

---

**Técnicas de Inteligência Artificial Aplicadas ao  
Mercado Financeiro**

*Gabriel de Souza Rosa*

---





*Gabriel de Souza Rosa*

# **Técnicas de Inteligência Artificial Aplicadas ao Mercado Financeiro**

Monografia apresentada ao Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), como requisito parcial para obtenção do título de bacharel em Engenharia de Computação.

Orientador: Prof. Dr. Alisson Marques da Silva

Coorientador: *Prof<sup>a</sup>. Dr<sup>a</sup>.* Charlene Cássia de Resende



Divinópolis  
27-06-2023



# Lista de Figuras

---

1.1	Porcentagem de pessoas por país que investem em ações. . . . .	18
1.2	Dados da bolsa de valores brasileira. . . . .	19
2.1	Tendências de uma série temporal. . . . .	20
2.2	Características cíclicas e sazonais de uma série temporal. . . . .	21
2.3	Gráfico de velas e suas características observáveis. . . . .	22
2.4	Processo de desenvolvimento dos trabalhos analisados. . . . .	22
2.5	Matriz de confusão. . . . .	28
3.1	Visão macro do processo de construção do conjunto de dados. . . . .	32
3.2	Visão macro do processo de seleção de variáveis. . . . .	34
3.3	Visão macro do processo de previsão. . . . .	35
3.4	Visão macro do processo de recomendação de investimentos. . . . .	36
4.1	Base de dados do ativo financeiro PETR3. . . . .	38
4.2	Base de dados do ativo financeiro WINFUT. . . . .	39
4.3	Base de dados do ativo financeiro WDOFUT. . . . .	39



# Lista de Tabelas

---

2.1	Transformação da saídas no domínio contínuo para o domínio discreto. . . . .	27
2.2	Síntese dos trabalhos relacionados. . . . .	31
4.1	Resultados da Classificação (PETR3 30 minutos) . . . . .	41
4.2	Resultados das Compras e Variação Financeira (PETR3 30 minutos) . . . . .	41
4.3	Resultados da Classificação (PETR3 60 minutos) . . . . .	42
4.4	Resultados das Compras e Variação Financeira (PETR3 60 minutos) . . . . .	42
4.5	Resultados da Classificação (WINFUT 30 minutos) . . . . .	43
4.6	Resultados das Compras e Variação Financeira (WINFUT 30 minutos) . . . . .	43
4.7	Resultados da Classificação (WINFUT 60 minutos) . . . . .	44
4.8	Resultados das Compras e Variação Financeira (WINFUT 60 minutos) . . . . .	45
4.9	Resultados da Classificação (WDOFUT 30 minutos) . . . . .	45
4.10	Resultados das Compras e Variação Financeira(WDOFUT 30 minutos) . . . . .	46
4.11	Resultados da Classificação (WDOFUT 60 minutos) . . . . .	46
4.12	Resultados das Compras e Variação Financeira (WDOFUT 60 minutos) . . . . .	47
5.1	Hiperparâmetros dos Modelos . . . . .	49





# Lista de Acrônimos

---

<b>%D</b>	Média Móvel Simples do Oscilador Estocástico
<b>%K</b>	Oscilador Estocástico
<b>%R</b>	Indicador <i>Williams</i>
<b>AC</b>	Acurácia
<b>ADC</b>	<i>Average Decision Committee</i>
<b>ADX</b>	<i>Average Directional Index</i>
<b>AF</b>	Análise Fundamentalista
<b>AMC</b>	<i>Average and Majority Committee</i>
<b>ANFIS</b>	<i>Adaptive Neuro-Fuzzy Inference System</i>
<b>AQ</b>	Análise Quantitativa
<b>ARIMA</b>	<i>Autoregressive Integrated Moving Average</i>
<b>AS</b>	Análise de Sentimento
<b>AT</b>	Análise Técnica
<b>CCI</b>	<i>Commodity Channel Index</i>
<b>CNN</b>	<i>Convolutional Neural Network</i>
<b>DBN</b>	<i>Deep Belief Network</i>
<b>DCDNN</b>	<i>Dynamic Correlation Deep Neural Network</i>
<b>DCFS</b>	<i>Deep Convolutional Fuzzy System</i>
<b>DJIA</b>	<i>Dow Jones Industrial Average</i>
<b>DT</b>	<i>Decision Trees</i>
<b>eGNN</b>	<i>evolving Granular Neural Network</i>
<b>ELM</b>	<i>Extreme Learning Machine</i>
<b>EMA</b>	<i>Exponential Moving Average</i>
<b>eMG</b>	<i>evolving Multivariable Gaussian Fuzzy System</i>
<b>eOGS</b>	<i>evolving Optimal Granular System</i>
<b>ETFs</b>	<i>Exchange-Traded Funds</i>
<b>F1</b>	F1 score
<b>FN</b>	Falso Negativo
<b>FP</b>	Falso Positivo
<b>GARCH</b>	<i>Generalized Autoregressive Conditional Heteroskedasticity</i>
<b>IA</b>	Inteligência Artificial
<b>KNN</b>	<i>K-Nearest Neighbors</i>
<b>LR</b>	<i>Logistic Regression</i>
<b>LSTM</b>	<i>Long Short-Term Memory</i>
<b>MACD</b>	<i>Moving Average Convergence Divergence</i>
<b>MAE</b>	<i>Mean Absolute Error</i>
<b>MAPE</b>	<i>Mean Absolute Percent Error</i>
<b>MDC</b>	<i>Majority Decision Committee</i>
<b>ME</b>	<i>Mean Error</i>
<b>MLP</b>	<i>Multi-Layer Perceptron</i>
<b>MSE</b>	<i>Mean Squared Error</i>
<b>OHLC</b>	<i>Open, High, Low e Close</i>

<b>PN</b>	Precisão Negativa
<b>PP</b>	Precisão Positiva
<b>RBF</b>	<i>Radial Basis Function</i>
<b>RF</b>	<i>Random Forest</i>
<b>RIM</b>	Regressão de Informação Mútua
<b>RMSE</b>	<i>Root Mean Squared Error</i>
<b>RNN</b>	<i>Recurrent Neural Networks</i>
<b>ROC</b>	<i>Rate of Change</i>
<b>R<sup>2</sup></b>	Coefficiente de Determinação
<b>SARIMA</b>	<i>Seasonal Autoregressive Integrated Moving Average</i>
<b>SMA</b>	<i>Simple Moving Average</i>
<b>SVM</b>	<i>Support Vector Machine</i>
<b>SVR</b>	<i>Support Vector Regression</i>
<b>TCN</b>	<i>Temporal Convolutional Network</i>
<b>TSI</b>	<i>True Strength Index</i>
<b>VH</b>	Valores Históricos
<b>VN</b>	Verdadeiro Negativo
<b>VP</b>	Verdadeiro Positivo
<b>X<sup>2</sup></b>	<i>Chi-Squared</i>

# Sumário

---

<b>1</b>	<b>Introdução</b>	<b>16</b>
1.1	Contextualização . . . . .	16
1.2	Motivação e Relevância . . . . .	17
1.3	Objetivo . . . . .	19
1.3.1	Objetivo Geral . . . . .	19
1.3.2	Objetivos Específicos . . . . .	19
1.4	Organização do Trabalho . . . . .	19
<b>2</b>	<b>Trabalhos Relacionados</b>	<b>20</b>
2.1	Séries Temporais . . . . .	20
2.2	Mercado Financeiro . . . . .	21
2.3	Abordagens . . . . .	22
2.4	Conjuntos de Dados . . . . .	23
2.5	Geração de Variáveis . . . . .	23
2.6	Seleção de Variáveis . . . . .	25
2.7	Técnicas de Predição . . . . .	26
2.8	Estratégias . . . . .	27
2.9	Critérios de Avaliação . . . . .	28
2.10	Resumo dos Trabalhos Relacionados . . . . .	30
<b>3</b>	<b>Abordagem Proposta</b>	<b>32</b>
3.1	Construção dos Conjuntos de Dados . . . . .	32
3.1.1	Extração de Atributos . . . . .	32
3.1.2	Geração de Variáveis . . . . .	33
3.2	Seleção de Variáveis . . . . .	33
3.3	Máquina de Previsão . . . . .	34
3.4	Recomendação de Investimento . . . . .	35
<b>4</b>	<b>Experimentos Computacionais</b>	<b>37</b>
4.1	Conjunto de Dados . . . . .	37
4.1.1	PETR3 . . . . .	37
4.1.2	WINFUT . . . . .	38
4.1.3	WDOFUT . . . . .	38
4.2	Modelos e Hiperparâmetros . . . . .	40
4.3	Experimentos e Métricas . . . . .	40
4.4	Resultados Experimentais . . . . .	40
4.4.1	PETR3 (30 Minutos) . . . . .	40
4.4.2	PETR3 (60 Minutos) . . . . .	41
4.4.3	WINFUT (30 Minutos) . . . . .	43
4.4.4	WINFUT (60 Minutos) . . . . .	44
4.4.5	WDOFUT (30 Minutos) . . . . .	45
4.4.6	WDOFUT (60 Minutos) . . . . .	46



# Resumo

---

Este estudo propõe a aplicação de técnicas de inteligência artificial (IA) no âmbito do mercado financeiro, visando prever e recomendar estrategicamente investimentos. O trabalho inicia com uma revisão histórica dos mercados financeiros, abordando a teoria da eficiência de mercado, que postula que os preços refletem todas as informações disponíveis, destacando ainda a influência das emoções e comportamentos na formação dos preços das ações. Uma investigação sobre a eficiência de mercados emergentes revela oportunidades de investimento devido a desvios sistemáticos nos preços. A literatura é abordada para destacar a evolução de modelos de análise técnica, fundamentalista, quantitativa e de sentimentos. O estudo segue revisando diversas técnicas de IA aplicadas à previsão de ativos financeiros, incluindo modelos lineares e aqueles fundamentados em IA. A metodologia proposta abrange a extração, geração e seleção de variáveis, culminando na implementação de uma máquina de previsão que combina modelos estatísticos, de classificação e de regressão por meio de técnicas de *ensemble*. Os resultados encontrados indicam a viabilidade de alcançar retornos significativos por meio de técnicas de previsão, sendo a máquina de previsão comparada favoravelmente com a técnica de *buy and hold*. Essa abordagem integrativa revela-se promissora para orientar decisões financeiras mais informadas e estratégicas no mercado.

Palavras-chave: Previsão de ações; Mercado financeiro; Inteligência artificial; Séries temporais; Análise de dados financeiros; Ensemble.



# Abstract

---

This study proposes the application of artificial intelligence (AI) techniques in the realm of the financial market, aiming to strategically predict and recommend investments. The work begins with a historical review of financial markets, addressing the efficient market theory, which posits that prices reflect all available information, also highlighting the influence of emotions and behaviors in shaping stock prices. An investigation into the efficiency of emerging markets reveals investment opportunities due to systematic deviations in prices. Literature is examined to emphasize the evolution of technical, fundamental, quantitative, and sentiment analysis models. The study then reviews various AI techniques applied to forecasting financial assets, including linear models and those grounded in AI. The proposed methodology encompasses the extraction, generation, and selection of variables, culminating in the implementation of a prediction machine that combines statistical, classification, and regression models through ensemble techniques. The results suggest the feasibility of achieving significant returns through forecasting techniques, with the prediction machine favorably compared to the buy-and-hold strategy. This integrative approach proves promising for guiding more informed and strategic financial decisions in the market.

Keywords: Stock prediction; Financial market; Artificial intelligence; Time series; Financial data analysis; Ensemble.

---

# Introdução

---

Neste capítulo, apresenta-se uma descrição inicial do problema abordado, juntamente com os principais desafios e motivações para a realização desta pesquisa. Na Seção 1.1, contextualiza-se o cenário do mercado de ações e os acontecimentos que contribuíram para o seu desenvolvimento. Em seguida, na Seção 1.2, são apresentadas a motivação e a relevância do estudo. Na Seção 1.3, descreve-se o objetivo geral do trabalho, juntamente com seus objetivos específicos. Por fim, na Seção 1.4, é apresentada a estrutura deste trabalho, destacando como as seções estão organizadas.

## 1.1 Contextualização

Os primeiros registros de mercados financeiros similares aos contemporâneos, surgiram no século XVII na Holanda, com a criação da Bolsa de Valores de Amsterdam, onde comerciantes e investidores se reuniam para negociar ações de empresas da Companhia das Índias Orientais (FRENCH, 2006). Desde então, o mercado de ações evoluiu e se expandiu para outros países, tornando-se uma importante fonte de financiamento para empresas e um meio para os investidores obterem lucros.

Com a expansão e grande aceitação do modelo financeiro, iniciou-se o mercado de índices no final do século XIX nos Estados Unidos, quando Charles H. Dow, fundador de *The Wall Street Journal*, deu nome ao primeiro índice de ações (STILLMAN, 1986), conhecido como *Dow Jones Industrial Average* (DJIA). O DJIA foi criado em 1886 com o intuito de acompanhar a performance do mercado de ações e fornecer uma medida do desempenho das empresas, aplicando uma média ponderada dos preços das ações de 12 empresas industriais americanas (SHOVEN; SIALM, 2000). Em seguida, outros índices de ações foram criados em todo o mundo, como o FTSE 100 no Reino Unido, o Nikkei no Japão, o S&P 500 nos Estados Unidos e o Ibovespa no Brasil. Guiando para o que se conhece nos dias atuais, em que há diversos índices de ações que representam diferentes setores da economia, indústrias e países.

Ao longo da história, os índices de ações surgiram como uma forma de investimento, o que levou ao desenvolvimento do mercado de índices (BOGLE, 1999). Esse mercado permite que os investidores comprem e vendam um conjunto de ações representadas por um índice, em vez de adquirir ações individuais. Essa forma de investimento é conhecida como *Exchange-Traded Funds* (ETFs) e tem se tornado popular devido à diversificação, custos reduzidos e facilidade de negociação (BEN-DAVID; FRANZONI; MOUSSAWI, 2017). Os ETFs permitem que os investidores tenham exposição a um portfólio diversificado de ativos por meio de uma única transação, proporcionando benefícios em termos de redução de riscos e acesso a diferentes setores e mercados (GAD; ANDRIKOPOULOS, 2019).

A teoria financeira subjacente ao mercado de índices conhecida como eficiência de mercado, afirma que os preços dos ativos financeiros refletem todas as informações disponíveis publicamente (SAMUELSON, 2016). Isso significa que é improvável obter lucros anormais de forma consistente baseado apenas em informações públicas (MALKIEL, 2003). Tal ocorrência é devido a grande procura de informações por parte dos investidores para obter vantagem competitiva no mercado, o que leva a uma rápida incorporação de informações aos preços dos ativos financeiros (FAMA, 1970). No entanto, há críticas à teoria da eficiência de mercado, uma vez que existem evidências empíricas de que os preços das ações podem ser influenciados por fatores emocionais e comportamentais, como o comportamento de manada e a aversão à perda (SHILLER, 2000). Possibilitando a ineficiências de mercado e oportunidades de investimento para aqueles que são capazes de identificá-las.

Em um estudo realizado por Cajueiro e Tabak (2004), investigou-se os coeficientes de *Hurst* das séries de



preços de ações de países desenvolvidos e emergentes com o objetivo de encontrar correlações entre as variações passadas e os estados futuros. O coeficiente de *Hurst* é utilizado para medir a dependência de longo prazo em uma série temporal, onde um valor igual a 0,5 indica uma série temporal aleatória, em que os valores passados não possuem influência sobre os valores futuros. Os resultados desse estudo indicaram que os coeficientes de *Hurst* das séries de países desenvolvidos estavam próximos de 0,5, o que é consistente com a hipótese de mercado eficiente (MUSSA et al., 2010). Por outro lado, os valores obtidos para os países emergentes se distanciaram mais desse valor, sugerindo que os mercados emergentes são menos eficientes que os mercados desenvolvidos. Este acontecimento pode ser explicado por diversos fatores, como a menor liquidez e transparência dos mercados emergentes, a menor qualidade das informações disponíveis, a presença de assimetrias de informação e a maior aparição de investidores não informados (NICHOLAS; ANDREI; ROBERT, 1998). Resultando em desvios sistemáticos dos preços em relação ao seu valor justo, e consequentemente a existência de oportunidades para os investidores obterem retornos consideráveis de forma consistente.

A busca por estratégias menos arriscadas e com maior potencial de retorno originou diversos modelos aplicáveis ao mercado financeiro, os quais posteriormente contribuíram para a exploração de evidências significativas que contestam a teoria de mercado eficiente (SAMUELSON, 2016). Esses modelos são baseados em:

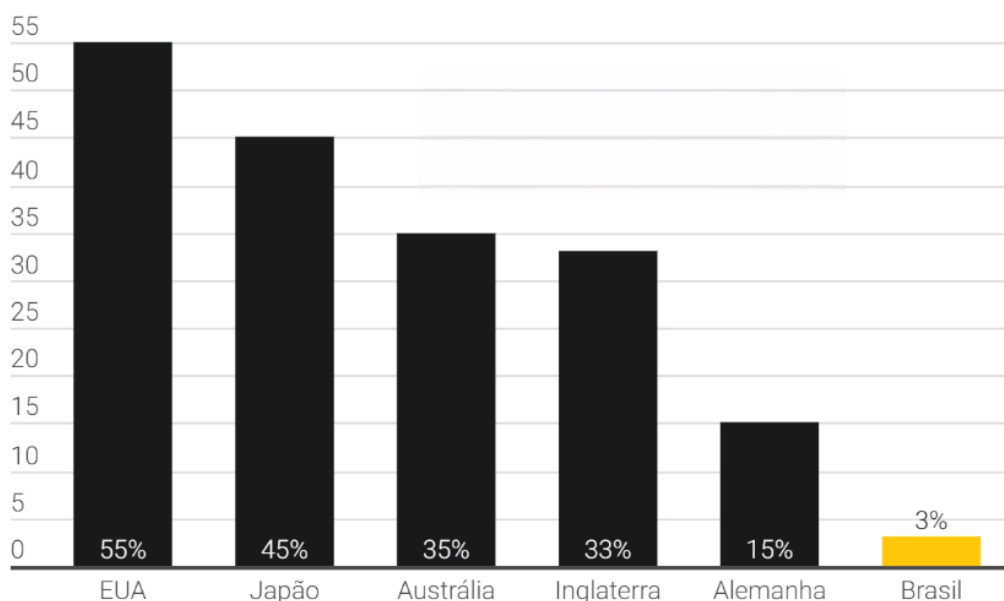
- **Análise Técnica (AT):** essa análise originou-se no ocidente por volta de 1900, quando Charles H. Dow, publicou uma série de editoriais sobre métodos de especulação com ações (MORRIS, 1994). Tais métodos surgiram com o objetivo de indicar padrões e tendências de mercado a partir de observações gráficas, baseado-se em informações passadas dos ativos financeiros.
- **Análise Fundamentalista (AF):** foi desenvolvida ao longo do século XX por Benjamin Graham (GRAHAM et al., 1934) com o objetivo de avaliar o preço justo (valor real) de ações que se respalda na expectativa de resultados futuros da empresa (CAVALCANTE; MISUMI; RUDGE, 2005). Tal análise se dá devido ao preço da ação na bolsa ser ocasionado de uma média de emoções e expectativas dos vendedores e compradores embutidos nesse meio.
- **Análise Quantitativa (AQ):** surgiu nas décadas de 1940 e 1950, impulsionada pelos primeiros computadores eletrônicos que permitiam processar grandes quantidades de dados de maneira rápida e eficiente (TING, 2015). O foco principal dessa abordagem é o uso de modelos matemáticos para identificar padrões e tendências no mercado.
- **Análise de Sentimento (AS):** surgiu posteriormente ao desenvolvimento teórico da linguística computacional e da Inteligência Artificial (IA), em meados da década de 1960. Porém, começou a ganhar destaque somente na década de 1990 com a chegada de programas de processamento de linguagem natural capazes de identificar a polaridade dos sentimentos expressos em textos (WILSON; WIEBE; HOFFMANN, 2009). Desde então, tal abordagem vem sendo utilizada em diferentes contextos, sendo um deles o mercado de ações e índices, focado em realizar previsões sobre a direção dos preços (BOLLEN; MAO; ZENG, 2011; RAO; SRIVASTAVA, 2012; NGUYEN; SHIRAI, 2015).

