

# UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

**Explorando a Inteligência Artificial na Previsão de  
Tendências e Preços no Contexto do Swing Trading:  
Um Estudo de Caso para Aprimorar Estratégias de  
Investimento**

**Claudinei Felix**

Monografia - MBA em Inteligência Artificial e Big Data



SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: \_\_\_\_\_

**Claudinei Felix**

**Explorando a Inteligência Artificial na Previsão de  
Tendências e Preços no Contexto do Swing Trading: Um  
Estudo de Caso para Aprimorar Estratégias de  
Investimento**

Monografia apresentada ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Inteligência Artificial e Big Data.

Área de concentração: Inteligência Artificial

Orientador: Prof.<sup>a</sup> Dr.<sup>a</sup> Marislei Nishijima

**Versão original**

**São Carlos**  
**2024**

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO,  
POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E  
PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi, ICMC/USP, com os dados fornecidos pelo(a) autor(a)

S856m	Felix, Claudinei Explorando a Inteligência Artificial na Previsão de Tendências e Preços no Contexto do Swing Trading: Um Estudo de Caso para Aprimorar Estratégias de Investimento / Claudinei Felix ; orientador Marislei Nishijima. – São Carlos, 2024. 40 p. : il. (algumas color.) ; 30 cm. Monografia (MBA em Inteligência Artificial e Big Data) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2024. 1. LaTeX. 2. abnTeX. 3. Classe USPSC. 4. Editoração de texto. 5. Normalização da documentação. 6. Tese. 7. Dissertação. 8. Documentos (elaboração). 9. Documentos eletrônicos. I. Nishijima, Marislei, orient. II. Título.
-------	---

**Claudinei Felix**

**Exploring Artificial Intelligence in Forecasting Trends and Prices in the Context of Swing Trading: A Case Study to Enhance Investment Strategies**

Monograph presented to the Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, as part of the requirements for obtaining the title of Specialist in Artificial Intelligence and Big Data.

Concentration area: Artificial Intelligence

Advisor: Prof. Dr. Marislei Nishijima

**Original version**

**São Carlos  
2024**



## RESUMO

Gôlo, M. P. S. **Explorando a Inteligência Artificial na Previsão de Tendências e Preços no Contexto do Swing Trading: Um Estudo de Caso para Aprimorar Estratégias de Investimento.** 2024. 40p. Monografia (MBA em Inteligência Artificial e Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2024.

Quais são as melhores oportunidades de compra e venda em operações de investimento na bolsa de valores Bovespa? O mercado acionário depende de múltiplos fatores, sejam eles políticos, econômicos, globais, climáticos, entre tantos outros. Quais seriam as melhores oportunidades de operar neste mercado usando “**Swing Trading**”? Este trabalho propõe uma investigação sobre a aplicação da inteligência artificial (**IA**) na previsão de melhores oportunidades nestas operações, usando conhecimento técnico de mercado. Operadores neste mercado utilizam duas abordagens principais como parâmetros para compras ou vendas na sua tomada de decisão. O método de análise técnica/gráfica usa o preço histórico das ações-alvo e de outros históricos (moedas, empresas, commodities, taxas de juros, etc.) que possam estar correlacionados com a ação-alvo. O segundo tipo de análise é qualitativa, envolve métricas mais difíceis de quantificar, como perfil da empresa, situação do mercado, fatores políticos e econômicos, informações textuais na forma de novos artigos financeiros, mídias sociais e até publicações de analistas econômicos. Este estudo abrange a análise de algoritmos de aprendizado de máquina e técnicas de processamento de dados para desenvolver um modelo preditivo que se baseia numa combinação destas abordagens mencionadas, mais especificamente, usa dados do mercado e os conhecimentos de operadores transformados em variáveis de entrada. Um exemplo, neste último caso, é o uso da análise técnica e padrões "**CandleStick**" para gerar variáveis de entrada, pois esta ferramenta fornece informações valiosas para investidores e profissionais financeiros interessados em aprimorar suas estratégias nas tomadas de decisão no mercado financeiro.

**Palavras-chave:** Aprendizado de máquina, BOVESPA, Candlestick, Swing Trading.



## **ABSTRACT**

Gôlo, M. P. S. **Exploring Artificial Intelligence in Forecasting Trends and Prices in the Context of Swing Trading: A Case Study to Enhance Investment Strategies.** 2024. 40p. Monograph (MBA in Artificial Intelligence and Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2024.

What are the best buying and selling opportunities in investment operations on the Bovespa stock exchange? The stock market depends on multiple factors, whether political, economic, global, climatic, among many others. What would be the best opportunities to operate in this market using "**Swing Trading**"? This work proposes an investigation into the application of artificial intelligence (AI) in predicting the best opportunities in these operations, using technical market knowledge. Operators in this market use two main approaches as parameters for buying or selling in their decision-making. The technical/graphical analysis method uses the historical price of the target stocks and other historical data (currencies, companies, commodities, interest rates, etc.) that may be correlated with the target stock. The second type of analysis is qualitative, involving metrics that are more difficult to quantify, such as company profile, market situation, political and economic factors, textual information in the form of new financial articles, social media, and even publications from economic analysts.

This study covers the analysis of machine learning algorithms and data processing techniques to develop a predictive model based on a combination of these mentioned approaches, more specifically, using market data and the knowledge of operators transformed into input variables. An example, in this last case, is the use of technical analysis and "**CandleStick**" patterns to generate input variables, as this tool provides valuable information for investors and financial professionals interested in improving their decision-making strategies in the financial market.

**Keywords:** Machine Learning, BOVESPA, Swing Trading.



## **LISTA DE FIGURAS**

Figura 1 – Índice de Força Relativa . . . . .	27
Figura 2 – Convergência e Divergência das Médias Móveis . . . . .	27
Figura 3 – Médias Móveis, MMA e MME . . . . .	28
Figura 4 – Fórmula Heikin-Ashi . . . . .	28
Figura 5 – Petrobras: Indicadores . . . . .	28
Figura 6 – Vale: Indicadores . . . . .	29
Figura 7 – Weg S.A.: Indicadores . . . . .	29
Figura 8 – Bradesco ON: Indicadores . . . . .	30
Figura 9 – BRF Brasil Foods: Indicadores . . . . .	30
Figura 10 – Variação percentual diária . . . . .	32
Figura 11 – Resultados dos algorítimos . . . . .	35



## **LISTA DE TABELAS**

Tabela 1 – Padrões Candlestick . . . . .	23
Tabela 2 – Conjuntos de Dados . . . . .	31
Tabela 3 – Desempenho do Algoritmo SVM com diferentes configurações de Kernel	33
Tabela 4 – Algoritmo XGB com diferentes parâmetros . . . . .	33
Tabela 5 – Desempenho do Algoritmo GaussianNB com diferentes Var smoothing	33
Tabela 6 – Algoritmo KNN com diferentes valores de K . . . . .	33
Tabela 7 – Algoritmo SGDClassifier com diferentes configurações . . . . .	34
Tabela 8 – Algoritmo LGBM com diferentes configurações . . . . .	34
Tabela 9 – Algoritmo MLP com diferentes configurações . . . . .	34
Tabela 10 – Melhores algoritmos de acordo com os parâmetros de precisão, recall e roc_auc. . . . .	34



