas globais. A volatilidade dos mercados, somada à quantidade massiva de dados não estruturados, exige que investidores adotem ferramentas e técnicas sofisticadas para realizar análises precisas. Métodos tradicionais, como a análise técnica e a análise fundamentalista, permanecem amplamente utilizados, no entanto, eles apresentam limitações importantes, especialmente em cenários marcados por alta volatilidade e incertezas (FAMA, 1970).

Nesse contexto, algoritmos de AM têm se mostrado promissores por sua capacidade de identificar padrões complexos e realizar previsões mais precisas com base em grandes volumes de dados históricos. Diferentemente das abordagens tradicionais, que dependem de modelos predefinidos, os algoritmos de AM são capazes de aprender com os dados e adaptar suas previsões conforme novas informações se tornam disponíveis. Técnicas como *Random Forest*, *XGBoost* e *Support Vector Machines* (SVM) têm ganhado popularidade no setor financeiro devido à sua eficiência e adaptabilidade. Ademais, o uso de AM é especialmente relevante para mercados voláteis, onde os modelos tradicionais falham em capturar a complexidade das flutuações. O XGBoost será o algoritmo utilizado neste trabalho devido a sua eficiência computacional, capacidade de lidar com grande volume de dados, escalabilidade, robustez e versatilidade.

Com a construção e validação desses modelos, espera-se fornecer uma ferramenta que auxilie investidores na tomada de decisão e aumente a precisão das previsões. Este estudo não busca apenas contribuir para o avanço do conhecimento acadêmico, mas também oferecer

informações valiosas que podem ser aplicadas por investidores em um mercado cada vez mais orientado por dados e tecnologia (GOODFELLOW; BENGIO; COURVILLE, 2016). Assim, a aplicação de AM na análise financeira se apresenta como uma alternativa promissora para enfrentar os desafios impostos pela complexidade e volatilidade dos mercados atuais.

1.1 Problemática

Com o avanço do mercado financeiro e a democratização do acesso a investimentos, cada vez mais pessoas têm ingressado nesse setor. No entanto, junto com essas oportunidades, surgem dificuldades consideráveis para investidores, especialmente os iniciantes (JUNIOR; REGONHA; CIPOLLA, 2021). Entre essas dificuldades, destaca-se a tomada de decisão em um ambiente de alta volatilidade e incerteza. A identificação de padrões e tendências no mercado exige uma combinação de técnicas e indicadores financeiros, muitos dos quais possuem limitações em ambientes dinâmicos.

Atualmente existe uma infinidade de indicadores a disposição de quem deseja fazer análises técnicas (GONÇALVES, 2018). Porém, o uso isolado dessas ferramentas pode resultar em previsões imprecisas, sobretudo quando o mercado é impactado por eventos inesperados, como mudanças políticas globais ou crises econômicas. Por este motivo, para uma análise mais precisa, é imprescindível que diversos indicadores sejam usados em harmonia, o que acaba gerando uma tarefa extremamente difícil.

A crescente complexidade do mercado e a quantidade massiva de dados disponíveis impõem desafios adicionais para investidores. O excesso de informações pode levar a uma sobrecarga cognitiva, dificultando a análise precisa e a identificação de padrões relevantes. Nesse contexto, técnicas de AM se destacam como uma solução potencial, pois são capazes de lidar com grandes volumes de dados, identificar padrões complexos e se adaptar rapidamente às mudanças de mercado.

1.2 Justificativa

Como explícito na seção anterior 1.1, os indicadores tradicionais, apesar de amplamente utilizados, mostram-se insuficientes e muitas vezes escassos. Essa limitação justifica a necessidade de explorar novos métodos para aprimorar a análise e a previsão de ativos financeiros. A IA e o AM emergem como alternativas robustas para lidar com a complexidade dos mercados, processando grandes volumes de dados e identificando padrões ocultos (JAIN; VANZARA, 2023).

O uso de IA não apenas melhora a precisão das previsões, mas também democratiza o acesso ao mercado financeiro, permitindo que investidores com diferentes níveis de conhecimento adotem estratégias baseadas em dados (GOODFELLOW; BENGIO; COURVILLE, 2016).

Ferramentas baseadas em IA, como por exemplo, algoritmos de aprendizado profundo, têm mostrado grande potencial para transformar a tomada de decisão no setor financeiro (VINOTHA et al., 2024). Além disso, essas técnicas oferecem aos investidores maior segurança na mitigação de riscos e na gestão de portfólios (CHOPRA; SHARMA, 2021).

Portanto, este trabalho é justificado pela necessidade de melhorar as ferramentas existentes, fornecendo uma solução prática e eficaz que auxilie investidores na tomada de decisões informadas e promova um mercado mais inclusivo e eficiente.

1.3 Objetivos

1.3.1 Objetivo Geral

Desenvolver algoritmos de AM de classificação capazes de analisar e prever a direção dos preços de ativos financeiros, proporcionando maior precisão na tomada de decisão para investidores, com base em dados históricos e indicadores financeiros.

1.3.2 Objetivos Específicos

- Coletar e organizar dados históricos financeiros de ativos selecionados, utilizando fontes como Yahoo Finance e indicadores técnicos reconhecidos.
- Implementar algoritmo de AM, mais especificamente o XGBoost, avaliando sua eficácia na previsão de preços.
- Validar os modelos por meio de simulações de investimento, aplicando diferentes cenários e mensurando o retorno e a mitigação de riscos.
- Avaliar as previsões utilizando métricas adequadas, como o acurácia e F1-Score, para garantir a confiabilidade dos resultados.

2 Fundamentação Teórica

Neste capítulo, serão apresentados os conceitos fundamentais relacionados ao AM, com ênfase no algoritmo XGBoost. Também será explicado o funcionamento de indicadores financeiros essenciais, que desempenham um papel central na análise técnica e serão aplicados nas estratégias ao longo deste trabalho.

2.1 Aprendizado de Máquina

O AM é uma subárea da IA que tem como objetivo criar algoritmos capazes de aprender a partir de dados e melhorar seu desempenho ao longo do tempo, sem a necessidade de serem explicitamente programados para cada tarefa específica (LUDERMIR, 2021). A construção de modelos de AM envolve três fases principais: pré-processamento dos dados, treinamento do modelo e avaliação com novos dados.

Como exemplo de funcionamento de AM, pode-se citar um algoritmo que tenta prever a raça de um grupo de cachorros, a partir de dados fornecidos sobre eles. O algoritmo tentará aprender a distinguir as raças, tendo em vista as informações disponíveis sobre cada cão, assim como o ser humano faz. Apesar de existirem inúmeros tipos de problemas que o AM pode ajudar a solucionar, é importante destacar que nenhum algoritmo é eficiente em todos os tipos de atividades, reforçando a necessidade de se aplicar mais de um algoritmo, como prova (KOHAVI; SOMMERFIELD; DOUGHERTY, 1997).

Atualmente, o aumento da capacidade computacional e a disponibilidade de grandes volumes de dados permitiram que algoritmos de AM fossem aplicados com maior eficiência em diferentes áreas, incluindo o mercado financeiro. No contexto econômico, esses algoritmos são utilizados para identificar padrões complexos, gerar previsões e automatizar decisões.

Além de sua flexibilidade e capacidade adaptativa, um dos principais benefícios do AM é a possibilidade de lidar com dados massivos e não estruturados, o que o torna uma ferramenta essencial para os mais diversos tipos de problemas. No entanto, desafios como *overfitting*, viés nos dados e a necessidade de interpretabilidade dos modelos permanecem, destacando a importância de rigorosos processos de validação e escolha adequada de algoritmos para cada contexto.

O AM pode ser classificado em três categorias: aprendizado supervisionado, não supervisionado e por reforço, como mostra a Figura 1. Cada um com a sua peculiaridade e método de agir. Nas subseções 2.1.1, 2.1.2 e 2.1.3 é explicado melhor cada um deles.