Como consequência, surgiram posteriormente modelos baseados em IA que combinam AT, AF, AQ e AS, com o objetivo de realizar operações mais precisas e assertivas no mercado de ações (AGRAWAL et al., 2022). Esses modelos incorporam algoritmos de aprendizado de máquina, permitindo a análise de grandes volumes de dados e o reconhecimento de padrões complexos (VADLAMUDI, 2017).

## 1.2 Motivação e Relevância

O crescente interesse e participação de pessoas físicas no mercado financeiro têm sido observados em todo o mundo. Em uma análise realizada pela Bolsa de Valores de Nova York, constatou-se que a globalização dos mercados financeiros e o aumento da renda das famílias em países em desenvolvimento são alguns dos principais fatores que impulsionam essa tendência (KANIEL; SAAR; TITMAN, 2008). Além disso, a facilidade de acesso à informação e o uso cada vez maior de tecnologias financeiras, como aplicativos de investimento, têm contribuído

Figura 1.1: Porcentagem de pessoas por país que investem em ações.



Fonte: Adaptado de XP (2023).

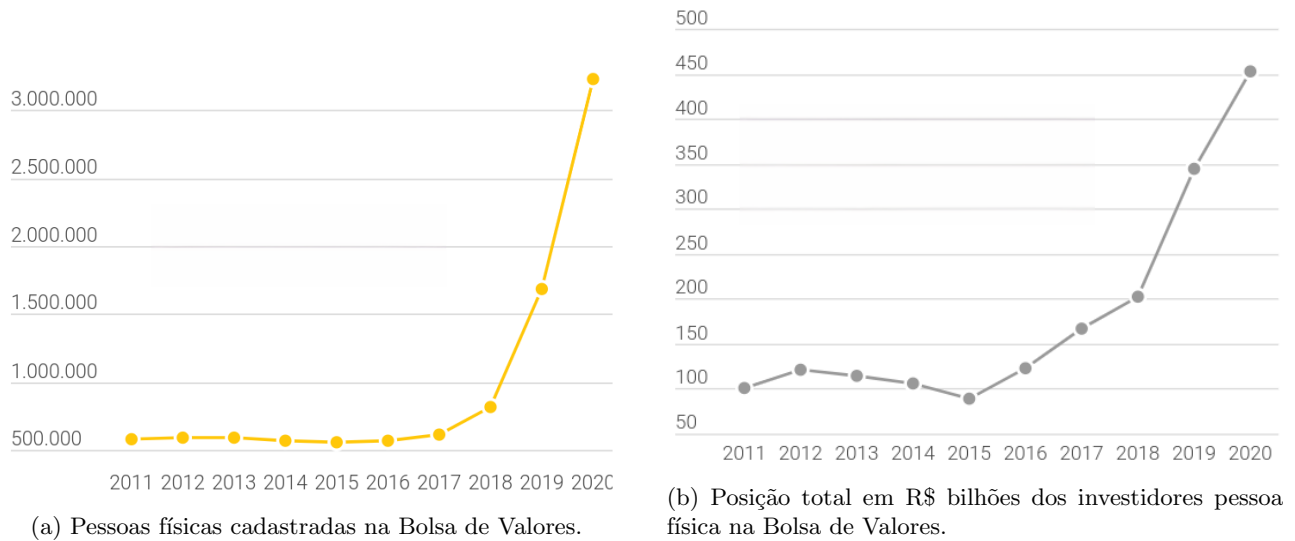
significativamente para o aumento da participação de pessoas físicas na bolsa de valores (UNCTAD, 2020). Essas mudanças refletem uma maior democratização do mercado financeiro, permitindo que mais pessoas tenham a oportunidade de investir e se envolver ativamente no processo de tomada de decisões financeiras.

O estudo realizado pela XP Investimentos (XP, 2023) revela que o número de pessoas físicas investindo em ações está diretamente relacionado à estabilidade do mercado e à qualidade de vida nos países analisados. A Figura 1.1 ilustra essa relação, confirmando que países com maior estabilidade econômica e política, como Estados Unidos e Japão, possuem uma maior proporção de investidores individuais. Por outro lado, nos mercados emergentes, o crescimento do número de investidores é mais acentuado devido à entrada de novos participantes. Esse fenômeno também tem sido observado na bolsa de valores brasileira (B3), com um aumento significativo no número de investidores pessoa física nos últimos anos. Em 2020, houve um acréscimo de mais de 2,8 milhões de investidores individuais em relação ao ano anterior, representando um crescimento de 93,7% em relação a 2019 (B3, 2023b). Esses incrementos podem ser visualizados na Figura 1.2a, o que reflete diretamente na posição total dos ativos como pode ser observado na Figura 1.2b.

Gonzalo e Granger (1995) mostraram que a previsão de séries temporais pode ser uma ferramenta útil na especulação de retornos de ações, com resultados significativos e positivos (GONZALO, 2010). Esta técnica tem sido explorada há décadas e é uma área de grande interesse para investidores em todo o mundo, uma vez que pode ser utilizada para melhorar a tomada de decisões no ato do investimento, mitigar riscos e gerar lucros. Sendo assim, é possível realizar o uso de técnicas de IA, aplicados a tarefas de previsão de sentido e/ou valor de ativos financeiros. Isso se dá devido à sua alta capacidade de identificar padrões dificilmente observados por seres humanos e de fornecer dados futuros com maior probabilidade de acontecimento, respaldados em padrões previamente computados (KORCZAK; HEMES, 2017).

Na literatura, várias pesquisas têm sido realizadas com o objetivo de prever valores futuros de ações. Esses estudos exploram diferentes abordagens, incluindo modelos matemáticos lineares que buscam identificar oportunidades em desvios da hipótese de mercado eficiente (RESENDE, 2021). Esses modelos lineares utilizam equações e métodos estatísticos para analisar os dados históricos e identificar padrões que possam sinalizar movimentos futuros no mercado financeiro. Adicionalmente, também são aplicados diferentes modelos de aprendizado de máquina na previsão de valores futuros de ações (FELIZARDO et al., 2019; NAMETALA et al., 2023). Esses modelos utilizam algoritmos e técnicas para aprender com dados históricos e fazer previsões com base em padrões identificados. Além disso, são exploradas abordagens híbridas que combinam vários modelos de previsão em um comitê, buscando obter um modelo final mais robusto e assertivo (KORCZAK; HEMES, 2017).

Figura 1.2: Dados da bolsa de valores brasileira.



Fonte: Adaptado de XP (2023).

Com o progresso tecnológico e o aumento do interesse dos investidores individuais, torna-se crucial explorar e desenvolver métodos que auxiliem na previsão de valores futuros de ações. Nesse contexto, o objetivo principal deste trabalho é contribuir para essa área de pesquisa, explorando diferentes modelos de previsão e aplicando-os tanto individualmente quanto em uma abordagem de comitê (*ensemble*). A proposta é construir um sistema de recomendação de investimentos que seja capaz de fornecer *insights* valiosos para os investidores, auxiliando-os na tomada de decisões financeiras mais informadas.

## 1.3 Objetivo

### 1.3.1 Objetivo Geral

O objetivo geral deste trabalho é aplicação de técnicas de IA para previsão e recomendação de investimentos no mercado financeiro.

### 1.3.2 Objetivos Específicos

Para alcançar o objetivo geral, propõe-se os seguintes objetivos específicos:

- revisar a literatura sobre mercado financeiro e técnicas de IA aplicadas a tarefas de previsão e classificação;
- propor e implementar uma abordagem para a previsão de valores e previsão de sentido de variação;
- avaliar o desempenho da abordagem proposta;
- propor e implementar uma abordagem para realizar a recomendação de compra e venda de ativos;
- avaliar os resultados da abordagem proposta.

## 1.4 Organização do Trabalho

Com o intuito de facilitar a compreensão da proposta, o estudo foi organizado em cinco capítulos, gradualmente introduzindo conceitos essenciais para a compreensão da abordagem. No presente capítulo, são apresentados o contexto do problema, as motivações, e os objetivos gerais e específicos da pesquisa. O Capítulo 2 explora estudos relacionados à previsão de valores no mercado financeiro. O Capítulo 3 detalha a abordagem proposta para a previsão de valores em séries temporais financeiras e apresenta recomendações de investimento. Os experimentos conduzidos para avaliar o desempenho da abordagem são descritos no Capítulo 4. Por fim, o Capítulo 5 conclui o estudo, discutindo possíveis direções para futuras pesquisas.

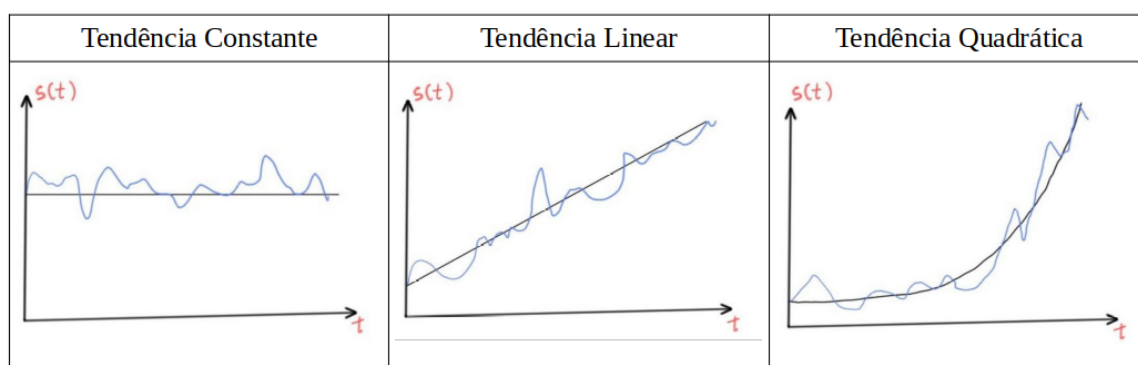
# Trabalhos Relacionados

Neste capítulo, são abordados conceitos fundamentais que regem o mercado financeiro, juntamente com a análise de trabalhos prévios focados na previsão e recomendação de compra e venda de ativos em diversos mercados. Inicialmente, na Seção 2.1, é apresentado o conceito de séries temporais, que posteriormente é relacionado com o mercado financeiro na Seção 2.2. Em seguida, na Seção 2.3, é descrita uma estrutura padrão seguida por grande parte dos trabalhos analisados. Na Seção 2.4, são apresentadas as variáveis utilizadas para a construção do conjunto de dados. A seguir, na Seção 2.5, é abordado o conceito de geração de variáveis a partir das já existentes. Na Seção 2.6, são discutidas técnicas de seleção das variáveis mais relevantes para o problema. Na Seção 2.7, são apresentados modelos computacionais e matemáticos utilizados para a predição dos valores ou tendências. Posteriormente, na Seção 2.8, são abordadas estratégias utilizadas para a tomada de decisão na bolsa. Na Seção 2.9, são discutidos alguns critérios de avaliação de desempenho. Por fim, na Seção 2.10 é apresentado um resumo dos trabalhos analisados.

## 2.1 Séries Temporais

Uma série temporal é uma sequência de observações coletadas em um determinado intervalo de tempo, que pode ser ou não igualmente espaçado, e que reflete a dependência serial dos dados. Portanto, uma boa representação de uma série temporal pode ser dada como:  $S_t, S_{t+1}, S_{t+2}, \dots, S_{t+n}$  em que  $S_t$  corresponde ao valor da série no instante  $t$ . Em razão desta sequência, uma série temporal pode assumir padrões determinísticos que são representados por uma ou mais variáveis matemáticas, ou pode ter um caráter estocástico, que inclui um componente aleatório na função geradora. Para analisar uma série temporal, uma abordagem comumente utilizada é a decomposição da série em componentes de tendência, ciclo e sazonalidade (PINO; MORETTIN; MENTZ, 1987). Isso ajuda a entender como a série evolui ao longo do tempo e pode fornecer informações úteis para modelagem e previsão.

Figura 2.1: Tendências de uma série temporal.

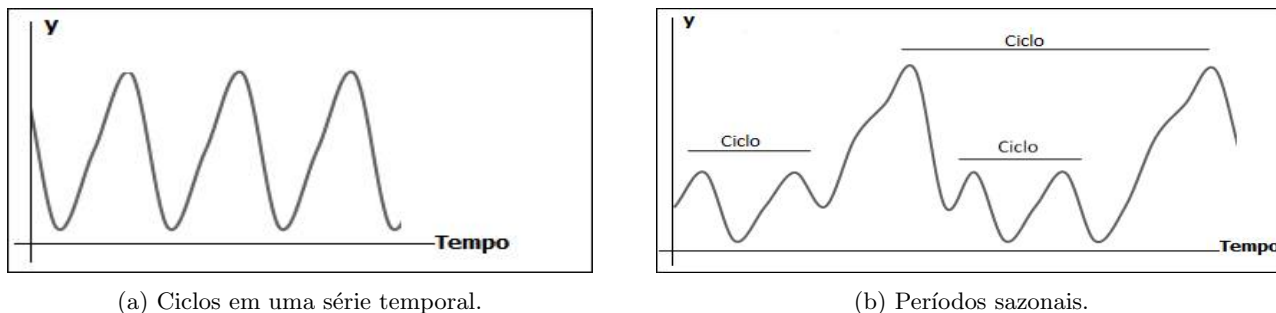


A tendência de uma série temporal é utilizada para descrever o comportamento da série ao longo do tempo, ou seja, verifica se há uma inclinação de alta, queda ou nula, além de indicar a velocidade dessas mudanças. Desse modo, existem diferentes formas matemáticas representativas destes movimentos da série temporal, como pode ser verificado na Figura 2.1. A tendência constante é aquela em que a série apresenta um comportamento linear ao longo do tempo, sem variações significativas. A tendência linear, por sua vez, é aquela em que a série apresenta uma inclinação de alta ou queda ao longo do tempo. Já a tendência quadrática é aquela em que a

série apresenta uma curvatura ao longo do tempo, podendo assumir comportamentos de concavidade para cima ou para baixo.

Já os ciclos são componentes de uma série temporal que visam identificar padrões que se repetem regularmente em um curto intervalo de tempo. Eles são caracterizados por movimentos cíclicos em torno da tendência com duração de períodos variáveis como pode ser visualizado na Figura 2.2a. Os ciclos são afetados por fatores externos, como mudanças econômicas, políticas, climáticas ou outros aspectos que regem a série abordada, e geralmente não são previsíveis a longo prazo.

Figura 2.2: Características cíclicas e sazonais de uma série temporal.



Fonte: Pinto (2009).

Por fim, entende-se a sazonalidade como as variações regulares que ocorrem em uma série temporal em intervalos fixos de tempo, como dias, semanas, meses ou anos. Um exemplo disso pode ser observado na Figura 2.2b, em que um padrão visual é repetido em um determinado intervalo de tempo. Essas variações podem ser influenciadas por fatores sazonais, como clima, feriados, eventos culturais, entre outros. A sazonalidade é uma componente importante de muitas séries temporais, e sua identificação é fundamental para uma análise precisa dos dados. A principal diferença entre componentes cíclicas e sazonais é que, enquanto nas cíclicas os movimentos são mais difíceis de prever, pois tendem a ser irregulares, na sazonalidade os movimentos ocorrem em intervalos regulares no tempo, tornando-os mais previsíveis.

## 2.2 Mercado Financeiro

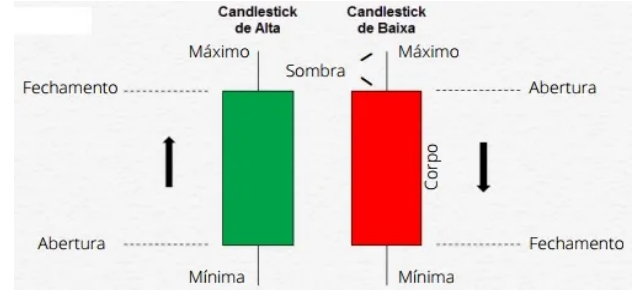
O mercado financeiro é um ambiente de negociação de diversos ativos financeiros, tais como ações, títulos, moedas, commodities e derivativos. Ele é composto por instituições financeiras, investidores, empresas, governos e outros agentes econômicos, que buscam negociar esses ativos visando lucros, proteção contra riscos e diversificação de investimentos. Os investidores (credores), que fornecem capital para o mercado financeiro, podem ser tanto pessoas físicas como jurídicas e podem investir em diferentes tipos de ativos, dependendo de seus objetivos, perfil de risco e estratégia. Já os capitadores (mutuários), que captam recursos, podem ser empresas, governo ou outras instituições que buscam financiamento para seus projetos ou operações (BRITO, 2019). As negociações no mercado financeiro geram uma grande quantidade de dados, que podem ser organizados e analisados como séries temporais devido a característica gradual e temporal do sistema, permitindo a identificação de tendências, ciclos e sazonalidades nos preços dos ativos ao longo do tempo.

Para facilitar a análise de séries temporais no mercado financeiro, conceitos como o gráfico de velas apresentado na Figura 2.3a tem sido amplamente adotado, pois consegue discretizar os dados de acordo com uma granulação sem perda de informações relevantes para as partes envolvidas (BULKOWSKI, 2012). Através deste gráfico, é possível obter informações como valor de abertura, valor máximo, valor mínimo e valor de fechamento de cada instante  $t$  de tempo, caracterizado do inglês *Open*, *High*, *Low* e *Close* (OHLC). Tais elementos podem ser observadas na Figura 2.3b. Além disso, outra informação relevante que pode ser correlacionada com esse gráfico é o volume de transações, que indica momentos de forte movimento no mercado ou baixa capacidade de conversão de capital. Respalda nessas informações, os investidores podem tomar decisões mais informadas sobre quando comprar ou vender ativos financeiros, buscando maximizar seus lucros e minimizar os riscos de seus investimentos.

Figura 2.3: Gráfico de velas e suas características observáveis.



(a) Representação de série temporal em gráfico de velas.



(b) Padrões observáveis em um candle.

Fonte: Trade Mental (2022).

Com base no modelo de gráfico de velas e na forma que os dados são discretizados em OHLC, surgiram diversas abordagens com o objetivo de obter melhores resultados no contexto de investimentos. Entre esses métodos, as alternativas baseadas em IA têm despertado grande interesse, devido à sua capacidade de identificar padrões complexos (DWIVEDI et al., 2021). Essa abordagem utiliza algoritmos de IA e dados históricos para analisar os padrões de mercado e tomar decisões de investimento de forma mais precisa e automatizada. Esses algoritmos são capazes de aprender com os dados históricos e adaptar suas estratégias com base nas mudanças da série temporal selecionada, oferecendo assim um potencial de obtenção de retornos superiores a média (STRADER et al., 2020).

## 2.3 Abordagens

Na tarefa de previsão da tendência ou do preço de um determinado ativo, uma variedade de técnicas podem ser empregadas. Entre elas, destacam-se a análise gráfica (MATSURA, 2017), análise de sentimento (IGARASHI; VALDEVIESO; IGARASHI, 2021) e análise fundamentada em dados numéricos (HALIL; DEMIRCI, 2019), esta última sendo o foco dos trabalhos analisados. Para isso, tais estudos seguem em sua maioria uma abordagem padrão, conforme ilustrado na Figura 2.4. Nesse processo, a primeira etapa consiste na extração de dados que apresentam uma boa correlação com a saída desejada, sendo essa etapa detalhada na Seção 2.4. Em seguida, são geradas novas variáveis por meio de modelos matemáticos, visando obter representações aprimoradas do problema no domínio temporal. As formulações matemáticas utilizadas para essa geração de variáveis podem ser encontradas na Seção 2.5. Posteriormente, ocorre a seleção de variáveis, empregando métodos estatísticos determinísticos e não determinísticos, incluindo modelos de IA, que são explorados com mais detalhes na Seção 2.6. Considerando que os estudos mencionados podem incorporar algumas ou todas as etapas da abordagem padrão (Figura 2.4), também foi observado em alguns trabalhos a utilização da normalização dos dados, com o objetivo de padronizar todos os valores em uma mesma escala.