## LISTA DE ABREVIATURAS E SIGLAS

SWT	<i>Swing Trading</i>
IBRX-100	<i>Indice das 100 ações mais negociadas na BM&amp;F BOVESPA</i>
RNN	Redes Neurais Recorrentes
RNA	Redes Neurais Artificiais
PETR4	<i>Petróleo Brasileiro S.A. (Petrobras)</i>
BM&F BOVESPA	Bolsa de Valores de São Paulo
IGPM	<i>Índice Geral de Preços ao Consumidor</i>
IPCA	<i>Índice Nacional de Preços ao Consumidor Amplo</i>
INPC	<i>Índice Nacional de Preços ao Consumidor</i>
SELIC	<i>Sistema Especial de Liquidação e Custódia</i>
MMCD	Média Móvel Convergente e Divergente
IFR	<i>Índice de Força Relativa</i>
MMS	Média Móvel Simples
MME	Média Móvel Exponencial
BB	Bandas de Bollinger
ROC	<i>Rate of Change</i>
OBV	Saldo de Volume
RMSE	<i>Erro Médio Mínimo Quadrático</i>



## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>19</b>
1.1	Contextualização, Motivação e Lacunas	19
1.2	Hipótesis e Objetivos	19
1.3	Organização do texto	19
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>21</b>
2.1	Mercado Financeiro Brasileiro	21
2.2	Contextualização sobre candlestick	22
2.2.1	O que são candlesticks e suas vantagens	23
<b>3</b>	<b>TRABALHOS RELACIONADOS</b>	<b>25</b>
<b>4</b>	<b>DADOS E MÉTODOS</b>	<b>27</b>
4.1	Dados	27
4.2	Métodos	27
<b>5</b>	<b>AVALIAÇÃO EXPERIMENTAL</b>	<b>31</b>
5.1	Conjuntos de Dados	31
5.2	Configuração Experimental	31
5.2.1	Pré Processamento dos dados	31
5.2.2	Extração de características	31
5.2.3	Redução de Dimensionalidade com PCA	32
5.2.4	Seleção de amostras de Testes e Treinamento	32
5.3	Resultados e Discussões	32
5.4	Dificuldades e Desafios	34
5.4.0.1	Dificuldades	34
5.4.0.2	Desafios Futuros	35
<b>6</b>	<b>CONCLUSÕES</b>	<b>37</b>
6.1	Desenvolvimento dos modelos	37
6.1.1	Melhores resultados	37
6.1.2	Melhor algoritmo → LGBM	37
	<b>Referências</b>	<b>39</b>



## 1 INTRODUÇÃO

### 1.1 Contextualização, Motivação e Lacunas

Qual a melhor oportunidade de entrada em uma operação de investimento na bolsa de valores Bovespa B3, quais são os melhores pontos de compra e venda levando em conta a estratégia *Swing Trading (SWT)* (EXAME, 2022); como o investimento é uma aposta em um preço futuro é necessário ter pontos de entrada e saída visando o melhor rendimento e também saídas definidas para minimizar as perdas ‘stops’, caso este preço não caminhe no sentido avaliado.

A estratégia de (**SWT**) visa capturar movimentos de preços no curto prazo, geralmente de alguns dias a algumas semanas (LIOU; HUANG, 2023), (EXAME, 2022) nela basicamente a ideia é comprar um ativo com preço baixo e vender mais caro com lucro quando se identifica uma tendência de alta ou ao contrário quando é de baixa; para isso, é necessário identificar pontos de entrada e saída o mais otimizado possível, bem como definir saídas deste posição para limitar as perdas em caso desta aposta não caminhar conforme o previsto.

Uma tarefa complicada principalmente para alguém sem conhecimento técnico, experiência ou mesmo tempo para analisar a grande quantidade de informações necessárias na tomada de decisão que mesmo sendo seguidas não garante o sucesso, pois o mercado acionário se comporta muitas vezes de forma aparentemente aleatória e não linear, impactado por diversos fatores sejam políticos, econômicos, sociais dentre muitos outros.

### 1.2 Hipótesis e Objetivos

Este trabalho tem por objetivo usar aprendizado de máquina, usando especificamente os algoritmos SVM, XGB, GaussianNB, KNN, SGClassifier e LGBM, para verificar dentre eles quais são os melhores para simular o conhecimento técnico de um operador na bolsa de valores identificando pontos de compra e venda neste mercado, os dados utilizados são índices WTI, algumas das moedas mais comuns (dollar, euro, yene e libra esterlina) índices econômicos mais relevantes no mercado financeiro, índice da Bovespa e de ações que compõem o IBX100 e moedas.

### 1.3 Organização do texto

Este trabalho está estruturado em cinco capítulos além desta introdução, conforme descrito a seguir.

No capítulo dois são apresentadas uma revisão bibliográfica sobre as áreas que

essa pesquisa envolve: aquisição e pré-processamento de dados, extração de características (indicadores técnicos dos índices), séries temporais e suas decomposições, aprendizado de máquina, aprendizado profundo e a abordagem por redes neurais. No capítulo três são descritos alguns dos trabalhos que relacionados a proposto desta estudo. No capítulo quatro são apresentadas os dados e as metodologias empregadas no desenvolvimento desta pesquisa: a formulação do problema, o fluxo de trabalho adotado, os detalhes de cada uma das etapas do fluxo. No capítulo cinco são apresentados os resultados obtidos através da avaliação experimental, um estudo de caso específico e as dificuldades e desafios observados ao longo do desenvolvimento deste projeto. No capítulo seis são apresentadas as conclusões e as possibilidades de contribuições e trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 Mercado Financeiro Brasileiro

O mercado financeiro de ações no Brasil é um ambiente dinâmico e atrativo e recentemente tem ganhado destaque como uma alternativa lucrativa de investimento. No entanto, é crucial estar ciente, além do conhecido risco, dos custos envolvidos, como taxas de corretagem, custódia, emolumentos e impostos. A Comissão de Valores Mobiliários (CVM) supervisiona rigorosamente este mercado (FILHO, 2020), estabelecendo regras e normas para garantir a transparência e segurança das operações do mercado de ações secundárias.

Para participar, os investidores devem abrir uma conta em uma corretora autorizada, preencher um questionário de perfil de investidor e assinar um termo de adesão e ciência de riscos. É fundamental que os investidores estejam bem informados sobre o funcionamento do mercado, os fundamentos das empresas em que pretendem investir e os riscos envolvidos. Em resumo, o mercado de ações oferece oportunidades interessantes, mas requer conhecimento, disciplina e cautela.

Para tentar prever o valor futuro de um ativo, existem duas abordagens principais que são as análises fundamentalista e técnica, esta última também conhecida como análise gráfica. (NTI; ADEKOYA; WEYORI, 2020), (MARKWAT *et al.*, 2008) (LYRIO *et al.*, 2015).