Figura 1 – Tipos de AM



Fonte: Elaborado pelo autor

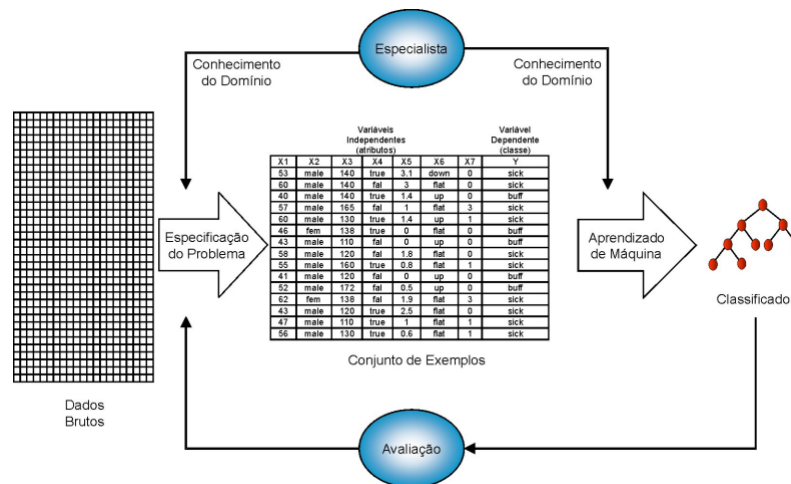
2.1.1 Aprendizado Supervisionado

O aprendizado supervisionado é um dos tipos mais comuns de AM e se destaca pela precisão em tarefas preditivas. Ele funciona utilizando dados rotulados, ou seja, informações de entrada associadas a uma saída conhecida. O objetivo é que o modelo aprenda a relacionar essas entradas e saídas, para que possa generalizar o conhecimento adquirido e fazer previsões em novos dados. Esse tipo de aprendizado é amplamente utilizado em tarefas de classificação, para valores discretos, e regressão, quando se deseja valores contínuos (MONARD; BARANAUSKAS, 2003).

Destaca-se ainda pela sua capacidade de realizar previsões precisas quando há dados de alta qualidade disponíveis. O que, no entanto, pode custar caro, tendo em vista que para isso ele necessita fortemente de grandes volumes de dados rotulados, algo relativamente custoso. Alguns cuidados também devem ser tomados para que o modelo não sofra de *overfitting*, isto é, se ajustar demais aos dados de treinamento, perdendo a capacidade de generalizar. Por isso, é essencial utilizar técnicas de regularização e validação cruzada para garantir um bom desempenho em dados não vistos.

Este será o aprendizado utilizado ao decorrer desse trabalho, mais especificamente com tarefas de classificação. Que, como já mencionado anteriormente, se refere a um tipo de tarefa que recebe atributos como entrada e partir deles tenta prever seu rótulo. A Figura 2 a seguir ilustra este processo.

Figura 2 – Funcionamento de um classificador.



Fonte: (MONARD; BARANAUSKAS, 2003).

2.1.2 Aprendizado Não Supervisionado

Diferente do aprendizado supervisionado, o aprendizado não supervisionado não trabalha com dados rotulados. O objetivo aqui é identificar padrões ou estruturas nos dados sem que o modelo saiba de antemão quais resultados procurar. O foco passa a se tornar a descoberta de relações implícitas entre as variáveis, e por isso é amplamente utilizado em tarefas como clusterização (agrupamento) e redução de dimensionalidade. Um exemplo clássico é a segmentação de clientes, em que o algoritmo agrupa consumidores com características semelhantes sem um rótulo definido.

A principal vantagem desse tipo de aprendizado é a capacidade de revelar padrões desconhecidos nos dados, proporcionando *insights* valiosos para empresas e pesquisadores. Contudo, como não há uma resposta correta explícita e a avaliação do modelo pode ser subjetiva, interpretar os resultados pode ser desafiador, e por este motivo, exige uma análise cuidadosa.

2.1.3 Aprendizado por Reforço

O aprendizado por reforço é um pouco diferente dos demais, pois é uma técnica baseada em um sistema de recompensas e punições. O modelo, chamado de agente, interage com o ambiente e aprende por tentativa e erro, recebendo uma recompensa positiva ou negativa com base em suas ações. O objetivo é maximizar a recompensa ao longo do tempo, desenvolvendo uma estratégia ou política ótima (SUTTON, 2018). Essa abordagem é frequentemente usada em cenários dinâmicos e complexos, como controle de robôs, jogos e veículos autônomos.

Embora o aprendizado por reforço seja promissor para resolver problemas complexos e dinâmicos, ele também apresenta desafios. Um dos maiores é o custo computacional elevado,

pois o agente precisa realizar muitas interações com o ambiente para aprender uma estratégia eficaz. Além disso, a escolha adequada da função de recompensa é essencial para evitar comportamentos indesejados por parte do agente. Apesar dessas dificuldades, esse tipo de aprendizado tem mostrado resultados impressionantes em aplicações reais, como o treinamento de robôs para operar em ambientes desconhecidos e o desenvolvimento de algoritmos que superaram campeões humanos em jogos complexos, como o xadrez.

2.2 XGBoost (*eXtreme Gradient Boosting*)

O XGBoost é o algoritmo de AM que será utilizado neste trabalho, para a previsão dos ativos financeiros. Ele é baseado em árvores de decisão, que se destaca pela sua eficiência e capacidade de lidar com grandes volumes de dados. Pertence à família de algoritmos de *boosting*, uma abordagem onde modelos fracos são combinados sequencialmente para formar um modelo robusto. A cada iteração, o XGBoost ajusta as previsões anteriores, minimizando o erro com base em gradientes da função de perda (CHEN; GUESTRIN, 2016).

A principal característica do XGBoost é seu foco em eficiência computacional e escalabilidade. Para isso, ele implementa várias otimizações, como a utilização de uma estrutura de colunas comprimidas, que armazena os dados já ordenados, facilitando o processo de divisão dos nós. O algoritmo também emprega uma técnica de subamostragem por colunas e regularização, o que ajuda a reduzir o *overfitting* e melhora o desempenho em cenários de dados complexos e ruidosos (CHEN; GUESTRIN, 2016).

Para descrever o funcionamento do XGBoost, que como já mencionado, se trata de um algoritmo em árvore, considere uma função de predição para um conjunto de N observações e K árvores, representada na Equação 2.1:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad f_k \in \mathcal{F}, \quad (2.1)$$

onde \hat{y}_i é a predição final para a observação i , x_i é o vetor de características, e \mathcal{F} representa o conjunto de todas as árvores de decisão possíveis. Cada f_k é uma árvore de decisão que contribui para a predição final.

O objetivo do XGBoost é minimizar uma função de perda L com regularização, dada pela Equação 2.2:

$$L(\phi) = \sum_{i=1}^N l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k), \quad (2.2)$$

em que l é a função de perda que mede a diferença entre a predição \hat{y}_i e o valor verdadeiro y_i . O termo de regularização $\Omega(f_k)$ penaliza a complexidade do modelo, contribuindo para a redução do *overfitting*. A regularização ocorre segundo a Equação 2.3:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2, \quad (2.3)$$

T é o número de folhas na árvore, γ controla a penalidade pelo número de folhas, λ controla a penalidade pelos pesos das folhas, e w_j são os pesos das folhas. Os parâmetros γ e λ são hiperparâmetros que podem ser ajustados para controlar a regularização do modelo.

O XGBoost utiliza uma técnica de expansão em segunda ordem para otimizar a função de perda. Na t -ésima iteração, a função objetivo $L^{(t)}$, como mostra a Equação 2.4, é aproximada por uma expansão de Taylor de segunda ordem:

$$L^{(t)} \approx \sum_{i=1}^N \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2 \right] + \Omega(f_t), \quad (2.4)$$

onde g_i e h_i representam o gradiente e o hessiano da função de perda em relação às previsões anteriores, e são definidos na Equação 2.5:

$$g_i = \frac{\partial l(y_i, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i^{(t-1)}}, \quad h_i = \frac{\partial^2 l(y_i, \hat{y}_i^{(t-1)})}{\partial (\hat{y}_i^{(t-1)})^2}. \quad (2.5)$$

Essas expressões permitem que o XGBoost ajuste as previsões anteriores com maior precisão, aproveitando informações de segunda ordem.

Além disso, o XGBoost lida eficientemente com valores ausentes. Durante a criação das árvores, ele ignora valores nulos por meio de um algoritmo de detecção de esparsidade, que permite continuar o processo de divisão sem prejuízo significativo.

Graças a essas inovações, o XGBoost é amplamente utilizado em competições como *Kaggle* e em aplicações que exigem previsões rápidas e precisas, como detecção de fraudes e classificação de texto (BENTÉJAC; CSÖRGŐ; MARTÍNEZ-MUÑOZ, 2021)

Embora muito poderoso, o XGBoost apresenta desafios, como a necessidade de ajuste fino de seus hiperparâmetros para obter o desempenho ideal. Parâmetros como a taxa de aprendizado (*learning rate*), a profundidade máxima das árvores e o número de árvores no *ensemble* precisam ser ajustados com cuidado para evitar tanto *overfitting* quanto *underfitting*. No entanto, uma vez bem ajustado, o XGBoost oferece resultados altamente precisos e generalizáveis.