Figura 2.4: Processo de desenvolvimento dos trabalhos analisados.



Com as variáveis selecionadas, inicia-se o processo de treinamento dos modelos utilizados em cada trabalho, conforme descrito na Seção 2.7. Após a obtenção dos resultados das previsões por meio desses modelos, é possível realizar uma análise mais aprofundada do seu desempenho. Alguns trabalhos utilizam a técnica de comitê (*ensemble*) que tende a aumentar a capacidade do modelo resultante (SAGI; ROKACH, 2018). Além disso, a ação final tomada com base nas previsões pode variar de acordo com os objetivos do pesquisador. Os métodos utilizados nessa etapa são abordados na Seção 2.8. Por fim, realiza-se uma análise do resultado por meio de métricas estatísticas, a fim de medir o desempenho do modelo ou da estratégia utilizada. Tais métricas são apresentadas na Seção 2.9.

## 2.4 Conjuntos de Dados

A qualidade e a quantidade de dados utilizados no treinamento do modelo de previsão, desempenham um papel fundamental na obtenção de resultados positivos. Para alcançar bons desdobramentos na tarefa de previsão, é necessário garantir que o conjunto de dados seja representativo da realidade e abranja o maior número possível de características do ativo financeiro em questão (KUMAR, 2020). Além disso, é importante realizar um pré-processamento adequado dos dados antes de utilizá-los no treinamento do modelo, pois a presença de dados incompletos, incorretos ou inconsistentes pode afetar a assertividade das previsões (KAUR; PANNU; MALHI, 2019).

Nesse sentido, técnicas de pré-processamento de dados são frequentemente aplicadas para lidar com esses problemas. Essas técnicas incluem a imputação de dados faltantes, a remoção de *outliers*, a padronização e normalização dos dados, a redução de dimensionalidade e a geração de novas variáveis por meio de cálculos matemáticos (TOMASEVIC; GVOZDENOVIC; VRANES, 2020). A escolha adequada do conjunto de dados e a aplicação correta das técnicas de pré-processamento são fatores determinantes para o sucesso da previsão utilizando métodos inteligentes.

Os estudos analisados nesta pesquisa utilizam dados brutos do volume de transação e OHLC em diversas granularidades, como minutos, horas ou dias. Por se tratar de uma série temporal, é comum que esses dados sejam coletados de forma contínua, sem lacunas ou dados faltantes. É importante ressaltar que a remoção de amostras não é recomendada, uma vez que a ordem das amostras é um fator crucial na construção das séries temporais. Além disso, outras variáveis são geradas a partir de cálculos matemáticos realizados com base nessa série, como descrito na Seção 2.5.

Certos trabalhos analisados nesta pesquisa optam por normalizar os dados, a fim de colocar os valores das variáveis dentro de um limite específico e reduzir as diferenças desproporcionais entre elas. Alguns pesquisadores realizam a normalização devido à capacidade do algoritmo em lidar apenas com dados normalizados (JIANG et al., 2022), enquanto outros aplicam essa técnica por terem observado melhorias nos resultados após a normalização (ANAND, 2021; WANG et al., 2022). É importante ressaltar que existem várias maneiras de normalizar um conjunto de dados, sendo as técnicas mais comuns a *Min-Max* (FELIZARDO et al., 2019; KUMAR; SINGH; JAIN, 2022), o *Z-Score* (FEI et al., 2021), a normalização por escala decimal (PATRO; SAHU, 2015), entre outras, que delimitam os valores dentro de um intervalo, geralmente de -1 a 1 ou 0 a 1 (KORCZAK; HEMES, 2017). Portanto, a normalização dos dados pode ser uma etapa essencial para permitir que o modelo compreenda melhor as relações entre as variáveis e, assim, realize previsões mais precisas.

## 2.5 Geração de Variáveis

Para melhorar o desempenho dos modelos, algumas abordagens utilizam técnicas de geração de variáveis adicionais derivadas dos dados brutos. Os métodos de geração de novos dados a partir dos dados brutos de OHLC são amplamente utilizados para fornecer informações adicionais sobre o ativo financeiro em questão.

Do conjunto de modelos matemáticos existentes com a capacidade de gerar dados relevantes para a previsão de valores de ativos financeiros, destacam-se dentre os trabalhos analisados:

- Valores Históricos (VH) - estabelece intervalos temporais conhecidos como janelas, abrangendo dados anteriores ao ponto de análise (AMARAL, 2021; ROSA, 2023), tais como  $S_{t-1}, S_{t-2}, \dots, S_{t-n}$ , em que

$S_{t-1}$  representa o valor da série temporal no instante  $t - 1$  e  $n$  é o tamanho da janela.

- *Simple Moving Average* (SMA) - um modelo matemático utilizado para suavizar a flutuação dos dados de preço ao longo do tempo e identificar tendências de forma mais clara (AMARAL, 2021; NAMETALA et al., 2023). A SMA pode ser obtida por:

$$SMA = \frac{1}{n} \sum_{i=0}^{n-1} S_{t-i}, \quad (2.1)$$

em que  $n$  é o tamanho da janela móvel e  $S_{t-i}$  representa o valor da série temporal no instante  $t - i$ .

- *Exponential Moving Average* (EMA) - um modelo matemático semelhante ao SMA, porém com mais peso atribuído aos preços mais recentes, tornando-o mais sensível às mudanças recentes nos dados (RESENDE, 2021). O EMA é calculado conforme:

$$EMA = \alpha(S_t - EMA_{t-1}) + EMA_{t-1}, \quad (2.2)$$

em que  $S_t$  é o valor da série temporal no instante  $t$ ,  $\alpha$  é o fator de suavização (frequentemente definido como  $\alpha = \frac{2}{n+1}$ ) e  $EMA_{t-1}$  representa o valor do EMA no instante  $t - 1$ .

- *Moving Average Convergence Divergence* (MACD) - identifica a direção e força de uma tendência predominante em um ativo financeiro (VEERAMANI; VENUGOPAL; MURUGANANDAN, 2022). O cálculo deste indicador é dado por:

$$MACD = EMA(w) - EMA(k), \quad (2.3)$$

em que  $w$  é a janela móvel curta e  $k$  é a janela móvel longa.

- *Commodity Channel Index* (CCI) - identifica pontos de reversão de tendência e avalia sua força (ALTAN; KARASU, 2019; HALIL; DEMIRCI, 2019). O cálculo deste indicador é realizado da seguinte forma:

$$CCI = \frac{TP - SMA(TP, c)}{0.015 \times DP(TP)}, \quad (2.4)$$

em que  $n$  é a janela de amostragem desejada,  $DP$  representa o desvio padrão da série e  $TP$  é calculado como  $TP = \frac{C+H+L}{3}$ , em que  $H$  é o ponto de máximo,  $L$  é o ponto de mínimo e  $C$  é o valor de fechamento do *candle* atual.

- *Average Directional Index* (ADX) - avalia a força de uma tendência predominante em um ativo financeiro, independentemente de sua direção (GAO, 2021). Este índice é calculado como se segue:

$$ADX = \frac{(n-1) \times EMA(TR) + TR}{n}, \quad (2.5)$$

em que  $n$  indica a janela de amostragem desejada e  $TR$  é obtido através de  $TR = \max(H - L, |H - C_{t-1}|, |L - C_{t-1}|)$ , em que  $H$  é o ponto de máximo do *candle* atual,  $L$  é o ponto de mínimo e  $C_{t-1}$  representa o valor de fechamento do *candle* anterior.

- *Rate of Change* (ROC) - mede a variação percentual de preço do ativo ao longo de um determinado período de tempo (AMPOMAH; QIN; NYAME, 2020; JIANG et al., 2020). Este indicador é computado por:

$$ROC = \frac{S_t - S_{t-n}}{S_{t-n}} \times 100, \quad (2.6)$$

considerando apenas o valor atual  $S_t$  e o valor de  $n$  estados passados  $S_{t-n}$  da série temporal.

- *True Strength Index* (TSI) - identifica a força e a direção de uma tendência em um determinado ativo financeiro, combinando a suavização do EMA com a taxa de variação ROC dos preços para fornecer sinais



de compra e venda (ANWAR; RAHMAN, 2019). Este índice é obtido através da seguinte fórmula:

$$TSI = \frac{EMA(EMA(PC, w), k)}{EMA(EMA(|PC|, w), k)} \times 100, \quad (2.7)$$

em que  $PC$  representa a variação do preço de fechamento ( $C_{t-1} - C_t$ ),  $w$  é a janela móvel curta e  $k$  é a janela móvel longa.

- Oscilador Estocástico (%K) - identifica a condição de sobre-compra e sobre-venda de um ativo financeiro, fornecendo sinais de compra e venda com base em movimentos de preço em relação à sua faixa de preço recente (FELIZARDO et al., 2019; VEERAMANI; VENUGOPAL; MURUGANANDAN, 2022). Podendo ser calculado pela equação a seguir:

$$\%K = \frac{C_t - L_{t-n}}{H_{t-n} - L_{t-n}}, \quad (2.8)$$

em que  $C_t$  é o valor de fechamento atual,  $L_{t-n}$  e  $H_{t-n}$  são o valor mínimo e máximo de  $n$  *candles* passados.

- Média Móvel Simples do Oscilador Estocástico (%D) - é uma média móvel de  $n$  períodos do %K, utilizada para fornecer um sinal mais suave e reduzir a volatilidade dos resultados (FELIZARDO et al., 2019; VEERAMANI; VENUGOPAL; MURUGANANDAN, 2022). O %D é obtido por:

$$\%D = \frac{\sum_{i=0}^{n-1} \%K_{t-i}}{n}, \quad (2.9)$$

em que  $\%K_{t-i}$  é o valor de %K no instante  $t - i$ .

- Indicador *Williams* (%R) - assim como o indicador %K, esse método tem como objetivo identificar condições de sobre-compra e sobre-venda de um ativo financeiro, porém com escalas invertidas (LEE et al., 2021). É calculado pela seguinte equação:

$$\%R = \frac{H_{t-n} - C_t}{H_{t-n} - L_{t-n}}, \quad (2.10)$$

onde  $C_t$  é o valor de fechamento atual,  $L_{t-n}$  e  $H_{t-n}$  representam o valor mínimo e máximo de  $n$  *candles* passados.

É importante observar que muitos dos métodos mencionados anteriormente podem ser empregados como modelos de predição. No entanto, este trabalho não se concentra nessa abordagem, uma vez que não é o foco da pesquisa.

## 2.6 Seleção de Variáveis

A seleção de variáveis desempenha um papel crucial no modelo de IA, pois um grande número de variáveis pode dificultar o treinamento e levar ao *overfitting*, prejudicando o desempenho do modelo. Apesar do processo de extração e criação gerar muitas variáveis descritivas, nem todas são igualmente úteis para a previsão, e a seleção de variáveis permite concentrar-se nos aspectos mais importantes e relevantes para essa tarefa (MEYER et al., 2019). Para isso, existem diversos métodos de seleção de variáveis, alguns voltados especificamente para a atividade de classificação, enquanto outros podem ser aplicados tanto para classificação quanto para regressão. Esses métodos auxiliam na identificação das variáveis mais relevantes e descartam aquelas que têm pouca influência no modelo final.

Uma parte dos trabalhos analisados utilizou métodos de seleção específicos para o problema de classificação. Dentre esses métodos, destacam-se:

- *Fisher* - aplicado por Peng et al. (2021), que calcula o score de *Fisher* para selecionar as características mais importantes com base na separação entre as classes (FISHER, 1936);
- *Gini* - empregado por Ji et al. (2022), que calcula o índice de impureza de *Gini* para cada *feature* em relação à variável de saída, sendo selecionadas aquelas com maior poder discriminativo (GINI, 1921);

- *Chi-Squared* ( $X^2$ ) - utilizado por Rosa (2023), que determina se existe uma relação significativa entre duas variáveis categóricas, auxiliando na seleção das características mais informativas para a classificação (PEARSON, 1900).

Esses métodos contribuem para identificar as características mais relevantes e reduzir a dimensionalidade do conjunto de dados, melhorando a eficiência e a interpretabilidade dos modelos de classificação.

Para abranger a classe dos algoritmos de regressão, foram utilizados diversos métodos de seleção de variáveis. Entre eles, foram empregadas abordagens estocásticas, que visam encontrar o melhor conjunto de variáveis por meio de métodos não-determinísticos, representados pelos modelos de IA. Esses métodos treinam uma determinada rede e avaliam qual conjunto de variáveis foi mais eficiente para a tarefa de previsão. Nesse contexto, os métodos mais utilizados foram:

- *Random Forest* (RF) aplicado por Aminimehr et al. (2022), que gera várias combinações de entradas de forma aleatória e treina uma árvore de decisão para cada combinação, selecionando assim as variáveis da árvore que obtêm os melhores resultados (BREIMAN, 2001);
- Lasso empregado por Sermpinis, Tsoukas e Zhang (2018), que separa as melhores variáveis com base em seus pesos na tarefa de previsão, zerando o peso das variáveis menos relevantes (MUTHUKRISHNAN; ROHINI, 2016);
- ElasticNet adotado por Zhang, Wahab e Wang (2023), que combina as técnicas de regularização L1 (Lasso) e L2 (Ridge), gerando assim um vetor ponderado pelo nível de significância de cada variável (AMINI; HU, 2021).

É importante ressaltar que há também métodos determinísticos baseados em conceitos estatísticos para essa tarefa, como o teste de *Kruskal-Wallis*, utilizado por Amaral (2021), que compara as medianas de duas ou mais amostras independentes para determinar se há diferença estatisticamente significativa entre elas (KRUSKAL; WALLIS, 1952); e o método de Regressão de Informação Mútua (RIM), empregado por Zhao, Anand e Wang (2019), que mede a dependência mútua entre cada recurso e a variável de destino, utilizando a entropia da informação (KRASKOV; STÖGBAUER; GRASSBERGER, 2004).

## 2.7 Técnicas de Predição

Existem diversas abordagens para o processo de predição de valores, desde modelos lineares até modelos não lineares. Cada uma dessas abordagens possui suas próprias características e cenários de atuação, resultando em um bom desempenho de alguns em relação a outros em cenários específicos. No entanto, não é possível afirmar que uma abordagem é totalmente superior à outra, uma vez que a escolha depende do problema em questão e dos dados disponíveis. Dentre os métodos explorados, pode-se citar as abordagens estatísticas, modelos baseados em IA.

Dentro da classe de modelos estatísticos, são exploradas diferentes abordagens para realizar previsões e análises. Além de ajustes lineares que buscam a linearidade local na vizinhança (RESENDE, 2021) e o uso de SMA (NAMETALA et al., 2023), também são empregados métodos como *Autoregressive Integrated Moving Average* (ARIMA) (FELIZARDO et al., 2019; GAO, 2021), *Seasonal Autoregressive Integrated Moving Average* (SARIMA) e *Generalized Autoregressive Conditional Heteroskedasticity* (GARCH) (NAMETALA et al., 2023). O modelo ARIMA combina elementos de autoregressão (AR), média móvel (MA) e diferenciação (I) para modelar padrões temporais e sazonalidade, enquanto o SARIMA incorpora componentes de sazonalidade em adição aos componentes de tendência e aleatoriedade. Já o modelo GARCH é usado para modelar a volatilidade condicional em séries temporais financeiras, capturando a natureza heterocedástica dos retornos. É importante ressaltar que muitos desses modelos estatísticos são utilizados em conjunto com abordagens de IA, permitindo explorar as vantagens de cada modelo e obter previsões mais precisas e confiáveis (NAMETALA et al., 2023).

Já a classe de *machine learning*<sup>1</sup> foi amplamente explorada nos trabalhos analisados, abrangendo uma variedade de modelos com diferentes naturezas. Dentre os modelos de natureza linear, destacam-se o uso do *Support Vector Machine* (SVM) (ALTAN; KARASU, 2019; ANAND, 2021) e *Logistic Regression* (LR) (PABUÇCU; ONGAN; ONGAN, 2023). Além disso, foram investigados modelos de natureza não linear, como *Multi-Layer Perceptron* (MLP) (KORCZAK; HEMES, 2017; SAGHI; REZAEI, 2021), *Extreme Learning Machine* (ELM) (JIANG et al., 2022), *evolving Granular Neural Network* (eGNN), *evolving Multivariable Gaussian Fuzzy System* (eMG), *evolving Optimal Granular System* (eOGS) (AMARAL, 2021), *WaveNet* (FELIZARDO et al., 2019), *Support Vector Regression* (SVR), *Radial Basis Function* (RBF), *Adaptive Neuro-Fuzzy Inference System* (ANFIS) (SAGHI; REZAEI, 2021), *Decision Trees* (DT) e *K-Nearest Neighbors* (KNN) (HALIL; DEMIRCI, 2019). Essa diversidade de modelos mostra a abrangência das abordagens de *machine learning* utilizadas nos estudos, permitindo explorar tanto a linearidade quanto a não linearidade dos dados para obter previsões mais precisas e acuradas.

Por fim, os modelos de *deep learning*<sup>2</sup> despertaram grande expectativa nos estudos analisados, devido à sua capacidade de reconhecer padrões complexos. Foram empregados diversos modelos, como *Long Short-Term Memory* (LSTM) (ANWAR; RAHMAN, 2019; ZHANG et al., 2021; NI; XU, 2021; GAO, 2021; LEE et al., 2021; YILMAZ; YILDIZTEPE, 2022), *Transformer* (WANG et al., 2022), *Convolutional Neural Network* (CNN) (ANAND, 2021), *Recurrent Neural Networks* (RNN) (ZHANG et al., 2021; GAO, 2021; ANAND, 2021; YILMAZ; YILDIZTEPE, 2022), *Temporal Convolutional Network* (TCN) (YILMAZ; YILDIZTEPE, 2022), *Deep Belief Network* (DBN) (ZHANG et al., 2021), *Deep Convolutional Fuzzy System* (DCFS) (WANG, 2019) e *Dynamic Correlation Deep Neural Network* (DCDNN) (NI; XU, 2021). Essa variedade de modelos demonstra a ampla gama de técnicas de *deep learning* aplicadas, explorando a capacidade dessas redes neurais em capturar informações de longo prazo, realizar análises sequenciais e lidar com dados de alta complexidade.

## 2.8 Estratégias

Após a previsão do valor de um ativo específico, torna-se possível tomar uma decisão embasada, e para isso, os estudos analisados apresentam diversas estratégias distintas com o objetivo de maximizar o retorno durante as operações.

Uma pesquisa conduzida por Salman e Kecman (2012) demonstrou a possibilidade de transformar modelos de regressão em modelos de classificação com base em três regras similares apresentadas na Tabela 2.1. Nesta abordagem, a classe -1 indica uma tendência de queda, 0 indica ausência de tendência, e 1 representa uma tendência de alta para a interação em questão.

Exemplo	Regra	Classe
A	$Y_t > Y_{t-1}$	1
	$Y_t = Y_{t-1}$	0
	$Y_t < Y_{t-1}$	-1
B	$Y_t \geq Y_{t-1}$	1
	$Y_t < Y_{t-1}$	-1
C	$Y_t > Y_{t-1}$	1
	$Y_t \leq Y_{t-1}$	-1

Tabela 2.1: Transformação da saídas no domínio contínuo para o domínio discreto.