1. A primeira é baseada na crença de que os preços das ações refletem os fundamentos da empresa, expectativas de mercado interno e externo, taxa cambial dentre outros. Assim, dados financeiros e econômicos são coletados e usados para estimar preços futuros.
2. A análise técnica/gráfica, ao contrário, pressupõe que todas as informações necessárias sobre o futuro dos preços das ações podem ser encontradas nos valores do passado. Assim, dados históricos sobre os preços devem ser analisados, pois o preço desconta tudo, nele são refletidas tendências que devem ser aproveitadas até que novas evidências mostrem o contrário.

Como este estudo é focado na estratégia de **SWT**, a análise fundamentalista não será abordada, portanto, o foco se dará na análise técnica que é baseada em três princípios: o mercado desconta tudo, os preços se movem em tendências e a história se repete ELLIOT WAVES (LIOU; HUANG, 2023) (EXAME, 2022) (MATSURA, 2017). Assim, sob a hipótese de que os preços refletem todas as informações disponíveis e, portanto, esta será a única variável de análise que será utilizada. Esta estratégia pressupõe que os preços

se movimentam em tendências, que podem ser de alta, baixa ou laterais, e também que a história se repete, ou seja, padrões de preços que ocorreram no passado tendem a se repetir no futuro. Na sequência, são listadas as vantagens e desvantagens dessa estratégia conforme os argumentos citados acima.

### Vantagens do (SWT)

- Permite aproveitar o fluxo natural dos mercados financeiros, aumentando seus retornos ao ganhar dinheiro no sobe e desce do mercado.
- Ao entrar e sair dos mercados, o indivíduo pode identificar mais oportunidades. Isso permite que diversificar o risco e comprometa menos capital.
- Os stops; *ordem de compra ou venda automática no caso de um determinado preço ser alcançado pela cotação de um ativo no mercado*; são geralmente menores do que em negociações de longo prazo. Isso permite que você coloque posições maiores
- Possui limites claros. O swing trader é um método mais técnico e normalmente tem uma área específica que considera como um sinal de que a negociação está funcionando contra ou a favor dela.

### Desvantagens do (SWT)

- O investidor pode ser frequentemente contrariado nas suas escolhas, simplesmente porque o mercado mostra suporte ou resistência em uma determinada área, não significa que eles vão respeitá-lo hoje.
- O investidor precisa ter bons conhecimentos em análise técnica. Embora não seja necessariamente uma “desvantagem”, isso significa trabalho extra.
- Requer uma mentalidade única em comparação com a negociação de longo prazo e mais nervos, pois arrisca ser stopado à medida que os recuos nessas faixas menores parecem ser mais violentos do que para alguém observando um gráfico semanal. Isso é uma questão psicológica e algo que a maioria dos operadores eventualmente lidará durante suas operações.

## 2.2 Contextualização sobre candlestick

Por ser uma ótima ferramenta SWT na identificação de padrões, análise de tendências, confirmação de sinais e gestão de risco, será utilizada como features binárias dos dados, identificando em cada dia os sinais de padrões como martelo, estrela-cadente, engolfo de alta e engolfo de baixa como exemplificado na figura 1. “Candlestick” é uma técnica amplamente utilizada na Análise Técnica para avaliar a movimentação dos preços de ativos (ZARGHAM; SAYEH, 1999; SANGSAWAD; FUNG, 2017; NG *et al.*, 2011; OHKAWA *et al.*, 2008; CHEN *et al.*, 2009), especialmente em operações de Day Trade e SWT. Esses gráficos são formados por candles também conhecidos como velas japonesas, que representam o preço de abertura, fechamento, além dos patamares máximos e mínimos

em um determinado período. Abaixo seguem algumas informações essenciais sobre os candlesticks e como eles podem ser aplicados no contexto de investimento em ações:

### 2.2.1 O que são candlesticks e suas vantagens

- Os candles permitem aos traders identificar potenciais pontos de entrada e saída em uma operação.
- Eles oferecem uma leitura mais precisa do mercado em comparação com gráficos de linhas.
- Cada candle representa um período específico (minutos, horas, dias, semanas etc.) e está no mesmo intervalo de tempo escolhido.
- Marcação de pontos de suporte e resistência.
- Leitura precisa de padrões gráficos de topos e fundos.
- Ajuste adequado dos stops gain e loss para controle de risco.
- Avaliação do sentimento dos investidores em relação à compra ou venda.
- Identificação de padrões que podem indicar reversão de tendência.
- A cor das velas (verde/vermelho) indica alta ou baixa.
- Os preços de abertura e fechamento são determinados pelas extremidades do corpo do candle.
- Os preços máximos e mínimos são marcados pelas extremidades das sombras.
- Cada candle representa um período específico.
- Existem diversos padrões de candlestick, cada um com sua funcionalidade.
- Alguns padrões comuns incluem o Martelo, Estrela da Manhã, Engolfo de Alta/Baixa, Harami, entre outros.

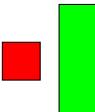
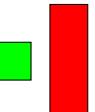
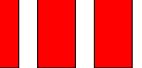
Martelo	Engolfo de Alta	Engolfo de Baixa	Harami
			
Estrela Cadente	Doji	Martelo Invertido	Três Corvos Negros
			

Tabela 1 – Padrões Candlestick



### 3 TRABALHOS RELACIONADOS

Muitos trabalhos têm sido desenvolvidos para prever o comportamento do mercado financeiro, utilizando técnicas de aprendizado de máquina e inteligência artificial. Boa parte usando Redes Neurais Recorrentes (RNN) (de Oliveira; NOBRE; ZáRATE, 2013)(LIOU; HUANG, 2023) (MARTINEZ *et al.*, 2009) LSTM (Long Short-Term Memory) para prever o preço das ações. Outros trabalhos utilizam técnicas de processamento de linguagem natural para analisar notícias e sentimentos do mercado. Além disso, há outros trabalhos que utilizam técnicas de processamento de imagens para analisar gráficos de ações e identificar padrões de comportamento. (ANGHINONI *et al.*, 2018)

Em (ANGHINONI *et al.*, 2018), o autor descreve a aplicação de Redes Neurais Artificiais (RNA) para prever o preço de fechamento das ações (PETR4) negociadas na BM&F BOVESPA. O estudo utilizou séries temporais de cotações históricas, indicadores técnicos e macroeconômicas. O método direto foi empregado para prever os preços de fechamento em horizontes de 1, 5, 15, 22, 37, 44, 66 e 110 dias. Neste artigo, o método envolve a inserção de uma série de dados na rede neural para obter uma previsão de preço futuro. O objetivo deste artigo é aplicar a RNA para prever o preço de fechamento das ações negociadas na BM&F BOVESPA em diversos horizontes padronizados, e as tendências de comportamento futuro dentro desses horizontes. Foram utilizadas séries temporais de cotações históricas de preços de ações, indicadores técnicos e séries macroeconômicas, utilizando como estudo de caso a ação (PETR4). Os resultados mostram que o método direto com horizontes de 1, 5, 15, 22, 37, 44, 66 e 110 dias de cotações segue o comportamento e tendência da ação.