2.3 Indicadores Financeiros

Os indicadores financeiros são ferramentas essenciais na análise técnica, e utilizadas para identificar padrões e tendências de preços no mercado. Eles auxiliam investidores e analistas na tomada de decisões, fornecendo sinais sobre momentos ideais de compra e venda. Cada indicador possui uma metodologia específica para analisar dados históricos de mercado, como

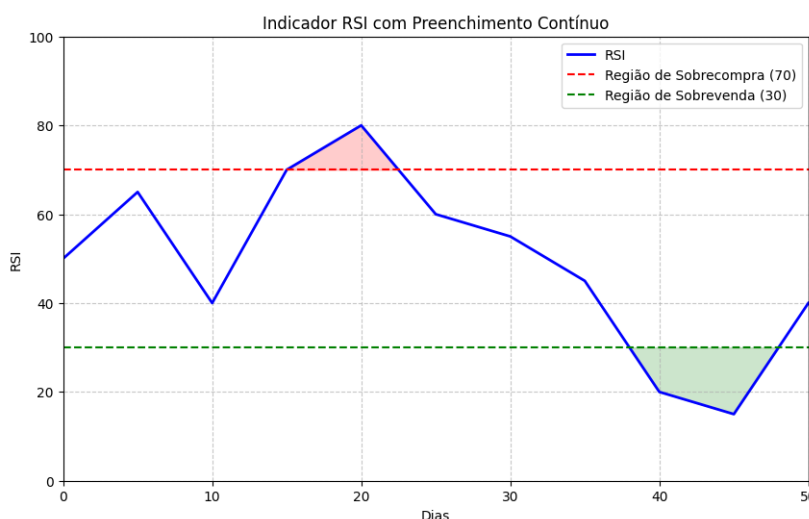
volume de negociação e preços de ativos. Nesta seção, será explorado alguns dos principais indicadores aplicados em estratégias de investimento e que serão utilizados por este estudo.

2.3.1 RSI (*Relative Strength Index*)

Desenvolvido por J. Welles Wilder, o RSI, ou Índice de Força Relativa, é um indicador de momento, muito usado no mercado de ativos, que mede a velocidade e a mudança dos movimentos de preços. Ele varia entre 0 e 100 e é utilizado para identificar condições de sobrecompra e sobrevenda. RSI é amplamente utilizado para prever reversões de tendência e avaliar a força dos movimentos de mercado. A Figura 3 demonstra seu funcionamento.

- **RSI > 70:** O ativo pode estar em sobrecompra (possível correção ou queda).
- **RSI < 30:** O ativo pode estar em sobrevenda (possível recuperação).

Figura 3 – Funcionamento do indicador RSI.



Fonte: Elaborado pelo autor.

O cálculo do RSI se baseia nos valores de ganho e perda média em um determinado período de tempo n , tipicamente definido como 14 dias, e é calculado segundo a Equação 2.6 e Equação 2.7:

$$\text{Ganho Médio} = \frac{\text{Soma dos Ganhos nos Últimos } n \text{ Períodos}}{n} \quad (2.6)$$

$$\text{Perda Média} = \frac{\text{Soma das Perdas nos Últimos } n \text{ Períodos}}{n} \quad (2.7)$$

Após obter os valores médios dos ganhos e das perdas, calculamos o *Relative Strength* da seguinte forma (Equação 2.8):

$$RS = \frac{\text{Ganho Médio}}{\text{Perda Média}} \quad (2.8)$$

Em seguida, o RSI é calculado com a fórmula (Equação 2.9):

$$RSI = 100 - \left(\frac{100}{1 + RS} \right) \quad (2.9)$$

O RSI é uma ferramenta útil e de fácil interpretação, porém por ser muito sensível ao mercado, é sempre recomendável combiná-lo com outros indicadores para confirmar os sinais.

2.3.2 MME (Média Móvel Exponencial)

A MME é uma média que atribui mais peso aos preços mais recentes, tornando-a mais sensível às variações do mercado. Ela é muito utilizada para identificar tendências e gerar sinais de compra e venda. É calculada de forma recursiva, com base na MME do período anterior e no preço atual, como na Equação 2.10.

$$MME_t = P_t \times \alpha + MME_{t-1} \times (1 - \alpha) \quad (2.10)$$

- α : Fator de suavização representado por $\alpha = \frac{2}{n+1}$.
- MME_t : MME atual.
- P_t : Preço de fechamento atual.
- MME_{t-1} : MME do período anterior.
- n : Número de períodos utilizado no cálculo (por exemplo, 12 ou 26).

2.3.3 MACD (*Moving Average Convergence Divergence*)

O MACD é um indicador que combina médias móveis para identificar mudanças na força, direção e duração de uma tendência. Foi desenvolvido por Gerald Appel em 1979, e combina duas linhas, a linha MACD e a linha de sinal. A primeira é calculada subtraindo uma MME de curto prazo, 12 períodos, de uma MME de longo prazo, 26 períodos. Já a segunda é calculada usando a MME da própria linha MACD.

O cruzamento entre a linha MACD e a linha de sinal é interpretado como um sinal de compra ou venda. Quando a linha MACD cruza a linha de sinal de baixo para cima indica um possível sinal de compra. Quando o contrário ocorre, ou seja, a linha MACD cruza a linha de sinal de cima para baixo, aponta um sinal de venda. Assim como o RSI, este indicador é usado para indicar tendências de alta ou baixo, ajudando os investidores a identificarem possíveis pontos de entrada e saída.

2.3.4 Oscilador Estocástico

O oscilador estocástico compara o preço de fechamento de um ativo com sua faixa de preços em um período específico. Ele é usado para identificar condições de sobrecompra e sobrevenda no mercado e varia entre 0 e 100. Para isso, utiliza duas linhas, a linha %K, que simboliza o preço momentâneo perante o intervalo definido, mostrada na Equação 2.11, e a linha %D, que é uma média móvel da linha %K, utilizada para suavizar o indicador e gerar sinais de compra ou venda, mostrada na Equação 2.12. A média móvel mais comum é de 3 períodos. A fórmula principal para o cálculo do estocástico é:

$$\%K = \frac{\text{Preço de Fechamento Atual} - \text{Mínimo mais Baixo do Período}}{\text{Máximo mais Alto do Período} - \text{Mínimo mais Baixo do Período}} \times 100 \quad (2.11)$$

$$\%D = \text{Média Móvel Simples de \%K (geralmente de 3 períodos)} \quad (2.12)$$

- Preço de Fechamento Atual: O preço de fechamento mais recente do ativo.
- Mínimo mais Baixo do Período: O menor preço atingido no período definido (geralmente 14 dias).
- Máximo mais Alto do Período: O maior preço atingido no mesmo período.

2.3.5 Cálculo das Bandas de Bollinger

As Bandas de Bollinger são responsáveis por medir a volatilidade de um ativo e indica se os preços estão relativamente altos ou baixos em relação a uma média móvel. As bandas consistem em três linhas principais: a banda superior, a banda inferior e a média móvel simples (linha central). A MMS é calculada segundo a Equação 2.13.

$$MMS = \frac{\sum_{i=1}^n P_i}{n} \quad (2.13)$$

- P_i : Preço de fechamento no período i .
- n : Número de períodos usados na média móvel (geralmente 20).

Com base nessa média, as bandas superior e inferior são calculadas usando o desvio padrão (σ) para medir a volatilidade do ativo, como na Equação 2.14 e na Equação 2.15.

$$\text{Banda Superior} = MMS + (k \times \sigma) \quad (2.14)$$

$$\text{Banda Inferior} = MMS - (k \times \sigma) \quad (2.15)$$

- k : Fator de desvio (geralmente $k = 2$).
- σ : Desvio padrão dos preços no período considerado.

As Bandas de Bollinger funcionam de modo que quando o preço se aproxima da banda superior, pode indicar que o ativo está sobrecomprado e quando se aproxima da banda inferior, indica possível sobrevenda. Ademais, a largura entre as bandas também se faz importante, já que quanto maior essa largura maior a volatilidade do mercado.

2.3.6 Índice de Canal de Commodities (CCI)

O Índice de Canal de Commodities mede a variação do preço de um ativo em relação à sua média ao longo de um período. Originalmente desenvolvido para identificar tendências de commodities, o CCI é amplamente utilizado em diversos mercados financeiros para identificar condições de sobrecompra e sobrevenda. A fórmula do CCI é mostrada na Equação 2.16:

$$CCI = \frac{P_t - \text{Média}_{TP}}{0,015 \times \text{Desvio Médio}} \quad (2.16)$$

- P_t : Preço típico atual, calculado como a média do preço máximo, mínimo e de fechamento (Equação 2.17):

$$P_t = \frac{\text{Máximo} + \text{Mínimo} + \text{Fechamento}}{3} \quad (2.17)$$

- Média_{TP} : Média dos preços típicos ao longo do período definido (geralmente 20 períodos).
- Desvio Médio: Média das diferenças absolutas entre cada P_t e a Média_{TP} no período.
- Constante 0,015: Usada para ajustar a escala do CCI, de forma que a maioria dos valores fique entre -100 e 100.