Nesse sentido, alguns estudos exploraram modelos robustos de regressão para, posteriormente, converter as saídas em classes de alta, baixa ou até mesmo constância dos valores (AMPOMAH; QIN; NYAME, 2020; JIANG et al., 2020; ZHANG et al., 2021). Essa abordagem permite uma interpretação mais direta e prática dos

<sup>1</sup>Modelos de Machine Learning são algoritmos e sistemas que aprendem padrões a partir de dados, permitindo que uma máquina tome decisões ou faça previsões sem ser explicitamente programada para uma tarefa específica. Esses modelos podem ser supervisionados, não supervisionados ou por reforço, dependendo do tipo de aprendizado envolvido.

<sup>2</sup>Deep Learning é uma subárea do Machine Learning que utiliza redes neurais artificiais profundas para realizar tarefas complexas de aprendizado e reconhecimento de padrões. Essas redes, inspiradas no funcionamento do cérebro humano, consistem em várias camadas de neurônios interconectados, permitindo a extração automática de características hierárquicas dos dados.

resultados, facilitando a tomada de decisões e a implementação de estratégias de investimento. Ao transformar as previsões em classes, os modelos podem fornecer informações sobre a direção do movimento do ativo financeiro, indicando se é provável que ele aumente, diminua ou permaneça estável.

Estratégias baseadas em classificação geralmente adotam a abordagem de compra quando o retorno é classificado como positivo e venda quando o retorno é classificado como negativo (WANG et al., 2022). Em alguns estudos que utilizam essa técnica, essas operações são executadas com todo o capital disponível na carteira de investimentos (KORCZAK; HEMES, 2017; RESENDE, 2021; VEERAMANI; VENUGOPAL; MURUGANANDAN, 2022).

No trabalho desenvolvido por Nametala et al. (2023), não é utilizado apenas um modelo de previsão, mas sim uma combinação de diferentes preditores. Para lidar com essa situação, estratégias são desenvolvidas para integrar as saídas desses diversos modelos de forma a maximizar os resultados obtidos. Portanto, foi criado um modelo de comitê (*ensemble*) para tomar decisões com base nas previsões dos modelos individuais. Nesse trabalho, são apresentadas três regras para obter o sinal final. A primeira regra é a *Average Decision Committee* (ADC), que verifica se os resultados dos modelos do comitê coincidem em relação à tendência (subir ou descer). Se os modelos indicarem a mesma direção, um sinal de subir ou descer é gerado; caso contrário, um sinal estável é retornado. A segunda regra é a *Majority Decision Committee* (MDC), que considera a contagem de votos e gera um sinal de subir ou descer se houver uma maioria de votos concordantes. Caso contrário, um sinal estável é retornado. Por fim, temos a *Average and Majority Committee* (AMC), que verifica se as duas regras anteriores estão em acordo. Se os resultados forem semelhantes, um sinal de subir ou descer é gerado; caso contrário, um sinal estável é gerado. Com base na saída da regra AMC, é possível decidir entre comprar ou vender o ativo financeiro. Essa escolha é feita de acordo com a interpretação do sinal gerado pela regra. Quando a saída indica um sinal de compra, significa que há uma tendência de alta no ativo, o que pode ser um momento favorável para realizar uma compra. Por outro lado, quando a saída indica um sinal de venda, significa que há uma tendência de queda no ativo, o que pode ser apropriado para realizar uma venda.

## 2.9 Critérios de Avaliação

As ferramentas de análise são responsáveis por fornecer parâmetros de comparação e visualização do desempenho das abordagens. Para tal, dois ramos da análise estatística são direcionados à tarefa de previsão com modelos de IA: a análise de problemas de classificação e a análise de problemas de regressão. A análise de problemas de classificação é voltada para a previsão de eventos discretos, como a indicação de compra ou venda de uma ação. Já a análise de problemas de regressão é voltada para previsão de valores contínuos, como o preço de uma ação em um determinado momento.

Figura 2.5: Matriz de confusão.

		Classe Predita	
		Subir	Descer
Classe Real	Subir	Verdadeiro Positivo (VP)	Falso Negativo (FN)
	Descer	Falso Positivo (FP)	Verdadeiro Negativo (VN)

Para avaliar a efetividade dos modelos de classificação, é comum utilizar métricas que mensurem a quantidade de classificações corretas, tais como:

- Precisão Positiva e Negativa - são métricas que medem a capacidade de um modelo em identificar corretamente amostras positivas e negativas, respectivamente (ZHANG et al., 2021). Essas métricas são calculadas usando uma matriz de confusão, que compara as classificações do modelo com as classes verdadeiras dos dados. A matriz de confusão pode ser visualizada na Figura 2.5. A equação para o cálculo da Precisão Positiva (PP), utiliza os dados de Verdadeiro Positivo (VP) e Falso Positivo (FP), sendo descrita como:

$$PP = \frac{TP}{TP + FP}. \quad (2.11)$$

Já para calcular a Precisão Negativa (PN), utiliza-se os valores de Verdadeiro Negativo (VN) e Falso Negativo (FN) conforme:

$$PN = \frac{TN}{TN + FN}. \quad (2.12)$$

- *Recall* - é uma métrica que mede a proporção de amostras positivas corretamente identificadas pelo modelo em relação ao total de amostras positivas presentes nos dados reais. Em outras palavras, o *recall* indica a capacidade do modelo de "lembrar" corretamente das amostras positivas (AMPOMAH; QIN; NYAME, 2020). Tal método pode ser calculado por:

$$Recall = \frac{TP}{TP + FN}. \quad (2.13)$$

- Acurácia (AC) - mede a proporção de amostras classificadas corretamente em relação ao total de amostras (AMPOMAH; QIN; NYAME, 2020; AMARAL, 2021; LEE et al., 2021). A fórmula para o cálculo da AC é dada por:

$$AC = \frac{TP + TN}{TP + TN + FP + FN}. \quad (2.14)$$

Note que a acurácia (AC) pode ser enganosa em conjuntos de dados desbalanceados, nos quais uma classe é predominante. Nesses casos, um modelo pode obter uma alta taxa de acertos simplesmente prevendo a classe predominante para todas as amostras. Para lidar com esse problema, é necessário utilizar outra métrica que avalie o desempenho dos modelos de classificação em termos de sensibilidade e equilíbrio entre Precisão e *Recall*, tal como o *F1 score* (F1) que é uma média harmônica da PP e do *Recall*. Essa métrica fornece uma medida única de desempenho de um modelo de classificação, equilibrando a importância tanto da PP quanto do *Recall* (JIANG et al., 2020). O cálculo da F1 é dada por:

$$F1 = 2 \frac{PP \cdot Recall}{PP + Recall}. \quad (2.15)$$

Já para os modelos de regressão, a efetividade é mensurada de diferentes formas, uma vez que é muito improvável obter a previsão exata do valor de saída. Nesse contexto, os trabalhos analisados utilizam das seguintes métricas:

- Coeficiente de Determinação ( $R^2$ ) - é uma medida da qualidade do ajuste do modelo aos dados observados (RESENDE, 2021; NI; XU, 2021). O valor de  $R^2$  varia de 0 a 1, indicando a proporção de variabilidade explicada pelo modelo e é dado por:

$$R^2 = 1 - \left( \frac{SSE}{SST} \right), \quad (2.16)$$

sendo que *SSE* é a soma dos quadrados dos resíduos do modelo, que representa a variação não explicada pelo modelo de regressão e o *SST* é a soma dos quadrados totais, que representa a variação total da variável dependente.

- *Mean Error* (ME) - verifica se as previsões estão consistentemente superestimando ou subestimando os valores reais (FELIZARDO et al., 2019). Pode ser calculado por:

$$ME = \frac{1}{n} \sum (y_{pred} - y_{real}), \quad (2.17)$$

no qual  $n$  é a quantidade de amostras,  $y_{pred}$  é o valor predito pelo modelo e  $y_{real}$  é a saída desejada.

- *Mean Absolute Error* (MAE) - quantifica o erro médio entre as previsões e os valores reais de uma variável sem se importar com o sentido do erro (AMINIMEHR et al., 2022; JIANG et al., 2022). O MAE pode ser obtido por:

$$MAE = \frac{1}{n} \sum (|y_{pred} - y_{real}|). \quad (2.18)$$

- *Mean Squared Error* (MSE) - avalia a dispersão das previsões em relação aos valores reais, penalizando erros maiores de forma quadrática (NI; XU, 2021; WANG et al., 2022). O MSE é calculado por:

$$MSE = \frac{1}{n} \sum (y_{pred} - y_{real})^2. \quad (2.19)$$

- *Mean Absolute Percent Error* (MAPE) - avalia a precisão de um modelo de previsão em termos percentuais (ANAND, 2021; YILMAZ; YILDIZTEPE, 2022). O MAPE pode ser obtido por:

$$MAPE = \frac{1}{n} \sum \left| \frac{y_{pred} - y_{real}}{y_{real}} \right| 100. \quad (2.20)$$

- *Root Mean Squared Error* (RMSE) - mede a diferença entre os valores previstos por um modelo e os valores reais, sendo expressa na mesma unidade dos valores originais (ALTAN; KARASU, 2019; PABUÇCU; ONGAN; ONGAN, 2023). O RMSE é calculado por:

$$RMSE = \sqrt{\frac{1}{n} \sum (y_{pred} - y_{real})^2}. \quad (2.21)$$

É importante ressaltar que diversos estudos analisaram os modelos por meio dos resultados obtidos após a aplicação da estratégia de investimento, utilizando uma variedade de métricas econômicas. Essas métricas incluem o retorno financeiro, a comparação do desempenho do modelo em relação à estratégia *buy and hold* (comprar e manter) (NAMETALA et al., 2023), o retorno percentual do valor investido (WANG et al., 2022), a avaliação do lucro e perda, o retorno anual esperado (RESENDE, 2021), o número de transações realizadas, o número de operações rentáveis (VEERAMANI; VENUGOPAL; MURUGANANDAN, 2022), o número de transações rentáveis consecutivas, a perda bruta e o lucro bruto (KORCZAK; HEMES, 2017). Essas métricas fornecem uma avaliação abrangente do desempenho dos modelos em termos de rentabilidade, consistência e eficácia na tomada de decisões de investimento. Além disso, no contexto da avaliação de desempenho dos modelos, também é possível aplicar testes estatísticos, como o T-test (KIM, 2015), teste de McNemar (MCNEMAR, 1947), teste de Friedman (SHELDON; FILLYAW; THOMPSON, 1996), entre outros, a fim de obter uma análise estatisticamente significativa dos resultados.

## 2.10 Resumo dos Trabalhos Relacionados

Os estudos analisados demonstraram uma diversidade de abordagens, utilizando uma ampla gama de modelos de IA. Esses modelos foram aplicados com o propósito de prever o valor e/ou a tendência dos ativos financeiros. A Tabela 2.2 fornece uma visão geral dos objetivos de cada estudo e dos modelos de previsão adotados em cada um deles.

Autor	Modelos	Objetivo
Korczak e Hemes (2017)	<i>Deep Learning</i> H2O, MLP e B&H.	Realizar previsões em um sistema de negociação de ações multiagente <i>A-Trader</i> .
Felizardo et al. (2019)	LSTM, WaveNet, SCM, RF e ARIMA.	Prever valores da série temporal do Bitcoin.

Autor	Modelos	Objetivo
Altan e Karasu (2019)	SVM.	Prever as taxas de câmbio USD/TRY e EUR/TRY.
Halil e Demirci (2019)	SVM, <i>Variant Decision Trees</i> , KNN e ANN.	Prever os movimentos futuros de preços do índice BIST 30 da bolsa de valores turca.
Ampomah, Qin e Nyame (2020)	RF, <i>Boosting</i> e <i>XGBoost</i> .	Aplicar técnicas de <i>ensemble</i> a fim de obter melhores resultados na predição de valores de ativos financeiros.
Jiang et al. (2020)	<i>Decision Tree</i> , SVM e ANN.	Desenvolver um <i>framework</i> de <i>Stacking</i> aprimorado para prever a direção do índice de preços das ações.
Amaral (2021)	eGNN, eMG e eOGS.	Realizar sugestões de compra e venda de criptoativos.
Resende (2021)	Ajuste Linear	Realizar a previsão de tendência de ativos financeiros.
Anand (2021)	CNN, LSTM, RNN, MLP e SVM.	Comparar o desempenho de redes de aprendizado profundo na tarefa de predição de séries temporais.
Saghi e Rezaee (2021)	MLP, RBF, ANFIS, GMDH e SVR.	Melhorar o desempenho de algoritmos de aprendizagem de máquina.
Zhang et al. (2021)	DBN e LSTM.	Apresentar um modelo de previsão de movimento de preços de ações.
Gao (2021)	ARIMA, ANN, RNN e LSTM.	Comparar diferentes técnicas de previsão de preços de ações.
Ni e Xu (2021)	DCDNN e LSTM.	Propor uma nova abordagem de seleção/correlação de <i>features</i> .
Lee et al. (2021)	LSTM.	Explorar a viabilidade e eficácia dos indicadores de análise técnica.
Aminimehr et al. (2022)	LSTM.	Prever o retorno diário do S&P 500.
Wang et al. (2022)	<i>Transformer</i> .	Realizar previsões no mercado de ações e comparar com <i>buy and hold</i> .
Veeramani, Venugopal e Muruganandan (2022)	FMCDM.	Avaliar o desempenho de diferentes métodos de tomada de decisão multicritério difusos na seleção de indicadores técnicos para a negociação diária.
Yilmaz e Yildiztepe (2022)	RNN, TCN, LSTM e <i>Gated Recurrent Unit Model</i> .	Avaliar o desempenho de modelos para previsão de retorno de ações em diferentes horizontes de previsão.
Jiang et al. (2022)	ELM e IHS.	Propor dois modelos de previsão de preços de ações em duas etapas, chamados de EMD-ELM-IHS e VMD-ELM-IHS.
Nametalala et al. (2023)	SMA, ARIMA, SARIMA, GARCH e MLP.	Utilizar técnicas de IA para a construção de um Robô de operações no mercado financeiro do Brasil.
Pabuçcu, Ogan e Ogan (2023)	SVM, ANN, NB, LR e RF.	Explorar diferentes técnicas de previsão de preços do Bitcoin.

Tabela 2.2: Síntese dos trabalhos relacionados.

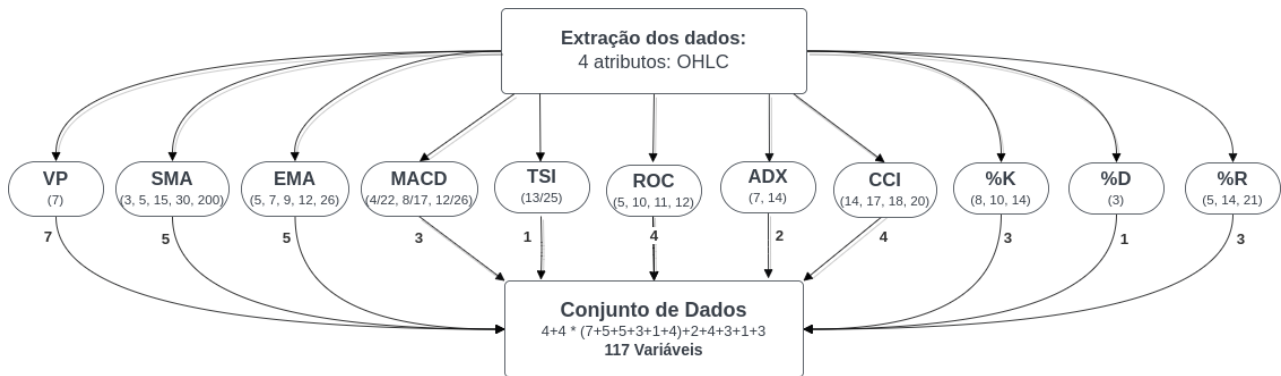
# Abordagem Proposta

Para o desenvolvimento deste trabalho, propõe-se quatro etapas principais que seguem a estrutura padrão adotada pelos trabalhos analisados, conforme ilustrado na Figura 2.4. Inicialmente, é realizada a construção do conjunto de dados, cujos processos são descritos em detalhes na Seção 3.1. Em seguida, ocorre a seleção de variáveis relevantes, abordada na Seção 3.2. Com as variáveis selecionadas, inicia-se o processo de previsão, que é detalhado na Seção 3.3. Por fim, é realizada a recomendação de investimento com base nos valores previstos, conforme descrito na Seção 3.4.

## 3.1 Construção dos Conjuntos de Dados

O processo de construção do conjunto de dados é dividido em duas etapas distintas. Primeiramente, ocorre a extração dos dados, conforme detalhado na Seção 3.1.1. Em seguida, é realizada a geração de novas variáveis a partir dos dados extraídos, como explicado na Seção 3.1.2. Nessa fase, são criadas variáveis adicionais que fornecem informações relevantes para o processo de previsão e análise. A Figura 3.1 apresenta uma visão geral do processo de construção do conjunto de dados, ilustrando de forma visual as etapas mencionadas. Esse processo garante a disponibilidade dos dados necessários e a preparação adequada das variáveis para as etapas subsequentes do trabalho.

Figura 3.1: Visão macro do processo de construção do conjunto de dados.



### 3.1.1 Extração de Atributos

O mercado de ações do Brasil possui um horário de funcionamento de segunda a sexta-feira, das 10:00 às 17:00 (B3, 2023a). Durante esse período, são geradas informações em diferentes granularidades, permitindo acompanhar a evolução dos preços e volumes de negociação ao longo do tempo. Uma plataforma amplamente utilizada para acessar essas informações é o *Investing*<sup>1</sup>, que disponibiliza dados históricos e em tempo real em diferentes intervalos, como diário, semanal e mensal. Essa plataforma abrange diversos mercados e modalidades de investimento em vários países, fornecendo informações relevantes, como os preços de abertura, máxima, mínima, fechamento (OHLC), volume negociado e percentual de mudança. É importante mencionar que existem outras fontes disponíveis para a coleta de dados, como a plataforma *Profit*<sup>2</sup>, a B3<sup>3</sup>, entre outras. No contexto deste trabalho, os dados históricos OHLC foram extraídos da plataforma *Profit* para análise e estudo.

<sup>1</sup><https://www.investing.com>

<sup>2</sup><https://www.nelogica.com.br/produtos/profit-ultra>

<sup>3</sup><https://www.b3.com.br/pt-br/market-data-e-indices/servicos-de-dados/market-data/historico/>



### 3.1.2 Geração de Variáveis

Após a extração dos dados, são aplicadas diversas técnicas para gerar variáveis relevantes na tarefa de previsão. Para esse propósito, são calculadas novas variáveis a partir de cada valor de OHLC, utilizando diferentes fórmulas econômicas ajustadas a parâmetros específicos abordados na literatura. Essas técnicas visam extrair informações adicionais dos dados e incluem:

- VH - é obtido por meio do deslocamento de 7 valores passados (AMARAL, 2021);
- SMA - calculada a partir da Equação (2.1) empregada aos intervalos de 3 (CHANTARAKASEMCHIT; NUCHITPRASITCHAI; NILSIAM, 2020), 5, 15, 30 (HANDAYANI et al., 2019) e 200 (ELLIS; PARBERY, 2005);
- EMA - tem seu calculo baseado na Equação (2.2), sendo aplicada aos períodos amostrais de 5, 7, 9 (ANBALAGAN; MAHESWARI, 2015), 12 (RESENDE, 2021) e 26 (ANANTHI; VIJAYAKUMAR, 2021);
- MACD - derivado da Equação (2.3) aplicada aos valores de janela móvel curta e janela móvel longa como 12/26 (ANBALAGAN; MAHESWARI, 2015; HANDAYANI et al., 2019), 8/17 e 4/22 (KANG, 2021), respectivamente;
- CCI - calculado a partir da Equação (2.4), considerando os períodos de amostragem de 14 (HALIL; DEMIRCI, 2019), 17, 18 (KARASU; ALTAN, 2022) e 20 (KELOTRA; PANDEY, 2020) para a janela de análise;
- ADX - obtido através da Equação (2.5) aplicada aos períodos de 7 (KELOTRA; PANDEY, 2020) e 14 (SHAMSEDDIN; ELMESKI, 2022) espaços amostrais;
- ROC - determinado a partir da Equação (2.6) empregada aos intervalos de 5, 10, 11 e 12 (KARASU; ALTAN, 2022);
- TSI - computado a partir da Equação (2.7) com base nos intervalos de 13 para a janela móvel curta e 25 para a janela móvel longa (NAYAK; MISHRA; RATH, 2015; ANWAR; RAHMAN, 2019);
- %K - calculado com base na Equação (2.8) utilizando os intervalos de tempo de 8 (NI et al., 2022), 10 (IJEGBWA et al., 2014) e 14 amostras (VEERAMANI; VENUGOPAL; MURUGANANDAN, 2022);
- %D - derivado da Equação (2.9) aplicada a janela de tempo de 3 amostras (IJEGBWA et al., 2014; VAIDYA, 2018);
- %R - tem seu calculo baseado na Equação (2.10), considerando as janelas de tempo de 5, 14 e 21 espaços amostrais (ALMEIDA; REYNOSO-MEZA; STEINER, 2016).