O Propósito em outro estudo (NETO; CALVALCANTI; REN, 2009) foi o de analisar e prever o comportamento das ações da (PETR4) no mercado acionário brasileiro (BOVESPA), utilizando séries temporais e exógenas correlacionadas. Este estudo propõe uma nova métrica de desempenho, a Soma das Perdas e Ganhos (SLG), para avaliar a eficácia das técnicas de predição no contexto de investimentos financeiros. Nele é discutido a aplicação de (RNA) para modelar e prever séries temporais, com foco no mercado de ações ele se concentra na previsão das ações da (PETR4), uma vez que elas têm um forte impacto na (BOVESPA). A análise foi realizada em uma série temporal de dez anos e também considerou outras séries que se correlacionam com a série como exógenas no procedimento de previsão.



## 4 DADOS E MÉTODOS

### 4.1 Dados

Neste estudo serão utilizadas bases de dados do índice IBX100 ((IBRX100..., ) que incluí 100 acoes que obedecem aos critérios estabelecidos deste índice, serão utilizados também o índices Ibovespa, SELIC, IGPM, IPCA e INPC, WEST TEXAS INTERMEDIATE - **WTI** referências usadas para estabelecer o preço do óleo no mercado internacional além das moedas euro, dollar, iene, libra esterlina.

Desta base serão extraídas características de analise técnica como indicadores IFR 1 "Indice de Forca Relativa", MACD 2 entre outros, estes dados serão rotulados manualmente como compra e venda e nenhum dos dois conforme as figuras 5, 6, 7, 8 e 9 que mostram estes dados em algumas das empresas que compõe o IBX100, nelas estão ilustradas o candles OHLC diários utilizando uma derivação gráfica em heikin-ashi, formula na figura 4.

### 4.2 Métodos

Nas figuras 1, 2 e 3 são mostradas algumas das formulas de analise técnica mais comuns utilizadas no mercado (MATSURA, 2017) e também neste estudo, formulas que mostram tendencias de preços e suas inversões ou mesmo possíveis pontos de máxima ou mínima dentro de ciclo temporal de preços.

$$\text{IFR} = 100 - \frac{100}{1 + \frac{\text{Média dos Ganhos}}{\text{Média das Perdas}}}$$

Figura 1 – Índice de Força Relativa

$$\text{MACD} = \text{EMA}_{\text{curto prazo}} - \text{EMA}_{\text{longo prazo}}$$

Figura 2 – Convergência e Divergência das Médias Móveis

$$\text{MMA}_n = \frac{P_1 + P_2 + \cdots + P_n}{n}$$

$$\text{MME}_t = \text{MME}_{t-1} + \alpha(P_t - \text{MME}_{t-1})$$

Figura 3 – Médias Móveis, MMA e MME

$$\begin{aligned}\text{HA\_Open} &= \frac{\text{HA\_Open}_{\text{anterior}} + \text{HA\_Close}_{\text{anterior}}}{2} \\ \text{HA\_Close} &= \frac{\text{Open} + \text{High} + \text{Low} + \text{Close}}{4} \\ \text{HA\_High} &= \max(\text{High}, \text{HA\_Open}, \text{HA\_Close}) \\ \text{HA\_Low} &= \min(\text{Low}, \text{HA\_Open}, \text{HA\_Close})\end{aligned}$$

Figura 4 – Fórmula Heikin-Ashi

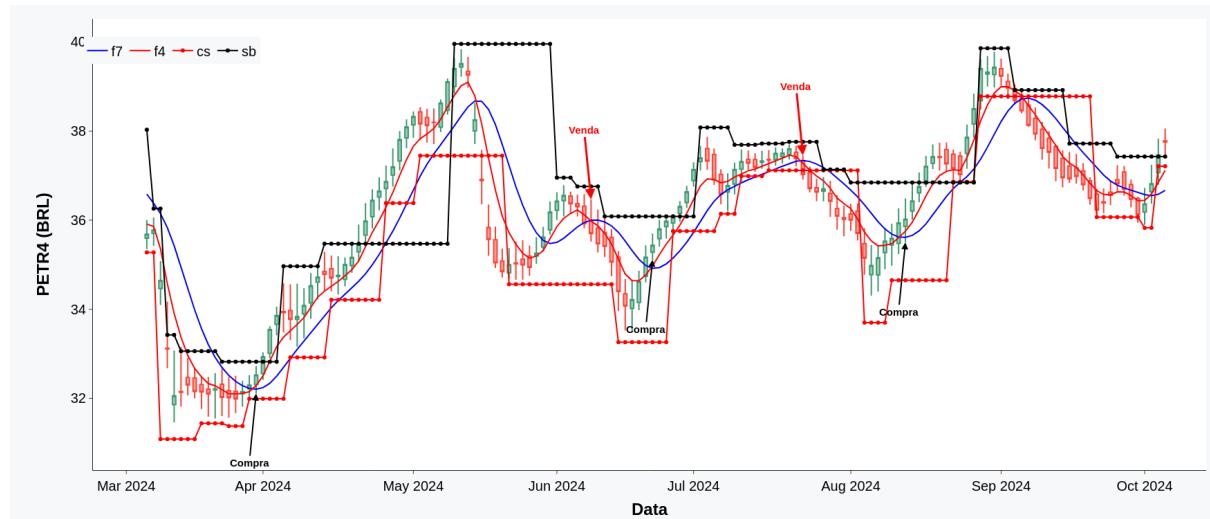


Figura 5 – Petrobras: Indicadores

Fonte: Elaborada pelo autor.

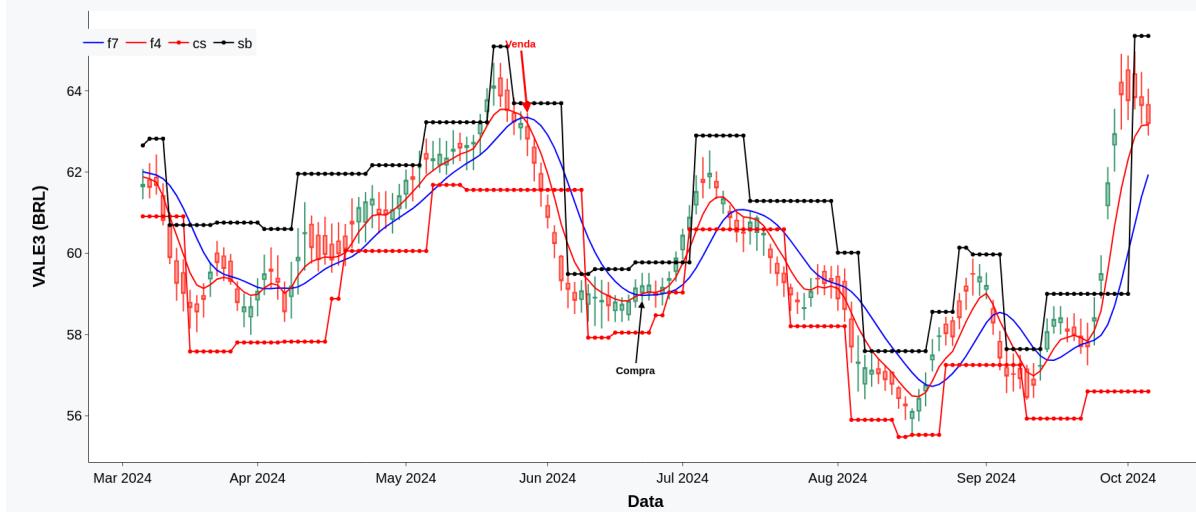


Figura 6 – Vale: Indicadores

Fonte: Elaborada pelo autor.