2.3.7 Cálculo do *Average True Range* (ATR)

ATR é um indicador de volatilidade que mede o intervalo médio de preço de um ativo ao longo de um determinado período. É usado para avaliar a volatilidade do mercado e é especialmente útil para definir pontos de stop-loss. Em uma fórmula recursiva, o ATR pode ser calculado como na Equação 2.18:

$$ATR_t = \frac{(ATR_{t-1} \times (n - 1)) + TR_t}{n} \quad (2.18)$$

Isso permite uma atualização contínua do ATR para cada novo período.

2.3.8 Cálculo do Parabolic SAR (*Stop and Reverse*)

O Parabolic SAR é um indicador técnico desenvolvido por J. Welles Wilder que ajuda a identificar a direção da tendência de um ativo e a definir pontos potenciais de reversão. Ele também é amplamente usado para definir pontos de entrada e saída em posições, uma vez que fornece sinais de *stop* e *reverse*. É calculado a partir da Equação 2.19.

$$SAR_{atual} = SAR_{anterior} + \alpha \times (EP - SAR_{anterior}) \quad (2.19)$$

- SAR_{atual} : Valor do Parabolic SAR no período atual.
- $SAR_{anterior}$: Valor do SAR no período anterior.
- α : Fator de aceleração (geralmente começa em 0,02 e aumenta em 0,02 a cada novo topo ou fundo atingido, até um limite de 0,2).
- EP (*Extreme Point*): O ponto extremo alcançado pelo preço durante a tendência atual (o mais alto em uma tendência de alta, o mais baixo em uma tendência de baixa).

2.3.9 Cálculo do *On-Balance Volume* (OBV)

Desenvolvido por Joseph Granville, o OBV é um indicador de volume. Ele mede a pressão de compra e venda ao acumular volume com base nas mudanças de preço, ajudando a identificar tendências e antecipar reversões, como na Equação 2.20.

$$OBV_t = \begin{cases} OBV_{t-1} + V_t & \text{se } P_t > P_{t-1} \\ OBV_{t-1} - V_t & \text{se } P_t < P_{t-1} \\ OBV_{t-1} & \text{se } P_t = P_{t-1} \end{cases} \quad (2.20)$$

- OBV_t : Valor do OBV no período atual.
- OBV_{t-1} : Valor do OBV no período anterior.
- V_t : Volume do período atual.
- P_t : Preço de fechamento atual.
- P_{t-1} : Preço de fechamento do período anterior.

2.3.10 Cálculo do *Percentage Price Oscillator* (PPO)

PPO é um indicador que mede a diferença percentual entre duas médias móveis exponenciais de preços. Ele é semelhante ao MACD, mas, em vez de mostrar a diferença

absoluta entre as médias, exibe a diferença em termos percentuais, como mostra a Equação 2.21, facilitando comparações entre ativos de diferentes preços.

$$PPO = \frac{\text{MME Rápida} - \text{MME Lenta}}{\text{MME Lenta}} \times 100 \quad (2.21)$$

- MME Rápida: A média móvel exponencial de curto prazo (geralmente 12 períodos).
- MME Lenta: A média móvel exponencial de longo prazo (geralmente 26 períodos).

2.3.11 Cálculo do *Average Directional Index* (ADX)

O ADX mede a força de uma tendência, independentemente de sua direção. Ele é amplamente utilizado para avaliar se um mercado está em uma tendência forte ou fraca, auxiliando na escolha de estratégias de trading adequadas.

2.3.12 Cálculo do Williams %R

Último indicador citado, o Williams %R, desenvolvido por Larry Williams, mede o nível do preço de fechamento em relação à faixa mais alta e mais baixa de preços de um ativo ao longo de um determinado período. Ele é usado para identificar condições de sobrecompra e sobrevenda, semelhante ao Estocástico, mas invertido, assim como na Equação 2.22.

$$\%R = \frac{\text{Máximo Alto} - \text{Preço de Fechamento}}{\text{Máximo Alto} - \text{Mínimo Baixo}} \times -100 \quad (2.22)$$

- Máximo Alto: O preço mais alto registrado no período definido (geralmente 14 períodos).
- Mínimo Baixo: O preço mais baixo registrado no período definido.
- Preço de Fechamento: O preço de fechamento atual do ativo.

3 Metodologia

Neste capítulo será abordado o método adotado para o desenvolvimento do presente trabalho, assim como, as ferramentas utilizadas, o conjunto de dados trabalhado, as métricas escolhidas para avaliar o modelo e todo o funcionamento do algoritmo desenvolvido.

3.1 Bibliotecas e Ferramentas

Para o desenvolvimento dos algoritmos de AM e análise dos dados financeiros, foram utilizadas diversas bibliotecas e ferramentas para auxiliarem nas etapas de coleta, manipulação, visualização e modelagem dos dados.

3.1.1 YahooQuery

Uma das principais bibliotecas para este trabalho, a YahooQuery permite o acesso direto à API do Yahoo Finance, possibilitando a coleta de uma ampla gama de dados financeiros, de forma rápida e eficiente. O usuário pode obter dados diretamente a partir de códigos específicos de ativos financeiros (por exemplo, PETR4.SA para a Petrobras), permitindo a aquisição de dados financeiros em tempo real e históricos.

A YahooQuery foi amplamente utilizada neste trabalho, por permitir o acesso rápido a dados confiáveis e detalhados, necessários para alimentar os modelos de AM. Sua integração com a linguagem Python facilita a coleta de grandes volumes de dados, bem como o pré-processamento e a análise desses dados, com alto grau de personalização.

Entre a vasta gama de informações disponibilizadas pela YahooQuery, destaca-se:

- **Dados de Preço:** Preço de abertura, fechamento, máxima e mínima diária, além do volume de negociação;
- **Indicadores Financeiros:** Indicadores técnicos e financeiros, como P/L (Preço sobre Lucro), P/VP (Preço sobre Valor Patrimonial), *dividend yield* e outros dados fundamentais;
- **Dados Históricos:** Histórico de preços de um ativo para diferentes períodos, podendo ser ajustado conforme a necessidade (diário, semanal, mensal);
- **Notícias e Eventos Corporativos:** Informações sobre dividendos, desdobramentos de ações e outras ações corporativas que impactam o preço do ativo;
- **Dados de Mercado:** Informação sobre índice de mercado, como o S&P 500, NASDAQ e outros índices relevantes.

3.1.2 Pandas

A biblioteca *Pandas* foi utilizada para manipulação e análise dos dados financeiros. Esta biblioteca permite o carregamento, transformação e análise de grandes volumes de dados de forma eficiente, com funcionalidades para filtragem, agregação e cálculo de estatísticas descritivas. Ela foi essencial para organizar e preparar o conjunto de dados antes da aplicação dos modelos de AM.

3.1.3 Matplotlib

Para visualização dos dados e dos resultados das previsões, utilizou-se a biblioteca *Matplotlib*. Ela foi usada para criar gráficos básicos de séries temporais e comparar visualmente os preços reais dos ativos com as previsões geradas pelos modelos.

3.1.4 Jupyter Notebook

Todo o desenvolvimento e execução do código foram realizados no ambiente *Jupyter Notebook*, que permite organizar o código em células, facilitando o teste e a documentação das etapas. Este ambiente também proporciona uma integração eficiente com as bibliotecas utilizadas, além de permitir a visualização interativa dos resultados.

3.2 Métricas Utilizadas

Para avaliar a performance dos modelos de AM neste projeto, foram selecionadas métricas que analisam tanto a precisão preditiva quanto o desempenho financeiro das estratégias desenvolvidas.

3.2.1 Acurácia

A *Acurácia* é amplamente utilizada para avaliar a proporção de previsões corretas em relação ao total de previsões realizadas.

Ela é medida segundo a Equação 3.1.

$$\text{Acurácia} = \frac{\text{Previsões Corretas}}{\text{Total de Previsões}} \times 100 \quad (3.1)$$

As previsões corretas referem-se aos casos em que a direção da previsão corresponde à direção real. Valores mais altos de acurácia indicam uma maior proporção de acertos nas previsões do modelo, no Quadro 1 tem-se um exemplo de um modelo com acurácia de 75%.

Quadro 1 – Exemplo Acurácia 75%

Feature 1 (Idade)	Feature 2 (Altura)	Target real	Target prevista
19	1,70	1	1
30	1,64	1	0
26	1,82	0	0
42	1,73	1	1

Fonte: Elaborado pelo autor.