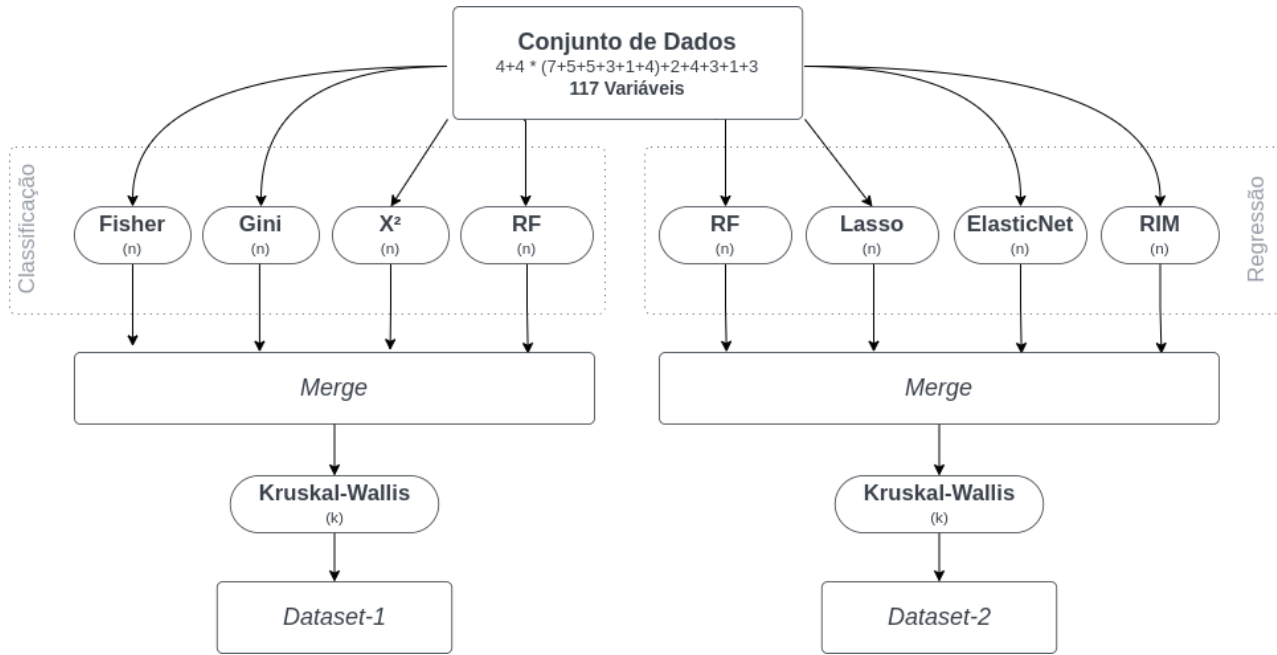
Como resultado desse processo, são geradas um total de 117 variáveis, o que proporciona uma variedade de padrões de entrada para os modelos de previsão.

## 3.2 Seleção de Variáveis

Com a base de dados gerada, é importante ressaltar que nem todas as variáveis possuem a mesma relevância para o processo de previsão, pois essa relevância pode depender tanto do objetivo do processo de previsão quanto dos modelos utilizados. Portanto, é necessário realizar a seleção de variáveis, a qual foi dividida em duas etapas neste estudo. Primeiramente, são empregados alguns métodos de seleção de variáveis descritos na Seção 2.6. Em seguida, as variáveis selecionadas são agrupadas e classificadas em dois conjuntos de dados distintos, denominados *Dataset-1* e *Dataset-2*. Uma representação geral desse processo pode ser observada na Figura 3.2.

Na primeira etapa do processo de seleção de variáveis, são aplicados dois grupos de algoritmos. O primeiro grupo é direcionado para modelos de classificação e inclui os métodos de Fisher, Gini,  $X^2$  e RF. Enquanto o segundo grupo é direcionado para modelos de regressão e inclui os métodos de RF, Lasso, ElasticNet e RIM.

Figura 3.2: Visão macro do processo de seleção de variáveis.



Cada um desses métodos, pertencentes aos dois grupos, seleciona  $n$  variáveis relevantes para a classificação ou regressão, respectivamente.

Por fim, na segunda etapa do processo de seleção de variáveis, é realizado um *merge* em cada grupo, unificando as variáveis selecionadas por cada método de seleção, sem que haja repetição. Em seguida, é aplicado o teste de *Kruskal-Wallis* em cada grupo para ranquear as variáveis mais relevantes provenientes do *merge*. Isso resulta em dois conjuntos de dados com  $k$  variáveis cada: o *Dataset-1*, referente ao grupo de classificação, e o *Dataset-2*, referente ao grupo de regressão. O uso do modelo de *Kruskal-Wallis* permite ordenar as variáveis de acordo com sua relevância, considerando as características específicas de cada grupo. Essa etapa visa consolidar os conjuntos de dados finais, contendo as variáveis mais importantes para cada tipo de tarefa, otimizando assim o processo de previsão.

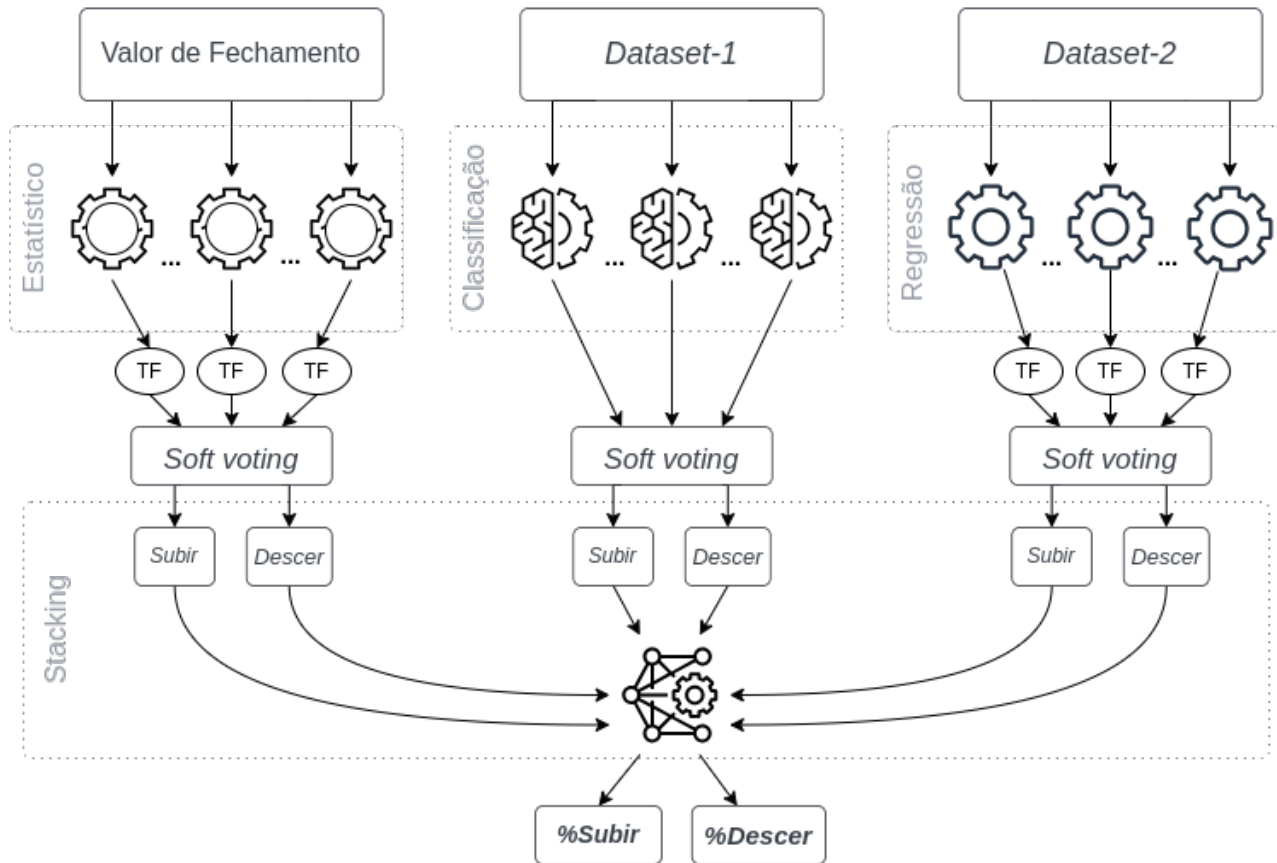
### 3.3 Máquina de Previsão

Para a construção da máquina de previsão proposta, foram empregadas duas técnicas de *ensemble*: o *Soft Voting* e o *Stacking*. A Figura 3.3 ilustra a forma como essas técnicas foram integradas.

O *Soft Voting*, conforme descrito por Wang et al. (2013), realiza uma votação ponderada, expressando a saída em percentuais correspondentes a cada classe. Nessa abordagem, diversos modelos são treinados de forma independente e suas previsões são combinadas usando pesos. O resultado final é obtido por meio de uma média ponderada das previsões de cada modelo, refletindo a confiança individual atribuída a cada classe. Já o método conhecido como *Stacking*, empregado por Dietterich et al. (2002), consiste na criação de camadas de modelos de predição interconectados. Nessa abordagem, os modelos de base são treinados coletivamente, e suas saídas são empregadas como entrada para um meta-modelo responsável pelo resultado final. O objetivo é combinar as previsões dos modelos de base de maneira a obter uma estimativa mais precisa e confiável.

Na abordagem proposta, são construídos três grupos de algoritmos distintos para a tarefa de previsão. O primeiro grupo consiste em modelos estatísticos, como por exemplo ARIMA, SARIMA e GARCH, que utilizam apenas o valor de fechamento de cada amostra como entrada e possuem uma função de transformação na saída de cada modelo, conforme descrita na Tabela 2.1. O segundo grupo é composto por técnicas de IA voltadas para classificação, como por exemplo SVM, LR e KNN, que recebem como entrada o *Dataset-1*, composto pelas variáveis relevantes selecionadas para tarefa de classificação. O terceiro grupo é formado por algoritmos de IA voltados para regressão, como por exemplo LSTM, RNN e CNN, que recebem como entrada o *Dataset-2*,

Figura 3.3: Visão macro do processo de previsão.



contendo as variáveis relevantes selecionadas para a tarefa de regressão e também possuem uma função de transformação na saída de cada modelo, de acordo com a Tabela 2.1.

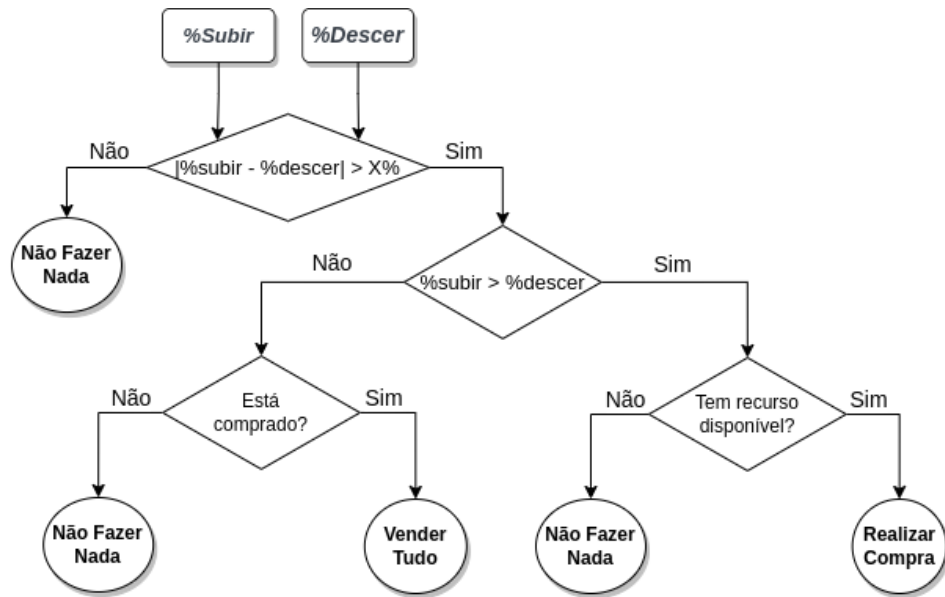
Em continuidade, as técnicas de *Soft Voting* e *Stacking* são aplicadas de forma sequencial. O *Soft Voting* combina as previsões de cada grupo de modelos, considerando a contribuição de cada um deles. Isso resulta em dois valores, sendo o grau de pertinência das classes 'subir' e 'descer'. Por sua vez, o *Stacking* utiliza um algoritmo de IA para combinar as saídas do *Soft Voting* de cada grupo, resultando em duas saídas referentes aos percentuais de correlação a cada classe. Por exemplo, considere três modelos de previsão pertencentes ao mesmo grupo, que prevêem as classes 1, 0 e 1, respectivamente. Utilizando um sistema de *soft voting*, as classes preditas pelos modelos são ponderadas, resultando em aproximadamente 66,5% de chance da próxima amostra pertencer à classe 1 e 33,5% de chance de pertencer à classe 0. Esse processo é repetido para cada conjunto de algoritmo: Estatístico, Classificação e Regressão. Em seguida, outro modelo genérico de IA recebe os percentuais de pertencimento de cada classe provenientes de cada conjunto de algoritmos, resultando em 6 entradas. Com base nessas 6 entradas, o modelo utilizado retorna dois valores: o percentual de pertencimento à classe 0 e 1.

### 3.4 Recomendação de Investimento

Tendo em mãos o percentual de subida ou descida obtido pela máquina de previsão proposta, inicia-se o processo de recomendação de compra e venda, conforme ilustrado na Figura 3.4, que objetiva reduzir operações de baixa confiabilidade e maximizar os retornos financeiros. A estratégia de recomendação é baseada nos resultados da máquina de previsão, visando identificar oportunidades de investimento com maior probabilidade de sucesso.

A estratégia de recomendação inicia com um filtro para verificar se a diferença percentual entre as classes de subida e descida é relevante. Isso é feito calculando a diferença nos graus de pertencimento a cada classe. Se a diferença não for significativa, ou seja, menor que um determinado valor X%, a recomendação é não realizar

Figura 3.4: Visão macro do processo de recomendação de investimentos.



nenhuma operação e manter a posição atual. Caso a diferença seja significativa, é feita uma comparação entre os sinais de 'subir' e 'descer' para determinar qual classe apresenta um percentual de pertinência maior. Isso ajuda a identificar a direção mais provável do próximo valor.

Se o sinal de subida for considerado mais relevante, ou seja, a classe 'subir' tem um percentual de pertinência maior, a estratégia sugere que a probabilidade de o próximo valor ser maior do que o atual é alta. Nesse caso, é verificado se há recurso disponível na carteira de investimentos. Se houver recurso disponível, a recomendação é realizar a compra do ativo. Caso contrário, não é feita nenhuma ação.

Por outro lado, se o sinal de subida não for superior ao de descida, é mais provável que o próximo valor seja menor do que o atual. Nesse cenário, é verificado se há algum ativo já comprado. Se houver ativos comprados, a recomendação é vender todos eles, indicando uma posição defensiva para evitar perdas potenciais. Caso não haja ativos comprados, nenhuma ação é recomendada.

# Experimentos Computacionais

Com base na abordagem metodológica apresentada no Capítulo 3, os experimentos computacionais foram conduzidos da seguinte maneira. Inicialmente, apresenta-se o processo de construção do conjunto de dados, conforme detalhado na Seção 4.1. Em seguida, na Seção 4.2, são definidos os modelos empregados na tarefa de previsão, bem como os ajustes de hiperparâmetros realizados. A Seção 4.3 discute como esses modelos serão avaliados e analisados. Por fim, os resultados experimentais são apresentados na Seção 4.4.

## 4.1 Conjunto de Dados

Para a realização dos experimentos computacionais, foram escolhidos três instrumentos financeiros: PETR3, WINFUT e WDOFUT. Amostras de cada ativo foram coletadas no período de 16/06/2021 até 16/06/2023 (2 anos), com frequências de 30 e 60 minutos. No entanto, vale ressaltar que, para o ativo WDOFUT com granularidade de 60 minutos, o intervalo considerado foi de 01/09/2021 até 01/09/2023, devido a limitações na plataforma utilizada para coleta. Essa abordagem resultou em um conjunto de dados consistente, totalizando 9156 amostras para a granularidade de 30 minutos e 4664 para a granularidade de 60 minutos. Posteriormente, foram criadas novas variáveis com base nos valores de OHLC, como descrito em detalhes na Seção 3.1.2. Essas variáveis foram então escolhidas para cada conjunto de modelos de previsão, conforme explicado na Seção 3.2, onde os parâmetros  $x$  e  $k$  foram definidos como 8 e 4, respectivamente.

Por fim, as seis bases de dados coletadas foram categorizadas cada uma em três segmentos distintos. Dessa forma, 10% da base de dados foi reservada para a otimização dos modelos de previsão, 70% destinou-se ao treinamento desses modelos, e os restantes 20% compuseram ao segmento de teste. No caso das bases de dados com granularidade de 30 minutos, essa distribuição compreendeu 916 amostras para otimização, 6409 para treinamento e 1831 para teste. Para as bases de dados com granularidade de 60 minutos, a distribuição foi de 466 amostras para otimização, 3265 para treinamento e 932 para teste. Essas alocações podem ser visualizadas nas figuras 4.1, 4.2 e 4.3, juntamente com a tendência de cada ativo e sua faixa de variação.