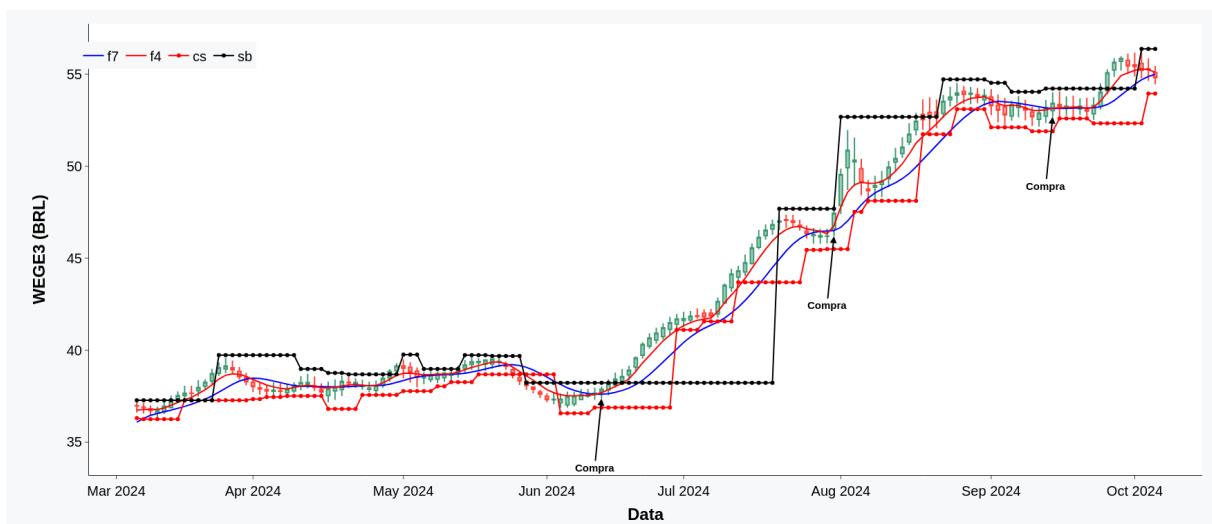


Figura 7 – Weg S.A.: Indicadores

Fonte: Elaborada pelo autor.

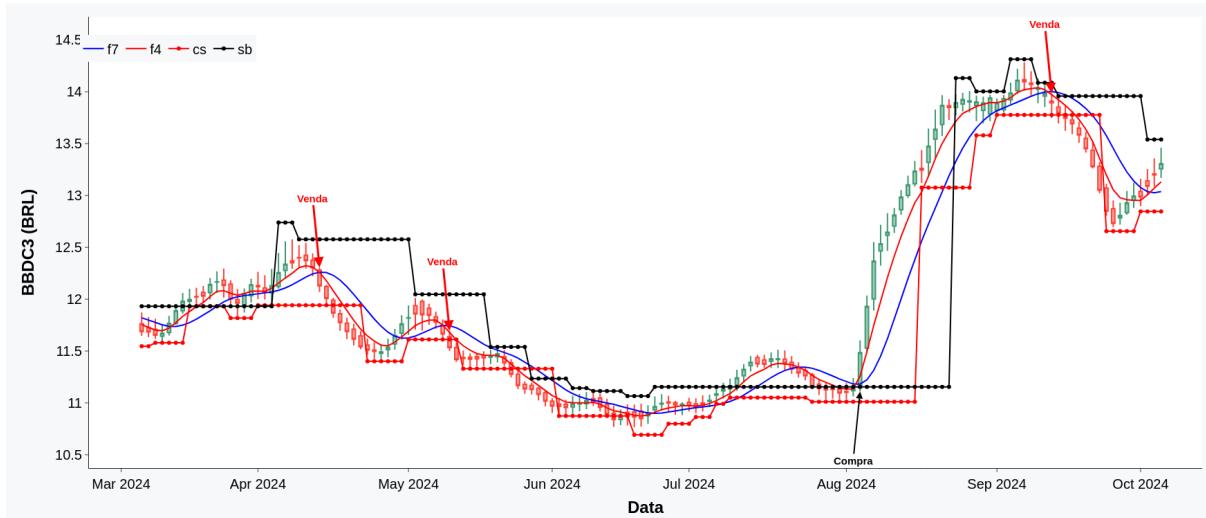


Figura 8 – Bradesco ON: Indicadores

Fonte: Elaborada pelo autor.

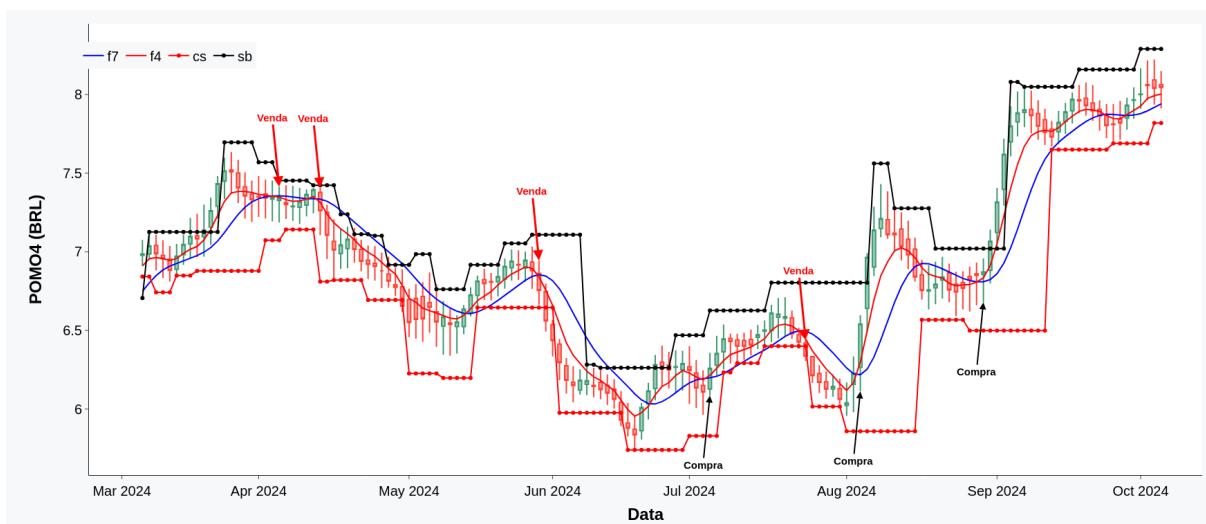


Figura 9 – BRF Brasil Foods: Indicadores

Fonte: Elaborada pelo autor.

## 5 AVALIAÇÃO EXPERIMENTAL

### 5.1 Conjuntos de Dados

Os dados históricos IBrX-100 e da Ibovespa foram obtidos a partir do site (QUOTEBR, ), neste conjunto de dados existem informações diárias desde 1986, estes dados consistem em abertura, máximo, mínimo, fechamento, volume e contrato abertos, neste estudo eles serão limitados entre 01/08/2014 e 01/08/2024 ou seja 10 anos. Foi coletado também através da biblioteca disponibilizada em (MOEDAS-DOCUMENTACAO..., ) outras informações tais como índices econômicos de referência são eles: selic, inpc, ipca e IG; moedas como o dollar, yene, euro, libra esterlina e também do petróleo cru (wti) (DIVYAA; SARAVANANE; KUMAR, 2024).

