
**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^a. Dr^a.* Charlene Cássia de Resende



Divinópolis-MG

15-03-2023

coloque aqui a dedicatoria

Agradecimientos

*"O investimento em conhecimento sempre rende os
melhores juros."*

Benjamin Franklin

Lista de Figuras

| | | |
|-----|---|----|
| 1.1 | Porcentagem de pessoas por país que investem em ações. | 4 |
| 1.2 | Dados da bolsa de valores brasileira | 5 |
| 2.1 | Tendências de uma série temporal. | 6 |
| 2.2 | Características cíclicas e sazonais de uma serie temporal | 7 |
| 2.3 | Gráfico de velas e suas características observáveis. | 8 |
| 2.4 | Processo de desenvolvimento dos trabalhos analisados. | 8 |
| 2.5 | Matriz de confusão. | 14 |
| 3.1 | Visão macro do processo de construção da base da dados. | 19 |
| 3.2 | Visão macro do processo de seleção de variáveis. | 21 |
| 3.3 | Visão macro do processo de previsão. | 22 |
| 3.4 | Visão macro do processo de recomendação de investimentos. | 23 |

Lista de Tabelas

| | | |
|-----|--|----|
| 2.1 | Transformação da saídas no domínio contínuo para o domínio discreto. | 13 |
| 2.2 | Trabalhos relacionados | 18 |
| 4.1 | Cronograma de atividades | 24 |

Lista de Acrônimos

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

| | |
|----------------------|--|
| RBF | <i>Radial Basis Function</i> |
| RF | <i>Random Forest</i> |
| RIM | Regressão de Informação Mútua |
| RMSE | <i>Root Mean Squared Error</i> |
| RNN | <i>Recurrent Neural Networks</i> |
| ROC | <i>Rate of Change</i> |
| R² | Coefficiente de Determinação |
| SARIMA | <i>Seasonal Autoregressive Integrated Moving Average</i> |
| SMA | <i>Simple Moving Average</i> |
| SVM | <i>Support Vector Machine</i> |
| SVR | <i>Support Vector Regression</i> |
| TCN | <i>Temporal Convolutional Network</i> |
| TN | <i>True Negative</i> |
| TP | <i>True Positive</i> |
| TSI | <i>True Strength Index</i> |
| VP | Valores Passados |
| X² | <i>Chi-Squared</i> |

Sumário

| | | |
|----------|--|-----------|
| 1 | Introdução | 2 |
| 1.1 | Contextualização | 2 |
| 1.2 | Motivação e Relevância | 3 |
| 1.3 | Objetivo | 5 |
| 1.3.1 | Objetivo Geral | 5 |
| 1.3.2 | Objetivos Específicos | 5 |
| 1.4 | Organização do Trabalho | 5 |
| 2 | Trabalhos Relacionados | 6 |
| 2.1 | Séries Temporais | 6 |
| 2.2 | Mercado Financeiro | 7 |
| 2.3 | Abordagem | 8 |
| 2.4 | Conjunto de Dados | 9 |
| 2.5 | Geração de Novas Variáveis | 9 |
| 2.6 | Seleção de Variáveis | 11 |
| 2.7 | Técnicas de Predição | 12 |
| 2.8 | Estratégias | 13 |
| 2.9 | Critérios de Avaliação | 14 |
| 2.10 | Resumo | 16 |
| 3 | Abordagem Proposta | 19 |
| 3.1 | Base de Dados | 19 |
| 3.1.1 | Extração | 19 |
| 3.1.2 | Geração de novas Variáveis | 20 |
| 3.2 | Seleção de Variáveis | 20 |
| 3.3 | Modelo de Previsão | 21 |
| 3.4 | Recomendação de Investimento | 22 |
| 4 | Análises e Resultados | 24 |
| 4.1 | Predição do Valor | 24 |
| 4.2 | Predição do Sentido | 24 |
| 4.3 | Consenso | 24 |
| 4.4 | Cronograma | 24 |
| 5 | Conclusão | 25 |
| 5.1 | Propostas de Continuidade | 25 |

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 que oferece (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 os investidores 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:

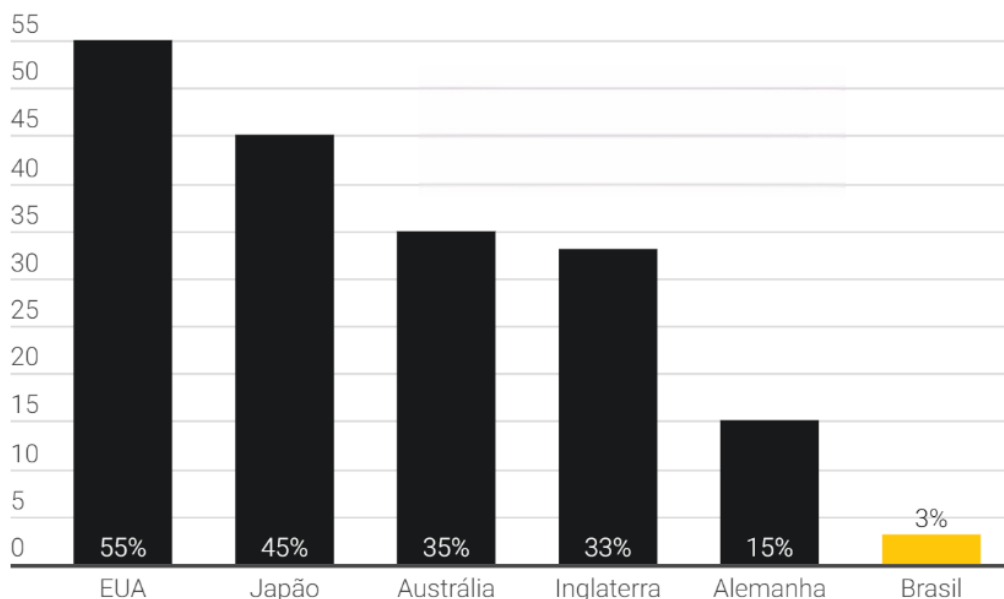
- Análise Técnica (AT): essa análise originou-se no ocidente por volta de 1900, quando Charles H. Dow, publicou uma série de editoriais sobre métodos de especulação com ações (MORRIS, 1994). Tais métodos surgiram com o objetivo de indicar padrões e tendências de mercado a partir de observações gráficas, 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 complexos 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.