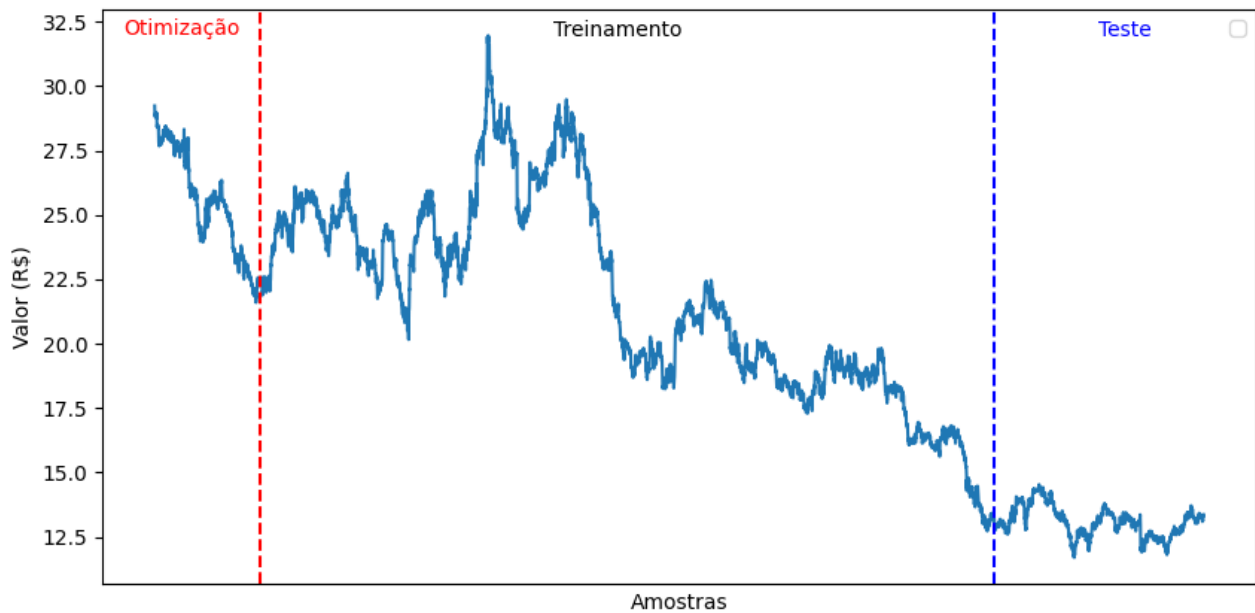
### 4.1.1 PETR3

No contexto do ativo financeiro PETR3, o processo de construção dos *datasets* (*dataset1* e *dataset2*) resultou em conjuntos distintos de variáveis, influenciados pelas diferentes granularidades presentes nas bases de dados. Dessa maneira, as variáveis que compõem cada conjunto de dados são as seguintes:

- ***dataset1* (30 minutos):** ADX com uma janela deslizante de 14, MACD obtido a partir do valor de fechamento com janelas deslizantes de 8 e 17 amostras, além de duas variáveis relacionadas ao ROC. Estas referem-se à derivada do valor máximo e do fechamento, ambas com janelas deslizantes de 10 amostras.
- ***dataset2* (30 minutos):** %K com uma janela deslizante de 8 amostras, TSI obtido a partir do valor de fechamento com janelas deslizantes de 13 e 25 amostras, e mais duas variáveis relacionadas ao %R com janelas deslizantes de 14 e 21 amostras.
- ***dataset1* (60 minutos):** ADX com uma janela deslizante de 14 amostras, ROC com janela deslizante de 10 amostras, calculado a partir do valor de fechamento, e duas variáveis relacionadas ao MACD, ambas derivadas do valor de fechamento. Uma delas com janelas deslizantes de 12 e 26 amostras, e a outra com uma janela deslizante de 8 e 17 amostras.

- **dataset2 (60 minutos):** ROC derivado do valor de abertura com uma janela deslizando de 12 amostras, %R com uma janela deslizando de 5 amostras, e duas variáveis associadas a %K, uma com janelas deslizando de 8 e 10 amostras.

Figura 4.1: Base de dados do ativo financeiro PETR3.



#### 4.1.2 WINFUT

Já no contexto do ativo financeiro WINFUT, o processo de construção dos *datasets* (*dataset1* e *dataset2*) também resultou em conjuntos distintos de variáveis. Dessa maneira, as variáveis que compõem cada conjunto de dados são as seguintes:

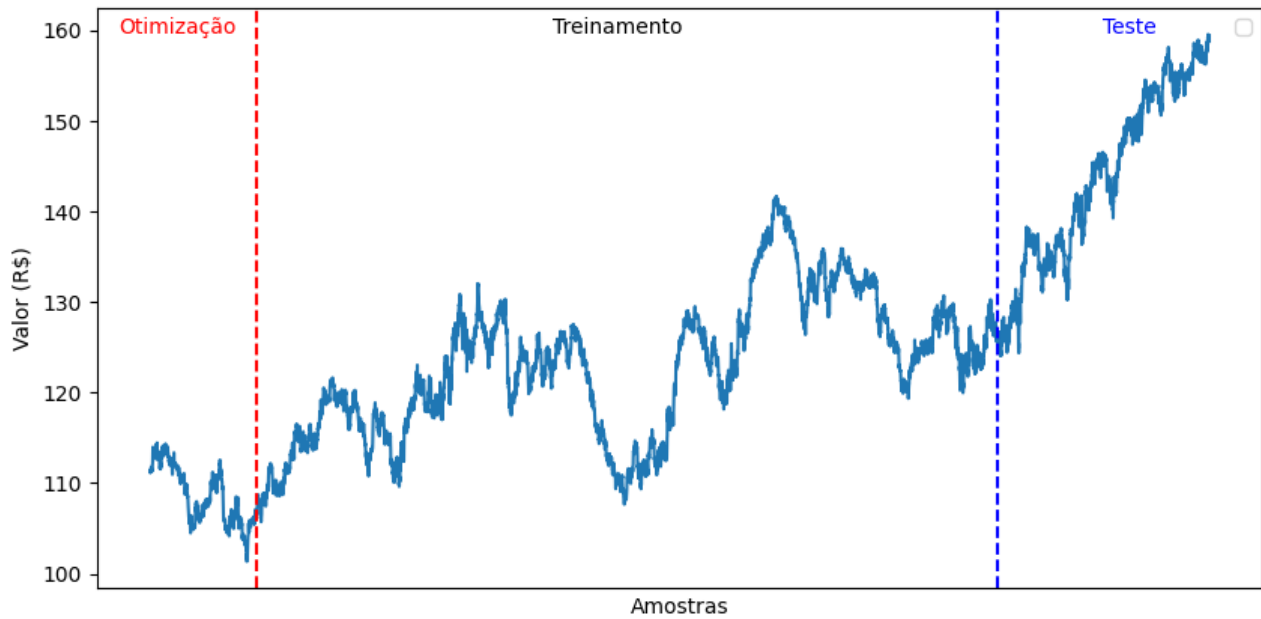
- **dataset1 (30 minutos):** ADX com uma janela deslizando de 14 amostras, MACD derivado do valor máximo com janelas deslizando de 12 e 26 amostras, %K com uma janela deslizando de 8 amostras, e CCI com uma janela deslizando de 18 amostras.
- **dataset2 (30 minutos):** %R com uma janela deslizando de 5 amostras, TSI derivado do valor de fechamento com janelas deslizando de 13 e 25 amostras, %K com uma janela deslizando de 8 amostras, e SMA com janela deslizando de 3 amostras.
- **dataset1 (60 minutos):** ADX com uma janela deslizando de 14 amostras, CCI com uma janela deslizando de 18 amostras, MACD derivado do valor máximo com janelas deslizando de 12 e 26 amostras, e %K com uma janela deslizando de 8 amostras.
- **dataset2 (60 minutos):** permanecem as variáveis do *dataset1* de 30 minutos, com exceção da variável %K, que neste *dataset* possui uma janela deslizando de 14 amostras.

#### 4.1.3 WDOFUT

Por fim, para o ativo financeiro WDOFUT, o processo de construção dos *datasets* (*dataset1* e *dataset2*) também conduziu à formação de conjuntos únicos de variáveis. Desse modo, as variáveis que integram cada conjunto de dados são as seguintes:

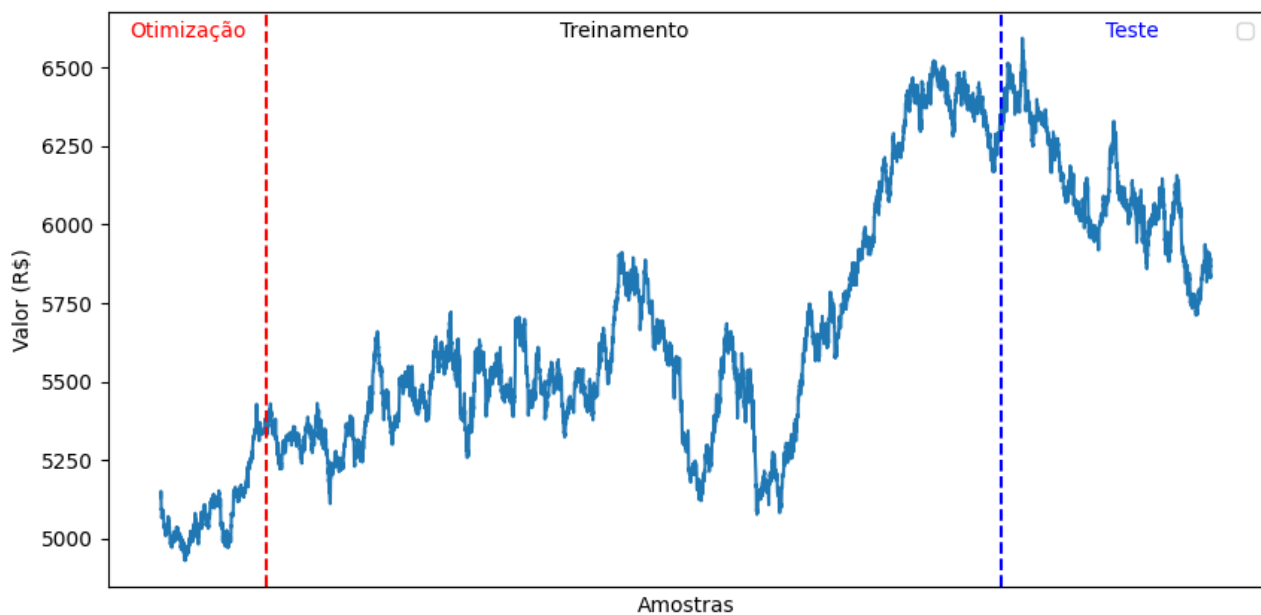
- **dataset1 (30 minutos):** ADX com uma janela deslizando de 14 amostras, MACD calculado a partir do valor máximo com janelas deslizando de 12 e 26 amostras, %K com uma janela deslizando de 8 amostras, e CCI com uma janela deslizando de 18 amostras.

Figura 4.2: Base de dados do ativo financeiro WINFUT.



- **dataset2 (30 minutos):** ADX com uma janela deslizando de 14 amostras, ROC obtida a partir do valor máximo com uma janela móvel de 10 amostras, e duas variáveis de MACD. Ambas são derivadas do valor máximo, sendo uma com janelas deslizantes de 8 e 17 amostras, e a outra com janelas deslizantes de 12 e 26 amostras.
- **dataset1 (60 minutos):** ADX com uma janela deslizando de 14 amostras, CCI com uma janela deslizando de 18 amostras, MACD obtido a partir do valor máximo com janelas deslizantes de 12 e 26 amostras, e %K com uma janela deslizando de 8 amostras.
- **dataset2 (60 minutos):** ADX com uma janela deslizando de 14 amostras, ROC calculada a partir do valor máximo com uma janela deslizando de 12 amostras, %K com uma janela deslizando de 10 amostras, e MACD derivado do valor máximo com janelas deslizantes de 12 e 26 amostras.

Figura 4.3: Base de dados do ativo financeiro WDOFUT.



## 4.2 Modelos e Hiperparâmetros

Para efetivar a proposta delineada na Seção 3.3, foram empregados 10 modelos distintos de previsão. No grupo de modelos estatísticos, destacam-se três escolhas específicas: ARIMA, SARIMA, e GARCH. O conjunto de modelos de classificação compreende outros três: SVM, KNN, e LR, enquanto o conjunto de modelos de regressão é representado por MLP, SVR, e RF. Além disso, concebeu-se um modelo final, uma MLP denominada "MLP OUT", cuja função é consolidar os resultados provenientes de cada ensemble dos três conjuntos mencionados anteriormente. A Tabela 5.1 no Apêndice A detalha a estrutura de cada modelo de previsão, juntamente com o processo de otimização implementado e as escolhas otimizadas para cada conjunto de dados coletado.

## 4.3 Experimentos e Métricas

As métricas de avaliação desempenham um papel crucial na análise metódica dos modelos propostos, sendo aplicadas a duas categorias distintas: análise de classificação e análise de recomendação de investimento. Na fase de análise de classificação, todos os modelos definidos são considerados, uma vez que a função de transformação, que converte resultados de regressão em classificação, é aplicada uniformemente a todos os modelos de regressão. Para a avaliação desses modelos, serão utilizadas métricas essenciais, como Acurácia, F1, e Matriz de Confusão. A Acurácia proporciona uma visão abrangente da precisão do modelo, enquanto a métrica F1 equilibra precisão e *recall*, garantindo uma avaliação robusta. A Matriz de Confusão oferece uma análise mais detalhada do desempenho do modelo em diferentes categorias, proporcionando uma compreensão completa de seu comportamento em situações de classificação.

Já na análise de recomendação de investimento, todos os modelos e ensembles definidos durante o experimento são avaliados com base na estratégia delineada na Seção 3.4. Isso permite a comparação do percentual de retorno de cada modelo ao final da execução, juntamente com a quantidade de operações de compra realizadas. Importante destacar que, nessa fase, foram incorporados resultados da estratégia *buy and hold* para efeitos de comparação. Essa análise final visa proporcionar *insights* valiosos sobre a eficácia prática de cada abordagem na execução da estratégia proposta.

## 4.4 Resultados Experimentais

Nesta seção, serão detalhados os resultados provenientes dos experimentos computacionais realizados para cada conjunto de dados, acompanhados de uma análise específica destinada a cada conjunto. Essa abordagem visa oferecer uma compreensão aprofundada das conclusões derivadas da aplicação prática dos modelos em cada cenário, contribuindo para uma interpretação robusta e informada dos resultados alcançados.

### 4.4.1 PETR3 (30 Minutos)

Os resultados derivados da análise da base de dados PETR3, com uma granularidade de 30 minutos, são apresentados nas seções subsequentes. Inicialmente, examina-se o desempenho individual de cada modelo no contexto de classificação, seguido por uma análise detalhada no cenário de recomendação de investimento.

#### Classificação

Através da Tabela 4.1, observa-se que, para esta base de dados, os modelos de classificação alcançaram os melhores resultados, destacando-se em termos de acurácia e F1, com o modelo LR liderando. Uma consideração importante surge ao analisar a matriz de confusão do modelo ARIMA, revelando um viés em suas classificações, comprometendo sua capacidade de identificar quedas futuras nos preços. Em síntese, o modelo de saída (MLP OUT) da abordagem descrita na Seção 3.3 demonstrou um desempenho notável, assemelhando-se ao conjunto de classificação.



Tabela 4.1: Resultados da Classificação (PETR3 30 minutos)

Modelo	Acurácia	F1-score	True Positives	True Negatives	False Positives	False Negatives
SVM	0,642	0,604	494	367	107	374
KNN	0,636	0,631	436	418	165	323
LR	0,648	0,667	398	472	203	269
MLP	0,477	0,492	300	340	301	401
SVR	0,495	0,516	302	362	299	379
RF	0,499	0,523	301	369	300	372
ARIMA	0,447	0,008	597	3	4	738
SARIMA	0,474	0,484	305	331	296	410
GARCH	0,504	0,564	245	431	356	310
Ensamble Classificação	0,636	0,631	436	418	165	323
Ensamble Regressão	0,477	0,492	300	340	301	401
Ensamble Estatístico	0,504	0,564	245	431	356	310
MLP OUT	0,636	0,631	436	418	165	323

### Recomendação de Investimento

No âmbito da recomendação de investimento, a análise da Tabela 4.2 revela que uma elevada acurácia nem sempre se traduz em resultados financeiros vantajosos. Destaca-se que o ARIMA alcançou o melhor desempenho, proporcionando uma variação percentual de 19,087% a mais em relação à estratégia *buy and hold*. Um resultado adicional de interesse é que a máquina de previsão proposta alcançou uma margem de 8,094% superior à estratégia *buy and hold*.

Tabela 4.2: Resultados das Compras e Variação Financeira (PETR3 30 minutos)

Modelo	Compras Realizadas	Valor Inicial	Valor Final	% Variação
SVM	275	1.000,00	886,37	-11,363
KNN	296	1.000,00	894,17	-10,583
LR	310	1.000,00	846,62	-15,338
MLP	412	1.000,00	728,40	-27,160
SVR	254	1.000,00	948,59	-5,141
RF	390	1.000,00	844,31	-15,569
ARIMA	7	1.000,00	1.009,41	0,941
SARIMA	338	1.000,00	844,08	-15,592
GARCH	36	1.000,00	876,31	-12,369
Ensamble Classificação	296	1.000,00	894,17	-10,583
Ensamble Regressão	412	1.000,00	728,40	-27,160
Ensamble Estatístico	36	1.000,00	876,31	-12,369
MLP OUT	296	1.000,00	894,17	-10,583
Buy and Hold	1	1.000,00	813,23	-18,677

#### 4.4.2 PETR3 (60 Minutos)

Os resultados provenientes da avaliação da base de dados PETR3, considerando uma granularidade de 60 minutos, estão expostos nas seções seguintes. Primeiramente, é realizada uma análise do desempenho singular de cada modelo no âmbito da classificação, seguida por uma exploração minuciosa no contexto da recomendação de investimento.

### Classificação

Ao analisar a Tabela 4.3, observa-se que os modelos baseados em técnicas de regressão alcançaram o melhor resultado médio em termos de acurácia na tarefa de classificação. Como consequência, o *ensemble* desse conjunto obteve os melhores resultados quando comparado com os outros *ensembles*.

Tabela 4.3: Resultados da Classificação (PETR3 60 minutos)

Modelo	Acurácia	F1-score	True Positives	True Negatives	False Positives	False Negatives
SVM	0,485	0,261	269	62	67	285
KNN	0,511	0,510	175	174	161	173
LR	0,474	0,426	191	133	145	214
MLP	0,534	0,531	185	180	151	167
SVR	0,499	0,500	170	171	166	176
RF	0,524	0,519	183	175	153	172
ARIMA	0,489	0,558	114	220	222	127
SARIMA	0,476	0,475	163	162	173	185
GARCH	0,502	0,558	128	215	208	132
Ensamble Classificação	0,511	0,510	175	174	161	173
Ensamble Regressão	0,531	0,452	231	132	105	215
Ensamble Estatístico	0,476	0,475	163	162	173	185
MLP OUT	0,511	0,510	175	174	161	173

### Recomendação de Investimento

Quanto à recomendação de investimento, ao analisar a Tabela 4.4, destaca-se que o KNN apresentou os resultados mais promissores, registrando uma variação percentual de 40,508 pontos superior à estratégia *buy and hold*. Esse desempenho positivo é compartilhado pelos resultados do *ensemble* de classificação e do MLP OUT, que seguem os mesmos padrões de variação.

Tabela 4.4: Resultados das Compras e Variação Financeira (PETR3 60 minutos)

Modelo	Compras Realizadas	Valor Inicial	Valor Final	% Variação
SVM	24	1.000,00	1.051,72	5,172
KNN	152	1.000,00	1.208,39	20,839
LR	37	1.000,00	905,28	-9,472
MLP	218	1.000,00	958,34	-4,166
SVR	185	1.000,00	972,77	-2,723
RF	220	1.000,00	837,75	-16,225
ARIMA	9	1.000,00	786,13	-21,387
SARIMA	174	1.000,00	948,66	-5,134
GARCH	23	1.000,00	818,98	-18,102
Ensamble Classificação	152	1.000,00	1.208,39	20,839
Ensamble Regressão	188	1.000,00	919,21	-8,079
Ensamble Estatístico	174	1.000,00	948,66	-5,134
MLP OUT	152	1.000,00	1.208,39	20,839
Buy and Hold	1	1.000,00	803,31	-19,669

### 4.4.3 WINFUT (30 Minutos)

Os resultados decorrentes da análise da base de dados WINFUT, considerando uma granularidade de 30 minutos, são apresentados nas seções subsequentes. Inicialmente, efetua-se uma análise do desempenho individual de cada modelo no contexto de classificação, seguida por uma exploração minuciosa na esfera da recomendação de investimento.

#### Classificação

Ao examinar a Tabela 4.5, percebe-se que a maioria dos resultados de classificação assume características de modelos aleatórios, com muitos deles apresentando acurácia inferior a 50%. Nesse cenário, destaca-se que os modelos de classificação foram os que alcançaram os melhores resultados, tanto em termos de acurácia quanto de F1. Um aspecto relevante revelado pela matriz de confusão é o viés presente no modelo ARIMA, que demonstra uma limitação em prever variações negativas.