3.2.2 F1 Score

O *F1 Score* combina precisão e sensibilidade (*recall*) em um único valor, proporcionando uma visão balanceada do desempenho do modelo. Ele é especialmente útil quando há um possível desequilíbrio entre classes, como alta e baixa. A fórmula do F1 Score é dada pela Equação 3.2.

$$\text{F1 Score} = 2 \times \frac{\text{Precisão} \times \text{Sensibilidade}}{\text{Precisão} + \text{Sensibilidade}} \quad (3.2)$$

A precisão é a proporção de previsões de alta corretas em relação ao total de previsões de alta, e a sensibilidade (ou *recall*) é a proporção de previsões corretas em relação ao total de movimentos reais. O F1 Score varia de 0 a 1, sendo que valores mais próximos de 1 indicam um bom desempenho em ambas as métricas.

3.2.3 Rendimento Final

Rendimento Final é uma métrica financeira simples e direta, que mede o percentual de retorno obtido ao final da aplicação da estratégia preditiva. Este cálculo fornece uma visão geral da lucratividade da estratégia, permitindo uma avaliação rápida de seu desempenho. Seu cálculo é descrito na Equação 3.3.

$$\text{Rendimento Final (\%)} = \frac{\text{Valor Final} - \text{Valor Inicial}}{\text{Valor Inicial}} \times 100 \quad (3.3)$$

3.2.4 Drawdown

O *Drawdown* representa a maior perda observada em uma sequência de investimentos, medindo a diferença entre um pico local e o menor valor subsequente no rendimento acumulado. O *drawdown* máximo é uma métrica importante para avaliar o risco da estratégia, sendo calculado pela fórmula da Equação 3.4.

$$\text{Drawdown} = \frac{\text{Valor Máximo} - \text{Valor Mínimo}}{\text{Valor Máximo}} \quad (3.4)$$

Essa métrica é especialmente relevante para investidores que desejam compreender os riscos associados a uma estratégia e a sua estabilidade ao longo do tempo. Valores menores de *drawdown* indicam menor risco de perda significativa ao longo do período de investimento.

3.3 Implementação do Modelo

Esta seção descreve a implementação do modelo de previsão, incluindo a coleta e preparação dos dados e a aplicação dos algoritmos de AM. A implementação foi organizada para garantir uma estrutura consistente, desde a obtenção dos dados até a avaliação do modelo com métricas específicas.

3.3.1 Conjunto de Dados

Para este trabalho, foram selecionados dados históricos de preços de fechamento de cinco ativos financeiros brasileiros: VALE3.SA (Vale), BBDC4.SA (Banco Bradesco), ITUB4.SA (Itaú Unibanco), JBSS3.SA (JBS) e WEGE.SA (WEG). Esses ativos foram escolhidos devido à sua relevância no mercado brasileiro e visando criar um portfólio variado.

Foram analisados os últimos dez anos de cada ativo financeiro e coletados a data, preço de fechamento, máxima, mínima e volume de cada um.

3.3.2 Pré-processamento dos Dados

O pré-processamento dos dados consistiu em limpar e estruturar as informações coletadas para garantir a qualidade da entrada nos modelos de AM. Entre as etapas de pré-processamento, destacam-se o tratamento de valores ausentes e a normalização dos indicadores para facilitar o treinamento.

3.3.3 Divisão dos Dados

Visando avaliar a capacidade preditiva do modelo, o conjunto de dados foi dividido em duas partes:

- **Treinamento:** 80% dos dados históricos, utilizados para ajustar os parâmetros do modelo.
- **Teste:** 20% restantes dos dados, reservados para avaliar a performance do modelo em dados novos.

A divisão foi feita de forma temporal, preservando a ordem cronológica dos dados para evitar vazamento de informação e garantir que a avaliação seja mais próxima do cenário real de previsão.

3.3.4 Treinamento do Modelo

Para a previsão dos movimentos de preços, foi utilizado o XGBoost. Foram elaborados três modelos que apresentam uma única diferença: os atributos de entrada. O primeiro recebeu como entrada apenas o indicador RSI. O segundo recebeu apenas o MACD. Já o terceiro modelo utilizou doze atributos, são eles os indicadores: RSI, MACD, MME, Oscilador Estocástico, Bandas de Bollinger, CCI, ATR, Parabolic SAR, OBV, PPO, ADX e Williams %R. Os modelos foram ajustados para prever três rótulos diferentes

- **0:** Quando o preço cai mais de 2%.
- **1:** Quando o preço sobe mais de 2%.
- **2:** Quando o preço se mantém na faixa de 2%.

Deste modo, a partir das variáveis de entrada o algoritmo tentar prever a tendência do mercado e, assim, identificar pontos de entrada e saída. O processo de treinamento incluiu ainda a definição dos hiperparâmetros, utilizando *RandomizedSearchCV* com o conjunto de dados de treinamento, realizado para maximizar a performance do modelo.

Por fim, objetivando analisar a eficácia do algoritmo, foi feita uma simulação com um capital inicial de dez mil reais divididos igualmente entre os ativos do portfólio. A simulação foi aplicada ao conjunto de dados de teste e consistia em usar todo o capital disponível na compra de determinada ação quando o modelo previa uma alta de mais de 2%. As ações compradas eram mantidas enquanto o seu preço se mantia na faixa de 2%, até o algoritmo prever uma queda maior que isso, e então vendia todas as ações. Esse processo era mantido até o final, com o intuito de avaliar o saldo final do investimento.

4 Resultados

Neste capítulo, são apresentados e analisados os resultados obtidos a partir dos experimentos realizados com os indicadores financeiros e algoritmo de AM aplicado no estudo. O desempenho de cada método é avaliado com base em métricas específicas, e os resultados são ilustrados por meio de gráficos e tabelas, facilitando a comparação e interpretação.

4.1 Modelo com RSI

Este foi o modelo treinado apenas com o RSI como indicador financeiro, o Quadro 2 apresenta as métricas de cada ativo analisado pelo modelo.

Quadro 2 – Métricas do modelo com RSI

Ativo	Acurácia	F1-Score	Rendimento
VALE3.SA	73.44%	0.72	3.09%
BBDC4.SA	80.28%	0.73	4.37%
ITUB4.SA	85.11%	0.80	9.24%
JBSS3.SA	65.79%	0.60	5.39%
WEGE3.SA	84.51%	0.79	52.75%

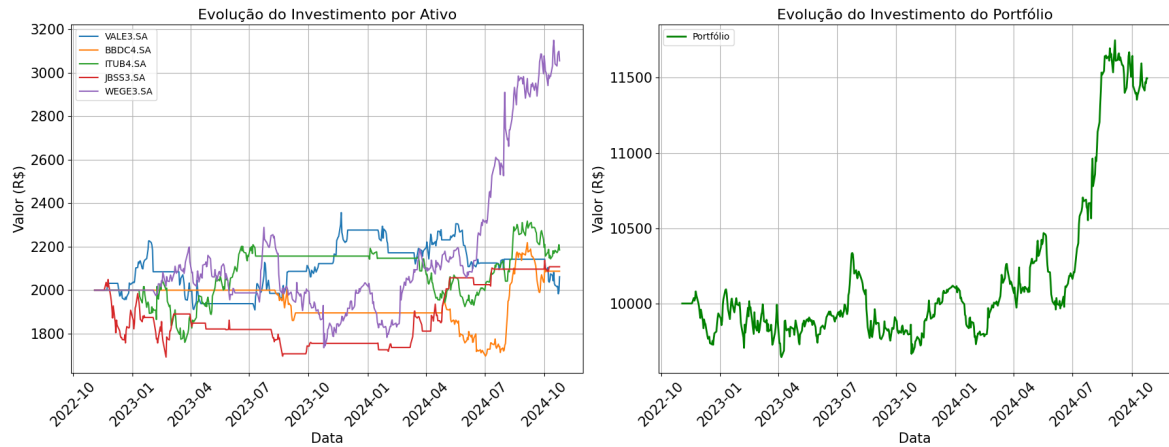
Fonte: Elaborado pelo autor.

Tendo como base o Quadro 2 é possível perceber uma acurácia e F1-Score relativamente altos, porém rendimentos menores do que o esperado em sua maioria, possivelmente reflexo de um modelo treinado a partir de um único indicador. As Figuras 4, 5 e 6, exemplificam através de gráficos o comportamento dos ativos e do modelo perante uma simulação no mercado financeiro. A Figura 4 mostra a evolução do capital aplicado em cada ativo e também no portfólio como um todo. Nela é possível verificar que o capital aplicado nas ações da WEG cresceu consideravelmente e que nenhum dos ativos apresentou prejuízo.

Seguindo a mesma lógica que a anterior, a Figura 5 tem como objetivo mostrar o percentual acumulado do capital ao longo do período em que este esteve aplicado nos ativos em questão. Novamente, observa-se um percentual muito maior para as ações da WEG. Vendo o portfólio como um todo, o retorno percentual acumulado se mostrou constante, até meados de julho de 2024, onde começou a apresentar um crescimento elevado.