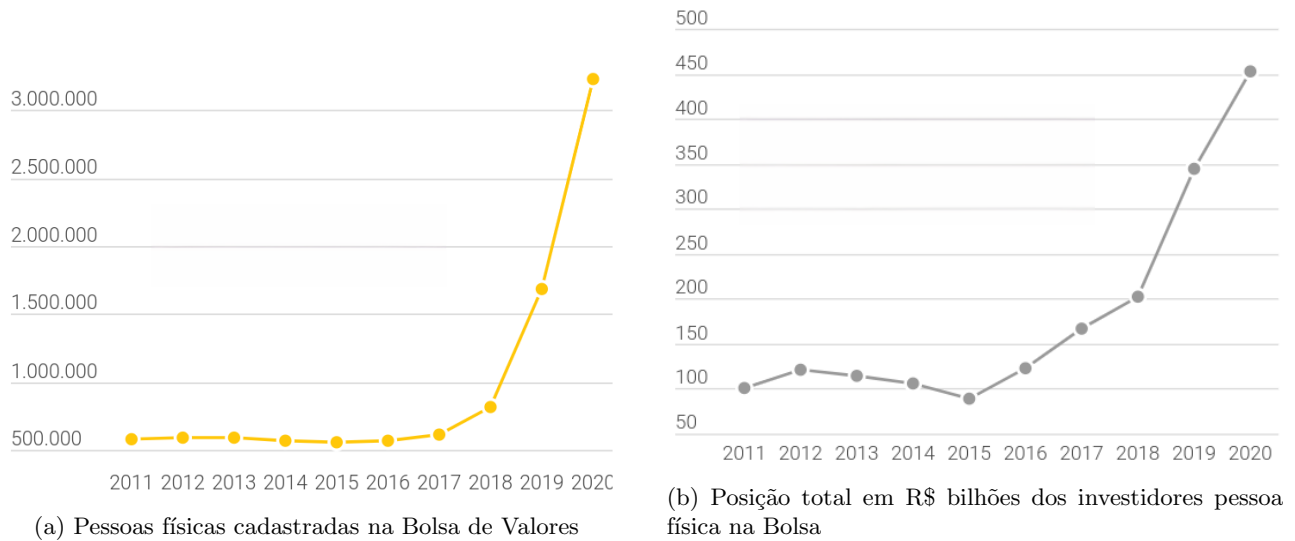
O estudo realizado pela XP Investimentos (XP, 2023) revela que o número de pessoas físicas investindo em ações está diretamente relacionado à estabilidade do mercado e à qualidade de vida nos países analisados. A Figura 1.1 ilustra essa relação, confirmando que países com maior estabilidade econômica 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 modelos 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 rapidamente 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, conforme mencionado por Nametala et al. (2023) e Felizardo et al. (2019). 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

Figura 1.2: Dados da bolsa de valores brasileira



Fonte: Adaptado de XP (2023).

comitê, buscando obter um modelo final mais robusto e assertivo (KORCZAK; HEMES, 2017).

Com o avanço da tecnologia e o crescente interesse dos investidores individuais, torna-se cada vez mais importante explorar e desenvolver métodos que possam ajudar na previsão de valores futuros de ações. Nesse contexto, o presente trabalho tem como objetivo principal contribuir para esse campo de pesquisa, explorando diferentes modelos de predição e aplicando-os a 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 do modelo proposto;
- 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

A fim de facilitar a compreensão do trabalho proposto, o estudo foi dividido em três seções que introduzem gradualmente os conceitos necessários para entender a proposta. Neste capítulo, são apresentados o contexto do problema, as motivações e os objetivos gerais e específicos para a realização da proposta. Em seguida, o Capítulo 2 aborda estudos relacionados às técnicas de previsão de valores no mercado financeiro. No Capítulo 3, é apresentada uma abordagem para a previsão de valores em series temporais financeira e recomendações de investimento.

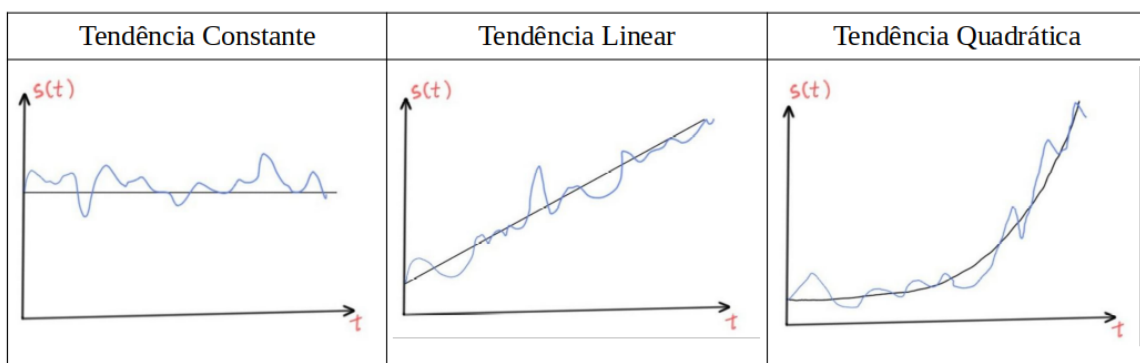
Trabalhos Relacionados

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 novas 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.