Tabela 2 – Conjuntos de Dados

Dataset Completo	Dataset de Treinamento	Testing Dataset
01/01/2014 - 01/01/2024	01/08/2014 - 01/08/2024	01/01/2022 - 01/01/2024

Fonte: Elaborada pelo autor.

### 5.2 Configuração Experimental

#### 5.2.1 Pré Processamento dos dados

Nos dados referentes a cotações IBrX-100, Ibovespa, Euro, Dollar, Iene, Libra Esterlina e (wti) não foram necessários tratamentos referentes a limpeza ou tratamento de dados corrompidos, faltantes ou inconsistentes uma vez que a fonte dos mesmos são confiáveis. De qualquer forma como toda fonte de dados é passível de erros e falhas os mesmos foram verificados quanto aos problemas, remoção de valores fora da curva (outliers) foi descartada devido a natureza dos dados trabalhados. Um exemplo outliers pode ser observado na figura 10 em que as variações diárias nos primeiros meses de 2020 "Explosão em casos de covid no mundo" extrapolam a média observada.

Dados dos índices econômicos precisaram de tratamento quanto a inconsistência dados valores e informações faltantes, principalmente devido a natureza temporal diária dos dados que foi trabalhado e estes índices estarem em outra escala temporal "semanal, mensal".

#### 5.2.2 Extração de características

Na extração de características foi utilizado indicadores técnicos, padrões de velas que são amplamente utilizados na análise técnica de ações, na tabela ?? são referenciadas alguns destes indicadores, são formulados bem conhecidas de uma maneira geral e podem

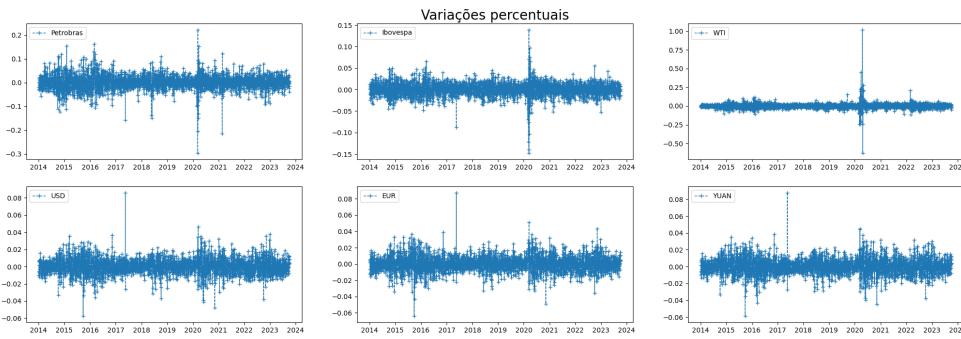


Figura 10 – Variação percentual diária

Fonte: Elaborada pelo autor.

ser encontradas em sites com o (WIKIPEDIA..., ) e disponíveis na rede, ao todo foram extraídas noventa e oito características de cada empresa, índice ou moeda utilizada no estudo. Nas figuras 5, 6, 7, 8 e 9 são demonstrados algumas destas características extraídas.

### 5.2.3 Redução de Dimensionalidade com PCA

Para reduzir a dimensionalidade dos dados, foi utilizado o PCA (Análise de Componentes Principais), que é uma técnica estatística essencial para a redução de dimensionalidade em conjuntos de dados. O PCA transforma variáveis correlacionadas em um conjunto não correlacionado chamado de componentes principais. Este método é útil quando há muitas variáveis correlacionadas que dificultam a análise dos dados, neste estudo a porcentagem PCA explicada foi de 95%.

### 5.2.4 Seleção de amostras de Testes e Treinamento

Dados dos índices econômicos precisaram de tratamento quanto a inconsistência dos dados, valores e informações faltantes, principalmente devido à natureza temporal diária dos dados que foi trabalhado e estes índices estarem em outra escala temporal "semanal, mensal".

## 5.3 Resultados e Discussões

De acordo com os experimentos realizados todos os algorítimos tiveram pouca mudança significativa no resultados diante da alteração nos parâmetros porem houve aumento significativo no tempo de processamento do algoritmo XGB, conforme se altera o tempo de processamento aumenta chegando a um tempo de 25x maior que com os parâmetros padrão. Levando se em conta o algoritmo XGB e LGBM praticamente ficaram com os mesmos resultados a escolha para o trabalho ficou com o LGBM que com menos de 20% do tempo, conforme 11 e 10, foi obtido o mesmo resultado.

Algoritmo SVM				
Parâmetro	Precisão	Recall	Roc_auc	Tempo
<b>Kernel rbf, gamma auto</b>	<b>0.866545</b>	<b>0.907341</b>	<b>0.939837</b>	<b>17.752407</b>
Kernel linear	0.828154	0.829122	0.894103	44.681905
Kernel sigmoide	0.731261	0.725632	0.797604	11.888488
Kernel polinomial grau 2	0.832419	0.969916	0.915237	13.117167
Kernel polinomial grau 3	0.787934	0.880866	0.875789	15.950303

Tabela 3 – Desempenho do Algoritmo SVM com diferentes configurações de Kernel

Algoritmo XGB				
Parâmetro	Precisão	Recall	Roc_auc	Tempo
default	0.952468	0.972323	0.984041	1.339032
n_estimators=1000, max_depth=10 e learn...	0.950030	0.973526	0.984334	25.190686
<b>n_estimators=1000 e max_depth=10</b>	<b>0.946374</b>	<b>0.968712</b>	<b>0.984893</b>	<b>6.805423</b>
n_estimators=1000, max_depth=10, learn...	0.948202	0.967509	0.984136	25.112977
n_estimators=1000	0.950030	0.971119	0.984588	6.388322

Tabela 4 – Algoritmo XGB com diferentes parâmetros

Algoritmo GaussianNB				
Parâmetro	Precisão	Recall	Roc_auc	Tempo
smoothing 1e-9	0.720293	0.953069	0.747973	0.024071
smoothing 1e-8	0.724558	0.947052	0.755559	0.020411
smoothing 1e-7	0.730652	0.937425	0.762979	0.019771
smoothing 1e-6	0.732480	0.931408	0.771159	0.019659
<b>smoothing 1e-5</b>	<b>0.733090</b>	<b>0.919374</b>	<b>0.775864</b>	<b>0.019790</b>

Tabela 5 – Desempenho do Algoritmo GaussianNB com diferentes Var smoothing

Algoritmo KNN				
Parâmetro	Precisão	Recall	Roc_auc	Tempo
K=1	0.764168	0.811071	0.763560	0.154772
K=3	0.792809	0.858002	0.849996	0.139704
K=5	0.798903	0.871239	0.875221	0.137880
K=7	0.811700	0.888087	0.886927	0.142555
<b>K=9</b>	<b>0.813528</b>	<b>0.892900</b>	<b>0.895060</b>	<b>0.143039</b>