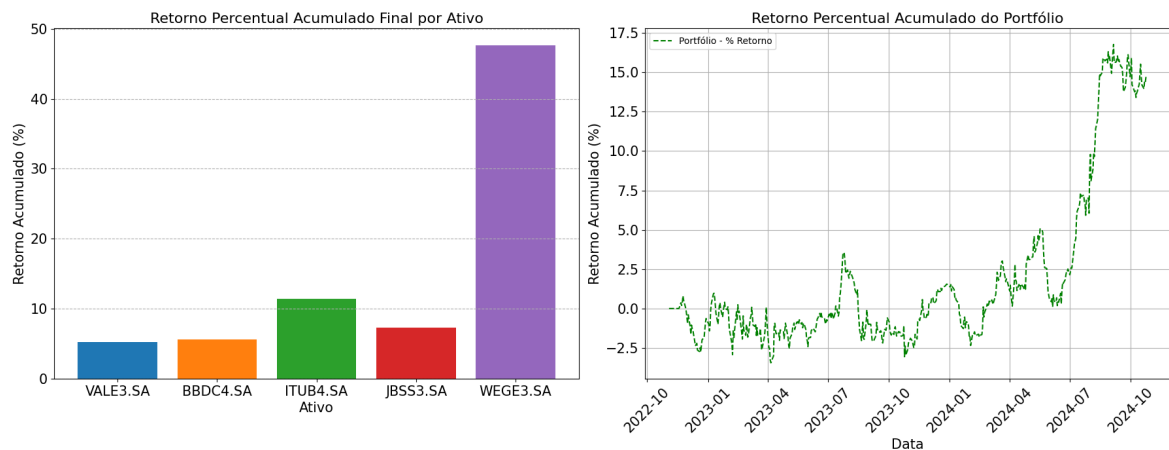
Já a Figura 6, representa outra métrica importante, o *drawdown*. As ações da WEG foram a que apresentou maior percentual de *drawdown*, indicando que apesar de obter mais lucros, também foi a que obteve mais risco associado a ela. Nos demais ativos os valores da métrica em questão se mostraram perto dos 15%.

Figura 4 – Evolução do investimento (ativo x portfólio) - RSI



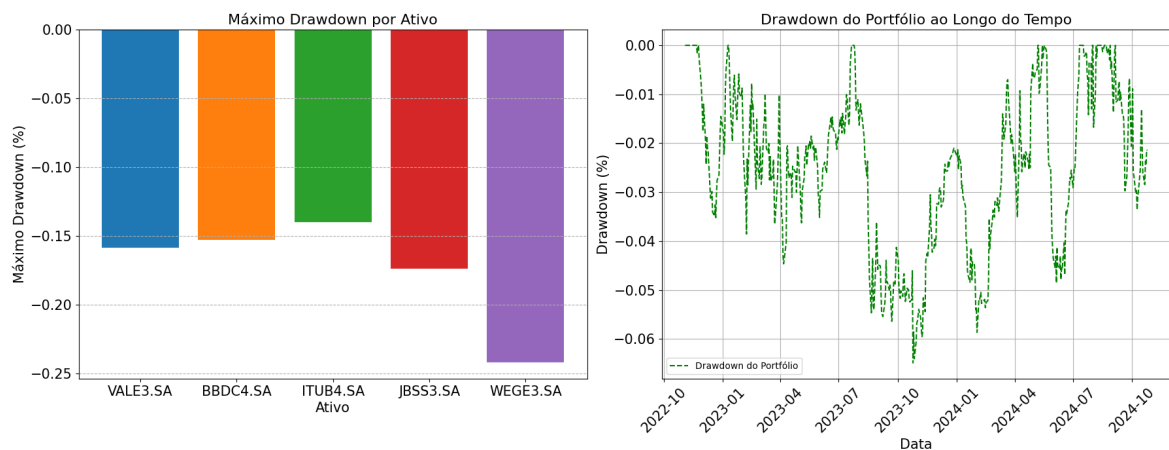
Fonte: elaborado pelo autor.

Figura 5 – Retorno percentual acumulado (ativo x portfólio) - RSI



Fonte: elaborado pelo autor.

Figura 6 – Drawdown (ativo x portfólio) - RSI



Fonte: elaborado pelo autor.

4.2 Modelo com MACD

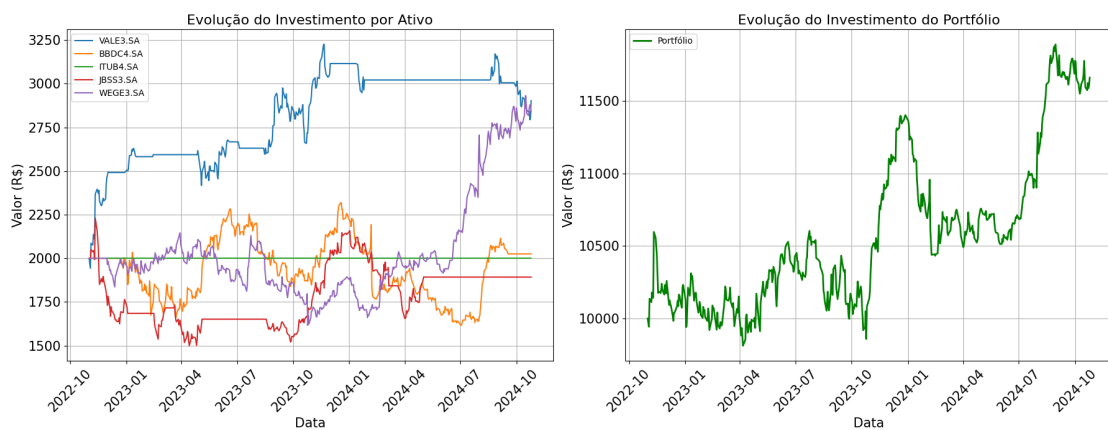
Assim como o anterior, este modelo foi treinado utilizando apenas um indicador, porém desta foi o MACD. O Quadro 3 mostra as métricas do modelo.

Quadro 3 – Métricas do modelo com MACD

Ativo	Acurácia	F1-Score	Rendimento
VALE3.SA	75.50%	0.72	45,14%
BBDC4.SA	79.52%	0.72	1,25%
ITUB4.SA	87.35%	0.81	00.00%
JBSS3.SA	68.47%	0.61	-5.38%
WEGE3.SA	84.74%	0.79	42.00%

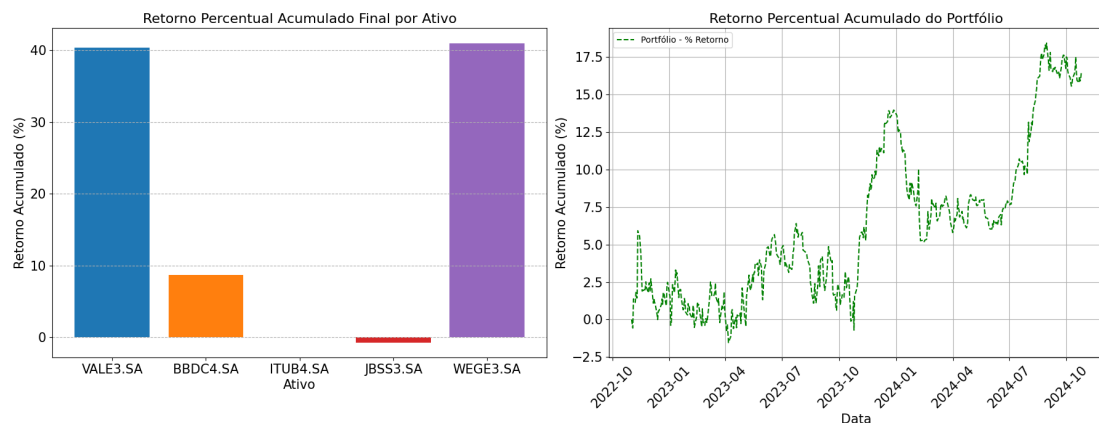
Fonte: Elaborado pelo autor.

Figura 7 – Evolução do investimento (ativo x portfólio) - MACD



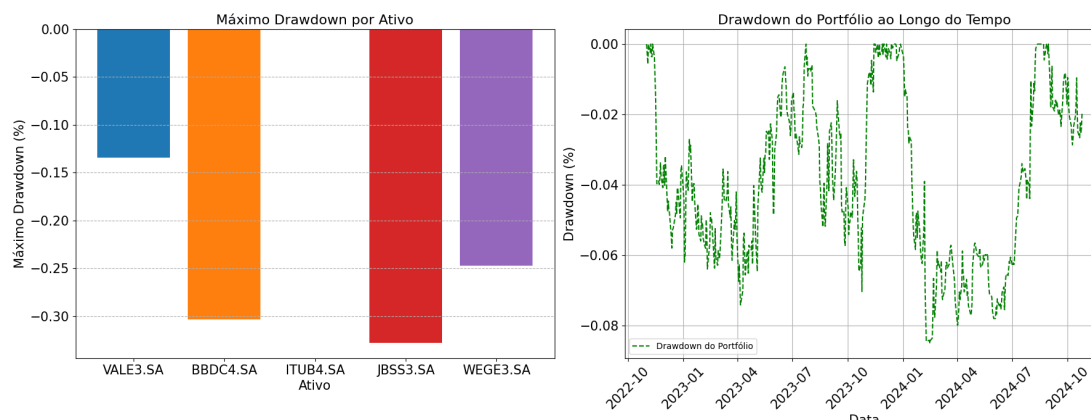
Fonte: elaborado pelo autor.