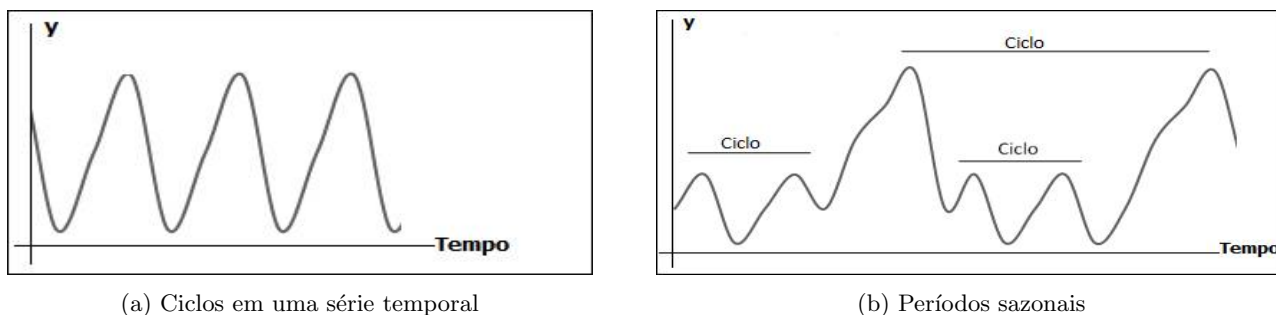


A tendência de uma série temporal é utilizada para descrever o comportamento da série ao longo do tempo, ou seja, verifica-se há uma inclinação de alta, queda ou nula, além de indicar a velocidade dessas mudanças. Desse modo, existem diferentes formas matemáticas representativas destes movimentos da 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 serie abordada, e geralmente não são previsíveis a longo prazo.

Figura 2.2: Características cíclicas e sazonais de uma serie temporal



Fonte: Pinto (2009)

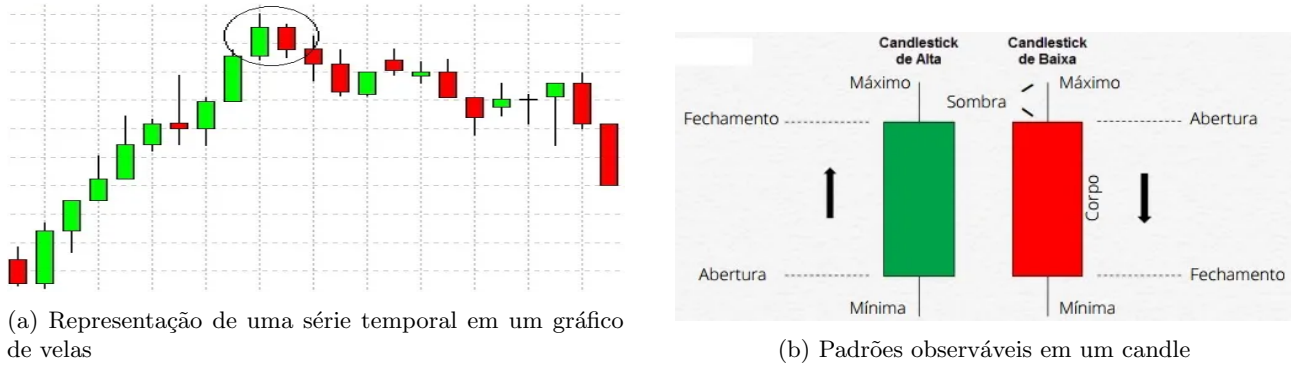
Por fim, entende-se a sazonalidade como as variações regulares que ocorrem em uma série temporal em intervalos fixos de tempo, como dias, semanas, meses ou anos. Um exemplo disso pode ser observado na Figura 2.2b, em que um padrão visual é repetido em um determinado intervalo de tempo. Essas variações 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 de gráfico de velas com o 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.



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 aprendizado de máquina 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.

2.3 Abordagem

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, esta última sendo o foco dos trabalhos analisados. Para isso, tais estudos seguem 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 podem ser explorados com mais detalhes na Seção 2.6. Em alguns casos, também é aplicada a normalização dos dados, a fim de colocar todos os valores em uma mesma faixa de 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 Subseção 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 Conjunto 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 de uma variável dentro de um limite específico e reduzir as diferenças desproporcionais entre as variáveis. 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*, a normalização por escala decimal, entre outras, que delimitam os valores dentro de um intervalo, geralmente de -1 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 Novas Variáveis

Para melhorar o desempenho dos modelos, algumas abordagens utilizam técnicas de geração de novas *features*, que são variáveis adicionais derivadas dos dados brutos. Os métodos de geração de novos dados a partir dos dados brutos 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 Passados (VP) - consiste em estabelecer intervalos temporais, conhecidos como janelas, que abrangem dados anteriores ao ponto de análise (AMARAL, 2021; ROSA, 2023), como: $S_{t-1}, S_{t-2}, \dots, S_{t-n}$, em

que S_{t-1} corresponde ao valor da serie temporal no instante $t - 1$ e n indica o tamanho da janela.

- *Simple Moving Average* (SMA) - é 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). O calculo da SMA é realizado por meio da seguinte equação:

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

no qual n representa o tamanho da janela móvel e S_{t-i} corresponde ao valor da serie temporal no instante $t - i$.

- *Exponential Moving Average* (EMA) - é um modelo matemático semelhante à SMA, pois também busca suavizar a flutuação dos dados de preço e identificar tendências. No entanto, a EMA atribui mais peso aos preços mais recentes, o que a torna mais sensível às mudanças recentes nos dados (RESENDE, 2021). O cálculo da EMA é realizado pela seguinte equação:

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

em que S_t corresponde ao valor da serie temporal no instante t e α é dito como o fator de suavização e é frequentemente definido como $\alpha = \frac{2}{n+1}$, sendo n o tamanho da janela móvel.