Tabela 6 – Algoritmo KNN com diferentes valores de K

<b>Algoritmo SGDClassifier</b>				
<b>Parâmetro</b>	<b>Precisão</b>	<b>Recall</b>	<b>Roc_auc</b>	<b>Tempo</b>
sgd_squared_error_01	0.787325	0.771360	0.868348	0.110788
sgd_squared_error_05	0.819013	0.809868	0.887020	0.179679
<b>sgd_squared_error_10</b>	<b>0.819622</b>	<b>0.813478</b>	<b>0.887662</b>	<b>0.203087</b>
sgd_squared_error_20	0.815966	0.814681	0.884736	0.170533
sgd_squared_error_50	0.809263	0.803851	0.880358	0.115501

Tabela 7 – Algoritmo SGDClassifier com diferentes configurações

<b>Algoritmo LGBM</b>				
<b>Parâmetro</b>	<b>Precisão</b>	<b>Recall</b>	<b>Roc_auc</b>	<b>Tempo</b>
Max depth=4	0.949421	0.972323	0.986184	0.424455
Max depth=8	0.947593	0.969916	0.985435	0.663882
<b>Max depth=16</b>	<b>0.947593</b>	<b>0.968712</b>	<b>0.986493</b>	<b>0.897113</b>
Max depth=32	0.944546	0.965102	0.986081	0.654769
Max depth=64	0.944546	0.965102	0.986081	0.748074

Tabela 8 – Algoritmo LGBM com diferentes configurações

<b>Algoritmo MLP</b>				
<b>Parâmetro</b>	<b>Precisão</b>	<b>Recall</b>	<b>Roc_auc</b>	<b>Tempo</b>
1 camada	0.895349	0.926594	0.965313	3.331924
2 camadas	0.907692	0.922984	0.963477	3.625266
3 camadas	0.881490	0.939832	0.958656	3.441999
4 camadas	0.903846	0.904934	0.961727	3.713768
<b>5 camadas</b>	<b>0.906619</b>	<b>0.922984</b>	<b>0.966791</b>	<b>4.593906</b>

Tabela 9 – Algoritmo MLP com diferentes configurações

<b>Melhores Algorítimos</b>				
<b>Parâmetro</b>	<b>Precisão</b>	<b>Recall</b>	<b>Roc_auc</b>	<b>Tempo</b>
SVM = Kernel rbf, gamma auto	0.866545	0.907341	0.939837	12.679423
XGB = XGB com n_estimators=1000 e max_depth=10	0.946374	0.968712	0.984893	6.805423
GaussianNB = Var smoothing 1e-5	0.733090	0.919374	0.775864	0.019790
KNN = K=9	0.813528	0.892900	0.895060	0.143039
SGDClassifier = sgd_squared_error_10	0.819622	0.813478	0.887662	0.203087
<b>LGBM = Max depth=16</b>	<b>0.947593</b>	<b>0.968712</b>	<b>0.986493</b>	<b>0.897113</b>
MLP = MLP 5 camadas	0.906619	0.922984	0.966791	4.593906

Tabela 10 – Melhores algoritmos de acordo com os parâmetros de precisão, recall e roc\_auc.

## 5.4 Dificuldades e Desafios

### 5.4.0.1 Dificuldades

Dentre as dificuldades encontradas vale destacar que as variações de preços muito fora da média, muitas vezes muito altas, são exemplos de problema difíceis para mapear,

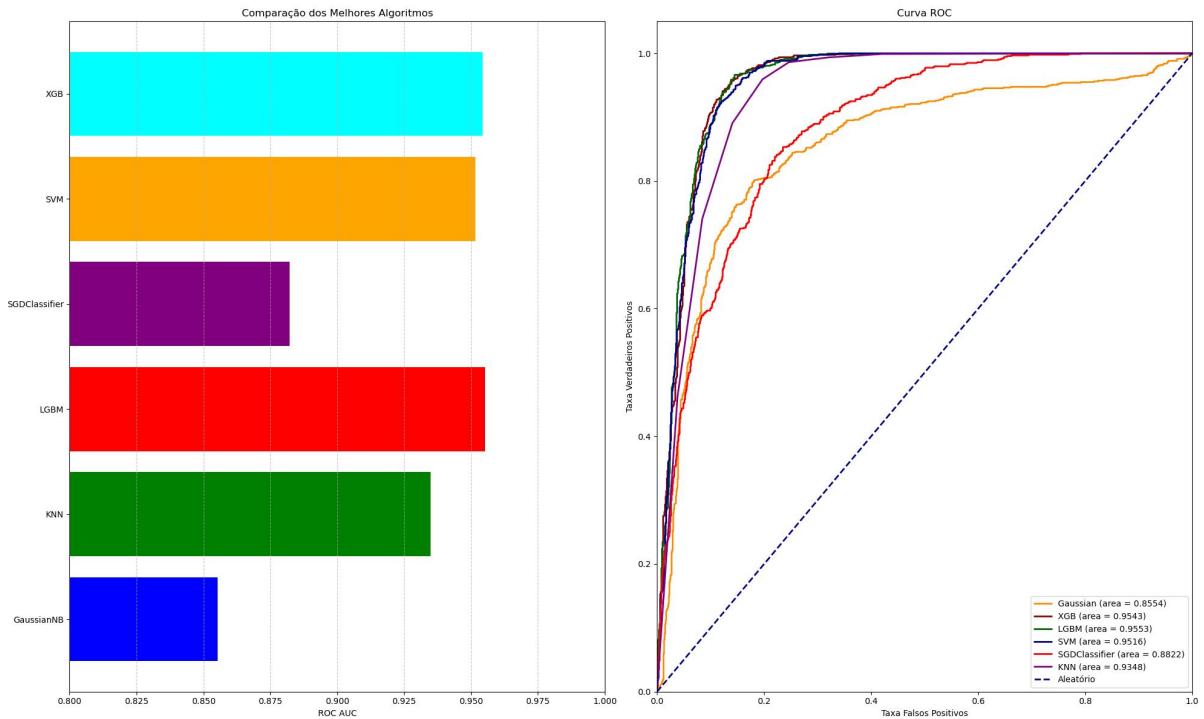


Figura 11 – Resultados dos algorítimos

Fonte: Elaborada pelo autor.

duas que vale destacar são MMX Mineração (MMXM3) e Americanas (AMER3) apenas para ficar nos casos mais recentes são casos de empresas que participarem em algum momento deste índice IBRX100 e que faliram ou deram grandes prejuízos aos seus acionistas compradores, eram empresas consideradas como bom investimento pelo mercado durante bom tempo. Neste caso somente um gerenciamento de risco bem utilizado pode minimizar as virtuais perdas.

#### 5.4.0.2 Desafios Futuros

Como o projeto teve como objetivo de simular a experiência de operações no mercado acionário baseado na sabedoria de um operador SWT, que observa múltiplas oportunidades e decisões em um mesmo dia ou semana, seria interessante no ponto de vista de um iniciante no mercado replicar este conhecimento, afim de diminuir sua taxa de erros e assim aumentar a lucratividade de suas operações. Se a mesma indicação fosse seguida por um outro valor que indicasse uma temperatura de quão boa seria aquela oportunidade em termos de lucratividade as decisões se tornariam mais controladas do ponto de vista de lucro ou perda, esta nova informação poderia extraída do conhecimento aprendido sobre a volatilidade daquela ação específica ou mesmo dos volumes negociados no mercado futuro da mesma naquele momento.