Tabela 4.5: Resultados da Classificação (WINFUT 30 minutos)

Modelo	Acurácia	F1-score	True Positives	True Negatives	False Positives	False Negatives
SVM	0,509	0,431	519	300	254	537
KNN	0,514	0,516	410	417	363	420
LR	0,512	0,496	437	387	336	450
MLP	0,470	0,475	371	386	402	451
SVR	0,506	0,520	382	432	391	405
RF	0,489	0,504	371	417	402	420
ARIMA	0,479	0,005	769	2	4	835
SARIMA	0,483	0,488	380	397	393	440
GARCH	0,515	0,455	503	326	270	511
Ensamble Classificação	0,514	0,516	410	417	363	420
Ensamble Regressão	0,465	0,393	469	279	304	558
Ensamble Estatístico	0,481	0,005	772	2	1	835
MLP OUT	0,514	0,516	410	417	363	420

#### Recomendação de Investimento

Do ponto de vista da recomendação de investimento, ao analisar a Tabela 4.6, destaca-se que o modelo SVR apresentou o melhor desempenho, registrando uma variação de 9,205% a mais no valor investido em comparação com a estratégia *buy and hold*. Esse fato contribuiu para que a saída da máquina de previsão (MLP OUT) alcançasse um resultado de 0,167% a mais na variação do valor investido em relação à estratégia *buy and hold*.

Tabela 4.6: Resultados das Compras e Variação Financeira (WINFUT 30 minutos)

Modelo	Compras Realizadas	Valor Inicial	Valor Final	% Variação
SVM	135	1.000,00	956,87	-4,313
KNN	397	1.000,00	1.180,38	18,038
LR	101	1.000,00	1.043,13	4,313
MLP	438	1.000,00	1.205,78	20,578
SVR	392	1.000,00	1.270,76	27,076
RF	492	1.000,00	1.002,91	0,291

Continua na próxima página

ARIMA	6	1.000,00	1.009,85	0,985
SARIMA	417	1.000,00	1.083,59	8,359
GARCH	65	1.000,00	1.221,47	22,147
Ensamble Classificação	397	1.000,00	1.180,38	18,038
Ensamble Regressão	382	1.000,00	1.115,02	11,502
Ensamble Estatístico	3	1.000,00	1.009,85	0,985
MLP OUT	397	1.000,00	1.180,38	18,038
Buy and Hold	1	1.000,00	1.178,71	17,871

#### 4.4.4 WINFUT (60 Minutos)

Os resultados provenientes da avaliação da base de dados WINFUT, com uma granularidade de 60 minutos, estão delineados nas seções que se seguem. Inicialmente, procede-se à análise do desempenho específico de cada modelo no domínio da classificação, seguida por uma exploração meticulosa no contexto da recomendação de investimento.

##### Classificação

Mediante a análise da Tabela 4.7, torna-se factível observar os resultados obtidos na fase de classificação, revelando desempenhos similares entre os modelos nas métricas analisadas.

Tabela 4.7: Resultados da Classificação (WINFUT 60 minutos)

Modelo	Acurácia	F1-score	True Positives	True Negatives	False Positives	False Negatives
SVM	0,501	0,599	103	299	277	123
KNN	0,519	0,558	172	244	208	178
LR	0,516	0,544	183	231	197	191
MLP	0,509	0,520	195	213	185	209
SVR	0,525	0,536	201	220	179	202
RF	0,480	0,493	182	203	198	219
ARIMA	0,507	0,517	196	211	184	211
SARIMA	0,512	0,515	203	208	177	214
GARCH	0,521	0,435	270	148	110	274
Ensamble Classificação	0,519	0,558	172	244	208	178
Ensamble Regressão	0,509	0,520	195	213	185	209
Ensamble Estatístico	0,507	0,517	196	211	184	211
MLP OUT	0,519	0,558	172	244	208	178

##### Recomendação de Investimento

Na análise do desempenho dos modelos no contexto da recomendação de investimento, a Tabela 4.8 apresenta os resultados dessa tarefa. Destaca-se que o modelo KNN alcançou os resultados mais expressivos, superando o *buy and hold* em 5.079 pontos percentuais na variação do investimento. Essa tendência positiva é compartilhada pelo *ensemble* de classificação e pelo MLP OUT, que atingem padrões semelhantes de variação.

Tabela 4.8: Resultados das Compras e Variação Financeira (WINFUT 60 minutos)

Modelo	Compras Realizadas	Valor Inicial	Valor Final	% Variação
SVM	54	1.000,00	1.088,75	8,875
KNN	175	1.000,00	1.233,30	23,330
LR	53	1.000,00	1.014,04	1,404
MLP	217	1.000,00	1.039,00	3,900
SVR	202	1.000,00	1.011,59	1,159
RF	249	1.000,00	1.132,08	13,208
ARIMA	209	1.000,00	1.073,09	7,309
SARIMA	194	1.000,00	1.031,59	3,159
GARCH	39	1.000,00	1.139,51	13,951
Ensamble Classificação	175	1.000,00	1.233,30	23,330
Ensamble Regressão	217	1.000,00	1.039,00	3,900
Ensamble Estatístico	209	1.000,00	1.073,09	7,309
MLP OUT	175	1.000,00	1.233,30	23,330
Buy and Hold	1	1.000,00	1.182,51	18,251

#### 4.4.5 WDOFUT (30 Minutos)

Os resultados originados da análise da base de dados WDOFUT, considerando uma granularidade de 30 minutos, estão expostos nas seções seguintes. Primeiramente, realiza-se a avaliação do desempenho particular de cada modelo no âmbito da classificação, seguida por uma exploração minuciosa no contexto da recomendação de investimento.

##### Classificação

Na análise apresentada na Tabela 4.9, os resultados da fase de classificação são revelados. Estes resultados sugerem características aleatórias no sistema de previsão, uma vez que todos os modelos atingiram acurácias próximas a 50%. Um ponto adicional a ser considerado é o viés potencial no modelo ARIMA, o que impede sua capacidade de antecipar quedas nos valores.

Tabela 4.9: Resultados da Classificação (WDOFUT 30 minutos)

Modelo	Acurácia	F1-score	True Positives	True Negatives	False Positives	False Negatives
SVM	0,503	0,615	172	638	640	160
KNN	0,500	0,517	375	430	437	368
LR	0,515	0,557	338	491	474	307
MLP	0,547	0,551	432	448	380	350
SVR	0,511	0,498	432	391	380	407
RF	0,532	0,516	456	401	356	397
ARIMA	0,504	0,000	812	0	0	798
SARIMA	0,516	0,520	407	423	405	375
GARCH	0,507	0,559	314	503	498	295
Ensamble Classificação	0,500	0,517	375	430	437	368
Ensamble Regressão	0,547	0,551	432	448	380	350
Ensamble Estatístico	0,504	0,000	812	0	0	798

Continua na próxima página

MLP OUT	0,500	0,517	375	430	437	368
---------	-------	-------	-----	-----	-----	-----

### Recomendação de Investimento

No âmbito da recomendação de investimento, a Tabela 4.10 destaca os resultados pertinentes a essa etapa. Observa-se que o modelo SVR obteve os melhores resultados, superando o *buy and hold* em 7,108% na variação percentual. Outro destaque relevante é o desempenho da saída da máquina de previsão (MLP OUT), que registrou uma diferença de 4,233% em relação aos resultados do *buy and hold* na variação percentual do valor investido.

Tabela 4.10: Resultados das Compras e Variação Financeira(WDOFUT 30 minutos)

Modelo	Compras Realizadas	Valor Inicial	Valor Final	% Variação
SVM	170	1.000,00	882,15	-11,785
KNN	382	1.000,00	978,12	-2,188
LR	191	1.000,00	869,55	-13,045
MLP	318	1.000,00	999,65	-0,035
SVR	260	1.000,00	1.006,87	0,687
RF	395	1.000,00	985,25	-1,475
ARIMA	0	1.000,00	1.000,00	0,000
SARIMA	427	1.000,00	947,32	-5,268
GARCH	61	1.000,00	1.048,91	4,891
Ensamble Classificação	382	1.000,00	978,12	-2,188
Ensamble Regressão	318	1.000,00	999,65	-0,035
Ensamble Estatístico	0	1.000,00	1.000,00	0,000
MLP OUT	382	1.000,00	978,12	-2,188
Buy and Hold	1	1.000,00	935,79	-6,421

#### 4.4.6 WDOFUT (60 Minutos)

Os resultados da análise da base de dados WDOFUT, com um intervalo de 60 minutos, estão apresentados nas seções a seguir. Primeiramente, é feita uma análise do desempenho de cada modelo no contexto da classificação, seguida por uma exploração cuidadosa no contexto da recomendação de investimento.

#### Classificação

Os resultados da fase de classificação são apresentados na Tabela 4.11, evidenciando os melhores desempenhos em acurácia e F1 para o conjunto de modelos de classificação. Por outro lado, ao analisar a matriz de confusão, é possível identificar que o modelo ARIMA está enviesado, perdendo, assim, a capacidade de antecipar possíveis quedas nos preços.

Tabela 4.11: Resultados da Classificação (WDOFUT 60 minutos)

Modelo	Acurácia	F1-score	True Positives	True Negatives	False Positives	False Negatives
SVM	0,551	0,644	116	326	276	84
KNN	0,520	0,529	201	216	191	194
LR	0,559	0,547	234	214	158	196
MLP	0,509	0,511	202	206	190	204

Continua na próxima página

SVR	0,475	0,468	196	185	196	225
RF	0,483	0,478	197	190	195	220
ARIMA	0,489	0,000	392	0	0	410
SARIMA	0,484	0,485	193	195	199	215
GARCH	0,491	0,400	258	136	134	274
Ensamble Classificação	0,520	0,529	201	216	191	194
Ensamble Regressão	0,483	0,478	197	190	195	220
Ensamble Estatístico	0,491	0,400	258	136	134	274
MLP OUT	0,520	0,529	201	216	191	194

### Recomendação de Investimento

Na etapa de recomendação de investimentos, os resultados são apresentados na Tabela 4.12. O modelo GARCH apresentou o melhor desempenho na variação percentual dentre os modelos explorados, com uma diferença de -0,887% em relação à estratégia *buy and hold*. Outro resultado notável foi o da máquina de previsão (MLP OUT), que apresentou uma diferença de -3,065% na variação percentual em relação à estratégia *buy and hold*.

Tabela 4.12: Resultados das Compras e Variação Financeira (WDOFUT 60 minutos)

Modelo	Compras Realizadas	Valor Inicial	Valor Final	% Variação
SVM	75	1.000,00	1.032,14	3,214
KNN	183	1.000,00	1.069,35	6,935
LR	81	1.000,00	1.055,00	5,500
MLP	201	1.000,00	991,72	-0,828
SVR	143	1.000,00	1.044,30	4,430
RF	202	1.000,00	1.058,17	5,817
ARIMA	0	1.000,00	1.000,00	0,000
SARIMA	201	1.000,00	988,82	-1,118
GARCH	29	1.000,00	1.091,13	9,113
Ensamble Classificação	183	1.000,00	1.069,35	6,935
Ensamble Regressão	202	1.000,00	1.058,17	5,817
Ensamble Estatístico	29	1.000,00	1.091,13	9,113
MLP OUT	183	1.000,00	1.069,35	6,935
Buy and Hold	1	1.000,00	1.100,00	10,000

---

# Conclusão

---

O cerne deste trabalho consistiu na aplicação de técnicas de inteligência artificial no contexto do mercado financeiro. Para alcançar esse objetivo, a pesquisa teve início com a apresentação dos conceitos fundamentais do mercado financeiro, percorrendo desde sua origem até análises mercadológicas de investimento. Além disso, uma revisão detalhada da literatura foi realizada, explorando as técnicas mais relevantes e contemporâneas para a previsão de ativos financeiros. Com base nesse embasamento teórico robusto, a proposta central do trabalho envolveu o desenvolvimento de uma máquina de previsão utilizando abordagens de *ensemble*. Essa máquina é alimentada com diversas métricas econômicas e tem a capacidade de fornecer a probabilidade de um determinado ativo financeiro apresentar movimento de alta ou baixa. Posteriormente, conduziu-se um processo de recomendação, visando maximizar os ganhos em cada operação, integrando assim efetivamente a análise preditiva com a tomada de decisões estratégicas.

Foram conduzidos experimentos com uma variedade de dez algoritmos de previsão, distribuídos em três categorias principais: classificação, representada pelos modelos KNN, SVM e LR; modelos estatísticos, incluindo ARIMA, SARIMA e GARCH; e modelos de regressão, como MLP, SVR e RF. Os resultados de cada categoria foram combinados usando uma abordagem de *ensemble*, e, por fim, um modelo de MLP foi empregado na saída da máquina de previsão. Os resultados desses algoritmos indicaram claramente a viabilidade de alcançar lucros substanciais. No entanto, é notável que a máquina de previsão proposta não se destacou como a melhor em todos os testes; no entanto, obteve resultados significativos que superaram a estratégia de *buy and hold* em 83,33% das instâncias. Um achado interessante é que os modelos que alcançaram a melhor acurácia e F1 não necessariamente alcançaram os melhores resultados na fase de recomendação de investimento, sugerindo que essas métricas podem não ser as mais apropriadas para o cenário específico de investimentos. Essa observação destaca a necessidade de considerar métricas mais alinhadas com os objetivos práticos de maximização de ganhos durante a tomada de decisões de investimento.

Para futuras melhorias deste trabalho, recomenda-se refinamentos no cálculo do erro no modelo de saída (MLP OUT), atribuindo maior peso às classificações corretas mais impactantes, especialmente aquelas associadas a maiores variações. Outra abordagem a ser considerada seria a substituição de modelos que não alcançaram resultados satisfatórios durante os experimentos, visando aprimorar a eficácia da máquina de previsão. Além disso, uma extensão interessante seria a exploração de uma classe de modelos evolutivos para avaliar o impacto do aprendizado contínuo nesse contexto específico. A incorporação de abordagens evolutivas poderia proporcionar insights valiosos sobre a adaptabilidade da máquina de previsão em face de mudanças dinâmicas no mercado financeiro.

Todo o código utilizado para realizar esses experimentos está disponível no Github<sup>1</sup> em um repositório denominado "TCC-AI". Essa disponibilidade facilita a replicação dos experimentos, permitindo a análise detalhada e a validação dos resultados por outros pesquisadores e profissionais interessados no campo da previsão financeira utilizando técnicas de inteligência artificial.

## Apêndice A - Tabela de Otimização

---

<sup>1</sup><https://github.com/Sr-Souza-dev>



Tabela 5.1: Hiperparâmetros dos Modelos

Modelo	Hiperparâmetro	Variação	PETR3 30m	PETR3 60m	WINFUT 30m	WINFUT 60m	WDOFUT 30m	WDOFUT 60m
SVM	kernel	linear e poly	poly	poly	poly	poly	poly	poly
	C	0 a 40; step = 15	15	15	15	15	30	30
	gamma	scale e auto	scale	scale	scale	scale	scale	scale
	degree	1, 2, 3 e 4	2	2	3	3	3	2
	coef0	1, 3 e 5	1	3	1	1	3	1
KNN	n neighbors	3, 9, 17 e 34	34	17	34	34	17	9
	weights	uniform e distance	uniform	distance	distance	uniform	uniform	uniform
	algorithm	auto, ball_tree e kd_tree	auto	auto	auto	auto	auto	auto
	p	1 e 2	1	1	1	1	1	1
LR	penalty	L1 e L2	L2	L2	L2	L2	L2	L2
	solver	newton- cg e lbfgs	newton- cg	newton- cg	newton- cg	lbfgs	newton- cg	lbfgs
	C	10, 100 e 1000	10	10	10	10	10	10
	max iter	50, 100 e 500	50	50	50	50	50	50

Continua na próxima página

	fit intercept	True e False	True	False	False	True	True	True
	class weight	balanced e None	None	balanced	None	balanced	balanced	balanced
	warm start	True e False	True	True	True	True	True	True
MLP	hidden layer	(50,50,50), (30,80,50,20) e (100,)	(50, 50, 50)	(30, 80, 50, 20)	(50, 50, 50)	(100,)	(50, 50, 50)	(50, 50, 50)
	activation	tanh, relu e logistic	relu	relu	logistic	logistic	logistic	logistic
	solver	sgd, adam e lbfgs	adam	adam	adam	adam	lbfgs	lbfgs
	alpha	0.0001 e 0.05	0.0001	0.05	0.05	0.0001	0.05	0.05
	learning rate	constant e adaptive	adaptive	adaptive	adaptive	constant	constant	constant
SVR	kernel	linear e poly	poly	poly	linear	linear	poly	linear
	C	0, 15 e 30	0.0	30.0	0.01	0.01	30.0	15.0
	gamma	scale	scale	scale	scale	scale	scale	scale
	degree	1, 2 e 3	2	2	1	1	3	1
	coef0	1 e 3	3	3	1	1	3	1
RF	n estimators	100, 200, 300 e 1000	200	100	300	1000	200	300
	max features	auto e sqrt	sqrt	sqrt	sqrt	sqrt	sqrt	sqrt

Continua na próxima página

	max depth	10, 20, 30, 40 e 50	10	20	10	10	10	10
	min samples split	2, 5 e 10	5	5	10	10	10	5
	min samples leaf	1, 2 e 4	4	4	1	4	2	4
	bootstrap	True e False	True	True	True	True	True	True
ARIMA	autoregressive (p)	0, 1 e 2	0	1	0	1	0	0
	diferencial (d)	0, 1 e 2	1	0	1	1	1	1
	media (q)	0, 1 e 2	0	0	0	1	0	0
SARIMA	autoregressive (p)	0, 1 e 2	0	0	0	0	0	1
	diferencial (d)	0, 1 e 2	1	1	1	1	1	0
	media (q)	0, 1 e 2	0	0	0	1	0	0
	sazonalidade (P)	1 e 2	1	1	1	1	1	1
	diferencial (D)	0 e 1	0	0	0	0	0	0
	media (Q)	1 e 2	1	1	1	1	1	1
	periodo (S)	5 e 12	12	5	5	5	5	5
GARCH	dists	normal, r e gaussian	t	normal	t	normal	normal	normal

Continua na próxima página

	autoregressive (p)	1, 2, 3 e 4	3	1	1	3	3	1
	media (q)	1, 2, 3 e 4	1	1	1	3	2	3
MLP OUT	layer 1	16, 32, 48 e 64	16	16	32	16	32	32
	layer 2	16, 32, 48 e 64	32	32	48	32	64	32
	layer 3	16, 32, 48 e 64	32	48	48	64	32	64
	layer 4	16, 32, 48 e 64	16	32	32	16	16	32

# Referências Bibliográficas

---

- AGRAWAL, M. et al. Stock prediction based on technical indicators using deep learning model. *Computers, Materials & Continua*, v. 70, n. 1, p. 287–304, 2022.
- ALMEIDA, R. de; REYNOSO-MEZA, G.; STEINER, M. T. A. Multi-objective optimization approach to stock market technical indicators. In: IEEE. *2016 IEEE congress on evolutionary computation (CEC)*. [S.l.], 2016. p. 3670–3677.
- ALTAN, A.; KARASU, S. The effect of kernel values in support vector machine to forecasting performance of financial time series. *The Journal of Cognitive Systems*, v. 4, n. 1, p. 17–21, 2019.
- AMARAL, V. L. do. *Sistemas Fuzzy Evolutivos na Previsão e Recomendação de Investimentos em Criptomoedas*. 128 p. Dissertação (Mestrado) — Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte - MG, 2021.
- AMINI, F.; HU, G. A two-layer feature selection method using genetic algorithm and elastic net. *Expert Systems with Applications*, Elsevier, v. 166, p. 114072, 2021.
- AMINIMEHR, A. et al. A comprehensive study of market prediction from efficient market hypothesis up to late intelligent market prediction approaches. *Computational Economics*, Springer, v. 60, n. 2, p. 781–815, 2022.
- AMPOMAH, E. K.; QIN, Z.; NYAME, G. Evaluation of tree-based ensemble machine learning models in predicting stock price direction of movement. *Information*, MDPI, v. 11, n. 6, p. 332, 2020.
- ANAND, C. Comparison of stock price prediction models using pre-trained neural networks. *Journal of Ubiquitous Computing and Communication Technologies (UCCT)*, v. 3, n. 02, p. 122–134, 2021.
- ANANTHI, M.; VIJAYAKUMAR, K. Retracted article: stock market analysis using candlestick regression and market trend prediction (ckrm). *Journal of Ambient Intelligence and Humanized Computing*, Springer, v. 12, n. 5, p. 4819–4826, 2021.
- ANBALAGAN, T.; MAHESWARI, S. U. Classification and prediction of stock market index based on fuzzy metagraph. *Procedia Computer Science*, v. 47, p. 214–221, 2015. ISSN 1877-0509.
- ANWAR, M. T.; RAHMAN, S. *Forecasting stock market prices using advanced tools of machine learning*. Tese (Doutorado) — Brac University, 2019.
- BEN-DAVID, I.; FRANZONI, F.; MOUSSAWI, R. Exchange-traded funds. *Annual Review of Financial Economics*, Annual Reviews, v. 9, p. 169–189, 2017.
- BOGLE, J. C. *Common sense on mutual funds: New imperatives for the intelligent investor*. [S.l.]: John Wiley & Sons, 1999.
- BOLLEN, J.; MAO, H.; ZENG, X. Twitter mood predicts the stock market. *Journal of computational science*, Elsevier, v. 2, n. 1, p. 1–8, 2011.
- BOLSA DE VALORES BRASILEIRA (B3). *Horários de negociação no mercado de ações (mercado de Bolsa)*. 2023. <[https://www.b3.com.br/pt\\_br/solucoes/plataformas/puma-trading-system/para-participantes-e-traders/horario-de-negociacao/acoes/](https://www.b3.com.br/pt_br/solucoes/plataformas/puma-trading-system/para-participantes-e-traders/horario-de-negociacao/acoes/)>.
- BOLSA DE VALORES BRASILEIRA (B3). *Número de investidores na B3 cresce mesmo em cenário de alta volatilidade*. 2023. <[https://www.b3.com.br/pt\\_br/noticias/numero-de-investidores-na-b3-cresce-mesmo-em-cenario-de-alta-volatilidade](https://www.b3.com.br/pt_br/noticias/numero-de-investidores-na-b3-cresce-mesmo-em-cenario-de-alta-volatilidade)>.
- BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, p. 5–32, 2001.
- BRITO, O. S. de. *Mercado financeiro*. [S.l.]: Saraiva Educação SA, 2019.
- BULKOWSKI, T. N. *Encyclopedia of candlestick charts*. [S.l.]: John Wiley & Sons, 2012.