- *Moving Average Convergence Divergence* (MACD) - identifica a direção e força de uma tendência predominante em um ativo financeiro (VEERAMANI; VENUGOPAL; MURUGANANDAN, 2022). Esse indicador pode ser calculado através da seguinte equação:

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

no qual w corresponde a janela móvel curta e k a janela móvel longa.

- *Comodity Channel Index* (CCI) - identifica pontos de reversão de tendência e avalia sua força (ALTAN; KARASU, 2019; HALIL; DEMIRCI, 2019). Tal indicador pode ser calculado com o uso da equação:

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

sendo que n é a janela de amostragem desejada, DP é o desvio padrão da serie e TP é dito como $TP = \frac{C+H+L}{3}$, sendo que H indica o ponto de máxima, L o ponto de mínima e C 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). Esse índice pode ser calculado pela equação:

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

em que n indica a janela de amostragem desejada e TR é dito como $TR = \max(H - L, |H - C_{t-1}|, |L - C_{t-1}|)$, sendo que H corresponde ao ponto de máxima do *candle* atual, L ao ponto de mínimo e C_{t-1} ao 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). Tal indicador pode ser calculado através da equação:

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

sendo considerado apenas o valor atual S_t e o valor de n estados passados S_{t-n} da serie temporal.

- *True Strength Index* (TSI) - identifica a força e a direção de uma tendência em um determinado ativo financeiro, combinando a suavização da EMA com a taxa de variação ROC dos preços para fornecer sinais de compra e venda (ANWAR; RAHMAN, 2019). Esse índice pode ser calculado com o uso da equação:

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

em que PC corresponde a variação do preço de fechamento ($C_{t-1} - C_t$), w indica a janela móvel curta e k 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 a sua faixa de preço recente (FELIZARDO et al., 2019; VEERAMANI; VENUGOPAL; MURUGANANDAN, 2022). Podendo ser calculado através da equação a seguir:

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

no qual C_t indica o valor de fechamento atual, L_{t-n} e H_{t-n} representam o valor mínimo e máximo de n *candles* passados.

- Média Móvel Simples do Oscilador Estocástico (%D) - proporciona uma visão mais suave da direção e força da tendência (PABUÇCU; ONGAN; ONGAN, 2023). Esse indicador pode ser calculado pela seguinte equação:

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

em que n corresponde a janela móvel.

- 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). Para tal, ele é calculado pela seguinte equação:

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

sendo que C_t indica 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 de *features* 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 regressão quanto para classificação. 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 escore de *Fisher* para selecionar as características mais importantes com base na separação entre as classes (FISHER, 1936);
- *Gini* - implementado 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; 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, podemos citar as abordagens estatísticas, modelos simples de *machine learning*, e até mesmo modelos mais complexos desse conjunto como os de *deep learning*, destacando a ampla variedade de opções disponíveis para aplicação nesse contexto.

Dentro da classe de modelos estatísticas, 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 aprendizado de máquina, permitindo explorar as vantagens de cada modelo e obter previsões mais precisas e confiáveis (NAMETALA et al., 2023).

Já a classe de *machine learning* 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), ANFIS (SAGHI; REZAEI, 2021), *Decision Trees* 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* 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 Sistem* (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 nas regras 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. Essa transformação permite a utilização de técnicas de classificação para identificar e agir apoiado em diferentes tendências do ativo.

| 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

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).

Já 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 propostas. 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.

| | | Valor Predito | |
|------|-----|--------------------------|--------------------------|
| | | Sim | Não |
| Real | Sim | Verdadeiro Positivo (TP) | Falso Negativo (FN) |
| | Não | Falso Positivo (FP) | Verdadeiro Negativo (TN) |

Fonte: Diego Nogari (2020)

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 formula para o calculo da Precisão Positiva (PP), utiliza-se os dados de *True Positive* (TP) e *False Positive* (FP), sendo descrita como:

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

Já para calcular a Precisão Negativa (PN), utiliza-se os valores de *True Negative* (TN) e *False Negative* (FN) conforme a seguinte equação:

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

- *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 pela seguinte equação:

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

- 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 formula 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 F-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). A formula para o calculo da F1 é dada por:

$$F1 = 2 \frac{PPRecal}{PP + Recal} \quad (2.15)$$

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 analisado 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. Um valor de R^2 igual a 1 indica que o modelo explica perfeitamente a variabilidade da variável dependente. Por outro lado, um valor de R^2 igual a 0 indica que o modelo não é capaz de explicar a variabilidade da variável dependente. Valores intermediários de R^2 indicam a proporção de variabilidade explicada pelo modelo, sendo que um valor mais próximo de 1 indica um ajuste melhor.
- *Mean Error* (ME) - verifica se as previsões estão consistentemente superestimando ou subestimando os valores reais (FELIZARDO et al., 2019). Tal métrica pode ser calculada através da equação a seguir:

$$ME = \frac{1}{n} \sum (y_{pred} - y_{real}) \quad (2.16)$$

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

- *Mean Absolute Error* (MAE) - diferente do ME, essa métrica 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). Sendo calculado por meio da seguinte equação:

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

em que todas as variáveis são as mesma correspondentes do método ME.

- *Mean Squared Error* (MSE) - avalia a dispersão das previsões em relação aos valores reais, penalizando error maiores de forma quadrática (NI; XU, 2021; WANG et al., 2022). Tal métrica pode ser calculada pela seguinte equação:

$$MSE = \frac{1}{n} \sum (y_{pred} - y_{real})^2 \quad (2.18)$$

em que todas as variáveis são as mesma correspondentes do método ME.