Em trabalhos futuros estudos levando em conta dados estruturados em tempo real

ou mesmo dados não estruturados como textos técnicos sobre o mercado de ações em blogs, redes sociais podem enriquecer o estudo.

## 6 CONCLUSÕES

### 6.1 Desenvolvimento dos modelos

#### 6.1.1 Melhores resultados

Dentre os algorítimos utilizados neste trabalho Gaussian NB, KNN, LGBM, SGD-Classifier, SVM e XGB cada um com múltiplas parametrizações distintas foram separadas as melhores para cada um deles e depois comparadas entre eles. Conforme dados demonstrados na 10 o melhor classificador encontrado para a tarefa foi o XGB com os parâmetros "n\_estimators=1000, max\_depth = 10 com learning\_rate = 0.01 e gamma=0.1" porém o tempo de execução do algoritmo é de quase 6 vezes o tempo executado no LGBM com melhores parametrizações e com resultado praticamente iguais.

#### 6.1.2 Melhor algorítimo → LGBM

De acordo com a tabelas 10 e 11 fica bem destacado quando se visualiza a curva ROC do algorítimo LGBM 97.8%, precisão de 94.4% e recall de 97,1% mostrou que o algoritmo se sobrepõe dentre os demais estudados, tanto em performance quanto em resultados.

Os resultados obtidos por este estudo, que pretende verificar se a possibilidade de uso de machine learning para simular o conhecimento técnico em operações SWT, mostrou através destes estudos que existe bons resultados para estas simulações. Vale ressaltar que mesmo o modelo melhor classificado nos testes não se pode deve adotar posições financeiras reais baseadas apenas no mesmo, sob risco de perdas significativas visto que o mercado de acionário é volátil e perdas percentuais expressivas podem se acumular em questão de horas.



## REFERÊNCIAS

- ANGHINONI, L. *et al.* Time series trend detection and forecasting using complex network topology analysis. In: **2018 International Joint Conference on Neural Networks (IJCNN)**. [S.l.: s.n.], 2018. p. 1–7. ISSN 2161-4407.
- CHEH, Y. *et al.* Constructing portfolio investment strategy based on time adapting genetic network programming. In: **2009 IEEE Congress on Evolutionary Computation**. [S.l.: s.n.], 2009. p. 2379–2386.
- de Oliveira, F. A.; NOBRE, C. N.; ZáRATE, L. E. Applying artificial neural networks to prediction of stock price and improvement of the directional prediction index – case study of petr4, petrobras, brazil. **Expert Systems with Applications**, v. 40, n. 18, p. 7596–7606, 2013. ISSN 0957-4174. Available at: <<https://www.sciencedirect.com/science/article/pii/S0957417413004703>>.
- DIVYAA, S.; SARAVANANE, R.; KUMAR, M. Predicting macro-economic factors of wti crude oil using monte carlo simulation. In: **2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)**. [S.l.: s.n.], 2024. p. 1–7.
- EXAME, R. **O que é swing trading e como funciona essa modalidade?** 2022. <<https://exame.com/invest/guia/o-que-e-swing-trading-e-como-funciona-essa-modalidade>> [Accessed: 03/03/2024].
- FILHO, J. R. de T. **Mercado de capitais brasileiro: uma introdução**. [S.l.: s.n.]: Cengage Learning, 2020.
- (IBRX100 B3). <[https://www.b3.com.br/pt\\_br/market-data-e-indices/indices/indices-amplos/indice-brasil-100-ibrx-100.htm](https://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/indice-brasil-100-ibrx-100.htm)>. [Accessed 20-08-2024].
- LIOU, H.-I.; HUANG, K.-C. A deep learning model for stock price prediction in swing trading. In: **2023 9th International Conference on Applied System Innovation (ICASI)**. [S.l.: s.n.], 2023. p. 154–156.
- LYRIO, M. V. L. *et al.* Análise da implementação de uma estratégia de investimento em ações baseada em um instrumento de apoio à decisão. **Contaduría y administración**, Elsevier, v. 60, n. 1, p. 113–143, 2015.
- MARKWAT, T. D. *et al.* The economic value of fundamental and technical information in emerging currency markets. In: ERS-2007-096-F&A, EFA 2008 ATHENS MEETINGS PAPER. [S.l.: s.n.], 2008.
- MARTINEZ, L. C. *et al.* From an artificial neural network to a stock market day-trading system: A case study on the bm&f bovespa. In: **2009 International Joint Conference on Neural Networks**. [S.l.: s.n.], 2009. p. 2006–2013.
- MATSURA, E. K. **Comprar ou vender?: como investir na bolsa utilizando análise gráfica**. [S.l.: s.n.]: Saraiva Educação SA, 2017.

- MOEDAS-DOCUMENTACAO python-bcb — wilsonfreitas.github.io. <<https://wilsonfreitas.github.io/python-bcb/currency.html>>. [Accessed 26-08-2024].
- NETO, M. C. A.; CALVALCANTI, G. D. C.; REN, T. I. Financial time series prediction using exogenous series and combined neural networks. In: **2009 International Joint Conference on Neural Networks**. [S.l.: s.n.], 2009. p. 149–156. ISSN 2161-4407.
- NG, W. W. Y. *et al.* Stock investment decision support for hong kong market using rbfnn based candlestick models. In: **2011 International Conference on Machine Learning and Cybernetics**. [S.l.: s.n.], 2011. v. 2, p. 538–543.
- NTI, I. K.; ADEKOYA, A. F.; WEYORI, B. A. A systematic review of fundamental and technical analysis of stock market predictions. **Artificial Intelligence Review**, v. 53, n. 4, p. 3007–3057, 2020. ISSN 1573-7462. Available at: <<https://doi.org/10.1007/s10462-019-09754-z>>.
- OHKAWA, E. *et al.* Evaluation of varying portfolio construction of stocks using genetic network programming with control nodes. In: **2008 SICE Annual Conference**. [S.l.: s.n.], 2008. p. 1231–1236.
- QUOTEBR. <<https://www.quotebr.com/>>. [Accessed 16-08-2024].
- SANGSAWAD, S.; FUNG, C. C. Extracting significant features based on candlestick patterns using unsupervised approach. In: **2017 2nd International Conference on Information Technology (INCIT)**. [S.l.: s.n.], 2017. p. 1–5.
- WIKIPEDIA, the free encyclopedia — en.wikipedia.org. <[https://en.wikipedia.org/wiki/Main\\_Page](https://en.wikipedia.org/wiki/Main_Page)>. [Accessed 26-08-2024].
- ZARGHAM, M.; SAYEH, M. A web-based information system for stock selection and evaluation. In: **Proceedings of International Workshop on Advance Issues of E-Commerce and Web-Based Information Systems. (Cat. No.PR00334)**. [S.l.: s.n.], 1999. p. 81–83.