- CAJUEIRO, D. O.; TABAK, B. M. The hurst exponent over time: testing the assertion that emerging markets are becoming more efficient. *Physica A: statistical mechanics and its applications*, Elsevier, v. 336, n. 3-4, p. 521–537, 2004.
- CAVALCANTE, F.; MISUMI, J. Y.; RUDGE, L. F. *Mercado de capitais: o que é, como funciona*. [S.l.]: Elsevier, 2005.
- CHANTARAKASEMCHIT, O.; NUCHITPRASITCHAI, S.; NILSIAM, Y. Forex rates prediction on eur/usd with simple moving average technique and financial factors. In: IEEE. *2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. [S.l.], 2020. p. 771–774.
- DIETTERICH, T. G. et al. Ensemble learning. *The handbook of brain theory and neural networks*, MIT press Cambridge, MA, USA, v. 2, n. 1, p. 110–125, 2002.
- DWIVEDI, Y. K. et al. Artificial intelligence (ai): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, Elsevier, v. 57, p. 101994, 2021.
- ELLIS, C. A.; PARBERY, S. A. Is smarter better? a comparison of adaptive, and simple moving average trading strategies. *Research in International Business and Finance*, v. 19, n. 3, p. 399–411, 2005. ISSN 0275-5319.
- EXPERT XP. *Pessoas físicas na Bolsa: Mercado de ações brasileiro em contínua evolução*. 2023. <<https://conteudos.xpi.com.br/acoes/relatorios/pessoas-fisicas-na-bolsa-mercado-de-acoes-brasileiro-em-continua-evolucao/>>.
- FAMA, E. F. Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, JSTOR, v. 25, n. 2, p. 383–417, 1970.
- FEI, N. et al. Z-score normalization, hubness, and few-shot learning. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. [S.l.: s.n.], 2021. p. 142–151.
- FELIZARDO, L. et al. Comparative study of bitcoin price prediction using wavenets, recurrent neural networks and other machine learning methods. In: IEEE. *2019 6th International Conference on Behavioral, Economic and Socio-Cultural Computing (BESC)*. [S.l.], 2019. p. 1–6.
- FISHER, R. A. The use of multiple measurements in taxonomic problems. *Annals of eugenics*, Wiley Online Library, v. 7, n. 2, p. 179–188, 1936.
- FRENCH, D. The dutch monetary environment during tulipmania. *Quarterly Journal of Austrian Economics*, SPRINGER SCIENCE+ BUSINESS MEDIA, v. 9, n. 1, p. 3, 2006.
- GAD, S.; ANDRIKOPOULOS, P. Diversification benefits of shari’ah compliant equity etfs in emerging markets. *Pacific-Basin Finance Journal*, Elsevier, v. 53, p. 133–144, 2019.
- GAO, Z. Stock price prediction with arima and deep learning models. In: IEEE. *2021 IEEE 6th International Conference on Big Data Analytics (ICBDA)*. [S.l.], 2021. p. 61–68.
- GINI, C. Measurement of inequality of incomes. *The economic journal*, Oxford University Press Oxford, UK, v. 31, n. 121, p. 124–125, 1921.
- GONZALO, J. The making of “estimation of common long-memory components in cointegrated systems”. *Journal of Financial Econometrics*, Oxford University Press, v. 8, n. 2, p. 174–176, 2010.
- GONZALO, J.; GRANGER, C. Estimation of common long-memory components in cointegrated systems. *Journal of Business & Economic Statistics*, Taylor & Francis, v. 13, n. 1, p. 27–35, 1995.
- GRAHAM et al. *Security Analysis: The Classic 1934 Edition*. [S.l.]: McGraw-Hill Education, 1934. ISBN 9780070244962.
- HALIL, R.; DEMIRCI, M. Predicting the turkish stock market bist 30 index using deep learning. *International Journal of Engineering Research and Development*, v. 11, n. 1, p. 253–265, 2019.
- HANDAYANI, I. et al. Longer time frame concept for foreign exchange trading indicator using matrix correlation technique. In: IEEE. *2019 Fourth International Conference on Informatics and Computing (ICIC)*. [S.l.], 2019. p. 1–5.

- IGARASHI, W.; VALDEVIESO, G. S.; IGARASHI, D. C. C. Análise de sentimentos e indicadores técnicos: uma análise da correlação dos preços de ativos com a polaridade de notícias do mercado de ações. *Brazilian Journal of Business*, v. 3, n. 1, p. 470–486, 2021.
- IJEGWA, A. D. et al. A predictive stock market technical analysis using fuzzy logic. *Comput. Inf. Sci.*, v. 7, n. 3, p. 1–17, 2014.
- Ji, G. et al. An adaptive feature selection schema using improved technical indicators for predicting stock price movements. *Expert Systems with Applications*, Elsevier, v. 200, p. 116941, 2022.
- JIANG, M. et al. The two-stage machine learning ensemble models for stock price prediction by combining mode decomposition, extreme learning machine and improved harmony search algorithm. *Annals of Operations Research*, Springer, p. 1–33, 2022.
- JIANG, M. et al. An improved stacking framework for stock index prediction by leveraging tree-based ensemble models and deep learning algorithms. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 541, p. 122–272, 2020.
- KANG, B.-K. Improving macd technical analysis by optimizing parameters and modifying trading rules: evidence from the japanese nikkei 225 futures market. *Journal of Risk and Financial Management*, MDPI, v. 14, n. 1, p. 37, 2021.
- KANIEL, R.; SAAR, G.; TITMAN, S. Individual investor trading and stock returns. *The Journal of finance*, Wiley Online Library, v. 63, n. 1, p. 273–310, 2008.
- KARASU, S.; ALTAN, A. Crude oil time series prediction model based on lstm network with chaotic henry gas solubility optimization. *Energy*, Elsevier, v. 242, p. 122964, 2022.
- KAUR, H.; PANNU, H. S.; MALHI, A. K. A systematic review on imbalanced data challenges in machine learning: Applications and solutions. *ACM Computing Surveys (CSUR)*, ACM New York, NY, USA, v. 52, n. 4, p. 1–36, 2019.
- KELOTRA, A.; PANDEY, P. Stock market prediction using optimized deep-convlstm model. *Big Data*, Mary Ann Liebert, Inc., publishers 140 Huguenot Street, 3rd Floor New ... , v. 8, n. 1, p. 5–24, 2020.
- KIM, T. K. T test as a parametric statistic. *Korean journal of anesthesiology*, The Korean Society of Anesthesiologists, v. 68, n. 6, p. 540–546, 2015.
- KORCZAK, J.; HEMES, M. Deep learning for financial time series forecasting in a-trader system. In: IEEE. *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*. [S.l.], 2017. p. 905–912.
- KRASKOV, A.; STÖGBAUER, H.; GRASSBERGER, P. Estimating mutual information. *Physical review E*, APS, v. 69, n. 6, p. 066138, 2004.
- KRUSKAL, W. H.; WALLIS, W. A. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, Taylor & Francis, v. 47, n. 260, p. 583–621, 1952.
- KUMAR, G.; SINGH, U. P.; JAIN, S. Swarm intelligence based hybrid neural network approach for stock price forecasting. *Computational Economics*, Springer, p. 1–49, 2022.
- KUMAR, T. S. Data mining based marketing decision support system using hybrid machine learning algorithm. *Journal of Artificial Intelligence*, v. 2, n. 03, p. 185–193, 2020.
- LEE, M.-C. et al. Exploring the effectiveness of deep neural networks with technical analysis applied to stock market prediction. *Computer Science and Information Systems*, v. 18, n. 2, p. 401–418, 2021.
- MALKIEL, B. G. *A random walk down Wall Street : the time-tested strategy for successful investing*. [S.l.]: New York: W.W. Norton & Company, 2019., 2003.
- MATSURA, E. K. *Comprar ou vender?: como investir na bolsa utilizando análise gráfica*. [S.l.]: Saraiva Educação SA, 2017.
- MCNEMAR, Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, Springer-Verlag New York, v. 12, n. 2, p. 153–157, 1947.
- MEYER, H. et al. Importance of spatial predictor variable selection in machine learning applications—moving from data reproduction to spatial prediction. *Ecological Modelling*, Elsevier, v. 411, p. 108815, 2019.

- MORRIS, G. L. *Candlestick Charting Explained: Timeless Techniques for Trading Stocks and Futures: Timeless Techniques for Trading Stocks and Futures*. [S.l.]: McGraw-Hill Companies, 1994. ISBN 978-0071461542.
- MUSSA, A. et al. Hipótese de mercados eficientes e finanças comportamentais: as discussões persistem. *FACEF Pesquisa-Desenvolvimento e Gestão*, v. 11, n. 1, 2010.
- MUTHUKRISHNAN, R.; ROHINI, R. Lasso: A feature selection technique in predictive modeling for machine learning. In: IEEE. *2016 IEEE international conference on advances in computer applications (ICACA)*. [S.l.], 2016. p. 18–20.
- NAMETALA, C. A. et al. Use of econometric predictors and artificial neural networks for the construction of stock market investment bots. *Computational Economics*, Springer, v. 61, n. 2, p. 743–773, 2023.
- NAYAK, R. K.; MISHRA, D.; RATH, A. K. A naïve svm-knn based stock market trend reversal analysis for indian benchmark indices. *Applied Soft Computing*, Elsevier, v. 35, p. 670–680, 2015.
- NGUYEN, T. H.; SHIRAI, K. Topic modeling based sentiment analysis on social media for stock market prediction. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. [S.l.: s.n.], 2015. p. 1354–1364.
- NI, J.; XU, Y. Forecasting the dynamic correlation of stock indices based on deep learning method. *Computational Economics*, Springer, p. 1–21, 2021.
- NI, Y. et al. Does board structure affect stock price overshooting informativeness measured by stochastic oscillator indicators? *International Journal of Finance & Economics*, Wiley Online Library, v. 27, n. 2, p. 2290–2302, 2022.
- NICHOLAS, B.; ANDREI, S.; ROBERT, V. A model of investor sentiment. We are grateful to the nsf for financial support, and to oliver blanchard, alon brav, john campbell (a referee), john cochrane, edward glaeser, jb heaton, danny kahneman, david laibson, owen lamont, drazen prelec, jay ritter (a referee), ken singleton, dick thaler, an anonymous referee, and the editor, bill schwert, for comments. 1. *Journal of Financial Economics*, Elsevier BV, v. 49, n. 3, p. 307–343, 1998.
- PABUÇCU, H.; ONGAN, S.; ONGAN, A. Forecasting the movements of bitcoin prices: an application of machine learning algorithms. *arXiv preprint arXiv:2303.04642*, 2023.
- PATRO, S.; SAHU, K. K. Normalization: A preprocessing stage. *arXiv preprint arXiv:1503.06462*, 2015.
- PEARSON, K. X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, Taylor & Francis, v. 50, n. 302, p. 157–175, 1900.
- PENG, Y. et al. Feature selection and deep neural networks for stock price direction forecasting using technical analysis indicators. *Machine Learning with Applications*, Elsevier, v. 5, p. 100060, 2021.
- PINO, F. A.; MORETTIN, P. A.; MENTZ, R. P. Modelling and forecasting linear combinations of time series. *International Statistical Review/Revue Internationale de Statistique*, JSTOR, p. 295–313, 1987.
- PINTO, W. D. O. *Previsão de séries temporais com redes neurais artificiais aplicada à demanda*. Monografia (Bacharel em Sistemas de Informação — Centro Universitário do Leste de Minas Gerais, Coronel Fabriciano, 2009).
- RAO, T.; SRIVASTAVA, S. Analyzing stock market movements using twitter sentiment analysis. IEEE Computer Society, 2012.
- RESENDE, C. C. de. *Modelos Matemático-computacional Para Previsão de Tendência de Preços: Abordagem com Múltiplos Ativos Buscando Oportunidades em Desvios da Hipótese de Mercado Eficiente*. 138 p. Tese (Doutorado) — Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte - MG, 2021.
- ROSA, G. Buying and selling decision in the brazilian stock exchange financial market by a neo fuzzy neuron (nfn) applied to the hurwicz criterion. In: *CSBC 2023 - BWAIF 2023*. João Pessoa, Paraíba: [s.n.], 2023.
- SAGHI, F.; REZAEI, M. J. Integrating wavelet decomposition and fuzzy transformation for improving the accuracy of forecasting crude oil price. *Computational Economics*, Springer, p. 1–33, 2021.



- SAGI, O.; ROKACH, L. Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Wiley Online Library, v. 8, n. 4, p. e1249, 2018.
- SALMAN, R.; KECMAN, V. Regression as classification. In: IEEE. *2012 Proceedings of IEEE Southeastcon*. [S.l.], 2012. p. 1–6.
- SAMUELSON, P. A. Proof that properly anticipated prices fluctuate randomly. In: *The world scientific handbook of futures markets*. [S.l.]: World Scientific, 2016. p. 25–38.
- SERMPINIS, G.; TSOUKAS, S.; ZHANG, P. Modelling market implied ratings using lasso variable selection techniques. *Journal of Empirical Finance*, Elsevier, v. 48, p. 19–35, 2018.
- SHAMSEDDIN, M. A.; ELMESKI, H. A. Mapping dynamic changes in hydrological time series using the average directional index. *International Journal of River Basin Management*, Taylor & Francis, v. 20, n. 1, p. 67–78, 2022.
- SHELDON, M. R.; FILLYAW, M. J.; THOMPSON, W. D. The use and interpretation of the friedman test in the analysis of ordinal-scale data in repeated measures designs. *Physiotherapy Research International*, Wiley Online Library, v. 1, n. 4, p. 221–228, 1996.
- SHILLER, R. J. *Irrational Exuberance*. [S.l.]: Princeton: Princeton University Press, 2000.
- SHOVEN, J. B.; SIALM, C. The dow jones industrial average: the impact of fixing its flaws. *Journal of Wealth Management*, Citeseer, v. 3, n. 3, p. 9–18, 2000.
- STILLMAN, R. J. *Dow Jones industrial average : history and role in an investment strategy*. [S.l.]: Homewood, Ill. : Dow Jones-Irwin, 1986.
- STRADER, T. J. et al. Machine learning stock market prediction studies: review and research directions. *Journal of International Technology and Information Management*, v. 28, n. 4, p. 63–83, 2020.
- TING, C. *Introduction To Quantitative Finance, An: A Three-principle Approach*. [S.l.]: World Scientific Publishing Company, 2015. ISBN 9789814704328.
- TOMASEVIC, N.; GVOZDENOVIC, N.; VRANES, S. An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers & education*, Elsevier, v. 143, p. 103676, 2020.
- TRADE MENTAL. *Padrões de Candlesticks – Os 10 Mais Importantes*. 2022. <<https://trademental.com/padroes-candlesticks/>>.
- UNCTAD. *International Production Beyond the Pandemic*. [S.l.]: World Investment Report 2020, 2020. ISBN 978-92-1-112985-4.
- VADLAMUDI, S. Stock market prediction using machine learning: A systematic literature review. *American Journal of Trade and Policy*, v. 4, n. 3, p. 123–128, 2017.
- VAIDYA, R. Stochastic and momentum analysis of nepalese stock market. *Journal of Nepalese Business Studies*, v. 11, n. 1, p. 14–22, 2018.
- VEERAMANI, C.; VENUGOPAL, R.; MURUGANANDAN, S. An exploration of the fuzzy inference system for the daily trading decision and its performance analysis based on fuzzy mcdm methods. *Computational Economics*, Springer, p. 1–28, 2022.
- WANG, C. et al. Stock market index prediction using deep transformer model. *Expert Systems with Applications*, Elsevier, v. 208, p. 118128, 2022.
- WANG, H. et al. Soft-voting clustering ensemble. In: SPRINGER. *Multiple Classifier Systems: 11th International Workshop, MCS 2013, Nanjing, China, May 15-17, 2013. Proceedings 11*. [S.l.], 2013. p. 307–318.
- WANG, L.-X. Fast training algorithms for deep convolutional fuzzy systems with application to stock index prediction. *IEEE Transactions on fuzzy systems*, IEEE, v. 28, n. 7, p. 1301–1314, 2019.
- WILSON, T.; WIEBE, J.; HOFFMANN, P. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational linguistics*, MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA, v. 35, n. 3, p. 399–433, 2009.
- YILMAZ, F. M.; YILDIZTEPE, E. Statistical evaluation of deep learning models for stock return forecasting. *Computational Economics*, Springer, p. 1–24, 2022.

- ZHANG, X. et al. Predicting stock price movement using a dbn-rnn. *Applied Artificial Intelligence*, Taylor & Francis, v. 35, n. 12, p. 876–892, 2021.
- ZHANG, Y.; WAHAB, M.; WANG, Y. Forecasting crude oil market volatility using variable selection and common factor. *International Journal of Forecasting*, Elsevier, v. 39, n. 1, p. 486–502, 2023.
- ZHAO, Z.; ANAND, R.; WANG, M. Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform. In: IEEE. *2019 IEEE international conference on data science and advanced analytics (DSAA)*. [S.l.], 2019. p. 442–452.