- *Mean Absolute Percent Error* (MAPE) - avalia a precisão de um modelo de previsão em termos percentuais (ANAND, 2021; YILMAZ; YILDIZTEPE, 2022). O cálculo dessa métrica por ser realizado através da seguinte equação:

$$MAPE = \frac{1}{n} \sum \left| \frac{y_{pred} - y_{real}}{y_{real}} \right| 100 \quad (2.19)$$

em que todas as variáveis são as mesma correspondentes do método ME.

- *Root Mean Squared Error* (RMSE) - assim como o MSE, essa métrica mede a diferença entre os valores previstos por um modelo e os valores real, porém, o RMSE tem a vantagem de ter a mesma unidade de medida dos valores originais, o que facilita a interpretação dos resultados (ALTAN; KARASU, 2019; PABUÇCU; ONGAN; ONGAN, 2023). Sendo calculada por meio da seguinte equação:

$$RMSE = \sqrt{\frac{1}{n} \sum (y_{pred} - y_{real})^2} \quad (2.20)$$

em que todas as variáveis são as mesma correspondentes do método ME.

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 estatísticas. Essas métricas incluem o retorno financeiro, a comparação do desempenho do modelo em relação à estratégia "buy and hold"(comprar e manter) (NAMEITALA et al., 2023), o retorno percentual do valor investido (WANG et al., 2022), a avaliação do lucro e perda (PNL), 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.

2.10 Resumo

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. |
| Anwar e Rahman (2019) | ANN, RNN, DNN, LSTM, RF, LR e SVM. | Construir um modelo de previsão utilizando algoritmos de IA. |
| 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. |
| Wang (2019) | DCFS. | Propor e avaliar algoritmos de treinamento rápido. |
| Ampomah, Qin e Nyame (2020) | RF, Boosting e XGBoost. | Aplicar técnicas de ensemble 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 framework de Stacking aprimorado para prever a direção do índice de preços das ações. |
| Nametala et al. (2023) | SMA, ARIMA, SARIMA, GARCH e MLP. | Utilizar técnicas de inteligência artificial para a construção de um Robô de operações no mercado financeiro do Brasil. |
| 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. |
| 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 "buy & hold". |
| 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. |
| Ni e Xu (2021) | DCDNN e LSTM. | Propor uma nova abordagem de seleção/correlação de 'features'. |

| Autor | Modelos | Objetivo |
|-----------------------------|------------------------|--|
| Pabuçcu, Ogan e Ogan (2023) | SVM, ANN, NB, LR e RF. | Explorar diferentes técnicas de previsão de preços do Bitcoin. |

Tabela 2.2: Trabalhos relacionados

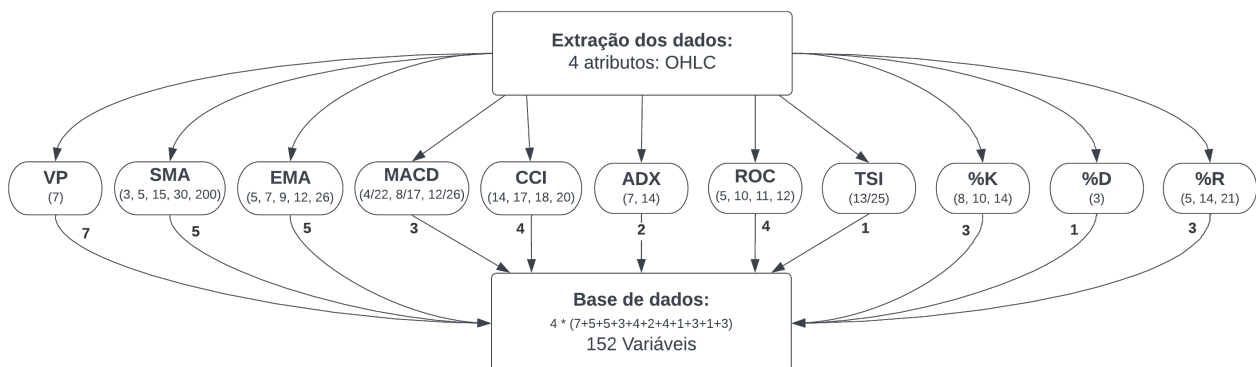
Abordagem Proposta

Para o desenvolvimento deste trabalho, propõe-se cinco 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 Base de Dados

O processo de construção da base 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 da base 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 da base da dados.



3.1.1 Extração

O mercado de ações do Brasil opera de segunda a sexta-feira, das 10:00 às 17:00 (B3, 2023a). Durante esse período, são geradas informações em diversas granularidades, permitindo acompanhar a evolução dos preços e volumes de negociação ao longo do tempo. Uma plataforma amplamente utilizada para acessar essas informações é a plataforma *Investing*¹, que disponibiliza dados históricos e em tempo real em diferentes intervalos, como diário, semanal e mensal. Além disso, essa plataforma abrange diversos mercados e modalidades de investimento em vários países, fornecendo informações relevantes, como os preços de OHLC, volume negociado e percentual de mudança. No contexto deste trabalho, os dados históricos de OHLC foram extraídos dessa plataforma para análise e estudo.