Figura 8 – Retorno percentual acumulado (ativo x portfólio) (ativo x portfólio) - MACD



Fonte: elaborado pelo autor.

Figura 9 – Drawdown (ativo x portfólio) - MACD



Fonte: elaborado pelo autor.

Analisando as métricas apresentadas pelo Quadro 3, é possível perceber bastante similaridade com as do modelo anterior, em questão de acurácia e F1-Score. Os gráficos da Figura 7, mostra a transformação do capital aplicado a cada ativo, do mês de Outubro de 2022 até Outubro de 2024. Contudo, ao observar os gráficos dispostos na sequência, fica evidente problemas na predição dos ativos, sobretudo na do Itaú Unibanco, que se manteve constante, e da JBS que teve uma pequena queda em relação ao capital inicial.

A Figura 8, representa o retorno percentual acumulado de cada ativo e também do portfólio como um todo. Nota-se, que a barra das ações da JBS estão abaixo do eixo X, o que representa o prejuízo do valor aplicado nelas. No entanto, o retorno do do portfólio foi positivo, o que significa que os outros ativos foram capazes de cobrir o prjuízo causado pelas ações da JBS. As ações do Itaú Unibanco não apresentam barra própria, já que por se manterem com o mesmo capital investido durante todo o período de aplicação, não tiveram retorno percentual, nem negativo, nem positivo.

Já a Figura 9, exhibe os gráficos do *drawdown*, importante para se ter uma análise sobre o risco de cada investimento. As barras laranja e vermelha, referentes as ações do Banco Bradesco e JBS, respectivamente, passaram da linha de 30% de *drawdown*, o que significa que em algum momento o investimento perdeu mais de 30% do capital máximo de cada ativo, um risco relativamente alto. Vendo o contexto geral do modelo, verifica-se que ele não foi capaz de atuar no mercado de maneira constante e eficaz, tendo problemas com a predição do ativos e gerando rendimentos incertos, altos para alguns ativos, porém baixos para outros.

4.3 Modelo completo

Último modelo desenvolvido, se caracteriza por ser o mais completo, se beneficiando da análise de doze indicadores técnicos diferentes e utilizados em conjunto. Suas métricas estão

no Quadro 4 e é possível notar um sutil aumento na acurácia e F1-Score.

Quadro 4 – Métricas do modelo completo

Ativo	Acurácia	F1-Score	Rendimento
VALE3.SA	80,36%	0.74	4.19%
BBDC4.SA	79,96%	0.73	13.97%
ITUB4.SA	87.45%	0.82	52.07%
JBSS3.SA	69.03%	0.60	49.29%
WEGE3.SA	85.43%	0.80	15.97%

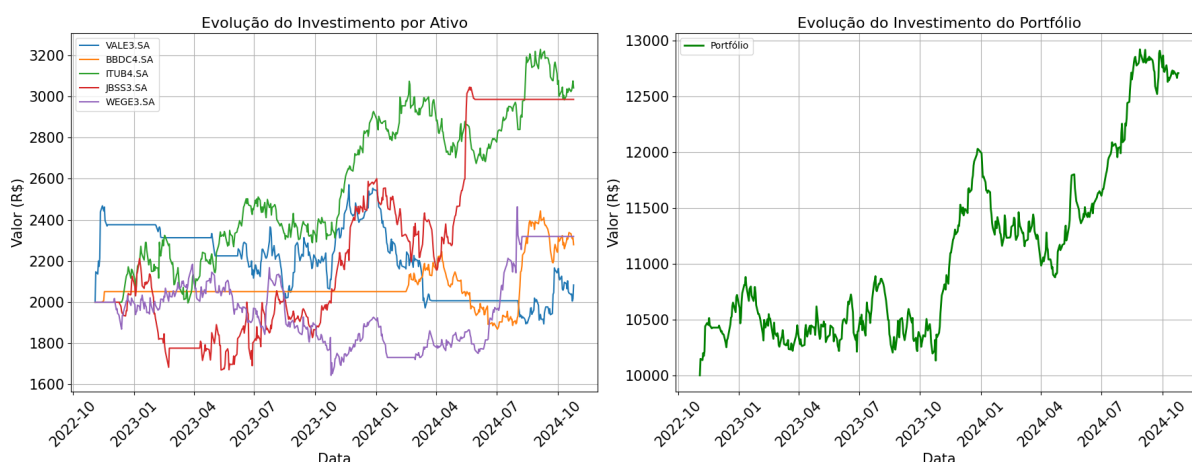
Fonte: Elaborado pelo autor.

Foi o algoritmo que mais gerou retorno ao fim do período de investimento e, a Figura 10, comprova este fato. Nela, nota-se que nenhum dos ativos obteve prejuízo, e destaca-se os ativos da JBS e Itaú Unibanco, que foram os mais beneficiados. Além disso, a evolução do portfólio seguiu uma linha crescente e positiva.

A Figura 11, com o retorno percentual acumulado dos ativos e do investimento, exibe taxas de retorno consideravelmente altas para os ativos, com as ações da JBS e Itaú Unibanco passando dos 40%. Deste modo, todos os ativos contribuíram positivamente para o crescimento do portfólio, onde o percentual acumulado ficou próximo de 25%.

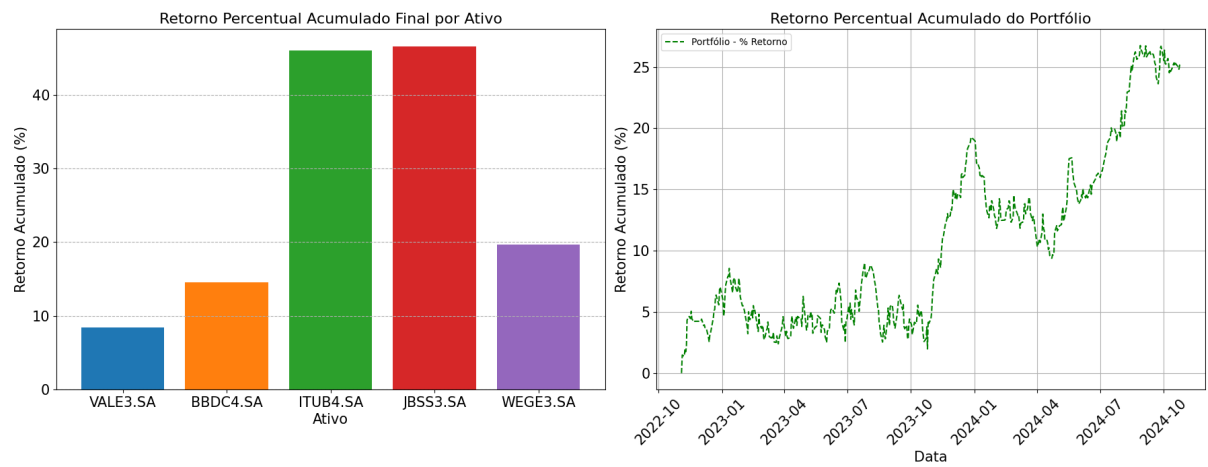
Na Figura 12, observa-se um *drawdown* entre 15% e 25% majoritariamente, não atingindo valores tão altos como o do modelo anterior, o que é um ponto positivo. Isso também, atribui um menor risco associado ao modelo, o tornando mais confiável. Nesse contexto, a partir dos dados apresentados, este modelo foi o que gerou melhores resultados e menores riscos. Isso comprova ainda, a eficácia do uso combinado de diferentes indicadores para obter análises mais precisas e consistentes.