¹<http://www.investing.com>

3.1.2 Geração de novas Variáveis

Após a extração dos dados, são aplicadas diversas técnicas para gerar novas variáveis relevantes na tarefa de previsão do preço de fechamento da próxima amostra. Para isso, são calculadas novas *features* para cada valor de OHLC, empregando-se diversas fórmulas estatísticas, como:

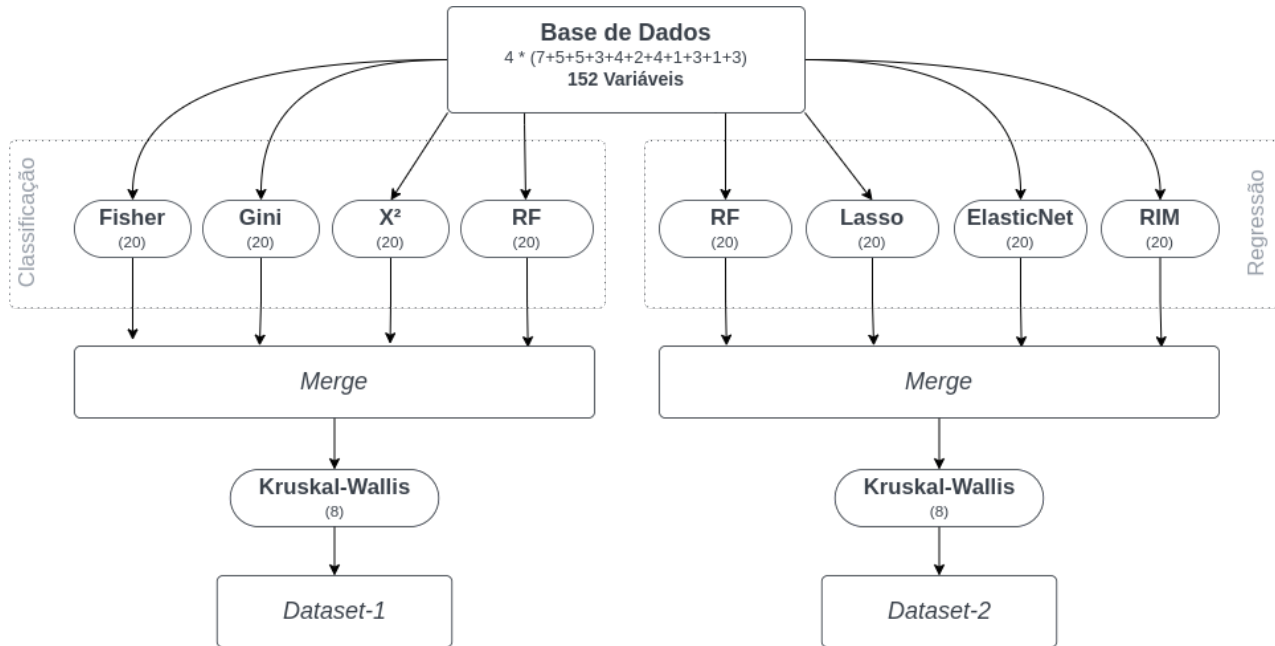
- VP - é obtido realizando o deslocamento de valores passados. Neste estudo, utiliza-se um deslocamento de 7 valores passados (AMARAL, 2021).
- SMA - calculado a partir da Equação (2.1). No presente estudo, utilizaram-se os intervalos de 3 (CHANTARAKASEMCHIT; NUCHITPRASITCHAI; NILSIAM, 2020), 5, 15, 30 (HANDAYANI et al., 2019) e 200 (ELLIS; PARBERY, 2005).
- EMA - tem seu calculo baseado na Equação (2.2). Neste estudo, foram adotados os períodos amostrais de 5, 7, 9 (ANBALAGAN; MAHESWARI, 2015), 12 (RESENDE, 2021) e 26 (ANANTHI; VIJAYAKUMAR, 2021).
- MACD - derivado da Equação (2.3). Nesta pesquisa, foram utilizados os 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). Neste trabalho, foram considerados 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 aplicação da Equação (2.5). Este estudo considera os períodos de 7 (KELOTRA; PANDEY, 2020) e 14 (SHAMSEDDIN; ELMESKI, 2022) espaços amostrais.
- ROC - determinado a partir da equação (2.6). O presente estudo adota os intervalos de 5, 10, 11 e 12 (KARASU; ALTAN, 2022) para o calculo do mesmo.
- TSI - calculado utilizando a Equação (2.7). Neste experimento, são adotados os 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). Neste estudo, são utilizados os intervalos de tempo de 8 (NI et al., 2022), 10 (IJEGWA et al., 2014) e 14 amostras (VEERAMANI; VENUGOPAL; MURUGANANDAN, 2022).
- %D - derivado da Equação (2.9). No presente trabalho, adotou-se a janela de tempo de 3 amostras (IJEGWA et al., 2014; VAIDYA, 2018).
- %R - tem seu calculo baseado na Equação (2.10). Neste estudo, são consideradas 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 152 novas variáveis, o que proporciona uma variedade de padrões de entrada para os modelos de predição.

3.2 Seleção de Variáveis

Com a base de dados gerada, é importante destacar que nem todas as variáveis possuem igual relevância para o processo de previsão. Portanto, é necessário realizar a seleção de variáveis, a qual foi dividida em duas etapas neste trabalho. Primeiramente, são aplicados métodos de seleção de variáveis descritos na Seção 2.6. Em seguida, ocorre a fusão de todas as variáveis selecionadas e a classificação delas em dois conjuntos de dados distintos, denominados *Dataset-1* e *Dataset-2*. Uma representação geral desse processo pode ser visualizada na Figura 3.2.

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



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. Já o segundo grupo é direcionado para modelos de regressão e inclui os métodos de RF, Lasso, ElasticNet e RIM. Cada um desses métodos, pertencentes aos dois grupos, seleciona 20 atributos 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 algoritmo, 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 8 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 modelo, otimizando assim o processo de previsão.

3.3 Modelo de Previsão

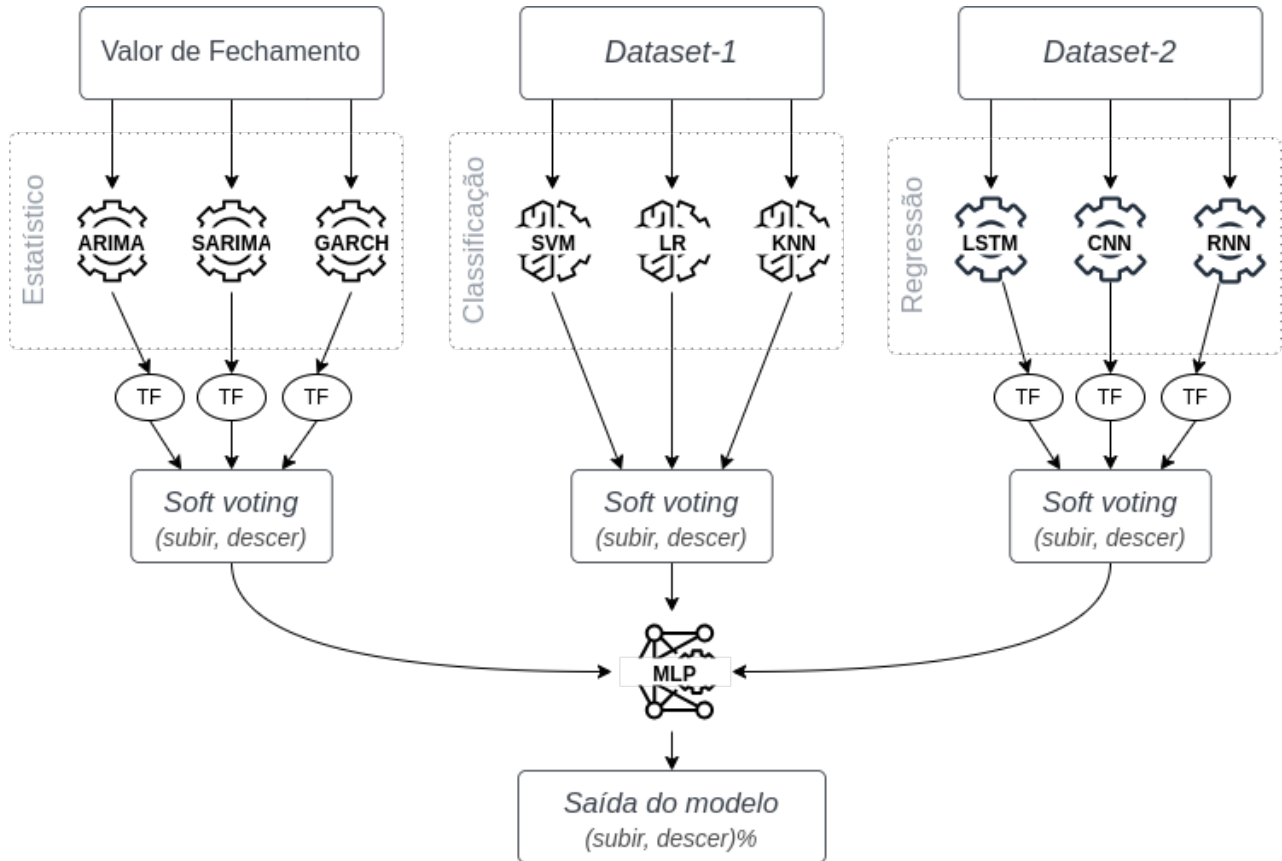
Para a construção do modelo de previsão proposto, foram utilizadas duas técnicas de *ensemble*: o *Soft Voting* e o *Stacking*.

O *Soft Voting*, como descrito por Wang et al. (2013), realiza a votação ponderada, expressando a saída em percentuais referentes a cada classe. Nessa técnica, vários modelos são treinados independentemente e suas previsões são combinadas usando pesos. O resultado final é uma média ponderada das previsões de cada modelo, refletindo a confiança de cada modelo na classificação de uma determinada classe.

Já o *Stacking*, introduzido por Dietterich et al. (2002), envolve a criação de camadas de modelos de aprendizado de máquina conectados. Nessa abordagem, os modelos de base são treinados em conjunto, e suas previsões são usadas como entrada para um meta-modelo que realiza a previsão final. O objetivo é combinar as previsões dos modelos de base de forma a obter uma previsão mais acurada.

No modelo proposto, são construídos três grupos de algoritmos distintos para realizar as previsões. O primeiro grupo consiste em modelos estatísticos, como 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 modelos de IA voltados para classificação, como SVM, LR e KNN, que recebem como entrada o *Dataset-1*, composto pelas variáveis relevantes

Figura 3.3: Visão macro do processo de previsão.



selecionadas para classificação. O terceiro grupo é formado por modelos de IA voltados para regressão, como LSTM, RNN e CNN, que recebem como entrada o *Dataset-2*, contendo as variáveis relevantes selecionadas para 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 uma lista com valores de pertinência para cada classe. Por sua vez, o *Stacking* utiliza uma rede MLP para combinar as saídas do *Soft Voting* de cada grupo, resultando em uma lista de percentuais de correlação a cada classe. Essa abordagem permite aproveitar a diversidade dos modelos utilizados e proporciona uma previsão mais robusta e confiável para a tarefa de classificação.

A proposta do modelo de previsão e a aplicação das técnicas de *ensemble* podem ser visualizadas na Figura 3.3, apresentando a sequência e interconexão dos diferentes estágios do processo.