Figura 10 – Evolução do investimento (ativo x portfólio) - Modelo completo



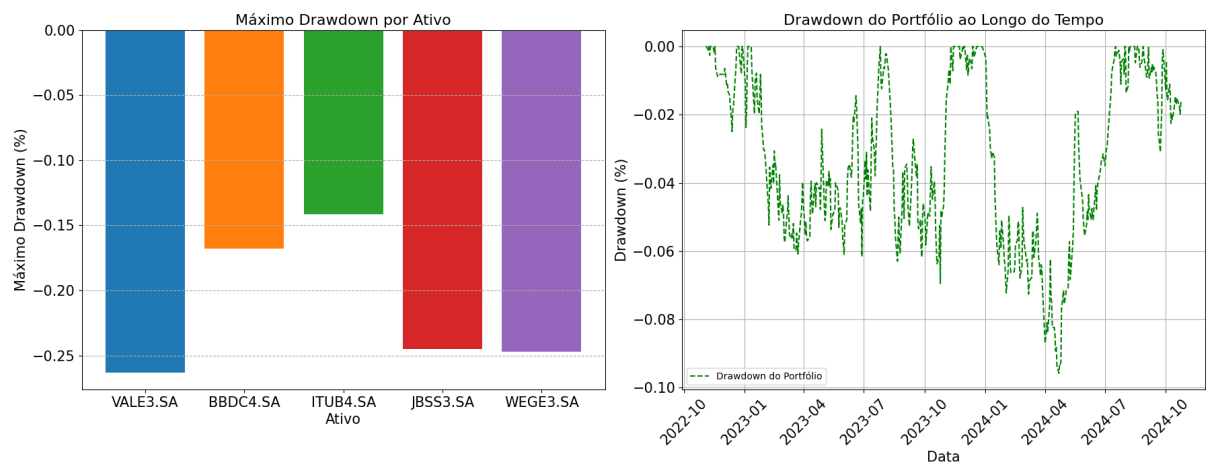
Fonte: elaborado pelo autor.

Figura 11 – Retorno percentual acumulado (ativo x portfólio) - Modelo completo



Fonte: elaborado pelo autor.

Figura 12 – Drawdown (ativo x portfólio) - Modelo completo



Fonte: elaborado pelo autor.

5 Conclusão

Em suma, este trabalho objetivou a criação de algoritmos de AM para prever o movimento de ativos financeiros, a partir da medição das tendências de mercado. Foram obtidos resultados que comprovam a capacidade de se obter bons retornos aplicando a técnica em questão, além de ser averiguado o fato de que a utilização de mais de um indicador em conjunto geram desisões mais consistentes.

A grande disposição de dados financeiros para o estudo, também foi fator determinando na realização do mesmo. Apoiado por eles, foi possível uma análise detalhada dos ativos utilizados, o que contribuiu altamente para o treinamento do modelo.

O estudo se faz bastante interessante para o mercado de ações, podendo colaborar positivamente na tomada de decisão de investidores, tanto experientes quanto iniciantes, pela sua alta facilidade de análise dos resultados. Todavia, é fundamental destacar, que assim como os outros indicadores de mercado, o algoritmo apresentado não deve ser utilizado isoladamente e como único fator para a tomada de decisão, e sim em conjunto, como uma métrica a mais.

5.1 Trabalhos futuros

Como este se trata de um tema bastante abrangente e complexo, existem diversos pontos que ainda podem vir a serem explorados dando continuidade ao trabalho. Há margem para melhora na acurácia dos resultados, assim como na consistência das predições, permitindo melhores tomadas de decisões. Ainda nesse contexto, outros tipos de algoritmos de AM podem ser empregados, como *Random Forest*, Máquinas de Vetores de Suporte e Redes Neurais Artificiais.

Também há espaço para a diversificação do portfólio, para que mais ativos sejam analisados e estudados. E, seguindo a mesma linha de raciocínio, o desenvolvimento de um ambiente, onde diferentes ativos possam ser avaliados, pode vir a ser uma maneira eficaz para a visualização dos resultados.

Referências

- AMORIM, K. A. F. de; LUCENA, G. K. F.; GIRÃO, L. F. d. A. P.; QUEIROZ, D. B. de. A influência da educação financeira na inserção dos investidores no mercado de capitais brasileiro: um estudo com discentes da área de negócios. *Race: revista de administração, contabilidade e economia*, Editora Unoesc, v. 17, n. 2, p. 567–590, 2018. Disponível em: <<https://dialnet.unirioja.es/servlet/articulo?codigo=6557495>> . Acesso em: 28 de outubro de 2024.
- BENTÉJAC, C.; CSÖRGŐ, A.; MARTÍNEZ-MUÑOZ, G. A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, Springer, v. 54, p. 1937–1967, 2021. Disponível em: <<https://link.springer.com/article/10.1007/S10462-020-09896-5>> . Acesso em: 28 de outubro de 2024.
- CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. [S.l.: s.n.], 2016. p. 785–794. Disponível em: <<https://dl.acm.org/doi/abs/10.1145/2939672.2939785>> . Acesso em: 28 de outubro de 2024.
- CHOPRA, R.; SHARMA, G. D. Application of artificial intelligence in stock market forecasting: a critique, review, and research agenda. *Journal of risk and financial management*, MDPI, v. 14, n. 11, p. 526, 2021. Disponível em: <<https://www.mdpi.com/1911-8074/14/11/526>> . Acesso em: 28 de outubro de 2024.
- FAMA, E. F. *Efficient Capital Markets: A Review of Theory and Empirical Work*. [S.l.: s.n.], 1970. v. 25. 383–417 p. Disponível em: <https://econpapers.repec.org/article/blajfinan/v_3a25_3ay_3a1970_3ai_3a2_3ap_3a383-417.htm> . Acesso em: 28 de outubro de 2024.
- GONÇALVES, L. I. O uso de indicadores técnicos como suporte à tomada de decisões no mercado financeiro. 2018. Disponível em: <<https://lume.ufrgs.br/handle/10183/193323>> . Acesso em: 28 de outubro de 2024.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. Disponível em: <<https://mitpress.mit.edu/9780262035613/deep-learning/>> . Acesso em: 28 de outubro de 2024.
- JAIN, R.; VANZARA, R. Emerging trends in ai-based stock market prediction: A comprehensive and systematic review. *MDPI Engineering Proceedings*, v. 56, n. 1, 2023. Disponível em: <<https://www.mdpi.com/2673-4591/56/1/254>> . Acesso em: 28 de outubro de 2024.
- JUNIOR, W. H.; REGONHA, C.; CIPOLLA, J. H. M. Dificuldades para um pequeno investidor se tornar eficiente e eficaz operando na modalidade day trade: um estudo de caso. 2021. Disponível em: <https://www.revistaifpsr.com/v1133_817.pdf> . Acesso em: 28 de outubro de 2024.
- KOHAVI, R.; SOMMERFIELD, D.; DOUGHERTY, J. Data mining using a machine learning library in c++. *International Journal on Artificial Intelligence Tools*, World Scientific, v. 6, n. 04, p. 537–566, 1997. Disponível em: <<https://www.worldscientific.com/doi/abs/10.1142/S021821309700027X>> . Acesso em: 28 de outubro de 2024.

LUDERMIR, T. B. Inteligência artificial e aprendizado de máquina: estado atual e tendências. *Estudos Avançados*, SciELO Brasil, v. 35, p. 85–94, 2021. Disponível em: <<https://www.scielo.br/j/ea/a/wXBdv8yHBV9xHz8qG5RCgZd/?lang=pt&format=html>> . Acesso em: 28 de outubro de 2024.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. *Sistemas inteligentes-Fundamentos e aplicações*, v. 1, n. 1, p. 32, 2003. Disponível em: <<https://dcm.ffclrp.usp.br/~augusto/publications/2003-sistemas-inteligentes-cap4.pdf>> . Acesso em: 28 de outubro de 2024.

NETO, D. d. S. C.; FERREIRA, A. S.; MASCARENHAS, C. C.; MAY, A. M.; VASSÃO, T. C.; RAÓ, E. M. A bolsa de valores e os novos investidores nos tempos atuais. *Revista Gestão em Foco*, v. 14, p. 310–321, 2022. Disponível em: <<https://portal.unisepe.com.br/unifia/wp-content/uploads/sites/10001/2022/03/A-BOLSA-DE-VALORES-E-OS-NOVOS-INVESTIDORES-309-a-320.pdf>> . Acesso em: 28 de outubro de 2024.

SUTTON, R. S. Reinforcement learning: An introduction. *A Bradford Book*, 2018. Disponível em: <<https://books.google.com.br/books?hl=pt-BR&lr=&id=sWV0DwAAQBAJ&oi=fnd&pg=PR7&dq=Reinforcement+Learning:+An+Introduction&ots=1-96q4frZb&sig=dYYWOafzCaMfQ-3HX1Bw4dPkCx8#v=onepage&q=Reinforcement%20Learning%3A%20An%20Introduction&f=false>> . Acesso em: 28 de outubro de 2024.

VINOTHA, C.; ALSULAMI, F.; SIDDIQUI, M. S.; SAXENA, R.; KANKARIYA, R. et al. Artificial intelligence in financial markets: Predictive models and risk management strategies. In: IEEE. *2024 Ninth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)*. [S.l.], 2024. p. 1–5. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/10568778>> . Acesso em: 28 de outubro de 2024.