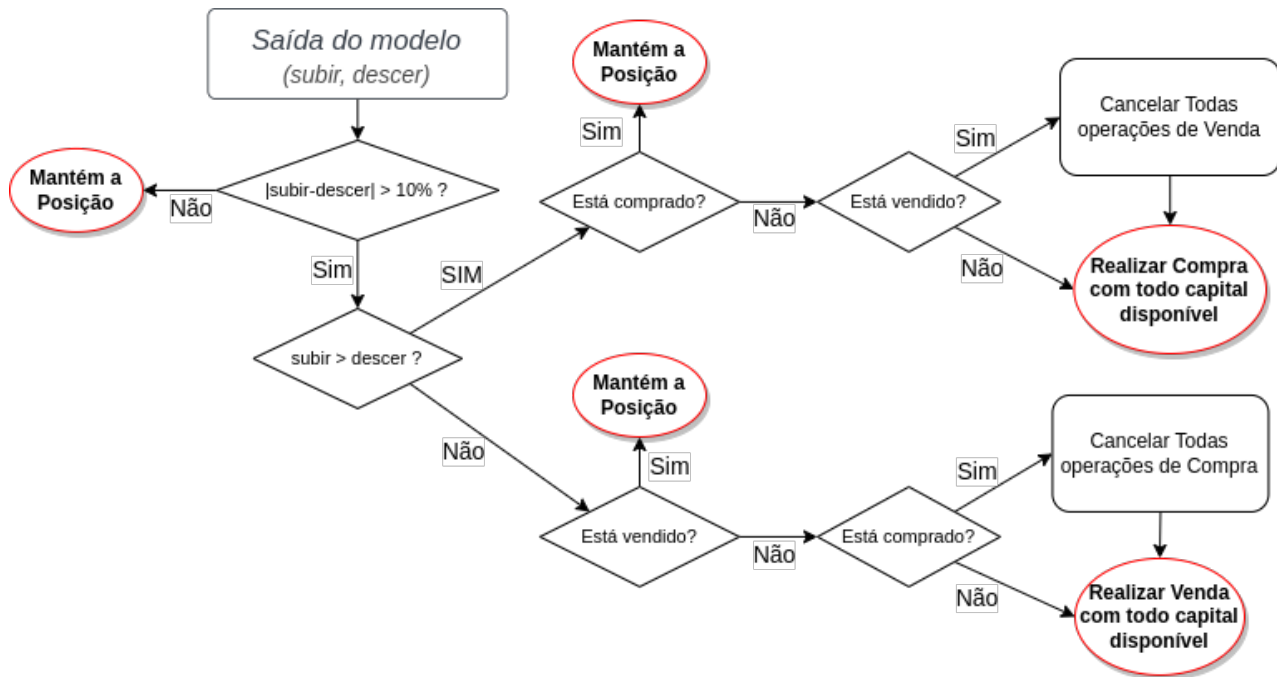
3.4 Recomendação de Investimento

Tendo em mãos o percentual de subida ou descida previsto pelo modelo proposto, inicia-se o processo de recomendação de compra e venda, conforme ilustrado na a Figura 3.4, que objetiva reduzir operações de baixa confiabilidade e maximizar os retornos financeiros. A estratégia de recomendação é baseada nos resultados do modelo 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. Se a diferença não for significativa, a recomendação é não realizar nenhuma operação e manter a posição atual. Caso a diferença seja relevante, é feita uma comparação para determinar qual classe possui maior percentual de pertinência.

Se o sinal de subida for mais relevante, então é mais provável que o próximo valor seja maior do que o atual. Nesse ponto, é verificado se há uma operação de compra em andamento. Se houver, é recomendado não

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



realizar nenhuma ação. Caso contrário, há uma verificação adicional para determinar se existe alguma operação de venda em andamento. Se houver, é recomendado encerrar todas as operações vigentes e realizar a compra utilizando todo o capital disponível. Se não houver operação de venda em andamento, é recomendado realizar a compra utilizando todo o capital.

Por outro lado, se o sinal de subida não for maior que o de descida, então é mais provável que o próximo valor seja menor do que o atual. Nessa etapa, é verificado se há uma operação de venda em andamento. Se houver, a recomendação é não realizar nenhuma ação. Caso contrário, é feita uma nova verificação para identificar se há uma operação de compra em andamento. Se houver, é recomendado encerrar todas as operações de compra e realizar uma venda utilizando todo o capital disponível. Caso não haja operação de compra em andamento, é recomendado realizar a venda utilizando todo o capital disponível.

Análises e Resultados

4.1 Predição do Valor

4.2 Predição do Sentido

4.3 Consenso

4.4 Cronograma

Esta seção detalha o cronograma a ser seguido sob orientação do Prof. Alisson Marques da Silva no cumprimento das exigências necessárias para a disciplina de TCC I. Para tanto, na Tabela 4.1 apresenta-se o cronograma de atividades proposto para realização deste projeto, considerando o prazo 1 semestres para sua integralização.

- **Atv-1:** Revisão da literatura;
- **Atv-2:** Escrita Monografia;
- **Atv-3:** Construir a base de dados;
- **Atv-4:** Implementar algoritmos de IA
- **Atv-5:** Aplicar e avaliar o desempenho dos algoritmos no mercado de índices.
- **Atv-6:** Defesa TCC 1
- **Atv-7:** Implementar algoritmos de recomendação
- **Atv-8:** Aplicar e avaliar o desempenho do algoritmo de recomendação desenvolvido
- **Atv-9:** Defesa TCC 2

| | Atv-1 | Atv-2 | Atv-3 | Atv-4 | Atv-5 | Atv-6 | Atv-7 | Atv-8 | Atv-9 |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Março | X | X | | | | | | | |
| Abril | X | X | X | | | | | | |
| Maio | | X | X | X | | | | | |
| Junho | | X | | X | X | | | | |
| Julho | | | | | X | X | | | |
| Agosto | | | | | | | X | | |
| Setembro | | X | | | | | X | X | |
| Outubro | | X | | | | | | X | |
| Novembro | | X | | | | | | | X |

Tabela 4.1: Cronograma de atividades

Conclusão

5.1 Propostas de Continuidade

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/.
- 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.
- 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.
- DIEGO NOGARI. *Performance de Machine Learning – Matriz de Confusão*. 2020. <https://diegonogare.net/2020/04/performance-de-machine-learning-matriz-de-confusao/>.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.